



Title	A framework for balancing information collection and data transmission
Author(s)	Chuai, J; Li, VOK
Citation	The IEEE 24th Annual International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC): Mobile and Wireless Networks, London, UK., 8-11 September 2013. In Conference Proceedings, 2013, p. 2208-2212
Issued Date	2013
URL	http://hdl.handle.net/10722/191606
Rights	IEEE International Symposium on Personal, Indoor and Mobile Radio Communications Proceedings. Copyright © I E E E.

A Framework for Balancing Information Collection and Data Transmission

Jie Chuai and Victor O.K. Li

Department of Electrical and Electronic Engineering,
The University of Hong Kong, Pokfulam, Hong Kong, China
Email: {jchuai, vli}@eee.hku.hk

Abstract—Network protocol performance is closely related to knowledge about the network state. However, acquiring such knowledge expends network bandwidth resource. Thus a trade off exists between the amount of bandwidth resource expended in acquiring knowledge about network state, and the improved protocol performance due to such knowledge. Previous work used rate distortion theory to calculate the minimum information required for certain network performance. However, this limit is asymptotic and might not be achievable due to the introduced infinite delay. This work develops a non-asymptotic framework to find a practical bound of the required information for certain network performance, and the strategies for implementing network information collection. The framework is illustrated by a wireless scheduling problem to show the quantitative relationship between collected traffic information and network throughput. Furthermore, we calculate the effective data rate by considering the overhead of network information collection, and determine the optimal resource allocation between information collection and data transmission.

I. INTRODUCTION

The performance of network communications is closely related to the acquired knowledge about the network state. If complete and accurate information is known to every entity (e.g., mobile nodes and base stations) in the network, optimal communication protocols could be designed to achieve the best performance. However, obtaining network information expends bandwidth resources, resulting in less resources for data transmission. Thus network resource allocation schemes should be designed for network information collection and data transmission. To tackle this problem, the fundamental question is “*How much information is required to achieve a given performance?*” By knowing this quantitative relationship, an optimal resource allocation strategy could be found.

Since information theory is a tool to measure information, it has been used in previous efforts [1]–[3] to address this issue. More precisely, rate distortion theory (see [4] for an introduction of rate distortion theory) was applied to give a theoretic bound of the amount of information required for a given network performance. Although rate distortion theory gives a systematic way to study this problem, it has limitations. In particular, rate distortion theory gives an *asymptotic* bound, which means infinite delay is introduced into the process of network information collection. This is not realistic since network communication works in real-time. In other words, the bound given by rate distortion theory might not be achievable when there is a delay constraint. Thus it is important to determine an *achievable* bound when the network information collection is conducted in real-time. In addition, the result

obtained by rate distortion theory only gives the quantity of information required; no implementation method is indicated. We desire to know *how to collect the information*.

The main contribution of this work is the development of a general framework to analyze the relationship between the quantity of known information and practically achievable network performance. We give not only a bound of required information but also the strategy of information collection. By applying this framework to specific network scenarios, we could design optimal resource allocation schemes for network information collection and data transmission.

The paper is organized as follows. Section II introduces the related work. Section III presents the framework for analyzing the minimum information required for a given network performance. Section IV applies the framework to a wireless scheduling problem to give the relationship between throughput and known traffic state information, and discusses the optimal channel sharing between information collection and data transmission by calculating the effective data rate. We discuss the applicability of our framework in Section V, and concludes this paper in Section VI.

II. RELATED WORK

Much work has been done to determine the limit of wireless network performance. For example, Gupta and Kumar calculated the capacity of wireless network [5]. [6] and [7] discussed the impact of interference on wireless network capacity. However, the above work assumed complete information (e.g., network connectivity, traffic patterns) of the network is available, which is not realistic.

Noting that network performance relies on the obtained knowledge of network state, some work discussed the effect of incomplete information on the performance of particular network protocols. For example, Hong and Li showed the quantitative relationship between network information and performance for a distance vector based routing algorithm [8] and a multi-hop wireless scheduling protocol [9].

Rate distortion theory was used to derive bounds of required information to achieve certain network performance, e.g., packet delivery ratio for different classes of routing protocols [1] [2]. Instead of analyzing specific protocols, Hong and Li [3] proposed a general framework which could be adopted to the overhead analysis of arbitrary networks and protocols. However, as mentioned in the previous section, the bounds are asymptotic which might be unachievable in practice, and no implementation strategy for information collection is given in the above work.

In this work, we also use information theory as the basis of analysis. We use the method of quantization to derive a practical bound. Some previous work has used this technique for transmitting network state information. For example, it is used to feed back channel state information from the receiver to the transmitter when the feedback link has a limited rate [10]. In [11], the authors applied this technique to paging systems to reduce location update cost at the expense of increased paging cost. Our work differs from the previous ones in that, we establish a general framework which could be applied to an arbitrary network scenario, and we aim at using this framework to derive the bound of minimum transmitted information for certain network performance requirements.

III. THE FRAMEWORK

In this section, we will introduce our framework for analyzing the relationship between the network performance and available network information. An example will be given in Section IV to illustrate the framework.

A. The Model

Network states (e.g., nodes locations, link states and traffic states) could be modeled as random variables with certain probability distributions, and network communication could be treated as a process of making decisions (e.g., scheduling the traffic) based on network states to achieve certain network performance requirements. Assume the true network state is X , and an observation Y associated with X is available. An entity (e.g., a central controller) observes Y and makes a decision Z . The resulted network performance (in terms of throughput, packet delivery ratio, etc.) is then a function of X and Z .

There are many ways for the observation Y to differ from the true state X , for example, erroneous measurements, noisy transmissions and incomplete information collection. In this paper, we only consider those due to incomplete information collection. The basic idea is that incomplete network information collection could be modeled as lossy information compression. As shown in Figure 1, the true network state X is first encoded into a codeword Y and transmitted, and the controller (i.e., the decoder) outputs a decision Z after observing Y . The above encoding-decoding process could be represented mathematically as

$$\begin{aligned} \text{Encoding function: } Y &= f_1(X) \\ \text{Decoding function: } Z &= f_2(Y) \end{aligned} \quad (1)$$

Using the Bayes decision rule, the optimal decision of the controller given the transmitted information is

$$z = f_2(y) = \arg \max_z \sum_x p(x|y)g(x, z) \quad (2)$$

where y is the received codeword, $g(x, z)$ is the network performance given network state x and a decision z .

The expected network performance is

$$G(f_1, f_2) = E[g(X, Z)] = \sum_{x \in \mathcal{X}} p(x)g(x, f_2(f_1(x))) \quad (3)$$

where \mathcal{X} is the alphabet (the set containing all the possible values) of X .

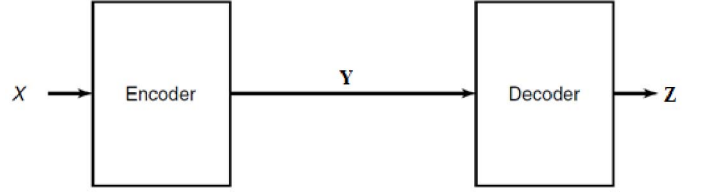


Fig. 1: Information Collection of a Network Protocol

B. Information Transmission Rate

The term *information transmission rate* refers to the bandwidth (in number of bits) used to transmit encoded network information Y for each network state. Assume the probability distribution of Y is $q(y)$. According to [4], the average number of bits R for each codeword by using optimal coding (e.g., Huffman coding) is bounded by

$$H(Y) \leq R < H(Y) + 1 \quad (4)$$

where $H(Y) = -\sum_y q(y) \log q(y)$ is the entropy of Y .

Since the entropy $H(Y)$ is the lower bound of R and it is a good estimate of the bandwidth occupied by the codewords, we use the following definition of information transmission rate in this paper.

Definition 1: For the system described in Figure 1, the information transmission rate is defined to be the entropy rate $H(Y)$.

C. Minimum Information Transmission Rate Required

Given a required network performance threshold G_{th} (e.g., a required throughput), the minimum information transmission rate is

$$\begin{aligned} R^*(G_{th}) &= \min_{f_1, f_2} H(Y) \\ \text{s.t. } G(f_1, f_2) &\geq G_{th} \end{aligned} \quad (5)$$

where $G(f_1, f_2) = E[g(X, Z)] = \sum_{x \in \mathcal{X}} p(x)g(x, f_2(f_1(x)))$.

Note that here we assume a higher value of $G(f_1, f_2)$ corresponds to a better performance. But the framework can also be applied to cases where a lower value gives better performance (e.g., when the performance metric is the average packet delay).

The above is equivalent to the optimal entropy-coded quantizer design problem, and could be solved by using the generalized Lloyd algorithm [12]. The optimal solutions f_1^* and f_2^* can then be implemented for the encoder and decoder. Thus the information collection strategy is readily available with this framework.

IV. EXAMPLE

In this section, we will use a two-node wireless scheduling problem to illustrate our framework. The approach can be generalized to N nodes.

A. Network Model

Assume two nodes 1 and 2 wish to send packets to the same receiver. Time is slotted and the length of each slot equals the time required to transmit a packet. In each time slot, if only one node transmits, there is a successful packet transmission; if two nodes transmit together, collision occurs and no packet is received successfully. Assume the traffic states of successive time slots are independent, i.e., we do not consider the correlation between time slots due to retransmissions. Assume X_1 and X_2 are the traffic states of nodes 1 and 2, respectively, in a time slot. X_1 and X_2 are independent and follow the same Bernoulli distribution, i.e.,

$$\begin{aligned} Pr(X_1 = 1) &= Pr(X_2 = 1) = p \\ Pr(X_1 = 0) &= Pr(X_2 = 0) = 1 - p = q \end{aligned} \quad (6)$$

The two nodes encode their traffic information separately into codewords Y_1 and Y_2 , i.e.,

$$Y_1 = f_1(X_1) \quad Y_2 = f_2(X_2) \quad (7)$$

The controller chooses an optimal scheduling decision Z from the set $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$ based on Y_1 and Y_2 . $(0, 0)$ means no node is allowed to transmit; $(1, 0)$ and $(0, 1)$ assigns the time slot to nodes 1 and 2, respectively; and $(1, 1)$ allows both nodes to transmit. The optimal decision z given y_1 and y_2 is

$$z = h(y_1, y_2) = \arg \max_z \sum_{x_1, x_2} p(x_1 x_2 | y_1 y_2) g(x_1 x_2, z) \quad (8)$$

The expected network performance is

$$G(f_1, f_2, h) = E[g(X_1 X_2, Z)] = \sum_{x_1, x_2} p(x_1 x_2) g(x_1 x_2, z) \quad (9)$$

where $z = h(f_1(x_1), f_2(x_2))$.

Thus given a throughput threshold G_{th} , the minimum information transmission rate is given by

$$\begin{aligned} R^*(G_{th}) &= \min_{f_1, f_2, h} (H(Y_1) + H(Y_2)) \\ s.t. \quad &G(f_1, f_2, h) \geq G_{th} \end{aligned} \quad (10)$$

B. Optimal Solutions

In this case, since a node only has two states, only four encoding strategies exist: each node can transmit no information (the codeword $Y_i = idle$, which means the channel is kept idle) or its true state ($Y_i = X_i$). The information transmission rates and throughputs (the probability of successful packet transmission in each time slot) of the four cases are discussed in the following:

- 1) $Y_1 = idle, Y_2 = idle$:

No traffic information is transmitted. The same decision is used all the time (thus the controller is not required). The optimal decision Z and the throughput G is

$$\begin{aligned} \text{For } p \geq 0.5, \quad Z &= (1, 0) \text{ or } (0, 1), \quad G = p \\ \text{For } p < 0.5, \quad Z &= (1, 1), \quad G = 2pq \end{aligned} \quad (11)$$

Transmission rate $R = 0$.

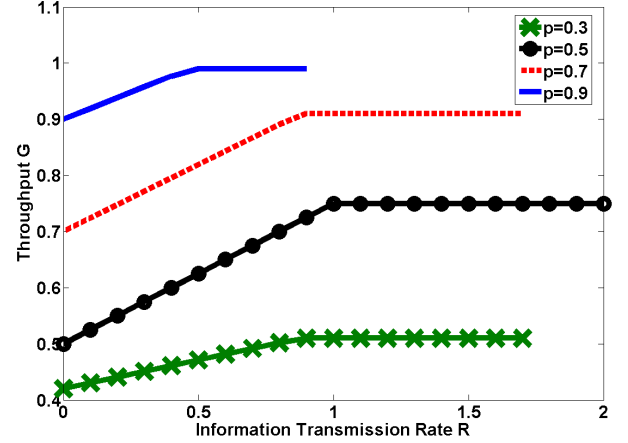


Fig. 2: Required transmission rate R for different values of G and p

- 2) $Y_1 = idle, Y_2 = X_2$:

The decoder knows the state of node 2. The optimal decisions and throughput are

$$\begin{aligned} Z(idle, 0) &= (1, 0) \quad \text{and} \quad Z(idle, 1) = (0, 1) \\ G &= Pr(X_2 = 0)Pr(X_1 = 1) + Pr(X_2 = 1) \\ &= 1 - q^2 \end{aligned} \quad (12)$$

Transmission rate $R = H(p)$.

- 3) $Y_1 = X_1, Y_2 = idle$

The decoder knows the state of node 1. The optimal decisions and throughput are

$$\begin{aligned} Z(0, idle) &= (0, 1) \quad \text{and} \quad Z(1, idle) = (1, 0) \\ G &= Pr(X_1 = 0)Pr(X_2 = 1) + Pr(X_1 = 1) \\ &= 1 - q^2 \end{aligned} \quad (13)$$

Transmission rate $R = H(p)$.

- 4) $Y_1 = X_1, Y_2 = X_2$

Complete information is obtained. One optimal decision and the corresponding throughput are

$$\begin{aligned} Z(0, 0) &= Z(0, 1) = (0, 1) \\ Z(1, 0) &= Z(1, 1) = (1, 0) \\ G &= 1 - Pr(X_1 = 0)Pr(X_2 = 0) = 1 - q^2 \end{aligned} \quad (14)$$

Transmission rate $R = 2H(p)$.

C. Plot of R - G Relationship

The relationship between information transmission rate R and network throughput G for different values of p is shown in Figure 2. As shown in Figure 2, for each p , throughput is at the minimum when no information is transmitted; and the maximum throughput is achieved when complete traffic information of one node, i.e., $H(p)$, is obtained. When transmission rate is above $H(p)$, there is no performance improvement due to this additional information. This is intuitive, since knowing the exact state of one node can give optimal decision for the two node scheduling problem. To achieve the maximum throughput, $p = 0.5$ requires the largest amount of information, since

the uncertainty of network state is the largest in this case¹; while $p = 0.9$ requires the smallest amount of information since $H(p)$ is the lowest among the four values of p . Note that for each p , the four encoding strategies correspond to transmission rates of 0, $H(p)$, $H(p)$ and $2H(p)$, respectively. The intermediate transmission rates could be obtained by time multiplexing between different encoding strategies. No causal² encoding strategy can achieve better performance than this linear bound [13].

With the linear relationship, we could get an equation of R - G relationship for each p :

$$1) \quad R \leq H(p)$$

$$G = \frac{pq}{H(p)}R + p \quad \text{for } p \geq 0.5 \quad (15)$$

$$G = \frac{p^2}{H(p)}R + 2pq \quad \text{for } p < 0.5 \quad (16)$$

$$2) \quad R > H(p)$$

$$G = 1 - q^2 \quad (17)$$

D. Comparison with the Asymptotic Bound

As mentioned, previous work used rate distortion theory to obtain the required information for a given network performance. This bound is asymptotic, which might not be achievable in practice. The framework in our paper provides a method to calculate the practical bound. The asymptotic and non-asymptotic bounds with $p = 0.5$ are compared in Figure 3.

The calculation of the asymptotic bound is given in [3]. In Figure 3, we use the performance metric, i.e., throughput, rather than the measure of performance degradation in [3]. In addition, the protocol in [3] is different from the one in our paper: instead of assigning the time slot to only one node, we allow two nodes to transmit together when p is smaller than 0.5. But for $p = 0.5$, the protocol in [3] is equivalent to ours.

As could be seen from Figure 3, for a given throughput, information required in practice is larger than that obtained by rate distortion theory. The two bounds coincide when $R = 0$ and 1, where no information and complete information of one node is transmitted, respectively. When R is between 0 and 1, the asymptotic bound achieves lower rate for a given performance by coding over an infinitely long block (i.e., infinite delay). However, in practice, when delay is not allowed we can only achieve the linear bound by time-multiplexing the zero information and full information transmission schemes.

E. Effective Data Rate

We have shown that the network performance increases with the increased amount of acquired network information. However, acquiring the information expends network bandwidth resource. As mentioned, we aim at finding the optimal

¹In fact, the required information transmission rate in Figure 2 is an underestimate of the actual bandwidth occupied. In this particular case, 1 bit is required to transmit the state information of one node by optimal coding regardless of the value of p , since the codeword length should be an integer. However, this difference does not affect the analytical procedures of our framework.

²The term *causal* means the produced decision z is a function of past and present input state x , and does not depend on future inputs. See [13] for details. The bound given by rate distortion theory is generally obtained by non-causal coding, and thus not achievable in practice.

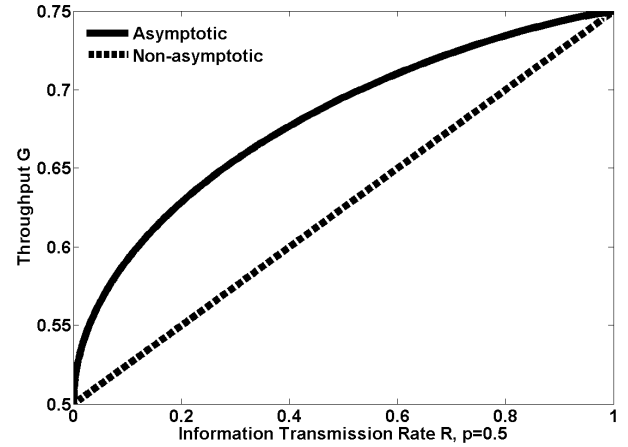


Fig. 3: Comparison of the asymptotic and non-asymptotic bounds

resource allocation between information collection and data transmission. Therefore, in this part we consider the effective data rate available with information collection. Assume there is a minislot of length R bits in front of each data slot of μ bits. Assume the throughput is G (the probability of a successful packet transmission in each time slot). Therefore, the effective data rate (in bits per channel use)³ is given by

$$R_e = \frac{\mu G}{\mu + R} \quad (18)$$

Combining Equations 15, 16 with 18, we get the relationship of R_e and R ,

$$R_e = \frac{\mu p + \frac{\mu p q}{H(p)} R}{\mu + R} \quad \text{for } p \geq 0.5 \quad (19)$$

$$R_e = \frac{2\mu p q + \frac{\mu p^2}{H(p)} R}{\mu + R} \quad \text{for } p < 0.5 \quad (20)$$

Differentiating R_e with respect to R yields the following conditions where R_e increases with R :

$$\mu q > H(p) \quad \text{for } p \geq 0.5 \quad (21)$$

$$\mu p > 2qH(p) \quad \text{for } p < 0.5 \quad (22)$$

Figure 4 shows the effective data rate versus R for $p = 0.8$ and various packet lengths. It is obvious that collecting traffic information is more useful when packets are longer, because the overhead due to information collection is smaller.

Figure 5 shows the minimum packet length required for R_e to increase with R for various values of p . For packet lengths smaller than this bound, no traffic information should be collected. The time slot is assigned to one node if $p \geq 0.5$, and both nodes are allowed to transmit if $p < 0.5$. For packet lengths greater than this bound, $H(p)$ bits should be transmitted to achieve the optimal effective data rate. Note

³A *channel use* means to transmit a symbol using the common channel. We assume binary symbols (i.e., 1 or 0) in this paper. Thus *1 bit per channel use* means 1 data bit is conveyed by transmitting a binary symbol through the channel.

that the bound reaches the minimum at $p = 0.5$ where the network state has the greatest uncertainty; when the uncertainty decreases, the performance improvement due to information collection decreases and the required packet length increases.

V. DISCUSSION

The approach we used could be applied to more complicated problems. For example, we can collect traffic information from N nodes and make scheduling decisions for multiple time slots. The basic steps of applying this model are as follows: first, identify the information source and the probability distribution; second, list the possible decisions that could be applied; third, determine the performance measure for each pair of network state and decision.

The effects of information delay and transmission errors are not considered in this work. Intuitively, with the presence of delay and error, the value of information will decrease and the resource allocation scheme will change accordingly. We plan to address these issues in our future work.

The network states at different time instants are assumed independent in this work. In reality, network states change over time, and states at different time instants are correlated generally. It would be interesting to design the resource allocation strategies when considering the temporal correlations of network states.

The above indicates that we could construct more sophisticated models to investigate the relationship between network performance and collected information. But the model in this work provides a basic framework to examine this problem, and gives insights on how available information affects network protocol performance.

VI. CONCLUSION

In this work, we developed a non-asymptotic framework to analyze the relationship between the amount of transmitted information and network performance. The basic idea is that network information is first compressed using quantization and then transmitted to a decision maker to make protocol decisions. We illustrated our framework by using a wireless scheduling problem for two nodes. We analyzed the relationship between network throughput and the amount of collected traffic state information. Besides the quantitative relationship, our framework also gave the implementation of information collection. We calculated the effective data rate by considering the overhead due to information collection. Based on these, we gave the optimal resource allocation between data transmission and information collection.

REFERENCES

- [1] N. Bisnik and A. A. Abouzeid, "Capacity deficit in mobile wireless ad hoc networks due to geographic routing overheads," in *Proc. IEEE INFOCOM*, May 2007, pp. 517–525.
- [2] D. Wang and A. A. Abouzeid, "Link state routing overhead in mobile ad hoc networks: A rate-distortion formulation," in *Proc. IEEE INFOCOM*, Apr. 2008, pp. 1337–1345.
- [3] J. Hong and V. O. K. Li, "Impact of information on network performance - an information-theoretic perspective," in *Proc. IEEE GLOBECOM*, Dec. 2009, pp. 1–6.
- [4] R. G. Gallager, *Information theory and reliable communication*. Springer-Verlag, 1970.
- [5] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar 2000.

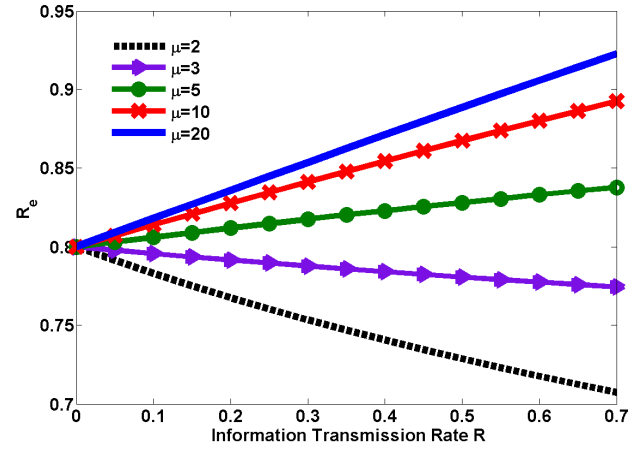


Fig. 4: R_e versus R with $p = 0.8$

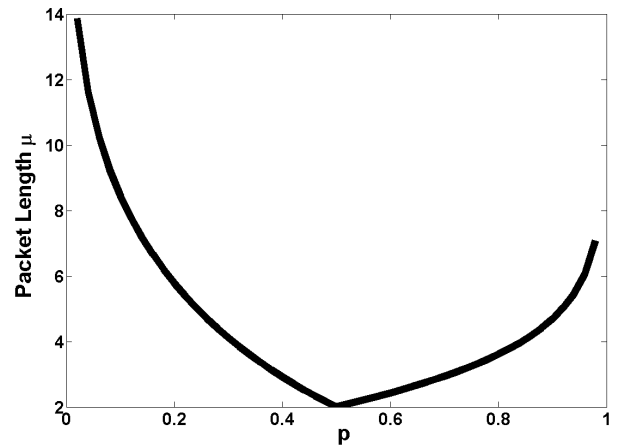


Fig. 5: Minimum packet length required for R_e increasing with R

- [6] K. Jain, J. Padhye, V. N. Padmanabhan, and L. L. Qiu, "Impact of interference on multi-hop wireless network performance," in *ACM MobiCom*, 2003, pp. 66–80.
- [7] M. Kodialam and T. Nandagopal, "The effect of interference on the capacity of multihop wireless networks," in *IEEE ISIT*, Jun 2004.
- [8] J. Hong and V. O. K. Li, "Effect of information on routing performance in multi-hop wireless networks," in *IEEE VTC*, May 2010, pp. 1–5.
- [9] —, "The effect of information on scheduling performance in multi-hop wireless networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 3238–3246, Oct. 2010.
- [10] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sept. 1999.
- [11] A. Roy, A. Misra, and S. K. Das, "Location update versus paging trade-off in cellular networks: An approach based on vector quantization," *IEEE Trans. Mobile Comput.*, vol. 6, no. 12, pp. 1426–1440, Dec. 2007.
- [12] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.
- [13] D. Neuhoff and R. Gilbert, "Causal source codes," *IEEE Trans. Inf. Theory*, vol. 28, no. 5, pp. 701–713, Sep 1982.