Exact Solution to Bandwidth Packing Problem with Queuing Delays

Navneet Vidyarthi^{a,b,∗}, Sachin Jayaswal^c, Vikranth Babu Tirumala Chetty^c

^aDepartment of Supply Chain and Business Technology Management John Molson School of Business, Concordia University, Montreal, QC, H3G 1M8, Canada ^bInteruniversity Research Center on Enterprise Networks, Logistics, and Transportation (CIRRELT)

c Indian Institute of Management, Vastrapur, Ahmedabad, Gujarat, 380 015 India

Abstract

The bandwidth packing problem seeks to select and route a set of calls from a given list, each with a pre-specified requirement for bandwidth, on an undirected communication network such that the revenue generated is maximized. In this paper, we present a model and an exact solution approach for the bandwidth packing problem with queuing delay costs under stochastic demand and congestion. We provide a more general model than available in the extant literature by assuming a general service time distribution on the links. The problem, under Poison call arrivals, is thus set up as a network of spatially distributed independent $M/G/1$ queues. However, the presence of delay cost in the objective function makes the resulting integer programming model nonlinear. We present an exact solution approach based on piecewise linearization and cutting plane algorithm. Computational results indicate that the proposed solution method provides optimal solution in reasonable computational times. Comparisons of our exact solution method with the Lagrangean relaxation based solution reported in the literature for the special case of exponential service times clearly demonstrate that our solution approach outperforms the latter, both in terms of the quality of solution and computational times. Using numerical examples, we demonstrate that the service time variability, if not correctly represented in the model, can result in a solution very different from the optimal.

Keywords: Bandwidth Packing; Telecommunications; Integer Programming; Queuing Delay; Linearization; Exact Approach; Cutting Plane Method

[∗]Corresponding author, Phone: +001-514-848-2424x2990, Fax: +001-514-848-2824

Email addresses: navneetv@jmsb.concordia.ca (Navneet Vidyarthi), sachin@iimahd.ernet.in (Sachin Jayaswal), tvikranth@iimahd.ernet.in (Vikranth Babu Tirumala Chetty)

1. Introduction

Technological improvements in the telecommunications industry have led to a massive growth of services like video-conferencing, social networking, collaborative computing, etc. At the same time, the arrival of cheaper and smarter devices have resulted in demand for faster and better services from the providers. This has increased the pressure on telecommunication firms to efficiently manage their limited bandwidth to provide satisfactory end-user services. In this context, one of the fundamental problems that arises is the Bandwidth Packing Problem (BPP). The BPP can be stated as: given a set of calls, and their associated potential revenues and bandwidth requirements (demand), arising at an instant on a telecommunication network with limited bandwidth on its links, (i) decide which of these calls to accept/reject, and (ii) select a single path (sequence of links) to route each selected call, such that the total revenue generated from the accepted calls is maximized without violating the bandwidth capacities on the links (Cox et al., 1991).

Several variants of BPP have been studied in the literature. For example, Amiri and Barkhi (2000) present multi-hour BPP to account for the variation in traffic between peak and off-peak hours of the day. Another version of BPP that involves scheduling of the selected calls within given time windows is reported by Amiri (2005). Amiri and Barkhi (2012) present an extension of BPP that has applications in telecommunication services like video conferencing and collaborative computing. They consider a case wherein each request from users consists of a set of calls between various pairs of nodes, and a request cannot be partially accepted/rejected. Recently, Bose (2009) has studied another version of the problem wherein the calls belong to two priority classes: the calls belonging to the higher priority class are shorter in length and generate more revenue but consume more bandwidth compared to the calls belonging to the lower priority class.

Other extensions of BPP account for the delays arising as a result of calls waiting at nodes due to congestion on the links. Excessive delays may arise if the solution to BPP, or its variants, result in certain links getting utilized close to their bandwidth capacities. Explicit consideration of such delays in the modeling and solution of BPP is important to guarantee quality service to customers. Amiri et al. (1999), Rolland et al. (1999), and Han et al. (2012) explicitly account for such network delays due to congestion by incorporating queuing delay terms in their model. All of these papers model the links on the network as a network of independent M/M/1 queues with the implicit underlying assumption that call arrivals are Poisson and their service times on links have exponential distribution. Amiri et al. (1999) discourage such delays in their model by penalizing them in the objective function,

while Rolland et al. (1999) and Han et al. (2012) impose a constraint to limit such delays. Bose (2009) extends the problem to a setting where calls may be classified into different priority classes. For this, he models each link as a preemptive priority $M/M/1$ queue. Amiri (2003) extends the multi-hour BPP, earlier studied by Amiri and Barkhi (2000), with delay guarantees. The problem presented by Gavish and Hantler (1983) is also related to BPP with delays due to congestion, although the acceptance/rejection of calls is not a decision in their problem.

The single path requirement in BPP, which arises in various telecommunication services like video teleconferencing, etc., makes the problem NP-hard (Parker and Ryan, 1993). As such, various solution methods are presented in the literature. Anderson et al. (1993); Laguna and Glover (1993), for instance, use Tabu Search metaheuristic, while Cox et al. (1991) apply Genetic Algorithms. Lagrangean relaxation has been a popular choice of solution method in the literature, reported by Gavish and Hantler (1983), Rolland et al. (1999), Amiri et al. (1999), Amiri and Barkhi (2000), Amiri (2003), Amiri (2005), and Amiri and Barkhi (2012). Branch-and-Price and Column Generation is used by Parker and Ryan (1993), while Park et al. (1996) and Villa and Hoffman (2006) report the use of Branch-and-Price-and-Cut and Column Generation. Han et al. (2012) use Branch-and-Price technique with their Dantzig-Wolfe decomposition based reformulation of their model.

From the review of literature, we observe that all the studies on BPP that account for delays on telecommunication links due to congestion are based on the simplifying assumption that call arrivals are Poisson and service times on links have exponential distribution (Gavish and Hantler, 1983; Amiri et al., 1999; Rolland et al., 1999; Amiri, 2003; Bose, 2009; Han et al., 2012). This is primarily to make the problem, which is already otherwise NP-hard, tractable. The current study is an attempt to overcome this limitation in the extant literature by presenting a more generalized model. Through this work, we make the following contributions to the literature on BPP:

- 1. We present a generalized model for BPP with queuing delay costs, where the links in the network are modeled as independent $M/G/1$ queues.
- 2. Using simple transformation and piecewise linearization of queuing delay cost function, we linearize the model and present an efficient and exact approach based on cutting plane algorithm to solve the resulting model.

The remainder of the paper is organized as follows. In Section 2, we formally describe the problem and present its non-linear integer programming formulation. Section 3 describes an approach to linearize the model, followed by an exact solution methodology to solve the resulting mixed integer linear programming problem (MILP). Illustrative example, computational results, and insights are reported in Section 4. Section 5 concludes with some directions for future research.

2. Problem Formulation

We introduce the following notations used to describe the problem.

In line with the literature (Gavish and Hantler, 1983; Amiri et al., 1999; Rolland et al., 1999; Han et al., 2012), we assume that the arrivals of calls/messages on the network occur according to a Poisson process. Further, links are assumed to have finite capacities Q_{ij} for transmission of messages, and that nodes have unlimited buffers to store messages waiting for transmission. However, unlike the existing literature, we allow the message lengths (in bits) to follow a general distribution with a mean $1/\mu$, standard deviation σ , and coefficient of variation $cv = \mu \sigma$. The service rate (in bits per second) of the link (i, j) is proportional to the capacity of the link Q_{ij} . Then, the service time per message on link (i, j) also follows a general distribution with a mean $1/\mu Q_{ij}$, standard deviation σ/Q_{ij} , and coefficient of variation $cv = \mu \sigma$. Each link is thus modeled as a single server M/G/1 queue, and the telecommunication network is modeled as a network of independent $M/G/1$ queues.

Assume the bits composing message $m \in M$ arrive at a rate d^m per unit time. Further, let $X_{ij}^m(X_{ji}^m) = 1$ if call m is routed through link (i, j) in the direction from i to j $(j \text{ to } i)$, 0 otherwise. Then, the arrival of bits on link (i, j) , due to superposition of Poisson processes, follows a Poisson process with a rate $\sum_{m\in M} d^m (X_{ij}^m + X_{ji}^m)$ per unit time, and the arrival rate of messages per unit time on link (i, j) is $\lambda_{ij} = \mu \sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m)$. The average utilization of link (i, j) is given by:

$$
\rho_{ij} = \frac{\lambda_{ij}}{\mu Q_{ij}} = \frac{\sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m)}{Q_{ij}} \tag{1}
$$

Under steady state conditions (ρ_{ij} < 1) and first-come first-serve (FCFS) queuing discipline, the mean sojourn time (waiting time in queue + service time) of a message on link (i, j) , which is modeled as an $M/G/1$ queue, is given by the Pollaczek-Khintchine (PK) formula as: $E[w_{ij}] = \left(\frac{1+cv^2}{2}\right)$ $\left(\frac{cv^2}{2}\right)\frac{\lambda_{ij}}{\mu Q_{ij}(\mu Q_{ij}-\lambda_{ij})}+\frac{1}{\mu Q_{ij}}$ $\frac{1}{\mu Q_{ij}}$. The expected network delay can be estimated as the weighted average of the expected delays on links: $\frac{1}{\Lambda} \sum_{(i,j)\in E} \lambda_{ij} E[w_{ij}]$, resulting in the following:

$$
E[W] = \frac{1}{\Lambda} \sum_{(i,j)\in E} \left\{ \left(\frac{1+cv^2}{2} \right) \frac{(\lambda_{ij})^2}{\mu Q_{ij}(\mu Q_{ij} - \lambda_{ij})} + \frac{\lambda_{ij}}{\mu Q_{ij}} \right\},\tag{2}
$$

where $\Lambda = \mu \sum_{m \in M} d^m$ is the total arrival rate of messages in the network. Substituting $\lambda_{ij} = \mu \sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m)$, as defined above, this can be further expressed as:

$$
E[W] = \frac{1}{\Lambda} \sum_{(i,j)\in E} \left\{ \left(\frac{1+cv^2}{2} \right) \frac{(\sum_{m\in M} d^m (X_{ij}^m + X_{ji}^m))^2}{Q_{ij} (Q_{ij} - \sum_{m\in M} d^m (X_{ij}^m + X_{ji}^m))} + \frac{\sum_{m\in M} d^m (X_{ij}^m + X_{ji}^m)}{Q_{ij}} \right\}
$$
(3)

Using the above notations, the problem BPP under queuing delay that we study can be stated as follows: given a set of calls M, their associated potential revenues $(r^m, m \in M)$ and bandwidth requirements $(d^m, m \in M)$, arising at an instant on an undirected telecommunication network consisting of nodes N and links E with fixed arc/link capacities $(Q_{ij}, (i, j) \in$ E), determine a subset of calls $M' \subseteq M$ and a subset of $E' \subseteq E$ for each $m \in M'$, such that the total net revenue minus queuing delay costs is maximized. Let $Y^m = 1$ if call m is accepted, 0 otherwise, then the mathematical model for BPP with queuing delays can be stated as:

 $[PN]$:

$$
\max Z(\mathbf{X}, \mathbf{Y}) = \sum_{m \in M} r^m Y^m - C \sum_{(i,j) \in E} \left\{ \left(\frac{1 + cv^2}{2} \right) \frac{(\sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m))^2}{Q_{ij} (Q_{ij} - \sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m)} + \frac{\sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m)}{Q_{ij}} \right\}
$$
(4)

s.t.
$$
\sum_{j \in N} X_{ij}^m - \sum_{j \in N} X_{ji}^m = \begin{cases} Y^m & \text{if } i = O(m); \\ -Y^m & \text{if } i = D(m); \\ 0 & \text{otherwise} \end{cases} \forall (i, j) \in E, m \in M
$$
 (5)

$$
\sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m) \le Q_{ij} \qquad \forall (i, j) \in E
$$
\n(6)

$$
X_{ij}^{m} \in \{0, 1\} \qquad \forall (i, j) \in E, m \in M \tag{7}
$$

$$
Y^m \in \{0, 1\} \qquad \qquad \forall m \in M \tag{8}
$$

The first term in the objective function (4) is the total revenue from accepted calls. The second term captures the average queuing delay cost due to all accepted calls, where $C = c/\Lambda$ (a constant). Constraint set (5) are the flow conservation equations on each link for each call. Constraint set (6) ensures that the total demand on each link is less than its bandwidth capacity, required for the stability of the queue $(\lambda_{ij} \leq \mu Q_{ij})$. Constraint sets (7) and (8) are binary restrictions on the variables. For $cv = 1$, the above formulation reduces to the $M/M/1$ model studied by Amiri et al. (1999) and others.

The formulation $[PN]$ is a non-linear integer program. In the following section, we present an approach to transform the above model, using auxiliary variables, into an MILP, and a cutting plane based method to solve it.

3. Solution Methodology

After rearranging the terms in (2) , $E[W]$ can be rewritten as:

$$
E[W] = \frac{1}{\Lambda} \sum_{(i,j)\in E} \frac{1}{2} \left\{ (1+cv^2) \frac{\lambda_{ij}}{\mu Q_{ij} - \lambda_{ij}} + (1-cv^2) \frac{\lambda_{ij}}{\mu Q_{ij}} \right\}
$$

= $\frac{1}{\Lambda} \sum_{(i,j)\in E} \frac{1}{2} \left\{ (1+cv^2) \frac{\sum_{m\in M} d^m(X_{ij}^m + X_{ji}^m)}{Q_{ij} - \sum_{m\in M} d^m(X_{ij}^m + X_{ji}^m)} + (1-cv^2) \frac{\sum_{m\in M} d^m(X_{ij}^m + X_{ji}^m)}{Q_{ij}} \right\}$

We define non-negative auxiliary variables R_{ij} , such that:

$$
R_{ij} = \frac{\lambda_{ij}}{\mu Q_{ij} - \lambda_{ij}} = \frac{\sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m)}{Q_{ij} - \sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m)}
$$
(9)

Then,

$$
\sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m) = \frac{R_{ij}}{1 + R_{ij}} Q_{ij}
$$
\n(10)

Substituting (9) in the expression for $E[W]$ above gives:

$$
E[W] = \frac{1}{\Lambda} \sum_{(i,j)\in E} \frac{1}{2} \left\{ \left(1 + cv^2 \right) R_{ij} + \left(1 - cv^2 \right) \frac{\sum_{m\in M} d^m (X_{ij}^m + X_{ji}^m)}{Q_{ij}} \right\}
$$

We use the following lemma to linearize $[PN]$.

Lemma 1: The function $f(R_{ij}) = \frac{R_{ij}}{1+R_{ij}}$ is concave in $R_{ij} \in [0, \infty)$. Proof:

Differentiating the function w.r.t. R_{ij} , we get the first derivative $\frac{\delta f}{\delta R_{ij}} = \frac{1}{(1+R)^{2}}$ $\frac{1}{(1+R_{ij})^2} > 0$, and the second derivative $\frac{\delta^2 f}{\delta R^2}$ $\frac{\delta^2 f}{\delta R_{ij}^2} = \frac{-2}{(1+R_i)}$ $\frac{-2}{(1+R_{ij})^3}$ < 0, which proves that the function is concave in R_{ij} for $R_{ij} > 0$.

Lemma 1 implies that the function $f(R_{ij}) = \frac{R_{ij}}{1+R_{ij}}$ can be approximated by a large set of piecewise linear functions that are tangent to $f(R_{ij})$ at points $\{R_{ij}^h\}_{h\in H}$, such that:

$$
\frac{R_{ij}}{1 + R_{ij}} = \min_{h \in H} \left\{ \frac{1}{(1 + R_{ij}^h)^2} R_{ij} + \frac{(R_{ij}^h)^2}{(1 + R_{ij}^h)^2} \right\}
$$

This is equivalent to the following set of constraints:

$$
\frac{R_{ij}}{1 + R_{ij}} \le \frac{1}{(1 + R_{ij}^h)^2} R_{ij} + \frac{(R_{ij}^h)^2}{(1 + R_{ij}^h)^2}, \qquad \forall (i, j) \in Eh \in H
$$

Using (10) , the above set of constraints can be rewritten as:

$$
\sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m) - \frac{Q_{ij}}{(1 + R_{ij}^h)^2} R_{ij} \le \frac{Q_{ij} (R_{ij}^h)^2}{(1 + R_{ij}^h)^2} \qquad \forall (i, j) \in E, h \in H \tag{11}
$$

provided $\exists h \in H$ such that (11) holds with equality.

The above substitutions result in the following linear MIP model:

$$
[PL(H)] : \max_{m \in M} \sum_{m \in M} r^m Y^m - \frac{C}{2} \sum_{(i,j) \in E} \left\{ \left(1 + cv^2 \right) R_{ij} + \left(1 - cv^2 \right) \frac{\sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m)}{Q_{ij}} \right\} \tag{12}
$$

s.t. (5) – (8), (11)

$$
R_{ij} \ge 0 \qquad \forall (i, j) \in E \tag{13}
$$

For equivalence between $[PN]$ and $[PL(H)]$, there should exist at least one $h \in H$ such that (11) holds with equality. Proposition 1 confirms that there indeed exists one such $h \in H$ at optimality.

Proposition 1: At least one of the constraints (11) in $[PL(H)]$ will be binding at optimality. Proof:

After rearranging the terms, (11) can be rewritten as:

$$
R_{ij} \ge (1 + R_{ij}^h)^2 \frac{\sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m)}{Q_{ij}} - (R_{ij}^h)^2 \tag{14}
$$

Since R_{ij} appears in the objective function with a negative coefficient, $[PL(H)]$ attains its optimum value only when R_{ij} is minimized. This implies that $\forall (i, j) \in E$, $\exists h \in H$ such that (14) holds with equality if $(1 + R_{ij}^h)^2 \frac{\sum_{m \in M} d^m(X_{ij}^m + X_{ji}^m)}{Q_{ij}}$ $\frac{\Gamma(A_{ij} + A_{ji})}{Q_{ij}} - (R_{ij}^h)^2 \ge 0$, else $R_{ij} = 0$. Further,

$$
0 \le (1 + R_{ij}^h)^2 \frac{\sum_{m \in M} d^m (X_{ij}^m + X_{ji}^m)}{Q_{ij}} - (R_{ij}^h)^2
$$

= $(1 + R_{ij}^h)^2 \rho_{ij} - (R_{ij}^h)^2$ (using (1))
= $(\rho_{ij} - 1)(R_{ij}^h)^2 + 2\rho_{ij} R_{ij}^h + \rho_{ij}$
 $\Leftrightarrow R_{ij}^h \in \left[0, \frac{\rho_{ij} + \sqrt{\rho_{ij}}}{1 - \rho_{ij}}\right] \forall h \in H \text{ (since } \rho_{ij} \le 1 \text{ and } R_{ij} \ge 0 \text{ using (9))}$

Thus, to prove that $\exists h \in H$ such that (11) holds with equality, we need to show that $R_{ij}^h \in \left[0, \frac{\rho_{ij} + \sqrt{\rho_{ij}}}{1-\rho_{ij}}\right]$. Since R_{ij}^h is an approximation to R_{ij} , we obtain:

$$
0 \le R_{ij}^h \approx R_{ij} = \frac{\lambda_{ij}}{\mu Q_{ij} - \lambda_{ij}} \qquad \text{(using (9))}
$$

$$
= \frac{\rho_{ij}}{1 - \rho_{ij}}
$$

$$
\le \frac{\rho_{ij} + \sqrt{\rho_{ij}}}{1 - \rho_{ij}}
$$

This proves that $\forall (i, j) \in E$, $\exists h \in H$ such that, at optimality, (11) always holds with equality.

Proposition 2: For every subset of points ${R_{ij}^h}_{h \in H^q \subseteq H}$, $v(PL(H^q))$ is an upper bound to $[PL(H)]$, and hence to $[PN]$, where $v(\bullet)$ is the optimal objective function value of the problem (\bullet) .

Proof:

Suppose, at any iteration, we use a subset of tangent points $\{R_{ij}^h\}_{h\in H^q\subseteq H}$, and solve the corresponding problem $[PL(H^q)]$, which yields the solution $(\mathbf{X}^q, \mathbf{Y}^q, \mathbf{R}^q)$ with the objective function value $v(PL(H^q))$. Since $[PL(H^q)]$ is a relaxation of the full problem $[PL(H)]$, $v(PL(H^q)) \ge v(PL(H))$, and hence $v(PL(H^q))$ provides an upper bound, given by:

$$
UB = v(PL(H^q)) = \sum_{m \in M} r^m Y^{mq} - \frac{C}{2} \sum_{(i,j) \in E} \left\{ \left(1 + cv^2 \right) R_{ij}^q + \left(1 - cv^2 \right) \frac{\sum_{m \in M} d^m (X_{ij}^{mq} + X_{ji}^{mq})}{Q_{ij}} \right\}
$$
(15)

п

Proposition 3: For every subset of points ${R_{ij}^h}_{h \in H^q \subseteq H}$, the lower bound to $[PN]$ is given by:

$$
LB = Z(\mathbf{X}^{\mathbf{q}}, \mathbf{Y}^{\mathbf{q}}) = \sum_{m \in M} r^m Y^{mq} - C \sum_{(i,j) \in E} \left\{ \left(\frac{1 + cv^2}{2} \right) \frac{(\sum_{m \in M} d^m (X_{ij}^{mq} + X_{ji}^{mq}))^2}{Q_{ij} (Q_{ij} - \sum_{m \in M} d^m (X_{ij}^{mq} + X_{ji}^{mq}))} + \frac{\sum_{m \in M} d^m (X_{ij}^{mq} + X_{ji}^{mq})}{Q_{ij}} \right\}
$$
(16)

where $(\mathbf{X}^q, \mathbf{Y}^q, \mathbf{R}^q)$ is the optimal solution to $[PL(H^q)]$.

Proof:

For every subset of points $\{R_{ij}^h\}_{h\in H^q\subseteq H}$, the solution $(\mathbf{X}^q, \mathbf{Y}^q, \mathbf{R}^q)$ to $[PL(H^q)]$ is also a feasible solution to $[PN]$, and hence the objective function (4) evaluated at the solution $(\mathbf{X}^q, \mathbf{Y}^q, \mathbf{R}^q)$, which is given by (16), gives a lower bound to $[PN]$.

3.1. Solution Algorithm

The model $[PL(H)]$ consists of a large number of constraints (11). However, not all of them need to be generated a priori. The solution algorithm starts with an initial subset $H^1 \subset H$. H^1 may be empty. However, our preliminary computational experiments show that starting with a non-empty H^1 helps in faster convergence of the algorithm. The resulting $[PL(H¹)]$ is solved, giving a solution $({\bf X}¹, {\bf Y}¹, {\bf R}¹)$. The upper bound $(UB¹)$ and the lower

bound $(LB¹)$ are computed using (15) and (16) respectively. The better of the last and the new lower bounds is retained as the new LB^1 . If UB^1 equals LB^1 within some accepted tolerance (ϵ), then $({\bf X}^1, {\bf Y}^1)$ is an optimal solution to $[PN]$, and the algorithm terminates. Else, a new set of points $R_{ij}^{h_{new}}$ is generated using the current solution (X^1, Y^1, R^1) as follows: $R_{ij}^{h_{new}} = \frac{\sum_{m \in M} d^m (X_{ij}^{m1} + X_{ji}^{m1})}{Q_{ij} - \sum_{m \in M} d^m (X_{ij}^{m1} + X_{ij}^{m1})}$ $\frac{\sum_{m\in M} a_{i} (\lambda_{ij} + \lambda_{ji})}{Q_{ij} - \sum_{m\in M} d^{m}(X_{ij}^{m-1} + X_{ji}^{m-1})}$. New cuts of the form (11) are generated using these points, and added to $[PL(H^1)]$ to arrive at $[PL(H^2)]$. Next, $[PL(H^2)]$ is solved, giving a new solution (X^2, Y^2, R^2) and UB^2 . The new lower bound is obtained as the greater of LB^1 and $Z(\mathbf{X}^2, \mathbf{Y}^2)$. If UB^2 equals LB^2 within the set tolerance (ϵ) , then the algorithm terminates with (X^2, Y^2) as an optimal solution. Else, the process is repeated until UB^q equals LB^q within the set tolerance for some iteration q . The complete algorithm is outlined below:

Algorithm 1 Solution Algorithm for $[PL(H)]$

1: $q \leftarrow 1; UB^{q-1} \leftarrow +\infty; LB^{q-1} \leftarrow -\infty;$ 2: Choose an initial set of points $\{R_{ij}^h\}_{h\in H^q}$ to approximate the function $R_{ij}/(1 + R_{ij})$ $\forall (i, j) \in E$. 3: while $(UB^{q-1} - LB^{q-1})/UB^{q-1} > \epsilon$ do 4: Solve $[PL(H^q)]$ to obtain $(\mathbf{X}^q, \mathbf{Y}^q, \mathbf{R}^q)$. 5: Update the upper bound: $UB^q \leftarrow v(PL(H^q)).$ 6: Update the lower bound: $LB^{q} \leftarrow \max\{LB^{q-1}, Z(\mathbf{X}^q, \mathbf{Y}^q)\}.$ 7: Compute new points: $R_{ij}^{h_{new}} = \frac{\sum_{m \in M} d^m (X_{ij}^{mq} + X_{ji}^{mq})}{Q_{ij} - \sum_{m \in M} d^m (X_{ij}^{mq} + X_{ji}^m)}$ $\frac{\sum_{m\in M}a\cdot(\overline{X_{ij}}+\overline{X_{ji}})}{Q_{ij}-\sum_{m\in M}d^m(\overline{X_{ij}}+X_{ji}^{mq})}$ $\forall (i,j)\in E$ 8: $H^{q+1} \leftarrow H^q \cup \{h_{new}\}$ 9: $q \leftarrow q + 1$ 10: end while

Proposition 4: Algorithm 1 to solve $[PL(H)]$ terminates in a finite number of iterations. Proof:

Given that $X_{ij}^m \in \{0,1\}$ and $R_{ij} = \frac{\lambda_{ij}}{\mu Q_{ij}}$ $\frac{\lambda_{ij}}{\mu Q_{ij}-\lambda_{ij}} = \frac{\sum_{m\in M} d^m(X^m_{ij}+X^m_{ji})}{Q_{ij}-\sum_{m\in M} d^m(X^m_{ii}+X)}$ $\frac{\sum_{m\in M} u^{(1)}(X_{ij}+X_{ji})}{Q_{ij}-\sum_{m\in M} d^m(X_{ij}^{m}+X_{ji}^{m})}$, the number of values that R_{ij} can take is finite. Therefore, in order to prove that Algorithm 1 is finite, it is sufficient to prove that the generated values of R_{ij}^h are not repeated.

Consider an iteration q, where Algorithm 1 has not yet converged, that is, $UB^q > LB^q$. Further, suppose $(\mathbf{X}^q, \mathbf{Y}^q)$ is the solution to $[PL(H^q)]$. Then, the new points $R_{ij}^{h_{new}}$ generated at iteration q are given by:

$$
R_{ij}^{h_{new}} = \frac{\sum_{m \in M} d^m (X_{ij}^{mq} + X_{ji}^{mq})}{Q_{ij} - \sum_{m \in M} d^m (X_{ij}^{mq} + X_{ji}^{mq})} \qquad \forall (i, j) \in E
$$

Suppose the values of $R_{ij}^{h_{new}}$ were already generated in one of the earlier iterations

 $\forall (i, j) \in E$. Then:

$$
(11) \Rightarrow \frac{R_{ij}^{h_{new}}}{1 + R_{ij}^{h_{new}}} \le \frac{1}{1 + R_{ij}^{h_{new}}} R_{ij}^q + \left(\frac{R_{ij}^{h_{new}}}{1 + R_{ij}^{h_{new}}}\right)^2 \qquad \forall (i, j) \in E
$$

$$
\Rightarrow R_{ij}^{h_{new}} \le R_{ij}^q \qquad \forall (i, j) \in E
$$

We now have:

$$
UB^{q} = \sum_{m \in M} r^{m} Y^{mq} - \frac{C}{2} \sum_{(i,j) \in E} \left\{ (1 + cv^{2}) R_{ij}^{q} + (1 - cv^{2}) \frac{\sum_{m \in M} d^{m} (X_{ij}^{mq} + X_{ji}^{mq})}{Q_{ij}} \right\}
$$

\n
$$
\leq \sum_{m \in M} r^{m} Y^{mq} - \frac{C}{2} \sum_{(i,j) \in E} \left\{ (1 + cv^{2}) R_{ij}^{h_{new}} + (1 - cv^{2}) \frac{\sum_{m \in M} d^{m} (X_{ij}^{mq} + X_{ji}^{mq})}{Q_{ij}} \right\}
$$

\n
$$
= \sum_{m \in M} r^{m} Y^{mq} - \frac{C}{2} \sum_{(i,j) \in E} \left\{ (1 + cv^{2}) \frac{\sum_{m \in M} d^{m} (X_{ij}^{mq} + X_{ji}^{mq})}{Q_{ij} - \sum_{m \in M} d^{m} (X_{ij}^{mq} + X_{ji}^{mq})} + (1 - cv^{2}) \frac{\sum_{m \in M} d^{m} (X_{ij}^{mq} + X_{ji}^{mq})}{Q_{ij}} \right\}
$$

\n
$$
= \sum_{m \in M} r^{m} Y^{mq} - C \sum_{(i,j) \in E} \left\{ \left(\frac{1 + cv^{2}}{2} \right) \frac{(\sum_{m \in M} d^{m} (X_{ij}^{mq} + X_{ji}^{mq}))}{Q_{ij} - \sum_{m \in M} d^{m} (X_{ij}^{mq} + X_{ji}^{mq})} \right\}
$$

\n
$$
\leq max \left(LB^{q-1}, \sum_{m \in M} r^{m} Y^{mq} - C \sum_{(i,j) \in E} \left\{ \left(\frac{1 + cv^{2}}{2} \right) \frac{(\sum_{m \in M} d^{m} (X_{ij}^{mq} + X_{ji}^{mq}))}{Q_{ij} (Q_{ij} - \sum_{m \in M} d^{m} (X_{ij}^{mq} + X_{ji}^{mq})} \right\}
$$

\n
$$
= LB^{q}
$$

\n
$$
= LB^{q}
$$

This contradicts our initial assumption $UB^q > LB^q$. Therefore, at a given iteration, at least one of the values of R_{ij}^h generated is different from all the previously generated values. Furthermore, the number of values that R_{ij}^h can take is finite, and hence the algorithm terminates in a finite number of iterations.

■

4. Computational Study

We report our computational experience with the solution methodology described in Section 3. The exact solution algorithm is coded in Visual C++, while $[PL(H^q)]$ at every iteration q is solved using IBM ILOG CPLEX 12.4. The experiments are conducted on a machine with the following specifications: Intel Core i5-3230M, 2.60 GHz CPU; 4.00 GB RAM; Windows 64-bit Operating System. In Section 4.1, using an illustrative example (adopted from Laguna and Glover, 1993) with 10 nodes and 20 calls, we demonstrate the impact of variability in service times of the links on the optimal selection of calls and their routes in the network. The computational performance of proposed solution approach on networks with varying sizes are presented in Section 4.2.

4.1. Illustrative Example

Figure 1 shows the network topology for a problem instance with 10 nodes. The bandwidth capacities (Q_{ij}) of different links on the network are given in Table 1. The call table listing the bandwidth requirements (d^m) and potential revenues (r^m) for 20 calls is shown in Table 2. The optimal solution obtained using the method described in Section 3 is presented in Table 3, which displays the optimal routing (collection of links) for each call that is accepted, as well as the total gross revenue (GR) and the total delay cost (DC) , for different values of coefficient of variation cv and unit delay cost (C) .

Table 3 demonstrates that the value of cv plays an important role in the call selection. For example, for $C = 5$, Call-9 is *rejected* at $cv = 0.5$, but gets *accepted* at higher values of $cv = 1, 1.5, 2$. On the other hand, for $C = 5$, Call-11 is *accepted* at $cv = 0.5$, but gets rejected at higher values of $cv = 1, 1.5, 2$. Call-16 exhibits an even more interesting pattern: for $C = 15$, it is accepted at $cv = 0.5$; rejected at $cv = 1, 1.5$; and again accepted at $cv = 2$. However, for $C = 20$, Call-16 is accepted at $cv = 0.5$; rejected at $cv = 1$; accepted at $cv = 1.5$; and again rejected at $cv = 2$. Table 3 further suggests that cv also plays a vital role in the route selection for the selected calls. For example, for $C = 5$, Call-16 is routed using links $0 - 8$; 7 − 0; 8 − 4 at $cv = 0.5, 1, 2$. However, the same call is routed using links $0 - 8$; $2-0$; $7-2$; $8-4$ at $cv=2$. Similar observations can be made for $\{Call -12; C=15\}$ and ${Call-19;C=20}.$ These results demonstrate the fact that service time variability plays a vital role in the optimal call and route selections in BPP, which, in turn, effect the total net revenue. This example thus illustrates the importance of accurately modelling service time variability for BPP.

Table 1: Bandwidth Capacities (Q_{ij}) of Links (i, j) for the Illustrative Example

					i 0 0 0 0 0 1 2 4 5 5 6 7		
					<i>j</i> 1 2 7 8 9 3 7 8 7 8 7 8		
Q_{ij} 25 35 40 20 15 10 20 15 10 15 10 10							

Call $\,m$	Origin node O(m)	Lable for the mastrative Example Destination node D(m)	Call demand d^m	Revenue r^m
$\mathbf{1}$	$\overline{0}$	$\overline{2}$	10	420
$\sqrt{2}$	θ	7	7	380
$\boldsymbol{3}$	0	$\overline{5}$	6	400
$\sqrt{4}$	0	$\overline{4}$	$\boldsymbol{6}$	390
$\overline{5}$	1	6	$\bf 5$	500
6	1	$\overline{5}$	$\overline{5}$	490
7	1	4	7	400
8	$\overline{2}$	9	$\overline{2}$	150
9	$\overline{2}$	3	4	450
10	$\overline{2}$	$\overline{4}$	8	500
11	3	$\overline{5}$	$\boldsymbol{6}$	850
12	$\overline{5}$	$\overline{2}$	3	200
13	6	$\boldsymbol{9}$	$\bf 5$	370
14	7	$\mathbf{1}$	6	500
15	7	9	$\overline{5}$	340
16	7	$\overline{4}$	$\overline{2}$	120
17	8	1	6	460
18	8	$\overline{2}$	8	450
19	9	$\overline{5}$	$\overline{5}$	360
20	9	$\mathbf{1}$	$\overline{5}$	170

Table 2: Call Table for the Illustrative Example

	$C=5$			$C=10$			L_{r} $C=15$				$C=20$					
	\underline{cv}		\underline{cv}			\underline{cv}				\underline{cv}						
Call m	$0.5\,$	$\mathbf{1}$	$1.5\,$	$\overline{2}$	$0.5\,$	$\mathbf{1}$	$1.5\,$	$\overline{2}$	$0.5\,$	$\mathbf{1}$	$1.5\,$	$\overline{2}$	$0.5\,$	$\mathbf{1}$	$1.5\,$	$\overline{2}$
$\mathbf{1}$	$0 - 2$	$0 - 2$	$0 - 2$	$0 - 2$	$0 - 2$	$0 - 2$	$0 - 2$	$0 - 2$	$0 - 2$	$0-2$	$0-2$	$0 - 2$	$0 - 2$	$0 - 2$	$0 - 2$	$0-2$
$\,2$	$0 - 7$	$0 - 7$	$0 - 7$	$0 - 7$	$0 - 7$	$0 - 7$	$0 - 7$	$0 - 7$	$0 - 7$	$0 - 7$	$0 - 7$	$0 - 7$	$0 - 7$	$0 - 7$	$0 - 7$	$0 - 7$
$\sqrt{3}$	$0 - 7;$ $7 - 5$	$0 - 7$; $7 - 5$	$0 - 7;$ $7 - 5$	$0 - 7;$ $7 - 5$	$0 - 7;$ $7-5$	$0 - 7$; $7 - 5$	$0-8;$ $8-5$	$0 - 8;$ $8-5$	$0 - 7;$ $7 - 5$	$0-8;$ $8-5\,$	$0-8;$ $8-5$	$0 - 7$; $7 - 5$	$0 - 7$; $7 - 5$	$0-8;$ $8-5$	$0 - 7;$ $7 - 5$	$0 - 7;$ $7 - 5$
$\overline{4}$	$0 - 8;$ $8 - 4$	$0 - 8;$ $8 - 4$	$0 - 8;$ $8 - 4$	$0 - 8;$ $8 - 4$	$0 - 8;$ $8 - 4$	$0 - 8;$ $8 - 4$	$0-8;$ $8 - 4$	$0 - 8;$ $8 - 4$	$0 - 8;$ $8 - 4$	$0 - 8;$ $8 - 4$	$0 - 8;$ $8 - 4$	$0 - 8;$ $8 - 4$	$0 - 8;$ $8 - 4$	$0-8;$ $8 - 4$	$0 - 8;$ $8 - 4$	$0 - 8;$ $8 - 4$
$\bf 5$	$0-7;$ $1-0;$ $7 - 6$	$0 - 7$; $1-0;$ $7-6$	$0 - 7;$ $1-0;$ $7-6$	$0 - 7;$ $1-0;$ $7-6$	$0 - 7;$ $1-0;$ $7-6$	$0 - 7;$ $1-0;$ $7-6$	$0 - 7;$ $1-0;$ $7-6$	$0 - 7;$ $1-0;$ $7-6$	$0 - 7;$ $1-0;$ $7-6$	$0 - 7;$ $1-0;$ $7-6$	$0 - 7;$ $1-0;$ $7-6$	$0 - 7;$ $1-0;$ $7-6$	$0 - 7$; $1-0;$ $7-6$	$0-7;$ $1-0;$ $7-6$	$0-7;$ $1-0;$ $7-6$	$0 - 7;$ $1-0;$ 7-6
$\,6\,$	$0 - 8;$ $1-0;$ $8 - 5$	$0 - 8;$ $1-0;$ $8 - 5$	$0-8;$ $1-0;$ $8-5$	$0 - 8;$ $1-0;$ $8-5$	$0 - 8;$ $1-0;$ $8 - 5$	$0 - 8;$ $1-0;$ $8 - 5$	$0 - 7;$ $1-0;$ $7 - 5$	$0 - 7;$ $1-0;$ $7-5$	$0 - 8;$ $1-0;$ $8 - 5$	$0 - 7;$ $1-0;$ $7-5$	$0 - 7;$ $1-0;$ $7-5$	$0 - 8;$ $1-0;$ $8-5$	$0-8;$ $1-0;$ $8 - 5$	$0-8;$ $1-0;$ $8-5$	$0 - 8;$ $1-0;$ $8-5$	$0-8;$ $1-0;$ $8-5$
$\,7$				\overline{a}												
$\,8\,$	$0-9;$ $2 - 0$	$0-9;$ $2 - 0$	$0-9;$ $2-0$	$0-9;$ $2-0$	$0-9;$ $2 - 0$	$0-9;$ $2 - 0$	$0-9;$ $2-0$	$0-9;$ $2-0$	$0-9;$ $2-0$	$0-9,$ $2-0$	$0-9;$ $2-0$	$0-9;$ $2-0$	$0-9;$ $2 - 0$	$0-9;$ $2-0$	$0-9;$ $2-0$	$0-9;$ $2-0$
9	$\overline{}$	$0-1;$ $1-3;$ $2-0$	$0-1;$ $1-3;$ $2-0$	$0-1;$ $1-3;$ $2-0$	$0-1;$ $1-3;$ $2-0$	$0-1;$ $1-3;$ $2 - 0$	$0-1;$ $1-3;$ $2-0$	$0-1;$ $1-3;$ $2-0$	$0-1;$ $1-3;$ $2-0$	$0-1;$ $1-3;$ $2 - 0$	$0-1;$ $1-3;$ $2-0$	$0-1;$ $1-3;$ $2-0$	$0-1;$ $1-3;$ $2-0$	$0-1;$ $1-3;$ $2-0$	$0-1,$ $1-3;$ $2 - 0$	$0-1;$ $1-3;$ $2-0$
$10\,$																
11	$0-8;$ $1-0;$ $3-1;$ $8-5$															
$12\,$	$5 - 7;$ $7\hbox{-}2$	$5 - 7;$ $7 - 2$	$5 - 7;$ $7 - 2$	$5 - 7;$ $7 - 2$	$5 - 7;$ $7 - 2$	$5 - 7;$ $7 - 2$	$5 - 7;$ $7 - 2$	$5 - 7;$ $7 - 2$	$5 - 7;$ $7 - 2$	$5 - 7;$ $7\hbox{-}2$	$5 - 7;$ $7\hbox{-}2$	$0-2;$ $5 - 8;$ $8-0$	$5 - 7;$ $7 - 2$	$5 - 7;$ $7 - 2$	$0-2;$ $5 - 8;$ $8-0$	$0-2;$ $5 - 8;$ $8-0$
13																
14	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$	$0-1;$ $7 - 0$
15	$0-9;$ $7 - 0$	$0-9;$ $7 - 0$	$0-9;$ $7 - 0$	$0-9;$ $7-0$	$0-9;$ $7 - 0$	$0-9;$ $7 - 0$	$0-9;$ $7 - 0$	$0-9;$ $7-0$	$0-9;$ $7 - 0$	$0-9;$ $7 - 0$	$0-9;$ $7-0$	$0-9;$ $7-0$	$0-9;$ $7 - 0$	$0-9;$ $7-0$	$0-9;$ $7 - 0$	$0-9;$ $7-0$
16	$0-8;$ $7-0;$ $8 - 4$	$0-8;$ $7-0;$ $8 - 4$	$0-8;$ $7-0;$ $8 - 4$	$0 - 8;$ $2-0;$ $7-2;$ $8 - 4$	$0-8;$ $7-0;$ $8 - 4$	$0-8;$ $7-0;$ $8 - 4$		\overline{a}	$0-8;$ $7-0;$ $8 - 4$			$0-8;$ $2-0;$ $7-2;$ $8 - 4$	$0-8;$ $7-0;$ $8 - 4$		$0-8;$ $2-0;$ $7-2;$ $8 - 4$	
17				$\qquad \qquad -$								\overline{a}				
18	$7-2;$ $8 - 7$	$7-2;$ $8 - 7$	$7-2;$ $8 - 7$	$7-2;$ $8 - 7$	$7-2;$ $8 - 7$	$7-2;$ $8 - 7$	$7-2;$ $8 - 7$	$7-2;$ $8 - 7$	$7-2;$ $8 - 7$	$7-2,$ $8 - 7$	$7-2;$ $8 - 7$	$7-2;$ $8 - 7$	$7-2;$ $8 - 7$	$7-2;$ $8 - 7$	$7-2;$ $8 - 7$	$7-2;$ $8 - 7$
19	$\qquad \qquad -$	$0-8;$ $8-5;$ $9 - 0$	$0-8;$ $8-5;$ $9 - 0$	$0-8;$ $8-5;$ $9 - 0$	$0-8;$ $8-5;$ $9 - 0$	$0-8;$ $8-5;$ $9 - 0$	$0-8;$ $8-5;$ $9 - 0$	$0-8;$ $8-5;$ $9 - 0$	$0-8;$ $8-5;$ $9 - 0$	$0-8;$ $8-5;$ $9 - 0$	$0-8;$ $8-5;$ $9-0$	$\qquad \qquad -$	$0-8;$ $8-5;$ $9 - 0$	$0-7;$ $7-5;$ $9 - 0$		
$20\,$				\overline{a}				\overline{a}								
Gross Revenue Delay Cost	5013 132	4948 128	4848 183	4707 266	4868 190	4747 $256\,$	4573 $306\,$	4368 429	4727 $\bf 285$	4563 $328\,$	4344 459	4073 451	4585 380	4407 437	4118 442	3842 $537\,$

Table 3: Solution Obtained for the Illustrative Example

Figure 2 shows the effect of varying cv and C , over a wider range of values, on the total gross revenue, total delay cost and the total net revenue ($NR = GR - DC$) for the above illustrative example. The figures suggest that as cv or C increases, the total net revenue decreases. This is expected since a higher cv or C causes either (i) a higher congestion related cost if the set of accepted calls remains unchanged; or (ii) calls with lower potential revenue getting accepted if they are associated with lower bandwidth demands. In either case, the total net revenue is expected to decrease. Further, when a higher cv or C causes the former (i), then the total gross revenue is expected to remain unchanged. However, when it causes the latter (ii), then the total gross revenue is also expected to decrease with an increase in cv or C . Hence, the total gross revenue in Figure 2 either remains unchanged or decreases with an increase in cv or C . However, the change in the total delay cost, as cv or C increases, is non-monotonic. This, although appears counter-intuitive, can be explained as follows. When a higher cv or C does not cause any change to the set of accepted calls, then the delay cost is expected to increase. However, when a higher cv or C causes calls with lower bandwidth demands getting accepted, then the total delay cost is expected to decrease due to a decrease in congestion in the network.

4.2. Computational Results

For our computational study, we adopt the data generation scheme as reported by Amiri et al. (1999) to generate 10 sets of networks for each value of $|N| = \{10, 20, 30, 40, 50\}$. For each of these networks, a call table is generated for $P = \{50, 60, 70, 80, 90\}$, where P is the percentage of the maximum possible types of calls (a call type is specified by an origindestination node pair) for the given network that are included in the call table. Thus, we have $10 \times 5 \times 5 = 250$ different problem sets. Each of these sets is solved for 4 different values of cv $(cv = 0, 0.5, 1, 1.5)$ and for 5 different values of C $(C = 0.5, 1, 5, 10, 15, 20)$, which together result in $250 \times 4 \times 5 = 5000$ problem instances. For each of the test instances, we start with a priori set $(H¹)$ of points to approximate the function $f(R_{ij}) = R_{ij}/(1 + R_{ij})$ using its tangents $\widehat{f}(R_{ij})$ at these points. These points are generated such that the approximation error measured by $\widehat{f}(R_{ij}) - f(R_{ij})$ is at most 0.001 (Elhedhli, 2005). Our initial computational experiments reveal that starting with an a priori set of points $(H¹)$ significantly improves the performance of the solution algorithm as it then requires fewer iterations/cuts and hence smaller CPU time for the algorithm to converge. The value of ϵ used in the convergence criterion is set at 10^{-6} in all the experiments.

The results of the computational experiments, which are averages over 10 different networks, are presented for each combination of $|N|$, P, C and cv in Table 4. The table reports

Figure 2: Gross Revenue (GR), Delay Cost (DC) and Net Revenue (NR) versus Coefficient of Variation of Service Times (cv) and Unit Delay Cost (C) for the Illustrative Example

the total gross revenue (GR) , delay cost (DC) expressed as percentage of the total gross revenue, CPU time (in seconds), and the minimum, maximum, and the average link utilizations. The results clearly demonstrate the stability and the efficiency of our proposed exact solution method over a wide range of problem instances: it succeeds in finding optimal solutions to several instances with different unit delay costs and service time variability within a couple of minutes, with the maximum CPU time being 2153 seconds (for $|N| = 50$; $P =$ 90; $C = 0.5$; $cv = 1$).

The efficiency of our solution algorithm is best highlighted by comparing its results, both the optimal objective function values and CPU times, with those from the Lagrangean relaxation based solution method reported by Amiri et al. (1999) for the special case of $cv = 1$ $(M/M/1)$ queue model for the links). For the completeness of the paper, the mathematical model and the Lagrangean relaxation based solution algorithm reported by Amiri et al. (1999) are briefly presented in the appendix. The comparison of the the results are presented in Table 5, which demonstrates that our proposed solution method is, on an average, 3 to 10 times faster than the Lagrangean relaxation approach. Moreover, our proposed method solves the problem to optimality whereas the Lagrangean relaxation leaves an optimality gap of 2 to 7% on an average, and 11% in the worst case.

Computation Time in seconds for the Proposed Exact Method

5. Conclusion

In this paper, we presented a model to analyze the impact of service time variability on the optimal call selection and routings in communication networks, commonly known as the bandwidth packing problem. We formulated a more generalized model of BPP with queuing delays by modelling the links, which process the calls arriving on the network, as $M/G/1$ queues. We presented a non-linear integer programming model, and linearized it using simple transformation and piecewise linear approximation. We further proposed an efficient solution approach, based on the cutting plane method, to solve the resulting linearized model to optimality. Through a computational study, we demonstrate the efficiency and the stability of the proposed solution algorithm in solving within minutes problem instances of the size of 50 nodes with varying service time variability delay costs. The proposed method also outperforms the Lagrangean relaxation approach, reported in the literature for the special case when services times on links are exponentially distributed.

The work reported in this paper can be extended in several ways. One such extension is to model the links as $GI/G/1$ queues, although the solution method for it is not immediately obvious. Another possible extension is to consider giving different priorities to calls from different classes of customers.

APPENDIX

We briefly present the mathematical model and the Lagrangian relaxation based solution approach reported by Amiri et al. (1999) for the special case when $cv = 1$ such that the links in the network are modeled as $M/M/1$ queues. For this, we introduce an additional set of variables W_{ij}^m as defined below:

$$
W_{ij}^m = \begin{cases} 1 & \text{if call } m \text{ uses link}(i,j) \text{in either direction;} \\ 0 & \text{otherwise.} \end{cases}
$$

The non-linear integer programming model of this problem is :

$$
[P_{M/M/1}] : \max \sum_{m \in M} r^m Y^m - C \sum_{(i,j) \in E} \frac{\sum_{m \in M} d^m W_{ij}^m}{Q_{ij} - \sum_{m \in M} d^m W_{ij}^m}
$$
(17)

$$
\text{s.t. } X_{ij}^m + X_{ji}^m \le W_{ij}^m \qquad \forall (i, j) \in E, m \in M \tag{18}
$$

$$
\sum_{m \in M} d^m W_{ij}^m \le Q_{ij} \qquad \forall (i, j) \in E \tag{19}
$$

$$
W_{ij}^{m}, \in \{0, 1\} \qquad \forall (i, j) \in E, m \in M \qquad (20)
$$

(5), (7), (8)

On dualizing the constraint set (18) using non-negative lagrangean multipliers $\alpha_{ij}^m \forall (i, j) \in E$ and $m \in M$, the problem $[P_{M/M/1}]$ decomposes into two sets of subproblems: (i) $[L1_{LR}^{m}]$ $\forall m \in M$; and (ii) $[L2_{LR}^E]$ $\forall (i, j) \in E$, as given below:

$$
[L1_{LR}^{m}] : \max r^{m}Y^{m} - \sum_{(i,j)\in E} \alpha_{ij}^{m}(X_{ij}^{m} + X_{ij}^{m})
$$
\n
$$
s.t. (5), (7), (8)
$$
\n
$$
[L2_{LR}^{E}] : \max \sum_{m\in M} \alpha_{ij}^{m}W_{ij}^{m} - C \frac{\sum_{m\in M} d^{m}W_{ij}^{m}}{Q_{ij} - \sum_{m\in M} d^{m}W_{ij}^{m}}
$$
\n
$$
s.t (19), (20)
$$
\n(22)

The solution algorithms to solve $[L1^m_{LR}]$, LP relaxation of $[L2^E_{LR}]$ and to generate feasible solutions are presented below:

The pseudocode to solve the BPP using Lagrangian Relaxation method is outlined below:

Algorithm 2 Solution Algorithm for $[L1^m_{LR}]$

1: Solve $[L1^m_{LR}]$ a shortest path problem with α^m_{ij} as the link costs 2: if $(r^m > \sum_{(i,j) \in E} \alpha^m_{ij} (X^m_{ij} + X^m_{ij})$ then 3: $(Y^m = 1)$ 4: else 5: $Y^m = 0$ and $X_{ij}^m = 0 \ \forall (i, j) \in E$ 6: end if

Algorithm 3 Solution Algorithm for LP Relaxation of $[L2_{LR}^E]$ for link (i, j)

1: Sort the calls $(m \in M)$ in non-increasing order of α_{ij}^m/d^m . Use index m' to represent the calls in this order.

$$
2\colon\thinspace m'\leftarrow 0
$$

3: while
$$
m' < |M|
$$
 do\n4: $m' \leftarrow m' + 1$ \n5: $S \leftarrow \sum_{k < m'} d^k W_{ij}^k$ \n6: $W_0 \leftarrow \min \left\{ 1, \frac{1}{d^{m'}} \left[(Q_{ij} - S) - \left(\frac{Cd^{m'} Q_{ij}}{\alpha_{ij}^{m'}} \right)^{1/2} \right] \right\}$ \n7: if $\alpha_{ij}^{m'} > 0$ and $W_0 > 0$ then\n8: $W_{ij}^{m'} \leftarrow W_0$ \n9: else\n10: $W_{ij}^{m'} \leftarrow 0$ \n11: end if\n12: if $W_{ij}^{m'} < 1$ then\n13: $W_{ij}^{m'} \leftarrow 0 \ \forall \{k : m' < k \le |M|\}$, and stop.\n14: end if\n15: end while

Algorithm 4 Solution Algorithm for Generating a Feasible Solution

1: $A_{ij} \leftarrow Q_{ij} \ \forall (i, j) \in E$

2: $DC \leftarrow 0$; ∆ ← 0

- 3: Sort the calls $(m \in M)$ in non-increasing order of $v(L1_{LR}^m)$ obtained from Algorithm 2. Use m' to represent the calls in this order.
- 4: Get the the values of $X_{ij}^{m'}$ $\forall m' \in M$, $(i, j) \in E$ obtained using Algorithm 2
- 5: $m' \leftarrow 0$
- 6: while $m' < |M|$ do 7: $m' \leftarrow m' + 1$ 8: if $d^{m'}(X_{ij}^{m'}+X_{ji}^{m'}) < A_{ij} \ \forall (i,j) \in E$ then 9: $\Delta \leftarrow C \sum_{(i,j)\in E}$ $\sum_{k' \leq m'} d^{k'} (X_{ij}^{k'} + X_{ji}^{k'})$ $\frac{c_1 k' \leq m'^{(a)} (X_{ij} + X_{ji}^k)}{Q_{ij} - d^{k'} (X_{ij}^{k'} + X_{ji}^{k'})} - DC$ 10: if $r^{m'} > \Delta$ then $11:$ $Y^{m'} \leftarrow 1$ 12: $A_{ij} \leftarrow A_{ij} - d^{m'}(X_{ij}^{m'} + X_{ji}^{m'}) \ \forall (i, j) \in E$ 13: $DC \leftarrow DC + \Delta$ 14: else $15:$ $y^{m'} \leftarrow 0$ and $X_{ij}^{m'} \leftarrow 0 \ \forall (i, j) \in E$ 16: end if 17: else $18:$ $y^{m'} \leftarrow 0$ and $X_{ij}^{m'} \leftarrow 0 \ \forall (i, j) \in E$ 19: end if

20: end while

Algorithm 5 Lagrangean Relaxation Based Solution Method

- 1: $\alpha_{ij}^m \leftarrow 0 \ \forall (i,j) \in E$ and $m \in M$; $UB \leftarrow +\infty$; $LB \leftarrow -\infty$; iter $\leftarrow 1$; max iter \leftarrow $500;\epsilon \leftarrow 10^{-6}$
- 2: while $(UB LB)/LB > \epsilon$ AND iter < max iter do
- 3: Solve $L1_{LR}^m \forall m \in M$ using Algorithm 2.
- 4: Solve $L2_{LR}^E \ \forall (i, j) \in E$ using Algorithm 3.
- 5: $UB \leftarrow \sum_{m \in M} v(L1_{LR}) + \sum_{(i,j) \in E} v(L2_{LR})$
- 6: Generate a feasible solution using Algorithm 4

$$
7: \quad LB \leftarrow v(P_{M/M/1})
$$

- 8: Update α_{ij}^{m} using sub-gradient method.
- 9: $iter \leftarrow iter + 1$

10: end while

Acknowledgements

This research was supported by the National Science and Engineering Research Council of Canada (NSERC) grant to the first author, and by the Research & Publication Grant, Indian Institute of Management Ahmedabad, to the second author.

References

- Amiri, A., 2003. The multi-hour bandwidth packing problem with response time guarantees. Information Technology and Management 4, 113–127.
- Amiri, A., 2005. The selection and scheduling of telecommunication calls with time windows. European Journal of Operational Research 167, 243–256.
- Amiri, A., Barkhi, R., 2000. The multi-hour bandwidth packing problem. Computer and Operations Research 27, 1–14.
- Amiri, A., Barkhi, R., 2012. The combinatorial bandwidth packing problem. European Journal of Operations Research 208, 37–45.
- Amiri, A., Rolland, E., Barkhi, R., 1999. Bandwidth packing with queuing delay costs: Bounding and heuristic procedures. European Journal of Operational Research 112, 635– 645.
- Anderson, C., Fraughnaugh, K., Parkner, M., Ryan, J., 1993. Path assignment for call routing: An application of tabu search. Annnals of Operations Research 41, 301–312.
- Bose, I., 2009. Bandwidth packing with priority classes. European Journal of Operational Research 192, 313–325.
- Cox, L., Davis, L., Qui, Y., 1991. Dynamic anticipatory routing in circuit-switched telecommunications networks, in: Davis, L. (Ed.), Handbook of Genetic Algorithms, Van Norstrand/Reinhold, New York. pp. 229–340.
- Elhedhli, S., 2005. Exact solution of a class of nonlinear knapsack problems. Operations Research Letters 33, 615–624.
- Gavish, B., Hantler, S., 1983. An algorithm for optimal route selection in sna networks. IEEE Transactions on Communications 31, 1154–1161.
- Han, J., Lee, K., Lee, C., Park, S., 2012. Exact algorithms for a bandwidth packing problem with queueing delay guarantees. INFORMS Journal on Computing doi:10.1287/ijoc.1120.0523.
- Laguna, M., Glover, F., 1993. Bandwidth packing: A tabu search approach. Management Science 39, 492–500.
- Park, K., Kang, S., Park, S., 1996. An integer programming approach to the bandwidth packing problem. Management Science 42, 1277–1291.
- Parker, M., Ryan, J., 1993. A column generation algorithm for bandwidth packing. Telecommunication Systems 2, 185–195.
- Rolland, E., Amiri, A., Barkhi, R., 1999. Queueing delay guarantees in bandwidth packing. Computers and Operations Research 26, 921–935.

Villa, C., Hoffman, K., 2006. A column-generation and branch-and-cut approach to the bandwidth-packing problem. Journal of Research of the National Institute of Standards and Technology 111, 161–185.