MG-Local: A Multivariable Control Framework for Optimal Wireless Resource Management

Chen Liu
Department of Computing Science
University of Alberta
Edmonton, Alberta, Canada
Email: cliu2@cs.ualberta.ca

M. H. MacGregor
Department of Computing Science
University of Alberta
Edmonton, Alberta, Canada
Email: macg@cs.ualberta.ca

Janelle Harms
Department of Computing Science
University of Alberta
Edmonton, Alberta, Canada
Email: harms@cs.ualberta.ca

***Abstract*—Competition for finite resources causes severe congestion and collisions in wireless networks. Without effective management, the network can become unstable, and users may experience very long delay, significant packet loss and poor throughput. In this paper, we propose a multivariable globalized-local (MG-Local) framework of resource management to find a balance between fair allocation and efficient utilization. This framework uses adaptive multivariable control to improve control effectiveness. Our design combines the advantages of both global and local optimization methods, and drives the system toward a global optimum by intelligently exploiting local information, without message passing. We demonstrate the effectiveness of this generic resource-management framework by applying it at the medium access control layer, which is the major performance bottleneck in wireless network [1]. Our experimental results show that our method significantly outperforms four other approaches in terms of throughput, packet loss rate, delay, and fairness.**

I. Introduction

Uncontrolled competition by multiple users can significantly degrade performance in wireless networks due to severe congestion and packet loss. To fill the gap between achieving effective resource management and requiring very low control-overhead, we propose a multivariable globalized-local (MG-Local) framework of resource management to find an optimal trade-off between resource allocation and utilization, without message passing. Important contributions of this work are:

- **Adaptive multivariable control**: We propose a multivariable mechanism to control system behavior more accurately than single variable control. In addition, our scheme is adaptive, and therefore able to handle the dynamic conditions that occur in wireless networks.
- **Optimal allocation-utilization trade-off**: We formulate a resource-management problem to jointly maximize resource utilization and minimize resource waste. Ideally, any allocated resource should be fully utilized. However, there is often a gap between resource allocation and utilization due to the lossy nature of wireless transmissions, inefficient protocol design, and the time-varying quality of wireless channels. Our method minimizes this allocation-utilization gap.
- **Configurable fairness criterion**: Our framework supports different fairness criteria such as proportional, max-min, equal-share or QoS-based fairness. Varying circumstances or objectives may require different criteria.
- **Zero control-message passing**: Our framework achieves performance competitive with distributed global optimization, but without message passing. This is desirable because message passing in wireless networks is unreliable and consumes resources.

MG-Local is a generic framework and can be applied to solve various problems. In this paper, we apply it at the medium-access-control (MAC) layer to manage network bandwidth to alleviate the major performance bottleneck in wireless networks. In Section II, we compare our method with existing work. Section III presents our MG-Local framework of resource management. Experiments and performance evaluation are presented in Section IV. Finally, we conclude the paper and outline future work in Section V.

II. Related Work

Optimal resource allocation in wireless networks is challenging. We categorize current methods as global optimization versus local tuning, and as deterministic versus stochastic optimization.

A. Global Optimization vs. Local Tuning

The classic method of network utility maximization belongs to this category, and has been widely applied [2][3]. The basic form of network utility maximization is given by Eq. 1. Here, U is a utility function that can be used to express user satisfaction, and to incorporate fairness. This maximization problem is also subject to a number of constraints. Although it can be solved exactly in a centralized manner, the amount of information that must be exchanged is impractical for wireless networks. Therefore, many efforts have been made to find efficient distributed solutions. The Lagrangian method transforms a constrained problem into an unconstrained one. The dual algorithm helps to separate the user and system aspects of resource allocation, and connects them via a Lagrangian multiplier, also called a shadow price. Users would like to optimize the trade-off between their utility and cost, while the system aims to maximize its revenue. Techniques such as log transformation [4] and cooperative game theory [5][6] help to decouple competing users into independent entities, and allow them to make control decisions locally. Although the required information passing is reduced to a smaller group

978-0-9836283-0-9 © 2011 ITC

This paper was peer reviewed by subject matter experts for publication in the Proceedings of ITC 2011

of competing users, message passing is still required for decision making. Depending on the size of the network, and the frequency of control, message passing may incur high overhead and have reliability issues. Although non-cooperative game theory can also be applied to obtain a Nash equilibrium without any communication between users [5][7], network-wide requirements such as fairness may be compromised due to users' selfish behavior.

$$\max \sum U_s(x_s),\ s.t.\ Rx \le C \tag{1}$$

Local tuning, in contrast, makes decisions based on local information. We further categorize local tuning methods as either pre-determined or dynamic.

Pre-determined methods control user behavior according to a fixed rule. For example, linear increase / linear decrease (LILD) [8] helps to reduce collisions when traffic is light. Multiplicative increase / linear decrease (MILD) [9] enables quick response to collisions when traffic is heavy. Exponential increase / exponential decrease (EIED) [10] achieves further improvement by allowing faster response to collisions and faster recovery from collisions. However, these approaches are only effective in a limited number of scenarios and cannot be applied generally [11].

Dynamic methods control user behavior by using feedback. For example, Deng et al. [12] tune the carrier sensing range to improve network throughput. Ma et al. differentiate causes of various collision scenarios, and tune corresponding control variables [13]. Although dynamic control is more adaptive than pre-determined approaches, local users may act in either a greedy or conservative manner depending on the aggressiveness of the control policy.

MG-Local resource management differs from global optimization and local tuning in three ways. First, MG-Local is formulated as a maximization of local utility to achieve a global optimum. The requirement of fairness, the impact of competition, and the existence of resource waste are explicitly incorporated into the formulation. Second, MG-Local does not restrict how fairness is defined, and supports various criteria. Third, besides fair allocation, MG-Local considers efficient resource utilization. Instead of being limited by the underlying scheduling scheme, it guides scheduling to reduce the allocation-utilization gap. In summary, MG-Local has the advantages of both global optimization and local tuning.

B. Deterministic vs. Stochastic Optimization

Deterministic optimization is based on steepest descent methods. Line search [14], gradient descent [14], and Newton's methods [14] can all be used. The major advantage of these methods is fast convergence when the underlying model is known. However, real systems are often complicated and dynamic in nature. It may not be possible to characterize dynamic features of complicated systems via a fixed model.

Stochastic optimization, in contrast, can be used when explicit system models are unavailable. Techniques such as the finite-difference stochastic approximation (FDSA) [15], and simultaneous perturbation stochastic approximation (SPSA) [16] can be used to estimate the gradient of each variable based on measured system responses. Although both FDSA and SPSA use perturbation to approximate gradients, SPSA significantly reduces the number of samples required to compute an approximated gradient [16]. However, SPSA can be slow to converge depending on the choice of step size, and other parameters [16].

The adaptive multivariable control component of MG-Local differs from both deterministic and stochastic optimization. We use a simple first-order regression model to capture the impact of selected variables as well as their interactions, and periodically update this model to reflect the dynamic nature of system behavior. Each variable is adjusted based on the gradient obtained from this adaptive model. Therefore, our method is more adaptive than deterministic optimization, and converges more quickly than stochastic techniques.

III. MG-Local Resource Management

MG-Local is composed of three components: adaptive multivariable control, G-Local optimization, and a configurable fairness interface. Effective resource management can be achieved by supporting fairness criteria specified via the configurable interface, driving the system toward an optimal trade-off between resource allocation and utilization via the G-Local optimization, and minimizing the allocation-utilization gap via adaptive multivariable control. This section presents the design of these components via the example of applying MG-Local at the MAC layer.

A. Adaptive Multivariable Control

The adaptive multivariable control of MG-Local is designed to achieve effective control of resource consumption, in order to reduce the allocation-utilization gap. In this work, resource consumption and waste are specified as the amount of bandwidth used for successful data transmissions (denoted x_i), and the amount of bandwidth wasted on packet collisions (denoted $coll_i$). Bandwidth consumption and waste are influenced by more than one factor. Therefore, we model x_i and $coll_i$ as functions of three variables that are carefully selected to handle different interference scenarios. This section presents our method in two steps: selecting effective control variables, and adaptively modeling system behavior. The method to update the control-variables is given in Section III-C.

1) Control-variable selection: To handle different interference scenarios, we introduce three control variables: transmission probability (p_i), collision-avoidance window ($Awin_i$), and collision-resolution window ($Rwin_i$).

Transmission probability p controls the probability that a node transmits when physical carrier sensing detects a busy medium. p is manipulated to reduce the occurrence of both hidden and exposed terminals. These two scenarios co-exist, and appear dynamically in wireless networks. It is very challenging to achieve accurate detection of both hidden and exposed terminals. For example, physical carrier sensing can effectively detect hidden terminals by decreasing its sensing threshold. However, at the same time, this increases the occurrence of exposed terminals. We overcome this problem by allowing node i to transmit with probability p_i when physical carrier sensing detects a busy medium. Depending on the severity of the exposed terminal problem, p_i can be adjusted accordingly. The transmission probability used by our method is different from the persistence probability of p-persistent CSMA, which is designed to reduce hidden terminals, but does not address exposed terminals. This is because after carrier sensing detects an idle medium, p-persistent CSMA allows a node to further backoff, instead of transmitting. In comparison, although request-to-send/clear-to-send (RTS/CTS) is designed to alleviate hidden and exposed terminals, the exchange of RTS/CTS control messages causes more complicated interference scenarios and further degrades network performance [17].

Contention avoidance window (**$Awin$**) specifies the maximum number of slots that a node can randomly select to wait before starting its transmission. We use $Awin$ to avoid potential collisions caused by simultaneous transmissions. It is similar to the contention window $cwin_i$ of CSMA, except that $Awin_i$ is only used when a packet is transmitted for the first time (the collision avoidance phase). During this phase, $Awin_i$ should be set to allow potentially simultaneous transmissions to start at different times, as well as to avoid unnecessary waiting time. In contrast, CSMA/CA sets $cwin_i$ to its minimum value $CWINMIN$ for collision avoidance, and exponentially increases $cwin_i$ upon each collision until either $cwin_i$ is greater than or equal to its maximum $CWINMAX$, or the retransmission limit is reached. Using the single variable $cwin_i$ for both collision avoidance and resolution forces an unnecessary trade-off between collisions and delay. For example, if $cwin_i$ is set small because of the limited number of simultaneous transmissions, it may not be increased quickly enough to avoid repeated collisions. Conversely, if $cwin_i$ is set large to reduce the number of retransmissions, unnecessary waiting time may be incurred for collision avoidance.

Contention resolution window (**$Rwin$**) is the contention window used to avoid repeated collisions of a packet. Besides separating collision resolution from avoidance, this variable also controls collisions caused by future transmissions. Although predicting interference from future transmissions is difficult, adjusting $Rwin_i$ helps to prevent repeated collisions of the same packet.

2) Adaptive multivariable modeling: So far, we have described the three variables that we use to control resource consumption (x_i) and waste ($coll_i$). Now we show how to model x_i and $coll_i$ as functions of these three variables. Our method considers both the accuracy of modeling, and the complexity of the computation.

a) First-order regression model: To reduce computational complexity, x_i and $coll_i$ are modeled via the first-order approximation given by Eq. 2 and 3. Although p_i, $Awin_i$ and $Rwin_i$ have different impacts on various interference scenarios, their impact is also correlated. Our model considers the independent impact of each variable, as well as their interactions. We intentionally ignore the impact from other nodes to avoid message passing. From our observation and performance evaluation, this model gives reasonable estimations of the gradient of each control variable. In the future, we will further improve model accuracy by applying the smoothed perturbation analysis (SPA) [18].

$$\begin{aligned} x_i &= R(p_i, Awin_i, Rwin_i) \\ &= e_1 \cdot p_i + e_2 \cdot Awin_i + e_3 \cdot Rwin_i \\ &+ e_4 \cdot p_i \cdot Awin_i + e_5 \cdot p_i \cdot Rwin_i \\ &+ e_6 \cdot Awin_i \cdot Rwin_i \ + e_7; \end{aligned} \tag{2}$$

$$\begin{aligned} coll_i &= CL(p_i, Awin_i, Rwin_i) \\ &= f_1 \cdot p_i + f_2 \cdot Awin_i + f_3 \cdot Rwin_i \\ &+ f_4 \cdot p_i \cdot Awin_i + f_5 \cdot p_i \cdot Rwin_i \\ &+ f_6 \cdot Awin_i \cdot Rwin_i \ + f_7; \end{aligned} \tag{3}$$

The coefficients of Eq. 2 and Eq. 3 are computed from a least-squares fit [19].

b) Periodic updating and noise processing: Eq. 2 and Eq. 3 are periodically updated by replacing old measurements of x_i and $coll_i$with fresh ones. We denote these measurements as x_i^m and $coll_i^m$. These periodic updates help the method to capture nonlinear system behavior, and adapt to dynamic system conditions. We also apply an exponential filter to alleviate the negative impact of noise in the measurements [20]. According to Eq. 4, a new measurement of x_i^m at time t (i.e. $x_i^m(t)$) is linearly combined with the previous measurement at time $t-1$ (i.e. $x_i^m(t-1)$). w is set to be 0.6 (an empirical value used in practice). Similarly, $coll_i^m$ is updated according to Eq. 5.

$$x_i^m(t) = w \cdot x_i^m(t) + (1-w) \cdot x_i^m(t-1); \tag{4}$$

$$coll_i^m(t) = w \cdot coll_i^m(t) + (1-w) \cdot coll_i^m(t-1); \tag{5}$$

B. G-Local Optimization

The second component of MG-Local is the G-Local optimization that we proposed previously [21], in which a single-variable control method is applied. This optimization aims to minimize the allocation-utilization gap, i.e. the amount of a resource allocated but not utilized. The gap between allocation and utilization is caused by the lossy nature of wireless transmissions, time-varying quality of wireless channels, and imperfect scheduling. We formulate G-Local optimization in Eq.6. It can be viewed as a joint solution for two sub-problems: 1) maximizing utilization, and 2) minimizing waste.

Maximizing resource utilization helps to reduce the gap by increasing a user's resource consumption, x_i, until the amount of resource utilized reaches a desired fair allocation. However, increasing consumption aggressively may cause more resource waste due to more collisions. Minimizing resource waste also helps to reduce the gap by decreasing x_i to avoid conflicts. However, over-conservative consumption may cause more resources to be wasted due to unnecessary idling. The G-Local optimization given by Eq. 6 balances these two forces, and finds an optimal trade-off between fair allocation and efficient utilization. In this equation, x_i is bounded by x_{min} (e.g. 0 pps), and x_{max} (e.g. the offered load).

$$\begin{aligned} &\max\ k \cdot U_i(x_i) - (1-k) \cdot (C_i(x_i) + W_i(x_i)) \\ &s.t.\ x_{min} < x_i < x_{max} \end{aligned} \tag{6}$$

G-Local optimization consists of three elements:

1) $U_i(x_i)$ represents the benefit or satisfaction of a local user i when its resource consumption is x_i. Currently, $U_i(x_i) = \log(x_i)$ ensures that user satisfaction increases with x_i. With this choice, the marginal benefit decreases as x_i increases.
2) $C_i(x_i)$ serves as a consumption cost, and plays two important roles. First, it controls the greedy behavior of a local user, by imposing a limit on the resource consumption x_i. Thus, the more a user consumes, the higher the cost it has to pay. Second, $C_i(x_i)$ drives x_i toward a desired fair allocation $\hat{x_i}$, in the form of $C_i(x_i) = p_{c,i} \cdot x_i$ that we previously derived [21]. $p_{c,i} = \frac{1}{\hat{x_i}}$ is the price of consuming one unit of bandwidth.
3) $W_i(x_i)$ is the cost of resource waste. We categorize resource waste as non-conflict-caused waste, and conflict-caused waste. Their corresponding costs are denoted $I_i(x_i)$ and $F_i(x_i)$. Non-conflict-caused waste is defined as the amount of resource that should be utilized, but is not. For example, some transmission opportunities can be wasted due to conservative scheduling. Its specific form is given in Eq.7. Conflict-caused waste is the fraction of channel bandwidth that is wasted on collisions. It is specified in Eq. 8. Considering both $F_i(x_i)$ and $I_i(x_i)$, we note that $W_i(x_i) = F_i(x_i) + I_i(x_i)$. $W_i(x_i)$ can be interpreted as the fraction of the channel that should have been effectively utilized by link i.

$$I_i(x_i) = \frac{\hat{x_i} - x_i}{B} \tag{7}$$

$$F_i(x_i) = \frac{coll_i}{B} \tag{8}$$

Furthermore, G-Local optimization can support different fairness criteria, by allowing corresponding functions to be interfaced via $\hat{x_i}$. In this paper, an equal-share fairness criteria with regard to interfering links is applied for proof-of-concept. According to this criteria, each link shares the channel bandwidth equally with interfering links. As a first step, we assume a static network topology; we will consider mobile scenarios in future work. We also assume saturated networks, where fairness is most desired. The fair share $r_{c,i} = \frac{B}{n_i^{share}}$ can be derived during the system startup procedure. n_i^{share} is the number of links that use the same wireless channel as link i, including link i itself. B is the channel bandwidth in units of packets per second, pps. To fulfill this allocation, $r_{c,i}$ is fed into the G-Local optimization via $\hat{x_i}$. Accordingly, $p_{c,i} = \frac{1}{r_{c,i}} = \frac{n_i^{share}}{B}$ is the consumption price. This price also has a valid physical meaning under this specific fairness criteria. The cost of sending 1 packet in 1 second is the actual transmission time of 1 packet, $\frac{1}{B}$, multiplied by the number of users sharing that channel. A more specific form of the optimization is given in Eq.9. It is derived by substituting $U_i(x_i)$, $C_i(x_i)$ and $W_i(x_i)$ with $\log(x_i)$, $\frac{1}{n_i^{share}} \cdot x_i$ and $\frac{1}{n_i^{share}} - \frac{x_i}{B} + \frac{coll_i}{B}$ respectively.

$$V_i(x_i) = k \cdot \log x_i - (1-k) \cdot (\frac{n_i^{share} - 1}{B} \cdot x_i + \frac{coll_i(x_i)}{B} + \frac{1}{n_i^{share}}) \tag{9}$$

Because $\log x_i$ is a concave function, and $\frac{n_i^{share}-1}{B} \cdot x_i + \frac{coll_i(x_i)}{B} + \frac{1}{n_i^{share}}$ is a non-decreasing function, Eq.9 is a strictly concave function, according to Lemma 3.2 in [22]. However, whether it has a unique optimal point depends on two conditions: $V_i(x_i) \to -\infty$ as $x_i \to 0$, and $V_i(x_i) \to -\infty$ as $x_i \to \infty$. This requirement has been accounted for in our formulation by the introduction of k, which determines whether these two conditions are satisfied. We dynamically set k as $1 - \frac{x_i \cdot B}{n_i^{share}}$ so that k decreases, when x_i increases.

C. MG-Local Control Policies

So far, we have introduced the optimization and control mechanisms of MG-Local. This section describes how these two mechanisms are integrated to achieve effective resource management. First, we transform the G-Local optimization Eq. 9 into the multivariable version Eq. 10, by substituting x_i and $coll_i$ with Eq. 2 and Eq. 3.

$$\begin{aligned} V(p_i, Awin_i, Rwin_i) &= k \cdot \log R(p_i, Awin_i, Rwin_i) \\ &- (1-k) \cdot (\frac{n_i^{share} - 1}{B} \cdot R(p_i, Awin_i, Rwin_i) \\ &+ \frac{CL(p_i, Awin_i, Rwin_i)}{B} + \frac{1}{n_i^{share}}) \end{aligned} \tag{10}$$

$$\begin{aligned} &\max V(p_i, Awin_i, Rwin_i) \\ &\quad s.t.\ 0 < p_i \leq 1; \\ &\qquad CAMIN \leq Awin_i \leq CAMAX; \\ &\qquad CRMIN \leq Rwin_i \leq CRMAX; \end{aligned} \tag{11}$$

Second, by applying the Lagrangian transformation, the constrained optimization Eq. 10 is converted to the unconstrained problem Eq.12, where λ_i^{p1}, λ_i^{p2}, λ_i^{ca1}, λ_i^{ca2}, λ_i^{cr1} and λ_i^{cr2} are the shadow prices. We denote these shadow prices as $\{\lambda_i\}$. The corresponding dual problem is given in Eq. 15.

$$\begin{aligned} &L(p_i, Awin_i, Rwin_i, \{\lambda_i\}) \\ &= V_i(p_i, Awin_i, Rwin_i) + \lambda_i^{p1} \cdot p_i - \lambda_i^{p2} \cdot (p_i - 1) \\ &- \lambda_i^{ca1} \cdot (CAMIN - Awin_i) - \lambda_i^{ca2} \cdot (Awin_i - CAMAX) \\ &- \lambda_i^{cr1} \cdot (CRMIN - Rwin_i) - \lambda_i^{cr2} \cdot (Rwin_i - CRMAX) \end{aligned} \tag{12}$$

$$\max L(p_i, Awin_i, Rwin_i, \{\lambda_i\}) \tag{13}$$

$$D(p_i, Awin_i, Rwin_i, \{\lambda_i\}) = \max \ L \tag{14}$$

$$\min \ D(p_i, Awin_i, Rwin_i, \{\lambda_i\}) \tag{15}$$

Third, we solve Eq. 13 by applying the method of gradient descent. The control policies for p_i, $Awin_i$ and $Rwin_i$ are given in Eq.16, Eq.17 and Eq.18. Each user applies these control policies periodically, until the algorithm converges.

$$p_i(t) = p_i(t-1) + k_p \cdot \frac{\partial L}{\partial p_i} \tag{16}$$

$$Awin_i(t) = Awin_i(t-1) + k_{ca} \cdot \frac{\partial L}{\partial Awin_i} \tag{17}$$

$$Rwin_i(t) = Rwin_i(t-1) + k_{cr} \cdot \frac{\partial L}{\partial Rwin_i} \tag{18}$$

Lastly, we solve the dual problem (Eq.15) to obtain the shadow prices. The solutions are $\lambda_i^{p1} = -h_1 \cdot {p_i}^+$, $\lambda_i^{p2} = h_2 \cdot (p_i - 1)^+$, $\lambda_i^{ca1} = h_3 \cdot (CAMIN - Awin_i)^+$, $\lambda_i^{ca2} = h_2 \cdot (Awin_i - CAMAX)^+$, $\lambda_i^{cr1} = h_4 \cdot (CRMIN - p_i)^+$ and $\lambda_i^{cr2} = h_5 \cdot (Rwin_i - CRMAX)^+$. The variables $\{h\}$ are scalars, and their values are determined experimentally.

IV. Experimental Results

In this section, we evaluate the performance of MG-Local via simulation in ns2. Experiments are designed to demonstrate the effectiveness of resource management in terms of fair allocation and efficient utilization, using four metrics. We measure the fairness of resource allocation via Jain's fairness index. This is calculated as $f(x) = \frac{(\sum x_i)^2}{n \cdot \sum x_i^2}$, and ranks the fairness of a resource allocation strategy between 0 and 1. Higher values indicate fairer allocation. We measure the efficiency of resource utilization using throughput and packet loss rate. Throughput is the total rate at which packets are successfully delivered to destinations in the network. Packet loss rate is the rate at which packets are lost due to collisions. We measure the trade-off made by our control method via end-to-end delay. End-to-end delay is the average time from packet generation to successful reception, for packets that are not lost due to collision. We compare the performance of MG-Local to four alternatives:

1) A stochastic estimation-based method of multivariable control, denoted SP in the figures. The SP estimation-based method obtains gradients from the simultaneous perturbation stochastic optimization proposed by Spall [16]. With this method, all control variables are simultaneously perturbed twice in a random manner during each control iteration. This yields two measured responses, the difference between them (denoted δy), and the distance between the two perturbations of each variable(denoted $\delta\theta$). The gradient of each variable is estimated as $\frac{\delta y}{\delta \theta}$ [16].
2) Single-variable control, denoted SVC in the figures [21]. This single-variable control scheme adjusts the transmission probability p_i, and requires no message passing.
3) Chiang et al.'s distributed global-optimization, denoted GB in the figures. Their method also uses single variable control. However, the global optimum is achieved at the expense of message passing [4].
4) CSMA/CA is used as a reference to examine whether the selected control variables behave as planned. Specifically, the separation of avoidance window ($Awin_i$) and resolution window ($Rwin_i$) is compared with the binary-exponential-backoff mechanism of CSMA in terms of collision and delay. We further differentiate two CSMA/CA configurations. The first reduces hidden terminals by setting the physical carrier sensing range to twice the transmission range, and is denoted CSMA/CA-HT in Table I. The second reduces exposed terminals by setting the carrier sensing range the same as transmission range. This method is denoted CSMA/CA-ET. We do not compare to CSMA with RTS/CTS, because it incurs higher overhead and the extra control packets create more complicated interference scenarios. Our previous work [23] and other work [17] have shown that CSMA without RTS/CTS can perform better.

We carried out two sets of experiments in different network topologies. The first topology is a 3-link network, and is designed to compare our adaptive multivariable control with a globally optimal allocation. The second set of experiments uses larger, randomly-generated networks to evaluate our adaptive multivariable control. We also consider the impact of traffic, and interference, by varying the offered load and number of source-destination pairs in each network. We replicate each experiment 10 times, and show the average performance and 95% confidence intervals. System parameters are categorized into fixed, changing, and control variables.

Fixed parameters are: a packet size of 512 bytes, transmit power of 0.2818 Watt, transmission range of 200 meters, physical carrier sensing range of 400 meters, and running time of 30 minutes. Changing parameters are offered load and interference level. Control variables are the transmission probability p_i, collision avoidance window $Awin_i$ and resolution window $Rwin_i$.

A. Three-Link Networks

With the 3-link topology shown in Fig 1, we can easily compute the globally-optimal and proportionally-fair allocation, which maximizes network utility for the ideal case, as formulated in Eq. 19. This formulation aims to maximize the total transmission rate of all three links in a proportionally fair manner, and is subject to a set of constraints. These constraints reflect the contention relationships between links 1, 2 and 3. Link 2 interferes with link 1 and 3, and, conversely, link 1 and link 3 interfere with link 2. The first constraint limits the total data rate of link 1 and link 2 to the maximum bandwidth of the channel they share. Similarly, the second constraint limits the sum rate of link 2 and 3. This problem can be easily solved and the optimum allocations for link 1, 2, and 3 are $\frac{B}{2}$, $\frac{B}{3}$

and $\frac{B}{2}$. We consider this ID result as the upper bound that all methods aim to achieve.

$$\begin{aligned} \max \ & \log(x_1) + \log(x_2) + \log(x_3) \\ s.t. \ & x_1 + x_2 \leq B; \\ & x_2 + x_3 \leq B; \end{aligned} \tag{19}$$

1) Fixed offered load: The first experiment fixes the offered load for each transmission pair. The selected offered load is 200 pps, and saturates the whole network. That is, the total offered load on each link exceeds its capacity. In this experiment, the source node of each link generates data at 200 pps, and the capacity of each link is 244 pps. The total offered loads that are imposed on link 1, 2, and 3 are 400 pps, 600 pps, and 400 pps, respectively. Our comparative results for the per-link allocation, and aggregate network performance are shown in Table I, and Figure 2, respectively.

First, our adaptive multivariable control (AM in Fig. 2) achieves a close-to-optimal allocation, compared to the ideal case. As shown in Table I, the ideal allocation for link 1, 2, and 3 is approximately 122 pps, 81 pps and 122 pps. Our method achieves 116 pps for link 1, 61 for link 2, and 114 pps for link 3. Fig. 2 shows that the fairness index of our adaptive multivariable control is very close to the ideal case. In addition, compared to the other four alternatives, our adaptive control utilizes resources more efficiently. This is demonstrated in Fig 2, where the gap between our method and the ideal case is the smallest, for both aggregate throughput and packet loss rate .

Second, compared to the estimation-based multivariable control (SP in Fig. 2), our adaptive multivariable control achieves a fairer allocation and higher utilization of resources. As shown in Table I, the per-link allocation of the estimation-based control is 120 pps for link 1, 25 pps for link 2, and 121 pps for link 3. Its low throughput on link 2 has a negative impact on both aggregate throughput and fairness (see Fig. 2). The estimation-based control has an inferior performance to our method because simultaneous perturbation can be slow to converge and can get stuck at a local maximum.

Third, compared to the single-control method (SVC in Fig. 2), our adaptive multivariable control achieves a similar level of fairness (as reflected by the Jain's index in Fig 2), and a much higher aggregate throughput. Moreover, both the single-variable control, and the estimation-based control have a higher packet loss rate than our method. The low resource utilization of single-variable control shows that it has a limited ability to deal with the complicated network behavior, and demonstrates the need for multivariable control.

Fourth, Chiang's distributed global-optimization (GB in Fig. 2) achieves similar results to the single-variable control (SVC in the figure), but at the cost of message passing. It is also a single-variable control, and results in lower throughput and higher packet loss than our adaptive multivariable control.

Fifth, CSMA/CA-ET aims to prevent node 1 and 3 from becoming exposed terminals. However, due to the inferior location of link 2, and the well-known unfair nature of CSMA [24], the excellent performance of link 1 and 3 comes at the expense of starving link 2. In contrast, CSMA/CA-ET sets the carrier sensing range to twice the transmission range to avoid hidden terminals. Although the fairness is improved over CSMA/CA-HT, nodes 1 and 2 waste resources because they are exposed terminals. In comparison, our adaptive multivariable control succeeds in handling both hidden and exposed terminals.

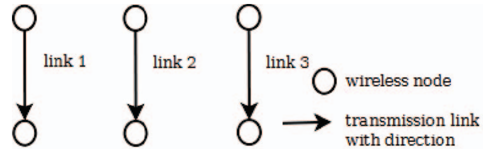

Fig. 1. 3-Link Network Topology

TABLE I
PER-LINK ALLOCATION

	link 1(pps)	link 2(pps)	link 3(pps)
ID	122.07	81.38	122.07
AM	116.32	61.68	114.54
SP	120.71	25.41	121.02
SVC	67.23	44.45	67.35
GB	68.93	42.65	69.21
CSMA/CA-ET	182.10	1.02	182.09
CSMA/CA-HT	65.37	55.55	65.30

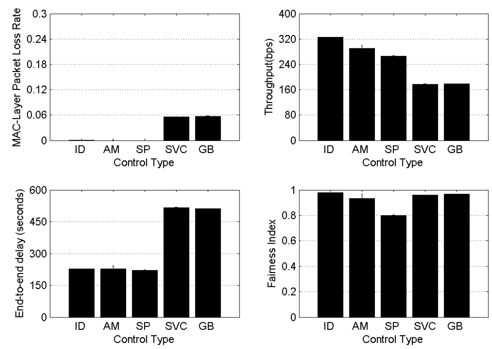

Fig. 2. Aggregate Performance Comparison

In summary, for the 3-link case, MG-Local outperforms the selected alternatives, and achieves effective resource management, improving both allocation fairness and utilization efficiency. It does this without message-passing.

2) Varying offered loads: In the previous experiment, we used a fixed offered load of 200 pps. We now examine results from using different levels of traffic load, from light (50 pps) and unsaturated (100 pps), to saturated (200 pps), and over-saturated (400 pps). The result of aggregate performance is shown in Fig.3. These results show that our adaptive multivariable control maintains its performance advantage at different levels of offered load. Our previous observation (with an offered load of 200 pps) in Section IV-A1 holds, even when the network is over-saturated (with an offered load of 400 pps).

When the traffic load is low (50 pps for each link), all methods achieve similar results because user demands for resources can be satisfied easily. However, at higher loads,

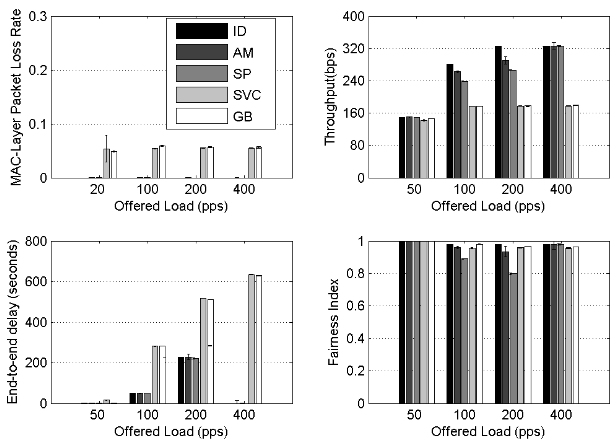

Fig. 3. Impact of Offered Load (3-Link Network)

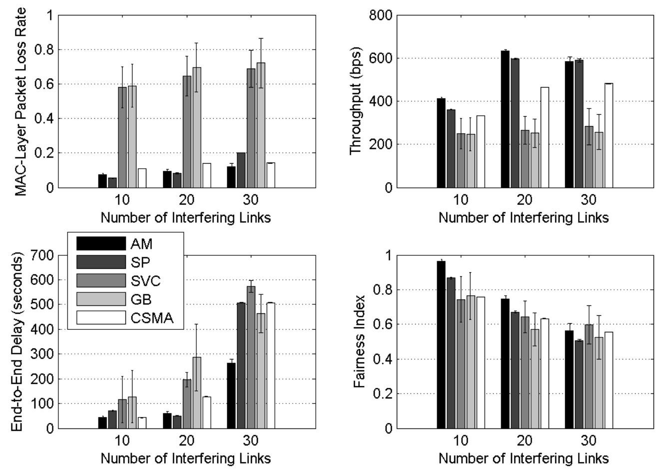

Fig. 4. Impact of Interference Level (Random 1-Hop Network)

the advantage of our adaptive multivariable control becomes obvious. Our method achieves lower packet loss rate and end-to-end delay, and higher aggregate throughput and fairness. This trend holds even when the offered load increases to over-saturate the whole network.

B. Large Random Networks

The second set of experiments is designed to evaluate our adaptive multivariable control in larger, randomly generated single-hop networks. We create different network topologies by deploying 200 nodes in a 800-by-800 m^2 field and selecting source-destination pairs randomly.

1) Varying interference level: The first experiment tests the influence of interference levels. To generate different interference levels, we vary the network size by creating 10, 20 and 30 source-destination pairs (also called links in this paper). Each experiment is replicated 10 times, and each replication uses a distinct topology by selecting the required number of links randomly. All other parameters are fixed, including the offered load (50 pps) and packet size (512 bytes). Fig. 4 shows the average performance and 95% confidence intervals for each control algorithm, over all replications.

Fig. 4 shows that the adaptive multivariable control (AM in the figure) achieves consistent performance, as it does in the three-link experiment. At each interference level, our method achieves the best trade-off between throughput and fairness, without compromising delay. Even when the interference level is the highest, at 30 links, the adaptive multivariable control yields much higher throughput and shorter delay than the other candidates, while achieving a similar fairness index to the others. In comparison with SP, our method achieves fast convergence.

Our adaptive multivariable control is a significant improvement over CSMA, due to the transmission probability p_i alleviating the exposed-terminal problem. We attribute the reduction of packet loss rate to the separate adjustment of the avoidance and resolution windows $Awin_i$ and $Rwin_i$. Also, replacing the binary exponential backoff with $Rwin_i$ helps to reduce the end-to-end delay. In contrast, the single-variable control (SVC in the figure) and Chiang's distributed global-optimization (GB in the figure) only adjust the transmission probability. Although exposed terminals can be alleviated, without appropriate adjustment of other factors (e.g. the contention window), these methods increase the occurrence of hidden terminals. As a result, they suffer from significant packet loss.

2) Varying offered load: The second experiment tests performance at offered loads of 50 pps, 100 pps, and 200 pps. The results show that our adaptive multivariable control again outperforms other alternatives. It achieves more efficient utilization, and fairer allocation (see Fig 5).

With an increasing offered load, the gap between the throughput of adaptive multivariable control (AM in the figure) and that of other candidates is more distinct, without trading off delay. In terms of fairness, the worst fairness index achieved by our method is as good as the single-variable control (SVC in the figure) and Chiang's distributed global-optimization (GB in the figure). MG-Local achieves higher throughput, lower packet loss rate, and shorter delay.

C. MG-Local Implementation and Application

In this paper, we apply MG-Local to IEEE 802.11 b, because its conventional antenna allows single packet per reception and this restriction poses more challenges on resource management. The implementation only requires each node to periodically measure the average data sending rate (x_i^m) and data loss rate ($coll_i^m$). These values are used to derive Eq. 2 and 3. Persistence probability p_i, $awin_i$ and $rwin_i$ are updated according to Eq. 16, 17 and 18 for each node.

MG-Local can also support smart antenna technologies. Although the advance of multiple-packet reception helps to alleviate the negative impact of interference, unfair resource allocation and inefficient resource utilization still exists if user competition for finite resources is not well managed. MG-Local can be applied based on a new analysis of the

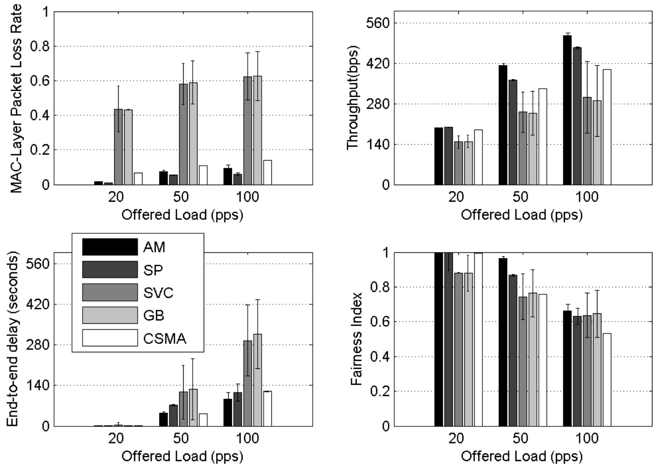

Fig. 5. Impact of Offered Load (Random 1-Hop Network)

causes of consumption conflicts and resource waste. Besides the physical and MAC layer, MG-Local can also be applied at other network layers to optimize protocol performance, to guide cross-layer design; to improve quality of service in terms of guaranteeing minimum bandwidth guarantee, congestion and admission control, as well as reliable transmission. Due to limited space, we cannot give detailed description.

V. Conclusions

We propose a novel MG-Local control framework to achieve effective resource management in wireless networks. This framework is composed of three elements: 1) an adaptive multivariable control, 2) a global-local optimization, and 3) configurable fairness. The adaptive multivariable control improves CSMA/CA's capability to handle both hidden and exposed terminals. Our optimization method finds an optimal trade-off between fair allocation and efficient utilization. By exploiting local information, the optimization limits the selfish behavior of local users, and drives the system toward the optimum without message passing. This overall framework of control plus optimization can support different fairness criteria. We compare our method with four other approaches. Our experiment results show that MG-Local significantly outperforms all four alternatives. In future work, we will extend our method in three directions. First, we will apply this framework to other types of wireless resources, including energy, frequency, etc. Second, we will combine it with different fairness criteria, to further demonstrate the effectiveness of the MG-Local framework. Third, we will consider mobile nodes and multi-hop forwarding.

References

[1] X. Lin, N. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1452–1463, 2006.

[2] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *The Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.

[3] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition: a mathematical theory of network architectures," in *Proceedings of the IEEE*, vol. 95, 2007, pp. 255–312.

[4] J.-W. Lee, M. Chiang, and A. R. Calderbank, "Optimal mac design based on utility maximization: Reverse and forward engineering," in *Proceedings of the 25th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'06)*, April 2006, pp. 1–13.

[5] Z. Han and K. J. R. Liu, *Resource Allocation for Wireless Networks: Basics, Techniques, and Applications*. Cambridge University Press, 2008.

[6] Z. Fang and B. Bensaou, "Design and implementation of a mac scheme for wireless ad-hoc networks based on a cooperative game framework," in *Proceedings of 2004 IEEE International Conference on Communications (ICC'04)*, June 2004, pp. 4034–4038.

[7] ——, "Fair bandwidth sharing algorithms based on game theory frameworks for wireless ad-hoc networks," in *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'04)*, March 2004, pp. 1284–1295.

[8] V. Vitas, "Throughput analysis of a linear bakcoff scheme in wireless LANs," *IEEE Electronic Letters*, vol. 39, no. 1, pp. 99–100, 2003.

[9] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: A media access protocol for wireless LANs," in *Proceedings of the Conference on Communications Architectures Protocols and Applications*, Octobor 1999, pp. 212–225.

[10] D. G. Jeong and W. S. Jeon, "Performance of an exponential backoff scheme for slotted-ALOHA protocol in local wireless environment," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 3, pp. 470–479, 1995.

[11] E. Kartsakli, J. Alonso-Zarate, A. Cateura, C. Verikoukis, and L. Alonso, *Contention-Based Collision-Resolution Medium Access Control Algorithms*. Hauppauge, NY, US: Nova Science Publishers Inc, 2009.

[12] J. Deng, B. Liang, and P. Varshney, "Tuning the carrier sensing range of IEEE 802.11 MAC," in *Proceedings of the 47th IEEE Global Telecommunications Conference (GLOBECOM 2004)*, December 2004, pp. 2987–2991.

[13] H. Ma, J. Zhu, S. Roy, and S. Y. Shin, "Joint transmit power and physical carrier sensing adaptation based on loss differentiation for high density IEEE 802.11 WLAN," *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 52, no. 9, pp. 1703–1720, 2008.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[15] J. C. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *The Annals of Mathematical Statistics*, vol. 23, pp. 462–466, 1952.

[16] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. on Automatic Control*, vol. 37, pp. 332–347, 1992.

[17] J. L. Sobrinho, R. de Haan, and J. M. Brazio, "Why RTS-CTS is not your ideal wireless LAN multiple access protocol," in *Proceedings of 2005 IEEE Wireless Communications and Networking Conference (WCNC'05)*, March 2005, pp. 81–87.

[18] C. G. Cassandras and V. Julka, "Scheduling policies using marked/phantom slot algorithms," *Queueing Systems*, vol. 20, no. 1-2, pp. 207–254, 1995.

[19] P. Lancaster and K. Salkauskas, *Curve and Surface Fitting: An Introduction*. London, UK: Academic Press, 1986.

[20] J.-P. Corriou, *Process Control: Theory and Applications*. London, UK: Springer-Verlag, 2004.

[21] C. Liu, J. Harms, and M. H. MacGregor, "G-local resource management: Achieving global optimization via local inference without message," in *the 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON'11) (submitted)*, 2011.

[22] R. Srikant, *The Mathematics of Internet Congestion Control*. USA: Birkhuser, 2004.

[23] C. Liu, J. Harms, and M. H. MacGregor, "Optimal control of spatial, temporal and bandwidth contention in wireless ad hoc networks," in *the 5th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob'09)*, October 2009, pp. 263–269.

[24] Y. Jian and S. Chen, "Can csma/ca networks be made fair?" in *Proceedings of the 14th ACM international conference on Mobile computing and networking (MobiCom'08)*, December 2008, pp. 235–246.