Cross-Layer Scheduling for Cooperative Multi-Hop Cognitive Radio Networks

Dongyue Xue, Eylem Ekici
Department of Electrical and Computer Engineering, Ohio State University, Columbus, OH 43202, USA
xued@ece.osu.edu, ekici@ece.osu.edu

ABSTRACT
The paper aims to design cross-layer optimal scheduling algorithms for cooperative multi-hop Cognitive Radio Networks (CRNs), where secondary users (SUs) assist primary users (PUs)’s multi-hop transmissions and in return gain authorization to access a share of the spectrum. We build two models for two different types of PUs, corresponding to elastic and inelastic service classes. For CRNs with elastic service, the PU maximizes its throughput while assigning a time-share of the channel to SUs proportional to SUs’ assistance. For the inelastic case, the PU is guaranteed a minimum utility. The proposed algorithm for elastic PU model can achieve arbitrarily close to the optimal PU throughput, while the proposed algorithm for inelastic PU model can achieve arbitrarily close to the optimal SU utility. Both algorithms provide deterministic upper-bounds for PU queue backlogs. In addition, we show a tradeoff between throughput/utility and PU’s average end-to-end delay upper-bounds for both algorithms. Furthermore, the algorithms work in both backlogged as well as arbitrary arrival rate systems.

Keywords
Congestion control, network scheduling, multi-hop wireless networks, cognitive radio networks, end-to-end delay guarantees

1. INTRODUCTION
In traditional networks, spectrum bands or channels are allocated to licensed users. However, such fixed spectrum assignment gives rise to the spectrum under-utilization problem as was reported by Federal Communication Commission (FCC) [1]. Cognitive Radio Networks (CRNs) [2] have recently emerged as a technology for unlicensed users, referred to as secondary users (SUs), to opportunistically utilize the spectrum assigned to licensed users, referred to as primary users (PUs). Researchers have been working on optimizing data rate and throughput of CRNs in single-hop settings [3]-[6]. However, these works are not readily extendable to multi-hop CRNs, since multi-hop transmission requires that the CR policies take into account scheduling and routing issues.

Back-pressure scheduling algorithms with Lyapunov optimization tools have been extensively investigated for generic wireless networks [11]-[12]. In addition to the seminal work [11], distributed and low-complexity algorithms have been proposed in the literature such as [13]-[15]. This technique has been applied to CRNs in [7]-[10]. Specifically, in [7], an optimal cross-layer scheduling algorithm has been proposed in a single-hop setting to maximize SU throughput subject to PU collision constraints. This single-hop setting is extended in [8], where aggregated utility is maximized subject to PU power constraints. In [9], a cooperative CRN is considered to optimize PU and SU utility, where SUs assist PU transmission in a two-hop relay scenario, which is not readily extendable to generic multi-hop CRNs. A multi-hop CRN scheduling algorithm is proposed in [10], without considering cooperation between PUs and SUs. To the best of our knowledge, no throughput/utility-optimal scheduling algorithms have been proposed in the literature for cooperative multi-hop CRNs.

In this paper, we propose two optimal cross-layer scheduling algorithms for a multi-hop cooperative CRN, where SUs relay data for a PU pair to gain access to the licensed spectrum. These two algorithms aim to solve the throughput/utility maximization problem under the so-called inelastic and elastic PU models. In the inelastic PU model, the PU pair is guaranteed a minimum utility and the SU utility is maximized. In this model, we consider an adaptive-routing scenario where the routes of the PU flow are not determined a priori, which is more general than a fixed-routing scenario. In the elastic PU model, the PU throughput is maximized using fixed routes while the SUs are guaranteed a throughput proportional to the PU data that they relay.

Salient contributions of our work with respect to the literature can be listed as follows: (1) Both inelastic and elastic algorithms can achieve a throughput/utility arbitrarily close to the optimal values. (2) The algorithms guarantee deterministically upper-bounded finite buffer sizes for PU queues in the CRN. (3) We identify a tradeoff between the throughput/utility and the average end-to-end delay upper-bounds for PU data: the inelastic algorithm achieves a PU delay upper-bound of order $O(\frac{N}{\epsilon})$, i.e., polynomial delay [24] is achieved, where $N$ denotes the number of nodes involved in PU relay and $\epsilon$ characterizes the difference between the achieved utility and the optimal utility; The elastic algo-
rithm achieves order optimal delay \cite{21,22}, i.e., the delay is upper-bounded by the first order of the number of hops in a route. (4) Both algorithms are extended from a backlogged source model to a model with arbitrary arrival rates at transport layer.

The rest of the paper is organized as follows: Section 2 introduces the network and PU models for the cooperative multi-hop CRN. In Section 3, we propose and analyze the inelastic algorithm. The elastic algorithm and its performances are provided in Section 4. In Section 5, we extend both algorithms to the model with arbitrary arrival rates at transport layer. We conclude our work in Section 6.

2. NETWORK MODEL

In this section, we first present the overall multi-hop cooperative CRN model, followed by analysis of the two PU models.

2.1 Overall Network Elements and Constraints

In this paper, we consider a multi-hop cooperative CRN where SUs relay PU data in return for the right to use the wireless spectrum. The multi-hop cooperative CRN in question can be divided into two subnetworks: a “PU relay subnetwork” and an “SU subnetwork”. The PU relay subnetwork is composed of one primary source node \((s_p, d_p)\), a corresponding primary destination node \((d_p)\), and a set of SUs \(S\) that relay the PU traffic between \(s_p\) and \(d_p\) over possibly multiple hops, where \(|S| = N\). We assume that the channel condition cannot support direct transmission between the PU pair, and thus PU data will be solely relayed by secondary nodes. Denote the node set of the PU relay subnetwork as \(N\). Then, the SU subnetwork is represented by \((N, \mathcal{L})\). Note that our model is readily extendable to the scenario of multiple PU pairs.

The SU subnetwork is composed of the set of SUs \(S\) that participate in PU data relaying, and the set of their one-hop secondary neighbors \(S'\) with which they communicate. For notational simplicity, we assume that \(S \cap S' = \emptyset\) and that there is a distinct SU \(l' \in S'\) that corresponds to every SU \(l \in S\). Then, the SU subnetwork is represented by \((S \cup S', \mathcal{L}')\), where \(\mathcal{L}' = \{(l, l'): l \in S, l' \in S'\}\) is the set of links in the SU subnetwork. Note that our analysis can readily be extended to cases where \(S \cap S' \neq \emptyset\).

Let \(\mathcal{V} = \mathcal{L} \cup \mathcal{L}'\). Then the CRN topology is represented by an interference graph \(G = (V, \mathcal{E})\) (or sometimes referred to as conflict graph). There is an edge in \(\mathcal{E}\) between two links in \(\mathcal{V}\) if the links interfere with each other when scheduled simultaneously. Furthermore, let \(\mu_{mn}\) be the scheduled link rate for PU data over link \((m, n) \in \mathcal{L}\), and denote the scheduled SU link rate as \(s_l\) over link \((l, l') \in \mathcal{L}'\). For analytical simplicity, we assume a scheduled link rate takes a value from \([0, 1]\). A link scheduled representation by a vector \((\mu_{mn}, s_l)\) is said to be feasible if any pair of two scheduled links does not belong to the interference edge set \(\mathcal{E}\). Let \(\mathcal{I}\) be the set of feasible link schedules. Then, a feasible link scheduler chooses a feasible link schedule \((\mu_{mn}(t), s_l(t))\) \(\in \mathcal{I}\) for each time slot \(t\). In addition, we assume that each node only possesses one transceiver that can only send or receive data from one neighbor node. Thus, \(\forall n \in N \setminus \{s_p\}\), the following inequality holds:

\[
\sum_{j \in (n, n) \in \mathcal{E}} \mu_{jn}(t) + \sum_{i \in (n, i) \in \mathcal{L}} \mu_{ni}(t) + 1_{\{n \in S\}} s_n(t) \leq 1, \forall t, (1)
\]

where \(1_{\{x\}}\) is the indicator function for event \(x\). Note that since \(s_p\) is the sender of the PU pair, we must have

\[
\sum_{n \in S} \mu_{ns_p} = 0, \forall t.
\]

In the following two subsections, we build two PU models corresponding to different PU service classes, namely, inelastic PU model and elastic PU model. In the inelastic PU model, adaptive-routing scenarios are considered and we maximize SU utility while PU is guaranteed a minimum utility. In the elastic model, we assume fixed-routing scenarios and maximize the PU throughput while SUs obtain a throughput proportional to the PU data that they relay.

2.2 Queueing Structure and Constraints for Inelastic PU Model

In the inelastic PU model, we denote by \(U_n(t)\) the queue backlog for PU packets at node \(n \in N\), where \(U_{dp}(t) = 0\) \(\forall t\). Let \(Q_l(t)\) be the queue backlog for PU packets corresponding to the SU pair associated with \(l \in S\). Now we define the stability of a generic queue with queue backlog \(X(t)\): \(X(t)\) is said to be stable if

\[
\limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} E[X(t)] < \infty.
\]

Therefore, the network is stable if queues \(U_n(t)\) and \(Q_l(t)\) are stable \(\forall n \in N\) and \(\forall l \in S\).

For the time being, we assume that PU and SU traffics are backlogged at the transport layer. Thus, a congestion controller is needed to admit packets into the network layer. Let \(\mu_{ps_p}(t)\) be the admitted PU arrival rate in time slot \(t\). Note that we can consider \(p\) in the subscript of \(\mu_{ps_p}(t)\) as the virtual node representing the PU transport layer and consider \((p, s_p)\) as the virtual link from transport layer to source PU, so we construct a new link set as \(\mathcal{L}^+ \triangleq \mathcal{L} \cup \{(p, s_p)\}\). Let \(A_l(t)\) denote the admitted SU arrival rates to the SU pair associated with secondary node \(l \in \mathcal{L}\). We assume \(\mu_{ps_p}(t) \leq M_p\) and \(A_l(t) \leq M_A\) \(\forall l \in S\), where \(M_p\) and \(M_A\) are the upper-bounds for admitted PU and SU arrival rates, respectively. For analytical simplicity, we assume that admitted packets are added to the queues at the end of time slot \(t\).

From the above analysis, we can develop the queueing dynamics for \(U_n(t)\), \(n \in N \setminus \{s_p\}\), as follows:

\[
U_n(t+1) \leq [U_n(t) - \sum_{i \in (n, i) \in \mathcal{L}} \mu_{ni}(t)]^++ \sum_{j \in (j, n) \in \mathcal{L}} \mu_{jn}(t), (2)
\]

where \([x]^+ = \max\{x, 0\}\) and \(\sum_{i \in (n, i) \in \mathcal{L}} \mu_{ni}(t)\) stands for the scheduled service rate. Note that \([x]^+\) is an inequality since a feasible scheduler can be designed independent of the queue backlog information. Specifically, the inequality holds when the actual arrival rate at node \(n\) is less than the scheduled arrival rate \(\sum_{j \in (j, n) \in \mathcal{L}} \mu_{jn}(t)\), i.e., some neighbor node \(j\) does not have packets for the scheduled transmission \(\mu_{jn}(t) = 1\). Similarly, \(Q_l(t)\) evolves as follows:

\[
Q_l(t+1) = [Q_l(t) - s_l(t)]^+ + A_l(t), (3)
\]

\[
\sum_{j \in (j, n) \in \mathcal{L}} \mu_{jn}(t) \leq \mu_{ps_p}(t), \forall t\]
We denote by \( f(x) \) and \( g_l(x) \) with \( l \in S \), respectively, the PU and SU utility functions of the time-average transmission rate. As convention, we assume that the utility functions are positive-valued, concave, strictly increasing and continuously differentiable, with \( f(0) = 0 \) and \( g_l(0) = 0 \) \( \forall \ l \in S \). Examples of utility functions include \( \theta' \log(1 + x) \) and \( \theta' x \), where \( \theta' > 0 \) is a weight for the utility functions. We assume the inelastic PU imposes a minimum utility constraint \( a_P \), i.e., the utility of the time-average PU transmission rate must be greater than or equal to \( a_P \).

According to [11][12], we define the capacity region \( \Lambda_1 \) of the inelastic CRN as the closure of all feasible arrival rate vectors consisting of an admitted PU arrival rate and \( N \) admitted SU arrival rates, where each feasible arrival rate vector is storable by some scheduler. Without loss of generality, we assume that there exists an SU rate vector \( (r_l)_{l \in S} \) such that \( f^{-1}(a_P), (r_l)_{l \in S} \) is strictly inside \( \Lambda_1 \), where \( f^{-1}(x) \) is the inverse function of the utility function \( f(x) \). To assist the analysis, we let \( (r^*_l)_{l \in S} \) be a solution to the following optimization problem:

\[
\max_{(r_l)_{l \in S} ; f^{-1}(a_P) + \varepsilon \subset (r_l)_{l \in S} \subset \Lambda_1} \sum_{l \in S} g_l(r_l),
\]

where \( \varepsilon > 0 \) can be chosen arbitrarily small. Then according to [13], we have:

\[
\lim_{\varepsilon \to 0} \sum_{l \in S} g_l(r^*_l, \varepsilon) = \sum_{l \in S} g_l(r^*_l),
\]

where \( (r^*_l)_{l \in S} \) is a solution to the following optimization:

\[
\max_{(r_l)_{l \in S} ; f^{-1}(a_P) \subset (r_l)_{l \in S} \subset \Lambda_1} \sum_{l \in S} g_l(r_l).
\]

In Section 3, we will propose an algorithm that satisfies the PU minimum utility constraint and can achieve SU utility arbitrarily close to the optimal value \( \sum_{l \in S} g_l(r^*_l) \), with a tradeoff between the SU utility and the average PU delay upper-bound.

### 2.3 Routing and Queueing Structure for Elastic PU Model

For the elastic PU model, we consider a fixed multi-path routing scenario, where the PU data transmission have \( K \) loopless pre-determined routes. We denote the path for the \( k \)-th route as \( P_k = (v_k^0, v_k^1, ..., v_{hk+k}^1, v_{hk+k+1}^1) \), where \( (H_k + 1) \) is the total number of hops in the PU relay subnetwork for route \( k \), where \( v_m^m \in N, \forall m \in \{0, 1, ..., H_k + 1\}, \forall k \in \{1, 2, ..., K\} \). Without loss of generality, we assume that each node \( l \in S \) is in at least one of the \( K \) routes, that is: \( \forall l \in S, \exists_k m \text{ s.t. } v_m^m = l \). Note that we always have \( v_0^k = s_P \) and \( v_{hk+k+1}^1 = d_P \), \( \forall k \in \{1, 2, ..., K\} \).

According to this routing structure, we construct PU queues \( U_k^k(t) \) along the nodes in the \( K \) routes, where \( 0 \leq m \leq H_k + 1 \) and \( 1 \leq k \leq K \). Note that, since \( v_{hk+k+1}^1 = d_P \), we have \( U_k^k(t) = 0 \), \( \forall t \), \( \forall k \in \{1, 2, ..., K\} \).

Similar to the inelastic model, we assume that PU and SU traffics are backlogged at the transport layer. Let \( \mu_{k,1,0}(t) \) be the admitted arrival rate from the PU transport layer to the source PU that is scheduled to pass through the \( k \)-th route. Note that, consistent with the elastic model, we assume that the sum of admitted PU arrival rates over \( K \) routes is upper-bounded by \( \mu_M \), i.e.,

\[
\sum_{k=1}^K \mu_{k,1,0}(t) \leq \mu_M, \forall t.
\]

In addition, we let \( \lambda_k, k = \{1, 2, ..., K\} \), be the time-average of \( \mu_{k,1,0}(t) \). Let \( \mu_{m,m+1}^k(t) \), \( 0 \leq m \leq H_k \), be the scheduled rate for the hop \((v_m^m, v_{m+1}^m)\) along the \( k \)-th path. Thus, \( U_k^k(t) \) evolves as follows:

\[
U_k^k(t+1) = \left[ U_k^k(t) - \mu_{m,m+1}^k(t)+ \mu_{m-1,m}^k(t), \ 0 \leq m \leq H_k \right],
\]

where the inequality holds if \( \mu_{m,1}^k = 1 \) and \( U_k^k(t) = 0, 1 \leq m \leq H_k \). Note that a link \( (m, n) \in E \) can be a hop in multiple routes, and hence we can only schedule the hop with rate one on each such route in any time slot.

Let \( \rho_k \) be the reward for SU routes when a packet is admitted to route \( k \), i.e., \( \rho_k \mu_{k,1,0}(t) \) packets will be admitted simultaneously to the SU queues corresponding to the nodes \( v_m^m, 1 \leq m \leq H_k \). Here, we assume that \( \rho_k \mu_{k,1,0}(t) \) takes integer values. Note that our analysis is readily extendable to fractional-valued \( \rho_k \mu_{k,1,0}(t) \) by constructing a counter that only admits \( \lfloor \rho_k \mu_{k,1,0}(t) \rfloor \) packets, where \( [x] \) is the floor function. Also note that the analysis can be extended to delayed rewards, i.e., a reward rate \( \rho_k \mu_{k,1,0}(t) \) is admitted to SU queue at \( t + \tau', \tau' \) is the delay in unit of time slots.

From the above analysis, the SU queueing dynamics for \( Q(t) \) can be expressed as follows:

\[
Q(t+1) = [Q(t) - s(t)] + \frac{K}{\sum_{k=1}^{H_k} \rho_k \mu_{k,1,0}(t) \mathbf{1}_{m:m} = 1] + \sum_{k=1}^{K} \rho_k \mu_{k,1,0}(t) \mathbf{1}_{m:m} = 1 \}
\]

where the second equality holds since each route is a loop-free.

The network is stable if queues \( U_k^k(t) \) and \( Q(t) \) are stable \( \forall m, k \). Then, we define the capacity region \( \Lambda_E \) of the elastic CRN as the closure of all feasible arrival rate vectors each storable by some scheduler. Note that a feasible arrival rate vector is in the form of

\[
((\lambda_k)_{k \in \{1, 2, ..., K\}}) = \sum_{k=1}^{K} \rho_k \lambda_k \mathbf{1}_{m:m} = 1) \in \Lambda_E
\]

where \( (\lambda_k)_{k \in \{1, 2, ..., K\}} \) represents the PU arrival rates per route and \( (\sum_{k=1}^{K} \rho_k \lambda_k \mathbf{1}_{m:m} = 1) \in \Lambda_E \) represents the SU arrival rates according to the reward mechanism. To assist the analysis, we let \( (\lambda^*_k)_{k \in \{1, 2, ..., K\}} \) be a solution to the following optimization problem:

\[
\max_{(\lambda_k)_{k \in \{1, 2, ..., K\}}} \sum_{k=1}^{K} \lambda_k
\]

\[
\text{s.t. } (\lambda_k) : \left( (\lambda_k + \epsilon) \sum_{k=1}^{K} \rho_k (\lambda_k + \epsilon) \mathbf{1}_{m:m} = 1) \right) \in \Lambda_E
\]

where \( \epsilon > 0 \) can be chosen arbitrarily small. Similarly, according to [13], we have:

\[
\lim_{\epsilon \to 0} \sum_{k=1}^{K} \lambda^*_k \epsilon = \sum_{k=1}^{K} \lambda^*_k
\]
where \((\lambda_k)_{k \in \{1, 2, \ldots, K\}}\) is a solution to the following optimization:

\[
\max_{(\lambda_k)_{k \in \{1, 2, \ldots, K\}}} \sum_{k=1}^{K} \lambda_k \\
\text{subject to } (\lambda_k, (\sum_{k=1}^{K} \lambda_k) 1_{(\exists m: \psi_m^\top = 1)}) \in \Lambda E
\]

In Section 4, we will propose an algorithm that can achieve a PU throughput arbitrarily close to the optimal value \(\sum_{k=1}^{K} \lambda_k\), with a tradeoff between the PU throughput and average PU/SU delay upper-bound.

3. INELASTIC ALGORITHM FOR THE CRN

In this section, we first introduce two types of virtual queues and their structures to assist the development of the inelastic algorithm. The inelastic algorithm is then introduced in Subsection 3.2.

3.1 Virtual Queues and Approaches

We construct a virtual queue \(U_p(t)\) at the PU transport layer with the following queue dynamics:

\[
U_p(t+1) = [U_p(t) - \mu_{sp}(t)]^+ + R(t),
\]

where \(R(t)\) denotes the virtual arrival rate to \(U_p(t)\) in time slot \(t\) which will be determined by the inelastic algorithm in the next subsection. Furthermore, let \(R(t)\) be upper-bounded by \(\mu_M\). When \(U_p(t)\) is stable, we know from queueing theory that the time-average admitted PU arrival rate \(\mu\) satisfies:

\[
\mu \triangleq \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mu_{sp}(t) \geq \gamma \triangleq \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} R(t).
\]

The virtual queue \(U_p(t)\), along with \(R(t)\), regulates the admitted PU arrival rate in the inelastic algorithm, in an attempt to guarantee an average end-to-end delay upper-bound, as will be stated in detail in the next subsection.

We construct another virtual service queue \(Z(t)\) at the PU source node \(s_p\) with the following queueing dynamics:

\[
Z(t+1) = [Z(t) - R(t)]^+ + f^{-1}(a_p).
\]

When \(Z(t)\) and \(U_p(t)\) are stable, we have \(f(\mu) \geq f(\gamma) \geq a_p\).

Specifically, the minimum utility constraint imposed by PU is satisfied when the two virtual queues are stable.

3.2 Inelastic Algorithm

We design a control parameter \(q_M\) indicating the buffer size for each PU queue in the CRN, with \(q_M \geq \mu_M\). The optimal inelastic algorithm consists of four parts, namely, SU congestion controller, \(R(t)\) controller, PU congestion controller and a link scheduler, described as follows.

1) SU Congestion Controller:

\[
\min_{0 \leq A_t(t) \leq A_M} \sum_{t \in S} \sum_{l \in S} Q_l(t) + V_l g(A_t(l)), \forall t \in S,
\]

where \(V_1 > 0\) is a control parameter in the algorithm. Note that we always have \(A_t(t)Q_l(t) - V_l g(A_t(l)) \leq 0\) under the SU congestion controller, since \(A_t(t) = 0\) is a valid candidate for the admitted arrival rate.

2) \(R(t)\) Regulator:

\[
\min_{0 \leq R(t) \leq \mu_M} R(t) [U_p(t) \frac{q_M - \mu_M}{q_M} - Z(t)].
\]

Specifically, when \(U_p(t) \frac{2M - \mu_M}{q_M} - Z(t) > 0\), the virtual rate \(R(t)\) is set to zero; otherwise, \(R(t) = \mu_M\).

3) PU Congestion Controller:

\[
\min_{0 \leq \mu_{sp}(t) \leq \mu_M} \mu_{sp}(t) [q_M - \mu_M - U_{sp}(t)].
\]

Specifically, when \(q_M - \mu_M - U_{sp}(t) \leq 0\), the admitted PU arrival rate \(\mu_{sp}(t)\) is set to zero; Otherwise, \(\mu_{sp}(t) = \mu_M\).

4) Link Rate Scheduler:

\[
\max_{(m,n) \in L} \sum_{l \in S} Q_l(t) s_l(t),
\]

with the constraint \((\{\mu_{mn}(t)\}_{(m,n) \in L}, \{s_l(t)\}_{l \in S}) \in Z\).

Note that when \(U_m(t) - U_n(t) \leq 0\), \(\{s_l(t)\}_{l \in S} \in Z\).

4. The inelastic algorithm has the following property:

**Proposition 1.**

\(U_n(t) \leq q_M, \forall t \in N\).

**Proof.** We can prove Proposition 1 by induction. Initially when \(t = 0\), \(U_n(0) = 0, \forall n \in N\). Assume in time slot \(t\) we have \(U_n(t) \leq q_M, \forall t \in N\). In the induction step, we consider two cases:

Case 1: \(n = s_p\). If \(U_{sp}(t) \leq q_M - \mu_M\), then since the admitted arrival rate to \(U_{sp}(t)\) is bounded by \(\mu_M\), we have \(U_{sp}(t+1) \leq U_{sp}(t) + \mu_M \leq q_M\). Otherwise, we have \(U_{sp}(t) > q_M - \mu_M\), and according to the PU congestion controller (11) we have \(\mu_{sp}(t) = 0\), from which we obtain \(U_{sp}(t+1) \leq U_{sp}(t) \leq q_M\).

Case 2: \(n \neq s_p\). If \(U_n(t) \leq q_M - 1\), then we have \(U_n(t+1) \leq U_n(t) + 1 \leq q_M\) according to (11) and the queueing dynamics (2). Otherwise, we have \(U_n(t) = q_M\) and \(U_n(t) \geq U_m(t) \forall m \in N\), and according to the link scheduler (12) we have \(\mu_{mn}(t) = 0 \forall j\) such that \((j, n) \in L\), from which we obtain \(U_n(t+1) \leq U_n(t) = q_M\) by the queueing dynamics (2).

Therefore, \(U_n(t+1) \leq q_M \forall t \in N\), i.e., the induction step holds, and the proposition is proved.

Now we present the main results of the inelastic algorithm in Theorem 1.

**Theorem 1.** Let \(\epsilon > 0\) be chosen arbitrarily small. Given that

\[
q_M > \frac{\mu_M^2 + N + 1}{\epsilon} + \mu_M,
\]

the inelastic algorithm ensures the following inequality on queue backlog:

\[
\lim \inf_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} E\{\sum_{l \in S} Q_l(t) + U_p(t) + Z(t)\} \leq \frac{B_1 + V_1 q_M}{\delta_1},
\]

where \(B_1 \triangleq \frac{1}{2} \mu_M^2 + \frac{1}{2} (f^{-1}(\mu_M))^2 + \frac{1}{2} \sigma^2(q_M - \mu_M) + \frac{1}{2} N + \frac{1}{2} N A_M^2 + \frac{1}{2} \mu_M q_M (N + 1), \delta_1\) is chosen such that \(0 < \delta_1 < \frac{\epsilon}{2q_M - \mu_M - q_M^2}, \) and \(g_M\) is defined as:

\[
g_M \triangleq \lim \inf_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} E\{\sum_{l \in S} g_l(A_t(l))\} - \sum_{l \in S} g_l(r_{1,e}).
\]
Furthermore, the inelastic algorithm achieves:

$$\sum_{l \in S} g_l(a_l) \geq \sum_{l \in S} g_l(r^*_l) - \frac{B_l}{V_l},$$  \quad (16)

where $a_l$ is defined as the time-average ensemble value of $A_l(t)$:

$$a_l \doteq \liminf_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} E[A_l(t)], \quad l \in S.$$

**Remark 1 (Network Stability):** The inequalities (13) from Proposition 1 and (15) from Theorem 1 indicate that the inelastic algorithm stabilizes the actual and virtual queues. As an immediate result, the network is stable and the minimum utility constraint is met. In addition, Proposition 1 ensures that the actual PU queues are deterministically bounded by the finite buffer size $q_M$.

**Remark 2 (Optimal Utility and Tradeoff with Delay):** The inequality (16) gives the lower-bound of the SU utility the inelastic algorithm can achieve. Since the constant $B_l$ is independent of the control parameter $V_l$, the algorithm can achieve a utility arbitrarily close to the optimal value $\sum_{l \in S} g_l(r^*_l)$ as $\epsilon$ is chosen arbitrarily small, with a tradeoff in the PU buffer size $q_M$ which is of order $O(\frac{H^2}{\epsilon})$ as shown in (14). By Little’s Theorem, the average end-to-end delay upper-bound is of order $O(\frac{H^2}{\epsilon})$. Note that it is easy to verify when applied to a fixed-routing scenario, the inelastic algorithm achieves an average end-to-end delay upper-bound of order $O(\frac{H^2}{\epsilon})$, where $H$ denotes the number of hops in the route.

**Remark 3 (Complexity of Algorithm):** In the inelastic algorithm, the SU congestion controller, the $R_l$ regulator and the PU congestion controller can operate locally at SU transport layer and source PU. The link rate scheduler is essentially a centralized maximal weight matching problem [11,16]. To reduce complexity of the link rate scheduler, suboptimal algorithms can be developed to at least achieve a fraction $\gamma$ of the optimal utility. These suboptimal algorithms include the well-studied Greedy Maximal Matching (GMM) [15] algorithm with $\gamma = \frac{1}{2}$ and the maximum weighted independent set (MWIS) problem such as GW-MAX and GWMIN proposed in [20] with $\gamma = \frac{1}{\Delta}$, where $\Delta$ is the maximum degree of the CRN topology.

**Remark 4 (Distributed Implementation of the Link Scheduler):** Distributed implementation can be developed in much the same way as in [14] to achieve a fraction of the optimal utility. In order to achieve a utility arbitrarily close to the optimal value with distributed implementation, we can employ random access techniques [21,23] in the link scheduler with fugacities [25] chosen as $\exp(\alpha Q_l(t))$ for link $(m,n) \in \mathcal{L}$ and $\exp(\alpha Q_l(t))$ for an SU link associated with $l \in \mathcal{S}$, where $U_p(t)$ is a local estimate of $U_p(t)$ and $\alpha$ is a positive weight. It can be shown that the distributed algorithm can still achieve an average PU end-to-end delay of order $O(\frac{H^2}{\epsilon})$ with the time-scale separation assumption [22,24]. Due to limited space, a detailed discussion is omitted.

We prove Theorem 1 in the next subsection.

### 3.3 Proof of Theorem 1

Before we proceed, we present Lemma 1 as follows to assist us in proving Theorem 1.

**Lemma 1:** For any feasible rate vector $\theta$, there exists a stationary randomized algorithm SI that stabilizes the network with SU admitted arrival rate $A_S(t) = r_t$, for $\forall t \in \mathcal{S}$, and PU admitted arrival rate $\mu_S(t) = \theta$, for $\forall t \in \mathcal{S}$, and schedule $\{(\mu_{mn}(t))_{(m,n) \in \mathcal{L}}, (\mu_S(t))_{t \in \mathcal{S}}\}$ independent of queue backlogs satisfying:

$$E\left\{ \sum_{i \in \mathcal{L}} \mu_i^S(t) - \sum_{j \in \mathcal{N}} \mu_j^S(t) \right\} = 0, \quad \forall t, \forall n \in \mathcal{N};$$

$$E\{ \nu^S(t) \} = r_t, \quad \forall t, \forall l \in \mathcal{S}.$$

Note that it is not necessary for the randomized algorithm SI to provide finite buffer size or delay guarantees. Similar formulations of stationary randomized algorithms and existence proofs have been presented in [3][11][13], so we omit the proof of Lemma 1 for brevity.

**Remark 5:** According to the SI algorithm in Lemma 1, we assign the virtual input rate as $R^S_l(t) = \mu^S_{p,t}(t) = \theta$, for $\forall t$. Hence, the time average of $R^S(t)$ satisfies $\bar{R}^S(t) = \theta$. Note that $\theta$, $r_{l,i}$ can take values as $(f^{-1}(a) + \frac{\epsilon}{2^t}, (r^*_l, \epsilon)_{i \in \mathcal{S}})$. We define the queue vector $Q_l(t)$ as:

$$Q_l(t) = ((U_{n,t})_{t \in \mathcal{N}}, (Q_l)_{l \in \mathcal{S}}, U_p(t), Z(t))$$

and define the Lyapunov function $L_l(Q_l(t))$ as follows:

$$L_l(Q_l(t)) \doteq \frac{1}{2} \left( \sum_{l \in \mathcal{S}} Q_l(t)^2 + \frac{q_M - \mu_M}{q_M} U_p(t)^2 + Z(t)^2 + \sum_{n \in \mathcal{N}} U_n(t)^2 U_p(t) \right),$$

where the last term of the above Lyapunov function takes a similar form as in [17][18]. Then, the corresponding Lyapunov drift is defined by:

$$\Delta_l(t) \doteq E\{L_l(Q^l(t + 1)) - L_l(Q_l(t))\}Q_l(t).$$

By squaring both sides of the queueing dynamics [2][3][4][8] and through algebra, we can obtain:

$$\Delta_l(t) - V_l \sum_{l \in \mathcal{S}} E\{g_l(A_l(t))Q_l(t)\} \leq B_l + \frac{\mu^2 + N + 1}{2q_M} U_p(t) - \sum_{n \in \mathcal{N}} E\left\{ \frac{U_n(t)U_p(t)}{q_M} \sum_{i \in \mathcal{S}, j \in \mathcal{N}} \mu_{ni}(t) \right\} - \sum_{j \in \mathcal{N}} \mu_{jn}(t)Q_j(t) - \bar{E}\{Z(r_{l,i}) - f^{-1}(a_{l,i})\}Q_l(t) - \frac{q_M - \mu_M}{q_M} U_p(t)(\mu_{p,t}(t) - R(t))Q_l(t) - \bar{E}\{ \sum_{l \in \mathcal{S}} Q_l(t)A_l(t) - A_l(t)\}Q_l(t) - V_l \sum_{l \in \mathcal{S}} E\{g_l(A_l(t))Q_l(t)\},$$

for $\forall l \in \mathcal{S}$.
where we also employ the following inequalities:

\[
\sum_{n \in N} \frac{U_n(t+1)^2U_p(t+1)}{q_M} \leq \left( \frac{R(t) + U_p(t)}{q_M} \right) \sum_{n \in N} U_n(t+1)^2
\]

\[
\leq \mu_M q_M (N + 1) + \frac{U_p(t)}{q_M} (\mu_M^2 + N + 1) + \frac{U_p(t)}{q_M} \sum_{n \in N} U_n(t)^2
\]

\[
-2 \frac{U_p(t)}{q_M} \sum_{n \in N} U_n(t) \left( \sum_{i \in \{n,i\} \in \mathcal{L}} \mu_{ni}(t) - \sum_{j \in \{n,j\} \in \mathcal{L}} \mu_{jn}(t) \right).
\]

Through algebra, we find the equivalent of (17):

\[
\Delta_l(t) - V_i \sum_{l \in S} \mathbb{E}\{g_i(A_i(t))|Q_l(t)\}
\]

\[
\leq B_1 + \frac{\mu_M^2 + N + 1}{2q_M} U_p(t) + f^{-1}(a_{\mu}) Z(t)
\]

\[
+ \sum_{l \in S} \mathbb{E}\{A_i(t)Q_l(t) - V_i g_i(A_i(t))|Q_l(t)\}
\]

\[
+ \mathbb{E}\{R(t) \{q_M - \mu_M \} U_p(t) - Z(t)\}|Q_l(t)\}
\]

\[
- \mathbb{E}\{\mu_{p_p}(t) \frac{U_p(t)}{q_M} (q_M - \mu_M - U_p(t))|Q_j(t)\}
\]

\[
- \mathbb{E}\{\sum_{l \in S} Q_l(t) s_i(t)\}
\]

\[
+ \sum_{(m,n) \in \mathcal{L}} \mu_{mn}(t) \left( \frac{U_m(t) - U_n(t)}{q_M} \right) |Q_l(t)\}.
\]

Note that the last four terms of the RHS of (18) are minimized by the SU congestion controller (9), the R(t) regulator (10), the PU congestion controller (11), and the link scheduler (12), respectively, over a set of feasible algorithms including the stationary randomized algorithm SI introduced in Lemma 1 and Remark 5. Then, we substitute into the fourth and fifth lines of the RHS of (18) (i.e., the third and fourth lines of (13)) a stationary randomized SI with admitted arrival rate vector \((f^{-1}(a) + \epsilon, (r_{i,n})_{i \in S})\), and we substitute into the last two terms the SI with admitted arrival rate vector \((f^{-1}(a) + \epsilon, (r_{i,n}^* + \epsilon)_{i \in S})\). After the above substitutions, we obtain:

\[
\Delta_l(t) - V_i \sum_{l \in S} \mathbb{E}\{g_i(A_i(t))|Q_l(t)\}
\]

\[
\leq B_1 + \frac{\epsilon (q_M - \mu_M) - \mu_M^2 - N - 1}{2q_M} U_p(t)
\]

\[
- \epsilon \sum_{l \in S} Q_l(t) - \frac{\epsilon}{2} Z(t) - V_i \sum_{l \in S} g_i(r_{i,n}^*)
\]

\[
\leq B_1 - \delta l \sum_{l \in S} Q_l(t) + U_p(t) + Z(t) - V_i \sum_{l \in S} g_i(r_{i,n}^*),
\]

where the second inequality holds when the condition \([\mu_{mn}(t)]_{(m,n) \in \mathcal{L}}\) in Theorem 1 is satisfied.

We take the expectation of both sides of (19) over \(Q_l(t)\) and take the time average on \(t = 0, 1, ..., T - 1\), which leads to

\[
\frac{\delta l}{T} \sum_{t=0}^{T-1} \mathbb{E}\{\sum_{l \in S} Q_l(t) + U_p(t) + Z(t)\}
\]

\[
\leq B_1 + \frac{\epsilon}{T} \sum_{t=0}^{T-1} \mathbb{E}\{\sum_{l \in S} g_i(A_i(t))\} - V_i \sum_{l \in S} g_i(r_{i,n}^*).
\]

By taking limsup of \(T\) on both sides of (20), we can prove (18). We can prove (16) by taking the limit of \(T\) on both sides of (20) and by employing the following fact from the concavity of the SU utility functions:

\[
\sum_{l \in S} g_i(A_i(t)) \geq \sum_{l \in S} \mathbb{E}\{g_i(A_i(t))\}.
\]

Therefore, Theorem 1 is proved.

### 4. ELASTIC ALGORITHM FOR THE CRN

In this section, we design the optimal elastic algorithm composed of two parts, namely, PU congestion controller and a hop/link scheduler, described in Subsection 4.1. Note that according to the fixed-routing structure in PU relay subnetwork introduced in Subsection 2.1, when developing the scheduler, we focus on the hop/link schedule \((\mu_{mn}(t)_{m,n})_{(m,n) \in \mathcal{L}}\) which is composed of a PU hop schedule and an SU link schedule. Note that each hop schedule \((\mu_{mn}(t)_{m,n})_{(m,n) \in \mathcal{L}}\) corresponds to a PU link schedule \((\mu_{mn}(t))_{(m,n) \in \mathcal{L}}\).

#### 4.1 Elastic Algorithm

1. **PU Congestion Controller:**

\[
\min_{k=1}^{K} \mu_{k-1,0}(t) \sum_{l \in S} Q_l(t) 1_{\{\exists m: v_{i,n}^* = l\}}
\]

\[
+ U_{i,n}^k(t) - V_2
\]

\[
\text{s.t. } \sum_{k=1}^{K} \mu_{k-1,0}(t) \leq \mu_M,
\]

where \(V_2\) is a control parameter in the algorithm. For time slot \(t\), define \(k^\top(t) = \arg \min_{k} \{\mu_{k-1,0}(t) \sum_{l \in S} Q_l(t) 1_{\{\exists m: v_{i,n}^* = l\}} + U_{i,n}^k(t)\}\). Specifically, from (21), we set

\[
\mu_{k-1,0}(t) = \left\{ \begin{array}{ll}
\mu_M, & \text{if } \delta k \sum_{l \in S} Q_l(t) 1_{\{\exists m: v_{i,n}^* = l\}} + U_{i,n}^k(t) \leq V_2,

0, & \text{otherwise}.
\end{array} \right.
\]

For \(k \neq k^\top\), we set \(\mu_{k-1,0}(t) = 0\).

2. **Hop/Link Scheduler:**

\[
\max_{k,m=0}^{K} \sum_{k=1}^{K} \mu_{m,m+1}(t)(U_{m}^k(t) - U_{m+1}^k(t))
\]

\[
+ \sum_{l \in S} Q_l(t) s_i(t),
\]

s.t. \(\{ (\mu_{mn}(t))_{(m,n) \in \mathcal{L}}\} \) in \(T\),

where the optimization is taken over all feasible \((\mu_{mn}(t))_{(m,n) \in \mathcal{L}}\) and we note that each hop schedule \((\mu_{mn}(t))_{(m,n) \in \mathcal{L}}\) corresponds to a PU link schedule \((\mu_{mn}(t))_{(m,n) \in \mathcal{L}}\). From (22), when \(U_{m}^k(t) - U_{m+1}^k(t) \leq 0\), \(m \in \{0,1, ..., H_k\}\), we set \(\mu_{m,m+1}(t) = 0\).

The elastic algorithm has the following property:
Proposition 2. \( \forall m \in \{0, 1, ..., H_k\}, \forall k \in \{1, 2, ..., K\} \), the following inequality holds:

\[
U^k_m(t) \leq U^k_M \triangleq \mu M + V_2.
\]  

(23)

Proof. Similar to the proof of Proposition 1, we prove Proposition 2 by induction. Initially when \( t = 0 \), \( U^k_m(0) = 0 \) \( \forall m, \forall k \). Now assume in time slot \( t \) we have \( U^k_m(t) \leq U^k_M, \forall m, \forall k \). In the induction step, we consider two cases: 

Case 1: \( m = 0 \). Given any route \( k \), if \( U^k_0(t) \leq V_2 \), then we have \( U^k_0(t + 1) \leq U^k_0(t) + \mu M \leq U^k_M \) according to queuing dynamics (4), where we recall that \( \mu^k_{1,0}(t) \leq \mu M \) from the constraint in PU congestion controller (21). Otherwise, we have \( V_2 < U^k_0(t) \leq U^k_M \), and hence we have

\[
\rho_k \sum_{l \in S} Q_l(t) \text{1}_{\{\exists m: v^m_k = l\}} + U^k_0(t) > V_2,
\]

which induces \( \mu^k_{1,0}(t) = 0 \) from the PU congestion controller (21), and it follows that \( U^k_0(t + 1) \leq U^k_0(t) \leq U^k_M \) by the queuing dynamics (4). 

Case 2: \( m \in \{1, 2, ..., H_k\} \), for any given route \( k \). If \( U^k_m(t) \leq U^k_M - 1 \), then we have \( U^k_m(t + 1) \leq U^k_m(t) + 1 \leq U^k_M \) according to queuing dynamics (4). Otherwise, we have \( U^k_m(t) = U^k_M \geq U^k_m(t - 1) \) and according to the hop/link scheduler we have \( \mu^k_{m-1,m}(t) = 0 \), from which we have \( U^k_m(t + 1) \leq U^k_m(t) = U^k_M \) by the queuing dynamics (4).

Therefore, \( U^k_m(t + 1) \leq U^k_M \) \( \forall m \in \{0, 1, ..., H_k\}, \forall k \in \{1, 2, ..., K\} \), i.e., the induction step holds, and the proposition is proved.

As a complement to Proposition 2, recall that given route \( k \), we always have \( U^k_{H_k+1}(t) = 0, \forall t \).

Now we present the main results of the elastic algorithm in Theorem 2.

Theorem 2. Let \( \epsilon > 0 \) be chosen arbitrarily small. The elastic algorithm ensures the following inequality on queue backlog:

\[
\limsup_{T \to \infty} \frac{1}{T} T \sum_{t = 0}^{T-1} \mathbb{E}\{\sum_{l \in S} Q_l(t)\} \leq \frac{B_2 + V_2 B_R}{\delta_2},
\]  

(24)

where \( B_2 \triangleq \frac{1}{2} K (N + 2) + \frac{1}{2} N \mu^2 \max \rho_k^2, \delta_2 \triangleq \epsilon \min \rho_k, \) and \( B_R \) is defined as:

\[
B_R \triangleq \limsup_{T \to \infty} \frac{1}{T} T \sum_{t = 0}^{T-1} \mathbb{E}\{K \sum_{k = 1}^{K} \mu^k_{m-1,0}(t) - \sum_{k = 1}^{K} \lambda^{*}_{k,\epsilon} \} \leq \mu M - \sum_{k = 1}^{K} \lambda^{*}_{k,\epsilon}.
\]

Furthermore, the inelastic algorithm achieves:

\[
\liminf_{T \to \infty} \frac{1}{T} T \sum_{t = 0}^{T-1} \mathbb{E}\{\mu^k_{1,0}(t)\} \geq \sum_{k = 1}^{K} \lambda^{*}_{k,\epsilon} - \frac{B_2}{V_2},
\]  

(25)

Remark 6 (Stability): The inequalities (24) from Proposition 2 and (25) from Theorem 2 indicate that PU and SU queues are all stable, and hence is the CRN. In addition, Proposition 2 ensures that PU queues maintained in each route are deterministically bounded by the finite buffer size \( U^k_M \).

Remark 7 (Optimal Throughput and Tradeoff with Delay): The inequality (25) gives the lower-bound of the throughput the elastic algorithm can achieve. Since the constant \( B_2 \) is independent of the control parameter \( V_2 \), the algorithm can achieve a PU throughput arbitrarily close to the optimal value \( \sum_{k = 1}^{K} \lambda^*_k \) as \( \epsilon \) can be chosen arbitrarily small and \( V_2 \) can be chosen arbitrarily large, with the following tradeoffs in PU and SU delay:

- The PU buffer size \( U_M \) is of order \( O(V_2) \) as shown in [24]. By Little’s Theorem, the PU’s average end-to-end delay over any given route \( k \) is of order \( O((H_k + 1)V_2) \) which is bounded by the first order of \( H_k \), i.e., the algorithm has order-optimal delay per route.

- From (21), the average SU buffer occupancy is of order \( O(\frac{1}{\epsilon}) \). And so is the SU average delay by Little’s Theorem. The average SU delay upper-bound has an extra term \( \frac{1}{\epsilon} \) in order compared with the average PU delay.

Remark 8 (Employing Delayed Queue Information): The PU congestion controller (21) is performed at the source PU. Thus, in order to account for the propagation delay of queue information \(( Q_l(t) )_{l \in S}\), we can replace \(( Q_l(t) \) in (21) by \(( Q_l(t - \tau) \) where \( \tau \) is an integer number that is larger than the maximum propagation delay from any node to a source. It is not difficult to show that Theorem 2 still holds with a different value of \( B_2 \), with similar proof techniques as in [18][19].

We prove Theorem 2 in the next subsection.

4.2 Proof of Theorem 2

Before we proceed, we present Lemma 2 as follows to assist us in proving Theorem 2.

Lemma 2. For any feasible rate vector

\[
((\lambda_k)_{k \in \{1, 2, ..., K\}}, \sum_{k = 1}^{K} \rho_k \lambda_k \text{1}_{\{\exists m: v^m_k = l\}} l \in S) \in \Lambda_E,
\]

there exists a stationary randomized algorithm SE that stabilizes the network with PU admitted arrival rates \( \mu^k_{m,SE}(t) = \lambda_k, \forall k \in \{1, 2, ..., K\} \) and a hop/link schedule \((\mu^k_{m,SE}(t))_{m,k}(s^E(t))_{l \in S}\) independent of queue backlog satisfying:

\[
\mathbb{E}\{\mu^k_{m-1,m}(t) - \mu^k_{m,m+1}(t)\} = 0, \forall t, \forall m, k,
\]

\[
\mathbb{E}\{s^E_l(t)\} = \sum_{k = 1}^{K} \rho_k \lambda_k \text{1}_{\{\exists m: v^m_k = l\}}, \forall t, \forall l \in S.
\]

Similar to Lemma 1, it is not necessary for the randomized algorithm SE to provide finite buffer size or delay guarantees. For brevity, we omit the proof of Lemma 2, and interested readers are referred to [8][11][13] for details. Note that \((\lambda_k)_{k \in \{1, 2, ..., K\}}\) can take values as \((\lambda^{*}_{k,\epsilon})_{k}\) and \((\lambda^{*}_{k,\epsilon} + \epsilon)\).

We define the queue vector \( Q_E(t) = ((U^k_m)_m,k, (Q_l)_{l \in S}) \) and define the Lyapunov function \( L_E(Q_E(t)) \) as follows:

\[
L_E(Q_E(t)) \triangleq \frac{1}{2} \left( \sum_{k = 1}^{K} \sum_{m = 0}^{H_k} (U^k_m(t))^2 + \sum_{l \in S} Q_l(t)^2 \right)
\]

Then, the corresponding Lyapunov drift is defined as

\[
\Delta E(t) \triangleq \mathbb{E}\{L_E(Q_E(t + 1)) - L_E(Q_E(t))|Q_E(t)\}.
\]
By squaring both sides of the queueing dynamics (15), we can obtain:
\[
\Delta E - V_2 E \left\{ \sum_{k=1}^{K} \mu_{k,1,0}(t) | Q_E(t) \right\} \\
\leq \frac{1}{2} \sum_{m=0}^{H_k} \sum_{k=1}^{K} \mathbb{E} \left[ \left( \mu_{m-m_1}^k(t) \right)^2 + \left( \mu_{m-m_1}^{k-1,0}(t) \right)^2 \right] \\
- 2U^k_{m}(t)(\mu_{m,m_1}^k(t) - \mu_{m-m_1}^{k-1,0}(t)) | Q_E(t) \right\} \\
+ \frac{1}{2} \sum_{l \in S} E \left\{ s_l(t)^2 + \sum_{k=1}^{K} \rho_k \mu_{k,1,0}(t) 1 \{ \exists m: v_{m-l}^l \}^2 \right\} \\
- 2Q_l(t)(s_l(t) - \sum_{k=1}^{K} \rho_k \mu_{k,1,0}(t) 1 \{ \exists m: v_{m-l}^l \} | Q_E(t) \right\} \\
- V_2 E \left\{ \sum_{k=1}^{K} \mu_{k,1,0}(t) | Q_E(t) \right\},
\]

from which we obtain:
\[
\Delta E - V_2 E \left\{ \sum_{k=1}^{K} \mu_{k,1,0}(t) | Q_E(t) \right\} \\
\leq B_2 - V_2 E \left\{ \sum_{k=1}^{K} \mu_{k,1,0}(t) | Q_E(t) \right\} \\
- \sum_{l \in S} E \left\{ Q_l(t)(s_l(t) - \sum_{k=1}^{K} \rho_k \mu_{k,1,0}(t) 1 \{ \exists m: v_{m-l}^l \}) | Q_E(t) \right\}.
\]

Through algebra, we find the equivalence of (26):
\[
\Delta E - V_2 E \left\{ \sum_{k=1}^{K} \mu_{k,1,0}(t) | Q_E(t) \right\} \\
\leq B_2 + \sum_{l \in S} E \left\{ Q_l(t)(s_l(t) - \sum_{k=1}^{K} \rho_k \mu_{k,1,0}(t) 1 \{ \exists m: v_{m-l}^l \}) | Q_E(t) \right\} - V_2 E \left\{ \sum_{k=1}^{K} \mu_{k,1,0}(t) | Q_E(t) \right\}.
\]

We take the expectation of both sides of (29) over $Q_E(t)$ and take the time average on $t = 0, 1, ..., T - 1$, which leads to
\[
\frac{1}{T} \sum_{k=1}^{T} E \left\{ \sum_{l \in S} Q_l(t) \right\} \\
\leq B_2 + \frac{V_2}{T} \sum_{l \in S} E \left\{ \sum_{k=1}^{K} \mu_{k,1,0}(t) \right\} - V_2 \sum_{k=1}^{K} \lambda_{k,e}.
\]

By taking limsup on both sides of (29), we can prove (21). By taking the liminf of $T$ on both sides of (29), we can prove (25). Therefore, Theorem 2 is proved.

5. ARBITRARY ARRIVAL RATES AT TRANSPORT LAYER

In the previous model description and algorithm development, we assumed that PU and SU packets are backlogged at the transport layer. In this section, we present optimal algorithms for the inelastic and elastic PU models, respectively, for arbitrary arrival rates at the transport layer.

At the transport layer, let $E_p(t)$ and $E_l(t)$, $l \in S$, be the PU and SU arrival rates at the beginning of time slot $t$, respectively. We assume that $E_p(t)$ and $E_l(t)$, $l \in S$, are i.i.d. with respect to time. For simplicity of analysis, we assume that the time average arrival rate vector, formed by the PU and SU arrival rates, is in the exterior of the capacity region, so that a congestion controller is needed. Let $W_p(t)$ and $W_l(t)$, $l \in S$, be the backlog of PU and SU data at the transport layer. PU and SU buffer sizes at the transport layer are denoted by $W_P$ and $W_S$, respectively. In the following subsections, we present modified algorithms that can handle arbitrary arrival rates at transport layer for inelastic and elastic PU models.

5.1 Inelastic Algorithm for Arbitrary Arrival Rates at Transport Layer

In the inelastic PU model, recalling that $A_l(t)$ is the admitted SU rate, we update the SU backlog $W_l(t)$ at the transport layer as follows:
\[
W_l(t+1) = \min \{ [W_l(t) + E_l(t) - A_l(t)]^+, W_S \}, \forall l \in S.
\]

Note that $W_l(t) = 0 \forall t \in S$ and $W_S = 0$ if there is no buffer at SU transport layer. Similarly, we update the PU backlog $W_p(t)$ at the transport layer as follows:
\[
W_p(t+1) = \min \{ [W_p(t) + E_p(t) - \mu_p(t)]^+, W_P \}.
\]
Note that $W_p(t) = 0 \forall t$ and $W_P = 0$ when there is no buffer at PU transport layer.

Following the idea introduced in [12], we construct a virtual queue $Y_p(t)$ at the PU transport layer with queueing dynamics:

$$ Y_p(t + 1) = [Y_p(t) - R(t)]^+ + u_p(t), $$

where $u_p(t)$ is an auxiliary variable associated with $Y_p(t)$. Similarly, we construct a virtual queue $Y_i(t)$, $i \in S$, at the SU transport layer with queueing dynamics:

$$ Y_i(t + 1) = [Y_i(t) - A_i(t)]^+ + u_i(t), $$

where $u_i(t)$ is an auxiliary variable associated with $Y_i(t)$, with $u_i$ being its time-average. Note that when $Y_i(t)$ is stable, the time-average of SU admitted rate $A_i(t)$ is greater than or equal to $u_i$. Thus, when $Y_i(t)$ is stable, $\forall i \in S$, if we can ensure that $\sum_{i \in S} g_i(u_i)$ is arbitrarily close to the optimal value $\sum_{i \in S} g_i(r_i^*)$, so is the SU utility.

Instead of (34), the queue state $Z(t)$ is now updated as:

$$ Z(t + 1) = [Z(t) - u_p(t)]^+ + f^{-1}(a_P). $$

Denote the time-average of $u_p(t)$ as $u_P$. Thus, when $Y_p(t)$, $U_P(t)$ and $Z(t)$ are stable, we have $f(\mu) \geq f_r(\tau) \geq f(u_p) \geq a_P$, where we recall that $\mu$ and $\tau$ are the time-average values of $\mu_{p+P}(t)$ and $R(t)$, respectively. Specifically, when $Y_p(t)$, $U_P(t)$ and $Z(t)$ are stable, the PU minimum utility constraint is met.

Now we provide the inelastic algorithm for arbitrary arrival rates at the transport layer:

1) **SU Congestion Controller:**

$$ u_i(t) = [A_i(t) - Q_i(t)]^+ + u_i(t) $$

we present the following theorem for the performance of the algorithm:

**Theorem 3.** Let $\epsilon > 0$ be chosen arbitrarily small. Given that $q_M > \mu_M + N+1 \epsilon \sum_{k=1}^K \mu_{k-1,0}(t)$, the inelastic algorithm ensures

$$ \limsup \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E} \{ (Q_i(t) + Y_i(t)) + U_p(t) + Y_p(t) + Z(t) \} \leq \frac{B_3 + \rho g M}{\delta_3}, $$

where $B_3 \triangleq B_1 + NA_2^2 + \mu_2^2$ and $\delta_3$ is a constant. Furthermore, the algorithm achieves:

$$ \sum_{i \in S} g_i(s) \geq \sum_{i \in S} g_i(r_i^*) - B_3 \frac{1}{V_1}, $$

where we recall that $(a_i)_{i \in S}$ is defined in Theorem 1.

The proof follows similar steps as the proof of Theorem 1. Due to space limitations, we omit the proof for Theorem 3 and the choice of $\delta_3$. Similar statements as presented in Remarks 1–4 also hold for the inelastic algorithm introduced in this subsection.

5.2 Elastic Algorithm for Arbitrary Arrival Rates at Transport Layer

In this subsection, the elastic algorithm for arbitrary arrival rates at transport layer and its performance are discussed. Similar to inelastic algorithm, we denote by $A_i(t)$ the admitted SU rate, which is upper-bounded by $A_M$. We update the SU backlogs $W_i(t)$ at the transport layer as (35). Similarly, we can update the PU backlog $W_p(t)$ at the transport layer as:

$$ W_p(t + 1) = \min \{ W_p(t) + E_p(t) - \sum_{k=1}^K \mu_{k-1,0}(t) \}^+, W_P \}. $$

Similar to the previous subsection, we construct virtual queues $Y_p(t)$ and $Y_i(t)$, $\forall i \in S$, with an auxiliary variable $u_p(t)$ associated with $Y_p(t)$. The virtual queues evolve as follows:

$$ Y_p(t + 1) = [Y_p(t) - \sum_{k=1}^K \mu_{k-1,0}(t)]^+ + u_p(t); $$

$$ Y_i(t + 1) = [Y_i(t) - A_i(t)]^+ $$

Note that when $Y_p(t)$ is stable, if we can ensure that $u_p$ is arbitrarily close to the optimal value $\sum_{k=1}^K \mu_{k-1,0}(t)$, then so is the PU throughput, where we recall that $u_p$ is the time average of $u_p(t)$. In addition, when $Y_i(t)$ is stable, $\forall i \in S$, the SUs’ throughput is proportional to the PU data that they relay.

Now we provide the elastic algorithm for arbitrary arrival rates at the transport layer:

1) **SU Congestion Controller:**

$$ u_i(t) = [A_i(t) - Q_i(t)]^+ + u_i(t) $$

2) **PU Congestion Controller:**

$$ \max u_p(t)(q_M + \mu_M - U_p(t)) $$

4) **Link Rate Scheduler:** The link scheduler is the same as (34) in Section 3.2.

It is not difficult to check that Proposition 1 still holds, and we present the following theorem for the performance of the algorithm:

**Theorem 3.** Let $\epsilon > 0$ be chosen arbitrarily small. Given that $q_M > \mu_M + N+1 \epsilon \sum_{k=1}^K \mu_{k-1,0}(t)$, the inelastic algorithm ensures
\[
\min \sum_{k=1}^{K} \mu_{k,0}(t) (p_k \sum_{i \in S} Y_i(t) 1_{\{\text{min}\ v_p^i = t\}} + U_k^b(t) - Y_p(t)) \\
\text{s.t. } 0 \leq \sum_{k=1}^{K} \mu_{k,0}(t) \leq \min\{W_p(t) + E_p(t), \mu_M\}.
\]

Note that (36) and (37) can be solved independently.

3) Link/Hop Rate Scheduler: The link/hop scheduler is the same as (22) in Section 4.1.

We present Proposition 3 and Theorem 4 to characterize the performance limits of QoS-constrained multi-hop CRNs. In our work, we aim at a statistically upper-bounded PU buffer sizes and hence the average end-to-end delay upper-bounds. Our work aims at a tradeoff in the deterministically upper-bound PU buffer sizes and hence the average end-to-end delay upper-bounds. Our work aims at a statistically upper-bounded PU buffer sizes and hence the average end-to-end delay upper-bounds.

6. CONCLUSIONS AND FUTURE WORKS

In this paper, two cross-layer scheduling algorithms for multi-hop cooperative cognitive radio networks are introduced. The algorithms can achieve arbitrarily close to the optimal throughput/utility, with a tradeoff in the deterministically upper-bound PU buffer sizes and hence the average end-to-end delay upper-bounds. Our work aims at a better understanding of the fundamental properties and performance limits of QoS-constrained multi-hop CRNs. In our future work, we will investigate distributed implementations and power management in CRNs.

7. REFERENCES

[1] “FCC Spectrum Policy Task Force: Report of the spectrum efficiency working group”, November 2002.
[2] I. Akyildiz, W. Lee, M. Vuran, and S. Mohanty, “Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey, Computer Networks”, in Computer Networks Journal (Elsevier), vol. 50, no. 11, pp. 2127-2159, 2006.
[3] L. Lai, H. Gamal, H. Jiang and H. Poor, “Cognitive Medium Access: Exploration, Exploitation and Competition”, to appear in IEEE/ACM Transactions on Networking, 2010.
[4] Q. Zhao, L. Tong, A. Swami and Y. Chen, “Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework”, in IEEE Journal on Selected Area in Communications, vol. 25, no. 3, pp. 589-600, April 2007.
[5] H. Su and X. Zhang, “Cross-layer based opportunistic MAC protocols for QoS provisioning over cognitive radio wireless networks,” in IEEE Journal on Selected Areas in Communications, vol. 26, no. 1, pp. 118-129, January 2008.
[6] H. Kim and K. G. Shin, “Efficient discovery of spectrum opportunities with MAC-layer sensing in cognitive radio networks,” in IEEE Transactions on Mobile Computing, vol. 7, no. 5, pp. 533-545, May 2008.
[7] R. Urgaonkar and M. Neely, “Opportunistic scheduling with reliability guarantees in cognitive radio networks”, in IEEE Transactions on Mobile Computing, vol. 8, no. 6, pp. 766-777, June 2009.
[8] M. Lotfizelahad, B. Liang, and E. Sousa, “Optimal control of constrained cognitive radio networks with dynamic population size”, in Proc. IEEE INFOCOM’10, March 2010.
[9] K. Khalil, M. Karaca, O. Ercetin, and E. Ekici, “Optimal scheduling in cooperate-to-join cognitive radio networks”, submitted to IEEE INFOCOM’11.
[10] D. Xue and E. Ekici, “Guaranteed opportunistic scheduling in multi-hop cognitive radio networks”, submitted to IEEE INFOCOM’11.
[11] L. Tassiulas and A. Ephremides, “Stability properties of constrained queuing systems and scheduling policies for maximum throughput in multihop radio networks,” in IEEE Trans. Autom. Control, vol. 37, no. 12, pp. 1936-1948, Dec. 1992.
[12] L. Georgiadis, M. Neely, and L. Tassiulas, “Resource allocation and cross-Layer control in wireless networks”, in Foundations and Trends in Networking, pp. 1-149, 2006.
[13] M. Neely, “Dynamic power allocation and routing for satellite and wireless networks with time varying channel”, Ph.D. dissertation, Mass. Inst. Technol. (MIT), Cambridge, MA, 2003.
[14] X. Wu, R. Srikant, and J. Perkins, “Scheduling efficiency of distributed greedy scheduling algorithms in wireless networks”, in IEEE Trans. Mobile Comput., vol. 6, no. 6, pp. 595-605, June 2007.
[15] C. Joo, X. Lin, and N. Shroff, “Greedy maximal matching: performance limits for arbitrary network graphs under the node-exclusive interference model”, in IEEE Transactions on Automatic Control, vol. 54, no. 12, pp. 2734-2744, December 2009.
[16] G. Sharma, R. Marumdar, and N. Shroff, “On the complexity of scheduling in wireless networks”, in Proc. of the 13th Annual International Conference on Mobile Computing and Networking (MobiCom’06), pp. 227-238, 2006.
[17] P. Giaccone, E. Leonardi, and D. Shah, “Throughput region of finite-buffered networks”, in IEEE Transactions on Parallel and Distributed Systems, vol. 18, no. 2, pp. 251-263, Feb. 2007.
[18] L. Le, E. Modiano, and N. Shroff, “Optimal control of wireless networks with finite buffers”, in Proc. IEEE INFOCOM’10, April 2010.
[19] L. Ying, R. Srikant and D. Towsley, “Cluster-based back-pressure routing algorithm”, in Proc. IEEE INFOCOM'08, April 2008.
[20] S. Saks, M. Togasaki, and K. Yamazaki, “A note on greedy algorithms for the maximum weighted independent set problem”, in Discrete Applied Mathematics, vol. 126, pp. 313-322, 2003.
[21] P. Huang, X. Lin, and C. Wang, “A low-complexity congestion control and scheduling algorithm for multi-hop wireless networks with order-optimal per-flow delay”, submitted to IEEE INFOCOM’11.
[22] M. Lotfizelahad and P. Marbach, “Throughput-optimal random access with order-optimal delay”, submitted to IEEE INFOCOM’11.
[23] L. Jiang and J. Walrand, “A distributed algorithm for throughput and utility maximization in wireless networks”, to appear in Transactions on Networking, 2010.
[24] J. Ni, B. Tan, and R. Srikant, “Q-CSMA: queue-length based CSMA/CA algorithms for achieving maximum throughput and low delay in wireless networks”, in IEEE INFOCOM’10, April 2010.
[25] S. Rajagopalan, D. Shah, and J. Shin, “Network adiabatic theorem: an efficient randomized protocol for contention resolution”, in ACM SIGMETRICS’09, pp. 133-144, 2009.
[26] L. Jiang, M. Leconte, J. Ni, R. Srikant, and J. Walrand, “Fast mixing of parallel Glauber dynamics and low-delay CSMA scheduling”, arXiv.org:1008.0227v1, August 2010.
[27] D. Shah, D. Tse, and J. Tsitsiklis, “Hardness of low delay network scheduling”, submitted to IEEE Transactions on Information Theory, 2009.