Journal of Universal Computer Science, vol. 12, no. 9 (2006), 1385-1392
submitted: 31/12/05, accepted: 12/5/06, appeared: 28/9/06 © J.UCS

Synthesis of Optimal Workflow Structure

József Tick
(John von Neumann Faculty of Informatics, Budapest Tech
Budapest, Hungary
tick@bmf.hu)

Zoltán Kovacs
(University of Szeged, Szeged, Hungary
kovacsz@inf.u-szeged.hu)

Ferenc Friedler
(Faculty of IT, University of Pannonia, Veszprem, Hungary
John von Neumann Faculty of Informatics, Budapest Tech
friedler@dcs.vein.hu)

Abstract: Optimal synthesis of workflow structures, the formerly undefined problem, has been introduced. Mathematical programming model is presented for determining the cost optimal workflow system of a given workflow problem. On the basis of a methodology developed for process network synthesis, effective solvers are available for the systematic synthesis of workflow systems.

Keywords: workflow, P-graph, network synthesis
Categories: F.4.1, G.1.6, H.1.0

1 Introduction

Workflow technology has become a general tool for a wide range of business and management applications. The major source of this success in the applications is the resultant higher efficiency of business and management systems. To satisfy the requirements defined by the applications, the mathematical basis has been developed mainly on Petri net theory [see e.g., Kiepuszewski *et al*., 03 and Aalst and Hofstede, 05]. Even though this mathematical methodology provides a foundation for optimal operations of workflows, it is incapable of optimal design of the structure of workflows.

The efficiency of a workflow system is highly depend on its structure or network. Since the same set of workflow problems can usually be solved by a large number of structurally different workflow systems with wide range of costs, the synthesis of the optimal network is crucial in practice.

In the present paper, a new mathematical model has been introduced for workflow system synthesis. The solution of this model is partly based on a methodology formerly developed for processing network synthesis, the P-graph framework. This framework includes a specific modeling techniques and effective algorithms for network synthesis.

Even though there are similarities between production processes and workflow processes, the differences prevent the direct applications of the tools of the design of production processes for the design of workflow systems.

2 Workflow System Synthesis Problem

Let R denote the set of different types of resources (e.g., specific information, or paper) that are available for the workflow processes to be designed from outside. Set R will be called as raw materials. The consumption of a raw material by the workflow process is subject to cost determined by a cost function depending on the volume of the consumed raw material. Let P denote the set of product documents, or products in brief, that must be generated by the workflow process. Let I be the set of all possible intermediate documents that can occur in the process as the output of some activities. Set M of materials, the union of sets R, I, and P, denotes all possible resources and types of documents of the process to be designed. An activity can be given as pair (α, β) where sets α and β are the inputs and outputs of the activity, respectively; they are subsets of M. The set of all plausible activities is denoted by A.

The structural component of a workflow synthesis problem can, therefore, be identified as triplet (P, R, A) on M. The cost of a workflow process that produces the product documents in the required quantity is given as the sum of the costs of the raw materials and the cost related to the activities appearing in the synthesized workflow process. The cost of an activity is determined as the sum of its running cost and investment cost assigned to the period of time examined. Naturally, both the running and investment costs depend on the “size” of the activity, i.e., the volume of the outputs of the activity. In the present paper, the workflow process with minimal cost is to be synthesized. Therefore, the optimization is not for minimizing the processing time, instead the focus is on how to set up the cost optimal system. It is supposed throughout this paper that there is an unlimited intermediate storage opportunity at any activity.

3 Structural Representation: P-Graph

For formally analyzing the structures in workflow synthesis, an unambiguous structural representation is required. Process graph or P-graph, which is directed bipartite graph, has been introduced for this purpose [see Friedlet *et al*. (1992)]. A brief description of P-graph is given below.

Let M be a given finite nonempty set of objects, usually materials, that can be transformed in the process under consideration. Transformation between two subsets of M occurs in an activity. It is necessary to link this activity to other activity through the elements of these two subsets of M.

Let A be the set of activities to be considered in the synthesis; then $A \subseteq \wp(M) \times \wp(M)$ where $A \cap M = \emptyset$. If (α, β) is an activity, i.e. $(\alpha, \beta) \in A$, then α is called the input set, and β, the output set of this activity. Pair (M, A) is termed a process graph or P-graph with the set of vertices $M \cup A$ and the set of arcs $\{(x, y): y = (\alpha, \beta) \in A$ and $x \in \alpha\} \cup \{(y, x): y = (\alpha, \beta) \in A$ and $x \in \beta\}$. P-graph (M, A) is defined to be a subgraph of (M', A'), i.e. $(M, A) \subseteq (M', A')$, if $M \subseteq M'$ and $A \subseteq A'$. The union of two P-graphs (M_1, A_1) and (M_2, A_2), $(M_1, A_1) \cup (M_2, A_2)$, is defined to be P-graph $((M_1 \cup M_2, A_1, \cup A_2)$. The indegree of vertex X, $d^-(X)$, is the number of arcs with endpoint X. If X is a material, then broadly speaking, $d^-(X)$ is the number of operating units producing material X.

Example 1. Let us suppose that set M_1 of materials and set A_1 of activities of P-graph (M_1, A_1) be given as

$M_1 = \{A, B, C, D, E, F, G, H, J, K \}$

and

$A_1 = \{(\{C\}, \{A, F\}), (\{D\}, \{B\}), (\{E, F\}, \{C\}), (\{F, G\}, \{C\}), (\{G, H\}, \{D\}), (\{H\}, \{B\}), (\{J\}, \{F\}), (\{K \}, \{G\}), (\{L\}, \{H\})\}$ [see also Table 1].

Activity #	Input	Output
1	C	A, F
2	D	B
3	E, F	C
4	F, G	C
5	G, H	D
6	H	B
7	J	F
8	K	G
9	K	G
10	L	H

Table 1: Plausible activities of Example 1

P-graph (M_1, A_1) is depicted in Figure 1. Note that the input and output sets of activity1, ({C}, {A, F}), are {C} and {A, F}, respectively, and that the indegree of vertex A, $d^-(A)$, is 1 since one activity produces material A.

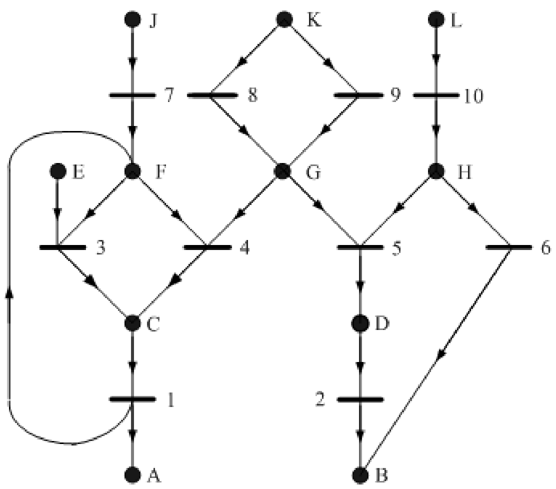

Figure 1: P-graph (M1, A1) where A, B, C, D, E, F, G, H, J, K, and L are materials, and 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 are the activities

P-graph (M, A) contains the interconnections among the activities of A. Each feasible activity network corresponds to a subgraph of (M, A). A product document can be generated by an appropriate network of the activities provided that the problem has at least one feasible solution. It is important to note that a product can usually be generated by different number and types of activities. For determining the optimal network of the workflow, all possible networks of each product must be taken into account.

The number of feasible networks is usually large, therefore, a systematic procedure is required for determining the optimal network. For that, the combinatorial properties of feasible networks is examined first.

Feasible process networks have some common combinatorial properties determining the term “combinatorially feasible network” as given in Definition 1 [Friedler *et al.*, 92a]. Since each feasible workflow network must have these combinatorial properties, the set of subgraphs of (M, A) can be reduced to the set of combinatorially feasible process networks or to the set of solution-networks in short.

Definition 1. Subnetwork (M', A') of (M, A) is called a solution-network of workflow synthesis problem (P, R, A) if

(S1) $P \subseteq M'$, i.e., every product is represented in graph (M', A'),

(S2) $\forall x \in M'$, $d^{-}(x) = 0$ iff $x \in R$, i.e., a vertex from M' has no input if and only if it represents a resource,

(S3) $\forall u \in A'$, $\exists$ path [u, v] in (M', A'), where $v \in P$, i.e., every vertex from A' has at least one path leading to a vertex representing a product, and

(S4) $\forall x \in M'$, $\exists (\alpha, \beta) \in A'$ such that $x \in (\alpha \cup \beta)$, i.e., any vertex from M' must be an input to or output from at least one vertex from A'.

The set of solution-networks is denoted by S(P, R, A); its important properties are expressed by the following theorem and lemma.

Theorem [Friedler *et al.*, 92a]. S(P, R, A) is closed under union.

Lemma [Friedler *et al.*, 92a]. If (M', A') $\in$ S(P, R, A), then M' is the union of all inputs and outputs of activities in A'.

If (M', O') $\in$ S(P, R, A); then, as a consequence of the theorem and the lemma, (M', O') is uniquely determined provided that set O' is given. On the basis of these terms, a specific structure including all activities and their connections can be defined.

Definition 2. Let us assume that S(P, R, A, O, f) $\neq \varnothing$; then, the union of all solution-networks of PNS problem (P, R, A), denoted by μ(P, R, A), is defined to be its maximal network.

Example 1 (revisited). If sets P and R are defined as P={A, B} and R={E, J, K, L}, then Figure 1 shows the maximal network of synthesis problem (P, R, O_1). One of the 50 solution-networks of this problem is given in Figure 2.

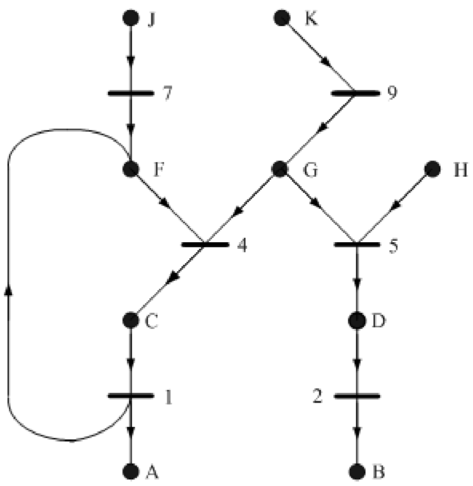

Figure 2: One of the solution-networks of workflow synthesis problem of Example 1

Since the set of solution-networks is finite and closed under union, the maximal network is also a solution-network, i.e., μ(P, R, A) $\in$ S(P, R, A).

Obviously, any activity not included in the maximal network should not be considered for the optimal solution. Since the structure of any optimal solution is a solution-network, the mathematical programming model of workflow system synthesis is to be based on the maximal network.

4 Mathematical Programming Model for Workflow Synthesis

Let us consider a workflow system synthesis problem in which the set of products is denoted by P; the set of resources, by R; and the set of available activities, by A = {a_1, a_2, . . ., a_n}. Moreover, let M = {m_1, m_2, . . ., m_l} be the set of the materials belonging to these activities, and assume that $P \cap R = \emptyset$, $P \subseteq M$, $R \subseteq M$, and $M \cap A = \emptyset$. Then, P-graph (M, A), termed the network of the problem, contains the interconnections among the activities. Furthermore, each feasible network, producing the given set P of products from the given set R of resources using activities from A, corresponds to a subgraph of (M, A), i.e., the network of the workflow network under consideration. For any $1 \leq j \leq n$, let $y_j = 1$ if a_j is contained in this subgraph and $y_j = 0$ otherwise. Thus, this subgraph is determined by vector $(y_1, y_2, \ldots, y_n)$.

Let $E = \{e_1, e_2, \ldots, e_r\}$ be the set of the arcs and assign to arc e_k the continuous variable x_k $(k = 1, 2, \ldots, r)$ representing the quantity of either the material consumed or the product produced. The function for which $\varphi(\{e_{i1}, e_{i2}, \ldots, e_{it}\}) = (x_{i1}, x_{i2}, \ldots, x_{it})$ holds for any subset $\{e_{i1}, e_{i2}, \ldots, e_{ir}\}$ of E is denoted by φ. Finally, variable z_j is assigned to activity a_j $(j = 1, 2, \ldots, n)$ for identification.

Activity represented by node a_j is linked to the system through interconnections represented by its arcs contained in $\omega(a_j)$. Especially, $\omega^-(a_j)$ includes the incoming arcs to vertex a_j, and $\omega^+(a_j)$ includes the outgoing arcs from vertex a_j. It follows that besides depending on y_j and z_j the constraint and cost of a_j depend only on the variables belonging to these arcs, i.e., on $\varphi(\omega^- a_j)) \cup \varphi(\omega^+(a_j))$. Consequently, the constraints on the cost of activity represented by node a_j can be expressed, respectively, by

$$g_j(y_j, \varphi(\omega^-(a_j)), \varphi(\omega^+(a_j)), z_j), \leq 0, \quad j = 1,2, \ldots, n$$
$$f_j(y_j, \varphi(\omega^-(a_j)), \varphi(\omega^+(a_j)), z_j), \leq 0, \quad j = 1,2, \ldots, n$$

where for a fixed value of y_j, g_j are nonlinear, differentiable functions on the practically interesting domain for $j = 1, 2, \ldots, n$. The cost of an activity is usually the sum of the running cost and investment cost of this activity divided by the payback time.

Similarly, the constraint on and cost function of vertex m_i can be given, respectively, as follows:

$$g_j'(\varphi(\omega^-(m_i)), \varphi(\omega^+(m_i))), \leq 0, \quad j = 1,2, \ldots, 1$$

and

$$f_j'(\varphi(\omega^-(m_i)), \varphi(\omega^+(m_i))), \leq 0, \quad j = 1,2, \ldots, 1$$

In practice, g' and f' are usually linear; the former represents the balances and specifications of the products, i.e., quantity and quality, and the latter, the cost of resources. This mathematical model usually contains binary or integer variables mainly related to the investment costs of the activities resulting in a mixed integer optimization problem.

5 Solution Procedure

The mathematical programming model of workflow network synthesis given in Chapter 4, can be generated systematically by the maximal network generation algorithm, algorithm MSG [Friedler *et al.*, 1993]. Since this algorithm is polynomial in the number of activities, it can determine the mathematical model of large workflow problems. Then, there are two options for solving this model. It can be solved by commercially available mixed integer programming solvers or specific solvers developed for the P-graph framework. The application of the specific solvers will be shown briefly.

Algorithm SSG [Friedler, 92b] is for generating all combinatorially feasible workflow networks. These networks can be simulated, evaluated and compared by available tools. Since the binary or integer variables of the mathematical programming model is usually related to the structure of the system, the mathematical model of the individual combinatorially feasible networks does not include any integer variable. Consequently, the evaluation can be performed by LP or NLP method.

If the number of plausible activities is large, the available general purpose mixed integer solvers become ineffective. Also, the exhaustive enumeration by algorithm SSG may require high level of computational effort. For solving complex problems, however, the combinatorial properties of feasible networks can be further exploited. The resultant procedure, the accelerated branch and bound method [Friedler *et al*., 96] developed for network synthesis reduces the search space to the combinatorially feasible networks, resulting in a high acceleration relative to basic branch-and-bound solvers [Peters, 03].

6 Concluding Remarks

The synthesis of cost optimal workflow structures has been introduced and formally defined. The mixed integer programming model of this problem has been identified. The solution of this model is proposed on the basis of P-graph based network algorithms.

References

[Aalst and Hofstede, 05] W.M.P. van der Aalst and A.H.M. ter Hofstede, YAWL: yet another workflow language, Information Systems, 30, 245-275, 2005.

[Friedler, 98] F. Friedler, L. T. Fan, B. Imreh, Process Network Synthesis: Problem Definition, Networks, 28, 119-124, 1998.

[Friedler, 96] F. Friedler, J. B. Varga, E. Feher, L. T. Fan, Combinatorially Accelerated Branch-and-Bound Method for Solving the MIP Model of Process Network Synthesis, State of the Art in Global Optimization, Computational Methods and Applications (Eds: C. A. Floudas and P. M. Pardalos), pp. 609-626, Kluwer Academic Publishers, Dordrecht, 1996.

[Friedler, 95] F. Friedler, J. B. Varga, L. T. Fan, Decision mapping: A Tool for Consistent and Complete Decisions in Process Synthesis, Chemical Engineering Science, 50(11), 1755-1768, 1995.

[Friedler, 93] F. Friedler, K. Tarjan, Y. W. Huang, L. T. Fan, Graph-Theoretic Approach to Process Synthesis: Polynomial Algorithm for Maximal Structure Generation, Computers and Chemical Engineering, 17(9), 929-942, 1998.

[Friedler, 92a] F. Friedler, K. Tarjan, Y. W. Huang, L. T. Fan, Graph-Theoretic Approach to Process Synthesis: Axioms and Theorems, Chemical Engineering Science, 47(8), 1973-1988, 1992.

[Friedler, 92b]Friedler, F., K. Tarjan, Y. W. Huang, and L. T. Fan, Combinatorial Algorithms for Process Synthesis, Computers Chem. Engng, **16**, S313-320 (1992).

[Kiepuszewski, 03] B. Kiepuszewski, A. H. M. ter Hofstede, W. M. P. van der Aalst, Fundamentals of Control Flow in Workflows, Acta Infromatica, 39(3), 143-209, 2003.

[Peters, 03] M. S. Peters, K. D. Timmerhaus, and R. E. West, Plant Design and Economics for Chemical Engineers, McGraw Hill, 5th Edition, 2003.