The HKU Scholars Hub   The University of Hong Kong   香港大學學術庫

Title	**Mathematical impact of information accuracy on network performance**
Author(s)	**Geng, Y; Li, VOK**
Citation	**The 2010 IEEE International Conference on Communications (ICC 2010), Cape Town, South Africa, 23-27 May 2010. In Proceedings of ICC, 2010, p. 1-8**
Issued Date	**2010**
URL	**http://hdl.handle.net/10722/126163**
Rights	**IEEE International Conference on Communications. Copyright © IEEE.**

Mathematical Impact of Information Accuracy on Network Performance

Yanhui Geng, *Member, IEEE*, Victor O.K. Li, *Fellow, IEEE*
Department of Electrical and Electronic Engineering
The University of Hong Kong, Hong Kong SAR, China
Email: {yhgeng, vli}@eee.hku.hk

Abstract—Many wireless network protocols have to deal with inaccurate information due to the lack of sufficient knowledge of the scenario or other limitations. Therefore, in this paper, we study information accuracy and investigate its quantitative impact on wireless network performance. First, we introduce an entropy-performance framework to model the relationship between wireless network performance and entropy, which characterizes the uncertainty of the input. Under this framework, we quantify the performance variations due to the availability of side information and find that the system performance improvement has a positive linear relationship with the amount of mutual information between input and side information. Subsequently, information accuracy is proposed to reflect the reliability of information. We show that information accuracy has a natural relationship with mutual information and we quantify the impact of information accuracy on wireless network performance.

Index Terms—information theory, network performance, side information, information accuracy

I. Introduction

In wireless networks, protocols and applications need some information fed back by network components to achieve better performance. A generally accepted intuition is that more information will lead to better performance, since the information can help reduce the uncertainty and allow the network to make better decision. But this result holds only when all reported information is 100% accurate, i.e., the event claimed in a message will occur definitely and the uncertainty will be totally eliminated with this information.

But sometimes, perhaps due to the lack of sufficient knowledge or the presence of inherent uncertainty in the network, we have to deal with inaccurate or estimated information, which may cause performance degradation. Thus, it is important to study the accuracy of information and its impact on wireless network performance. In this paper, we will show how the wireless network performance will change with side information available to the system and how the information accuracy of such information will affect the wireless network performance quantitatively. To the best of our knowledge, this is the first such study.

The rest of this paper is organized as follows. Related work is discussed in Section II. Section III introduces the entropy-performance framework and derives some theoretical results for later analysis. Then in Section IV, we investigate the impact of information accuracy on wireless network performance. Wireless scheduling is used to illustrate our analysis in Section V. Finally we summarize this paper in Section VI.

II. Related Work

In 1956, J. L. Kelly [1] studied the gambling game of horse racing and found that the increased doubling rate (a measure of wealth growth rate) due to side information was the same as the mutual information between side information and the winning probability distribution of the horses. In the field of communication, there are related work on studying side information and most of them focuses on the source coding and lossless compression problems [2]. In networking research, there also exists some research on side information. In [3] and [4], routing and asymptotic performance analysis with side information in sensor networks are studied, respectively.

There are also some previous efforts on the accuracy of information. [5] studies how to utilize inaccurate information to support energy-efficient routing in mobile Ad-hoc networks. In [6], the authors discuss how to design scheduling algorithms in sensor networks to reduce energy consumption while maintaining the predefined information accuracy. In [7], the authors study how to achieve greater information accuracy by using more sensing nodes jointly in wireless sensor networks.

From the related work introduced above, we can see that the study of information accuracy in networking research mainly focuses on specific application scenarios. Investigation of general networks is not considered and a theoretical study of the quantitative relationship between information accuracy and wireless network performance is still lacking. This is the motivation of our research in this paper.

III. Entropy-Performance Model

A. Guessing and Matching Network Models

In many applications of networking, one has to deal with some inputs with uncertainties, e.g. there may be uncertainty in node positions in a scheduling problem, due to the lack of sufficient knowledge or some inherent randomness. Under this situation, the system will make decisions based on estimates of the inputs. The model of system with uncertainty is shown in Fig. 1.
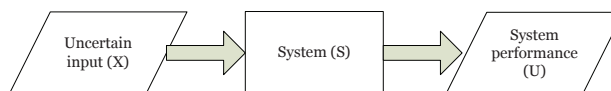

Fig. 1. Model of system with uncertainty

The total uncertainty sensed by the system without side information is $H(X)$. We use $U(H(X))$ to denote the original system performance, which is a function of entropy $H(X)$ introduced by uncertain input X. We can investigate the property of system performance function U from the perspective of guessing error. A large value of $H(X)$ means that the degree of uncertainty of input X is high, and the system cannot guess the value of X easily, i.e. the guessing error probability is high. High guessing error probability usually means bad system performance.

A representative type of network applications involves iterative operations and the system performance is accumulated; thus, $U_n = \sum_{i=1}^{n} U_i$. In each iteration, if the system's guessing of the uncertain input is correct and hence a correct decision is made, the system will receive a performance reward θ ($\theta \geq 0$); otherwise, the system will receive a performance punishment δ ($\delta \geq 0$). We call this kind of network systems "guessing and matching" systems. Transmission scheduling in wireless networks is a typical "guessing and matching" application, in which one scheduling action in a time slot forms one iteration, and the number of successful transmissions is accumulated as a performance metric. By setting $\theta = 1$ and $\delta = 0$, this scheduling problem fits into our "guessing and matching" network model, since usually a successfully scheduled transmission will contribute to the system throughput by a certain amount while a failed one will not cause performance punishment.

If the system experiences a guessing error probability P_e in each iteration, we can have the following theorem.

Theorem 3.1: For a "guessing and matching" system with guessing error probability P_e, the average system performance is $U(P_e) = \theta - (\theta + \delta)P_e$.

Proof: Since in the long run, the system will make a fraction $1 - P_e$ of correct decisions and a fraction P_e of wrong decisions, the average performance is,

$$U(P_e) = \frac{1}{n}U_n = (1 - P_e)\theta - P_e \cdot \delta$$
$$= \theta - (\theta + \delta)P_e \qquad (1)$$

∎

Corollary 3.1: The optimal system performance of a "guessing and matching" system, denoted by U^*, is,

$$U^* = \frac{1}{n}U_n(P_e = 0) = \theta, \qquad (2)$$

and the optimum is achieved when the guessing error probability P_e is zero, i.e., the system is guessing error-free.

Proof: The result can be obtained directly from Theorem 3.1, since the maximal $U(P_e)$ is achieved when $P_e = 0$ according to (1). ∎

B. Entropy-Performance Relationship

Now we will relate the entropy of input $H(X)$ with the system's guessing error probability P_e. Let $X = \{s_1, s_2, \ldots, s_N\}$ be the state space of the system, and $P_1 \geq P_2 \geq \cdots \geq P_N$, where $P_i = P(s_i), i = 1, 2, \ldots, N$. Then the best guess of X is $X' = s_1$ and the resulting error probability is $P_e = 1 - P_1$.

According to Fano's inequality [8], which is elaborated in detail in [9]. We can get (3).

$$H(P_e) + P_e \log|X| \geq H(X) \qquad (3)$$

Then we have,

$$P_e \geq \frac{H(X) - 1}{\log|X|} \qquad (4)$$

It means that the smallest guessing error probability P_e^{min} with given entropy $H(X)$ is,

$$P_e^{min} = \frac{H(X) - 1}{\log|X|} \qquad (5)$$

Put (4) into (1) and we now have,

$$U(H(X)) \leq \theta - (\theta + \delta)\frac{H(X) - 1}{\log|X|} \qquad (6)$$

Inequality (6) reveals the upper bound of the performance of wireless network systems with uncertain input X characterized by entropy $H(X)$.

Theorem 3.2: For a "guessing and matching" network system with uncertain input X characterized by $H(X)$, its optimal average system performance, denoted by $\hat{U}(H(X))$, is,

$$\hat{U}(H(X)) = \max U(H(X)) = U(P_e^{min})$$
$$= \theta - (\theta + \delta)\frac{H(X) - 1}{\log|X|} \qquad (7)$$

We can tell from the above theorem that the relationship between $\hat{U}(H(X))$ and $H(X)$ is negative linear with the number of states $|X|$ fixed.

With the number of possible states fixed, e.g., $|X| = N$, the minimal entropy of X is zero when any state is assigned with probability one while all other states are assigned with probability zero, and hence the corresponding maximal system performance is achieved,

$$\hat{U}^{max}(H(X)) = \theta + \frac{\theta + \delta}{\log N} \qquad (8)$$

The maximal entropy of X is $\log N$ and it is achieved when all states are assigned with equal probability $1/N$. The corresponding minimal system performance is as follows,

$$\hat{U}^{min}(H(X)) = \theta - (\theta + \delta)\frac{\log N - 1}{\log N} \qquad (9)$$

We can also extend the above arguments to more general wireless networks. Consider a general relationship f between U and P_e determined by the network scenario, denoted as,

$$U(P_e) = f(P_e) \qquad (10)$$

As shown in (5), there always exists a positive linear relationship between P_e^{min} and $H(X)$, denoted as,

$$P_e^{min} = L(H(X)) \qquad (11)$$

then,

$$U(P_e^{min}) = f(P_e^{min}) = f(L(H(X)))$$
$$= g(H(X)) \qquad (12)$$

Since replacing the variable X of the mapping function f with a linear function L with variable X will not alter the properties or the original function f such as the non-decreasing trend, we have the following corollary.

Corollary 3.2: In a general network system, the properties of the function f mapping P_e to system performance U are identical with the function g mapping entropy $H(X)$ to the system performance U.

IV. IMPACT OF INFORMATION ACCURACY ON WIRELESS NETWORK PERFORMANCE

A. Investigation of Side Information

In Section III, we have introduced the system model and quantitatively modeled the relationship between the entropy of the system's uncertain input X and the wireless network system U from the perspective of guessing error probability. It is obvious that the network system can perform better if some side information is available. This kind of information may be fed back by certain network components, such as nodes that can sense the system state. To facilitate our analysis of information accuracy, we will study the scenarios when side information Y is provided to the system and investigate the impact of side information Y on the system performance U. The system model with side information is shown in Fig. 2.
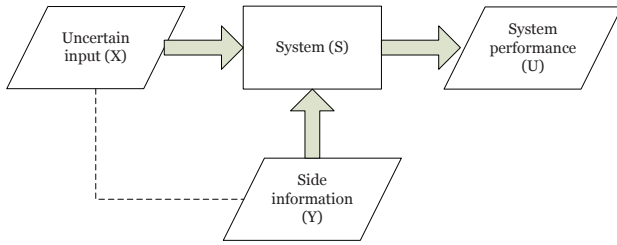

Fig. 2.   Model of system with side information

With side information Y provided, now the system will handle the uncertain input X with the assistance of Y. $X' = X|Y$ reflects the total uncertainty finally sensed by the system. The quantitative relationship between the two correlated random variables can be measured using conditional entropy $H(X|Y)$ or $I(X;Y)$. A higher correlation of X and Y means that Y contains more information about X, and hence implies a smaller $H(X|Y)$ value and a larger $I(X;Y)$ value.

We start the analysis with the best expected result, i.e., we want to eliminate all uncertainty introduced by X. To achieve this goal, we should deduce the value of X from Y with zero probability of error. According to Fano's inequality, $P_e = 0$ implies that $H(X|Y) = 0$. We know that the conditional entropy of the random variable X given another random variable Y is zero if and only if X is a function of Y. Therefore, we have the following lemma.

Lemma 4.1: If side information Y satisfies $H(X|Y) = 0$, the optimal performance of the wireless network systems, with an uncertain input X containing N states, will be improved from

$$\hat{U}(H(X)) = \theta - (\theta + \delta)\frac{H(X) - 1}{\log N}$$

to

$$\hat{U}(H(X')) = \hat{U}(H(X|Y)) = \theta + \frac{\theta + \delta}{\log N}$$

Now consider another extreme case.

Lemma 4.2: If side information Y is independent of the input X, the system performance will remain unchanged.

Proof: The independence between X and Y implies,

$$H(X') = H(X|Y) = H(X) \tag{13}$$

which means the uncertainty degree sensed by the system remains unchanged. Hence,

$$\hat{U}(H(X')) = \hat{U}(H(X)) \tag{14}$$

∎

Lemma 4.3: In general, side information Y is correlated with but not a function of the input X, and the mutual information between X and Y is $I(X;Y)$, then the optimal system performance will be improved by,

$$\Delta = \frac{\theta + \delta}{\log N}I(X;Y) \tag{15}$$

Proof: With side information Y, the total uncertainty sensed by the system changes from $H(X)$ to $H(X') = H(X|Y)$. According to (7), we have,

$$\Delta = \hat{U}(H(X|Y)) - \hat{U}(H(X))$$
$$= (\theta + \delta)\frac{H(X) - H(X|Y)}{\log N}$$
$$= \frac{\theta + \delta}{\log N}I(X;Y) \tag{16}$$

∎

$I(X;Y)$ takes a value within range $[0, H(X)]$ and the relationship between performance improvement and the mutual information is positive linear.

The maximal system improvement is achieved when $I(X;Y) = H(X)$ and it is,

$$\Delta_{max} = \frac{(\theta + \delta)H(X)}{\log N} \tag{17}$$

The most important observation we made from the above deduced lemmas forms the following theorem.

Theorem 4.1: The performance improvement of "guessing and matching" wireless networks with uncertain input X due to the availability of side information Y has a positive linear relationship with the amount of mutual information between X and Y.

Theorem 4.1 not only verifies the intuition of the impact of side information on system performance, i.e., the more information carried by Y about input X, the greater performance improvement achieved, but also quantifies the relationship between them for the "guessing and matching" wireless networks. With these results, we can better understand the impact of side information when it can be acquired and fed back to the system under some mechanisms, and hence analyze and predict the system behavior such as performance variations after the change of available information.

B. Impact of Information Accuracy

Information accuracy is defined as the degree of closeness of a measured quantity to its actual (true) value. We use probability to quantify the closeness and a larger probability means a higher accuracy. Let random variable X (with N possible states) be the actual object we want to investigate and information X' be the measured value of X. The information accuracy of X, denoted by q, is defined as follows,

$$q = P(X' = s_i | X = s_i), \forall i \in [1, 2, \ldots, N] \quad (18)$$

It means that information $X = s_i$ passing through an inaccurate channel has a probability q to remain being s_i, and a probability $\frac{1-q}{N-1}$ to be any other state $s_j, j \neq i$, the latter assuming the wrong guess will be mapped equally likely to any other states.

Now we will analyze the impact of information with information accuracy q on system performance, where $0 \leq q \leq 1$. The value of q means $P(Y = s_i | X = s_i) = q$ for any $i \in [1, 2, \ldots, N]$. For example, if the actual value of X is $X = s_1$, then the reported result Y is s_1 with possibility q, and it may also take values from s_2 to s_N, with equal probability $\frac{1-q}{N-1}$. This is shown in Fig. 3 below.
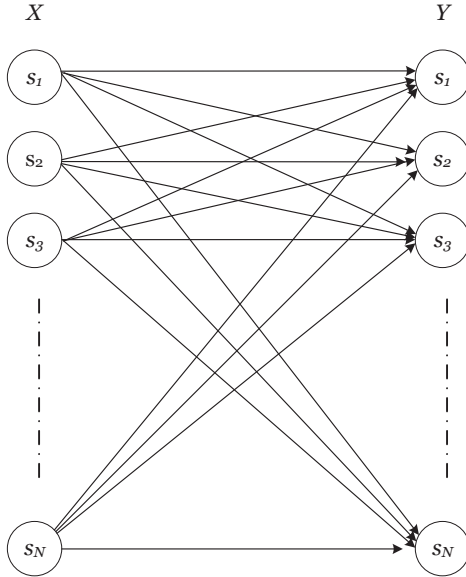

Fig. 3. Mapping of X to Y with information accuracy q

Based on this mapping relationship, we can calculate the probability distribution $P(Y)$ and the conditional probability distribution $P(X|Y)$ as follows. The event $X = s_i$ is abbreviated to X_i in the following and Y_i denotes "Y tells the system that the current state of uncertain input X is s_i."

$$P(Y_i) = \sum_{j=1}^{N} P(X_j)P(Y_i|X_j)$$

$$= P_i \cdot q + \frac{1-q}{N-1}(1 - P_i) \quad (19)$$

$$P_{i=j}(X_i|Y_j) = \frac{P_j q}{P_j q + \frac{1-q}{N-1}(1 - P_j)} \quad (20)$$

$$P_{i \neq j}(X_i|Y_j) = \frac{P_i \frac{1-q}{N-1}}{P_j q + \frac{1-q}{N-1}(1 - P_j)} \quad (21)$$

With the above equations, we can calculate the conditional entropy $H(X|Y)$.

$$H(X|Y) = \sum_{i=1}^{N} P(Y_i)H(X|Y_i) \quad (22)$$

Due to the complexity, we first focus on the calculation of the first term of the R.H.S. of (22).

$$P(Y_1)H(X|Y_1) = P_1 q \log \frac{P_1 q + \frac{1-q}{N-1}(1 - P_1)}{P_1 q}$$
$$+ \frac{P_2(1-q)}{N-1} \log \frac{P_1 q + \frac{1-q}{N-1}(1 - P_1)}{\frac{P_2(1-q)}{N-1}}$$
$$\vdots$$
$$+ \frac{P_N(1-q)}{N-1} \log \frac{P_1 q + \frac{1-q}{N-1}(1 - P_1)}{\frac{P_N(1-q)}{N-1}} \quad (23)$$

After some manipulations, the expression above becomes,

$$P(Y_1)H(X|Y_1)$$
$$= \left(P_1 q + \frac{1-q}{N-1}(1 - P_1)\right) \log \left(P_1 q + \frac{1-q}{N-1}(1 - P_1)\right)$$
$$- P_1 q \log P_1 - P_1 q \log q + \frac{1-q}{N-1}(H(X) + P_1 \log P_1)$$
$$- \frac{1-q}{N-1}(1 - P_1) \log \frac{1-q}{N-1} \quad (24)$$

Similarly, we can get the expression of $P(Y_i)H(X|Y_i)$ for each i, put it into (22), with,

$$\left(P_1 q + \frac{1-q}{N-1}(1 - P_1)\right) \log \left(P_1 q + \frac{1-q}{N-1}(1 - P_1)\right) \leq 1$$

and after some transformations, we can have the following inequality.

$$H(X|Y) \geq H(X) - N - q \log q - (1-q) \log \frac{1-q}{N-1} \quad (25)$$

Now,

$$I(X;Y) = H(X) - H(X|Y)$$
$$\leq N + q \log q + (1-q) \log \frac{1-q}{N-1} \quad (26)$$

According to Lemma 4.3, we can get the following theorem.

Theorem 4.2: If side information Y is an information source with information accuracy q on the uncertain input X, the system performance can be improved by,

$$\Delta \leq \frac{\theta + \delta}{\log N}\left(N + q \log q + (1-q) \log \frac{1-q}{N-1}\right) \quad (27)$$

Up to now, we have proved that there exists a natural relationship between information accuracy q and mutual information $I(X;Y)$. Furthermore, with the deduced result about mutual information and system performance in "guessing and matching" network systems, we have also quantified the impact of information accuracy on the system performance.

Theorem 4.2 also implies that higher information accuracy does not necessarily mean larger mutual information or better system performance since the relationship between q and the corresponding maximal $I(X;Y)$ is not monotonic increasing. Detailed quantitative relationship can be further derived according the framework proposed in [10].

V. ILLUSTRATION

We use numerical results of a wireless scheduling application to illustrate the impact of information accuracy on network performance. In this scenario, a base station (BS) is attempting to transmit to multiple wireless nodes. The channel condition of each node may be characterized by one of N states, such as, good, bad, very bad, etc. Side information of each node's channel condition is provided to BS with information accuracy q. If BS guesses the channel condition correctly and hence schedules a successful transmission, it will contribute 1 unit to system performance; while a wrong guess will schedule a failed transmission and consumes 0.5 unit of system resource. Therefore, this scenario fits into our model with $\theta = 1$ and $\delta = 0.5$.

According to Theorem 4.2, the analytical results of the relationship between information accuracy q ($0 \le q \le 1$) and performance variation δ with $N = 2, 5, 10$ are derived and plotted in Fig. 4. As we can see, with different values of N, the degrees of impact of information accuracy on performance are different. In all cases, minimal Δ does not happen when $q = 0$; because totally incorrect information actually tells the system something useful. In particular, when $N = 2$, information with $q = 0$ is identical to information with $q = 1$ since BS can deduce the correct channel condition by complementing the reported information. The minimal Δ occurs when $q = 0.5$, because this information cannot tell anything useful since two states will happen with equal possibilities. However, as N increases, information with small values of q is less useful and increasing q will generally generate better performance. This is not true in small scale networks. For example, as shown in Fig. 4, when $N = 2$, if we are presently operating at $q = 0.2$, we might as well stay at this operating region unless we can increase q beyond 0.8, because increasing q from 0.2 to 0.5 actually decreases the performance, and only when q is higher than 0.8 do we get better performance.

VI. CONCLUSION

In this paper, we first introduce a model of "guessing and matching" network systems and then propose a framework to capture and analyze the relationship between entropy and network performance in this network model. Under this entropy-performance framework, we quantify the performance variations due to the availability of side information and find that
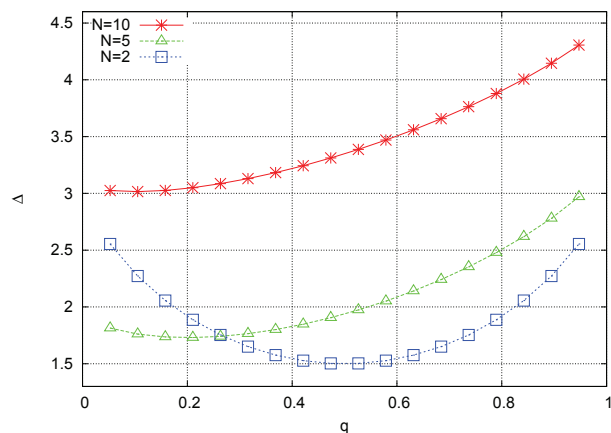

Fig. 4. Relationship between Δ and q

the performance improvement has a positive linear relationship with the amount of mutual information between the system input and the side information. Based on these results, we have further determined the quantitative relationship between information accuracy and mutual information, and the resulted relationship between information accuracy and the network performance. These theoretical results are insightful and they can be utilized to analyze and predict the behaviors of network systems which need to deal with inaccurate information. Practical wireless protocols are usually more complex than our introduced model due to complex control and fairness considerations. Therefore, the relationship between information accuracy and performance in more realistic wireless network models will be investigated in our future work.

REFERENCES

[1] J. Kelly, J., "A new interpretation of information rate," *Information Theory, IRE Transactions on*, vol. 2, no. 3, pp. 185–189, September 1956.
[2] E. Martinian, G. Wornell, and R. Zamir, "Source coding with distortion side information," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4638–4665, Oct. 2008.
[3] H. Luo and G. Pottie, "A study on combined routing and source coding with explicit side information in sensor networks," in *Proc. IEEE GLOBECOM '05*, vol. 5, Dec. 2005, pp. 2981–2985.
[4] W. P. Tay, J. Tsitsiklis, and M. Win, "Asymptotic performance of a censoring sensor network," *IEEE Transactions on Information Theory*, vol. 53, no. 11, pp. 4191–4209, Nov. 2007.
[5] R. Alhalimi and T. Kunz, "Energy level accuracy in mobile ad-hoc networks using OLSR," in *Proc. ICST WICON '07*, Brussels, Belgium, 2007, pp. 1–9.
[6] H. Li, S. Jiang, and G. Wei, "A theoretical framework on power consumption subject to information accuracy in wireless sensor networks," in *Proc. 2nd International Conference on Mobile Technology, Applications and Systems*, Nov. 2005, pp. 1–6.
[7] K. Cai, G. Wei, and H. Li, "Information accuracy versus jointly sensing nodes in wireless sensor networks," in *Proc. IEEE Asia Pacific Conference on Circuits and Systems*, 30 2008-Dec. 3 2008, pp. 1050–1053.
[8] R. M. Fano, *Transmission of information: A statistical theory of communications*. M.I.T. Press, 1961.
[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, June 2006.
[10] Y. Geng and V. O.K. Li, "A framework to analyze network performance based on information quality," in *Proc. IEEE ICC 2010*, Cape Town, South Africa, May 2010.