Theoretical Analysis on A Traffic-Based Routing Algorithm of Mobile Agents

Wenyu Qu[1], Hong Shen[1], and Yingwei Jin[2]
[1]Graduate School of Information Science
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Tatsunokuchi, Ishikawa, 923-1292, Japan
[2] Dalian University of Technology Press
Dalian University of Technology
No 2, Linggong Road, Ganjinzi District, Dalian, 116024, China

Abstract— **In this paper, we propose a mobile agent-based routing algorithm in which the traffic cost is considered. We define a traffic cost function for each link based on known traffic information and find the probability distribution that mobile agents may select a neighboring node and move to. We theoretically analyze the probability distribution and provide the optimal probability distribution that makes inference on the known traffic information and approximates to a unbiased distribution.**

Keywords: Mobile agents, routing, traffic information, unbiased distribution.

I. Introduction

With the rapid growth of the Internet and dramatic advances in computer technology, computers are no longer isolated computational machines. People communicate with the outer world through wireless networks, LANs, and the Internet. The widespread popularity of the WWW (World Wide Web) demands new paradigms for building computer systems. The deployment of mobile agents, which are small decision-making programs capable of migrating autonomously from node to node in a computer network, is an important representative of these new paradigms and is an effective way to reduce network load and latency [14].

In [7], Milojicic described that mobile agents are autonomous, adaptive, reactive, mobile, cooperative, interactive, and delegated software entities. The key idea underlying mobile agents is to bring the computation to the data rather than the data to the computation [18]. The application of mobile agents in network routing has attracted significant attention [6], [17], [21]. Successful examples of mobile agent applications can be found in [12], [13]. The use of mobile agents in applications ranging from electronic commerce to distributed computation has also been studied extensively.

Routing is a key factor for network performance. It is the process of moving a packet of data from source to destination. Once request for sending a packet is received, the router should recommend the optimal path (or the shortest path) for sending this packet over the network. As searching for the optimal path in a stationary network is already a difficult problem, the searching for the optimal path in a dynamical network or mobile network will be much more difficult. Mobile agent-based routing algorithm is a promising option for use in these environments [4]. In a mobile agent-based routing algorithm, a group of mobile agents build paths between pair of nodes, exploring the network concurrently and exchanging data to update routing tables [8], [10]. Once a request for sending a packet is received from a server, a number of mobile agents are generated and dispatched to the network. These agents roam around the network and gather relevant information. Once an agent accomplishes its task, the collected information is sent back to the server. When a certain number of those agents have come back, the server selects the optimal path by certain criterion and sends the packet to the destination along the new path. At the same time, the server updates its routing table by the information of the new path.

It can be seen that in a large communication network such as Internet, agents have to be generated frequently and dispatched to the network. Thus, they will certainly consume a certain amount of bandwidth of each link in the network. If there are too many agents migrating through one or several links at the same time, they will introduce too much transferring overhead to the links. Eventually, these links will be busy and indirectly block the network traffic. Therefore, there is a need of developing routing algorithms that consider about the traffic load. Since the state of different links may change dynamically over time, the agents have to dynamically adapt themselves to the environment, which increases the difficulty for both algorithm design and theoretical analysis. In [3], the network state is monitored by launching an agent at regular intervals from a source to a certain destination. In [5], the agent was enabled to estimate queuing delay without waiting inside data packet queues. In [15], the authors showed that the information needed in [3], [5] for each destination is difficult to obtain in real networks. In [1], a mechanism of handling routing table entries at the neighbors of crashed routers was proposed which significantly improved the algorithm proposed in [3], [5]. In [2], the authors formulated a method of mobile agent planning, which is analogous to the travelling salesman problem [9] to decide the sequence of nodes to be visited by minimizing the total execution time until the desired information is found.

In this paper, we propose an agent-based routing algorithm in which the traffic cost for each link is considered. To balance

Proceedings of the 2005 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'05)

the traffic load on each link, we introduce the maximum entropy theory into our algorithm to find an optimal probability distribution that makes inference on the known traffic information and balances the traffic load. Theoretical analysis shows that our derived probability distribution that an agent on an intermediate node may select a neighboring node and move to satisfies these two requirements. The remainder of this paper is structured as follows. Section II presents our algorithm. Section III introduces the maximum entropy theory. Section IV provides theoretical analysis, and Section V gives our conclusions.

II. The Mobile Agent-Based Routing Algorithm

For a large agent-driven network, suppose that agents can be generated from every node in the network, and each node in the network provides to mobile agents an execution environment. A node from which mobile agents are generated is called the server of these agents. At any time, requests may be keyed in the network. Once a request for sending a packet to a destination (point-to-point) or to multiple destinations (multicast) is received from a server, the server will generate a number of mobile agents. Each agent carries the addresses of its server, its destination, the previous node it jumped from, and some control information for routings such as life-span limit and hop counter. All these data can be contained in several lines of Java code; thus, the size of a mobile agent is very small, resulting in great reduction on network load and latency. After being generated, these agents move out from the server and roam in the network. When an agent reaches a node, it checks whether the host node is its destination. If the current host node of an agent is not the destination, the agent performs a random walk based on the traffic situation of links to the neighboring nodes. That is, it selects a neighboring node according to the costs of the neighboring links and the number of users of each link. A link with lower cost and fewer user will be selected with priority. Once an agent has reached the destination, it will go back to the server along the path searched, update the routing tables on the nodes along the path, and submit its report about the searched path to the server. When a certain number of those agents have come back, the server selects the optimal path by certain criterion and sends the packet to the destination along the new path. At the same time, the server updates its routing table by the information of the new path. To eliminate unnecessary searching in the network, a life-span limit is assigned to each agent. An agent will die if it cannot find its destination in its life-span limit. Moreover, if an agent cannot return to its server in two times the life-span limit (e.g., its return route is interrupted due to a link/node failure), the agent also will die.

Let $G = \{V, E\}$ be a graph corresponding to a fixed network, where $V = \{\nu_1, \nu_2, \cdots\}$ is the set of vertices (hosts) and E is the set of edges. In this paper, we assume that the topology of a network is a connected graph in order to ensure that communication are able to be made between any two host machines. $NB(i)$ is the set of neighboring nodes of node ν_i and $|NB(i)|$ is the number of nodes in $NB(i)$. Originally, each node has no information about its neighboring nodes and links, and no agent passed it on the return trip. Therefore, each vertex in set $NB(i)$ has the same probability to be selected, i.e., $1/|NB(i)|$. This uniform probability distribution of agents' neighboring node-selection will be updated with time going. The new probability distribution should satisfies two constrains:

1) It makes inference on all the known traffic information.
2) It is unbiased. That is, the probability should mostly balance the traffic cost on each link.

To find a probability distribution that both makes inference on the known information and approximates to the unbiased (uniform) distribution, We mathematically model the constrains as follows. The effect of the known traffic information on the agent's migrating decision making can be expressed by a min-max problem as follows:

$$\min_{x \in R^n} f^{(j)}_{\max}(x), \tag{1}$$

where x is a random variable with n entries, which denotes n items to be considered to the cost of a link. The objective function $f^{(j)}_{\max}(x)$ is the traffic cost function defined as follows:

$$f^{(j)}_{\max}(x) \equiv \max_{i \in NB(j)} \{f_{ji}(x)\}. \tag{2}$$

Here, $f_{ji}(x)$ is the traffic cost function from node ν_j to ν_i. Without losing generalization, we assume that functions $f_{ji}(i \in NB(j))$ are differentiable. Obviously, the maximum value function $f^{(j)}_{\max}(x)$ is an undifferentiable function.

At the same time, the unbiased requirement is expressed by the maximum entropy function (as shown in the next section). Solving the combinatorial optimization problem results in a probability distribution that can be expressed as follows:

$$p_{ji} = \frac{\exp\{\theta f_{ji}(x)\}}{\sum_{l \in NB(j)} \exp\{\theta f_{jl}(x)\}},\ j = 1, 2, \cdots; i \in NB(j), \tag{3}$$

where p_{ji} is the probability that an agent on node ν_j migrates to node ν_i, $\theta \geq 0$ is a weight coefficient defined according to the effect of the known traffic state of the network.

In Section III, we will detail the reduction and show the rationality of this probability distribution.

III. The Maximum Entropy Theory

In [19], Shannon first introduced the concept of entropy into informatics as a measurement of uncertainty. Suppose that there are a set of possible events whose probabilities of occurrence are $\lambda_1, \lambda_2, \cdots, \lambda_n$. These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome? Shannon pointed that if there is such a measure, say $H(\lambda_1, \lambda_2, \cdots, \lambda_n)$, it should have the following properties:

1) H should be continuous on λ_i.
2) If all λ_i are equal, i.e., $\lambda_i = 1/n$, then H should be a monotonically increasing function of n.

3) If a choice is broken down into two successive choices, the original H should be the weighted sum of the individual values of H.

In [19], it is proved that the entropy function $H = -k\sum_{i=1}^{n}\lambda_i \ln \lambda_i$ is the only function that can satisfies all the requirements, where k is a positive constant decided by measurement units. Usually, k is set to be 1. The Shannon entropy has the following properties:

1) $H_n(\lambda_1, \lambda_2, \cdots, \lambda_n) \geq 0$;
2) If $\lambda_k = 1$ and $\lambda_i = 0$ ($i = 1, 2, \cdots, n; i \neq k$), then $H_n(\lambda_1, \lambda_2, \cdots, \lambda_n) = 0$;
3) $H_{n+1}(\lambda_1, \lambda_2, \cdots, \lambda_n, \lambda_{n+1} = 0) = H_n(\lambda_1, \lambda_2, \cdots, \lambda_n)$;
4) $H_n(\lambda_1, \lambda_2, \cdots, \lambda_n) \leq H_n(1/n, 1/n, \cdots, 1/n) = \ln n$;
5) $H_n(\lambda_1, \lambda_2, \cdots, \lambda_n)$ is a symmetrical concave function on all variables.

where $H = -\sum_{i=1}^{n}\lambda_i \ln \lambda_i$.

E. T. Jaynes found that in many probabilistic executions, the resulting probability distribution cannot foreknown; thus, the entropy cannot be calculated. But he also claimed that the probability distribution could be induced by the accumulated test data such as the mean and the variance. In [11], E.T.Jaynes proposed the maximum entropy theory: "in making inference on the basis of partial information we must use that probability distribution which has maximum entropy subject to whatever is known. This is the only unbiased assignment we can make; to use any other would amount to arbitrary assumption of information which by hypothesis we do not have". Notice that "entropy" is a measurement of the degree of uncertainty and the great the entropy's value, the less known information, the maximum entropy theory can be mathematically expressed as follows:

$$\begin{cases} \max \quad H = -\sum_{i=1}^{N}\lambda_i \ln \lambda_i \\ s.t. \quad \sum_{i=1}^{N}\lambda_i = 1; \\ \qquad \sum_{i=1}^{N}\lambda_i g_j(x_i) = E\left[g_j\right], j = 1, 2, \cdots, m; \\ \qquad \lambda_i \geq 0, i = 1, 2, \cdots, N, \end{cases} \tag{4}$$

where $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n)$, $g_j(j = 1, 2, \cdots, m)$ is some predefined constrained function, and $E[\cdot]$ is the mean of these constrained function.

Templeman et al. [20] first applied maximum entropy theory to solve optimization problems in which the objective function is unanimously approximated by a smooth one. By solving the resulting problem, an approximate solution of the original problem can be obtained. The purpose of deploying maximum entropy theory in agents' searching process is to find a probability distribution that both satisfies the known routing information and mostly approximate to the unbiased (uniform) distribution.

IV. Theoretical Analysis

Let's look at the following Lagrange function:

$$\ell_j(x, p_j) = \sum_{i \in NB(j)} p_{ji} f_{ji}(x) \ \ \forall x \in R^n, p_j \in \Delta_j, \tag{5}$$

where $p_j = (p_{j1}, p_{j2}, \cdots, p_{j,|NB(j)|})^T$ is the vector of Lagrange multiplier, Δ_j is a simplex set defined as follows:

$$\Delta_j \equiv \left\{ p_j \in R^{|NB(j)|} \,\middle|\, \sum_{i \in NB(j)} p_{ji} = 1, p_{ji} \geq 0 \right\}. \tag{6}$$

It is easy to see that no matter which value the multiplier vector p_j is chosen, the value of the Lagrange function $\ell_j(x, p_j)$ is less than or equal to the maximum value function $f_{\max}^{(j)}(x)$, i.e.,

$$\ell_j(x, p_j) \leq f_{\max}^{(j)}(x). \tag{7}$$

From the definition of Lagrange function $\ell_j(x, p_j)$, we have the following lemma:

Lemma 1: The maximum value function $f_{\max}^{(j)}(x)$, defined in (2), can be expressed as follows:

$$f_{\max}^{(j)}(x) = \sup_{p_j \in \Delta_j} \ell_j(x, p_j) = \max_{p_j \in \Delta_j} \ell_j(x, p_j). \tag{8}$$

Proof: For $\forall x \in R^n$ and $\forall p_j \in \Delta_j$, it is easy to see that

$$\sum_{i \in NB(j)} p_{ji} f_{ji}(x) \leq f_{\max}^{(j)}(x). \tag{9}$$

Therefore,

$$\sup_{p_j \in \Delta_j} \ell_j(x, p_j) \leq f_{\max}^{(j)}(x). \tag{10}$$

Let $I_{\max}^{(j)}(x)$ be the indicator set of element functions $f_{ji}(x)(i \in NB(j))$ that equal to the maximum value function $f_{\max}^{(j)}$ at point x, i.e.,

$$I_{\max}^{(j)}(x) := \{k | f_{jk}(x) = f_{\max}^{(j)}(x)\}. \tag{11}$$

If $k \in I_{\max}^{(j)}(x)$, then for arbitrary $x \in R^n$ and $p_j \in \Delta_j$, we have

$$\sup_{p_j \in \Delta_j} \ell_j(x, p_j) \geq \sum_{i \in NB(j)} \bar{p}_{ji} f_{ji}(x) = f_{\max}^{(j)}(x)b \tag{12}$$

where

$$\bar{p}_{ji} = \begin{cases} 1, & i = k; \\ 0, & i \neq k. \end{cases} \tag{13}$$

From (10) and (12), the first equality in (8) is hold. Consider that Δ_j is a tight set and $\ell_j(x, p_j)$ is a continuous function on p_j, the second equality in (8) is also hold. ■

From Lemma 1, it can be seen that since the Lagrange function $\ell_j(x, p_j)$ is a linear function on variable p_j, (8) has multi-solutions. Therefore, function $f_{\max}^{(j)}(x)$ defined in (2) is an undifferentiable function.

From Lemma 1, it can also be seen see that since the multiplier vector p_j is limited inside the simplex Δ_j, the Lagrange function $\ell_j(x, p_j)$ can be interpreted as a convex combination of all element functions $f_{ji}(x)$ $(i \in NB(j))$

and multipliers p_{ji} are the combination coefficients. Therefore, Problem (1) can be solved by solving an equivalent problem of finding a set of value of $p_{ji}, (i \in NB(j))$ such that the Lagrange function $\ell_j(x, p_j)$ approximates to the maximum value function, i.e., to find the optimal combination $\hat{p}_j$ from all combinations that satisfies (6) such that (7) becomes the following equality:

$$\ell_j(x, \hat{p}_j) = \sum_{i \in NB(j)} \hat{p}_{ji} f_{ji}(x) = f_{\max}^{(j)}(x). \qquad (14)$$

On the other hand, if the Lagrange multipliers p_{ji} $(i \in NB(j))$, also called the combination coefficients, is endued with a probability sense, i.e., describing them as the corresponding probabilities such that the element function $f_{ji}(x)$ becomes the maximum value function $f_{\max}^{(j)}(x)$, then from the concept of probability, Problem (1) can be transferred into a maximized problem of finding the optimal probability distribution that satisfies:

$$\begin{cases} \max\limits_{p_j \in R^{|NB(j)|}} & \ell_j(x, p_j) \\ s.t. & \sum\limits_{i \in NB(j)} p_{ji} = 1; \\ & p_{ji} \geq 0, \quad i \in NB(j). \end{cases} \qquad (15)$$

Now, we begin to find a smooth function to approximate to the maximum value function. According to the analysis above, there are two object functions to be maximized:

1) To maximize the Lagrange function through selecting the optimal multiplier vector;
2) To maximize the entropy function by finding an unbiased probability distribution.

Consider the maximum entropy theory that introduced in the previous section, the optimal probability distribution should also satisfies:

$$\begin{cases} \max\limits_{p_j \in R^{|NB(j)|}} & H(p_j) = -\sum\limits_{i \in NB(j)} p_{ji} \ln p_{ji} \\ s.t. & \sum\limits_{i \in NB(j)} p_{ji} = 1; \\ & p_{ji} \geq 0, \quad i \in NB(j). \end{cases} \qquad (16)$$

Therefore, the problem to be solved is a multi-objective problem as follows:

$$\begin{cases} \max\limits_{p_j \in R^{|NB(j)|}} & \{\ell_j(x, p_j), H(p_j)\} \\ s.t. & \sum\limits_{i \in NB(j)} p_{ji} = 1; \\ & p_{ji} \geq 0, \quad i \in NB(j). \end{cases} \qquad (17)$$

By the weighting coefficient method, the multi-object problem (17) can be transformed into a single-object problem as follows:

$$\begin{cases} \max\limits_{p_j} & L_\theta^{(j)}(x, p_j) = \sum\limits_{i \in NB(j)} p_{ji} f_{ji}(x) \\ & \qquad - \dfrac{1}{\theta} \sum\limits_{i \in NB(j)} p_{ji} \ln p_{ji}; \\ s.t. & \sum\limits_{i \in NB(j)} p_{ji} = 1; \\ & p_{ji} \geq 0, \quad i \in NB(j), \end{cases} \qquad (18)$$

where $\theta \geq 0$ is a weighting coefficient. Obviously, when θ is small, the second item of the object function $L_\theta^{(j)}(x, p_j)$ is dominative. Then the gained probability distribution mainly reflects the requirement of unbiased distribution. With the increase of θ's value, the effect of the first item increases; thus, the object of maximizing the Lagrange function becomes dominative.

To solve Problem (18), we first consider the following problem:

$$\sup_{p_j \in \Delta_j} \left\{ L_\theta^{(j)}(x, p_j) := \ell_j(x, p_j) - \theta^{-1} \sum_{i \in NB(j)} p_{ji} \ln p_{ji} \right\}. \qquad (19)$$

Based on the knowledge of convex analysis and the property of entropy function, we can prove that the function defined by (19) has the following property:

Theorem 1: The function $F_\theta^{(j)}(x)$ defined by (19) is differentiable and uniformly approximate to function $f_{\max}^{(j)}(x)$ on the whole space R^n.

Proof: From the definition of indicator function, Problem (19) can be reduced as

$$\sup_{p_j \in R^{|NB(j)|}} \left\{ \sum_{i \in NB(j)} p_{ji} f_{ji}(x) - \theta^{-1} \sum_{i \in NB(j)} p_{ji} \ln p_{ji} - \delta(p_j | \Delta_j) \right\}, \qquad (20)$$

where $\delta(p_j | \Delta_j)$ is an indicator function on the closed convex set Δ_j. From the strictly convex property of entropy function $\sum_{i \in NB(j)} p_{ji} \ln p_{ji}$, it can be seen that for arbitrary fixed $x \in R^n$, the object function of the maximum problem (20) is a closed normal strictly concave function on variable p_j and the effective region is the tight set Δ_j. Therefore, from the Weiertrass theory, the maximum problem exists an unique solution $p_j^*(x, \theta)$ and reaches its finite optimal value on the unique solution. That is, Problem (19) defines a real value function on R^n as follows:

$$F_\theta^{(j)}(x) := \sum_{p_j \in \Delta_j} L_\theta^{(j)}(x, p_j) = L_\theta^{(j)}\left[x, p_j^*(x, \theta)\right]. \qquad (21)$$

Consider that function $-\sum_{i \in NB(j)} p_{ji} \ln p_{ji}$ is unnegative on the bounded closed convex set Δ_j and has an upper bound $(\ln m)/\theta$, that is

$$1 \leq -\sum_{i \in NB(j)} p_{ji} \ln p_{ji} \leq \frac{\ln m}{\theta}, \forall p_j \in \Delta_j, \qquad (22)$$

we have

$$\sup_{p_j\in\Delta_j} \ell_j(x,p_j) \le \sup_{p_j\in\Delta_j} L_\theta^{(j)}(x,p_j) \le \sup_{p_j\in\Delta_j} \ell_j(x,p_j) + \frac{\ln m}{\theta}. \tag{23}$$

Thus, from (8) and (21), we have

$$f_{\max}^{(j)}(x) \le F_\theta^{(j)}(x) \le f_{\max}^{(j)}(x) + \frac{\ln m}{\theta}. \tag{24}$$

This indicates that the function $F_\theta^{(j)}(x)$, defined by the maximum problem (19), is uniformly approximate to $f_{\max}^{(j)}(x)$ on the whole space R^n.

At the same time, if function $K(p_j)$ is defined as follows:

$$K(p_j) := \begin{cases} \sum_{i\in NB(j)} p_{ji}\ln p_{ji} & p_{ji}\ge 0; \\ +\infty & p_{ji} < 0, \end{cases} \tag{25}$$

then it is a closed normal strictly convex function on $R^{|NB(j)|}$ and $\mathrm{ri}(\mathrm{dom}K)\cap \mathrm{ri}(\Delta_j) \ne \phi$. Therefore, from (20) and the definition of convex conjugate function, we have

$$\begin{aligned} F_\theta^{(j)}(x) &= \theta^{-1}\cdot (K+\delta)^*(\theta F_j(x)) \\ &= \theta^{-1}\cdot (K^*\diamond \delta^*)(\theta F_j(x))), \end{aligned} \tag{26}$$

where $F_j(x) := (f_{ji}(x))^T_{i\in NB(j)}$ is a vector function and K^* is the convex conjugate function of K. Since $K \in Leg(R^m_+)$, i.e., $K(p_j)$ is a Legendre convex function, $K^* \in Leg(\mathrm{int}(\mathrm{dom}K^*))$ and $(K^*\diamond\delta^*)(\cdot)$ is essentially smooth. Thus, from (26) and the property that $F_\theta^{(j)}(x)$ is a real value function on R^n, we have $\mathrm{dom}(K^*\diamond\delta^*) = R^{|NB(j)|}$. Due to the continuous differentiable property of $F_j(x)$, $F_\theta^{(j)}(x)$ is a smooth function.

Combine the above two aspects, the theorem is proven. ■

In the following, we will prove some properties of function $F_\theta^{(j)}(x)$.

Theorem 2: For $\forall x\in R^n$, function $F_\theta^{(j)}(x)$ has the following properties:

1) $f_{\max}^{(j)}(x) \le F_\theta^{(j)}(x) \le f_{\max}^{(j)}(x) + (\ln m)/\theta$.
2) $\lim_{\theta\to\infty} F_\theta^{(j)}(x) = f_{max}^{(j)}(x)$.
3) If all the functions f_{ji} $(i = 1,2,\cdots,m)$ in the original problem (1) are convex, $F_\theta^{(j)}(x)$ is a convex function too.
4) $\nabla_x F_\theta^{(j)}(x) = \sum_{i\in NB(j)} \hat{p}_{ji}(x)\nabla_x f_{ji}(x)$.
5) $-(\ln m)/\theta^2 \le \partial F_\theta^{(j)}(x)/\partial\theta \le 0$.
6) $F_\theta^{(j)}(x) \le F_\vartheta^{(j)}(x)$, $\forall \theta\le\vartheta$.

Proof: 1. Here, we provide a different proof from Theorem 1. From the expression of $F_\theta^{(j)}(x)$ in (39), we have

$$F_\theta^{(j)}(x) = f_{\max}^{(j)}(x) + \frac{1}{\theta}\ln\left\{\sum_{i\in NB(j)} \exp\left[\theta\left(f_{ji}(x) - f_{\max}^{(j)}(x)\right)\right]\right\}. \tag{27}$$

From the definition of maximum value function $f_{\max}^{(j)}(x)$ in (2), we have

$$1 \le \sum_{i\in NB(j)} \exp\left[\theta\left(f_{ji}(x) - f_{\max}^{(j)}(x)\right)\right] \le m. \tag{28}$$

Substitute this inequality into (27), we have

$$f_{\max}^{(j)}(x) + \frac{1}{\theta}\ln 1 \le F_\theta^{(j)}(x) \le f_{\max}^{(j)}(x) + \frac{1}{\theta}\ln m. \tag{29}$$

2. Take limitation on both side of (29), this property is proven.
3. For any $x, y\in R^n$ and $\alpha\in(0,1)$, since all the functions f_{ji} $(i\in NB(j))$ are convex, we have

$$\begin{aligned} & F_\theta^{(j)}(\alpha x + (1-\alpha)y) \\ &= \frac{1}{\theta}\ln\left\{\sum_{i\in NB(j)} \exp\left[\theta f_{ji}\left(\alpha x + (1-\alpha)y\right)\right]\right\} \\ &\le \frac{1}{\theta}\ln\left\{\sum_{i\in NB(j)} \exp\left[\theta\left(\alpha f_{ji}(x) + (1-\alpha)f_{ji}(y)\right)\right]\right\} \\ &= \frac{1}{\theta}\ln\left\{\sum_{i\in NB(j)} \left(\exp\left[\theta f_{ji}(x)\right]\right)^\alpha \left(\exp\left[\theta f_{ji}(y)\right]\right)^{1-\alpha}\right\}. \end{aligned}$$

Applying to the Hölder inequality, we have

$$\begin{aligned} & \sum_{i\in NB(j)} \left(\exp\left[\theta f_{ji}(x)\right]\right)^\alpha \left(\exp\left[\theta f_{ji}(y)\right]\right)^{1-\alpha} \\ &\le \left\{\sum_{i\in NB(j)} \exp\left[\theta f_{ji}(x)\right]\right\}^\alpha \cdot \left\{\sum_{i\in NB(j)} \exp\left[\theta f_{ji}(y)\right]\right\}^{1-\alpha}. \end{aligned}$$

Combined the above two relationships, we have

$$\begin{aligned} & F_\theta^{(j)}\left(\alpha x + (1-\alpha)y\right) \\ &\le \frac{\alpha}{\theta}\ln\left\{\sum_{i\in NB(j)} \exp\left[\theta f_{ji}(x)\right]\right\} \\ &\quad + \frac{1-\alpha}{\theta}\ln\left\{\sum_{i\in NB(i)} \exp\left[\theta f_{ji}(y)\right]\right\} \\ &= \alpha F_\theta^{(j)}(x) + (1-\alpha)F_\theta^{(j)}(y). \end{aligned} \tag{30}$$

Hence, function $F_\theta^{(j)}(x)$ is a convex function.
4. Take derivation about x on both side of (27), we have

$$\nabla_x F_\theta^{(j)}(x) = \nabla_x f_{\max}^{(j)}(x) + \frac{1}{\theta}\nabla_x \ln\left\{\sum_{i\in NB(j)} \exp\left[\theta\left(f_{ji}(x) - f_{\max}^{(j)}(x)\right)\right]\right\}. \tag{31}$$

Since

$$\begin{aligned}
&\nabla_x \ln \left\{ \sum_{i \in NB(j)} \exp\left[\theta\left(f_{ji}(x) - f_{\max}^{(j)}(x)\right)\right] \right\} \\
&= \frac{\sum_{i \in NB(j)} \nabla_x \exp\left[\theta\left(f_{ji}(x) - f_{\max}^{(j)}(x)\right)\right]}{\sum_{i \in NB(j)} \exp\left[\theta\left(f_{ji}(x) - f_{\max}^{(j)}(x)\right)\right]} \\
&= \sum_{i \in NB(j)} \frac{\exp\left[\theta\left(f_{ji}(x) - f_{\max}^{(j)}(x)\right)\right]}{\sum_{i \in NB(j)} \exp\left[\theta\left(f_{ji}(x) - f_{\max}^{(j)}(x)\right)\right]} \\
&\quad \cdot \nabla_x \left[\theta\left(f_{ji}(x) - f_{\max}^{(j)}(x)\right)\right] \\
&= \theta \sum_{i \in NB(j)} \hat{p}_{ji}(x) \nabla_x f_{ji}(x) - \nabla_x f_{\max}^{(j)}(x),
\end{aligned} \tag{32}$$

where $\hat{p}_{ji}x$ is defined as (38). Substitute this results into (31), this property is proven.

5. From the expression of function $F_\theta^{(j)}(x)$ in (39) and the definition of $\hat{p}_{ji}(x)$ in (38), we have

$$\begin{aligned}
&\frac{\partial F_\theta^{(j)}(x)}{\partial \theta} \\
&\quad = -\frac{F_\theta^{(j)}(x)}{\theta} + \frac{\sum_{i \in NB(j)} \exp\left(\theta f_{ji}(x)\right) f_{ji}(x)}{\theta \sum_{i \in NB(j)} \exp\left[\theta f_{ji}(x)\right]} \\
&\quad = \theta^{-1} \left[\sum_{i \in NB(j)} \hat{p}_{ji}(x) f_{ji}(x) - F_\theta^{(j)}(x) \right].
\end{aligned} \tag{33}$$

According to (18) and (39), we have

$$\begin{aligned}
F_\theta^{(j)}(x) = &\sum_{i \in NB(j)} p_{ji}(x) f_{ji}(x) \\
&-\theta^{-1} \sum_{i \in NB(j)} p_{ji}(x) \ln\left[p_{ji}(x)\right].
\end{aligned} \tag{34}$$

Therefore,

$$\partial F_\theta^{(j)}(x)/\partial\theta = \theta^{-2} \sum_{i \in NB(j)} p_{ji}(x) \ln\left[p_{ji}(x)\right]. \tag{35}$$

Thus, from the following inequality:

$$-\frac{\ln m}{\theta} \le \theta^{-2} \sum_{i=1}^{m} p_{ji}(x) \ln\left[p_{ji}(x)\right] \le 0, \quad \forall x \in R^n, \tag{36}$$

this property is proven.

6. From property 5, we can see that function $F_\theta^{(j)}(x)$ is a decreasing function on θ, thus, this property is straight forward for property 5. ■

Item 1 in Theorem 2 provides error bounds of function $F_\theta^{(j)}(x)$, and item 2 shows that function $F_\theta^{(j)}(x)$ uniformly approximates to function $f_{\max}^{(j)}(x)$. Item 3 shows the convex property of function $F_\theta^{(j)}(x)$, and item 4 is for the continuity and the differentiability of function $F_\theta^{(j)}(x)$. Item 5 provides both upper bound and lower bound of the derivation of function $F_\theta^{(j)}(x)$ on p, and item 6 shows that function $F_\theta^{(j)}(x)$ is a monotonously decrease function on θ.

Since function $F_\theta^{(j)}(x)$ uniformly converges to the objective function $f_{\max}^{(j)}(x)$, solving the original problem (1) is equivalent to solving the following problem:

$$\min_{x \in R^n} F_\theta^{(j)}(x). \tag{37}$$

As function $F_\theta^{(j)}(x)$ is differentiable, the optimal solution, $\hat{p}_j(x)$, of Ea. (18) can be easily derived from applying the $K-T$ condition as follows:

$$\hat{p}_{ji}(x) = \frac{\exp\{\theta f_{ji}(x)\}}{\sum_{l \in NB(j)} \exp\{\theta f_{jl}(x)\}}, \quad i \in NB(j). \tag{38}$$

Substitute the analytical solution $\hat{p}_j(x)$ of the multiplier p_j in the objective function of (18), we have

$$\begin{aligned}
F_\theta^{(j)}(x) &= L_\theta^{(j)}\left(x, \hat{p}_j(x)\right) \\
&= \frac{1}{\theta} \ln \left\{ \sum_{i \in NB(j)} \exp\left[\theta f_{ji}(x)\right] \right\}.
\end{aligned} \tag{39}$$

V. Conclusion

Routing is a key factor for network routing, and mobile agent-based routing is newly proposed for use in large dynamic network. In this paper, we proposed a new mobile agent-based routing algorithm in which the balance of traffic cost is considered. On each node, there is a probability distribution for an agent to select one of the neighboring nodes and move to. We macroscopically define a traffic cost function for each link according to the known traffic information and found a probability distribution that not only makes inference on the known information but also balances the traffic costs. Theoretical analysis provided some properties of the approximating function (including the convergence property) and showed the rationality of our probability distribution.

Acknowledgements

This research is conducted as a program for the “21st Century COE Program” by Ministry of Education, Culture, Sports, Science and Technology.

References

[1] B. Baran and R. Sosa. “A new approach for antnet routing,” *Proc. of the 9th Int. Conf. on Computer, Communications and Networks*, 2000.

[2] B. Brewington, R. Gray, K. Moizumi, D. Kotz, G. Cybenko, and D. Rus. “Mobile Agents in Distributed Information Retrieval,” *Intelligent Information Agents: Agents-Based Information Discovery and Management on the Internet*, M. Klusch, ed., Springer-Verlag, Berlin, chapter 15, pp. 355-395, 1999.

[3] G. D. Caro and M. Dorigo. “AntNet: Distributed Stigmergetic Control for Communications Networks,” *Journal of Artificial Intelligence Research*, Vol. 9, pp. 317-365, 1998.

[4] G. D. Caro and M. Dorigo. “Mobile Agents for Adaptive Routing,” *Proc. of the 31st Hawaii Int. Conf. on System Sciences*, pp. 74-83, 1998.

[5] G. D. Caro and M. Dorigo. “Two Ant Colony Algorithms for Best-Effort Routing in Datagram Networks,” *Proc. of the 10th IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS’98)*, pp. 541-546, 1998.

[6] J. Claessens, B. Preneel, and J. Vandewalle. “(How) Can Mobile Agents Do Secure Electronic Transactions on Untrusted Hosts? A Survey of the Security Issues and the Current Solutions,” *ACM Trans. on Internet Technology*, Vol. 3, No. 1, pp. 28-48, 2003.

[7] D. Milojicic. "Guest Editor's Introduction: Agent Systems and Applications," *IEEE Concurrency*, Vol.8, No. 2, pp. 22-23, 2000.
[8] M. Dorigo, V. Maniezzo, A. Colorni, "The Ant System: Optimization by a colony of cooperating agents," *IEEE Trans. on Systems, Man, and Cybernetics - Part B*, Vol. 26, No. 1, pp. 29-41, 1996.
[9] M. Garey and D. Johnson. "Computers and Intractability: A Guide to the Theorey of NP-Completeness," Freeman, 1979.
[10] P. S. Heck and S. Ghosh. "A Study of Synthetic Creativity through Behavior Modeling and Simulation of an Ant Colony," *IEEE Intelligent Systems*, Vol. 15, No. 6, pp. 58-66, 2000.
[11] E. T. Joynes. "Information theory and statistical mechanics," *The Physical Review*. No. 108, pp. 171-190, 1957.
[12] G. Karjoth, D. Lange, and M. Oshima. "A Security Model for Aglets," *IEEE Internet Computing*, Vol. 1, No. 4, pp. 68-77, 1997.
[13] D. Lange and M. Oshima. "Programming and Developing Java Mobile Agents with Aglets," *Addison Wesley*, 1998.
[14] D. Lange and M. Oshima. "Seven Good Reasons for Mobile Agents," *Communications of the ACM*, Vol. 42, pp. 88-89, 1999.
[15] S. Liang, A. Zincir-Heywood, and M. Heywood. "The effect of routing under local information using a social insect metaphor," *Proc. of the IEEE Congress on Evolutionary Computing*, 2002.
[16] D. Milojicic. "Trend Wars: Mobile Agent Applications," *IEEE Concurrency*, 7(3), pp. 80-90, Jul.-Sept. 1999.
5 (SRDS02), Oct. 2002.
[17] S. Pleisch and A. Schiper. "Fault-Tolerant Mobile Agent Execution," *IEEE Trans. on Computers*, Vol. 52, No. 2, pp. 209-222, 2003.
[18] D. Schoder and T. Eymann. "The Real Challenges of Mobile Agents," *Commun. of ACM*, Vol. 43, No. 6, pp. 111-112, Jun. 2000.
[19] C. E. Shannon. "A Mathematical Theory of Communication," *Bell System Technical Journal*, Vol. 27, No. 3, pp. 379-428, 1948.
[20] A. B. Templeman and X. Li. "A maximum entropy approach to constrained non-linear programming," *Engineering Optimization*, Vol. 12, No. 3, pp. 191-205, 1987.
[21] D. Xu, J. Yin, Y. Deng, and J. Ding. "A Formal Architectural Model for Logical Agent Mobility," *IEEE Trans. on Software Engineering*, Vol. 29, No. 1, pp. 31-45, 2003.