A Field Service Support System Using the Computer Analysis of Networks of Queues (CAN-Q) Model

H.T. Papadopoulos and J. Darzentas

Dept. of Mathematics, University of the Aegean,
GR-832 00 - Karlovasi, Samos, Greece

The Field Service (FS) Organization of many companies constitutes a vital Department playing an important role to their success. Field Service managers need some tools to analyse the impact of their decisions on customer service level, which must be as high as possible (at least 95%), and inventory cost, trying to reduce it to a minimum level (its ratio to the total FS revenue must be less than 10%). Such a tool is presented in this paper. We applied a closed queueing network model developed by Solberg, called CAN-Q (acronym for Computer Analysis of Networks of Queues). This model, although originally developed for modelling Flexible Manufacturing Systems, has been applied to the FS organization of a subsidiary of multinational computer company, in Greece, and it has proved to be very efficient from the computational point of view and constitutes a powerful tool for the FS managers, providing them with some useful performance measures. Its successful application gives rise to try some other queueing network models, available from the literature, in combination with various inventory control models to help FS managers solve their critical problems.

Keywords: *Strategic planning; queueing networks; field service; decision support systems.*

1. Introduction

Many industries, multinational and national companies set as their first priority the so-called 'customer satisfaction', as they know that only in this way they can maintain or even increase their market share. There are many factors that contribute to this end. To name a few, the quality of the products they sell, the quality of the service they provide, both administrative and technical, among others. The FS Department has the responsibility of the service maintenance of all the different products, which are sold by sales forces of the company and which many times may come from different manufacturers, making the management of the field service a difficult task. Field Service managers have to cope with the two conflicting objectives: (a) to maintain high level of customer service, which must be above 95%, and (b) keep the spares inventory level as low as possible. An efficient measurement for this, is to keep the ratio of the spares inventory cost to the total revenue of the FS organization less than 10%, usually between 5%-8%. Some other important decisions that the FS manager has to take are how many FS engineers to have and how to allocate them to the company's customer population.

In this paper, we attempted to approach the FS manager's problem by implementing the computer analysis of networks queues (CAN-Q) model, which is a closed queueing network (CQN) model introduced by Solberg [5] mainly to model the flexible manufacturing systems (FMS) (for a detailed analysis of FMS as CQNs the interested reader is addressed to Buzacott and Shanthikumar [1], Chapter 8, Gershwin [3], Chapters 6 and 9, and to Papadopoulos *et al.* [4], Chapter 3, among others). The idea of modelling the field service problem as a closed queueing network is justified as the number of customers of a FS organization is usually constant, like the number of parts/pallets circulating within a FMS. Further, we decided to implement CAN-Q and not any other CQN model, in order to exploit the existing software programme, developed by Solberg's research team. We applied this model to the data of the FS organization of a multinational computer subsidiary, in Athens, and we saw that it performs quite satisfactorily. Unfortunately, due to confidentiality reasons, numerical data and results cannot be presented.

This paper is organized as follows. In the next section, we present the structure of a real FS organization having in mind that of the (multinational) computer subsidiary, and we make some remarks relative to the function and operations of this organization. Then, in the following section, we give the development of the model. Last section concludes the paper and recommendation for further research is suggested. Finally, in the Appendix, the analysis of the CQN model is given with the implementation of the CAN-Q algorithm (Solberg's model).

2. The Structure of a FS Organization of a Computer Company

In this section, we describe shortly the structure and the operation of a FS organization of a computer (Multinational) subsidiary, as this motivated and formed the basis for the development of our FS support system. We believe that the structure and operation of the FS Departments of other industries are quite similar (e.g. photocopier, communication companies, etc.). For this reason, we strongly recommend the application of the proposed model, described in the next section, to these companies as well, with slight modifications and adjustments. In a computer (subsidiary of a multinational, or national of medium size) company, the FS organization consists mainly of two Departments:

- (a) the FS contracts Sales Department, which brings revenue to the organization by selling service contracts and other services (such as preparation of computing rooms to accommodate the computing systems), to the customers. Other source of revenue is the revenue transfer coming from the Sales Department as a certain percentage on the sales of hardware (H/W) and Software (S/W) products, covering the one-year period during which the products are under guarantee, and
- (b) the Technical Support Department, which consists of the engineers (technicians), for both H/W and S/W products. This Department together with the administrator, the secretaries, the call hander(s), the spare parts inventory, the various travelling and training expenses, and of course the wages of all the FS employees, constitute the expense part of the FS organization.

The operation of the technical support Department is done as follows. The customer who has a problem calls the call-handler in the FS Support Department and this person longs the call in the so-called FS Log-Book, checks the contract of the customer and passes the information to the dispatcher who arranges for an engineer to take care of this call, according to the terms and conditions of the contract of this customer. There are various classes of customers, depending on the type of their contract with the company. A typical classification, for example, is:

1. *Class-1 customers:* These are customers with 24-hour coverage and they are assigned the highest priority, the company being obliged to service these customers immediately, having an engineer stand-by, the non-working hours, even on the weekends and holidays. Example of such a customer is a production plant (e.g., cement industries, etc.) which maintains 3 shifts per 24 hours.
2. *Class-2 customers:* These are high-priority customers too, where according to the contract, the FS engineer must go and fix their problem within 2 hours from the call. A representative example of such a class of customers is a bank which wants to maintain a reliable on-line system.
3. *Class-3 customers:* These are also priority customers, where according to their contract, the FS engineer must go and fix their problem within 4 hours from the call. An insurance company may belong to this class of customers.
4. *Class-4 customers:* These are the normal customers, where according to their contract, the FS engineer must go and fix their problem within 8 hours from the call. The majority of the customers population of the company belong to this class (e.g. Universities, research institutes, various private commercial companies, etc.).

5. *Class-5 customers:* These are customers with an *elementary* service maintainance contract, that cannot afford a normal contract and try to be covered somehow at least from the spare parts point of view, where according to their contract, the FS engineer must go and fix their problem within 16 hours from the call. Of course, the cost for such a contract is lower than a normal (say 65% of the price of a normal contract). The company tries not to sell such contracts and for this, only a few customers belong to this class.

6. *Class-6 customers:* These are customers *without* contract, the so-called *per-call customers.* The rules of the company dictate the engineers to give them very low priority and depending on the work-load, they are obliged to go and fix their problem within a week, over-charging them, of course, in order to force them sign a normal service maintainance contract.

The FS manager together with the FS Unit managers (managers of the H/W and S/W engineers) are obliged to service all these customers trying to always be consistent with the terms and conditions of their contract.

One of the problems they face is how to allocate the engineers to the various customers. Usually the allocation is done depending on the account (customer class) and the type of the system (super, mini, PCs-UNIX, DOS, VMS, etc.) and the part of the computer (CPU, peripherals (discs & tapes), printers, terminals, etc.). It is very common the case where more than one engineers are specialized on a specific area or part of the computer or type of operating system, for back-up reasons and for easily dispatching the calls among them.

FS managers, restricted by the budget, don't have the luxury of hiring new engineers. Instead, they invest more money on training the existing staff in different areas (e.g., both for mini systems and PCs, and H/W and S/W products) in order to decrease the idle time of the engineers and increase their utilization, efficiency and productivity, in general.

Concerning the inventory investment and the spare parts stock management, this is an important factor affecting the probability an engineer will not have the necessary parts. In this work, it's not our aim to deal with stock control models. We just make some remarks.

To handle all the various classes of customers, from the stock control point of view, we say that, first of all, this is the Logistic manager's (the materials manager's, more specifically) responsibility. The manufacturer, which in a multinational company, is the 'mother'-company, issues special Logistic and Field Service plans for all the products, providing a *recommended spares list* (RSL) for various levels of service (LOS), usually 95% and 98%. Then, depending on the consumption of the spares (forecasted consumption index for the new products, based on the MTBF given by the Quality Department of the manufacturing plant), the re-ordering policy is chosen, by setting the appropriate target stock levels (TSL) for each line item. In either case, the priority of the customers is always taken into account. For example, for class-1 customers, it's not rare to keep dedicated stock, which depending on the distance of the customer's site, it may be stocked at customer's place, after negotiations with the customer. For class-1 and class-2 customers, the so-called *option-swap material* solution is also applied. For example, if the customer is a production plant or a bank and the problem cannot be fixed within 1-2 hours, the FS engineer replaces the whole CPU unit or the peripheral, by a similar one, stocked for this purpose and for this particular customer, at the warehouse (W/H), and after the repair - which takes place at the Company's premise, - the engineer returns the customer's option and takes back the company's one; this is the option-swap material. In this way, the customer doesn't remain 'down' for a long time and he/she is satisfied with this solution.

A critical point, tested experimentally, for a couple of years, at the Greek subsidiary of the computer company, is to *demolish the kits* and maintain single line items stock, in order to increase the availability of the spare parts and thus the level of service to the engineers (and to the customers). Of course, some kits must be maintained, for certain products that are highly consumed and their spares are highly demanded, e.g., for specific models of printers, terminals, etc.

Another important issue related to the Logistics and Field Service strategy, is to establish a *local repair center* at Company's premises (close to the warehouse), or at the sub-contractor's site. In this way, the lead time of a spare part's order is reduced greatly, reducing the spares stock level, to some good extent. Of course, some cost is involved here and this constitutes another decision variable for the FS manager. It is suggested the establishment of such a local repair centre, at some minimum cost, in the beginning, by not hiring new engineers, but utilizing the existing ones, on a rotational basis, starting with the repair of peripherals and PCs. Further investigation is needed on whether is beneficial to further invest on the repair of more complicated options (e.g., CPUs), for which expensive tools and equipment are needed. Data from other more developed subsidiaries of the same computer company (e.g., in Israel and Italy) supports this investment combined with the subcontracting of the engineering repair staff, due to the restrictions in hiring new employees, directed by the Headquarters.

To increase the utilization of the FS engineers, decreasing their idle time, is something that can be done at some cost in two ways: (i) keeping extra spares *stock on vans* that are moving in various areas of the city (where the most critical customers are), - this is the *mobile stock*, - paying money for security, as well, and (ii) having *dedicated courier service* at the warehouse, delivering in this way the spares at the customer allowing the engineers to move from customer to customer, without being obliged to visit the W/H to collect the good spares, for the new call, or to return the defectives consumed at a former call. This solution is very useful, since otherwise, in cities like Ahtens, with heavy traffic, the time of the engineers is splitted almost equally at travelling and repairing!

3. Model Development and Application

The model we decided that it best suits to the FS organization of the computer company, described in the previous section, is the CAN-Q, a closed queueing network (CQN) model introduced by Solberg [6]. Of course, this model was developed primarily to model FMS with one type of 'customers'. We modified it appropriately to reflect better the needs of our FS organization. This model is illustrated in Figure 1.

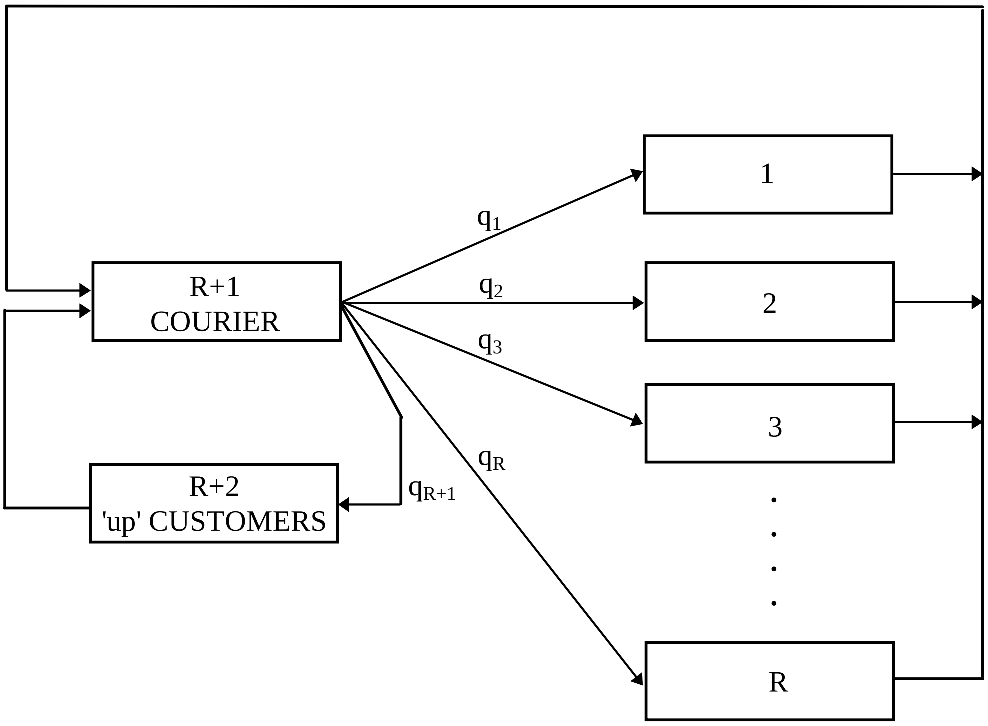

Figure 1: A CQN model for the FS organization

The actual layout of this network has as follows. The customer population of the company is splitted into R different classes of customers (in our case R=6), depending on either the terms and conditions of the service maintainance contract or the type of the system (CPU, memory boards, discs/tapes, terminals, PCs, etc.).

It is the FS unit manager's responsibility to allocate the FS engineers among the customers, depending on their experience, training, knowledge, etc.

In Figure 1, stations 1,2,...,R, represent the customers of class 1,2,...,R, respectively, that have failed. Note R+2 models all the 'up'-customers of all the classes. Whenever an 'up'-customer goes down, he/she calls the company, more specifically the call-handler who logs the call at the so-called FS Log Book and informs the FS dispatcher (the FS Unit manager usually plays this role), who in turn allocates an available and appropriate engineer (depending on the customer class and the type of the problem). The engineer then goes to the customer's site to fix the problem, having first either visited the warehouse (W/H) or called the W/H attendant to bring him/her the spare part(s) he/she needs for the job, via a courier (this is represented by station R+1 in figure 1). For modelling purposes, when an 'up'-customer goes down, he/she informs the call-handler and he/she is transferred to one of the R stations - to receive service by a FS engineer, with probability q_i, (i=1,...,R),- depending on his/her class (q_{R+1} represents the probability that the customer becomes 'up'). After having received the (repair) service the 'down'-customer becomes an 'up'-customer again and he/she is transferred to the R+2 station via the transporter station (node R+1). In reality, customers are at their sites and the courier transfers the good spare parts at their sites and after the service completion, the courier brings the defective parts back to the warehouse, plus any good parts that were not used for that particular call.

It has to be noticed that with the CAN-Q model, the 'up'-customers (node R+2) are not modelled. But it is very easy to do that and derive, say, the expected number of 'up'-customers, by the formula:

$$E[N_{R+2}] = N - \sum_{i=1}^{R+1} E[N_i],$$

where N denotes the total number of customers, which is known to the company and $E[N_i]$'s are calculated from the model (see the Appendix for the analysis of the CAN-Q model).

Stations 1,...,R+1, may be modelled as single-server or multi-server nodes, depending on the work-load of a particular class of customers (e.g., class-5 and class-6 customers are only a few and the respective stations are not only single-server stations but even they are pooled to justify the FS engineer utilization, and this facility (sensitivity analysis) is provided by the CAN-Q model).

The reason we applied the CAN-Q algorithm as a FS support system, was not only to exploit the S/W programme, which is available in FORTRAN, but also because the running time for solving typical realistic problems is very low (a few seconds only) and neither storage requirements nor execution time pose any difficulties. The output of this model includes the relative utilization and the station (i.e. FS engineer) utilization, the expected number of customers of any class (at each node), the average number of customers (of any class) waiting in queue for service, the average waiting time spent in the system (for repair) by any 'down'-customer, etc.

Remarks: Nodes 1 and 2 that model the repair operation of the high priority customers may be modelled as self-service queueing systems, i.e., $M/M/\infty$. The only case where these two classes of the highest priority customers are not served immediately, is when the spare part (which is always available at the W/H or at customer's site) is dead-on-arrival (DOA), i.e., although is brand-new and sealed, there is a problem from manufacturing. This case is very rare, and the FS manager tries to find a solution for the critical customer to resolve the problem, implementing any unusual method he/she can imagine!

All the rest classes of customers (3-6) are served at nodes 3-6, respectively, which are modelled as $M/M/c/K$ queueing systems, with the customers being serviced according to the FCFS service discipline.

Node R+2 is a fictitious queueing station accommodating all the 'up' customers of the company (something like the 'negative' customers), consisting of customers of all classes whose system/option is up and running. Notation concerned with the mean service rates (or mean service times) of all the nodes of the CAN-Q network (of Figure 1), as well as the definition of the various performance measures and the solution of the model, are given at the Appendix.

Applying the data of the Greek subsidiary of the multinational computer company, using the proposed CQN model, we were able to derive numerical results with a remarkable accuracy, - compared against the real data (the deviation varied from 2% to 13%), - concerning the various performance measures, such as the expected number of customers at each node and the expected utilization of the FS engineers, or equivalently the expected idle time of the FS engineers, and the production rate or equivalently the average waiting time spent by any customer in the system (meaning the total time elapsed from the time the customer called the call-handler till the time his/her problem was fixed). We discovered, based always on the real data, that for class-6 (per-call) and class-5 (with an elementary service contract) customers, the results for the FS engineers utilization were very very low, meaning that their idleness was very high, and this means that the FS manager must not allocate engineers only for these classes of customers. And this happens, indeed, in reality.

The advantage of the proposed model is that it models separately the transporter (courier), which is very realistic, as travelling time especially in cities with heavy traffic (like Athens) is not negligible at all.

4. Conclutions and Further Research

Applying the computer analysis of networks of queue (CAN-Q) algorithm, introduced by Solberg [6], we developed a closed queueing network model to form the basis for field service support system, by slightly modifying the CAN-Q model which was originally developed for modelling FMS. Running this model with the data of the FS organization of a subsidiary of a computer multinational company in Athens, Greece, we saw that it gives very accurate results. The proposed model offers the FS manager the possibility to estimate various useful performance measures such as the mean sojourn (repair) time of a customer of any class, the average production rate (how many customer calls are handled and completed per certain unit time), the utilization of the FS engineers, or equivalently their idleness, the expected number of customers (of any class) that are 'down' at any time and the maximum number of customers that may be allocated per FS engineer to maintain a high level of customer service (usually above 95%).

Although we focused on the application of the proposed CQN model to the small Greek subsidiary of the computer (multinational) company, we strongly believe that this model is applicable to many other companies with similar structure of their field service organizations, such as the photocopier and the (tele)communication industries, among others. The successful application of CAN-Q as a FS support system, give rise to try some other queueing network models, available from the literature, in combination with various inventory control models to help FS managers solve their critical problems.

A quite interesting and useful area for further research would be the development of a *simple*, *easy to use* (by the FS manager who usually doesn't know anything about queueing theory), *total system cost model*, incorporating all the possible decision variables into it, such as the number of the FS engineers required, their allocation to the various customer classes, the effect of the spares stock policy on the level of service (LOS), the FS engineers utilization or the percentage of their idle time, the cost for training them, and how much time a customer waits from the time of a breakdown till his/her problem has been fixed (total sojourn/repair time), and if the terms and conditions of his/her contract are being met, etc.

Appendix

Here, we first give the notation used in the CAN-Q network and the (mathematical) definition of the various performance measures that are derived from this model, and then we just present the solution of this model adopted to the FS organization of the computer company. The details of their derivation are omitted as these may be found in Solberg [5] & [6].

Notation:

N = the total number of customers (of all R = 6 classes) in the system (both 'up' and 'down');

N_{R+2} = the number of 'up'-customers (of any class i=1,...,R);

N_i = the number of 'down'-customers (of any class i=1,...,R) that are waiting for or in repair at the respective i-station;

c_i = the number of servers (FS engineers) at station i (i=1,...,R,R+1);

ì_i = the average service (repair) rate of station i (i=1,...,R,R+1), when it is busy (i.e., $1/\text{ì}_i$ is the average processing (repair) time at station i);

q_i = [For modelling purposes] the probability that the courier transfers the customer of class i (i=1,...,R) to station i for repair. [In reality] q_i is the probability the courier delivers the spare part(s) to the customer of class i (at station i). This probability, q_i, is nothing else but the failure frequency of a customer of class i;

q_{R+1} = the probability that the 'down'-customer of any class i (i=1,...,R) has been repaired, becoming an 'up'-customer;

Definition of some performance measures:

- The *utilization of the ith station*, denoted by u_i, is defined as, the long run average number of busy FS engineers at station i. Mathematically, this is given by

$$u_i = \frac{q_i \cdot \text{ì}_{R+1}}{\text{ì}_i} u_{R+1}, \quad \text{for all } i \neq R+1 \ (i=1,...,R).$$

The *utilization per FS engineer* of the ith station is defined as the fraction of time that each FS engineer is busy, and is given by u_i/c_i.

- The *relative utilizations* of the i stations, denoted by r_i, are defined by the fractions appearing in the above formula, i.e.,

$$\begin{aligned} r_i &= \frac{q_i \cdot \text{ì}_{R+1}}{\text{ì}_i} = \frac{u_i}{u_{R+1}}, \text{ for } i=1,...,R, \\ &= 1, \text{ for } i=R+1. \end{aligned}$$

- The *production rate* of the FS Department (concerned with the repair branch) is defined as the steady-state average number of 'down'-customers that are being repaired per time unit (e.g., per day or week or month). This is equal to the mean rate of flow of repaired customers out of the transporter station (courier, i.e., station R+1). This is denoted by X and is expressed by the formula

$$X = (q_{R+1}\cdot\mu_{R+1})\cdot u_{R+1}.$$

- The *average time spent in the system* by any 'down'-customer, denoted by W, is defined as the average total repair time, including transportation and any queueing time (waiting for the FS engineer to come), and is given by the well-known Little formula:

$$W = \frac{N}{X}.$$

The solution:

The state of the system modelled by the network depicted in Figure 1, is indicated by the vector $(N_1,...,N_{R+1})$, and the steady-state probabilities, for single-server stations, are given by

$$P(N_1,...,N_{R+1}) = \frac{1}{G(R+1,N)}\, r_1^{N_1}\cdot r_2^{N_2}\cdots r_{R+1}^{N_{R+1}}.$$

where G(R+1,N) is a normalized constant defined so that all of the probabilities $P(N_1,...,N_{R+1})$ sum to one. Buzen [2] provided a very efficient recursive technique for its calculation.

If one or more of the stations has multiple FS engineers, a slight modification is necessary. For any such station i, having c_i FS engineers, the term in the product corresponding to that station (that is, $r_i^{N_i}$) must be replaced by $r_i^{N_i}$ / $A_i(N_i)$, where,

$$A_i(N_i) = \begin{cases} N_i! & \text{if} \quad N_i \le c_i, \\ c_i!\,c_i^{N_i - c_i} & \text{if} \quad N_i > c_i. \end{cases}$$

The utilization of the transport station (courier), u_{R+1}, is of particular interest. This is given by (for a proof see Solberg [6])

$$u_{R+1} = \frac{G(R+1,N\text{-}1)}{G(R+1,N)}.$$

As soon as u_{R+1} is determined, the average production rate of the FS Department, concerned with the repair operations, the expected time spent in the system by any 'down'-customer (of any class) and the utilizations of the stations are given, respectively, by

$$X = q_{R+1}\cdot\mu_{R+1}\cdot\frac{G(R+1,N\text{-}1)}{G(R+1,N)},$$

$$W = \frac{N\cdot G(R+1,N)}{q_{R+1}\cdot\mu_{R+1}\cdot G(R+1,N\text{-}1)},$$

$$u_i = r_i \frac{G(R+1,N-1)}{G(R+1,N)}\ .$$

The marginal probability distribution for the number of customers at the ith station (i-class customers), provided that it is a single-server station, is given by

$$P(N_i = \nu) = \frac{r_i^\nu}{G(R+1,N)}\,[G(R+1,N-\nu) - r_i G(R+1, N-\nu-1)].$$

Station R+1, i.e., the courier, need not be a single-server station. In this case, the marginal probability distribution is given by

$$P(N_{R+1} = \nu) = \frac{r_{R+1}^\nu \cdot G(R, N-\nu)}{A_{R+1}(\nu)\cdot G(R+1, N)}\,.$$

If any of the other stations have multiple servers (more than one FS engineer), it is possible to obtain their marginal distributions. To do so requires permuting the indices of the stations in such a way that the one of interest becomes R+1, recomputing the matrix G and using the above equation.

The marginal distributions can be used to compute *average queue lengths* for each station. For single-server stations, it holds:

$$E[N_i] = \sum_{\nu=1}^{N} r_i^\nu \cdot \frac{G(R+1, N-\nu)}{G(R+1, N)}\ ,$$

and

$$E[N_{R+1}] = \sum_{\nu=1}^{N} \nu r_{R+1}^\nu \cdot \frac{G(R, N-\nu)}{A_{R+1}(\nu)\cdot G(R+1, N)}\,.$$

It also holds:

$$E[\text{number in queue of station i}] = E[N_i]-u_i,$$

and the *idleness* of a station i (i=1,...,R) is given by the probability

$$P[N_i = 0] = 1- u_i,$$

and for the courier (station R+1):

$$P[N_{R+1} = 0] = \frac{G(R,N)}{G(R+1,N)}\ .$$

Now, concerning the *allocation of the FS engineers* to the customers population, this may be easily controlled by the FS manager, by finding the maximum N that satisfies the inequality:

$$E[N_{R+2}] = N - \sum_{i=1}^{R+1} E[N_i] \geq 0.95N,$$

when he/she wants to maintain a 95% level of service to his/her customers, which is an acceptable level for customer satisfaction.

References

[1] Buzacott J.A. and Shanthikumar J.G. (1993) *Stochastic models of manufacturing systems.* Prentice Hall, New Jersey.

[2] Buzen J.P. (1973) Computational algorithms for closed queueing networks with exponential servers. *Comm.* ACM, 16, 527-531.

[3] Gershwin S.B. (1994) *Manufacturing systems engineering.* Prentice Hall, New Jersey.

[4] Papadopoulos H.T., Heavey C. and Browne J. (1993) *Queueing Theory in Manufacturing Systems Analysis and Design.* Chapman and Hall, London.

[5] Solberg J.J. (1976) Optimal design and control of computerized manufacturing systems. In *Proceedings* AIIE 1976 *Systems Engineering Conference.*

[6] Solberg J.J. (1977) A mathematical model of computerized manufacturing systems. In *Proceedings 4th International Conference on Production Research, Tokyo.*