

Enhancing Integrated Communication Network Service Using a Queueing Model

Mohammad Sedaghat, Fairleigh Dickinson University, NJ, USA
Gary Bronson, Fairleigh Dickinson University, NJ, USA

ABSTRACT

This paper describes a strategy for efficiently constructing the demand servicing process of an Integrated Communication Network (ICN). Performance analysis, service problems, and relief action of ICNs are provided. End-to-end statistical performance parameters are first used to measure network compliance over a given fixed period. If any of the performance objectives are not satisfied, a servicing function determines the corrective action required to maintain service quality. The advantage of this network model is its efficiency and flexibility in handling a variety of services and applications. Enhanced network service and ICN traffic problems are solved using adaptive queueing models.

Keywords: Integrated Communication Network; Service System; Queueing System; Networking

INTRODUCTION

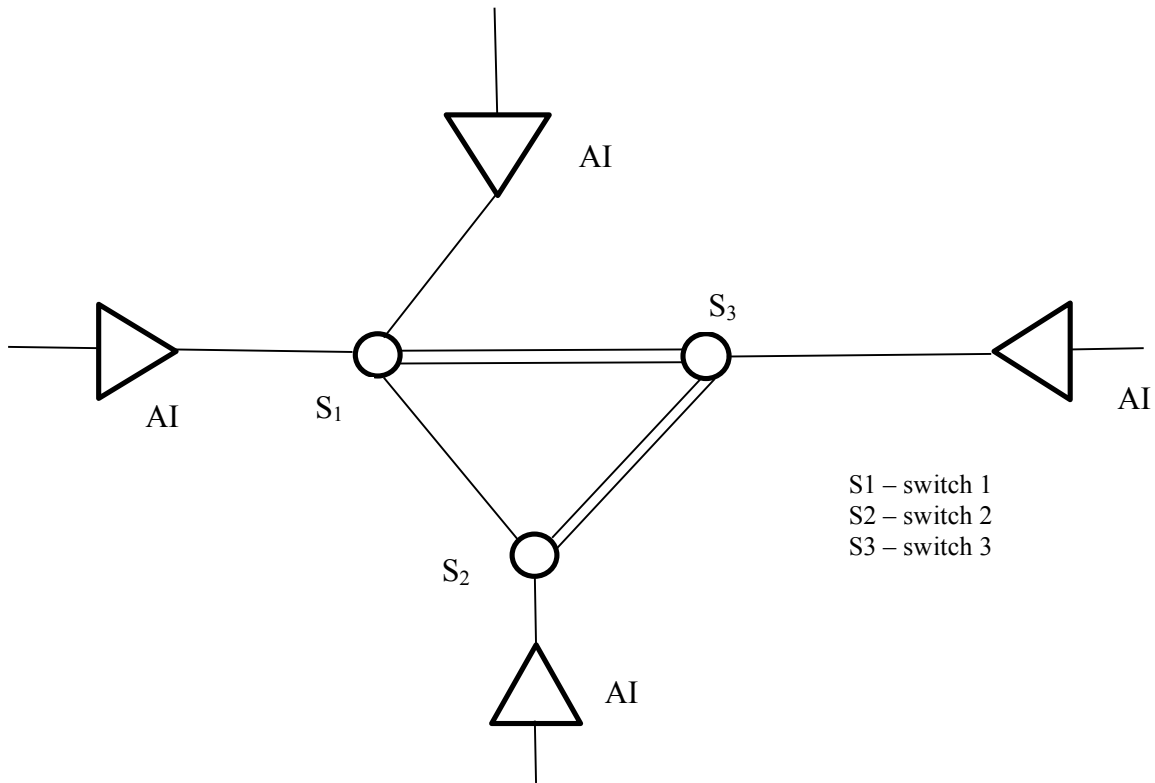
An Integrated Communication Network (ICN) carries both voice and data traffic in a packet format on a common set of facilities and switches. Figure 1 depicts a schematic diagram of an ICN. The network consists of two categories of components: the Access Interface (AI) components and the Packet Cross Connect System (PCCS). The AI provides access points to the network and serves as the system's packetizer and its depacketizer. The PCCS consists of the system nodes; as such, it provides the switching elements and serves as the cross-connect points for linking traffic moving through the system.

Designing, deploying, and operating an ICN service based system requires a series of centralized and interrelated network administration functions. These functions include forecasting and planning, demand servicing, and congestion management. Forecasting and planning functions provide the long-term strategy for network evolution to meet future traffic demands. Demand servicing monitors the network's performance, determining corrective actions for maintaining service quality. This real-time congestion and control management ensures maximum network operating efficiency during congestion periods. These congestion periods result from over loads and failure and are similar to the network flows described by Ahuja, Magnanti and Orlin (1993), Todinov (2013), and Nicholson and Zhang (2016).

Collection of raw measurements for demand servicing occurs at the Access Interfaces (AIs) (as noted previously, these are the network's access points). The Packet Cross Connect System (PCCS) is the switching element, which, as a telecommunications packet-switching network, has been considered by Dumas and Schwartz (2009) and Hochfelder and Misha (1999).

In the general case, for both localized and global servicing, the service problem's severity determines the applicable control action. Initialization of global controls occurs whenever localized control is not effective, or there is high network traffic patterns. The subject of this paper are these global controls, involving routing changes and network resizing. Based on a queueing network model (Warland, 1998) and (Brooks Kar & Mendonca, 2016), an ICN model providing efficient service processing is constructed.

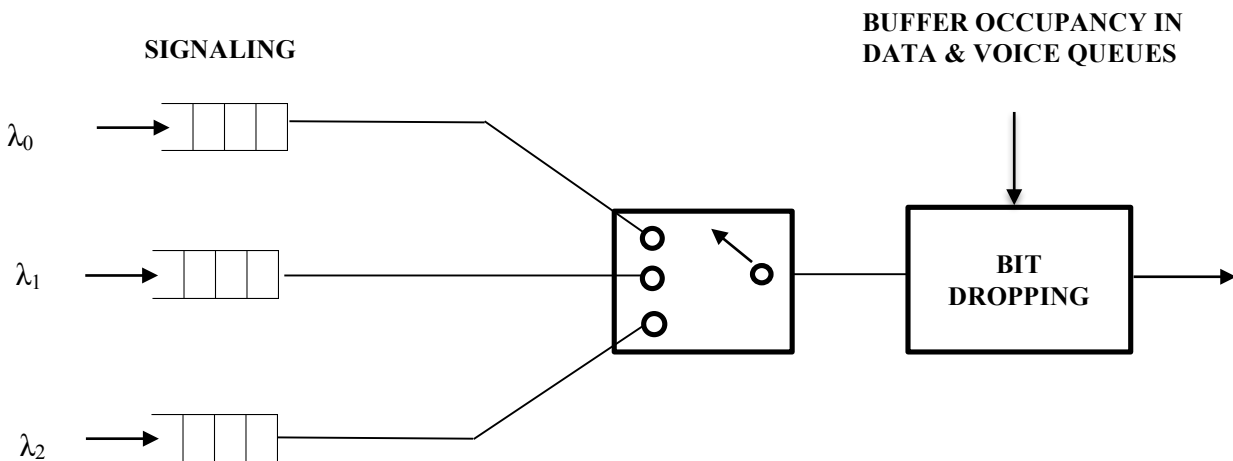
Figure 1. A Model of an ICN



ICN QUEUING MODEL

Using a queuing model, Figure 2 shows three (3) queues for services by each node. Here, queue 0 handles the signaling and control packets, queue 1 handles voice packets, and queue 2 handles data packets. As the signaling and control packet queue, queue 0 has non-preemptory priority over the dynamically served queues 1 and 2.

Figure 2. Multiplexer Service Scheme at each Switch



A routing algorithm (Dib, Manier, Moalic & Caminada, 2017) is used to establish the appropriate path along a network; however, any single node can be routed by no more than one High-Speed Link (HSL) group to any other given node. It is the information in the packet header, which includes priority and first intermediate node field that guides the packet through the network. Here, the network routing administrator loads it into the nodes as part of the provisioning process. For routing purposes, we now consider the grade of service criteria for the voice and data traffic. In an ICN, failure to meet a grade of service criteria reveals itself in end-to-end packet degradation (speed or quality), which indicates that servicing action must be initiated. Thus, the first step in this servicing action is to locate the “problem” path. Once this is located, the “problem” link within this path is then determined.

LOCATING THE PROBLEM LINK WITHIN THE PROBLEM PATH

A *problem path* is one in which end-to-end performance for either or both voice and data do not meet their specified criteria. Location of the *problem link* in this path can be determined using the following steps:

Step 1: Calculate the component objectives of voice and data for each link along that path using the following formula.

$$Q_{dol} = \frac{D_o - P_d - S}{N} \quad (\text{Eq.1})$$

where

- Q_{dol} = Queuing delay objective of link,
- D_o = End-to-end delay objective,
- P_d = Propagation delay,
- S = Switching delay,
- N = Number of nodes in the path

Note that a queuing delay through the problem path can occur at N places; confinement to a single node within the problem path need not occur.

Step 2: Compare the queuing delay for data and voice for each problem path link with their component objectives, as provided in Eq. 1. Using an M/M/1/K queuing system, the delay behavior can be approximated (Gross & Harris, 1998). A determination is then made of the link or links that do not meet their objective(s) by comparing the current performance of each link with their component objectives. These are the problem links.

CORRECTIVE ACTION

Corrective action involves routing changes and possibly network resizing. If routing changes are insufficient to solve all of the performance problems, then a new determination is made of the required number of links for each group to meet all of the relevant performance objectives for voice and data. Formulating this as a nonlinear optimization problem, the required network routing and sizing to clear the path degradation can be determined. Due to the complexity of this approach, we have developed a simpler algorithm to solve the service problems. To develop this algorithm, a spare bandwidth calculation and path selection methodology is first required.

SPARE BANDWIDTH CALCULATION

Calculating a system’s spare bandwidth requires determining the traffic through a high-speed link (HSL) for a time-constant of one hour, over a fixed period. Letting λ_b represent the 95th percentile hourly traffic, in bits per second, going through an HSL designated as b , then the spare bandwidth for this HSL, denoted as BW_b , is simply the difference between the provisioned bandwidth for the link and λ_b . Letting PBW_b denote the provisioned bandwidth of the HSL, in bits per second, the relationship between these variables is:

$$BW_b = PBW_b - \lambda_b \quad (\text{Eq.2})$$

PATH SELECTION METHODOLOGY

The path selection algorithm (Ahuja, et al. 1993) and (Panaousis, et al. 2017), determines the path that achieves near-optimum resource utilization for each path from its source to its destination. A path is the set of links and nodes traversed through the network. The objective is to find the candidate link or path that has enough spare bandwidth to support the total bandwidth requirement of rerouted traffic. There are two components in this path selection:

- Determination of the set of nodes to be traversed in the network
- Selection of an HSL in each of the HSL groups that make up the path (an HSL group consists of one or more HSLs between two adjacent nodes).

In this path selection, the goal is to distribute traffic load, in terms of bandwidth, evenly among all HSLs in an HSL group. This path selection depends on several factors, including:

- The bandwidth requirement of the path
- The average spare bandwidth available (per HSL) of each HSL group in each candidate path,
- The spare bandwidth available on each HSL in each candidate path,
- The length of each HSL in each candidate path, and
- The number of nodes traversed.

The first procedure for finding the required path is to identify a set of candidate paths between the two end nodes. The next step is to find links in the same HSL group: When traffic rerouting on one of these links is not possible, investigation of a direct path is undertaken. Should rerouting on a direct path not be possible, a search of two-link candidate paths between end nodes commences. Examination of three-link candidate paths begins when no two-link path are found. Finally, additional trunks are required when there are no three-link paths available. For paths involving more than two nodes, the following steps will determine a path.

Step 1: A selected path has the least cost metric for HSL groups (length divided by spare bandwidth). Among the qualified paths, the selected path is the one having the least cost in terms of HSL groups.

Step 2: Individuals HSLs from the HSL groups of the selected path are determined on a cost metric for the HSLs. Examination of other qualified paths begins only if there is not enough spare bandwidth on any HSL group in the selected path. This examination continue until a path on every group with spare bandwidth is found; otherwise, trunks are added.

THE SERVICE ALGORITHM

Having determined the required spare bandwidth and a path that can provide this bandwidth, the derivation of an algorithm for solving the service problem follows:

Letting

- ρ_1 = utilization factor of voice transmission,
- ρ_2 = utilization factor of data transmission,

we allow a link utilization factor of 90%, that is

$$\rho_1 + \rho_2 = .90$$

Keeping total link occupancy at 90% ensures the delay for voice and the mean data packet delay will satisfy their component objectives, as calculated from Eq. 1. If the utilization factor for the problem link is greater than one, then the only solution is to reroute traffic away from the problem link. Resolving the over-one utilization problem requires moving loads from the problem link.

DETERMINING LOADS TO BE REMOVED FROM THE PROBLEM LINK

The determination of problem link loads for removal uses the following notation:

λ_v^* and λ_d^* = the amounts of voice and data traffic, respectively, that can be carried on a problem link while meeting voice and data performance objectives.
 λ_v and λ_d = the amounts of voice and data traffic, respectively, that can be carried on a problem link while meeting voice and data performance objectives.

Thus, the overload for data is

$$\lambda_d - \lambda_d^*$$

and the overload for voice is

$$\lambda_v - \lambda_v^*$$

By keeping the total link utilization at 90% ($\rho_1 + \rho_2 = 90\%$), the delay for voice and data will meet their component objective, with a safety factor of 10%. When the actual utilization of link exceeds 90%, that is, ($\rho_1 + \rho_2 = a$), then the overload for voice becomes

$$p_1/a * (a - .90) = p_1 \left[\frac{a - .90}{a} \right],$$

and the overload for data packets becomes

$$p_2/a * (a - .90) = p_2 \left[\frac{a - .90}{a} \right],$$

both of which are proportional to their current utilizations, ρ_1 and ρ_2 , respectively.

The first choice to reduce these overloads is to reroute the extra loads from the problem link to a target path that can accommodate the overload.

DETERMINING A TARGET PATH

To reroute traffic, the HSL group that directly connects two nodes, designated here as node A and node Z is first displayed, if such a group is available. Where there is sufficient spare bandwidth on one or more HSLs in the group, chose the direct HSL group having the highest spare bandwidth. Otherwise, make successive path locations, either locating a suitable target path or determining that no such links are available. These determinations follow, in the order indicated:

1. Same HSL group selection

Reroute the overload from the problem link to another link in the same HSL group, if the candidate link has enough spare bandwidth. Then the amount of load that the candidate link can support is equal to the spare bandwidth of the link dived by the average bandwidth required by a path.

2. Direct Path Selection

To reroute traffic, the HSL group that directly connects the starting and ending nodes of the problem path is displayed, if such a group is available. If there is sufficient spare bandwidth on one or more HSLs in the group, choose the direct HSL group with the HSL that has the highest spare bandwidth in the selected.

3. Two-Link Path Selection (3 nodes)

Choosing from a set of three-node paths in which there is sufficient spare bandwidth, requires the following two steps:

Step 1: First, the nodes of a path are determined based on the cost metric c_g for HSL group g , which includes two parameters: the average spare bandwidth of an HSL in an HSL group that has spare bandwidths, and the distance between two nodes. Letting the cost metric in path g be denoted by c_g , we have.

$$c_g = \frac{d_g}{BW_{avg}^g} \quad (\text{Eq.3})$$

where

d_g = the length of the HSL group,
 BW_{avg}^g = the average spared bandwidth of the HSL group

As shown in Eq.3, the cost metric, c_g , is proportional to the length of the HSL group and inversely proportional to the available average spare bandwidth available in the group. Note that d_g for HSL group g is the same as d_1 for HSL 1 in HSL group g . c_g reflects the undesirability of using HSL group g as part of the path for the traffic. Thus, the lower the value of c_g (which may result from short distance and/or high average spare bandwidth), the more desirable group g is within its HSL group.

The total cost metric for a path, C_p , is simply the summation of the costs of its constituent HSL groups, c_g ,

$$C_p = \sum_{g \in p} c_g \quad (\text{Eq.4})$$

The selected path is the one having the smallest cost metric for each path under investigation.

Three-Link Path (4 nodes)

If no eligible two-link paths (three-nodes) exist, repetition of the above procedure proceeds for a set of four-node paths. Trunks are added if there are no four-node path with sufficient spare bandwidth.

ADDING TRUNKS

When the rerouting of loads from the problem path to another path is not achievable, the only choice is to add additional network trunks. The cost of an additional trunk is proportional to the length of the trunk; therefore, to minimize cost, the cost of adding each trunk in the candidate group is determined, followed by selection of the least expensive trunk addition. The following steps achieve this:

Step 1: Calculate the cost of adding trunks within the problem link group.

Step 2: Consider the cost of adding trunks in the direct path.

Step 3: Compare the cost of additional trunks in two HSL paths, and select the smallest cost trunk.

Step 4: Select the trunk having the smallest cost of each of the previous steps and add a sufficient number of trunks to accommodate the overload.

The additional trunks in a target HSL group can be determined by:

$$Y = \frac{b}{o*s} - t \tag{Eq.5}$$

where

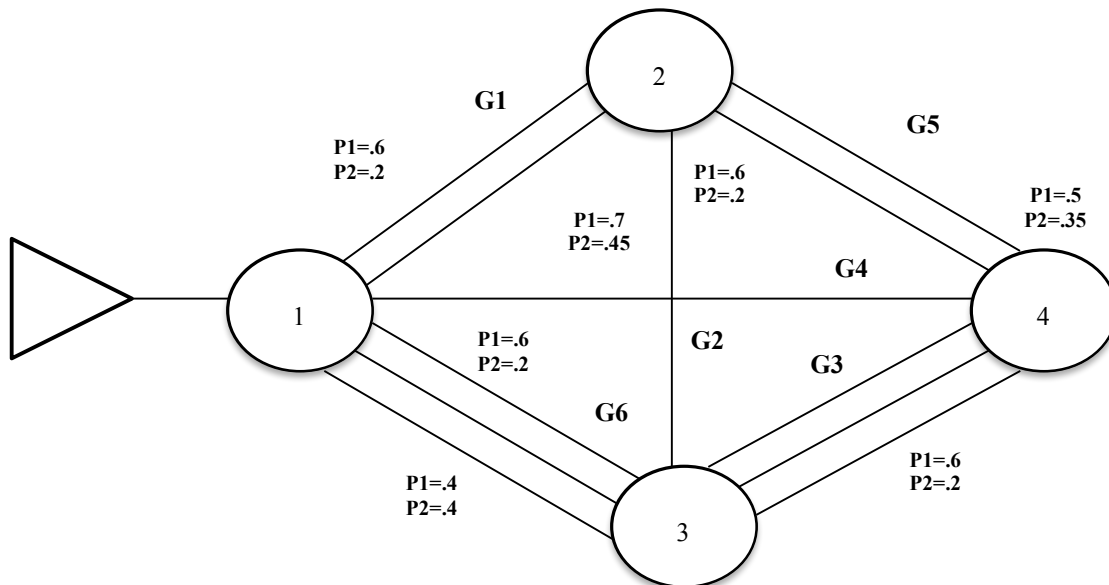
- y = number of trunks to add,
- b = maximum bits per second load offered to the HSL group,
- o = objective utilization for one trunk in the HSL group,
- s = speed of a trunk in the HSL group,
- t = total available trunks in the HSL group.

Once the added trunks are determined, load rerouting takes place. This rerouting is from the problem link to the group with the newly added trunks, with the load being equally distributed among the links in this group.

AN EXAMPLE

Figure 3 illustrates a network consisting of 4 nodes. Assume that the length of HSL group 1, group 2, group 3, group 4, group 5, and group 6 are 1000, 500, 1500, 900, 400, and 300 miles, respectively. For these distances, the utilization factors for both (ρ_1) and data traffic (ρ_2) are as shown in the figure.

Figure 3.



In Figure 3, problem path is 1-2-3-4 and the problem link is link 1 of group 2, because $\rho_1 + \rho_2 = 1.15$. The objective, as previously noted, is to keep the total utilization within a link at or below 90%, which ensures that delays for voice and data are within their component objectives. The removal of 25% utilization from this path achieves this maximum 90% utilization goal. For the data shown in Figure 3, this reduction should consist of 15% voice call $[(.7)/(1.15*.25) = .15]$ and 10% data $[(.45)/(1.15*.25) = .1]$, utilization.

Providing link 2 of group 2 with a 90% utilization factor is accomplished by rerouting 10% data utilization from link 1 of group 2 to link 2 of group 2. Here, the utilization factor of group 3 remains the same. Due to the delay sensitivity of voice traffic, rerouting 5% of voice traffic from 1-2-3-4 to group 4 reduces the utilization factor of path 1-2-3-4 by 5%. Rerouting of the remaining 10% of voice traffic from the problem link then takes place. Candidate

paths for this rerouting are 1-2-4 and 1-3-4. Both paths have enough spare capacities to support 10% additional voice traffic. Using Eq. 4 the cost metric for path 1-2-4, C_{124} , is $1000/.30+400/.10=7333$, and the cost metric, C_{134} , for path 1-3-4 is $300/.20+1500/.30= 6500$. Thus, path 1-3-4 becomes the selected path because of its lower cost metric.

Whenever there is insufficient spare capacities in the network to support the extra loads of the problem link, the cost of adding an additional HSL in the candidate groups is determined, with selection of the least expensive addition. In our example (Figure 3), the problem link is link 1 of group 2. The following steps are undertaken whenever there are insufficient capacities for rerouting traffic to reduce a link's utilization to 90%:

Step 1: Determine the cost of adding an HSL in the same group, $c_2=500*c$

Step 2: Determine the cost of adding HSL in group 5, $c_5=400*c$

Step 3: Determine the cost of adding HSL in group 4, $c_4=900*c$

Step 4: Determine the cost of adding HSL in group 6, $c_6=200*c$,

where

c = the cost per mile of an HSL

In our example, meeting the 90% criteria requires rerouting 25% utilization from link 1 of group 2. Thus, if all c 's are equal, then alternative 4 has the least cost. Therefore, adding one HSL has to group 6 in handles the extra load of path 1-2-3-4. The loads must them be distributed evenly within the group 6 links.

Should two groups in a path be overloaded (group 2 and group 3 in path 1-2-3-4), then the following alternatives must be considered:

Step 1: The cost of adding one HSL in group 2 and one HSL in group 3 is $Tc=2000*c$

Step 2: The cost of adding one HSL in group 5 is $Tc=400*c$

Step 3: The cost of adding one HSL in group 4 is $Tc=900*c$

Step 4: The cost of adding one HSL in group 6 and one HSL in group 3 is $Tc=1800*c$

Here, If all the c 's are equal, alternative 2 is the indicated selection because it has the smallest cost. Thus, adding one HSL to group 5, and rerouting 25% of the traffic load from path 1-2-3-4 to path 1-2-4 accomplishes the 90% goal.

CONCLUSIONS

A strategy, based on queuing theory, is developed and presented for the demand servicing process (traffic and service) of an ICN. This demand servicing process requires collecting and processing raw measurements to determine statistics on various end-to-end performance parameters. If any performance objectives, such as mean voice delay and mean data packet delay, are not satisfied, this servicing strategy determines the required corrective action for maintaining service quality. When rerouting changes are insufficient to restore service quality, the servicing involves changing traffic routing followed by network resizing. The advantage of this network strategy is its efficiency and flexibility in handling a variety of services and applications.

ACKNOWLEDGMENTS

The authors are grateful to Aparna Awasthi graduate Assistant for providing computer assistance in the tabulation of exact solutions for the scenarios presented in the paper.

AUTHOR BIOGRAPHIES

Mohammad Sedaghat is a Professor in the Department of Information Systems and Decision Sciences at Fairleigh Dickinson University. He obtained his Ph.D. in Operations Research from Polytechnic University, New York. His research interests include queuing theory and stochastic processes. His research papers have appeared in *Computers and Industrial Engineering*, *International Journal of Operations and Quantitative* among others. E-mail: sedaghat@fd.edu (Corresponding author)

Gary Bronson is a Professor in the Department of Information Systems and Decision Sciences at Fairleigh Dickinson University. He obtained his Ph.D. in Electrical Engineering from Stevens Institute of Technology, Hoboken, NJ. He was awarded the University award for outstanding research and publications, and is the author of over fifteen textbooks on programming languages. He was a senior project engineer at Lockheed Electronics and an invited lecture to Bell Laboratory's in-house continuing education program.

REFERENCES

- Ahuja, R.K., Magnanti, T.I., & Orlin, J.B. (1993). *Network Flows; Theory, Algorithms, and Applications*. Englewood Cliffs, NJ: Prentice Hall
- Brooks, J.D., Kar, K. & Mendonca, D.J. (2016). Allocation of flows in closed bipartite queueing networks. *European Journal of Operational Research*, 255, 333-344
- Dib, O., Manier, M., Moalic, L. & Caminada, A. (2017), *Combining VNS with Genetic Algorithm to solve the one-to-one routing issue in road networks*. *Computers & Operations Research*, 78, 420-430.
- Dumas, M.B. & Schwartz, M. (2009). *Principles of Computer Networks and Communications*. New Jersey: Pearson Education.
- Hochfelder, D. & Mischa, S. (1999). *Telecommunications Networks*. September 17, 1999.
- Gross, D. & Harris, C.M. (1998). *Fundamentals of Queuing Theory*: New York, NY: Wiley.
- Holtznider, B. & Jaffe, B.D. (2001). *IT Manager's Handbook*. New York, NY: Morgan Kaufmann Publishers.
- Nicholson, C.D. & Zhang, W. (2016). Optimal network flow. *Computers and Industrial Engineering Journal*, 99, 260-268.
- Panaousis, E., Karapistoli, E., Elsemary, H., Alpcan, T., Khuzani, M.H.R., & Economides, A.A. (2017). Game theoretic path selection to support security in device-to-device communications. *Ad Hoc Networks*, 56, 28-42.
- Parthasarathy, P.R., Selvaraju, N., & Manimaran, G. (1999). A Paired Queuing System Arising in Multimedia Synchronization. *Mathematical and Computer Modelling*, 30, 133–140
- Takahashi, M., Osawa, H. & Takahashi T., (2000). On a Synchronization Queue with Two Finite Buffers. *Queueing Systems*, 36, 107-123.
- Todinov, M.T. (2013). *Flow Networks. Analysis and Optimization of Repairable Flow Networks. Network with Disturbed Flows*. New York, NY., Elsevier Inc.
- Trivedi, K.S. (2002). *Probability and Statistics with Reliability, Queuing and Computer Science Applications*. New York, NY: Wiley.
- Walrand, J. (1998). *An Introduction to Queuing Networks*, Englewood Cliffs, New Jersey: Prentice Hall.

NOTES