Random Access in Wireless Networks: How Much Aggressiveness Can Cause Instability?

Javad Ghaderi†, Sem Borst‡, Phil Whiting‡
†University of Illinois, Urbana-Champaign, ‡Alcatel-Lucent Bell Labs, Murray Hill

ABSTRACT

Random access schemes are simple and inherently distributed, yet capable of matching the optimal throughput performance of centralized scheduling algorithms. The throughput optimality however has been established for activation rules that are relatively sluggish, and may yield excessive queues and delays. More aggressive/persistent access schemes have the potential to improve the delay performance, but it is not clear if they can offer any universal throughput optimality guarantees. In this paper, we identify a limit on the aggressiveness of nodes, beyond which instability is bound to occur in a broad class of networks.

1. INTRODUCTION

Emerging wireless networks typically lack any centralized access control entity, and instead vitally rely on the individual nodes to operate autonomously and to efficiently share the medium in a distributed fashion. This requires the nodes to schedule their individual transmissions and decide on the use of a shared medium based on knowledge that is locally available or only involves limited exchange of information. A popular mechanism for distributed medium access control is provided by the so-called Carrier-Sense Multiple-Access (CSMA) protocol, various incarnations of which are implemented in IEEE 802.11 networks. In the CSMA protocol each node attempts to access the medium after a certain back-off time, but nodes that sense activity of interfering nodes freeze their back-off timer until the medium is sensed idle.

Despite their asynchronous and distributed nature, CSMA-like algorithms have been shown to offer the capability of achieving the full capacity region and thus match the optimal throughput performance of centralized scheduling mechanisms operating in slotted time [5, 6]. Based on this observation, various ingenious algorithms have been developed for finding the back-off rates that yield a particular target throughput vector or that optimize a certain concave throughput utility function in scenarios with saturated buffers [5, 8]. In the same spirit, several effective approaches have been devised for adapting the transmission lengths based on backlog information, and been shown to guarantee maximum stability [4,11]. Here, as in the latter approach, we consider adapting transmission lengths by considering a weight for each node as a function of its backlog. The aggressiveness of a node, in not releasing the medium once it starts a transmission, is controlled through its weight.

The maximum-stability guarantees were originally established under the condition that the weights of the various nodes behave as $\log\log(\cdot)$ functions of the backlogs [11]. Unfortunately, such weights can induce excessive backlogs and delays, which has triggered a strong interest in developing approaches for improving the delay performance [7, 10, 12]. Motivated by this issue, Ghaderi & Srikant [3] showed that it is in fact sufficient for weights to behave as $\log(\cdot)$ functions of the backlogs, divided by an arbitrarily slowly increasing function. These results indicate that the maximum-stability guarantees are preserved for weights that are essentially logarithmic for all practical values of the backlogs, although asymptotically the weight must grow slower than any logarithmic function of the backlog.

In the present paper, we explore the scope for using more aggressive weight functions in order to improve the delay performance while preserving the maximum-stability guarantees. We will consider the fluid limits of the system where dynamics are scaled in both space and time. Fluid limits may be interpreted as first-order approximations of the original stochastic process, and provide valuable qualitative insight and a powerful approach for establishing (in)stability properties. As observed in [1], qualitatively different types of fluid limits can arise, depending on the structure of the interference graph, in conjunction with the functional shape of the weights. For weight functions which grow slower than $\log(\cdot)$, "fast mixing" is guaranteed in any general topology, where the activity process evolves on a much faster time scale than the scaled backlogs. Qualitatively similar fluid limits can arise for more aggressive weight functions as well, provided the topology is benign. However, aggressive weight functions can cause "sluggish mixing", where the activity process evolves on a much slower time scale than the scaled backlogs, yielding random oscillatory fluid limits. Such fluid limits can force the system into inefficient states for extended periods of time and produce instability.

Main contribution: We will demonstrate instability for weights that grow faster than $\gamma\log(\cdot)$, for any $\gamma > 1$, but our proof arguments suggest that it can occur for any $\gamma > 0$, in networks with sufficiently many nodes. In other words, "the near-logarithmic growth condition" on the weights is a fundamental limit on the aggressiveness of nodes to ensure maximum stability in any general topology.

The rigorous statement and proof of the fluid limits as well as all subsequent proofs are omitted in this extended abstract. We refer the interested reader to [2] for details.

2. MODEL DESCRIPTION

We consider a network of several nodes sharing a wireless medium according to a random-access mechanism. The network is represented by an undirected graph $G = (V, E)$ where the set of vertices $V = \{1, \ldots, N\}$ correspond to the various nodes and the set of edges $E \subseteq V \times V$ indicate which pairs of nodes interfere. Define $S \subseteq \{0, 1\}^N$ as the set of feasible joint activity states, i.e., the incidence vectors of all the independent sets of the interference graph, and denote by $\mathcal{C} = \mathrm{conv}(S)$ the capacity region.

Packets arrive at node i as a Poisson process of rate λ_i. The packet transmission times at node i are independent and exponentially distributed with mean $1/\mu_i$. Denote by $\rho_i = \lambda_i/\mu_i$ the load of node i.

Let $U(t) \in S$ represent the joint activity state of the network at time t, with $U_i(t)$ indicating whether node i is active at time t or not. Denote by $Q_i(t)$ the backlog at node i at time t, i.e., the number of packets waiting for transmission or in the process of being transmitted.

When a node ends an activity period (consisting of possibly several back-to-back packet transmissions), it starts a back-off period. The back-off times of node i are independent and exponentially distributed with mean 1. At the end of the back-off period, the node starts a new transmission only if none of its neighbors are active, otherwise it starts a new back-off. When a transmission of node i ends at time t, it releases the medium and begins a back-off period with probability $\exp(-w(Q_i(t^-)))$, or starts the next transmission otherwise.

We are interested to determine under what conditions the system is stable, i.e., the Markov process $\{(U(t), Q(t))\}_{t\geq 0}$ is positive-recurrent. It is easily seen that $(\rho_1, \ldots, \rho_N) < \sigma \in \mathcal{C}$ is a necessary condition for that to be the case. In [3], it is shown that this condition is in fact also sufficient for weight functions of the form $w(Q_i) = \log(1 + Q_i)/g_i(Q_i)$, where $g_i(Q_i)$ is allowed to increase to infinity at an arbitrarily slow rate. In particular, $w(Q_i) = \log^{1-\epsilon}(1 + Q_i)$, for any small positive ϵ, yields stability.

Results in [1] suggest that more aggressive weight functions can improve the delay performance. In view of these results, we will examine to what extent the sufficient stability conditions of [3] are tight. We will particularly consider functions that are more aggressive and analyze fluid limits of the system, as introduced in the next section.

3. FLUID LIMITS AND INSTABILITY RESULTS

In order to obtain fluid limits, the original stochastic process is scaled in both space and time. More specifically, we consider a sequence of processes $\{(U^R(t), Q^R(t))\}_{t\geq 0}$ indexed by a sequence of positive integers R, each governed by similar statistical laws as the original process, where the initial states satisfy $\sum_{i=1}^N Q_i^R(0) = R$ and $Q_i^R(0)/R \to q_i$ as $R \to \infty$. The process $\{(U^R(Rt), \frac{1}{R}Q^R(Rt))\}_{t\geq 0}$ is referred to as the fluid-scaled version of the process $\{(U^R(t), Q^R(t)\}_{t\geq 0}$. Any (possibly random) weak limit $\{q(t)\}_{t\geq 0}$ of the sequence $\{\frac{1}{R}Q^R(t)\}_{t\geq 0}$, as $R \to \infty$, is called a fluid limit.

We now proceed to demonstrate the potential for "aggressive" weight functions to cause instability. The potential for instability arises in a broad class of networks, but here we focus on a "nearly" complete 3-partite graph with two nodes in each component as shown in Figure 1. The intuitive explanation for the potential instability may be described as follows. Denote $\rho_0 = \max\{\rho_1, \rho_2\}$, and assume $\rho_3 > \rho_4$ and $\rho_5 < \rho_6$. The fraction of time that at least one of the nodes 1, 2, 3 and 6 is served, must be no less than $\rho = \rho_0 + \rho_3 + \rho_6$ in order for these nodes to be stable. During some of these periods nodes 4 or 5 may also be served, but not simultaneously, i.e., schedule M_4 cannot be used. In other words, the system cannot be stable if schedule M_4 is used for a fraction of the time larger than $1 - \rho$. As it turns out, however, when the weight function is sufficiently aggressive, e.g., $w(\cdot) = \gamma \log(\cdot)$, with $\gamma > 1$, schedule M_4 is in fact persistently used for a fraction of the time that does not tend to 0 as ρ approaches 1, which forces the system to be unstable.

Although the above arguments indicate that invoking schedule M_4 is a recipe for trouble, the reason may not be directly evident from the system dynamics, since no obvious inefficiency occurs as long as the queues of nodes 4 and 5 are non-empty. However, the fact that the Lyapunov function $L(t) = \sum_{k=1}^3 \max_{i\in M_k} q_i(t)$ may increase while serving nodes 4 and 5, when $q_3(t) \geq q_4(t)$ and $q_5(t) \leq q_6(t)$, is already highly suggestive.

Indeed, serving nodes 4 and 5 may make their queues smaller than those of nodes 3 and 6, leaving these queues to be served by themselves at a later stage, at which point inefficiency inevitably occurs. In fact, it can be proved that after some finite amount of time T_0, the fluid limit process enters a *natural state*, when $q_3(t) \geq q_4(t)$ and $q_6(t) \geq q_5(t)$, with equality only when both sides are zero. As described above, instability is bound to occur when schedule M_4 is used repeatedly for substantial periods of time while the fluid limit process is in a natural state. It is intuitively plausible that such events occur repeatedly with positive probability, but a rigorous proof that this leads to instability is far from simple. Such a proof requires detailed analysis of the underlying stochastic process (in our case via fluid limits), and its conclusion crucially depends on the weight function. Indeed, the stability results in [3, 4, 11] indirectly indicate that our network is *not* rendered unstable for sufficiently cautious weight functions.

We now characterize the dynamics of the fluid limit process. For a weight function $\gamma \log(1 + x)$, $\gamma > 1$, the fluid limit process follows oscillatory piece-wise linear trajectories, with random switches. In particular, in the fluid limit a node must completely empty almost surely before it releases the medium. Thus, a transition from one component to another occurs when one or both queues in that component hit zero.

More specifically, any fluid limit $\{q(t)\}_{t\geq 0}$ can be shown to have the following properties. For notational convenience, we henceforth assume $\mu_i \equiv 1$.

1. *Piecewise linearity:* For each node i, there are countably many positive time intervals $[z_L, z_H)$, $z_L < z_H < \infty$ such that $q_i(z_L) = q_i(z_H) = 0$, $q_i(z) > 0$, $z \in (z_L, z_H)$, as a consequence of continuity. Define $\tau_Z = (1 - \lambda_i)z_H + \lambda_i z_L$, then almost surely,

$$\begin{aligned} q_i(s) &= \lambda_i(s - z_L), \quad s \in [z_L, \tau_Z] \\ q_i(s) &= (\tau_Z - z_L) - (1 - \lambda_i)(s - z_L), \quad s \in [\tau_Z, z_H] \end{aligned} \tag{1}$$

and, in case $z_H = \infty$, only the first equation in (1) holds. Linearity as in (1) follows from the choice of $\gamma > 1$. Thus the limiting measure is governed by the joint distribution of such endpoints, termed *switching epochs*, as (1) determines the fluid path by continuity. In general, such fluid limits are

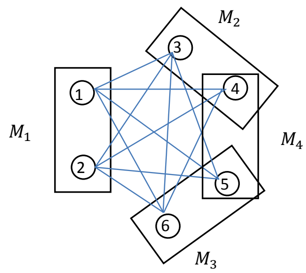

Figure 1: A network with 4 possible schedules obtained by removing 1 edge from a 3-partite complete network.

not Markovian. Finally there is no 'idling' and queues in the same component drain together, if at all.

2. *Switching*: Following the completion of M_1 (of strictly positive duration), the schedules M_2, M_3 are each begun with probability 3/8 and M_4 with probability 1/4. Similarly following the completion of M_2 for a positive period *strictly after* T_0, schedule M_1 is begun with probability $\bar{p} < 1/2$, M_3 with probability $1-\bar{p}$, and M_4 with probability 0. Note that M_3 has an advantage because node 4 drains first under M_2. Moreover the outcome is affected by the queue size at node 3 at the time when it first backs off and hence is affected by the backoff parameter. Similarly for switching out of M_3. Finally following the completion of a strictly positive M_4 period, switching occurs to M_2 or to M_3 according to which of nodes 4 and 5 drain first. In case they drain together, one of these two events will occur and M_1 with probability 0.

Let $(\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6) = \rho(\kappa_1, \kappa_2, \kappa_3, \kappa_4, \kappa_5, \kappa_6)$ with $\max\{\kappa_1, \kappa_2\} + \kappa_3 + \kappa_6 = 1$ and $\kappa_3 > \kappa_4$ and $\kappa_6 > \kappa_5$. The fluid limit process can then be shown to be unstable in the sense that $L(t) = \sum_{k=1}^{3} \max_{i \in M_k} q_i(t) \to \infty$ as $t \to \infty$, as stated in the next theorem.

THEOREM 1. *Consider the network of Figure 1, with* $w(x) = \gamma \log(1+x)$ *and* $\gamma > 1$. *For any* $m > 1$, *there exists a constant* $\rho^* = \rho^*(\kappa, m) < 1$, *such that for all* $\rho \in (\rho^*, 1]$,

$$\limsup_{t\to\infty} \mathbb{E}\left\{\frac{t}{L^m(t)}\right\} = 0,$$

for any initial state $q(0)$ *with* $|L(0)| = 1$.

COROLLARY 1. *For any* $m > 1$, *there exists a constant* $\rho^* = \rho^*(\kappa, m) < 1$, *such that for all* $\rho \in (\rho^*, 1]$,

$$\liminf_{t\to\infty} \frac{L(t)}{t^{1/m}} = \infty,$$

almost surely.

The original stochastic process is said to be unstable when $\{(U(t), Q(t))\}_{t\geq 0}$ is transient, and $\|Q(t)\| \to \infty$ almost surely for any initial state $Q(0)$. Exploiting similar arguments as in Meyn [9], the instability of the original stochastic process can be deduced from the instability of the fluid limit process.

For the sake of transparency, here we focused on the specific six-node network of Figure 1, and weights $\gamma \log(\cdot)$ with $\gamma > 1$. Similar instability issues can however arise in a far broader class of interference graphs, as we have discussed in [2]. The proof arguments further suggest that instability can in fact occur for any $\gamma > 1/K$ for network sizes of order K. For example, a graph obtained by duplicating each node of the six-node network K times is unstable for any $\gamma > 1/K$. The duplication is formally described as follows. Let $\mathcal{I}(i)$ denote the set of neighbors of i. For each node $1 \leq i \leq 6$, add K duplicate nodes $d_1^{(i)}, \ldots, d_K^{(i)}$ to the graph, with the same arrival rate λ_i and the same initial queue length $Q_i(0)$, such that each node $d_j^{(i)}$ is connected to all the neighbors of node i and their duplicates, i.e., $\mathcal{I}(d_j^{(i)}) = \mathcal{I}(i) \cup_{l \in \mathcal{I}(i)} \{d_1^{(l)}, \cdots, d_k^{(l)}\}$, for all $1 \leq j \leq K$. We define $D_i^{(K)} := \{i, d_1^{(i)}, \ldots, d_K^{(i)}\}$ as the duplicate collection of node i. Essentially, for $\gamma > 1/K$, each duplicate collection acts as a super node with $\gamma > 1$: although a node alone may backoff, not all the nodes in the duplicate collection will backoff simultaneously, thus the collection will not release the channel until the entire backlogs of the nodes are cleared (as in the six-node network).

4. REFERENCES

[1] N. Bouman, S.C. Borst, J.S.H. van Leeuwaarden, A. Proutière (2011). Backlog-based random access in wireless networks: fluid limits and delay issues. In: *Proc. ITC 23*, 39–46.

[2] J. Ghaderi, S.C. Borst, P.A. Whiting (2013). Queue-based random-access algorithms: fluid limits and stability issues, *Stochastic Systems*, submitted, *arXiv:1302.5945*.

[3] J. Ghaderi, R. Srikant (2010). On the design of efficient CSMA algorithms for wireless networks. In: *Proc. CDC 2010.*

[4] L. Jiang, D. Shah, J. Shin, J. Walrand (2010). Distributed random access algorithm: scheduling and congestion control. *IEEE Trans. Inform. Theory* **56 (12)**, 6182–6207.

[5] L. Jiang, J. Walrand (2010). A distributed CSMA algorithm for throughput and utility maximization in wireless networks. *IEEE/ACM Trans. Netw.* **18 (3)**, 960–972.

[6] J. Liu, Y. Yi, A. Proutière, M. Chiang, H.V. Poor (2008). Maximizing utility via random access without message passing. Tech. Rep. MSR-TR-2008-128, Microsoft Research.

[7] M. Lotfinezhad, P. Marbach (2011). Throughput-optimal random access with order-optimal delay. In: *Proc. Infocom 2011.*

[8] P. Marbach, A. Eryilmaz (2008). A backlog-based CSMA mechanism to achieve fairness and throughput-optimality in multihop wireless networks. In: *Proc. Allerton 2008.*

[9] S.P. Meyn (1995). Transience of multiclass queueing networks via fluid limit models. *Ann. Appl. Prob.* **5**, 946–957.

[10] J. Ni, B. Tan, R. Srikant (2010). Q-CSMA: queue-length based CSMA/CA algorithms for achieving maximum throughput and low delay in wireless networks. In: *Proc. Infocom 2010 Mini-Conf.*

[11] S. Rajagopalan, D. Shah, J. Shin (2009). Network adiabatic theorem: an efficient randomized protocol for content resolution. In: *Proc. ACM SIGMETRICS/Performance 2009.*

[12] D. Shah, J. Shin (2010). Delay-optimal queue-based CSMA. In: *Proc. ACM SIGMETRICS 2010.*