Structure of the Submarking-Reachability Problem and Network Programming

M. A. COMEAU, MEMBER, IEEE, AND K. THULASIRAMAN, SENIOR MEMBER, IEEE

Abstract —Using a linear programming formulation, a unified treatment of the submarking-reachability problem for both capacitated and uncapacitated marked graphs is presented. In both cases the problem reduces to that of testing feasibility of the dual transshipment problem of operations research. An algorithm called REACH is presented for the feasibility testing problem and its worst-case time complexity is O(mn), where m and n are, respectively, the number of edges and the number of nodes in the marked graph. The place of this work in the context of general network programming problems is highlighted.

I. Introduction

A PETRI NET is a general abstract algebraic structure originally developed by Carl Adam Petri as a model for information flow in systems exhibiting asynchronism and parallelism [1]. Petri net modeling has applications in computer communication, operating systems, operations research, artificial intelligence as well as physiological models of the brain. The generality of the Petri net makes modeling of complex systems possible. However, the feasibility of analysis becomes questionable and in many cases the problems are *NP*-complete, with solutions sometimes undecidable. As a result, several restricted classes of the Petri net have been introduced and studied. These include the computation graph [2], the marked graph [3], [4], and the state graph.

Our study in this paper is concerned with marked graphs. A number of papers have appeared in the literature exploring several problems related to marked graphs [3]–[10]. In [3], Commoner *et al.* have presented, among other things, an algorithmic approach to the reachability problem and have also studied the maximum-marking problem. In [4], Murata has presented a circuit-theoretic approach for the study of the reachability problem. Recently, Kumagai, Kodama, and Kitagawa [5], [6] have introduced and studied the submarking-reachability problem. Related to the problem of reachability is that of controllability which has been considered in [7] and [8]. The maximum-weight marking problem is discussed in [9]. Certain structural properties of Petri nets are presented in [10].

In this paper we study the structure of the submarking-reachability problem and its relationship to a problem in operations research. Whereas, in the reachability problem, a final marking is specified on each edge of the given marked graph G, in the submarking-reachability problem, a final marking is specified only on a subset of the edges of G and no marking is specified for the remaining edges. The submarking-reachability problem concerns determining whether or not there exists a marking with the specified final token distribution in the reachable space of a given initial marking of G. In their pioneering work, Kumagai, Kodama, and Kitagawa [5] have provided an approach for the study of the submarking-reachability problem and have described an approach for constructing a marking with the specified final submarking that is reachable from an initial marking, whenever it exists. However, their study does not fully expose the structure of the problem. As a result, extension of their approach to the study of the submarking-reachability problem for the capacitated case has not been easy [6].

Our paper is organized as follows. First, we formulate the submarking-reachability problem as a linear program and demonstrate how to solve this problem after reducing it to an equivalent smaller problem by relaxing the feasibility constraints and by introducing an intermediate state (Sections III–VI). We then define (Section VII) the submarking-reachability problem for the capacitated case and show that a similar reduction and solution technique applies to this case too, thereby unifying the study of the capacitated and uncapacitated cases. In Section VIII, we point out the link between the submarking-reachability problem and the problem of testing feasibility of the dual transshipment problem and present an algorithmic solution (algorithm REACH) to the problem. The place of algorithm REACH in the general context of combinatorial optimization problems is highlighted.

In the following section we summarize the main definitions and results relating to marked graphs.

II. Preliminaries

A *marked graph* is a directed graph $G=(V,E)$ with vertex set V, edge set E, a nonnegative integer column vector M associated with E, called a *marking*, *state* or *token distribution* of G, and a state-transition function $\delta_i(M)$ mapping M into a new marking M' resulting from firing vertex $i\in V$. The transition function subtracts one

Manuscript received October 6, 1986. The work presented in this paper is based on part of the doctoral thesis of M.A. Comeau submitted to the Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada, in June 1986. This paper was recommended by Associate Editor A. N. Venetsanopoulos.

M. A. Comeau is with the Centre de Recherche Informatique de Montreal, Montreal, Canada.

K. Thulasiraman is with the Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada.

IEEE Log Number 8717750.

0098-4094/88/0100-0089$01.00

token from M on each edge incident into i and adds one token to M on each edge incident out of i, to obtain M'. Since M' must be nonnegative, $\delta_i(M)$ can be applied only if M has a positive token count on each edge incident into vertex i. In other words, to be *legally fired*, a vertex must have at least one token on each of its input edges. A vertex is said to be *enabled* under a marking M if it is legally firable under M. The *enabling number* of vertex i is the minimum of the token counts on the edges incident into i.

A marking M' is *reachable* from a marking M if some sequence of legal firings will transform M into M'. The *reachability set* $R(M)$ of a marking M is defined as the set of all markings reachable from M. Since the null sequence is trivially legal, $M \in R(M)$.

A marked graph G is *live* under a marking M if each vertex $i \in G$ can be enabled through some legal firing sequence starting from M. The marking M is then called a *live marking*. Liveness is characterized in the following theorem [3] where the term *dead subgraph* refers to a token-free directed circuit.

Theorem 1: A marked graph G is live under a marking M if and only if G contains no dead subgraphs under M. ■

Let G be a marked graph with an initial marking M_0 and let $M \in R(M_0)$. Then, the *differential* marking $\Delta_M \triangleq M - M_0$ satisfies Kirchhoff's voltage law in G [4]. Thus if B_f is a fundamental-circuit matrix of G, then

$$B_f \Delta_M = 0. \tag{1}$$

In view of (1), we can consider the elements of Δ_M as voltages of the corresponding edges of G. Using well-known network-theoretic results, we can determine a set of *node voltages* $\{\sigma_1, \sigma_2, \cdots, \sigma_n\}$ of G such that $\sigma_i - \sigma_j = \Delta_M(e)$ where $e = (i, j) \in E$ is the edge directed from vertex i to vertex j and $\Delta_M(e)$ is the component of Δ_M corresponding to edge e. Let $\Sigma \triangleq [\sigma_1, \sigma_2, \cdots, \sigma_n]'$ denote the column vector of σ_i's. Note that

$$A'\Sigma = \Delta_M \tag{2}$$

where A is the incidence matrix of G. If the smallest entry in Σ is not a zero, then we can obtain such a vector Σ_0 by adding or subtracting an appropriate number to all entries in Σ. It is easy to show that Σ_0 also satisfies (2) and it is unique for a given value of Δ_M. The vector Σ_0 is called the *minimum nonnegative firing-count* vector and the σ_i's are called *firing numbers* or *firing counts*. The ith element of Σ_0 indicates the minimum number of times vertex i would fire in a firing sequence leading from M_0 to M. The vertex with zero firing count will be referred to as a *datum*.

A nonnegative firing-count vector Σ is said to be *executable* from M_0 if a legal firing sequence exists starting from M_0 and its firing-count vector is Σ. Note that for any Δ_M satisfying (1), existence of Σ satisfying (2) is guaranteed. However, this alone does not guarantee executability of Σ. Executability of Σ_0, or, equivalently, reachability of M from M_0, is characterized in the following theorem [4].

Theorem 2 (Reachability Theorem): Let M_0 and M_r be two markings of a marked graph G. Let $\Delta_M = M_r - M_0$. M_r is reachable from M_0 if and only if

i) $B_f \Delta_M = 0$, and
ii) σ_k is zero for each vertex belonging to a dead subgraph of G, where

$$\Sigma_0 = [\sigma_1, \sigma_2, \cdots, \sigma_n]' \geqslant 0$$

is the minimum nonnegative solution of

$$A'\Sigma = \Delta_m.$$ ■

A *capacitated marked graph* is a marked graph $G = (V, E)$ in which a lowerbound $L(e)$ and an upperbound $U(e)$ are specified on the token count $M(e)$ of each edge $e \in E$, for all markings of G. A marking M of G is called *feasible* if and only if $L(e) \leqslant M(e) \leqslant U(e)$, $\forall e \in E$. We state this feasibility in vector notation as $L \leqslant M \leqslant U$. This definition reduces to the original definition when $L(e) = 0$ and $U(e) = \infty$, $\forall e \in E$.

The *enabling number* of a vertex $v \in V$ under a marking M of a capacitated marked graph G is defined as

$$\mu_v = \min\left\{ \min_{e \in E_v^-} \{M(e) - L(e)\},\ \min_{e \in E_v^+} \{U(e) - M(e)\} \right\} \tag{3}$$

where E_v^- and E_v^+ are the input and output edge sets of vertex v, respectively. This reduces to the usual definition for uncapacitated graphs when $L(e) = 0$ and $U(e) = \infty$, $\forall e \in E$. A vertex v is *enabled* or *legally firable* under a marking M is its enabling number under M is greater than zero.

Let $C \subseteq E$ be a circuit of G and define an arbitrary orientation for C. Let C_+ and C_- denote the subsets of C consisting of all edges following and opposing the orientation, respectively. A *dead subgraph* of a capacitated marked graph G under a marking M is either

i) an edge $e \in E$ with $L(e) = M(e) = U(e)$, or
ii) a circuit $C = C_+ \cup C_- \subseteq E$ such that either

$$M(e) = L(e), \forall e \in C_+ \quad \text{and } M(e) = U(e), \forall e \in C_-.$$

or

$$M(e) = U(e), \forall e \in C_+ \quad \text{and } M(e) = L(e), \forall e \in C_-.$$

A capacitated marked graph is *live* under a marking M if each vertex $i \in G$ can be enabled through some legal firing sequence starting at M. Liveness of a capacitated marked graph is characterized in the following theorem.

Theorem 3: A capacitated marked graph G is live under a marking M if and only if G has no dead subgraphs under M. ■

Necessary and sufficient conditions for reachability of capacitated marked graphs are given in the following theorem.

Theorem 4 (Capacitated-Reachability Theorem): Let M_0 and M_r be two feasible markings of a marked graph G. Let

$\Delta_M = M_r - M_0$. M_r is reachable from M_0 if and only if

i) $B_f \Delta_M = 0$, and

ii) σ_k is zero for each vertex belonging to a dead subgraph of G, where

$$\Sigma_0 = [\sigma_1, \sigma_2, \cdots, \sigma_n]' \geqslant 0$$

is the minimum nonnegative solution of

$$A'\Sigma = \Delta_M. \quad \blacksquare$$

Finally, the concept of enabling of a node can be extended to that of a subgraph. This leads to the notion of a *diakoptic* transition in a marked graph $G = (V, E)$ as explained below.

Let S and $\bar{S} = V - S$ be a partition of V and let $\langle S, \bar{S} \rangle$ denote the *cut* $\langle S, \bar{S} \rangle_+ \cup \langle S, \bar{S} \rangle_-$ consisting of the forward cut edges $\langle S, \bar{S} \rangle_+$ directed from S to $\bar{S}$ and the backward cut edges $\langle S, \bar{S} \rangle_-$ directed from $\bar{S}$ to S. Let $G(S)$ be the subgraph induced on S by removing $\langle S, \bar{S} \rangle$ from G. Similarly, let $G(\bar{S})$ be the subgraph induced on $\bar{S}$ by removing $\langle S, \bar{S} \rangle$ from G. If $G(S)$ and $G(\bar{S})$ are both connected, then $\langle S, \bar{S} \rangle$ is called a *cutset* of G. Let us assume $\langle S, \bar{S} \rangle$ is an arbitrary cut of G.

The enabling number of $G(S)$ is defined as

$$\mu(G(S)) \triangleq \min\left\{ \min_{e \in \langle S, \bar{S} \rangle_-} \{M(e) - L(e)\}, \min_{e \in \langle S, \bar{S} \rangle_+} \{U(e) - M(e)\} \right\} \quad (4)$$

If the graph G is uncapacitated then $\mu(G(S))$ reduces to

$$\mu(G(S)) \triangleq \min_{e \in \langle S, \bar{S} \rangle_-} \{M(e)\}. \quad (5)$$

An *elementary diakoptic firing* of a vertex-induced subgraph $G(\cdot)$ of a marked graph G is any legal firing sequence confined to vertices in $G(\cdot)$ which fires each vertex in $G(\cdot)$ exactly once. Note that this definition includes subgraphs $G(\cdot)$ consisting of disjoint components.

Theorem 5 (Diakoptic-Transition Theorem): An elementary diakoptic firing of a vertex-induced subgraph $G(S)$ of a marked graph G is legal under a live marking M if and only if $\mu(G(S)) > 0$. $\blacksquare$

Proof of the above theorem may be found in [9].

III. Formulation of the Submarking-Reachability Problem

We are given a marked graph $G = (V, E)$ with an initial marking, M_0. The edge set is partitioned into the *controlled set* E_C for which a final marking $M_l(E_C)$ is specified and the *free set* E_F for which a final marking $M_l(E_F)$ is not specified. The partitioning of E into E_C and E_F partitions V into V_C and V_F, where V_C is the set of all vertices of G incident to an edge of E_C and $V_F = V - V_C$. The vertices in V_C are called the *controlled vertices* and those in V_F are called the *free vertices*. The incidence matrix A of G partitions as follows:

$$A = \begin{array}{c} \\ \left[\begin{array}{cc} A_{CC} & A_{CF} \\ 0 & A_{FF} \end{array}\right] \end{array} \begin{array}{l} \\ V_C \\ V_F \end{array}, \quad \text{columns } E_C \;\; E_F.$$

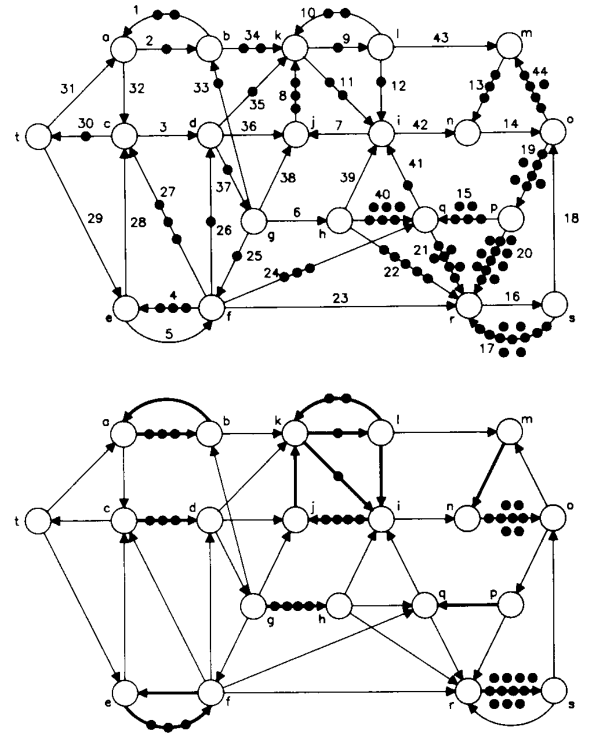

Fig. 1. a) A marked graph with initial marking M_0. b) The final submarking.

As an example, Fig. 1(a) illustrates a marked graph G with an initial marking M_0. Here, edges are numbered $1, 2, 3, \cdots, 44$. The final submarking is shown in Fig. 1(b). The controlled edges are drawn as heavy lines to distinguish them from the remaining free edges. The controlled vertex set $V_C = \{a, b, \cdots, s\}$ and the free vertex set $V_F = \{t\}$.

Let $\Sigma = [\Sigma_C', \Sigma_F']'$ be a firing-count vector associated with V, partitioned according to V_C and V_F, respectively. Every executable Σ on G from the initial marking M_0 results in a marking M given by the state equation $M = M_0 + A'\Sigma$. Partitioning the state equation accordingly gives

$$M(E_C) = M_0(E_C) + A_{CC}'\Sigma_C \quad (6)$$

$$M(E_F) = M_0(E_F) + A_{CF}'\Sigma_C + A_{FF}'\Sigma_F. \quad (7)$$

The *submarking-reachability* problem is to decide whether or not $M_l(E_C)$ is reachable from M_0 and if so, determine the minimum firing-count vector Σ leading from M_0 to some marking M for which $M(E_C) = M_l(E_C)$ and $M(E_F)$ is feasible; i.e., $M(E_F) \geqslant 0$. Clearly, the minimum firing-count vector realizing such an M from M_0 must satisfy the dead-subgraph condition of the reachability theorem (Theorem 2).

Since the dead-subgraph condition must be satisfied and can be checked easily after M is known, then we may simply neglect this restriction for now by considering live problems only. At first, this seems as if it could be computationally difficult if there exist multiple markings M satisfying (6) and (7). However, by further specifying that

M should be the *nearest* such marking to M_0, in the sense that the minimum firing-count vector realizing M from M_0 is minimum among all minimum firing-count vectors realizing such a marking from M_0, we render the solution to the submarking-reachability problem unique. If the minimum firing-count vector realizing the nearest M, whenever it exists, does not satisfy the dead-subgraph condition of the reachability theorem, then, nor will any other. Thus we may state the submarking-reachability problem for live marked graphs as the linear program

$$\begin{aligned} &\text{minimize } \Sigma \\ &\text{subject to} \\ &A_{CC}^t \Sigma_C = M_l(E_C) - M_0(E_C) \\ &A_{CF}^t \Sigma_C + A_{FF}^t \Sigma_F \geqslant -M_0(E_F) \\ &\Sigma \geqslant 0 \end{aligned} \tag{8}$$

in which the notation "minimize Σ" means "minimize $\Sigma_{i=1}^n \sigma_i$", henceforth.

IV. Structure of the Problem

Let us investigate the structure of (8). The edge induced subgraph $G_C = G(E_C) = (V_C, E_C)$ may consist of a number of maximally connected subgraphs or *components* and consequently, the equality constraints in (8) decompose into disjoint subsystems–one for each component of G_C. Let G_C consist of r components $G_C^1 = (V_C^1, E_C^1)$, $G_C^2 = (V_C^2, E_C^2), \cdots, G_C^r = (V_C^r, E_C^r)$. Then, the incidence matrix of G has the following structure:

$$A = \begin{array}{c} \\ \end{array}\left[\begin{array}{cccccc|c} E_C^1 & E_C^2 & \cdots\ E_C^k & \cdots & E_C^r & E_F & \\ A_{CC}^1 & & & & & & V_C^1 \\ & A_{CC}^2 & & 0 & & & V_C^2 \\ & & \ddots & & & A_{CF} & \vdots \\ 0 & & A_{CC}^k & & & & V_C^k \\ & & & \ddots & & & \vdots \\ & & & & A_{CC}^r & & V_C^r \\ \hline & & 0 & & & A_{FF} & V_F \end{array}\right] \tag{9}$$

and hence, the equality constraints in (8) decompose to

$$\left(A_{CC}^k\right)^t \Sigma_C^k = M_l\left(E_C^k\right) - M_0\left(E_C^k\right), \qquad k = 1, 2, \cdots, r \tag{10}$$

where A_{CC}^k is the incidence matrix of $G_C^k = (V_C^k, E_C^k)$ and Σ_C^k is the firing-count vector associated with V_C^k.

To illustrate this structure, Fig. 2 shows the controlled subgraph $G_C = (V_C, E_C)$ of the graph $G = (V, E)$ shown in Fig. 1, obtained by simply removing the free edges and the free vertices from G. Each of the isolated portions of G_C is a maximally connected subgraph or component of G_C.

The solution to each subsystem in (10) is unique to within an additive constant, if it exists. Each one of these subsystems corresponds to an independent reachability problem on the corresponding controlled component. Clearly, if $M_l(E_C^k)$ is not independently reachable from $M_0(E_C^k)$ on any component $G_C^k = (V_C^k, E_C^k)$ of $G_C = (V_C, E_C)$, then $M_l(E_C)$ is not reachable from M_0 on G. Thus in order to solve the submarking-reachability problem, we must first solve a number of reachability subproblems.

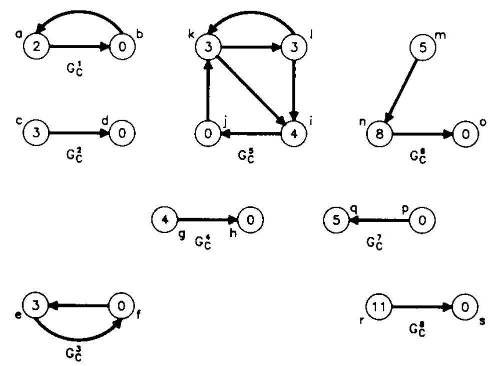

Fig. 2. The controlled subgraph.

Assuming that (10) has a solution, let $\hat{\Sigma}_C^k$ denote the minimum nonnegative solution for the kth subsystem. Clearly, $\hat{\Sigma}_C^k$ leads from $M_0(E_C^k)$ to $M_l(E_C^k)$. Each $\hat{\Sigma}_C^k$ contains at least one zero entry and every solution to the kth subsystem in (10) can be expressed as

$$\Sigma_C^k = \hat{\Sigma}_C^k + [\gamma_k, \gamma_k, \cdots, \gamma_k]_{(1 \times |V_C^k|)}^t \tag{11}$$

for some scalar constant γ_k. As an example, we have given in Fig. 2 (the controlled subgraph) the components of $\hat{\Sigma}$ within the corresponding vertices. In order to obtain a uniform notation, let $w = |V_F|$ and consider each free vertex $i \in V_F$ as a trivial controlled component of G with no controlled edges. Thus the minimum firing-count of vertex i is $\hat{\Sigma}_F^i = \hat{\sigma}_i = 0$ and, obviously, any firing count of vertex i can be expressed as $\Sigma_F^i = \hat{\Sigma}_F^i + \gamma_i$. We now have a γ_i for every controlled component of G, including the trivial components. Let $\hat{\Sigma}_C = [(\hat{\Sigma}_C^1)^t, (\hat{\Sigma}_C^2)^t, \cdots, (\hat{\Sigma}_C^r)^t]^t$ denote the vector of the $\hat{\Sigma}_C^k$ vectors, let $\hat{\Sigma}_F = [\hat{\Sigma}_F^1, \hat{\Sigma}_F^2, \cdots, \hat{\Sigma}_F^w]^t = 0$ and let $\hat{\Sigma} = [\hat{\Sigma}_C^t, \hat{\Sigma}_F^t]^t$. Then, any solution to (8) can be expressed as

$$\Sigma = \hat{\Sigma} + \kappa\Gamma \tag{12}$$

where κ is a binary matrix of size $(n \times (r + w))$, whose (i, j)th entry is 1 if vertex i of G is in component j of G and 0, otherwise. The column vector

$$\Gamma = [\gamma_1, \gamma_2, \cdots, \gamma_{r+w}]^t \tag{13}$$

has an entry γ_j for each component j of G. We now proceed to pose (8) in terms of Γ, with $\hat{\Sigma}$ fixed.

V. Reduction of the Problem

Having obtained $\hat{\Sigma}$, let us take G to the marking $\tilde{M}$ reachable from M_0, through a possibly illegal firing sequence executing $\hat{\Sigma}$. Thus

$$\tilde{M} = M_0 + A^t\hat{\Sigma}. \tag{14}$$

Clearly, $\tilde{M}(E_C) = M_l(E_C)$. If $\tilde{M}(E_F) \geqslant 0$, then $\hat{\Sigma}$ is the unique solution to (8); otherwise, $\tilde{M}$ is an infeasible state. Suppose that $\tilde{M}$ is infeasible. Then, our approach is to pull G out of $\tilde{M}$ and move it to a feasible state M of G through additional firings or determine that this is not possible. Assuming such a feasible state M is reachable, then, we can write M as

$$M = M_0 + A^t(\hat{\Sigma} + \kappa\Gamma)$$
$$= \tilde{M} + A^t\kappa\Gamma. \qquad (15)$$

Since we require that $M(E_C) = M_l(E_C)$ it follows that $M(E_C) = \tilde{M}(E_C)$. So, we get from (15),

$$\left(A_{CC}^k\right)^t\Gamma_k = 0, \qquad k = 1, 2, \cdots, r \qquad (16)$$

where

$$\Gamma_k = [\gamma_k, \gamma_k, \cdots, \gamma_k]^t. \qquad (17)$$

But, (16) is satisfied for any arbitrary value of γ_k. Thus in any firing sequence leading from $\tilde{M}$ to the feasible state M, all the vertices in each component G_C^k will be fired an equal number of times, namely γ_k times. So, we may consider such a firing sequence as a sequence of *diakoptic firings*. During this firing process, the markings on the edges in G_C^k are not altered and only those on the remaining edges may change. So, we need to focus our attention on the edges connecting different components including the trivial components. This suggests that we direct our attention only to the contracted graph $\tilde{G}$ obtained by short-circuiting all the vertices within each component and removing all the edges of E_C.

The graph $\tilde{G}$ may have free-edge self-loops. If any of these self-loops has a negative marking under $\tilde{M}$ then (8) is infeasible since the state of a self-loop can never change under a diakoptic firing. This simple observation implies that we may remove all self-loops from $\tilde{G}$ and proceed from there, provided all self-loops are feasibly marked under $\tilde{M}$. Thus assume $\tilde{G}$ is free of self-loops. Also, $\tilde{G}$ may contain parallel edges. Let E_{ij} denote the set of all edges directed from vertex $i \in \tilde{G}$ to vertex $j \in \tilde{G}$. For each edge $e \in E_{ij}$, we have $M(e) = \tilde{M}(e) + \gamma_i - \gamma_j$. Thus the feasibility condition is

$$\gamma_i - \gamma_j \geqslant -\tilde{M}(e), \qquad \forall e \in E_{ij}. \qquad (18)$$

Clearly, feasibility is satisfied for all $e \in E_{ij}$ if and only if

$$\gamma_i - \gamma_j \geqslant \max_{e \in E_{ij}} \{-\tilde{M}(e)\}. \qquad (19)$$

Therefore, for any i and j, $i \neq j$, we may remove all the parallel edges, except the minimally marked one, that is, the one for which the right-hand side of (19) is obtained.

At this point $\tilde{G}$ has no self-loops nor parallel edges. If $\tilde{G}$ is disconnected then the problem breaks into smaller subproblems. Hence, we assume $\tilde{G}$ is connected. With the contracted graph established, let us relax the notation to help simplify our presentation. That is, from this point on, let $\tilde{A}$ denote the incidence matrix of $\tilde{G}$ and let M and $\tilde{M}$ mean the same as before, but for edges of $\tilde{G}$ only. Then (8) reduces to

$$\text{minimize } \Gamma$$
$$\text{subject to } \tilde{A}^t\Gamma \geqslant -\tilde{M}, \qquad (20)$$
$$\Gamma \geqslant 0.$$

We wish to emphasize the fact that in the above formulation of the submarking-reachability problem, we do not distinguish between the firings of the controlled components, namely, the G_C^k's, and the firings of the free vertices. Also, in any firing sequence leading from $\tilde{M}$ to the feasible state M, firing a vertex of $\tilde{G}$ corresponds to a diakoptic firing of the corresponding component in G.

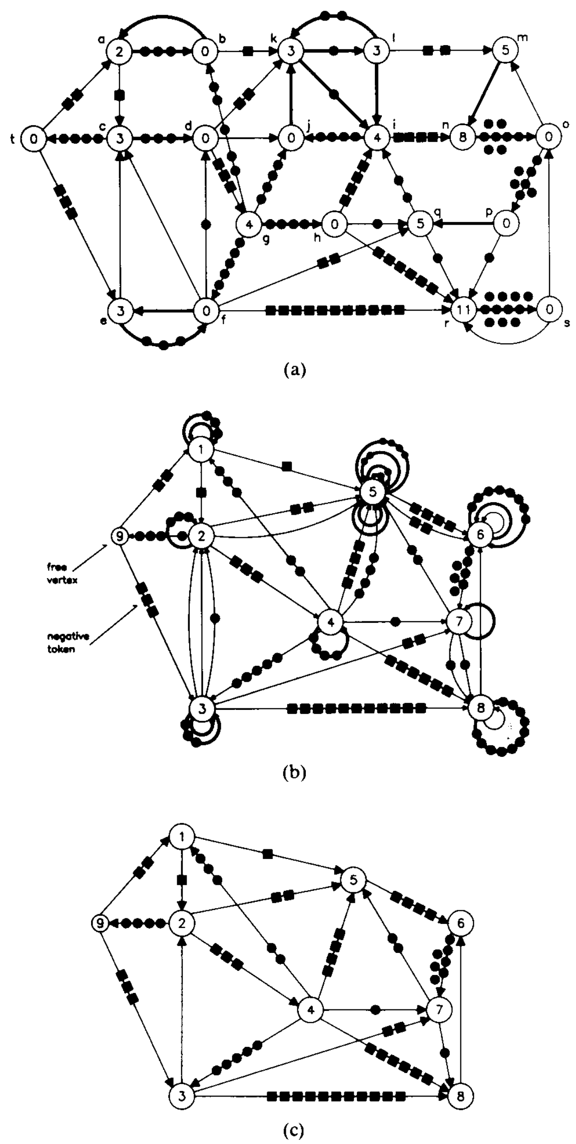

Fig. 3. a) The graph in state $\tilde{M}$. b) Graph after contraction. c) The contracted graph $\tilde{G}$.

Returning to our example, Fig. 3 illustrates the contraction of the marked graph. Fig. 3(a) shows the marked graph of Fig. 1 in state $\tilde{M}$. The number within each vertex is the firing count of that vertex in the minimum solution of (10). Fig. 3(b) shows the graph which results after short-circuiting the vertices within each component. All self-loops are feasibly marked and Fig. 3(c) shows the contracted graph $\tilde{G}$ with the reduced intermediate state $\tilde{M}$, obtained by removing all self-loops and all but the minimally marked parallel edges.

VI. A Solution to the Submarking-Reachability Problem

As stated previously, if $\tilde{M} \geqslant 0$, then the solution to (20) is $\Gamma = 0$. Thus we are interested in the case where $\tilde{M} \ngeqslant 0$. With this in mind, we proceed as follows. Let $\tilde{G} = (\tilde{V}, \tilde{E})$. Let p_{ij} denote a directed path leading from vertex $i \in \tilde{V}$ to vertex $j \in \tilde{V}$. Also, let p_{ii} denote a directed circuit through vertex $i \in \tilde{V}$. Clearly, if we sum the inequality constraints in (20) along any directed circuit

$$p_{ii} = \{(i, u_1), (u_1, u_2), \cdots, (u_n, i)\} \in \tilde{G}$$

we obtain

$$0 = \gamma_i - \gamma_{u_1} + \gamma_{u_1} - \gamma_{u_2} + \cdots + \gamma_{u_{n-1}} - \gamma_{u_n} + \gamma_{u_n} - \gamma_i \geqslant - \sum_{e \in p_{ii}} \tilde{M}(e).$$

Thus a necessary condition for the existence of a solution to (20) is simply

$$\sum_{e \in C} \tilde{M}(e) \geqslant 0, \qquad \forall \text{ directed circuits } C \text{ in } \tilde{G}. \quad (21)$$

In other words, for the existence of a solution to (20) it is necessary that every directed circuit in $\tilde{G}$ has a nonnegative token count.

We now establish a simple lowerbound on the residual firing count γ_i of the contracted component G_C^i. If we sum the constraints in (20) along a directed path $p_{ij} \triangleq \{(i, u_1), (u_1, u_2), \cdots, (u_{n-1}, u_n), (u_n, j)\} \in \tilde{G}$ leading from vertex $i \in \tilde{V}$ to vertex $j \in \tilde{V}$ through the intermediate vertices $\{u_1, u_2, \cdots, u_n\}$, then we obtain

$$\gamma_i - \gamma_{u_1} + \gamma_{u_1} - \gamma_{u_2} + \cdots + \gamma_{u_{n-1}} - \gamma_{u_n} + \gamma_{u_n} - \gamma_j \geqslant - \sum_{e \in p_{ij}} \tilde{M}(e)$$

or simply,

$$\gamma_i - \gamma_j \geqslant - \sum_{e \in p_{ij}} \tilde{M}(e) \quad (22)$$

and hence, $\gamma_i \geqslant \gamma_j - \sum_{e \in p_{ij}} \tilde{M}(e)$. Imposing the nonnegativity requirement on Γ, namely $\gamma_j \geqslant 0$, we obtain the following.

Lemma 1: $\gamma_i \geqslant \max\{0, -\sum_{e \in p_{ij}} \tilde{M}(e)\}, \quad \forall p_{ij} \in \tilde{G}$.

Using this lower bound on the residual firing numbers, we establish the solution to (20), whenever it exists, in the following theorem.

Theorem 6 (Submarking-Reachability Theorem): Let the token count of every directed circuit in $\tilde{G}$ be nonnegative under $\tilde{M}$. If d_{ij} denotes the shortest distance from vertex $i \in \tilde{G}$ to vertex $j \in \tilde{G}$ under $\tilde{M}(\tilde{E})$, then the unique solution to (20) is given by

$$\gamma_i \triangleq \max\left\{0, - \min_j \{d_{ij}\}\right\}, \qquad \forall i \in \tilde{V}. \quad (23)$$

Proof: Let p_{ij} denote a shortest-path from vertex i to vertex j in $\tilde{G}$, under $\tilde{M}(\tilde{E})$, with distance $d_{ij} = \sum_{e \in p_{ij}} \tilde{M}(e)$. We must prove feasibility and optimality of the solution (23). Note that the assumption that $\tilde{G}$ under $\tilde{M}$ has no directed circuit of negative token count guarantees that all the shortest distances exist and they satisfy the triangle inequality. Thus $d_{ij} + d_{jk} \geqslant d_{ik}$ for distinct i, j, and k.

Feasibility: We prove the feasibility of (23) by showing that such an assignment, whenever it exists, results in a feasible final marking of $\tilde{G}$. We prove this by a series of contradictions. Assume that the assignments indicated in (23) do not correspond to a feasible final marking of $\tilde{G}$. Then, there exists at least one edge $e = (i, j) \in \tilde{E}$ such that

$$\gamma_i - \gamma_j < - \tilde{M}(e). \quad (24)$$

Now, consider the assignments of γ_i and γ_j as in (23). There are four cases to consider.

Case 1: $\min_k\{d_{ik}\} \geqslant 0$, $\min_k\{d_{jk}\} \geqslant 0$.

In this case, $\gamma_i = \gamma_j = 0$. Since $d_{ik} \geqslant 0$, $\forall k \in \tilde{V}$, it follows that $d_{ij} \geqslant 0$. However, with $\gamma_i = \gamma_j = 0$, assumption (24) implies $\tilde{M}(e) < 0$, contradicting that $d_{ij} \geqslant 0$.

Case 2: $\min_k\{d_{ik}\} \geqslant 0$, $\min_k\{d_{jk}\} < 0$.

In this case, $\gamma_i = 0$. Let the minimization in (23) occur on vertex $v \in \tilde{V}$ for the assignment of γ_j. That is, vertex v is a closest one to vertex j in $\tilde{G}$ under the intermediate marking $\tilde{M}$. By hypothesis, $d_{ik} \geqslant 0$, $\forall k \in \tilde{V}$ and specifically, $d_{iv} \geqslant 0$. The triangle inequality property of the shortest distances requires that $d_{iv} \leqslant \tilde{M}(e) + d_{jv}$. Combining this with the requirement $d_{iv} \geqslant 0$ implies that $\tilde{M}(e) + d_{jv} \geqslant 0$ or $d_{jv} \geqslant - \tilde{M}(e)$. Now, the assumption (24) establishes the contradiction by requiring that $d_{jv} < - \tilde{M}(e)$.

Case 3: $\min_k\{d_{ik}\} < 0$, $\min_k\{d_{jk}\} \geqslant 0$.

Proof in this case follows as in Case 2.

Case 4: $\min_k\{d_{ik}\} < 0$, $\min_k\{d_{jk}\} < 0$.

Let the minimization in (23) occur on vertex $u \in \tilde{V}$ and vertex $v \in \tilde{V}$ for the assignments of γ_i and γ_j, respectively, where possibly $u = v$. Then vertex u is a closest one to vertex i and vertex v is a closest one to vertex j. Thus $d_{iu} \leqslant d_{ik}$, $\forall k \in \tilde{V}$ and, specifically, $d_{iu} \leqslant d_{iv}$. The triangle inequality property requires that $d_{iv} \leqslant \tilde{M}(e) + d_{jv}$. Hence, $d_{iu} \leqslant \tilde{M}(e) + d_{jv}$. Finally, assumption (24) implies $d_{iu} > \tilde{M}(e) + d_{jv}$, establishing the contradiction.

Thus we have established the feasibility of the solution (23) to (20).

Optimality: With the feasibility of (23) established, optimality of (23) follows from Lemma 1. ■

The existence conditions follow easily as a corollary of Theorem 6.

Corollary 6.1: The solution of (20) defined in Theorem 6 exists if and only if $\sum_{e \in C} \tilde{M}(e) \geqslant 0$ for all directed circuits of $\tilde{G}$.

Proof: The existence of solution (23) is predicated on the existence of the shortest distances which, in turn, exist if and only if no negative-length directed circuit is present in $\tilde{G}$ under $\tilde{M}$. ■

As in Section II, a vertex of $\tilde{G}$ is called a *datum* vertex of $\tilde{G}$ if it need not be fired in reaching the nearest feasible marking of $\tilde{G}$, whenever it exists. A datum vertex is characterized in the following Corollary of Theorem 6.

Corollary 6.2: A vertex $i \in \tilde{V}$ is a datum vertex of $\tilde{G}$ if and only if $\sum_{e \in p_{ij}} \tilde{M}(e) \geqslant 0$ for all directed paths p_{ij} from vertex i in $\tilde{G}$.

Proof: From Theorem 6, $\gamma_i = 0$ if and only if $d_{ij} \geqslant 0$, $\forall j \in \tilde{V}$ which is clearly equivalent to the condition

$$\sum_{e \in p_{ij}} \tilde{M}(e) \geqslant 0, \qquad \forall p_{ij} \in \tilde{G}, \qquad \forall j \in \tilde{V}. \qquad \blacksquare$$

Finally, we have the following.

Corollary 6.3: If the solution to (20) exists, then $\tilde{G}$ has at least one datum under $\tilde{M}$.

Proof: Consider any vertex i. Let u be a vertex closest to i so that $d_{iu} \leqslant d_{iv}$, for $v \in \tilde{V}$. We claim vertex u is a datum. If not, then $d_{uj} < 0$ for some $j \in \tilde{V}$. Since $d_{iu} \leqslant d_{ij}$, we get, using the triangle inequality, $d_{iu} \leqslant d_{ij} \leqslant d_{iu} + d_{uj}$. So, $d_{uj} \geqslant 0$, contradicting the assumption $d_{uj} < 0$. $\blacksquare$

We can now construct the solution to the submarking-reachability problem by combining the component solutions with the solution to the reduced problem. The minimum firing-count vector realizing the overall solution is defined as

$$\Sigma = \hat{\Sigma} + \kappa \Gamma \tag{25}$$

where $\hat{\Sigma}$ is the minimum nonnegative solution of (10) and Γ is the solution to (20) as given in Theorem 6. The corresponding final marking of G is simply

$$M = M_0 + A'\Sigma. \tag{26}$$

Returning to our example, we have shown, again, the contracted graph in Fig. 4. Here the weights on the edges refer to the token counts. To calculate γ_i's, we first obtain the shortest distances between all pairs of vertices using Floyd's Algorithm [11]. These distances are given below in matrix form

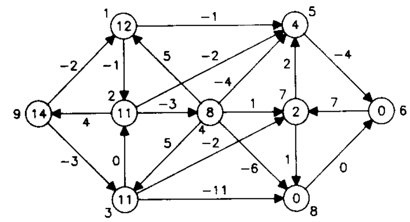

Fig. 4. The contracted graph $\tilde{G}$ redrawn.

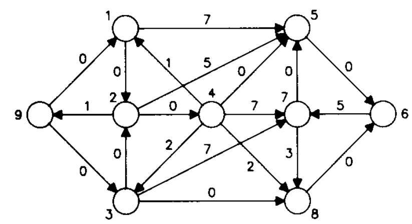

Fig. 5. A feasible marking of the contracted graph.

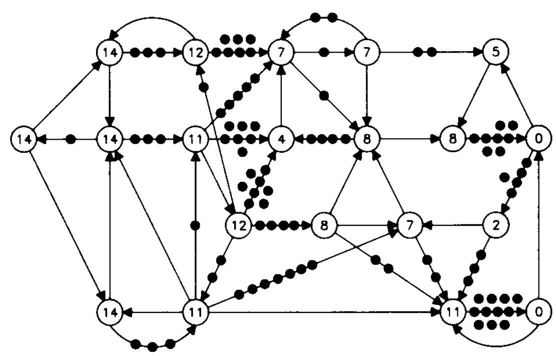

Fig. 6. A feasible marking of the original graph.

$$[d_{ij}] = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{array} \begin{array}{c} \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \begin{bmatrix} 0 & -1 & 0 & -4 & -8 & -12 & -5 & -11 & 3 \\ 2 & 0 & 1 & -3 & -7 & -11 & -4 & -10 & 4 \\ 2 & 0 & 0 & -3 & -4 & -11 & -4 & -11 & 4 \\ 5 & 4 & 5 & 0 & -4 & -8 & -1 & -6 & 8 \\ \infty & \infty & \infty & \infty & 0 & -4 & 3 & 4 & \infty \\ \infty & \infty & \infty & \infty & 9 & 0 & 7 & 8 & \infty \\ \infty & \infty & \infty & \infty & 2 & -2 & 0 & 1 & \infty \\ \infty & \infty & \infty & \infty & 9 & 0 & 7 & 0 & \infty \\ -2 & -3 & -3 & -6 & -10 & -14 & -7 & -14 & 0 \end{bmatrix} \end{array}.$$

Using this matrix we get γ_i's as follows:

$$\gamma_1 = 12, \quad \gamma_2 = 11, \quad \gamma_3 = 11$$
$$\gamma_4 = 8, \quad \gamma_5 = 4, \quad \gamma_6 = 0$$
$$\gamma_7 = 2, \quad \gamma_8 = 0, \quad \gamma_9 = 14.$$

In Fig. 4 we have indicated γ_i's within the circles representing the vertices. The marking shown in Fig. 5 is a feasible marking of the contracted graph reachable from the marking shown in Fig. 4, and it is obtained using the γ_i's shown in Fig. 4. Using the marking in Fig. 5, the specified final marking on the controlled edges, and the intermediate marking of the free-edge self-loops and computing the final marking on the parallel edges that were removed, we obtain the feasible marking in Fig. 6. The firing numbers realizing this marking are indicated within circles in this figure.

VII. The Capacitated Submarking-Reachability Problem

Recall that a capacitated marked graph is a marked graph $G=(V,E)$ with an integer lowerbound $L(e)$ and an integer upperbound $U(e)$ specified on the token count $M(e)$ of each edge $e \in E$. A feasible marking M of G is one satisfying $L(e) \leqslant M(e) \leqslant U(e)$, $\forall e \in E$ or simply, $L \leqslant M \leqslant U$ in column-vector format. We assume that $L < U$. The capacitated submarking-reachability problem is defined in a fashion similar to the uncapacitated version. The graph G is marked with a feasible initial marking M_0 and the edge set E is partitioned into controlled edges E_C and free edges E_F induced by specifying a final feasible submarking $M_l(E_C)$ for G. Thus the problem partitions as described in Section IV.

As before, we will consider live marked graphs first and note that the arguments expressed in Section III are also valid in the context of the capacitated submarking-reachability problem. Using (6) and (7) and the feasibility requirement, the capacitated submarking-reachability problem is equivalent to the linear program

$$\begin{aligned} &\text{minimize } \Sigma \\ &\quad \text{subject to} \\ &\quad A_{CC}^t \Sigma_C = M_l(E_C) - M_0(E_C) \\ &\quad L(E_F) - M_0(E_F) \leqslant A_{CF}^t \Sigma_C + A_{FF}^t \Sigma_F \\ &\qquad\qquad \leqslant U(E_F) - M_0(E_F) \\ &\quad \Sigma \geqslant 0 \end{aligned} \tag{27}$$

where $L(E_F)$ and $U(E_F)$ denote the lower and upper bound vectors associated with $M(E_F)$.

7.1. Structure, Decomposition, and Reduction of the Problem

The topological partitioning of the marked graph is exactly as described in Section IV. The incidence matrix A of G has the form shown in (9). The decomposition into subproblems is exactly as described in Section IV. In order to solve the capacitated submarking-reachability problem, we must first solve a number of capacitated reachability subproblems.

Once the solution to the capacitated reachability subproblems has been obtained, reduction of the capacitated-submarking-reachability problem follows exactly as described in Section V, where we need only clarify how to handle parallel edges which may form when the controlled components of G are contracted.

In the notation of Section IV, let $\hat{\Sigma}$ denote the minimum feasible solution, when it exists, to the capacitated reachability subproblems and let $\tilde{M} \triangleq M_0 + A^t\hat{\Sigma}$ be the corresponding intermediate state. If $L \leqslant \tilde{M} \leqslant U$ then $\hat{\Sigma}$ is the unique solution to (27). Otherwise, we simply have a state $\tilde{M}$ reachable from M_0 which violates the feasibility condition. Let $\tilde{G}=(\tilde{V},\tilde{E})$ denote the contracted graph which results after contracting the controlled components of G and removing the self-loops which form in this contraction. Note that if any free self-loop is not feasibly marked, then (27) is infeasible. Let E_{ij} denote the set of all edges directed from vertex $i \in \tilde{V}$ to vertex $j \in \tilde{V}$. The feasibility condition is

$$L(e) - \tilde{M}(e) \leqslant \gamma_i - \gamma_j \leqslant U(e) - \tilde{M}(e), \qquad \forall e \in E_{ij}$$

which is clearly covered by the single constraint

$$\max_{e \in E_{ij}} \{L(e) - \tilde{M}(e)\} \leqslant \gamma_i - \gamma_j \leqslant \min_{e \in E_{ij}} \{U(e) - \tilde{M}(e)\}. \tag{28}$$

At this point, a significant difference arises between the capacitated and uncapacitated problems. Although the constraint (28) reduces to the constraint (19) when $L(e) = 0$ and $U(e) = \infty$, $\forall e \in E$ and it is always consistent as defined, the covering constraint (28) may not be consistent in general. Specifically, when some edges are finitely upperbounded, there is no guarantee that

$$\max_{e \in E_{ij}} \{L(e) - \tilde{M}(e)\} \leqslant \min_{e \in E_{ij}} \{U(e) - \tilde{M}(e)\} \tag{29}$$

since the maximization and the minimization in (27) will, in general, occur on distinct edges. Clearly, if condition (29) is not satisfied for any set of parallel edges $E_{ij} \subseteq \tilde{E}$, then $M_l(E_C)$ is not reachable from M_0 on G and hence, when reducing the capacitated submarking-reachability problem, we must further test the consistency condition (29) for all parallel edge sets E_{ij} which form when the controlled components of G are contracted.

Assuming (29) is satisfied for all parallel edge sets $E_{ij} \subseteq \tilde{E}$, we must resolve an ambiguity which may arise in removing parallel edges. Clearly, if both the minimization and the maximization in (28) occur on the same edge $e \in E_{ij}$ then edge e represents a most constraining edge of E_{ij} since its associated inequality is equivalent to condition (28) and thus all edges of E_{ij} but edge e may be removed from $\tilde{G}$. The ambiguity arises when the maximization occurs on, say, edge $e_L \in E_{ij}$ and the minimization occurs on some other edge, say, $e_U \in E_{ij}$. Clearly, all edges of E_{ij} but e_L and e_U may be removed from G. The problem is that neither edge e_L nor edge e_U alone suffices to represent inequality (28). The obvious next step is to replace the pair of edges $\{e_L, e_U\}$ with a single *artificial edge*, say, e_A which does represent (28). If $\tilde{M}(e_L) = \tilde{M}(e_U)$, then we may simply define $L(e_A) \triangleq L(e_L)$, $U(e_A) \triangleq U(e_U)$ and $\tilde{M}(e_A) = \tilde{M}(e_L) = \tilde{M}(e_U)$ and replace E_{ij} with the artificial edge e_A. The apparent difficulty with this approach arises when $\tilde{M}(e_L) \neq \tilde{M}(e_U)$, which is the general case. We simply need an artificial edge e_A which will represent inequality (28) and this can be achieved by defining the parameters and intermediate state of edge e_A as

$$\begin{aligned} L(e_A) &\triangleq \max_{e \in E_{ij}} \{L(e) - \tilde{M}(e)\} \\ U(e_A) &\triangleq \min_{e \in E_{ij}} \{U(e) - \tilde{M}(e)\} \\ \tilde{M}(e_A) &\triangleq 0. \end{aligned} \tag{30}$$

Clearly, if all edges in E_{ij} are replaced with an artificial edge e_A whose parameters and state under $\tilde{M}$ are as defined in (30), then the inequality associated with edge e_A is exactly the constraint (28). Hence, parallel edges present no real difficulty in extending our solution to the capacitated problem. The important point is that the existence of a solution to the capacitated submarking-reachability problem is also contingent upon the satisfaction of condition (29), whereas, the corresponding condition is always satisfied for the uncapacitated submarking-reachability problem.

Following the notation of Section V, the reduced problem is then expressed as the linear program

$$\begin{aligned}&\text{minimize } \Gamma\\ &\text{subject to } L(\tilde{E}) - \tilde{M}(\tilde{E}) \leqslant \tilde{A}^t\Gamma \leqslant U(\tilde{E}) - \tilde{M}(\tilde{E})\\ &\Gamma \geqslant 0 \qquad (31)\end{aligned}$$

where $\tilde{A}$ is the incidence matrix of $\tilde{G} = (\tilde{V}, \tilde{E})$ which may contain artificial edges as described above.

7.2. A Solution to the Capacitated Submarking-Reachability Problem

We now proceed to solve (31) using our solution to (20). If $L(\tilde{E}) \leqslant \tilde{M}(\tilde{E}) \leqslant U(\tilde{E})$, then it is easy to see that the solution to (31) is simply $\Gamma = 0$. Therefore, we consider the case when $\tilde{M}(\tilde{E})$ is not a feasible marking of $\tilde{G}$. Again, let p_{ij} denote a directed path from vertex $i \in \tilde{G}$ to vertex $j \in \tilde{G}$ and p_{ii} denote a directed circuit through vertex $i \in \tilde{G}$. Summing the inequalities in (31) along a directed circuit $p_{ii} = \{(i, u_1), (u_1, u_2), \cdots, (u_{n-1}, u_n), (u_n, i)\}$ in $\tilde{G}$ through the vertices $\{i, u_1, u_2, \cdots, u_n\}$ gives

$$\sum_{e \in p_{ii}} (L(e) - \tilde{M}(e)) \leqslant \gamma_i - \gamma_{u_1} + \gamma_{u_1} - \gamma_{u_2} + \cdots + \gamma_{u_{n-1}} - \gamma_{u_n} + \gamma_{u_n} - \gamma_i \leqslant \sum_{e \in p_{ii}} (U(e) - \tilde{M}(e))$$

or simply,

$$\sum_{e \in p_{ii}} L(e) \leqslant \sum_{e \in p_{ii}} \tilde{M}(e) \leqslant \sum_{e \in p_{ii}} U(e) \qquad (32)$$

as a necessary condition for the existence of a solution to (31) and, as expected, a sufficient condition is

$$\sum_{e \in C} L(e) \leqslant \sum_{e \in C} \tilde{M}(e) \leqslant \sum_{e \in C} U(e)$$

for every directed circuit C in $\tilde{G}$.

We establish the solution to (31) by first transforming this into an equivalent uncapacitated *auxiliary problem* and then invoking Theorem 6. The solution follows easily once (31) is written in the canonical form as

$$\begin{aligned}&\text{minimize } \Gamma\\ &\text{subject to } \tilde{A}^t\Gamma \geqslant L(\tilde{E}) - \tilde{M}(\tilde{E})\\ &-\tilde{A}^t\Gamma \geqslant \tilde{M}(\tilde{E}) - U(\tilde{E})\\ &\Gamma \geqslant 0. \qquad (33)\end{aligned}$$

Now, if A is the incidence matrix of a directed graph G, then $-A$ is the incidence matrix of the graph G' obtained by reversing the directions of all edges of G. Hence, $-\tilde{A}$ is the incidence matrix of the graph $\tilde{G}'$ obtained by reversing the directions of all edges of $\tilde{G}$. Let $\tilde{E}' \triangleq \{(i, j) | (j, i) \in \tilde{E}\}$ denote the edge set of the reversed graph $\tilde{G}' \triangleq (\tilde{V}, \tilde{E}')$. Construct the matrix $A \triangleq [\tilde{A} | -\tilde{A}]$ by horizontally concatenating $\tilde{A}$ and $-\tilde{A}$. Define a column vector N of dimension $2|\tilde{E}|$ as

$$\begin{aligned}N(\tilde{E}) &\triangleq \tilde{M}(\tilde{E}) - L(\tilde{E})\\ N(\tilde{E}') &\triangleq U(\tilde{E}) - \tilde{M}(\tilde{E}). \qquad (34)\end{aligned}$$

Then, (33) is equivalent to the auxiliary program

$$\begin{aligned}&\text{minimize } \Gamma\\ &\text{subject to } \hat{A}^t\Gamma \geqslant -N\\ &\Gamma \geqslant 0 \qquad (35)\end{aligned}$$

where the matrix $\hat{A}$ is the incidence matrix of the *auxiliary graph* $\hat{G} \triangleq \tilde{G} \cup \tilde{G}' = (\tilde{V}, \tilde{E} \cup \tilde{E}')$ obtained by superimposing $\tilde{G}'$ on $\tilde{G}$. For each edge $e \in \hat{G}$, the marking $N(e)$ is defined as in (34). Clearly, in structure, (35) is equivalent to (20). Thus at least from this point on, the capacitated submarking-reachability problem reduces to an equivalent uncapacitated problem on a larger graph. So, we state the solution to (31), assuming it exists, in the following theorem. Proof of correctness follows from that of Theorem 6.

Theorem 7: Let $\hat{G}$ have no directed circuit of negative token count under the marking N. If $\hat{d}_{ij}$ denotes the shortest distance from vertex i to vertex j in the auxiliary graph $\hat{G} = (\hat{V}, \hat{E}) \triangleq (\tilde{V}, \tilde{E} \cup \tilde{E}')$ under N, then the unique solution to (35) and, hence, (31) is given by

$$\gamma_i \triangleq \max\left\{0, -\min_j \left\{\hat{d}_{ij}\right\}\right\}, \qquad \forall i \in \hat{V}. \ \blacksquare \quad (36)$$

Corollary 7.1: The solution to (31), as given in (36), exists if and only if $\sum_{e \in C} N(e) \geqslant 0$ for every directed circuit C in $\hat{G}$. ■

As in the uncapacitated case, we can combine the solutions for the controlled components with the solution for (35) and obtain the firing numbers realizing the specified final marking from the given initial marking.

VIII. Submarking-Reachability and Network Programming

The approach we have adopted in the previous sections has resulted in a unified treatment of the submarking-reachability problem for both the capacitated and uncapacitated cases. As we will see soon, this approach also enables us to see the link between the submarking-reachability problem and the problem of testing feasibility of the dual transshipment problem.

In this section, we first point out the equivalence between the submarking-reachability problem and the problem of testing feasibility of the dual transshipment problem. We then present in the next section an algorithm for this problem, prove its correctness and termination whenever a solution exists and finally, establish its complexity.

Given a graph $G = (V, E)$ on n vertices and m edges, let A denote the incidence matrix of G. Then, *the dual trans-*

shipment problem is as follows [12]:

$$\begin{aligned} &\text{maximize } \Omega\Sigma \\ &\text{subject to } A^t\Sigma \geqslant -M_0 \\ &\Sigma \geqslant 0 \end{aligned} \tag{37}$$

where Σ is a column vector of dimension n and Ω, called the *weight vector*, is a row vector, also of dimension n. Note that Ω and M_0 are specified.

The above problem can also be formulated as [9]

$$\begin{aligned} &\text{maximize } WM \\ &\text{subject to } B_f M = Z_{\bar{T}} \\ &M \geqslant 0 \end{aligned} \tag{38}$$

where B_f is a fundamental-circuit matrix of G, M is a column vector of dimension m, and W is a row vector of dimension m. Here, $Z_{\bar{T}}$ and W are specified.

First, we note that the constraint part of (37) is exactly the same as that of the (20). So, if we look upon M_0 as an initial marking of G and Σ as a firing-count vector, then the problem of testing feasibility of (37) is exactly the same as the problem of determining a set of firing numbers for the vertices of G which transform M_0, which may be an infeasible marking, to a feasible marking M. Here, we consider all the edges of G as free edges.

On the other hand, suppose the dual transshipment problem appears as in (38). Then, we can identify each entry of $Z_{\bar{T}}$ as a marking on the corresponding chord of the cospanning tree $\bar{T}$ which defines the fundamental-circuit matrix B_f. Again, testing the feasiability of (38) reduces to the problem of determining firing numbers which transform $Z_{\bar{T}}$ to a feasible marking M. This follows from the fact that vertex operations do not change the algebraic sum of the tokens in any circuit.

The above discussion shows that testing the feasibility of the dual-transshipment problem is the same as the submarking-reachability problem, where we are required to take G from an infeasible marking to a feasible one.

IX. Algorithm Reach

The equivalence shown in the previous section underlines the importance of designing an efficient algorithm to solve (20). The solution to this problem is given in Theorem 6. All that we need is to design an algorithm to determine γ_i's as given in this theorem. We propose the following algorithm for this purpose.

Algorithm REACH: Let M be the given initial marking of the graph G with no negative-length directed circuits under M.

While there exists an edge $e = (i, j) \in E$ with $M(e) < 0$ **do** Fire vertex i, $-M(e)$ times, updating M.

The rest of this section is concerned with the proof of correctness and termination of this algorithm and its complexity analysis.

9.1. Proof of Correctness and Termination

In the following, the length $l(p_{ij})$ of a directed path p_{ij} in G will refer to the sum of the markings under M of all the edges in p_{ij}. γ_i's are as defined in Theorem 6. By the *size* of p_{ij}, we refer to the number of edges in p_{ij}. Also, d_{ij} is the length of a shortest path from i to j.

Theorem 8: If $\gamma_i > 0$, Algorithm REACH fires vertex i at least γ_i times.

Proof: We prove the theorem by showing that if there exists a directed negative-length path p_{ij}, then Algorithm REACH fires i at least $|l(p_{ij})|$ times. Proof is by induction on the size of p_{ij}.

Clearly, the result is true if the size of p_{ij} is 1. Assume the result to be true for all negative-length directed paths of size $\leqslant k$. Consider any negative-length directed path p_{ij} of size $k+1$. Let (j', j) be the last edge in p_{ij}.

Case 1: Marking $M(j', j) \geqslant 0$.

In this case, $l(p_{ij'}) = l(p_{ij}) - M(j', j) < 0$. But, $p_{ij'}$ is a path of size k. Hence, by the induction hypothesis, vertex i will be fired at least

$$|l(p_{ij'})| > |l(p_{ij})|$$

times.

Case 2: Marking $M(j', j) < 0$.

In this case, at some step during the execution of Algorithm REACH, vertex j' will have been fired at least $|M(j', j)|$ times. Let M' be the marking at that step. Then, $M'(j', j) \geqslant 0$.

Assume that the vertices i and j' have been fired σ_i and $\sigma_{j'}$ times in reaching M' from M. We shall denote the length of p_{ij} under the marking M' by $l'(p_{ij})$. Clearly, $\sigma_{j'} \geqslant |M(j', j)|$. There is nothing to prove if $\sigma_i \geqslant |l(p_{ij})|$. So, assume that $\sigma_i < |l(p_{ij})|$. Then,

$$\begin{aligned} l'(p_{ij'}) &= l(p_{ij'}) + \sigma_i - \sigma_{j'} \\ &\leqslant l(p_{ij'}) + \sigma_i - |M(j', j)| \\ &= l(p_{ij}) + \sigma_i < 0. \end{aligned}$$

So, under the marking M', the length of $p_{ij'}$ is negative. But, $p_{ij'}$ is of size k. Hence, invoking the induction hypothesis, we find that vertex i will be fired at least $|l(p_{ij})| - \sigma_i$ times after M' has been reached. Thus the algorithm will fire i at least $|l(p_{ij})|$ times starting from M. ∎

Recall (Corollary 6.2) that a vertex i is called a datum in G if under the initial marking M there exists no negative-length directed path originating at i. In other words, $\gamma_i = 0$ if i is a datum.

Next, we note that, at each *step*, Algorithm REACH examines an edge and performs an appropriate vertex firing operation. So, this algorithm may be considered as performing a sequence of vertex firing operations. Let M_i denote the marking of G at the end of the ith step or ith firing operation. Thus Algorithm REACH, starting at M, takes G through a sequence of markings $M_1, M_2, \cdots, M_k, \cdots$. In the following, $d_{ij}^{(k)}$ will refer to the length of a shortest path from i to j in G under M_k. Thus

$$\gamma_i^{(k)} = \max\left\{0, -\min_j \left\{d_{ij}^{(k)}\right\}\right\}.$$

Similarly, $l^k(p_{ij})$ will refer to the length of p_{ij} under M_k.

Consider any vertex i for which $\gamma_i > 0$. Then, let i' denote a vertex such that $\gamma_i = |d_{ii'}|$. As we have seen in Section VI (see proof of Corollary 6.3), each such i' is a datum vertex. Finally, we note that firing a vertex i affects only the value of γ_i in the new marking. So, if the vertices $v_1, \cdots, v_k$ have been fired $\sigma_1, \sigma_2, \cdots, \sigma_k$ times to take G to the marking M_k, then under M_k,

$$\gamma_i^{(k)} = \gamma_i - \sigma_i, \qquad \forall i \in V. \tag{39}$$

Furthermore, for each i, vertex i' continues to be a datum of i under all the markings generated by Algorithm REACH. Thus if

$$\gamma_i^{(k)} > 0$$

then

$$\gamma_i^{(k)} = |d_{ii'}^{(k)}|. \tag{40}$$

If $\gamma_i^{(k)} = 0$, then vertex i is a datum under marking M_k. These crucial properties of Algorithm REACH prove the following.

Theorem 9: Algorithm REACH never fires a datum. ∎

Theorem 10: If G has no negative-length directed circuits under marking M, then Algorithm REACH terminates in a finite number of steps after firing every vertex exactly γ_i times.

Proof: By Theorem 8, Algorithm REACH fires each vertex i at least γ_i times. By Theorem 9, a datum is never fired in this algorithm. This means that each step results in reducing the value of exactly one γ_i. Thus all the γ_i's will eventually be reduced to zero, or equivalently, every vertex i will be fired exactly a total of γ_i times in no more than $\Sigma_{i=1}^{k}\gamma_i$ steps. When all the γ_i's reduce to zero, then in the resulting marking there will be no edges with negative tokens and Algorithm REACH will terminate. ∎

9.2. Complexity Analysis of Algorithm REACH

To bound the number of computational steps required by Algorithm REACH, we implement the algorithm as follows.

First, order the edges as $e_1, e_2, \cdots, e_m$, where m is the number of edges in G. Then, execute the algorithm by first examining e_1, then e_2 and so on, and firing the vertices the appropriate number of times. After the first such *sweep*, perform additional sweeps until an entire sweep results in no firings.

Theorem 11: Assume that G has no negative-length directed circuits under the initial marking M. Consider any vertex i for which $\gamma_i > 0$. If the size of the path $p_{ii'}$ is k, then the vertex i will have been fired γ_i times at the end of the kth sweep. (Note: i' is a datum vertex and $|l(p_{ii'})| = \gamma_i$.)

Proof: Let $p_{ii'} = i, i_1, i_2, \cdots, i_{k-1}, i'$. At the beginning of the first sweep, the marking on the edge (i_{k-1}, i') is $-\gamma_{i_{k-1}}$. Also, i' is never fired. So, at the end of the first sweep, vertex i_{k-1} will have been fired $\gamma_{i_{k-1}}$ times. At the beginning of the second sweep i_{k-1} is a datum and the marking on the edge (i_{k-2}, i_{k-1}) will be $-\gamma_{i_{k-2}}$. So, during this sweep vertex i_{k-2} will be fired $\gamma_{i_{k-2}}$ times. Repeating these arguments, we can see that at the end of the kth sweep, vertex i will have been fired γ_i times. ∎

Theorem 12: If G has no negative-length directed circuits under M, then Algorithm REACH will terminate in no more than n sweeps, where n is the number of vertices in G.

Proof: Each directed path $p_{ii'}$ in G is of size $\leqslant n-1$. So, by Theorem 11, each vertex i will be fired a total of γ_i times in no more than $n-1$ sweeps and the theorem follows. ∎

During each sweep, m edges are examined. Examining an edge requires computing the value of its current marking and firing a vertex. These operations take constant time. Thus each sweep takes $O(m)$ time and, hence, we have the following theorem.

Theorem 13: The complexity of Algorithm REACH is $O(mn)$, if G has no negative-length directed circuits under the initial marking M. ∎

As an example, it may be verified that Algorithm REACH when applied to the graph G of Fig. 4 terminates in the feasible marking shown in Fig. 5. The number of times each vertex is fired by Algorithm REACH is shown in Fig. 4.

We now show how to incorporate, in Algorithm REACH, a mechanism to detect the presence of a negative-length directed circuit in the graph under a given initial marking.

We associate with each vertex i a label denoted by LABEL(i). To begin with, the label of every vertex is set equal to zero. Every time an edge (i, j) with a negative marking is encountered and, as a result, vertex i is fired, we update LABEL(i) setting it equal to j. If, during the nth sweep, the label of any vertex, say i, changes, then it indicates the presence of a negative-length directed circuit. This directed circuit can be obtained by tracing the label values starting at i.

X. Summary and Conclusion

Using a linear programming formulation, we have presented a unified treatment of the submarking-reachability problem for both capacitated and uncapacitated marked graphs. In both cases, the problem is shown to reduce to that of taking a marked graph from an infeasible marking to a feasible one. This is precisely the same as the problem of determining feasibility of the dual transshipment problem which generalizes a wide range of network programming problems. This relationship of the submarking-reachability problem to network programming has motivated the need for an algorithmic solution, resulting in the design of Algorithm REACH presented in Section IX. Algorithm REACH tests the feasibility of the dual transshipment problem in $O(mn)$ time. This is in contrast to the complexity $O(n^3)$ of the well-known Malhotra, Kumar, and Maheshwari's algorithm [13], [14] for the maximum flow problem which can be used to test the feasibility of the transshipment problem.

It has been shown in [9] that starting from any feasible marking a basic marking can be constructed in $O(n^2)$ time. Combining this with the complexity of Algorithm

REACH, we conclude that a basic feasible solution of the dual transshipment can be constructed in $O(mn)$ time.

References

[1] J. L. Peterson, "Petri nets," *Computing Surveys*, vol. 9, pp. 223–252, 1977.

[2] R. M. Karp and R. E. Miller, "Properties of a model for parallel computations: determinancy, termination and queueing," *SIAM J. Appl. Math.*, vol. 14, pp. 1390–1411, 1966.

[3] F. Commoner, A. W. Holt, S. Even, and A. Pneuli, "Marked directed graphs," *J. Comput. Syst. Sci.*, vol. 5, pp. 511–523, 1971.

[4] T. Murata, "Circuit-theoretic analysis and synthesis of marked graphs," *IEEE Trans. Circuits Syst.*, vol. CAS-24, pp. 400–405, 1977.

[5] S. Kumagai, S. Kodama, and M. Kitagawa, "Submarking reachability of marked graphs," *IEEE Trans. Circuits Syst.*, vol. CAS-31, pp. 159–164, 1984.

[6] ——, "Submarking reachability of marked graphs with token capacity constraints," *Trans. IECE, Japan*, vol. E67, pp. 373–378, 1984.

[7] T. Murata, "State equation, controllability and maximal matchings of Petri nets," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 412–416, 1977.

[8] Y. Kajitani and T. Kawakami, "Subgraph reachability of marked graphs," in *Proc. 18th Ann. All. Conf. on Communication, Control and Computing*, pp. 507–513, 1980.

[9] K. Thulasiraman and M. A. Comeau, "Maximum-weight markings in marked graphs: Algorithms and interpretations based on the simplex method," *IEEE Trans. Circuits Syst.*, vol. CAS-34, pp. 1535–1545, Dec. 1987.

[10] J. L. Johnson and T. Murata, "Structure matrices for Petri nets and their applications," *J. Franklin Inst.*, pp. 299–309, 1985.

[11] M. N. S. Swamy and K. Thulasiraman, *Graphs, Networks and Algorithms*. New York: Wiley-Interscience, 1981.

[12] V. Chvatal, *Linear Programming*. New York: W. H. Freeman, 1983.

[13] V. M. Malhotra, M. P. Kumar and S. N. Maheshwari, "An $O(|V|^3)$ algorithm for maximum flows in networks," *Inform. Process. Lett.*, vol. 7, pp. 277–278, 1978.

[14] S. Even, *Graph Algorithms*. Computer Science Press, 1979.

Marc Comeau received the Bachelor's degree in electrical engineering from the Technical University of Nova Scotia, Canada, and the Ph.D. degree from Concordia University, Montreal, Canada, in 1981 and 1986, respectively.

He is a researcher at le Centre de recherche informatique de Montréal, Inc. (CRIM). He is also associated with Concordia University as an Adjunct Assistant Professor in the Department of Electrical and Computer Engineering. His research interests include graph-theoretic modeling, distributed computing, discrete optimization, queueing systems and system performance evaluation.

K. Thulasiraman was born in India in 1942. He received the Bachelor's and Master's degrees in electrical engineering from the University of Madras, Madras, India, in 1963 and 1965, respectively, and the Ph.D. degree in electrical engineering from the Indian Institute of Technology, Madras, in 1968.

He joined the Indian Institute of Technology, Madras in 1965, where he was associated with the Department of Electrical Engineering from 1965 to 1973 and with the Department of Computer Science from 1973 to 1981. He was promoted to the rank of Professor in January 1977. After serving for a year (1981–1982) at the Department of Electrical Engineering, Technical University of Nova Scotia, Halifax, Canada, he joined Concordia University, Montreal, as Professor at the Department of Mechanical Engineering where he was involved in the development of programs in Industrial Engineering at the undergraduate and graduate levels. Since 1984 he has been with the Department of Electrical and Computer Engineering at Concordia University. Earlier, he had held visiting positions at Concordia University during the periods of 1970–1972, 1975–1976, and 1979–1980. He has published over 50 technical papers on different aspects of Electrical Network Theory, Graph Theory and Design and Analysis of Algorithms. He has also coauthored the book, *Graphs, Networks and Algorithms* (New York: Wiley-Interscience, 1981). Recently, he has been awarded a Senior Fellowship by the Japan Society for Promotion of Science. Under this fellowship he will be visiting the Tokyo Institute of Technology, Tokyo during March-July, 1988. Dr. Thulasiraman's current research interests are in Network and Systems Theory, Graph Theory, Parallel and Distributed Computations, Operations Research and Computational Graph Theory with Applications in VLSI Design Automation, Computer Networks, Communication Network Planning, etc.