Designing Manufacturing Facility Layouts to Mitigate Congestion

Min Zhang, Rajan Batta, and Rakesh Nagi

Department of Industrial and Systems Engineering, 438 Bell Hall,
University at Buffalo (SUNY), Buffalo, NY, 14260, USA

November 2009

Abstract

When workflow congestion is prevalent, minimizing total expected material handling time is more appropriate than minimizing a distanced-based objective in manufacturing facility layout design. We present a model labeled Full Assignment Problem with Congestion (FAPC), which simultaneously optimizes the layout and flow routing. FAPC is a generalization of the Quadratic Assignment Problem (QAP), a classical problem for the location of a set of indivisible economical activities. A branch-and-price algorithm is proposed and a computational study is performed to verify its effectiveness as a solution methodology for the FAPC. A numerical study confirms the benefits of simultaneous consideration of layout and routing when confronted with workflow congestion. A detailed simulation for a case problem is presented to verify the overall benefits of incorporating congestion in layout/routing. A critique of FAPC with two alternative models is also provided. Three conclusions are offered from our work. First, a combination of re-layout and re-routing is a more powerful way to mitigate the impact of workflow congestion rather than using just the re-layout or just the re-routing options. Second, it is important to model workflow congestion in a manufacturing facility – namely, ignoring it can result in a significantly poor design. Third, the QAP layout is dominated by the FAPC layout for situations of medium workflow intensity.

Keywords: Facility Redesign, Workflow Congestion, Material Handling, Flow Re-routing.

1 Introduction and Motivation

Minimizing workflow congestion is an important concern in a manufacturing facility. Using alternative paths as a means to alleviate congestion is the subject of a recent paper by Zhang *et al.* (2009). They propose the notion of workflow congestion in the context of material handling equipment interruptions in a manufacturing or warehousing facility and use a combination of probabilistic and physics-based models for modeling workflow interruptions and evaluation of the expected link travel time. The major finding in their paper is that re-routing of traffic in a congested facility can significantly alleviate congestion delays and improve the efficiency of material movement. When re-routing is not sufficient, the more drastic (and expensive) step of re-layout is needed because concurrent re-routing and re-layout allow greatest flexibility to alleviate workflow congestion.

Layout design and redesign have been extensive research areas including continuous and discrete models. We will restrict our following literature review to discrete layout models. The Quadratic Assignment Problem (QAP) has been a prevalent discrete model for facility layout problems where the material handling cost is captured as the sum of products of flow between a pair of facilities and the appropriate distance between their assigned locations. While QAP allows the distance between locations to be pre-computed to reflect different application needs, the distance is generally the shortest path distance. The implicit assumption in the QAP is that there is no interaction among the shortest paths and therefore in practice, congestion may be exacerbated in a QAP layout using the shortest path distances.

The Quartic Assignment Problem (QrAP) is the first quantitative attempt to incorporate workflow interference in facility layout design (Chiang *et al.*, 2002 and 2006). QrAP is an extension of QAP with an objective function of higher degree, where cost is accrued when "pairs" of material handling flows interact – simultaneous interaction of more than two flows is not considered. In the Full Assignment Problem (FAP) proposed by Nagi (2006), a material handling network is incorporated which connects a set of locations to be occupied by a set of facilities and congestion (interaction between flow) is limited by placing a bound on the amount of flow that can occur on a link. In addition, FAP is an integrated model in the sense that the layout and flow distances are determined by the model simultaneously. We notice that attentions have been paid to integrated design of layout and routing in various contexts (Montreuil 1991, Ioannou 2007, Castillo and Peters 2003, 2004, Yang and Peters 1997, Kulturel-Konak *et al.* 2004). For example, Montreuil (1991) introduced a model similar to FAP for the continuous layout case. No solution methods are provided to these complex models. Norman *et al.* (2001) integrates design of block layout and selection of I/O stations assuming uncapacitated aisles. Ioannou (2007) presents a formulation of integrating design of layout and material handling system. It is noteworthy that Ioannou also proposed to place bounds on flow links in the network to avoid creating congested links in his model for concurrent layout and material handling system design. While these models are remarkable in establishing the integrated nature of layout and material handling system, integrated design of layout and material handling system is not the objective of this paper. An existing layout and aisle network are assumed and our work is to develop a discrete model to investigate (1) the potential of designing layout and flow routing simultaneously to alleviate congestion, (2) the impact of re-layout on alleviating congestion compared with re-routing of flow.

The paper is organized as follows. We first introduce a mathematical model referred to as the Full Assignment Problem with Congestion (FAPC) in Section 2. A branch-and-price algorithm is introduced to solve the FAPC in Section 3. Computational results are reported in Section 4. Section 5 presents a numerical study whose goal is to ascertain the benefits of simultaneously considering layout and flow routing. Section 6 contains a case study based on simulation model runs − its goal is to ascertain the benefits of incorporating congestion in layout and routing design. Section 7 provides a comparison of the QAP, QrAP and FAPC approaches, focusing on both data requirements and computational requirements. Finally we present our summary and conclusions in Section 8.

2 The FAPC Model

The FAPC model incorporates material handling congestion as a criterion and aims at reducing congestion by allowing both re-layout and flow re-routing. The current paper will employ the result developed in Zhang *et al.* (2009) and a brief introduction is presented. Their paper discusses the notion of workflow congestion in the context of material handling equipment interruptions in a manufacturing facility where a vehicle may be interrupted (1) at intersections due to the stop-sign rule, (2) by another approaching vehicle, (3) by pedestrians crossing aisles, or (4) at pickup/drop-off locations. A probabilistic model is used to describe the process of encountering interruptions as a vehicle is traveling through a link. A physics-based model is used to calculate the travel time in the presence of interruptions. Combination of these two models allows an approximate estimation of the expected link travel time. The interruption rate is then assumed to be a three-step function of flow. This assumption companioned with the result of the expected link travel time given an interruption rate leads to the simplification that travel time through a link is a three-step function[1] of flow as shown in Figure 1 (Note that the notation link $\{i, j\}$ implies an undirected link, whereas link (i, j) implies a directed link.). Their study indicates that using alternative paths for routing flow is effective to alleviate congestion.

The current paper utilizes the result above that travel time is a three-step function. Given an existing layout and aisle network, the FAPC model is to explore the potential of alleviating workflow congestion by redesign of the layout and flow routing simultaneously. The following variables are introduced. z_{qm} is binary variables to determine if department

[1]The preliminary work in Zhang *et al.* (2009) indicates that the three-step function is sufficiently approximate.

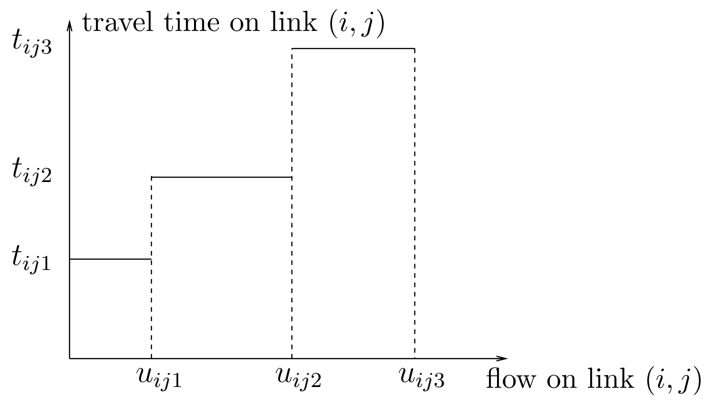

Figure 1: Travel time versus flow on link (i,j)

q is relocated at location m. The travel time as a three-step function implies that there are three options of setting capacities on each link and the expected travel time differs under each option of capacity. Therefore, the binary variable y_{ijl} $(l = 1, 2, 3)$ is to reflect the choice of different capacities on each link. In addition, the model requires continuous flow variables, x^k_{ijl} associated with each link and each commodity. A commodity refers to material being moved between a specific origin-destination pair and the commodity flows are splittable over multiple paths from origin to destination. It is practical to split the commodity flows in a material handling system driven by human operated forklifts especially when the volume of the commodity flows are measured by the number of transportation trips. The simulation result in Zhang *et al.* (2009) reveals that spreading flow over alternative paths can reduce travel time in medium traffic situations. Note that the activity of re-layout is constrained by the existing layout and aisle network. There are four situations if department q is to be located at location m: (1) department q can be exactly fitted to location m, (2) department q can be squeezed in location m, (3) location m is more than spacious to accommodate department q, (4) location m is too small to accommodate department q. Situation (3) requires additional travel time for all flow into and out of department q from the I/O point of department q (newly located at location m) to the aisle. The total additional travel time due to situation (3) is denoted by σ_{qm} (σ_{qm} is equal to 0 for the other three situations). Situation (4) would cause overlap between the department and aisles, hence cause change of aisles as shown in Figure (2). We assume that nodes are neither created, nor destroyed due to change of aisles. As Figure (2) reflects, the change of aisles will cause all flow through link (i,j) to experience extra (reduced) travel time (e.g., links (1,5), (2,5), (3,5) and (4,5)). A simplification assumption is made that the extra (reduced) time is a fixed amount, denoted by e_{qmij} (e_{qmij} is equal to 0 if placement of department q at location m will not result in change of link (i,j)). Additionally, e_{qmij} is approximated by observing the change due to the individual placement of department q at location m. This approximation is necessary because it is very difficult to estimate e_{qmij} if we want to consider interaction among individual department placements. These approximations allow us to focus on investigating the impact of considering workflow congestion in simultaneous design of flow routing and layout.

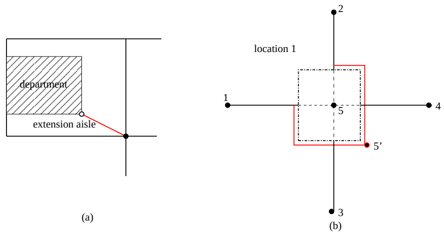

Figure 2: Aisle change

Variables:

x^k_{ijl} the amount of the k^{th} commodity flow that travels through link (i, j) under capacity option l,

$$y_{ijl} = \begin{cases} 1 & \text{if link } \{i, j\} \text{ has capacity option } l \text{ selected,} \\ 0 & \text{otherwise,} \end{cases}$$

$$z_{qm} = \begin{cases} 1 & \text{if department } q \text{ is located at location } m, \\ 0 & \text{otherwise.} \end{cases}$$

Data:

N set of nodes in the aisle network (indexed with i, j)
$\bar{A}$ set of undirected links (indexed with $\{i, j\}$)
A set of directed links (indexed with (i, j))
M set of locations where departments can be placed (indexed with m)
c_m the node index in N for the pickup/drop-off point of location $m \in M$
Q set of departments (indexed with q),
K set of commodities (indexed with k),
o_k the index of origin department of commodity k,
s_k the index of destination department of commodity k,
F_k the amount of flow for commodity k,
$V(i)$ set of neighbors of node i in the aisle network,
u_{ijl} capacity of link $\{i, j\}$ under capacity option l,
t_{ijl} travel time for link (i, j) under capacity option l,
σ_{qm} total additional travel time for all flow into and out of department q that is located at location m
e_{qmij} extra (reduced) travel time on link (i, j) due to aisle change when department q is located at location m

The mathematical formulation is as follows:

$$
\begin{aligned}
L_A: \quad & \min \sum_k \sum_{(i,j)\in A} \sum_{l=1}^3 \tau_{ijl}^k + \sum_{q\in Q}\sum_{m\in M} \sigma_{qm} z_{qm} && && (1)\\
s.t. \quad & \tau_{ijl}^k \geq t_{ijl} x_{ijl}^k + && && \\
& \sum_{q\in Q}\sum_{m\in M} e_{qmij} z_{qm} && \forall k \in K, \forall (i,j)\in A, \forall l = 1,2,3, && (2)\\
& \sum_{k\in K} (x_{ijl}^k + x_{jil}^k) \leq u_{ijl} y_{ijl} && \forall \{i,j\} \in \bar{A}, \forall l = 1,2,3, && (3)\\
& \sum_{l=1}^3 y_{ijl} = 1 && \forall \{i,j\} \in \bar{A}, && (4)\\
& \sum_{j\in V(c_m)} \sum_{l=1}^3 x_{c_m jl}^k - \sum_{j\in V(c_m)} \sum_{l=1}^3 x_{jc_m l}^k && && \\
& = F_k z_{o_k m} - F_k z_{s_k m} && \forall k\in K, \forall m \in M, && (5)\\
& \sum_{j\in V(i)} \sum_{l=1}^3 x_{ijl}^k - \sum_{j\in V(i)} \sum_{l=1}^3 x_{jil}^k = 0 && \forall k \in K, \forall i \in N \setminus \cup_{m\in M} c_m, && (6)\\
& \sum_{j\in V(c_m)} \sum_{l=1}^3 x_{c_m jl}^k \geq F_k z_{o_k m} && \forall k\in K, \forall m \in M, && (7)\\
& \sum_{j\in V(c_m)} \sum_{l=1}^3 x_{jc_m l}^k \geq F_k z_{s_k m} && \forall k\in K, \forall m \in M, && (8)\\
& \sum_{m\in M} z_{qm} = 1 && \forall q \in Q, && (9)\\
& \sum_{q\in Q} z_{qm} = 1 && \forall m \in M, && (10)\\
& x_{ijl}^k \geq 0 && \forall k \in K, \forall (i,j)\in A, \forall l = 1,2,3, && (11)\\
& y_{ijl} \in \{0,1\} && \forall \{i,j\} \in \bar{A}, \forall l = 1,2,3, && (12)\\
& z_{qm} \in \{0,1\} && \forall q \in Q, \forall m \in M. && (13)
\end{aligned}
$$

The objective function (1) intends to minimize the total expected material handling time. The implicit assumption here is that travel time is a reasonable indicator of congestion. Variables τ_{ijl}^k's and constraints (2) are introduced to retain linearity of the model. Constraints (3) are the capacity constraints. Constraints (4) indicate that only one capacity option can be selected for a link. Constraints (5) and (6) are normal flow conservation constraints. Specifically, constraints (5) restrict that the net flow at node c_m is equal to F_k if department o_k is located at location m, or $-F_k$ if department s_k is located at location m, or 0 if neither o_k nor s_k is located at m (i.e. node c_m is a transshipment node for commodity k). Constraints (6) restrict that the net flow at a node that is neither a pickup nor a drop-off point should be equal to 0. Constraints (7) and (8) are in fact redundant but valid constraints that serve to tighten the formulation. Our numerical experience indicates that constraints (7) and (8) effectively tighten the formulation 47.4% on average compared with the LP relaxation of L_A without these constraints. Constraints (9) and (10) are assignment constraints indicating that one location must accommodate only one department, and one department must occupy only one location.

3 Branch-and-Price Approach

It is well known that QAP is NP-hard, and in practice has been proven to be among the most difficult discrete optimization problems for exact solution within reasonable time. Problems of size greater than 20 remain challenging (Anstreicher *et al.*, 2002). Therefore a variety of heuristic approaches have been proposed for its solution, including tabu search (Taillard, 1991), simulated annealing (Connolly, 1990), greedy randomized adaptive search procedure (Resende *et al.*, 1996), genetic algorithm (Tate and Smith, 1995) and ant systems (Gambardella *et al.*, 1999). FAPC is even more difficult than QAP because it needs to tackle a variant of the multicommodity flow problem for every layout in the solution space that may contain $(n-1)!$ permutations as in the QAP case. These facts necessitate quick and effective heuristic methods.

We determine to use the Branch-and-Price (BNP) algorithm to solve FAPC. Column Generation (CG) is a powerful pricing framework for solving large scale combinatorial problems based on the decomposition principle. One of the advantages of CG is that it starts with a small set of "good" columns and generates better columns systematically through a pricing scheme. However, CG may not provide an integral solution and, in these situations finding an optimal or even integer solution requires branching. The BNP algorithm is a generalization of the BNB algorithm and it employs CG as a bounding scheme at each node instead of LP relaxation. The CG algorithm is first described because the BNP algorithm builds on CG.

3.1 Column Generation

To develop a CG algorithm for the FAPC, the formulation (L_A) is decomposed into a master and sub-problems. The master problem, with an initial set of columns, is solved as a relaxed LP problem, wherein a column represents a complete pattern of allocating departments to locations. Additional columns are generated by solving the pricing problem/sub-problem. The sub-problem uses dual multipliers from the master problem to generate a new column with a favorable reduced cost. Thus the sub-problem generates allocation patterns that improve the present LP relaxation solution of the master problem.

We define some additional terms that are used in the master problem and sub-problem for the CG framework.

H_z set of columns for assignment options,

$Z^h_{qm} = \begin{cases} 1 & \text{if department } q \text{ is located at location } m \text{ in column } h, h \in H_z, \\ 0 & \text{otherwise,} \end{cases}$

$\delta_h = \begin{cases} 1 & \text{if the } h^{th} \ (h \in H_z) \text{ column is selected,} \\ 0 & \text{otherwise.} \end{cases}$

H_y set of columns for capacity options,

$Y^h_{ijl} = \begin{cases} 1 & \text{if capacity option } l \text{ is selected for link } \{i,j\} \text{ in column } h, h \in H_y, \\ 0 & \text{otherwise,} \end{cases}$

$\eta_h = \begin{cases} 1 & \text{if the } h^{th} \ (h \in H_y) \text{ column is selected,} \\ 0 & \text{otherwise.} \end{cases}$

δ_h is a decision variable and represents the column or pattern consisting of a complete instance of z_{qm} variables where $z_{qm} \leq I_{qm}$. The term Z^h_{qm} is the value of z_{qm} in the h^{th} column. η_h is another decision variable and represents the column or pattern consisting of a complete instance of y_{ijl} variables. The term Y^h_{ijl} is the value of y_{ijl} in the h^{th} column. Note that Z^h_{qm}'s and Y^h_{ijl}'s are constants in the master problem.

(1) Restricted Master Problem (RMP):

$$
\begin{aligned}
L_{A\_RMP}: \quad & \min \textstyle\sum_{k\in K}\sum_{(i,j)\in A}\sum_{l=1}^{3}\tau^k_{ijl} + \\
& \textstyle\sum_{h\in H_z}(\sum_{q\in Q}\sum_{m\in M}\sigma_{qm}Z^h_{qm})\delta_h && (14)\\
s.t. \quad & \tau^k_{ijl} \geq t_{ijl}x^k_{ijl} + \\
& \textstyle\sum_{h\in H_z}(\sum_{q\in Q}\sum_{m\in M}e_{qmij}Z^h_{qm})\delta_h && \forall k\in K, \forall (i,j)\in A, \forall l=1,2,3, \quad (15)\\
& u_{ijl}\textstyle\sum_{h\in H_y}(Y^h_{ijl}\eta_h) \geq \sum_{k\in K}(x^k_{ijl}+x^k_{jil}) && \forall \{i,j\}\in \bar{A}, \forall l=1,2,3, \quad (16)\\
& \textstyle\sum_{h\in H_y}\eta_h = 1 && (17)\\
& \textstyle\sum_{j\in V(c_m)}\sum_{l=1}^{3}x^k_{c_mjl} - \sum_{j\in V(c_m)}\sum_{l=1}^{3}x^k_{jc_ml} \\
& = F_k\textstyle\sum_{h\in H_z}Z^h_{o_km}\delta_h - F_k\sum_{h\in H_z}Z^h_{s_km}\delta_h && \forall k\in K, \forall m\in M, \quad (18)\\
& \textstyle\sum_{j\in V(c_m)}\sum_{l=1}^{3}x^k_{c_mjl} \geq F_k\sum_{h\in H_z}Z^h_{o_km}\delta_h && \forall k\in K, \forall m\in M, \quad (19)\\
& \textstyle\sum_{j\in V(c_m)}\sum_{l=1}^{3}x^k_{jc_ml} \geq F_k\sum_{h\in H_z}Z^h_{s_km}\delta_h && \forall k\in K, \forall m\in M, \quad (20)\\
& \textstyle\sum_{j\in V(i)}\sum_{l=1}^{3}x^k_{ijl} - \sum_{j\in V(i)}\sum_{l=1}^{3}x^k_{jil} = 0 && \forall k\in K, \forall i\in N\setminus \cup_{m\in M}c_m, \quad (21)\\
& \textstyle\sum_{h\in H_z}\delta_h = 1 && (22)\\
& x^k_{ijl} \geq 0 && \forall k\in K, \forall (i,j)\in A, \forall l=1,2,3, \quad (23)\\
& \eta_h \in \{0,1\} && \forall h\in H_y, \quad (24)\\
& \delta_h \in \{0,1\} && \forall h\in H_z. \quad (25)
\end{aligned}
$$

The objective function is the same as (1) and seeks to minimize the overall expected travel time. Constraints (15), (16) and (18) through (21) are the same as constraints (2), (3) and (5) to (8) of (L_A). Constraint (17) ensures that only one pattern of link capacities can be selected.

Constraint (22) states that only one pattern of departmental assignment to locations can be selected. Constraints (23) are non-negativity constraints. Finally, Constraints (24) and (25) are binary constraints. If all the possible columns are introduced in the master problem, the problem is equivalent to (L_A). However, this needs enumeration of numerous patterns of link capacities as well as assignment patterns, and the latter one increases exponentially with the number of departments. Consequently we start with a small set of columns and generate them systematically through iterations between the master and sub problems. With such a set of columns the master problem is solved as a relaxed LP, i.e., both η_h's and δ_h's are treated as continuous variables with a value between 0 and 1.

The general CG scheme is stated as follows. The relaxed master problem when solved optimally generates a set of dual multipliers corresponding to each constraint. The set of these dual multipliers is utilized to form the objective of the sub-problem. Solving the sub-problem generates a column with favorable reduced cost that is to be added to the master problem. In general, there is one master problem and one sub-problem. However in our problem, there are two types of binary variables in the master problem that are independent of each other. Therefore we have two sub-problems that generate new columns independently of each other. Let ω_{ijl} and ρ_1 be the multipliers corresponding to constraints (16) and (17). These multipliers are utilized to form the objective of sub-problem 1 represented as follows.

$$
\begin{aligned}
SP_1: \quad & \max \textstyle\sum_{\{i,j\}\in\bar{A}} \sum_{l=1}^{3} u_{ijl}\omega_{ijl}y_{ijl}, & & \\
s.t. \quad & \textstyle\sum_{l=1}^{3} y_{ijl} = 1 & & \forall \{i,j\} \in \bar{A}, \\
& y_{ijl} \in \{0,1\} & & \forall \{i,j\} \in \bar{A}, l = 1,2,3.
\end{aligned}
$$

Sub-problem 1 has integral property and is trivial to solve by setting $y_{ij\hat{l}} = 1$ and $y_{ijl} = 0$ if $l \neq \hat{l}$ for $\{i,j\} \in \bar{A}$, where $\hat{l} = argmax_l\{u_{ijl}\omega_{ijl}\}$.

The dual multipliers corresponding to constraints (15), (18), (19), (20), and (22) are denoted by ϵ_{ijl}, α_{km}, β_{km}, γ_{km}, and ρ_2. Sub-problem 2 utilizes these dual multipliers to generate a column with a favorable negative reduced cost, and it is a linear assignment problem as follows that can be solved polynomially.

$$
\begin{aligned}
SP_2: \min \quad & \textstyle\sum_{q\in Q} \sum_{m\in M} \bar{C}_{qm} z_{qm}, & & \\
s.t. \quad & \textstyle\sum_{m\in M} z_{qm} = 1 & & \forall q \in Q, \\
& \textstyle\sum_{q\in Q} z_{qm} = 1 & & \forall m \in M, \\
& z_{qm} \in \{0,1\} & & \forall q \in Q, \forall m \in M,
\end{aligned}
$$

where $\bar{C}_{qm}$ is obtained from the equation below:

$$\bar{C}_{qm} = \sigma_{km} + \sum_{k \in K} F_k(\alpha_{km}O_{kq} - \alpha_{km}D_{kq} + \beta_{km}O_{kq} + \gamma_{km}D_{kq}) + \sum_{k \in K} \sum_{(i,j) \in A} \sum_{l=1,2,3} (e_{qmij}\epsilon_{ijl}). \quad (26)$$

In the equation above, O_{kq} is equal to 1 if q is the origin department of the k^{th} commodity, otherwise it is 0; similarly D_{kq} is equal to 1 if q is the destination department of the k^{th} commodity, otherwise it is 0.

In the above CG scheme (referred to as *CG1*), two sub-problems are formed independently to generate favorable columns for capacity selection on links and location assignment respectively. We also consider another CG scheme (referred to as *CG2*) where y_{ijl}'s remain in the master problem as they do in the original formulation. Therefore sub-problem 1 in *CG1* is no longer required in the CG framework, and sub-problem 2 (SP_2) still remains to generate favorable columns for departmental assignment to locations as it does in the *CG1* scheme. In summary, in the *CG2* scheme, constraint (17) in the master problem (L_{A_RMP}) is replaced by constraints (4), and there is only one sub-problem (SP_2). Our preliminary experience indicates that *CG1* achieves a tighter lower bound than *CG2*, which makes the BNP algorithm based on the *CG1* scheme more effective. Therefore we only consider *CG1* in the following BNP algorithm.

3.2 Branch-and-Price Algorithm

In our case, at the termination of a CG procedure, if the δ_h variables are not binary then the solution to the master problem is not the solution to the problem presented in (L_A). To find the optimal solution in this situation we employ the BNP algorithm whose implementation issues are discussed in the following.

(1) Branching Strategy: Branching strategies in 0-1 integer programs involve fixing a variable or a set of variables to 0 or 1. There are two types of binary variables and the BNP algorithm in our case only branches on z_{qm} variables. An integral solution of δ_h variables indicates that a layout is selected, and this layout is used as an input to the model in Zhang *et al.* (2009) to determine the near optimal values of y_{ijl} variables associated with this layout[2]. In other words, we stop further branching on y_{ijl} variables when the BNP algorithm reaches a node with δ_h variables being integral. The branching variable is determined by calculating the weighted fractional z_{qm} value that is closest to 0.5. The weighted z_{qm} value for a particular q and m is given by the sum of the product of Z_{qm}^h value and δ_h variable value. The following equation determines the next branching variable:

[2]Zhang *et al.* (2009) uses a Lagrangean heuristic to determine near optimal y_{ijl} values.

$$(q,m) = argmin_{q,m}|\sum_h(\delta_h Z^h_{qm}) - 0.5|.$$

Using the branching variable z_{qm} obtained, we set $z_{qm} = 1$ on one branch and z_{qm}=0 on the other. Fixing the variable z_{qm} to one restricts the master problem to consider only those columns that have variable $z_{qm} = 1$. In the sub-problem this has a different meaning. Setting the variable $z_{qm} = 1$ in the sub-problem ensures that all the feasible solutions generated at that node will have this particular $z_{qm} = 1$. These ensure that consistent solutions are generated by the sub-problem, or infeasibility is detected.

(2) Tree Traversal Strategy: When the BNP algorithm is searching the solution space, best-first search is used as the node selection strategy. In best first search the node with the smallest lower bound is considered for further exploration first. The choice is motivated by the fact that best first search offers faster improvement of the lower bound (upper bound) of a minimization (maximization) problem.

(3) Updating Lower and Upper Bounds: All the open nodes to be explored are stored in a list in ascending order regarding the lower bound value at each node. Therefore the lower bound, denoted by z_{LB}, is updated with the lower bound value at the head node in the list. As for the upper bound, if an integer solution (i.e., δ_h variables are binary) is found at a particular node, the corresponding objective value is a feasible upper bound to the original problem. Let z_{UB} be the best upper bound found to this point. As we advance deeper into the tree, z_{UB} is updated whenever a superior integer solution is found at a leaf node.

4 Computational Performance

We first investigate the difficulty of the FAPC model compared with the QAP model. Then the performance of the BNP algorithm is investigated. The examples used to conduct computational tests in this section are generated in the following manner. The networks are non-full grid networks where length of a link $\{i,j\} \in \bar{A}$, denoted by d_{ij}, and the row and column distances of the grid is randomly generated from the set $\{40, 50, 60, 80, 100, 120\}$. The capacity options for link $\{i,j\}$, u_{ijl}, are generated in the following manner: $u_{ij1} = 2d_{ij}$, $u_{ij2} = 4d_{ij}$ and $u_{ij3} = 6d_{ij}$. The three levels of travel times, t_{ijl}, are generated in a different manner: $t_{ij1} \sim U(1.1d_{ij}, 1.2d_{ij})$, $t_{ij2} \sim U(1.4d_{ij}, 1.6d_{ij})$ and $t_{ij3} \sim U(1.8d_{ij}, 2.2d_{ij})$, where U denotes a uniform distribution. The origin and destination nodes of a commodity are randomly selected. Flow amount is an integer randomly drawn from $[24, 72]$. Values of e_{qmij} and σ_{qm} are difficult to generate in a random manner because they are dependent on spatial requirement of a department and spatial area of a location. Therefore in this section, e_{qmij}

and σ_{qm} are assumed to be 0 without impacting the result of computational performance. This would make constraints (15) become equality constraints automatically.

The FAPC formulation generalizes the QAP because every QAP instance can be easily mapped into the FAPC. In order to fairly compare FAPC and QAP, we present FAPC as a linear formulation for QAP in the following that is referred to as Linearized QAP (LQAP):

$$
\begin{aligned}
\textit{LQAP}: \quad & \min \sum_{q_1 \in Q} \sum_{q_2 \in Q} \sum_{m_1 \in M} \sum_{m_2 \in M} d_{m_1 m_2} y_{q_1 q_2 m_1 m_2}, && && (27) \\
s.t. \quad & \sum_{m_2 \in M} y_{q_1 q_2 m_1 m_2} = F_{q_1 q_2} z_{q_1 m_1} && \forall q_1, q_2 \in Q, m_1 \in M, && (28) \\
& \sum_{m_1 \in M} y_{q_1 q_2 m_1 m_2} = F_{q_1 q_2} z_{q_2 m_2} && \forall q_1, q_2 \in Q, m_2 \in M, && (29) \\
& \sum_{m \in M} z_{qm} = 1 && \forall q \in Q, && (30) \\
& \sum_{q \in Q} z_{qm} = 1 && \forall m \in M, && (31) \\
& y_{q_1 q_2 m_1 m_2} \geq 0 && \forall q_1, q_2 \in Q, m_1, m_2 \in M, && (32) \\
& z_{qm} \in \{0, 1\} && \forall q \in Q, m \in M, && (33)
\end{aligned}
$$

where $d_{m_1 m_2}$ denotes the distance between locations m_1 and m_2, $F_{q_1 q_2}$ is the flow amount from department q_1 to department q_2, z_{qm}'s are the binary variables as defined earlier, and $y_{q_1 q_2 m_1 m_2}$ is the continuous flow variable that is equal to $F_{q_1 q_2}$ if q_1 is located at m_1 and q_2 is located at m_2, otherwise it is 0.

CPLEX 9.1 MIP Solver is employed to solve FAPC and LQAP[3]. Table 1 summarizes the results where the $*$ symbol on the superscript of problem number indicates that CPLEX solves this problem to optimality. As seen from Table 1, the time[4] required by CPLEX to solve FAPC increases significantly with increasing problem size and additionally, CPLEX consumes considerably more time to solve FAPC than LQAP when the problem size increases. This is not unexpected because finding the optimal solution to FAPC requires not only examining, in the worst case, a set of layout solutions with (n-1)! permutations, but also deals with a variant of the capacitated multicommodity flow problem for every layout.

4.1 Column Generation

Table 2 provides the computational results of the column generation method for FAPC. The terms GAP_{LB} and GAP_{UB} in Table 2 are obtained from $\frac{\text{optimal value - lower bound}}{\text{optimal value}}$ and

[3]The parameters in the CPLEX MIP Solver are fine tuned and the following settings were found to be beneficial: (1) the mip emphasis is set as *integer feasibility*, (2) the backtrack strategy is set as 1, (3) the nodeselect strategy is set as *best estimate*, (4) the variable strategy is set as *strong branching*, and (5) the probing level is set as the maximum level of 3.

[4]Times reported in this paper are CPU seconds based on a PC with Intel Pentium D processor (3400 MHz) and 2G RAM operating on Windows XP platform.

Table 1: Comparison of CPLEX solution times (in seconds) for FAPC and LQAP

prob #	dept #	node #	link #	comm #	FAPC	LQAP
1*	10	20	27	25	866	6
2*	12	21	29	30	1961	100
3*	13	22	30	32	4717	351
4*	14	23	33	35	33386	1884
5*	15	23	34	38	62866	2601

$\frac{\text{upper bound - optimal value}}{\text{optimal value}}$, respectively. When the CG method converges, there is a final master problem with a set of columns and each column represents a layout. The lower bound is equal to the relaxed LP value of this final master problem. For those columns in the final master problem, we evaluate all these columns using the model in Zhang *et al.* (2009) and choose the best (smallest) value resulting from the best column as the upper bound. Table 2 reveals that the time for the *CG1* scheme to converge increases significantly from within seconds to within minutes as problem size increases. Additionally, the lower bound and upper bound obtained from the CG method still remain undesirable, which justifies further exploration using the BNP method.

Table 2: Column Generation Method for FAPC

prob	dept	node	link	comm	time (sec)		GAP_{LB} (%)		GAP_{UB}(%)
#	#	#	#	#	CG1	LP relaxation	CG1	LP relaxation	(CG1)
1*	10	20	27	25	3.5	0.6	41.6	47.0	49.5
2*	12	21	29	30	22.2	1.1	40.6	46.6	53.8
3*	13	22	30	32	20.9	1.1	43.2	47.6	53.2
4*	14	23	33	35	57.2	1.8	49.1	53.7	57.1
5*	15	23	34	38	116.2	2.3	41.1	46.4	51.6

4.2 Branch-and-Price

The BNP algorithm is coded in C$^\sharp$ and Table 3 displays the computational results where the symbol "–" indicates that no integer solution is achieved, and "GAP" is derived from $\frac{\text{upper bound - lower bound}}{\text{lower bound}}$. An immediate observation is that the BNP algorithm explores less nodes than CPLEX. The reason is that CPLEX branches on both z_{qm} and y_{ijl} variables, whereas the BNP algorithm branches only on z_{qm} variables. The BNP algorithm utilizes the fact that the problem reduces to the problem in Zhang *et al.* (2009) for a given layout, and CPLEX is not able to utilize this information. For small problems 1 to 4, the BNP algorithm achieves an optimal solution within shorter time than CPLEX except problem 3. For larger problems 5 to 7, the BNP algorithm provides better integer solutions than CPLEX within 3 hours and also proves optimality gap less than 8%. For even larger problems 8 to 12, the BNP algorithm does not show significant performance especially with regards to the optimality gap achieved within given time. This is because the BNP algorithm requires considerably more time to explore a node as problem size increases. However, the BNP algorithm is advantageous for providing integer solutions within reasonable time. For instance for problems 10 to 12, the BNP algorithm provides integer solutions, whereas CPLEX does not within the same amount of time.

Table 3: BNP Results

prob	dept	node	link	comm	objective value		nodes explored		solution time (sec)		GAP
#	#	#	#	#	CPLEX	BNP1	CPLEX	BNP1	CPLEX	BNP1	(%)
1*	10	20	27	25	129097.1	129097.1	3716	864	866	552	0
2*	12	21	29	30	102467.4	102467.4	6563	1344	1961	1470	0
3*	13	22	30	32	122778.9	122778.9	12357	3410	4717	5477	0
4*	14	23	33	35	136926.9	136926.9	59675	8750	33386	11345	0
5	15	23	34	38	144175.0	138246.5	7481	2937	3 hrs	3 hrs	3.77
6	16	25	37	40	149564.3	133252.7	6518	2319	3 hrs	3 hrs	4.22
7	18	37	52	45	185484.0	181733.6	5962	1894	3 hrs	3 hrs	7.93
8	20	38	55	50	233570.1	225819.2	3938	1738	3 hrs	3 hrs	14.69
9	22	40	58	55	367657.0	385174.4	1901	952	3 hrs	3 hrs	29.15
10	25	50	68	65	–	463718.4	1758	192	6 hrs	6 hrs	76.33
11	28	59	81	70	–	574164.6	1477	79	6 hrs	6 hrs	221.94
12	30	62	85	75	–	647155.8	1019	45	6 hrs	6 hrs	351.10

5 Numerical Study to Ascertain Benefits of Simultaneous Layout and Routing Design

To answer the question of the benefits of simultaneously considering layout and routing we consider the problem instance whose planar network is shown in Figure 3 and flow matrix in Table 4. The problem instance is randomly generated. The case used in Zhang *et al.* (2009) is not used in this paper because it is special due to the fact that most departments are approximately of equal area. Values of e_{qmij} and σ_{qm} are also assumed to be 0 because the result is subject to our determination of their values. The permutations of the QAP and FAPC layout are $(2, 5, 7, 9, 3, 8, 4, 1, 0, 6)$ and $(5, 2, 0, 1, 8, 3, 4, 9, 7, 6)$ respectively.

We have chosen not to use the QrAP layout for comparison. The reason is that the QrAP layout depends upon the choice of the parameter ω (as defined in Chiang *et al.* 2002, 2006), and there appears to be no formal way of determining ω and connecting it to our flow intensity modeling.

The solution from the FAPC model provides a layout that would generally be different from the QAP layout. We use "FAPC" to denote this layout, and "QAP" to denote the

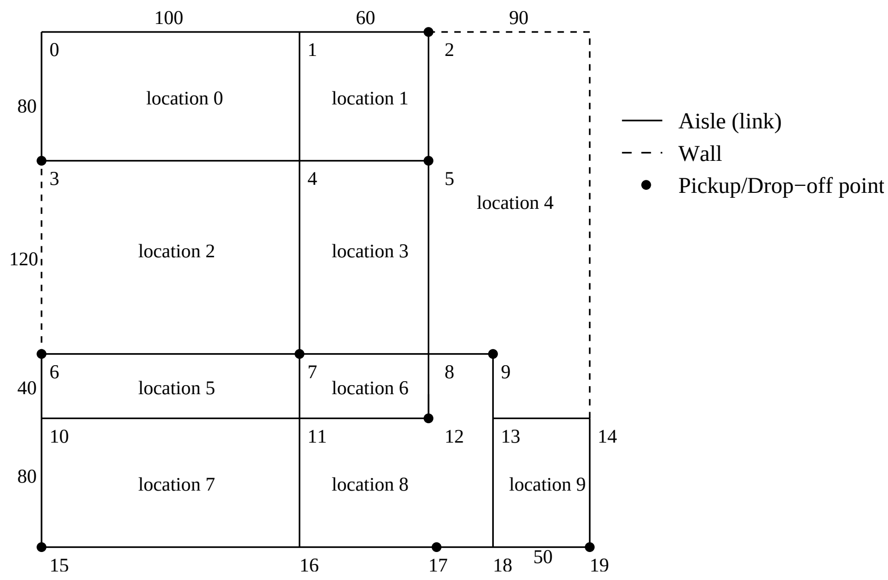

Figure 3: Layout used in numerical study

layout from solving the traditional QAP model. Given a layout, two different flow routing policies could be employed: the shortest path routing policy "SP" and the alternative path routing policy "AP". Consequently, four scenarios (two different routing policies with two different layouts) are to be investigated, and these scenarios are referred to as "FAPC-SP", "FAPC-AP", "QAP-SP" and "QAP-AP". Note that there are several ways to implement the AP approach in operation. One way is to let drivers select path 1 the first 2 times and path 2 the next 3 times, if variables x^k_{ijl}'s indicate that 40% and 60% of an o-d flow should be on path 1 and path 2 respectively, and the cycle can repeat until all delivery requests for the same o-d flow are served. Table 5 displays the expected material handling times for the four scenarios with increasing flow. θ in Table 5 denotes the multiplier to the flow matrix that exhibits the required flow amount over a certain period. The larger the value of θ, the larger the overall flow intensity in the system.

The procedure to evaluate the expected material handling time varies under different scenarios. The column labeled "FAPC-AP" displays the objective value resulting from the FAPC model (the FAPC model ensures that flow routing follows the "AP" policy). Given a layout, we apply the procedure developed in Zhang *et al.* (2009) to determine the best choice of alternative paths for flow routing. Using the QAP layout as input to this model, we obtain results displayed in the "QAP-AP" column. For the "FAPC-SP" and "QAP-SP" scenarios, only one shortest path is allowed for transporting flow between each location pair, and one is randomly selected if multiple shortest paths exist. The overall expected material handling times are evaluated using the function shown in Figure 1. The "−" symbol in Table

Table 4: Flow requirements over an eighteen-hour period, used in numerical study

FR/TO	0	1	2	3	4	5	6	7	8	9	
0											0
1	255		120				60				435
2	60									150	210
3						60				60	120
4	120									150	270
5		90	105				105		150		450
6	120				120			105			345
7					150						150
8		90			150						240
9	210	90			90		90				480
Sum	765	270	225	0	510	60	255	105	150	360	2700

5 indicates infeasibility.

We now highlight the main findings from Table 5. First, alternative flow routing allows a facility with a given layout to accommodate more flow, and further flow capacity can be achieved by seeking an alternative layout. This is implied from the observation that the "QAP-SP" scenario reaches infeasibility when θ is 1.4, the "FAPC-SP" scenario reaches infeasibility when θ is 1.5, and the other two scenarios with alternative flow routing are still feasible when θ is 1.6.

The last five columns in Table 5 display the percentage of reduced travel time for one scenario versus another. The improvement with aggregate effects of both layout change and alternative flow routing is reflected in the column labeled "D vs A". The columns labeled "B vs A" and "D vs C" display the travel time improvement due to the singular effect of alternative flow routing for the QAP and FAPC layouts, respectively. The column labeled "C vs A" is the travel time improvement due to the singular effect of layout change from the QAP to the FAPC layout. Scenarios "QAP-AP" and "FAPC-AP" allow alternative flow routing in the QAP and FAPC layout. Thus the column labeled "D vs B" investigates the effect of alternative flow routing for the FAPC layout compared with the QAP layout.

The general trend observed is that the improvement increases with increasing flow intensity. In addition, the aggregate effects of allowing both layout change and alternative flow routing are more effective than the singular effect of layout change or alternative flow routing. Furthermore, the singular effect of alternative flow routing is more effective than that of layout change. For example, when θ is 1.3, an 8.9% improvement is achieved if allowing both layout change and alternative flow routing, whereas allowing alternative flow routing in the QAP and FAPC layouts yields a 4.1% and 7.0% improvement respectively, and allowing layout change from the QAP layout to the FAPC layout yields a 2.1% improvement. We also observe a positive improvement of 5.0% in the last column. This implies that alternative flow routing is more effective in reducing overall travel time for a FAPC layout than a QAP layout.

Table 5: Results of Numerical Study

θ	total travel time (sec)				travel time improvement (%)				
	QAP-SP(A)	QAP-AP(B)	FAPC-SP(C)	FAPC-AP(D)	D vs A	B vs A	D vs C	C vs A	D vs B
1.0	563385.2	560022.0	562481.5	545680.3	3.1	0.6	3.0	0.2	2.6
1.1	623847.4	617660.0	623085.3	604960.5	4.6	2.6	2.9	1.7	2.1
1.2	737246.6	706550.6	717253.0	689500.4	6.7	4.3	4.0	2.8	2.5
1.3	817453.7	783800.6	800358.2	744470.5	8.9	4.1	7.0	2.1	5.0
1.4	967248.3	926621.0	945339.1	879400.3	9.9	5.1	7.9	3.2	5.1
1.5	–	1077801.0	1101627.4	1014471.0	NA	NA	7.9	NA	5.9
1.6	–	1182340.0	–	1109721.0	NA	NA	NA	NA	6.1

We recognize that other example layouts might yield different numerical results. The above example is one representative case of the problems we studied. The other case studied is special as all departments are of equal area. For this case, it is found out that the scenarios of QAP-AP and FAPC-AP may result in similar performance. This is because there are multiple paths of equal length in a layout where all departments are of equal area. It is also found out that the AP approach is better than the SP approach in both cases. Thus, even though we cannot conclusively state the improvement percentage yielded by the FAPC approach, we can conclusively state that the combination of layout and routing design by the FAPC formulation dominates the traditional QAP-SP approach for moderate congestion levels.

6 Simulation Study to Ascertain Benefits of Incorporating Congestion

We used the same simulator in Zhang *et al.* (2009) to which details can be referred. We only provide a brief description of simulation model in this paper. The model simulates the material handling traffic in the presence of varying speed and four types of interruptions. One of the distinguished features is that the model considers acceleration/deceleration factors to simulate forklifts' travel. Forklifts are defined to be in one of the following four states: (1) accelerating, (2) traveling at normal cruise speed, (3) decelerating, or (4) stationary. The simulator is a discrete time model as time is discretized into small intervals. When the forklift should schedule a deceleration action is determined by the logic that if the forklift does not start deceleration at current time instant and it will result in collision, the forklift should start deceleration at current time instant, otherwise not. In the simulation, the closest dispatching rule, the rules of positioning idle vehicles at release point and serving transport request based on first come first serve principle are implemented. Parameters used in the simulation are summarized in Table (6).

Table 6: Parameters in Simulation Model

planning period	18 hours
discretized time interval	0.1 seconds
interval between pickup requests for each o-d pair	exponential distribution a rate = $\frac{\text{flow amount of this o-d pair}}{\text{the length of the planning period}}$
normal cruise speed	3 feet/second
constant acceleration and deceleration rates	1.5 feet/second^2
time of vehicle crossing an intersections	uniform distribution between 8 seconds and 12 seconds
delay time due to a pedestrian or a vehicle interruption	uniform distribution between 6 seconds and 10 seconds
unloading and loading times	40 seconds

6.1 Benefit of a Combined Routing and Layout Design

We used our simulation model on the case problem presented in Section 5 to answer the broad question of the benefit of modeling congestion in re-layout via an experiment. The procedures of this experiment are as follows: (1) obtain the QAP layout, (2) increase the multiplier of the flow matrix and run the simulation with the QAP layout and shortest path routing policy, (3) perform a regression analysis to obtain the value of the congestion parameter and subsequently obtain the travel time values on each link as shown in Figure (1), (4) increase the multiplier of the flow matrix and solve model (L_A) to obtain the FAPC layouts and their corresponding alternative flow routes. We explore two scenarios. The first scenario (Table 7) have the number of forklifts in operation fixed. The second scenario (Table 8) assumes that the number of forklifts in operation will assure that the average utilization rate of forklifts is between 70% and 80% − a reasonable workload.

Tables 7 and 8 display the percentages of reduction in travel time and blocking time for one scenario versus another. There are different ways to view the results. First of all, we notice that the aggregate effects of both layout change and alternative flow routing are more beneficial than the singular effect of allowing layout change or alternative flow routing for intermediate flow intensity in a facility. This is indicated by the observation that the improvement in the column labeled "D vs A" is greater than the other columns except when θ is either 1.0 or 1.6. (where columns labeled "D vs A" reflect the improvement with aggregate effects of both layout change and alternative flow routing, those labeled "B vs A" and "D vs C" display the improvement due to the singular effect of alternative flow routing for the QAP and FAPC layouts respectively, and the columns labeled "C vs A" reveal the improvement due to the singular effect of layout change from QAP to FAPC).

We also note that it is more effective to reduce congestion by alternative flow routing than layout change. For example in Table 7, when θ is 1.3, the improvement on travel time is 9.4%, 4.0%, 7.7% and 1.9% respectively in columns labeled "D vs A", "B vs A", "D vs C" and "C vs A". This means that the improvement from alternative flow routing for the QAP and FAPC layout is 4.0% and 7.7% respectively, whereas an improvement of 1.9% is observed if only layout change is allowed.

In addition, columns labeled "B vs A", "D vs C" reveal that the improvement from alternative flow routing in the QAP layout is not as significant as that in the FAPC layout. This is because alternative paths are significantly longer than the shortest paths in a typical QAP layout.

6.2 Under What Flow Intensity Level Are Re-layout and Re-routing Beneficial?

Looking at a single column with increasing value of θ in Tables 7 and 8, we observe that the improvement increases with increasing flow, then decreases as flow further increases. For example, in the column labeled "D vs A" of Table 7, the blocking time improvement starts with 15.4%, improvement increases with increasing flow initially, then reaches the peak of 35.3% improvement when θ is 1.3. After reaching the peak, the blocking time improvement decreases as θ becomes larger and it reduces to 13.1% at the maximum flow intensity level (when θ is 1.6). The same general trend is observed for all other columns in Table 7. Furthermore, Table 8 exhibits the same behavior.

In summary, a FAPC layout with alternative flow routing achieves less overall travel time than a QAP layout with alternative flow routing, particularly when the flow intensity is intermediate. In addition, for intermediate flow intensity the aggregate effects of both layout change and alternative flow routing are more effective than the singular effect of layout change or alternative flow routing.

7 Comparison of QAP, QrAP and FAPC Models

The FAPC model developed in this paper takes material handling congestion into consideration, and is capable of determining layout and flow routing simultaneously. The FAPC model is more sophisticated than both the QAP and QrAP models in two ways: (1) The QAP and

Table 7: Simulation Result (Scenario 1): improvement

θ	tasks ♯	forklifts ♯	percentage of saved time (%)									
			travel time					blocking time				
			D vs A (1)	B vs A (2)	D vs C (3)	C vs A (4)	D vs B (5)	D vs A (1)	B vs A (2)	D vs C (3)	C vs A (4)	D vs B (5)
1.0	2690	25	2.1	3.0	1.9	0.2	-0.8	15.4	13.6	9.1	6.9	2.1
1.1	2970	25	4.9	3.3	3.9	1.0	1.7	22.7	14.2	17.0	6.9	10.0
1.2	3235	25	7.1	3.7	4.5	2.8	3.6	27.2	14.5	20.8	8.0	14.8
1.3	3519	25	9.4	4.0	7.7	1.9	5.6	35.3	15.9	27.0	11.4	23.0
1.4	3801	25	7.6	5.1	5.0	2.7	2.7	27.7	18.4	20.2	9.4	11.5
1.5	4026	25	2.5	1.6	1.4	1.2	1.0	13.8	7.7	7.4	6.9	6.6
1.6	4048	25	1.0	0.1	1.2	-0.3	0.8	13.1	7.3	7.3	6.2	6.3
A: QAP-SP; B: QAP-AP; C: FAPC-SP; D: FAPC-AP												

Table 8: Simulation Result (Scenario 2): improvement

θ	tasks ♯	forklifts ♯	percentage of saved time (%)									
			travel time					blocking time				
			D vs A (1)	B vs A (2)	D vs C (3)	C vs A (4)	D vs B (5)	D vs A (1)	B vs A (2)	D vs C (3)	C vs A (4)	D vs B (5)
1.0	2690	18	0.7	1.5	1.1	- 0.4	-0.8	11.5	10.2	7.3	4.5	1.4
1.1	2970	20	3.9	2.2	2.4	1.5	1.6	19.7	10.8	13.0	7.8	10.1
1.2	3235	22	4.7	2.7	3.4	1.3	2.1	24.7	13.7	16.4	9.9	12.7
1.3	3519	24	8.1	2.6	4.0	4.2	5.7	32.6	12.4	18.7	17.2	23.2
1.4	3801	27	5.4	3.3	2.3	3.2	2.2	20.2	11.1	9.6	11.8	10.2
1.5	4026	30	3.6	2.1	1.9	1.7	1.6	16.0	8.3	11.2	5.4	8.4
1.6	4343	33	-1.5	0.8	1.2	-2.8	-2.4	8.5	6.0	5.2	3.5	2.7
A: QAP-SP; B: QAP-AP; C: FAPC-SP; D: FAPC-AP												

QrAP models require less data input. The QrAP model requires a congestion parameter ω in addition to the distance and flow requirement tables. The FAPC model requires an off-line simulation model to determine the average interruption rate as a function of flow, which is then used to determine link travel time as a step function of flow (Figure 1). (2) The computational difficulty of solving the QAP and QrAP models are similar. However, the FAPC model requires more computational effort than the QAP model, as Table 1 reflects. We have compared the FAPC layout with the QAP layout via a simulation model. We did not compare the FAPC layout with the QrAP layout because we were not able to develop a suitable value of the parameter ω for the QrAP layout. Table 9 provides a summary of the comparisons between the QAP, QrAP and FAPC models.

Table 9: Comparison Between Layout Models

	QAP	QrAP	FAPC
Data requirement	Distance matrix Flow matrix	Distance matrix Flow matrix ω: penalty due to a pair of opposed parallel flow	Distance matrix Flow matrix λ: interruption rate as a function of flow
Computational difficulty	Very difficult	As difficult as QAP	More difficult than QAP Simulation needed to estimate coefficients
Modeling sophistication	No workflow congestion considered Distance modeled equivalent with time No re-routing of flow permitted	Workflow congestion modeled only using pairs of flows No re-routing of flow permitted	Workflow Congestion modeled using all the flows Re-routing of flow permitted
Which is better?	Recommended for both low and high workflow intensity	Not studied due to difficulty with determining appropriate ω value	Recommended for medium workflow intensity

8 Summary, Conclusions, Discussion and Further Research

In this paper, we develop the FAPC model, a generalization of the QAP model − with additional modeling capability that incorporates material handling congestion costs and simultaneous consideration of layout and flow routing. A branch-and-price algorithm is specifically designed for this problem. The BNP algorithm generally performs better than CPLEX 9.1 when problem size ranges from 10 to 14 departments. Both CPLEX 9.1 and the BNP algorithm are incapable of solving the FAPC model within reasonable time (assumed to be 3 hours in this paper) for problems with size greater than 20. For these problems, CPLEX 9.1 cannot find a feasible solution within reasonable time, however, the BNP algorithm is at least able to provide feasible solutions.

A comparison between the FAPC and QAP models is attempted via both a numerical and a detailed simulation study. The results show that the aggregate effects of allowing both layout change and alternative flow routing are more beneficial than the singular effect of allowing layout change or alternative flow routing. This concludes that the aggregate effects of layout change and alternative flow routing allow the greatest flexibility to mitigate workflow congestion. However, the singular effect of rerouting of flow is more beneficial than the singular effect of layout change. In the special case where all departments are of equal area, there are multiple alternate paths of equal length and thus, both the FAPC and QAP layouts that allow alternative flow routing would result in similar performance in regards to workflow congestion. In the case where departments are of different area, the FAPC layout is superior to the QAP layout especially when there is intermediate flow intensity.

Before the future research directions are presented, a discussion of the strengths and limitations of our model might be helpful. One of the strengths is that the material handling time, which is a function of the distance of the chosen path and associated congestion level are calculated simultaneously when solution of layout and routing is searched. This allows the model to take congestion into consideration. The other strength is that the facilities do not need to be of equal area: smaller facilities in larger areas can have an extension aisle to connect to the network; and larger facilities can either be prohibited from occupying certain sites or they can introduce a local change in the aisle structure with accompanying alteration in link travel times. While, the objective of this paper is to explore the impact of congestion on layout and flow routing rather than study a space filling layout problem. The contribution of studying the impact of congestion on layout and flow routing continues. One way to relax the assumptions of modeling changed aisles is to let the aisle network be optimized concurrently with the layout and flow routing, which will be one of the future research directions.

One of our suggested directions for future research is to take aisle redesign into consideration. In other words, the aggregate effects of three factors: flow re-routing, layout change and aisle redesign are to be explored on worklflow congestion. Interesting comparisons can be conducted to investigate the aggregate effects of two of the factors against the three factors. Another suggested future research direction is to take into consideration both cost incurred due to relocation and the cost saved due reducing congestion. This should be considered over a certain period to assure that the costs of excess material handling due to congestion warrant a costly rearrangement of layout. Some of the developed results in dynamic layout research (Rosenblatt 1986, Kochhar and Heragu 1999, etc.) will be useful. In addition, radio frequency identification technology can be applied to forklifts so that a centralized traffic control system like in the AGV system can be implemented to globally manage traffic.

Acknowledgment

This work was supported by the National Science Foundation via grant number DMI-0300370.

References

Anstreicher, K., N. Brixius, J. Goux, and J. Linderoth (2002). Solving large quadratic assignment problems on computational grids. *Mathematical Programming 91*(3), 563–588.

Castillo, I. and B. Peters (2003). An extended distance-based facility layout problem. *International Journal of Production Research 41*(11), 2451–2479.

Castillo, I. and B. Peters (2004). Integrating design and production planning considerations in multi-bay manufacturing facility layout. *European Journal of Operational Research 157*(3), 671–687.

Chiang, W., P. Kouvelis, and T. Urban (2002). Incorporating workflow interference in facility layout design: the quartic assignment problem. *Management Science 48*(4), 584–590.

Chiang, W., P. Kouvelis, and T. Urban (2006). Single and multi-objective facility layout with workflow interference considerations. *European Journal of Operational Research 174*(3), 1414–1426.

Connolly, D. (1990). An improved annealing scheme for the QAP. *European Journal of Operational Research 46*, 93–100.

Gambardella, L., E. Taillard, and M. Dorigo (1999). Ant colonies for the quadratic assignment problem. *Journal of the Operational Research Society 50*(2), 167–176.

Ioannou, G. (2007). An integrated model and a decomposition-based approach for concurrent layout and material handling system design. *Computers and Industrial Engineering 52*(4), 459–485.

Kochhar, J. and S. Heragu (1999). Facility layout design in a changing environment. *International Journal of Production Research 37*(10), 2429–2446.

Kulturel-Konak, S., A. Smith, and B. Norman (2004). Layout optimization considering production uncertainty and routing flexibility. *International Journal of Production Research 42*(21), 4475–4493.

Montreuil, B. (1991). A modeling framework for integrating layout design and flow network design. In J. White and I. Pence (Eds.), *Progress in material handling and logistics*, Volume 2, pp. 95–115. Springer-Verlag.

Nagi, R. (2006). Material handling congestion and the facility layout problem. 9th International Material Handling Research Colloquium (IMHRC 2006), Salt Lake City, UH, pp. 216–221.

Norman, B., R. Arapoglu, and A. Smith (2001). Integrated facilities design using a contour distance metric. *IIE transactions 33*, 337–344.

Resende, M., P. Pardalos, and Y. Li (1996). Algorithm 754: Fortran subroutines for approximate solution of dense quadratic assignment problems using GRASP. *ACM Transactions on Mathematical Software 22*, 104–118.

Rosenblatt, M. (1986). The dynamics of plant layout. *Management Science 32*(1), 76–86.

Taillard, E. (1991). Robust tabu search for the quadratic assignment problem. *Parallel Computing 17*, 443–455.

Tate, D. and A. Smith (1995). A genetic approach to the quadratic assignment problem. *Computers and Operations Reseach 22*(1), 73–83.

Yang, T. and B. Peters (1997). Integrated facility layout and material handling system design in semiconductor fabrication facilities. *IEEE Transactions 10*(3), 360–369.

Zhang, M., R. Batta, and R. Nagi (2009). Modeling of congestion and optimization of flow routing in a manufacturing/warehouse facility. *Management Science 55*(2), 267–280.