On the Quality and Timeliness of Fusion in a Random Access Sensor Network

Abdurrahim Soganli, Ozgur Ercetin, and Mujdat Cetin, *Senior Member, IEEE*

***Abstract*—In this letter, we investigate the trade-off between the quality of multi-sensor fusion decision and the delay of making that decision in a random access sensor network in a particular setting where the correlation between two sensor observations decreases with the increasing distance between these sensors. We propose a system operation wherein the nodes located in a small neighborhood of a sensor node that has successfully transmitted its decision to the fusion center are deactivated to reduce the contention over the random access channel. We propose a utility function that involves the quality and timeliness of the decision and determine the optimal size of this neighborhood analytically under a Markovian system operation model and numerically by simulating the actual system operation. These experiments demonstrate how our model can be used to set this neighborhood size before data collection to achieve a desired trade-off between decision quality and timeliness.**

***Index Terms*—Data fusion, decentralized detection, wireless sensor networks.**

I. Introduction

MULTI-SENSOR DATA FUSION (MDF) combines information acquired from multiple spatially distributed sensors for detection/estimation of a phenomenon of interest [1], [2]. In this letter, we quantitatively analyze the trade-off between the quality of multi-sensor fusion decision and the delay of making that decision in a random access sensor network in a particular setting where the correlation between two sensor observations reduces with the increasing distance between these sensors. Consider the following Decision Fusion Network (DFN) to shed light on the problem. Assume that sensor nodes measure the CO_2 level over an area of interest, and a sensor node decides to produce an alarm decision if its measurement level is above a given threshold. Also assume that the CO_2 level is varying smoothly within the area of interest, so the decisions of the individual sensors in a sufficiently small neighborhood are highly correlated. Let the fusion center be tasked with determining if the average CO_2 level in the area is above a threshold or not. Unless one is willing to wait for the reception of the decisions of the entire set of sensors, it is of interest to come up with sequential strategies of *sampling* the spatially distributed sensors in such a way to capture the diversity of the observations. Thus, at each instance we consider the distance relationship of sensor nodes. Once a sensor sends its decision, sensors located within a distance of τ from this sensor are deactivated for the current decision period which is until the time point of the decision of the FC. We consider an objective function that captures the desired trade-off between the quality and timeliness of the decision. The goal then is to find the optimal neighborhood size parameter τ with respect to that function. We propose a Markovian system model operation based on which we determine τ analytically. Simulation of the actual system operation demonstrates that our analytical approach produces an accurate value for the optimal τ, which suggests our model could be used during system set up to achieve the desired trade-off between decision quality and timeliness during system operation.

The quality of decision of the FC is usually represented in terms of its *global error probability*, which is a function of its false alarm and missed detection probabilities [3]. The decision quality mainly depends on the locations of the successfully transmitting sensors and the number of sensor decisions collected. An equally important issue in multi-sensor fusion that is often neglected in prior works is the repercussion of the length of the duration until a final decision is made by the FC. For example, some fusion applications, such as the estimation of room temperature, can tolerate a longer delay than others, e.g., the detection of a fire or intrusion. There is often a trade-off between how soon the decision is made and how reliable this decision is. The problem is further exacerbated if data is carried over a common random access communication channel, where the delay is not fixed due to possible collisions of transmissions from sensor nodes.

The effects of sharing a common access communication channel in DFNs were previously investigated in the literature [3]–[9]. In order to control the access of large number of spatially distributed nodes, a random access protocol such as Aloha or CSMA [4], [10] is employed, since these protocols require little coordination. Over random access channels, transmissions occasionally collide with collided packets being retransmitted until they are successful. In [5] the transmission probabilities of sensor nodes are determined based on their reliabilities, and in [6], sensors are grouped with respect to the informativeness of their data, with priority given to more informative groups of sensors. A reliability-based algorithm is also proposed in [7], and [8] where the decision is made periodically for a given period. Although the delay of making a decision can be made arbitrarily small, the quality of the decision depends on whether sufficient data over the field of interest has been collected or not. In [9], by using compressed sensing theory and the location knowledge of the decisions, complete sensor decisions are recovered by using randomly selected decisions via l_1 minimization. In this work, unlike the previously proposed methods, we develop a utility based

Manuscript received December 25, 2014; accepted January 18, 2015. Date of publication January 27, 2015; date of current version February 04, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Peter Willett.

The authors are with the Faculty of Engineering and Natural Sciences, Sabanci University, Tuzla, Istanbul 34950 Turkey (e-mail: soganli@sabanciuniv.edu; oercetin@sabanciuniv.edu; mcetin@sabanciuniv.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2015.2396573

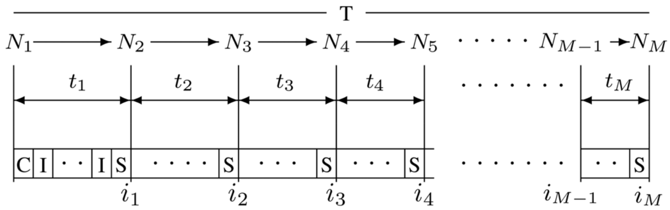

Fig. 1. Timing diagram of the operation of the system. *S, I, C* represent success, idle, and collision events, respectively.

optimization framework involving the quality and timeliness of the decision to quantitatively decide on the system parameter giving the best performance.

II. System Operation

We consider a sensor network of N nodes and a single FC randomly distributed over an area of interest with radius R units. All sensors sample the environment and make decisions when the FC sends an observation request to the sensors. We assume that all sensor nodes are numbered and they know their distance to one another. The nodes attempt to send their decisions via CSMA based multiple access channel. In particular, we consider both slotted and unslotted CSMA in our analysis. After a successful transmission, the FC acknowledges the nodes about which node transmitted its decision successfully. Sensor nodes within τ units to the successful node refrain from transmitting in subsequent epochs.

The operation of the system is illustrated in Fig. 1. Initially, the FC requests sensor nodes to make observations. After receiving the FC request, all nodes become backlogged and they transmit with probability q. Time is divided into *epochs* of random length each of which ends after a successful transmission. Let i_m be the sensor node which successfully transmitted its decision in the mth epoch. Let t_m be the length of the mth epoch, i.e., elapsed time between the $(m-1)$th and mth successful transmissions. After a successful transmission, a number of sensor nodes that are in the vicinity of the successfully transmitting sensor drop their packets, and remain silent until the FC makes its decision. Accordingly, channel contention is reduced at each consecutive epoch. The final epoch is reached when there is no sensor node remaining to transmit its decision to the FC. The FC makes its decision after receiving the Mth successful transmission, where M is a random variable depending on the ordered set of sensor nodes successfully transmitting in each subsequent epoch. Finally, N_m is the number of active nodes between $(m-1)$th and mth successful transmissions.

III. Optimization Framework

In this section, we define an optimization problem and seek the best distance threshold τ with respect to an objective function that balances the quality and timeliness of sensor fusion. The quality of sensor fusion can be measured in terms of the global error probability, which is a function of the false alarm and missed detection probabilities. Let H_0 and H_1 denote the two hypotheses that the FC is testing. The *a priori* probabilities of H_0 and H_1 are assumed to be constant and known. The local decision rules are fixed and every sensor has the same false alarm probability and missed detection probability. For many natural phenomena, decisions of the sensors in a small neighborhood are highly correlated. Hence, the relationship between the number of local decisions received and the global error probability reflects the *principle of diminishing returns*, i.e., the benefit of adding an element to a set is non-increasing as a function of the number of elements in the set [11]. For many sensor fusion applications the quality and the timeliness of the fusion decision are not equally important. Some applications may tolerate a small increase in global error probability in exchange for reducing the delay of the fusion decision. Although there is little prior work on the utility of making a fusion decision quickly, we assume that the utility of fusion decision decays exponentially with respect to its delay. Such a utility function is especially appropriate for applications such as intrusion detection, target identification or emergency home care systems.

In our work, we examine the trade-off between the quality and the timeliness of the sensor fusion by varying the number of sensor nodes that refrain from transmitting after each epoch (by adjusting τ). Hence, we define the utility functions in terms of τ. Let $\overline{M}(\tau)$ and $\overline{T}(\tau)$ be the average number of sensor nodes successfully transmitting to the FC until a decision is made and the average duration of time it takes to do so, respectively. Fig. 2 illustrates $\overline{M}(\tau)$ and $\overline{T}(\tau)$ with respect to τ, obtained by numerical analysis for a particular sensor network set up we have used in experiments. For the CSMA protocol, two distinct values of the normalized propagation delay $\beta = 0.0026$ and $\beta = 0.01$ are used[1]. Note that both $\overline{M}(\tau)$ and $\overline{T}(\tau)$ monotonically decrease, since increasing τ also increases the number of sensor nodes remaining silent after a successful transmission. Also, let $Q(\overline{M}(\tau))$ be the *quality utility* of the FC for a sensor fusion decision made with $\overline{M}(\tau)$ local sensor decisions. The *delay utility* of this fusion decision is given by $R(\overline{T}(\tau))$. If sensor nodes provide similar performance, $Q(\overline{M}(\tau))$ should be a decreasing function of τ, since the global error probability will increase when the FC makes a decision with fewer number of local sensor decisions. On the other hand, $R(\overline{T}(\tau))$ should be an increasing function of τ, since the FC can make a decision quicker as τ increases. Fig. 3 illustrates the variation of our generic choices for $Q(\cdot)$ and $R(\cdot)$ with respect to τ. In practice, these utility functions can be chosen considering the requirements and constraints of a particular application. We aim to determine the value of τ which balances the utilities of quality and timeliness of the fusion when the optimization problem is defined as follows:

$$\tau^* = \arg\max_{\tau} Q(\overline{M}(\tau))R(\overline{T}(\tau))^{\alpha}, \tag{1}$$

where α balances the weight of the quality and timeliness utilities, and it can be selected to reflect the characteristics of a particular sensor network application. In the subsequent section, we investigate the analytical calculation of $\overline{M}$ and $\overline{T}$.

IV. Analytical Results

According to our system operation, after each successful transmission, the nodes that are at a distance τ or less to the successful node become deactivated. However, at each epoch there may be some nodes that are already deactivated in the previous epochs and still at a distance of τ or less to the successful node in the current epoch. The possible overlapping between consecutive deactivated nodes makes the calculation

[1]These values correspond to 256 and 64 bit packet sizes for the IEEE 802.11 protocol with 2 Mb/s data rate. Dynamic β values can also be used according to the contention in the channel [12].

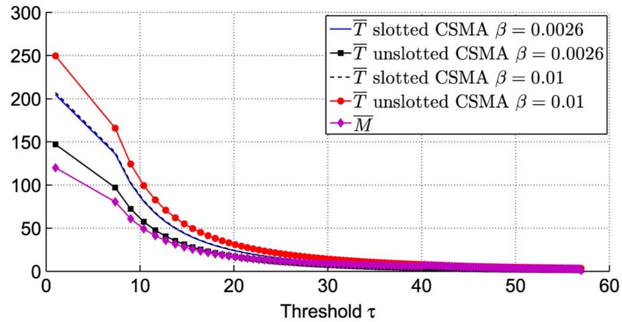

Fig. 2. Expected value of M and T with respect to threshold τ.

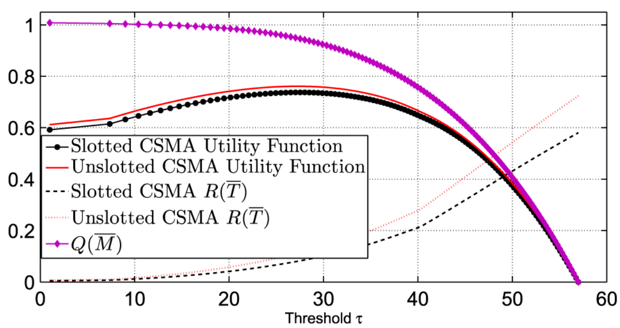

Fig. 3. $Q(\overline{M})$ and $R(\overline{T})$ with respect to threshold τ, and corresponding overall concave utility function.

of the number of active nodes in each epoch complicated. In the subsequent analysis, we assume that the number of nodes that are deactivated follows a uniform distribution in $[1, \gamma]$ where γ is a function of τ and sensor density. Let N_m be the number of active nodes in the mth epoch. Hence, the conditional distribution of N_m given N_{m-1} when $N_{m-1} > \gamma$ is given as:

$$f_{N_m|N_{m-1}}(n) = \begin{cases} \frac{1}{\gamma}, & N_{m-1} - \gamma \leq n \leq N_{m-1} - 1 \\ & \text{and} \quad N_{m-1} \geq \gamma \\ 0, & \text{otherwise} \end{cases}, \quad (2)$$

where $\frac{\gamma+1}{2}$ is the expected number of deactivated nodes in a radius of τ. If $N_{m-1} \leq \gamma$, then $f_{N_m|N_{m-1}}(n) = 0$, for all n. Note that this implies that when fewer than γ nodes remain, our analysis model assumes that they are within a distance of τ from one another. With the aforementioned assumption on the number of deactivated sensor nodes at the end of each epoch, our system can be modeled as a Markov chain as shown in Fig. 4. The states of the Markov chain are the number of active nodes, and state transitions occur at the end of each epoch according to the conditional distribution given in (2). Hence, the Markov chain has $\gamma + 1$ absorbing states, and $N - \gamma$ transient states. Note that at each state transition, one successful sensor decision is transmitted to the FC. Hence, the expected number of steps before being absorbed when starting in transient state N gives the average number of successful sensor decisions received by the FC, i.e., $\overline{M}$.

A. Number of Successful Transmissions

Let h_n be the number of steps until absorption, when there are n active nodes in the system. Note that $\overline{M} = h_N + 1$.

Lemma 1: For the Markov chain given in Fig. 4, *h_n is calculated as follows*:

$$h_n = \frac{1}{\gamma} \sum_{i=1}^{\gamma} h_{n-i} + 1 \quad \text{s.t.} \quad n > \gamma \quad (3)$$

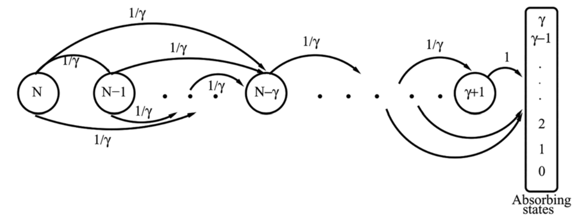

Fig. 4. Markov chain for the proposed model.

Proof: In a finite absorbing chain, starting from a transient state, the chain makes a finite number of visits to some transient states before its eventual absorption into one of the absorbing states. Hence the expected absorption time of the chain, starting from transient state i initially, is the sum of the expected numbers of visits made to transient states. We perform first-step analysis, by conditioning on the first step the chain makes after moving away from a given initial state to obtain the result in (3). □

B. Successful Transmission Probability

In slotted CSMA, the successful transmission probability during the mth epoch is $p_s(m) = N_m q(1-q)^{N_m - 1}$, where q is the retransmission probability of an active sensor node when channel is sensed idle. Note that the probability of the slot being idle or having a collision is $1 - p_s(m)$. Let β be the maximum round-trip propagation delay between any two nodes normalized with respect to packet duration. Following the throughput analysis of the slotted CSMA given in [10], the expected duration between two consecutive successful transmission is determined as $\frac{\beta+1-(1-q)^{N_m}}{p_s(m)}$. The optimal throughput is achieved when the retransmission probability is equal to the reciprocal of the number of active nodes in the system. Hence, in the following we assume that the retransmission probability during the mth epoch is $q_m = \frac{1}{N_m}$. The optimal probability of successful transmission during the mth epoch is $p_s(m) = (1 - \frac{1}{N_m})^{N_m - 1}$. Since the true value of N_m is unknown, each sensor node in our system needs to estimate the number of active nodes during each epoch using methods such as Bayesian broadcast [13] or modified stochastic gradient [14]. Note that a sub-optimal value of transmission probability will result in a lower probability of success, which in turn will result in a longer delay.

In unslotted CSMA, each node attempts to retransmit with a probability that is exponentially distributed with an average rate of x. The successful transmission probability is $p_{us}(m) = e^{-\beta N_m x}$, and the expected time between two consecutive successful transmissions can be obtained as $\frac{\beta+1+\frac{1}{N_m x}}{p_{us}(m)}$ by following the analysis in [10]. Note that for optimal operation of unslotted CSMA, the perfect knowledge about the active nodes is not required, and for β small this protocol significantly reduces the collisions as compared to the slotted CSMA[2].

C. Delay of the FC Decision

The average delay until the FC decision is $d = c + \overline{T} + f$ where c is the time required for sensor nodes to collect measurements and make their decisions, $\overline{T}$ is the transmission time of sensor decisions, and f is the processing time of the FC.

[2]Note that our analysis remains valid for any other medium access protocol when the probability of success and the expected duration between the successful transmissions are replaced with their appropriate values.

In general, c depends on the hardware of the sensor nodes and complexity of the sampling task and $\overline{T}$ is usually much longer than f. In the numerical results, we neglect f and c to focus on the effect of our proposed method. Let $Z(n)$ be the average number of slots until successful transmission, where n is the number of active nodes. Also define T_n as the average time to reach an absorbing state, starting from a transient state n. The value of $\overline{T}$ is determined in the following lemma by noting that $\overline{T} = T_N + Z(\gamma)$.

Lemma 2: For the Markov chain defined in Fig. 4, T_n *is expressed as*:

$$T_n = \frac{1}{\gamma}\sum_{i=1}^{\gamma} T_{n-i} + Z(n) \quad \text{s.t.} \quad n > \gamma \tag{4}$$

Proof: The result follows from the first-step analysis. □

V. Numerical Results

In this section, we compare our analytical results with those obtained from simulations for different values of α and β. In our simulation model, there are $N = 120$ sensor nodes distributed uniformly randomly in a square area of side-length $R = 100$. For sensor networks with uniformly randomly distributed nodes, the number of deactivated sensor nodes in one successful transmission is proportional to $\frac{\tau^2}{R^2}$. Thus, we use the following relationship between τ and γ, i.e., $\gamma = \lfloor N\frac{\tau^2}{R^2} \rfloor$ in order to compare the analytical and simulation results.

In our analysis, we use the following utility functions, which are also depicted in Fig. 3:

$$Q(\overline{M}(\tau)) = 1 + \frac{\overline{M}(\tau)}{N^2} - \frac{\gamma^2}{N^2},$$
$$R(\overline{T}(\tau)) = \overline{T}^{-1}(\tau),$$

Note that the second term of the quality utility tends to increase with the number of successful transmissions while the third term penalizes excessive threshold distances. These utility functions can be adapted to particular networks and applications by changing the parameters involved. Also note that we set $x = 1$ which is the normalized transmission time.

We run the simulations for 200 times and observed that the deviations are rather small. We depict the average of these results in the figures. In Fig. 5, the objective function with respect to different settings for both slotted and unslotted CSMA protocols are shown. For $\alpha = 0.1$, we observe that the analytical and simulation results are in good agreement, and the optimum distance threshold τ, or equivalently the average number of deactivated nodes γ can be determined reliably based on our analytical model. We also evaluate the performance of our proposed method for two different values of the normalized propagation delay β. Note that for small β, the aggregate utility of unslotted CSMA is better than that of slotted CSMA. This is because, when all nodes are backlogged the idle period is shorter if the propagation delay is much smaller than the transmission delay. We also compare the simulation results for three different values of α. The optimal value of τ shifts to the right, since for large α the contribution of the timeliness function in the objective function is smaller.

Finally, we compare the performance of our proposed approach to the *conventional* case when $\tau = 0$, i.e., only the sensor node whose transmission is successful is deactivated. Hence, with the conventional approach the FC makes its decision only after receiving decisions from all sensor nodes. Table I provides the optimal values of τ^*, $\overline{M}$, $\overline{T}$, and the utility for our proposed and conventional approaches obtained by simulations as well as analytically. Our proposed approach provides approximately an order of magnitude improvement in $\overline{M}$ and $\overline{T}$, and better utility as compared to the conventional approach. Note that the value of $\overline{T}$ obtained by the simulations is higher than the value obtained from the analytical model, since in reality the number of sensor nodes deactivated after each successful transmission is usually smaller than the number of nodes deactivated in the analytical model.

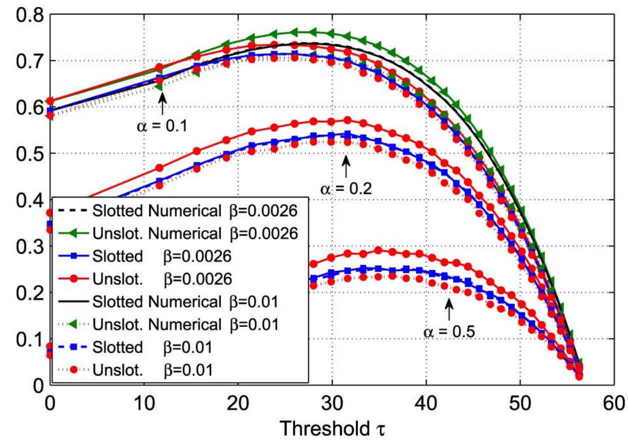

Fig. 5. Comparison of simulation and numerical result of the model.

TABLE I
COMPARISON OF PROPOSED METHODS WITH CONVENTIONAL APPROACH FOR $\alpha = 0.1$ AND $\beta = 0.026$. THE MAIN ENTRY IN EACH CELL IS FROM SIMULATIONS, WHEREAS THE VALUE IN PARENTHESIS IS CALCULATED FROM THE ANALYTICAL MODEL

Method	τ^*	$\overline{M}$	$\overline{T}$	Utility
Slotted	24.95 (27.53)	12.15 (11.88)	19.92 (15.09)	0.73 (0.71)
Unslotted	24.95 (27.03)	12.05 (11.88)	14.89 (10.90)	0.76 (0.74)
Slotted Conv.	0	120 (120)	205 (204)	0.59 (0.59)
Unslotted Conv.	0	120 (120)	146 (147)	0.61 (0.61)

VI. Conclusion

In this letter, we analyzed the trade-off between the quality of multi-sensor fusion decision and the delay of making that decision in a random access sensor network in a particular setting where the correlation between two sensor observations reduces with the increasing distance between these sensors. We exploited this correlation structure to deactivate a set of neighborhood sensors located within a distance τ from a successfully transmitting sensor node. We proposed utility functions for the quality and timeliness of the decision and determined the optimal value of τ analytically under an idealized system model and numerically by simulation the operation of the actual system. Our results indicate that the timeliness of sensor fusion decision is as important as the quality of the decision, and this trade-off can be addressed by designing algorithms spanning multiple (e.g., application and MAC) layers of the network stack.

REFERENCES

[1] D. Hall, "Multisensor data fusion," in *ARL Rev.*, 2001.

[2] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Inf. Fusion*, vol. 14, no. 1, pp. 28–44, 2013.

[3] Y. Yuan and M. Kam, "Distributed decision fusion with a random-access channel for sensor network applications," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 4, pp. 1339–1344, Aug. 2004.

[4] Q. Tian and E. J. Coyle, "Optimal distributed detection in clustered wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3892–3904, 2007.

[5] T. Chang, T. Hsu, and Y. P. Hong, "Exploiting data-dependent transmission control and mac timing information for distributed detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1369–1382, Mar. 2010.

[6] D. Xu and Y. Yao, "Contention-based transmission for decentralized detection," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1334–1342, Apr. 2012.

[7] S. Laitrakun and E. J. Coyle, "Optimizing the collection of local decisions for time-constrained distributed detection in wsns," in *Proc. 2013 IEEE INFOCOM*, 2013, pp. 1923–1931.

[8] S. Laitrakun and E. J. Coyle, "Reliability-based splitting algorithms for time-constrained distributed detection in wsns," in *IEEE Int. Conf. Distributed Computing in Sensor Systems (DCOSS)*, 2013, pp. 207–214.

[9] F. Fazel, M. Fazel, and M. Stojanovic, "Random access compressed sensing for energy-efficient underwater sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1660–1670, Sep. 2011.

[10] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1992.

[11] P. Samuelson and W. D. Nordhaus, *Microeconomics*, 19th ed. New York, NY, USA: Mc Graw-Hill, 2009.

[12] P. Marbach, A. Eryilmaz, and A. Ozdaglar, "Asynchronous csma policies in multihop wireless networks with primary interference constraints," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3644–3676, Jun. 2011.

[13] R. L. Rivest, "Network control by bayesian broadcast," *IEEE Trans. Inf. Theory*, vol. 33, no. 3, pp. 323–328, May 1987.

[14] S. A. Shah and C. F N. Cowan, "Modified stochastic gradient algorithm using nonquadratic cost functions for data echo cancellation," *Proc. Inst. Elect. Eng., Vis. Image Signal Process.*, vol. 142, no. 3, pp. 187–191, Jun. 1995.