Computers & Industrial Engineering 56 (2009) 1414–1418

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Optimal routing policy of a stochastic-flow network

Yi-Kuei Lin *

Department of Industrial Management, National Taiwan University of Science and Technology, No. 43, Section 4, Keelung Road, Taipei 106, Taiwan, ROC

ARTICLE INFO

Article history:
Received 22 February 2008
Received in revised form 2 September 2008
Accepted 3 September 2008
Available online 11 September 2008

Keywords:
Routing policy
Optimal
Stochastic-flow network
Quickest path
System reliability

ABSTRACT

Under the assumption that each arc's capacity of the network is deterministic, the quickest path problem is to find a path sending a given amount of data from the source to the sink such that the transmission time is minimized. However, in many real-life networks such as computer systems, telecommunication systems, etc., the capacity of each arc is stochastic due to failure, maintenance, etc. Such a network is named a stochastic-flow network. Hence, the minimum transmission time is not fixed. We try to evaluate the probability that d units of data can be sent through the stochastic-flow network within the time constraint according to a routing policy. Such a probability is named the system reliability, which is a performance index to measure the system quality. This paper mainly finds the optimal routing policy with highest system reliability. The solution procedure is presented to calculate the system reliability with respect to a routing policy. An efficient algorithm is subsequently proposed to derive the optimal routing policy.

1. Introduction

In operations research, computer science, networking, and other areas, the shortest path problem is one of the well-known and practical problems. When goods or commodities are transmitted through a flow network, it is desirable to adopt the shortest path, least cost path, largest capacity path, shortest delay path, or some combination of multiple criteria (Ahuja, 1988; Bodin, Golden, Assad, & Ball, 1982; Fredman & Tarjan, 1987; Golden & Magnanti, 1977), which are all variants of the shortest path problem. From the point of view of quality management and decision making, it is an important task to reduce the transmission time through the network, especially through computer and telecommunication networks. Hence, a version of the shortest path problem called the quickest path problem is proposed by Chen and Chin (1990). This problem is to find a path (named the quickest path) sending a given amount of data with minimum transmission time, where each arc has two attributes; the capacity and the lead time (Chen & Chin, 1990; Hung & Chen, 1992; Martins & Santos, 1997; Park, Lee, & Park, 2004). More specifically, the capacity and the lead time of each arc are both assumed to be deterministic. Since then, several variants of quickest path problems are proposed; constrained quickest path problem (Chen & Hung, 1994; Chen & Tang, 1998), the first k quickest paths problem (Chen, 1993, 1994; Clímaco, Pascoal, Craveirinha, & Captivo, 2007), and all-pairs quickest path problem (Chen & Hung, 1993; Lee & Papadopoulou, 1993).

However, the capacity of each arc is stochastic in many real flow networks due to failure, maintenance, etc. That is, each arc has several possible capacities or states. Such a network is named a stochastic-flow network (Jane, Lin, & Yuan, 1993; Lin, 2003, 2004, 2007a, 2007b; Lin, Jane, & Yuan, 1995; Xue, 1985; Yeh, 2002, 2007, 2008). For instance, a computer system with each arc denoting the transmission medium and each node denoting station of servers is a typical stochastic-flow network. In fact, each transmission medium is combined with several physical lines (coaxial cables, fiber optics, etc.), and each physical line has only success or failure state. That implies a transmission medium has several states in which state k means k physical lines are successful. Hence, the minimum transmission time is not fixed for a stochastic-flow network. We need to solve the problem that how to transmit a specified amount of data through a stochastic-flow network within the time constraint.

The data are transmitted through some specified MPs (minimal paths) which are decided in advance by the system administrator, where a MP is a path without loops. The administrator plans the routing policy, which specifies the first priority MP and the second priority MP. The second priority MP is responsible for the transmission duty if the first priority MP fails. The purpose of this paper is to evaluate the probability that the stochastic-flow network can send d units of data within time T according to a routing policy. Such a probability is named the system reliability, which is a performance index to measure the system quality. The remainder of this paper is organized as follows. In Section 2, the probability that a specified MP sends d units of data within the time constraint can be calculated. The routing policy with level 2 and the system reliability

* Tel.: +886 2 27303277; fax: +886 2 27376344.
E-mail address: yklin@mail.ntust.edu.tw

0360-8352/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cie.2008.09.002

according to the routing policy are both discussed in Section 3. We present the algorithm to determine the optimal routing policy with highest system reliability in Section 4. Computational complexity analysis of the algorithms is shown in Section 5. The extension to evaluate the system reliability according to the routing policy with level 3 is discussed in Section 6.

1.1. Notations and assumptions

$G \equiv (N, A, L, M)$ denotes a stochastic-flow network with a source s and a sink t where N denotes the set of nodes, $\{A \equiv ai | 1 \geq i \geq n\}$ denotes the set of arcs, $L \equiv (l_1, l_2, \ldots, l_n)$ with l_i denoting the lead time of a_i, and $M = (M_1, M_2, \ldots, M_n)$ with M_i denoting the maximal capacity of a_i. The capacity is the maximal number of data sent through the medium (an arc or a path) per unit of time. The (current) capacity of arc a_i, denoted by x_i, takes possible values $0 = b_{i1} < b_{i2} < \cdots < b_{ir_i} = M_i$, where b_{ij} is an integer for $j = 1, 2, \ldots, r_i$. The vector $X \equiv (x_1, x_2, \ldots, xn)$ is called the capacity vector of G. Such a G is assumed to further satisfy the following assumptions:

1. Each node is perfectly reliable.
2. All data are sent through one MP.
3. The capacity of each arc is stochastic with a given probability distribution.
4. The capacities of different arcs are statistically independent.

Vectors operations are made according to the following rules:

$$Y \geq X(y_1, y_2, \ldots, y_n) \geq (x_1, x_2, \ldots, x_n) : y_i \geq x_i \text{ for each } i = 1, 2, \ldots, n$$
$$Y \geq X(y_1, y_2, \ldots, y_n) > (x_1, x_2, \ldots, x_n) : Y \geq X \text{ and } y_i > x_i \text{ for at least one } i$$

2. Model formulation

Suppose there are m MPs: $P_1, P_2, \ldots, P_m$. With respect to each MP $P_j = \{a_{j1}, a_{j2}, \ldots, a_{jn_j}\}, j = 1, 2, \ldots, m$, the capacity of P_j under the capacity vector X is $\min_{1 \leq k \leq n_j}(x_{jk})$. The transmission time to send d units of data through P_j under the capacity vector X, denoted by $\psi(d, X, P_j)$, is

$$\text{lead time of } P_j + \left\lceil \frac{d}{\text{the capacity of } P_j} \right\rceil = \sum_{k=1}^{n_j} l_{jk} + \left\lceil \frac{d}{\min_{1 \leq k \leq n_j} x_{jk}} \right\rceil, \quad (1)$$

where $\lceil x \rceil$ is the smallest integer such that $\lceil x \rceil \geq x$. It contradicts the time constraint if $\psi(d, X, P_j) > T$.

Any capacity vector X with $\psi(d, X, P_j) \leqslant T$ means that X can send d units of data through P_j within time constraint T. Let Ω_j be the set of such X and $\Omega_{j,\min} = \{X | X \text{ is minimal in } \Omega_j\}$. Then $X \in \Omega_{j,\min}$ is called (d, T, P_j)-QP, equivalently, X is the (d, T, P_j)-QP if and only if (i) $\psi(d, X, P_j) \leqslant T$ and (ii) $\psi(d, Y, P_j) > T$ for any capacity vector Y with $Y < X$. Hence, we have the following lemma.

Lemma 1. *If X is the (d, T, P_j)-QP, then $\psi(d, X, P_j) \leqslant T$ for any $Y \geqslant X$.*

Proof. If $X \leqslant Y$, then $x_{jk} \leqslant y_{jk}$ for each $a_{jk} \in P_j$, and $\min_{1 \leq k \leq n_j} x_{jk} \leq \min_{1 \leq k \leq n_j} y_{jk}$. Thus,

$$\left\lceil \frac{d}{\min_{1 \leq k \leq n_j} x_{jk}} \right\rceil \geq \left\lceil \frac{d}{\min_{1 \leq k \leq n_j} y_{jk}} \right\rceil. \text{ So } T \geq \psi(d, X, P_j) \geq \psi(d, Y, P_j).$$

□

For a specified MP P_j, the (d, T, P_j)-QP can be generated as the following algorithm.

Algorithm 1.

Step 1. Find the minimal capacity v_j of P_j such that d units of data can be sent through P_j within T. That is, find the smallest integer v_j such that

$$\sum_{k=1}^{n_j} l_{jk} + \left\lceil \frac{d}{v_j} \right\rceil \leq T \quad (2)$$

Step 2. [Generate the (d, T, P_j)-QP]

If $v_j \leq \min_{1 \leq k \leq n_j}(M_{jk})$, then the minimal capacity vector $Z_j = (z_1, z_2, \ldots, z_n)$ such that the network sends d units of data under T is obtained according to

$$\begin{cases} z_i = \text{the minimal capacity } u \text{ } of \ a_i \text{ such that } u \geq v_j & \text{if } a_i \in P_j, \\ z_i = 0 & \text{if } a_i \notin P_j. \end{cases} \quad (3)$$

Otherwise, Z_j does not exist.

To guarantee that the (d, T, P_j)-QP is generated from Algorithm 1, the following lemma is necessary.

Lemma 2. *The (d, T, P_j)-QP is the Z_j generated from Algorithm 1.*

Proof. We claim that the obtained Z_j from the proposed algorithm is the (d, T, P_j)-QP. (i) It is trivial that $\psi(d, Z_j, P_j) \leqslant T$. (ii) If $X < Z_j$ where $X = (x_1, x_2, \ldots, x_n)$, then there exists an arc $a_u \in P_j$ such that $x_u < z_u$. The capacity of P_j under X is thus less than that under Z_j since the setting from Step 2. Hence, $\psi(d, X, P_j) > T$. By (i) and (ii), we conclude that Z_j is the (d, T, P_j)-QP. □

3. Routing policy

The network administrator decides the routing policy in advance to indicate the first priority MP, the second priority MP (or named alternative MP), etc. The alternative MP will be responsible for the transmission duty if the first priority MP fails. The level of routing policy is called 2 if an alternative MP is standing by. Similarly, the level is 3 if two alternative MPs are standing by. That is, the second priority MP takes charge if the first priority MP fails, and the third priority MP takes charge if the second priority MP fails. A MP fails if and only if at least one arc in it fails. We first focus on the reliability evaluation for routing policy with level 2. Without loss of generality, let P_1 and P_2 be the first and the second priority MPs, respectively. Let E_j denote the event that P_j fails, and S_j denote the event that P_j can send d units of data within time T, $j = 1, 2$. Then

$$\Pr(E_j) = \Pr(x_i = 0 \text{ for at least one } a_i \in P_j) = 1 - \prod_{i:a_i \in P_j} \Pr(x_i \geq 1), \quad j = 1, 2. \quad (4)$$

Lemma 2 indicates that $\Omega_{j,\min} = \{Z_j\}$ where Z_j is generated from Algorithm 1. Thus

$$\Pr(S_j) = \Pr\{\Omega_j\} = \Pr\{X | X \geqslant Zj\} = \prod_{i:a_i \in P_j} \Pr(x_i \geqslant z_i), \quad j = 1, 2. \quad (5)$$

The system reliability $R_{d,T}$ is the probability that the network can send d units of data within time T. Under the routing policy with level 2, the system reliability $R_{d,T}$ is

$$R_{d,T} = \Pr(S_1) + \Pr(S_2 | E_1) \times \Pr(E_1) = \Pr(S_1) + \Pr(S_2) \times Pr(E_1) = \prod_{i:a_i \in P_1} \Pr(x_i \geqslant z_i) + \prod_{i:a_i \in P_2} \Pr(x_i \geqslant z_i) \times \left(1 - \prod_{i:a_i \in P_1} \Pr(x_i \geqslant 1)\right). \quad (6)$$