2H1-OS4-6

On Computational Complexity of Pathway-Inspired Networking

Jian-Qin Liu

Kobe Advanced ICT Research Center
National Institute of Information and Communications Technology

In order to understand cell communication mechanisms based on the dynamics of signal transduction, in this paper an algorithm for constructing a pathway network based on binary states whose structure is equivalent to a well-known one's in linear network coding is proposed. The corresponding computational complexity of enzyme nodes in the pathway network is briefly discussed so as to improve the efficiency of the networking process.

1. Introduction

In essence, the signaling pathways are biochemical processes. By means of bioinformatics, an information processing model can be formulated by mathematics which can help us to understand and elucidate the biochemical mechanism of cellular functions. Graph theory and nonlinear dynamics are two supporting tools for the formalization of signaling pathways. Here linear network coding is suggested to formalize the structure of information processing of signaling pathways. As a promising paradigm of information theory, linear network coding [Li 2003, Li 2005] shows advantages in the research and application of communication networks. Based on the abstract structure of networks which rely on the store-and-forward principle [Li 2009], in this paper an equivalence between a network structure given by linear network coding in Fig. 1(b) of [Li 2003] and a formalized network structure of signaling pathway networks [Liu 2007] is established. The parallelism of signaling pathways in cells shows strong potential in efficiency, which form a striking contrast to its counterpart in computer communications. The computational complexity, as we know, is an important aspect in the algorithm design and deserves to be studied.

2. Preliminaries

In the cell, the biochemical reactions of molecules such as DNA, RNA, proteins and other molecules are the basis of cell communications, which influence major functions of the cell. One of the most fundamental signals of the cell communication is concentration, which is an analog signal. The biochemical reactions can be described by the equations such as Michaelis-Menten equation. By observing the temporal process of the chemical signals in the cell, the synchronization and non-synchronization mechanism of cellular signaling can be analyzed.

In cells, there are two kinds of reversible switches of molecular signals, i.e., the switch of phosphorylation and dephosphorylation and the switch of GTP-bound and GDP-bound states of GTPases. With respect to these reversible switches mentioned above, kinase and phosphatase activate the phosphorylation and dephosphorylation processes, respectively. And GEF and GAP activate the processes of GTP-bound and GDP-bound states of GTPases, respectively. The two states provide a binary representation by moleware. Here, the phosphorylation and GTP-bound state are defined as 1, the dephosphorylation and GDP-bound state are defined as 0. The molecule with the state of phosphorylation or dephosphorylation and GTPase with the state of the GTP-bound or the GDP-bound can be used to represent the variable. Bioinformatics models based on the above-mentioned switches can help us to understand the phenomena in biology. For example, based on the GTP/GDP and the kinase/phosphatase switches, it is possible to use a pathway-inspired networking algorithm to analyze the feedback mechanism of the switching process for mitosis/meiosis in fission yeast.

Contact: Jian-Qin Liu, Kobe Advanced ICT Research Center, National Institute of Information and Communications Technology, 588-2, Iwaoka, Iwaoka-cho, Nishi-ku, 651-2492, Kobe, Hyogo, Japan Phone 078-969-2227, Facsimile 078-969-3928, E-mail liu@nict.go.jp

3. Information Processing Mechanism of Signaling Pathways by Linear Network Coding

3.1 Constructing a Pathway-inspired Network

Here we adopt the notations of Fig. 1(b) in [Li 2003] and use the structure given in the same figure as the reference to formalize the signaling pathways. Let *path (A, E)* be a predicate that denotes a pathway whose input is *A* and output is *E*. A predicate-based logic equals to a kind of primitive in programming. A pathway network is constructed as shown in Fig.1. The construction algorithm is proposed as follows:

Step 1: Source *S* is set as the starting point of the intra-cell communication process of signaling pathways and is mapped to the output of a switch of phosphorylation and dephosphorylation processes. Here the binary values of data bits b1 and b2 (1 and 0) are defined by the phosphorylation and dephosphorylation states, respectively. The two states can also be mapped to the GTP-bound or GDP-bound states of GTPases.

Step 2: The channel from *S* to *T* and the channel from *T* to *W* are simplified as one process, that is, they are mapped to one (active or passive) transportation process of the cell and won't change its state of a signaling molecule, which is used to represent *b1*. Let the molecules at *S* node be denoted as *m(b1)* and *m(b2)*. We have *path (m(b1), m(b1))* for the channel from *S* to *T* denoted as *channel* (*S*,*T*) and the channel from *T* to *W* denoted as *channel* (*T*, *W*). The channel from *S* to *U* and then

from *U* to *W* are defined in a similar way and is denoted as *path (m(b2), m(b2))*.

Step 3: *b1* and *b2* meet at node *W*, which construct a block of data. The molecules *m(b1)* and *m(b2)* will be bound to the molecule which represents the node *W* and is denoted as *m(W)* in different sites of *m(W)*. The binding operation for *b1* and *b2* corresponds to the encoding operation in linear network coding. The notation of *b1+b2* in linear network coding is transformed into the compound of *m(b1)-m(W)-m(b2)*.

Step 4: The channels from *W* to *X* and from *X* to *Y* are unified and defined as a transportation process denoted as *path(m(b1)-m(W)-m(b2)*, *m(b1)-m(W)-m(b2))*. The channels from *W* to *X* and from *X* to *Z* are unified and defined in the same way.

Step 5: At *Y* and *Z*, a pathway is used to separate the bound *b1* and *b2* into the individual state of *b1* and *b2*, respectively.

Since the sites of the compound are the references to separate the *m(b1)* and *m(b2)* from the compound *m(b1)-m(W)-m(b2)*, the channel from *T* to *Y* and the channel from *U* to *Z* are equivalently reflected here, i.e., the information of the sites for *Y* equals to the reference of the channel from *T* to *Y* and the information of the sites for *Z* equals to the reference of the channel from *X* to *Z*. We use *site(m(b1)* to represent the channel from *T* to *Y* and *site(m(b2))* to represent the channel from *U* to *Z*. Accordingly, the decoding in nonlinear network coding is defined.

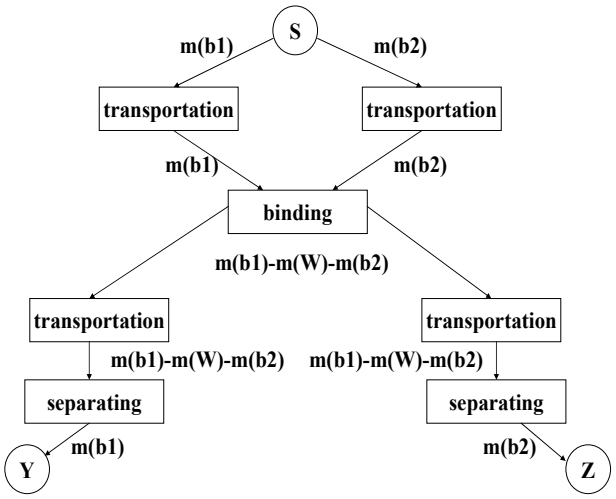

Figure 1 Pathway Network

In concept the structure of nonlinear network coding is mapped into an information network of signaling pathways. Besides the modules of pathways (links) and molecules (nodes), enzymes are needed for the activation of the pathways in biochemistry. This gives arise to a question on how many enzymes are needed for activating the pathways that correspond to the structure of linear network coding.

3.2 Computational Complexity of the Networking Process Inspired by the Signaling Pathways

In order to make a minimum configuration of pathways for linear network coding that guarantees the synchronization for encoding and decoding, at least three enzymes – *e(TW, UW)*, *e(XY, XZ)*, and *e(TY, UZ)* are needed for the synchronization between *channel (T,W)* and *channel (U,W)*, and between *channel (X,Y)* and *channel (X,Z)*, as well as for the synchronization of *e(TW, UW)* and *e(XY, XZ)*, respectively. *e(TY, UZ)* equals to the operation for the synchronization of *channel (T,Y)* and *channel (U,Z)*.

On the synchronization by the enzymes, the speed difference of the two pathways will be positive or negative. The enzyme will have two roles – to increase the speed of one pathway or to decrease its speed when compared with another pathway in the pair of channels mentioned above. Let the three enzymes be X_1, X_2, and X_3. The two states of the enzymes are defined as X_i and $\neg X_i$, respectively. Assuming that we have m units of pathways and each corresponds to an above-mentioned network of linear network coding, the combinatorial form of the enzymes for activating the synchronization of the molecules in the pathways selected from a set with n size is a 3-SAT problem, which is a NP one. Let l be the number of the enzymes used in the unit of pathways. The computational complexity denoted as $Q(m,n,l)$ under the condition of minimum number of enzymes in the unit is the total number of the states of the enzymes for activating the synchronization in terms of information theory supposing the related cellular signaling processes are message transmission processes. We have that

$$Q(m,n,3) < 2^n . \quad (1)$$

Considering the mechanism of signaling pathway that one phosphorylation can activate another dephosphorylation or vice versa, the complexity will be reduced as follows:

$$Q(m,n,2) < Q(m,n,3). \quad (2)$$

If the enzymes can activate the pathways simultaneously, then we have that

$$\max Q(m,n,3) = n. \quad (3)$$

4. Conclusion

In order to fill the gap between the biochemistry of cells and the informatics of computer communication networks and to provide an information-theoretic tool for bioinformatics in general, a configuration of signal pathways for the linear network coding is designed by a construction algorithm. The corresponding computational complexity is briefly discussed in order to quantitatively describe the efficiency of cell communications.

References

[Li 2003] S.-Y. R. Li, R.W. Yeung, N. Cai, Linear network coding, IEEE Transactions on Information Theory, Vol.49, No.2, IEEE, February 2003, pp.371-381.

[Li 2005] S.-Y. R. Li, N. Cai, R. W. Yeung, "On theory of linear network coding," Proceeding of the 2005 IEEE International Symposium on Information Theory (ISIT 2005), Sep. 4-9, 2005, pp. 273-277.

[Li 2009] S.-Y. R. Li, Network coding theory via commutative algebra, 2009, http://www.tims.ntu.edu.tw/ntumath2009/pdf/S.-Y.%20R.%20Li%20.pdf.

[Liu 2007] J.-Q. Liu, On information-theoretical formalization of intracellular communications based on linear network coding, SICE Annual Conference 2007 (Sept.17-20, 2007, Takamatsu, Japan), pp. 909-912.