Proceedings of the 17th International Symposium on Mathematical
Theory of Networks and Systems, Kyoto, Japan, July 24-28, 2006

Analysis of an Implementable Application Layer Scheme for Flow Control over Wireless Networks

Alessandro Abate, Minghua Chen and Shankar Sastry
Department of Electrical Engineering and Computer Sciences
University of California, at Berkeley
{aabate,minghua,sastry}@eecs.berkeley.edu

Abstract— **This paper deals with the problem of congestion control and packets exchange on a wireless network. The mathematical model of the protocol is inspired by and extends a known fluid flow scheme for the control of congestion on a wired network. The necessity to introduce a specific wireless model is motivated by the presence of channel error; often this error (due to intrinsic noise or channel corruption) is not known exactly. This motivates the approximation of parts of the structure of the model with binary functions, whose switching point can be precisely known. These new discontinuous elements, while in practice greatly simplifying the structure of the algorithm (they carry a single bit of information), complicate the theoretical analysis of its dynamical properties. We therefore approximate them with continuous functions with proper limiting behavior: they thus preserve the simple shape and yield themselves to analysis as well. Given this setup, we then investigate the important issues of existence and uniqueness of the equilibrium for the dynamical system, and of local asymptotic stability. Furthermore, we show that this equilibrium solves a concave net utility optimization problem, of which the classical one for wired networks is a special case. The take away point of this work is that the scheme we propose to handle the traffic on a wireless network is not only innovative and meaningful, but has also the potential to be modified and translated into practical implementation.**

I. Introduction

Transmission Control Protocol (TCP) has recently been the focus of much research (originated, among the many contributions, in [1]–[3]). Not long ago, this practical scheme has been dynamically modeled via a system of continuous time differential equations that describe the evolution of the rates (that is, the number of bits per second) of a set of users that exchange information over a network. This is an instance of fluid flow model [4], [5]. The study of this model advances the understanding of the intrinsic characteristics and dynamical properties of the system. Investigating this scheme has nevertheless proven to be a rather challenging task, mostly because of the presence of strong non linearities in the functions that come into play, and because of the distributed nature of the scheme. Moreover, the multiple couplings between its entities (senders, receivers and links) hampers the global understanding of its behavior.

The current fluid flow models for TCP are limited to the case of wired networks [4] [5]. Fundamental properties such as uniqueness of equilibria and stability have been studied [6], [7] and conditions for achieving robustness to disturbances [8] and to delays [9] have been introduced.

Quite recently some researchers have turned their attention to the wireless scenario. This new setting poses new, unconfronted challenges, due to the presence of intrinsic noise and channel errors at the link level. An algorithm known as MULTFRC [10] and proposed for video streaming over wireless networks has yielded a scheme for TCP-friendly rate control, or TFRC, for wireless networks. In [11], a corresponding continuous-time model is introduced and studied. Many properties, such as global stability, robustness conditions to delays and to disturbances, have been derived [12], [13].

This paper takes a further step: the presence of channel error is the cause of imperfect feedback from the network to the users; these errors prevent the exact measurement of the congestion status on the network. This motivates the introduction of a simplifying approximation for that part of the model which is affected by noise. This approximation, in the form of a step function that switches at a known (or computable) point, is on the one hand simpler, but on the other hand discontinuous. It is thus quite hard to do analysis on the modified scheme. This calls for the introduction of some continuous approximations, the limiting behavior of which is studied. It shall be argued and motivated that the approximated scheme represents an implementable version of the proposed algorithm for wireless networks.

The paper unfolds as follows: after a brief explanation of fluid-flow models for wireline networks and a concise introduction to the TCP scheme for wireless ones in Sec. II, we propose its related modification and the corresponding continuous approximations (Sec. II-E). In Section III, a series of facts will elucidate the existence and uniqueness of the equilibrium for the approximation of the modified system. Furthermore, local stability for the scheme will be proved and limiting behaviors explained. It will then be shown that the equilibria of the modified model are the solution of a concave net utility optimization problem, of which the generic one proposed by Kelly for TCP on wired networks [4] is a special case. In Section IV, discussions and a description of future work will close up the paper.

II. A Practical Flow Control Scheme

In this section we first present the dynamical model of the well known general flow control problem first introduced

This work was supported by NSF grant ANI-9905799, the AFOSR contract F49620-00-1-0327 and the NSF grant CCR-0225610.

by Kelly *et. al.* [4]. Starting from the wired scenario, we motivate and build up the extensions for the more challenging wireless case; a modification to the model for this setup is discussed in order to simplify it and enable its practical implementation.

A. Wired Networks

A communication network is described via a set J of *links* and a set R of *users* (sender-receiver pairs). Each $j \in J$ has a finite capacity $C_j < \infty$. The network interconnections are described via a routing matrix $A = (a_{jr}, j \in J, r \in R)$, where $a_{jr} = 1$ if $j \in r$, $a_{jr} = 0$ else. A fluid-flow, continuous-time model for the TCP scheme [4] has been proposed in order to facilitate the analysis of the properties of the protocol. To each user a sending rate $x_r \geq 0$ and a utility function $U_r(x_r)$ are associated. $U_r(x_r)$ is assumed to be increasing, strictly concave and $\mathcal{C}^1$. The exchange of information between users over the links can be interpreted as a concave maximization problem [14], [15] depending on the aggregate utility functions for the rates and on some costs on the links:

$$\max \sum_{r \in R} U_r(x_r) - \sum_{j \in J} P_j \left(\sum_{s: j \in s} x_s \right), \tag{1}$$

where the cost functions $P_j(\cdot)$ are defined as

$$P_j(y) = \int_0^y p_j(z)\, dz. \tag{2}$$

The terms $p_j(y)$ can be interpreted as "prices" at the link and are assumed to be non-negative, continuous and increasing functions; they represent some congestion measure and, as can be inferred from their structure, they have a local dependence on the aggregate rate passing through the link. As in [5], in this paper we shall stick to the following "packet loss rate":

$$p_j(y) = \frac{(y - C_j)^+}{y}. \tag{3}$$

Flow control can then be regarded as a dynamical system evolving according to Problem (1), that is having an equilibrium which is the solution of (1). User r will accrue a packet loss rate which, under our assumptions of small p_j, can be approximated as $\sum_{j \in r} p_j(\sum_{s: j \in s} x_s)$. The rate control scheme has the following shape, for $r \in R$:

$$\frac{d}{dt} x_r(t) = k_r \left(w_r^o - x_r(t) \sum_{j \in r} p_j \left(\sum_{s: j \in s} x_s(t) \right) \right), \tag{4}$$

with k_r a positive scale factor affecting the adaptation rate, and the constant w_r^o can be physically interpreted as the number of connections that the user establishes with the network; as discussed, the congestion signal (packet loss rate) depends on the sum of the prices along all the links that are crossed by the user. Interpreting the model (4) as a dynamical relation, it is easy to express its equilibrium in an implicit form. In [4] it is shown by Lyapunov arguments that this equilibrium is unique and asymptotically stable. Moreover, the schemes can be endowed, under conditions over their parameters, with many interesting properties (as an example, robustness to delays).

B. Wireless Networks

Wireless channels are affected by errors, due to the corruptibility of the signals flowing through them and to the presence of noise. This directly influences the packet loss at each link in a TCP-like setting. We thus encompass this fact within a new price function for, say, link $j \in J$:

$$q_j \left(\sum_{s: j \in s} x_s(t) \right) \triangleq p_j \left(\sum_{s: j \in s} x_s(t) \right) + \epsilon_j. \tag{5}$$

This function accounts for both the congestion measure (presence of the term p_j), as well as the non-negative term ϵ_j that accounts for the channel error. The TCP model (4) will then depend on this new function q_j. It is again easy to calculate the equilibrium of this new dynamic relation. Interpreting this fact through an underlying optimization problem, as in (1), shows that the new equilibrium will be suboptimal. This fact motivates the introduction of an enhancement to the wireless scheme, as described in the following section.

C. A new Control Scheme for Wireless Networks

In [11]–[13], we introduced two extensions of the TCP scheme, both aimed at compensating its suboptimality when employed in wireless networks. In this paper we shall focus on one of these two proposed schemes, the "dynamic update." Assume the term ω_r is time dependent, $\omega_r(t)$, and evolves according to:

$$\frac{d}{dt} w_r(t) = c \left(w_r^o - w_r(t) \frac{\sum_{j \in r} p_j(\sum_{s: j \in s} x_s(t))}{\sum_{j \in r} q_j(\sum_{s: j \in s} x_s(t))} \right). \tag{6}$$

We can interpret this dynamical relation, in a fluid-flow sense, as the modification of the number of connections that the user has with the network. It is easy to compute the equilibria $\{x_r^*, \omega_r^*\}$ of this new scheme and check that the "optimum point" of the first component (the rates) is the same as that of (4). Aiming at a dynamical analysis of this scheme, in [11] we showed that the interconnection is globally asymptotically stable, under the realisitc assumption that the two dynamical relations evolve in two different time scales. In [12] and [13] we instead investigated the robustness of the scheme to delays and studied its resilience against disturbances. It is important to stress that this scheme can be easily implemented by adjusting the number of connections which an application opens in a real network. Therefore, it is an *application layer based approach* and it is easy to deploy, as it does not require changes on the network infrastructure or its protocol.

D. A Practical Discretization of the Scheme

From the structure of Eqn. (6) we can gather that the implementation of the control law for w_r depends on the precise measurement of the ratio $\frac{\sum_{j \in r} p_j(\sum_{s: j \in s} x_s(t))}{\sum_{j \in r} q_j(\sum_{s: j \in s} x_s(t))}$, which is the portion of the packet loss rate that is exclusively caused by congestion. From an end-to-end point of view, users can infer which packet is lost only by observing a discontinuity in the sequence number that is carried by every packet[1]; the reason of the loss (congestion or channel error) would not be given though. Therefore users can only precisely measure the end-to-end packet loss rate, i.e. $\sum_{j\in r} q_j(\sum_{s:j\in s} x_s(t))$, but not the quantity due to congestion, i.e. $\sum_{j\in r} p_j(\sum_{s:j\in s} x_s(t))$.

In principle, users can have the ability to exactly measure $\sum_{j\in r} p_j(\sum_{s:j\in s} x_s(t))$, provided more information is gathered from the network infrastructure. As an example, the routers and the base stations can generate an Explicit-Loss-Notification (ELN) marking[2] on consecutive packets when they understand that the current packet is lost due to the wireless transmission. Therefore if the users observe a lost packet, they can check the ELN bit on the successive packet to see whether the loss is caused by congestions or by channel error. This way, users can get a precise measure of $\sum_{j\in r} p_j(\sum_{s:j\in s} x_s(t))$, and therefore a better estimate of the above ratio. Other solutions are based on end-to-end statistics, or there exist schemes that are not using packet loss as a congestion measure: for instance, TCP Vegas quantifies the congestion on a measure of the queueing delay. However, to our best knowledge, none of the real world network infrastructures currently employ these functionalities. Even worst, it is quite hard to add these enhancemente in every router and base station, and it may break the end-to-end principle Internet relies on.

All of the above motivates to seek a better way to control the quantities w_r based on some alternative that is easy for users to measure. We first gauge how the ratio affects the system performance in (6):

- If a route r is underutilized, then the ratio is zero; this implies that the number of connections $w_r(t)$ increases in order to boost the user rate $x_r(t)$, which makes the system pursue full utilization on the r^{th} route;
- If the route r is fully utilized, i.e. if any one of its link is congested, than the ratio takes a value between zero and one, finely adjusting $w_r(t)$, and hence $x_r(t)$, to make the system pursue the maximum utility.

This behavior suggests the idea of replacing the ratio with an *indicator function*. Specifically, introducing the vector quantity $x(t) = [x_1(t), \ldots, x_{\mathrm{card}(R)}(t)]^T$, let us define the following quantity:

$$\begin{aligned} I_r(x) &= \mathrm{Ind}\left(\sum_{j\in r} \frac{(y_j(t)-C_j)^+}{y_j} > 0\right) \qquad (7) \\ &= \begin{cases} 1, & \text{if route } r \text{ is congested at time } t; \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Here $y_j(t) = \sum_{s:j\in s} x_s(t)$ is the aggregate rate flowing through link j. From this definition, we can observe that

- $I_r(x)$ has exactly the same behavior as the ratio when route r is underutilized, therefore, replacing the ratio with $I_r(x)$ will not affect the system's thrust to pursue full utilization.
- If any of the links of route r is congested, $I_r(x)$ does not have the exact same behavior as the ratio; instead, it assumes the value one to push down $w_r(t)$, so as to decrease $x_r(t)$ in order to avoid further congestion on the route.

Unlike the ratio in (6), the value of the indicator function can be easily and accurately estimated by each user. In fact, its value is directly correlated to changes on the round trip time (RTT) for each user[3]. Physically, RTT consists of the round trip propagation delay and round trip queuing delay. For a given route, assuming the backward path is congestion-free, i.e. the incoming rates to the sender are less than the links capacities, the round trip propagation delay is a fixed value, and the queuing delay is zero if the forward path is not congested. If the forward path is congested, the queuing delay increases to positive values, and if the path is continuously congested the value keeps increasing to a maximum until the buffer is overflowed. Hence, *an increase in RTT is due to the presence of forward congestion* (and therefore increasing queueing delay)[4]; the increase itself is symptomatic of the indicator function assuming a value of one. On the other hand, if there is no increase in RTT, then most probably the route is not congested, which means that he indicator function is likely to be equal to zero.

System (4-5), endowed with this new term, is modified as:

$$\begin{cases} \frac{d}{dt}x_r(t) = k_r\left(w_r(t) - x_r(t)\sum_{j\in r}(\epsilon_j + p_j(y_j(t)))\right); \\ \frac{d}{dt}w_r(t) = c\left(w_r^o - w_r(t)I_r(x)\right). \end{cases} \qquad (8)$$

The model is a nonlinear, coupled system with discontinuities introduced by the terms $I_r(x), r \in R$. The discontinuities make it harder to analyze the system, which does not fit into the classical framework for analysis previously employed; it would instead require the study of solutions in the *Filippov* sense [16]. We instead decide to tackle this problem by approximating the term $I_r(x), r \in R$ with continuous functions; hence we get a continuous approximated version of the system in (8), which we describe in the following.

E. Continuous Approximations of the System and the Two Time Scales Assumption

The parameter-dependent function we use to approximate $I_r(x)$ in (8) is the following, where $\beta > 0$:

$$f_r^\beta(x) = \frac{e^{\sum_{j\in r}\ln\left(1+e^{\beta\frac{y_j-C_j}{y_j}}\right)} - 1}{1 + e^{\sum_{j\in r}\ln\left(1+e^{\beta\frac{y_j-C_j}{y_j}}\right)}}, \quad r \in R. \qquad (9)$$

[1]In practice, the sender waits for three duplicate acknowledgements asking for the retransmission of the missing packet, before it asserts that the packet is lost.

[2]Along with ELN, there exist schemes known as Explicit-Congestion-Notification (ECN) that, as intuitive, work similarly.

[3]Here we call RTT the sum of the time it takes a packet to go from sender to receiver, and back.

[4]Again, under the slack assumption that the incoming rates to the sender are less than the links capacities.

we also approximate the non-smooth quantity $p_j(y_j(t)) = (y_j(t) - C_j)^+/y_j(t)$ in (8) with the following function:

$$g_j^\beta(y_j(t)) = \frac{1}{\beta}\ln\left(1 + e^{\beta\frac{y_j(t)-C_j}{y_j(t)}}\right), \quad j \in J. \tag{10}$$

It should be clear that $f_r^\beta(x) \to I_r(x)$ and $g_j^\beta(y_j(t)) \to p_j(y_j(t))$ as $\beta \to \infty$.

The corresponding approximated system is, $\forall r \in R$,

$$\begin{cases} \frac{d}{dt}x_r(t) = k_r\left(w_r(t) - x_r(t)\sum_{j\in r}(\epsilon_j + g_j^\beta(y_j(t)))\right); \\ \frac{d}{dt}w_r(t) = c\left(w_r^o - w_r(t)f_r^\beta(x)\right). \end{cases} \tag{11}$$

Since the approximated system in (11) is continuous, we can then analyze its equilibrium and stability for arbitrary values of β. As $\beta \to \infty$, the system in (11) approaches the original system in (8). Therefore, the logic is to analyze the properties of the system in (11) and, by letting $\beta \to \infty$, we expect to reveal those of the interconnection in (8).

The approximated system in (11), although continuous, is still complex to analyze. Similar to the model in (4-6), it is a nonlinear, coupled, multivariable system, and the two equations are not exactly symmetrical even though they might appear to be so.

In [11] we argued that in the actual TCP schemes the rate of change of the quantity w_r, representing the number of connections that a user has with the network, is dimensionally less than that of x_r, representing the source sending rate. Therefore, inspired by the control literature on singular perturbation systems [16], we carefully make a key assumption to enable the time-decoupling of the system: *the dynamics corresponding to x_r and w_r evolve in two different time scales; the first in a faster one, while the second in a slower one*. This helps us derive strong results for the overall interconnection.

The two time scales assumption applied to the approximated system in (11) highlights two kinds of dynamics: a fast one, which is described in the *boundary-layer system*, and a slow one, which is encompassed in the *reduced-order system.* The fast interconnection is described, $\forall r \in R$, as

$$\begin{cases} \frac{d}{dt}x_r(t) = k_r\left(w_r(t) - x_r(t)\sum_{j\in r}(\epsilon_j + g_j^\beta(y_j(t)))\right); \\ w_r(t) = \text{constant}. \end{cases} \tag{12}$$

In the slower timescale, we instead have the following dynamics, $\forall r \in R$:

$$\begin{cases} x_r(t) = \frac{w_r(t)}{\sum_{j\in r}(\epsilon_j + g_j^\beta(y_j(t)))}; \\ \frac{d}{dt}w_r(t) = c\left(w_r^o - w_r(t)f_r^\beta(x(t))\right). \end{cases} \tag{13}$$

Under the two times scale setting, the behavior of the system can be described as follows. On the fast timescale, w_r can be thought as being held constant, and the entire system can be expressed as the boundary system shown in (12). This system is nothing but a slight modification of Kelly's control system on wired network (as expressed in (4)), except for the term w_r^o replaced by the "constant" $w_r(t)$ and the price function $p_j(y_j(t))$ replaced by $\sum_{j\in r}(\epsilon_j + g_j^\beta(y_j(t)))$; the behavior of the boundary system can thus be easily inferred from the known results of the system in (4). It has a unique and globally exponentially stable equilibrium, which is a function of w_r. Particularly, on the fast timescale, x_r converges to the equilibrium manifold defined as follows:

$$x_r(t) = \frac{w_r(t)}{\sum_{j\in r}(\epsilon_j + g_j(y_j(t)))}, \quad r \in R. \tag{14}$$

On the slow timescale, x_r has already converged to the equilibrium manifold, and the system collapses into the reduced system described in (13). Its behavior determines how the approximated system evolves in the long run; therefore, together with the boundary layer system, it fully characterizes behavior of the system for all possible times. Motivated by the above considerations, we shall mainly focus on investigating the reduced system in (13).

III. Analysis and Simulations Results

A. Existence, Uniqueness of The Equilibrium, Its Local Stability and The Related Optimization Problem

In this section we show that the system in (11) has a unique equilibrium, and that this equilibrium is locally exponentially stable. We start from showing that any existing equilibrium is locally stable in a neighborhood; then, thanks to this fact and together with some results from the Poincare-Hopf Index Theorem [16], we conclude that there can be only one equilibrium.

Before stating the main results, the following fact is introduced. Let us remind the definition of the vector quantity $x(t) = [x_1(t), \ldots, x_{\text{card}(R)}(t)]^T$, and similarly for $w(t)$.

Lemma 1: The equilibrium manifold shown in (14) is a one-to-one mapping between $x(t)$ and $w(t)$; moreover, the following holds on the manifold:

$$\dot{w} = D(x)\dot{x},$$

where

$$D(x) = \text{diag}(x)\Big(\text{diag}\Big(\sum_{j\in r}\frac{\epsilon_j + g_j(y_j)}{x_r}\Big) + A^T\text{diag}(g_j'(y_j))A\Big)$$

is a product of two positive definite matrices, and as such all its eigenvalues are positive.

Proof: Refer to Appendix A. ■

Remark 1: Lemma 1 implies that within the reduced system (13), analyzing the behavior of the system with respect to x is equivalent to carrying out the analysis with respect to w; as a matter of fact, both of them, as well as their derivatives, are in a one-to-one relationship.

Based on the two time scales decomposition and on singular perturbation theory [16], [17], showing that for the approximated system in (11) any possible equilibrium is locally exponentially stable follows from the fact that both the boundary system and the reduced system are locally exponentially stable around the equilibrium. We first claim the following lemma for the reduced system:

Lemma 2: Assume that for the reduced system in (13), x^e is one of its possible equilibria; then x^e is locally exponentially stable for any $\beta > 0$.

Proof: Refer to Appendix B. ■

Furthermore, exploiting the fact that the boundary layer system is locally exponentially stable [18], we apply arguments used in [16] and in [17] for the stability of singular perturbation non-linear system to infer that any equilibrium of the composite system shown in (11) is locally exponentially stable. This fact is stated in the following:

Theorem 1: x^e, the equilibrium for the composite system shown in (11) with arbitrary $\beta > 0$, is locally exponentially stable.

Thus far we have shown that any existing equilibrium is locally exponentially stable. Another important question to address is how many equilibria there are for the system. The answer is stated in the following:

Theorem 2: For any arbitrary $\beta > 0$, the approximated system (11) has a unique equilibrium.

Proof: Refer to Appendix C. ■

Remark 2: Theorem 1 and Theorem 2 state the existence of a unique equilibrium and ensure its locally stability for the continuous approximated system in (11), for any value of β. At the limit as $\beta \to \infty$, the approximated system approaches the original discontinuous system in (8). Therefore, for large β, we expect the approximated system to have a very close behavior with the original system, except at the discontinuity point $y_j(t) = C_j$.

In the following, we motivate how the unique equilibrium solves a concave optimization problem, which is a modification of the one proposed for the wired case in Eqn. (1).

Theorem 3: For any arbitrary $\beta > 0$, the unique equilibrium of the approximate system in (11), denoted by (x^e, w^e), solves the following concave optimization problem

$$\max_{x \geq 0} \sum_{r \in R} U_r(x_r) - \sum_{j \in J} \int_0^{y_j} g_j(z)\, dz, \qquad (15)$$

with $U_r, r \in R$ being the concave function:

$$U_r(x_r) = \int_0^{x_r} h_r^{-1}\left(\frac{w_r^o}{\nu}\right) d\nu, \quad r \in R,$$

where $h_r^{-1}, r \in R$, is the inverse of the monotonically increasing function h_r:

$$h_r(z) \triangleq \left(\sum_{j \in r} \epsilon_j + z\right) f_r(z) = \left(\sum_{j \in r} \epsilon_j + z\right) \frac{e^{\beta z} - 1}{e^{\beta z} + 1}.$$

Proof: First it is easy to see the net utility function in (15) is concave. Then the claim follows by setting to zero the derivative of the net utility function with respect to x. ■

One observation for Theorem 3 is in order: the unique equilibrium for the system in (11) for the wireless scenario solves a concave optimization problem which is similar to the general one (Eqn. (1)) solved for the wired network [4], but with different utility functions $U_r(x_r)$ for each user. More precisely, while the $U_r(x_r)$ in the wired network case is only a function of x_r, in wireless scenario it is also a function of $\sum_{j \in r} \epsilon_j$, that is the packet loss rate associated with the route r. In fact, if we let $\beta \to \infty$ and $\epsilon_j = 0, \forall j \in J$, i.e. if we tend to the wired network scenario, we have $h_r(z) = z$, and thus the optimization problem in (15) becomes identical to that of the wired network optimization one. In this case, the equilibrium x^e is exactly the same as x^*, implying the optimization problem in the wired network is merely a special case of that in (15).

Regarding the actual implementation of the proposed scheme in (8), it is necessary to discretize continuous quantities. For instance, controlling w_r is implemented by adjusting the number of connections, which has to be an integer number; controlling x_r is implemented by adjusting the number of finite packets to be sent out in a time interval. Therefore, it is very unlikely that the system will operate exactly at those points of discontinuity. From this point of view, the analysis based on the approximated system is accurate enough to predict and interpret the performance of the actual implementation of the algorithm.

From a theoretical point of view, the existence of a unique locally stable equilibrium encourages our effort to show that in fact the equilibrium is *globally* asymptotically stable; indeed we have already seen that the whole setting can be interpreted as a utility maximization problem that holds globally.

B. Simulations

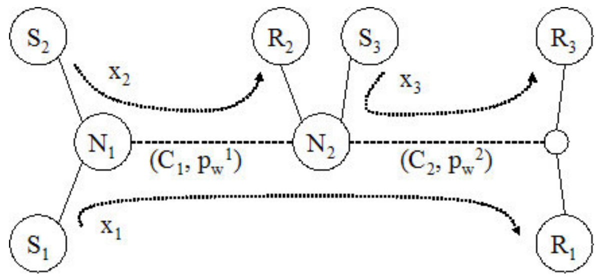

Fig. 1. Simulation topology.

In the following we present the output of some simulations. They show how the performance of the modified scheme closely matches that of the original scheme, in which we assumed full information of the feedback signals from each link. The topology is presented in Fig. 1 and matches that in [11]. The two time scales assumption has been taken in consideration by properly setting the multiplicative constants in the differential equations: as a result, it can be observed that the changes of w_r are slower than those of the rates x_r. The initial conditions for Figure 2 are precisely those of Fig. 5 in [11]. The reader should compare these two plots to convince himself of the similarity of the results. Due to the discretizations we introduced, the current outcomes display some oscillations (see for instance Fig. 2-second) that were not present in the original scheme; for this reason, we have refined the integration step and thus necessarily increased the simulation time. This oscillating behavior happens around the optimum for the system (see Fig. 2-first plot), which matches that of [11]; we discussed that this optimum corresponds to the full utilization of the links (observe the oscillations of the congestion measures, (see

Fig. 2-third)), hence at conditions close to the discontinuity points in the vector fields.

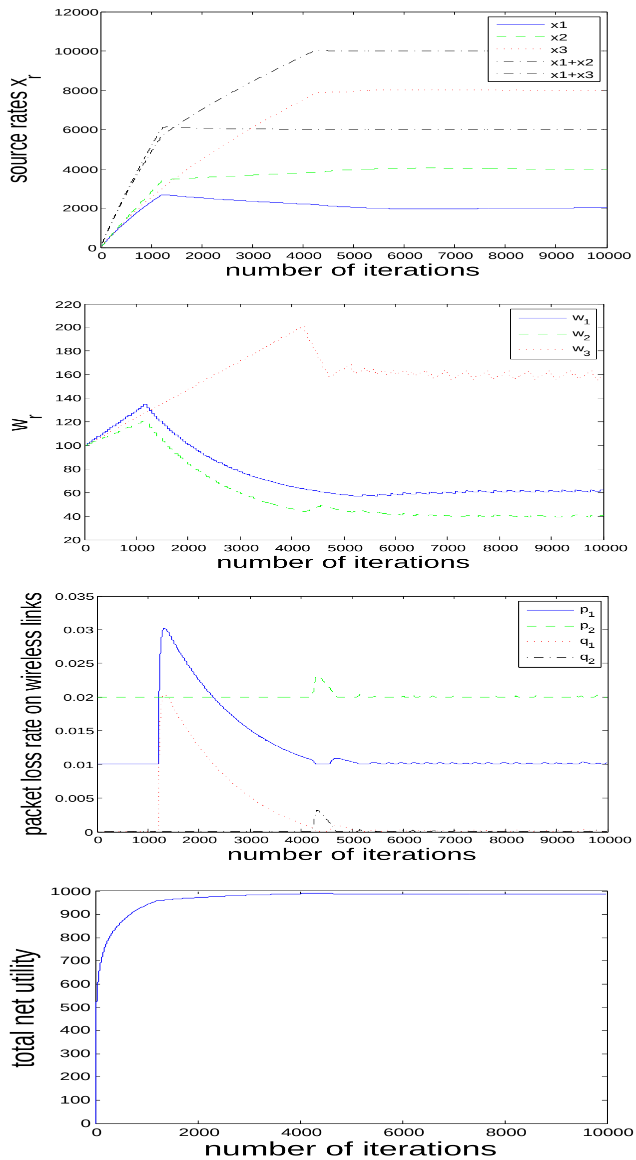

Fig. 2. Simulations for the modified "dynamic-update" scheme. Convergence of rates $x_r(t), r = 1, 2, 3$ (top figure), $w_r(t), r = 1, 2, 3$ (second plot), packet loss rate $p_j(\cdot)$ and $q_j(\cdot), j = 1, 2$ (third plot), and net utility (bottom figure); the initial rates have been set to the value 0.

IV. Conclusions and Future Work

Standing upon the results presented in [11], where a fluid-flow approximation as a dynamical scheme for controlling the flow over packet-switched wireless networks was proposed and analyzed, this paper introduces an alternative of such model for the wireless scenario. The model is obtained by introducing an indicator function. This simplification is motivated by the necessity to apply the scheme to real world networks, which present inaccurate feedback to the end-users; the new, 1-bit scheme is still an application layer based approach, which therefore does not require any change in the network infrastructure and protocol. The modified model, although easier to implement than its precursor, comes at the cost of introducing some discontinuities in the dynamics, which complicate the theoretical analysis. Therefore, we propose an approximation based on some continuous, parameter-dependent functions which, at the limit, coincide with the discontinuous ones. The new functions yield themselves to some analysis: we prove the existence and uniqueness of the equilibrium of the interconnected systems, solving a concave net utility optimization problem, of which the generic one proposed by Kelly *et. al.*, [4], is a special case. Moreover, we show that this scheme, on a neighborhood of the equilibrium, is exponentially stable. These results are accurate enough to predict and interpret the performance in reality, and are interesting enough to encourage continuing efforts in theoretical aspects.

Given the parallel with the model in [11], the investigaton of the global asymptotical stability of the unique equilibrium holds promising results; furthermore, interpreting the properties of the equilibrium from the network optimization standpoint, such as fairness between users and route utilization, may give important further insights. The delay stability and the robustness to stochastic disturbance are also interesting to investigate from both a practical as well as a theoretical point of view.

V. Acknowledgements

The authors would like to thank Ao Tang for the helpful exchange of e-mails, and ChenKe Xu for the discussions.

References

[1] V. Jacobson, "Congestion avoidance and control," *Proceedings of ACM SIGCOMM*, pp. 314–329, 1998.

[2] S. Floyd, "Tcp and explicit congestion notification," *ACM Computer Communication Review*, vol. 24(5), Oct 1994.

[3] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Trans. on Networking*, vol. 1(4), pp. 397–413, Aug 1993.

[4] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness, and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, Dec 1999.

[5] F. Kelly, "Fairness and stability of end-to-end congestion control," *European Journal of Control*, vol. 9, pp. 159–176, 2003.

[6] Z. Wang and F. Paganini, "Global stability with time-delay of a primal-dual congestion control," *IEEE Conference on Decision and Control*, Dec 2003.

[7] F. Paganini, Z. Wang, J. C. Doyle, and S. H. Low, "Congestion control for high performance, stability and fairness in general network," *ACM/IEEE Trans. on Networking*, to appear.

[8] R. Johari and D. Tan, "End-to-end congestion control for the internet: delays and stability," *IEEE/ACM Trans. on Networking*, vol. 9, no.6, pp. 818–832, Dec 2001.

[9] G. Vinnicombe, "On the stability of end-to-end congestion control for the internet," *University of Cambridge Technical Report CUED/F-INFENG/TR.398*, 2001.

[10] M. Chen and A. Zakhor, "Rate control for streaming video over wireless," *Proceeding of Infocom 2004*, 2004.

[11] M. Chen, A. Abate, A. Zakhor, and S. Sastry, "Stability and delay considerations for flow control over wireless networks," *UCB EECS ERL Technical Report*, 2005.

[12] M. Chen, A. Abate, and S. Sastry, "New congestion control schemes over wireless networks: Stability analysis," *Proceedings of the* 16^{th} *IFAC World Congress, Prague*, 2005.

[13] A. Abate, M. Chen, and S. Sastry, "New congestion control schemes over wireless networks: Delay sensitivity analysis and simulations," *Proceedings of the* 16^{th} *IFAC World Congress, Prague*, 2005.
[14] S. H. Low and D. E. Lapsley, "Optimzation flow control i: Basic algorithm and convergenece," *IEEE/ACM Trans. on Networking*, pp. 861–875, 1999.
[15] S. H. Low, "A duality model of tcp and queue management algorithms," *IEEE/ACM Trans. on Networking*, Aug 2003.
[16] S. Sastry, *Nonlinear Systems: Analysis, Stability and Control*. Springer Verlag, 1999.
[17] H. Khalil, *Nonlinear Systems (3rd edition)*. Prentice Hall, 2001.
[18] S. Kunniyur and R. Srikant, "A time scale decomposition approach to adaptive ecn marking," *Proceeding of IEEE Infocom*, vol. 1330-1339, Mar 2001.
[19] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
[20] J. Marsden and M. Hoffman, *Elementary Classical Analysis. Second edition*. W.H.Freeman and Company, New York, 1993.

APPENDIX

A. Proof of Lemma 1

Proof: Focusing on the manifold described by Eqn. (14), in order to prove the desired result we want to show that one $x(t)$ results in one $w(t)$, and conversely that one $w(t)$ results in one $x(t)$.

- It is easy to see from (14) that one $x(t)$ maps to a unique $w(t)$.
- Now we show that, given $w(t)$, there is only one value for $x(t)$ satisfying (14). Given $w(t) = w$, (14) is the maximum for the following strictly concave function of x over $\mathbb{R}^{\mathrm{card}(R)}$:

$$U(x) = \sum_{r\in R} w_r \log x_r - \sum_{j\in J} \int_0^{\sum_{s:j\in s} x_s} \left(\epsilon_j + \frac{(y - C_j)^+}{y}\right) dy.$$

The strict concavity implies that the maximum exists over $\mathbb{R}^{\mathrm{card}(R)}$ and is unique; hence there can only be one value for $x(t) = x$ satisfying (14). Therefore, one $w(t)$ maps to only one $x(t)$.

The relation between $\dot{w}$ and $\dot{x}$ is derived as in [11]. For the last claim of the proposition, refer to known results from [19]. ∎

B. Proof of Lemma 2

Proof: Around the equilibrium of the reduced system, let $x_r(t) = x_r^e + z_r(t)$ and denote $D(x^e)$ as $\tilde{D}$; linearizing around this point, we have that, $\forall r \in R$,

$$\begin{aligned}
\dot{z}(t) &= c\tilde{D}^{-1}\Bigg(\frac{1}{x_r^e} f_r(x^e) \sum_{j\in r}\left(\epsilon_j + g_j(y_j^e)\right) z_r(t) \\
&\quad + \sum_{j\in r}\left(\epsilon_j + g_j(y_j^e)\right)\mu_r^e \sum_{j\in r} \beta g_j'(y_j^e) \sum_{s:j\in s} z_s(t) \\
&\quad + f_r(x^e) \sum_{j\in r} g_j'(y_j^e) \sum_{s:j\in s} z_s(t)\Bigg), \\
&= -c\tilde{D}^{-1}\Bigg(diag\left(f_r(x^e)\right)\tilde{D} + diag\Big(\sum_{j\in r}\left(\epsilon_j + g_j(y_j^e)\right)\Big) \\
&\quad \cdot diag\left(\beta\mu_r^e\right) A^T diag(g_j'(y_j^e)) A\Bigg) z(t), \qquad (16)
\end{aligned}$$

where

$$\mu_r^e = \frac{e^{\sum_{j\in r}\ln\left(1+e^{\beta\frac{y_j^e - C_j}{y_j}}\right)}}{\left(1 + e^{\sum_{j\in r}\ln\left(1+e^{\beta\frac{y_j^e - C_j}{y_j}}\right)}\right)^2} > 0, \quad r \in R,$$

and

$$g_j'(y_j^e) = \frac{C_j}{(y_j^e)^2}\frac{e^{\beta\frac{y_j^e - C_j}{y_j^e}}}{1 + e^{\beta\frac{y_j^e - C_j}{y_j^e}}} > 0, \quad j \in J.$$

Denote

$$\begin{aligned}
E &= diag\left(f_r(x^e)\right)\tilde{D} \\
&+ diag\left(\sum_{j\in r}\left(\epsilon_j + g_j(y_j^e)\right)\beta\mu_r^e\right) A^T diag(g_j'(y_j^e)) A.
\end{aligned}$$

Then by simple arguments, the system in (16) is stable if and only if $\tilde{D}^{-1}E$ has all positive eigenvalues. We now show that this requirement is verified.

First notice that this is equivalent to show that the eigenvalues of $E\tilde{D}^{-1}$ are positive since $E\tilde{D}^{-1}$ is similar to $\tilde{D}^{-1}E$. Define $G = diag\left(\sum_{j\in r}\frac{\epsilon_j + g_j(y_j^e)}{x_r^e}\right)$. Then

$$E\tilde{D}^{-1} = diag(f_r(x^e)) + G \cdot diag\left(x_r^e\beta\mu_r^e\right) A^T diag(g_j'(y_j^e)) A\tilde{D}^{-1}.$$

At the same time we notice that

$$\begin{aligned}
&\tilde{D}\left[A^T diag(g_j'(y_j^e))A\right]^{-1} \\
&= \left(G + A^T diag(g_j'(y_j^e))A\right)\left(A^T diag(g_j'(y_j^e))A\right)^{-1} \\
&= G\left(\left(A^T diag(g_j'(y_j^e))A\right)^{-1} + \left(diag(\sum_{j\in r}\frac{\epsilon_j + g_j(y_j^e)}{x_r^e})\right)^{-1}\right).
\end{aligned}$$

Define the terms inside the brackets as B; we can then have the following expression for $E\tilde{D}^{-1}$:

$$\begin{aligned}
E\tilde{D}^{-1} &= diag(f_r(x^e)) + G \cdot diag\left(\beta\mu_r^e\right) B^{-1} G^{-1} \\
&= G \cdot diag(\beta x_r^e \mu_r^e)\left(diag\left(\frac{f_r(x^e)}{\beta x_r^e \mu_r^e}\right) + B^{-1}\right) G^{-1}.
\end{aligned}$$

We claim that $E\tilde{D}^{-1}$ has positive eigenvalues, due to the following three facts:

- $B \succ 0$, since it is a sum of two positive definite matrices; hence $diag\left(\frac{f_r(x^e)}{x_r^e\beta\mu_r^e}\right) + B^{-1} \succ 0$.
- $diag(\beta x_r^e\mu_r^e)\left(diag\left(\frac{f_r(x^e)}{x_r^e\beta\mu_r^e}\right) + B^{-1}\right)$ has positive eigenvalues, because it is the product of two positive definite matrices [19];
- $E\tilde{D}^{-1}$ has positive eigenvalues, because it is similar to $diag(\beta x_r^e\mu_r^e)\left(diag\left(\frac{f_r(x^e)}{\beta\mu_r^e}\right) + B^{-1}\right)$.

Finally, $\tilde{D}^{-1}E$ has positive eigenvalues and hence the system in (16) is exponentially stable for arbitrary $\beta > 0$. ∎

C. Proof of Theorem 2

Proof: First, any equilibrium (x^e, w^e) of the system in (11) must lie on the equilibrium manifold defined by (14). We also know that on this manifold the entire system collapses to a lower-dimension reduced system shown in (13). Therefore, it is equivalent to investigate the reduced system for the existence and uniqueness of the equilibrium.

Here, we apply the Poincare-Hopf Index Theorem to claim that at least one equilibrium exists in the reduced system; then we apply Lemma 2 to conclude that the number of equilibria must be one.

Fact 1: (*Poincare-Hopf Index Theorem*) Let $\mathcal{D}$ be an open subset of $\mathbb{R}, N > 0$, and $\nu : \mathcal{D}^N \to \mathbb{R}^N$ be a smooth vector field, with nonsingular Jacobian matrix $\partial\nu/\partial p$ at every equilibrium p. If there is a $\mathcal{G} \subseteq \mathcal{D}^N$ such that every trajectory moves inward of region $\mathcal{G}$, then the sum of the indices of the equilibria in $\mathcal{G}$ is $(-1)^N$.

To apply Poincare-Hopf Index Theorem, we need to construct a proper vector field and the corresponding region $\mathcal{G}$. For the reduced system, it is equivalent to investigate either $w(t)$ or $x(t)$ as they are connected by a one-to-one mapping.

We claim that the vector field defined by

$$\nu(w(t)) := \dot{w}(t) = c\left([w_r^o]_{r\in R} - [w_r(t) f_r(x(t))]_{r\in R}\right) \tag{17}$$

is the one we want. To see that, first note $\nu(w(t))$ can be expressed as a function of $x(t)$, the Jacobian matrix can be expressed as

$$\partial\nu/\partial w = \partial\nu/\partial x \cdot \partial x/\partial w.$$

We have shown that if x^e is an equilibrium of system in (13), then x^e is locally stable, indicating $\partial\nu/\partial x$ is nonsingular at the equilibrium. Also remember that x and w are related by a one-to-one mapping, hence $\partial x/\partial w$ is nonsingular. It follows that $\partial\nu/\partial w$ is nonsingular at the equilibrium.[5]

We now start to construct the region $\mathcal{G}$. First note the following facts:

- if route r is not congested, $g_j(y_j(t)) \le \frac{1}{\beta}\ln 2$; so

  $$x_r(t) \ge \frac{w_r(t)}{\sum_{j\in r}(\epsilon_j + \frac{1}{\beta}\ln 2)}.$$

  As we increase $w_r(t)$, $x_r(t)$ will eventually reach the value $\min_{j\in r} C_j$ and route r will be congested (the existence of cross traffic can only add to the congestion of the route). Hence we claim that if $w_r(t)$ is sufficiently large, the route will be congested, regardless of the traffic pattern in the network.
- if route r is congested, at least on one link j of the route, the aggregate arriving rate $y_j(t)$ exceeds the link capacity C_j, therefore

  $$f_r(x(t)) \ge 2\frac{e^{\ln 2}}{1 + e^{\ln 2}} - 1 = 1/3.$$

  Hence $w_r^o - w_r(t) f_r(x(t)) < w_r^e - w_r(t)/3$ as long as route r is congested.

It follows that, as $w_r(t)$ becomes sufficiently large, the route r must be congested. There must exist one w_r^{max} such that $w_r^o - w_r^{max} f_r(x) < w_r^o - w_r^{max}/3 < 0$. Then the region $\mathcal{G}$ can be defined as

$$\mathcal{G} = [0, w_1^{max}] \times [0, w_2^{max}] \cdots [0, w_{\mathrm{card}(R)}^{max}].$$

We check the value of the vector field on the boundary of $\mathcal{G}$:

- if $w_r(t) = 0$, then $\dot{w}_r(t) > 0$, according to (17).
- if $w_r(t) = w_r^{max}$, then by the definition of w_r^{max} and according to (17), $\dot{w}_r(t) < 0$.

Therefore, every point on the boundary $\mathcal{G}$ will move inward.

Before we use the Poincare-Hopf Index Theorem, the following Lemma says there are only finite number of equilibria inside $\mathcal{G}$.

Lemma 3: Let M denote the number of equilibriums inside $\mathcal{G}$, and $0 < w_i^{eq} < w^{max}$ represents the ith equilibrium, then $M < \infty$.

Proof: Any equilibrium is locally exponentially stable by Lemma 2, and hence is locally unique in an open set around it. The set of equilibria, denoted by $\mathcal{E} = \{w^{eq} : w_r^o - w_r^{eq} f_r(w^{eq}) = 0, r \in R\}$, is closed and bounded (i.e., compact) since the $w_r^{eq} f_r(w^{eq})$ is continuous and w^{eq} is bounded. The union of those disjoint open sets, each including one locally unique equilibrium $w^{eq} \in \mathcal{E}$, forms a covering of $\mathcal{E}$. By [20], we claim the number of these disjoint open sets must be finite. Therefore M is finite. ■

Hence by the Poincare-Hopf Index Theorem, and noticing that M is finite, we have the following equations, indicating that there is at least one equilibrium inside region $\mathcal{G}$ and

$$Index(\mathcal{G}) = (-1)^{\mathrm{card}(R)} = \sum_{i=1}^{M} Index(w_i^{eq}),$$

where we have again used the quantity $\mathrm{card}(R)$, the dimension of $w(t)$.

But every w_i^{eq} is locally stable, hence the Jocobian matrix at the equilibrium w_i^{eq}, denoted by $J(w_i^{eq})$, has all its eigenvalues be negative. Therefore

$$Index(w_i^{eq}) = sgn(Det(J(w_i^{eq}))) = (-1)^{\mathrm{card}(R)}.$$

Therefore, we can see these two equations imply $M = 1$. Together with the fact that any point outside $\mathcal{G}$ can not be an equilibrium, we conclude there is only one equilibrium for system in (13).

Finally, as the reduced order system has only one unique equilibrium on the equilibrium manifold, we conclude the system (11) has a unique equilibrium, for arbitrary $\beta > 0$. ■

[5] $rank(A) + rank(B) - k \le rank(AB) \le \min(rank(A), rank(B))$.