APPLICATION OF QUEUING THEORY FOR AVAILABILITY ASSESSMENT IN AIRSPACE CONTROL SYSTEMS

Walter Nogueira Pizzo
Paulo Sérgio Cugnasca[1]

Polytechnic School
University of São Paulo

Abstract

This paper provides useful information to build availability models for computer systems used in airspace control centers, based on analytical models provided by queuing theories. A general model is first presented, referencing a published case study, where the authors described the use of queuing models to establish availability parameters related to a data center operation and its management issues. In addition, some considerations are introduced to extend this general model in order to propose its application for the specific computer systems used in integrated airspace control centers, where operational control could depend on human controllers, responsible for civil air traffic control or military air defense operations. This model could also be applicable where process controls require intensive use of human-machine interfaces (HMI). From this study, further extensions could be also developed, intended to similar applications on other specific cases of process control computer systems.

[1] Authors Contact:
GAS/PCS/EPUSP.
Av. Prof. Luciano Gualberto, Trav.3, 158
Prédio da Engenharia de Eletricidade
CEP: 05508-010 - São Paulo – SP, Brasil
Phone Number: +55 11 3091-5401
Email: paulo.cugnasca@poli.usp.br

1. INTRODUCTION

This article presents useful information to build availability models for computer systems in airspace control centers, where many activities related to both civil air traffic and military operations control are conducted with direct interest for a broad array of management, including planning for human resources dimensioning, technical resources and sustainability evaluation for operational needs within those centers.

Based on a closed queuing model, it is first presented a general model for the operation and the maintenance of a generic computer center (data center). The concepts and analytical models described helps to solve problems often addressed to management and resources allocation within this kind of data centers, where main characteristics are the intensive usage of storage or processing resources in a variety of applications such as storage serves, transactional commerce or banking systems, automation and process control systems, involving closed loop control, where computer systems respond for the entire process control, without human intervention, as occurs in some industrial plants and transportation systems like some railway and subway control systems.

After that, some considerations on operational characteristics of airspace control centers are presented, with the objective of extending the general model application and to make it more adequate for these specific cases, since the control of the operations in aeronautical processes is not integrally executed by computational systems, being instead dependent on interventions of human operators. In this aspect, it is important that the model considers the existence of operational states that are not completely automatized, either for the inherent need of some type of human action, due to safety and reliability issues, or due to the characteristics of fault tolerance, in order to assure system availability even in degraded operational conditions, or still for the definition of different classes of machines and their distinct demands of service.

Thus, this work presents information examples of applications of the queuing theory to address practical questions of sizing and availability assessments, important issues for airspace control systems, indicating useful techniques for the management of critical mission control centers, where many aspects related to human operation, fault tolerance, degraded operation, and demand of service maintenance are basic concerns.

As an example, in Brazil, the structure illustrated in Figure 1, refers to an Integrated Center of the Air Defense and Air Traffic Control - CINDACTA (DEPV, 1999b). Many

other air traffic control centers in the world operate like this one, performing equivalent services and relying on human intervention in the control process. Their operational structures also derive from the recommendations of normative entities such as the Federal Aviation Administration (FAA) of the United States of America (FAA, 2006) and the Organization of International Civil Aviation (ICAO) (ICAO, 1996).



Figure 1 – Operation of an Air Traffic Control Center
(photo: Força Aérea Magazine)

2. OPERATIONAL MODEL OF A COMPUTER CENTER

In this section, a model to assess the operational availability of a computer center is presented in summary, as illustrated by the referenced case study (Menascé et al., 2004), in which the authors built a model based on the queuing theory, in order to solve problems of sizing analysis and parameters assessment of the availability of a data center.

2.1. Operational and Maintenance Model of a Data Center

In the cited case study (Menascé et al., 2004), the authors considered the adoption of the following model: an hypothetical data center operates with M machines and a staff of N dedicated people to the attendance of systems and equipments' failure and maintenance. A diagnostic system is considered to automatically perform the following functions:

- Detect failures in any of the M machines;
- Maintain a queue of machines waiting for repairing;
- Log the instant when a machine failed;
- Log the events where one technician starts/finishes the maintenance of each machine.

In this type of configuration of computer centers, management is interested in keeping high levels of availability, by means of the maintenance of high reliability (reduced failure) rates, as well as optimizing the maintenance services of the machines, with automatic diagnose systems, specialized technicians and well defined processes to identify or to locate faults, efficient execution of the repairs and quick return to the operation after servicing any failed equipment submitted to maintenance. One of the fundamental management problems of this data center is to size the necessary staff to service the operation,

in order to estimate the necessary number of machines to guarantee a certain level of operational reliability. This means to provide a nominal service level as expected or defined by formal service level agreements, until a minimum service level established for a degraded case. These parameters are related to the number of maintenance people (staff sizing) and to their technical skills, and are reflected in the mean time to repair failed machines (MTTR), in order to guarantee the desirable service availability level.

As indicated in Figure 2, a queuing network model can represent this operation. The following items are considered in this case: a) all the machines are identical and operate independently, resulting in the same failure rate λ for each machine, where $\lambda = 1/MTTF$ (mean time to failure); b) each one of the M machines presents only two possible states ("operational" or "failure"); c) a diagnostic mechanism executes periodical checks over all the machines in operation and, when a failure is detected, the system automatically signals the maintenance staff, indicating the machine that must enter in a queue to wait for repair; d) once in the queue, the machine waits for one of the N people of the maintenance staff; and e) once repaired, it returns immediately to the pool of machines in the operational status.

As mentioned in the cited case study, the management of this type of data center is interested in answering the following questions:

I. Given the failure rate, the number M of machines, the number N of the maintenance people, and the average time to repair the machines *MTTR*, what is the probability that exactly j machines are operational at any given time?

II. Given the same previous parameters, what is the probability that at least j machines are operational?

III. Given the failure rate λ, the number M of machines and the repair rate μ ($\mu = 1/MTTR$), what is the number N of necessary maintenance people to guarantee that at least j machines are operational with a given probability?

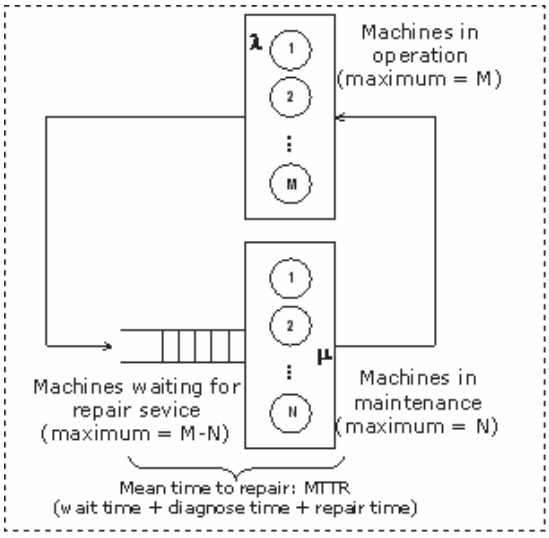


Figure 2: Queuing model for the operational-maintenance states of a computer center

In the cited case, it is admitted that failures of all the machines occur independently, with the same failure rate λ. It is also considered that the repair rate μ, equivalent to the inverse of the mean time to repair (1/*MTTR)*, is identical for all the types of repair services and are independent of the technician who executes the service. In the case that different failure rates are observed for each machine or group of machines, a more complex model could be elaborated, considering multiple class queuing models (Menascé et al., 2004) for the distinct demands of services. In case of different technician repair rates, a heterogeneous multi-server model could be elaborated, to represent each individual repair rate.

2.2. Solving the queuing network model

As illustrated by the same case study (Menascé et al., 2004), the solution for the closed queuing network model presented in Figure 2, can be modeled by a Markov Chain (Menascé et al., 2004; Shooman, 2002), where each state corresponds to the situation in which there are *k* failed machines out of the total of *M* machines, and with a maximum of *N* machines in maintenance, as illustrates Figure 3.

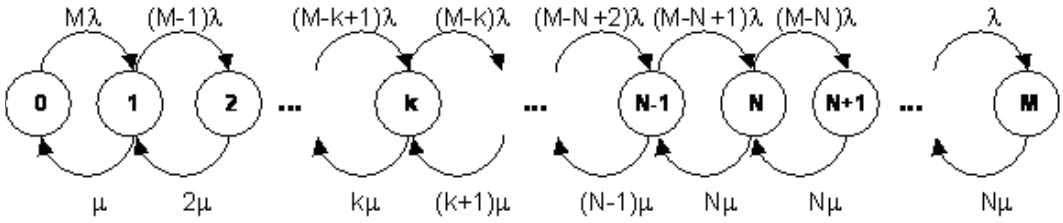


Figure 3: Markov chain model for a data center with M machines

The transition from the state *k* to the state *k+1* occurs when a machine fails, event that occurs with a fail rate λ multiplied by the number *M-k* of machines in operation. In a similar way, a transition from the state *k* to the state *k-1* takes place whenever a machine is repaired, a process that occurs at a repair rate μ times the number of machines being repaired *k*, limited to a maximum of *N.*μ, as the maximum number of machines in maintenance is limited to the number *N*, corresponding to the maximum size of the maintenance staff.

Since failures are independent, the aggregate failure rate is *M.*λ, when all the *M* machines are operational. This rate decreases to at least 1.λ, when only one machine remains in operation. Thus, in state k, the following failure rate is obtained:

$\lambda_k = (M-k).\lambda$, for k = 0, 1, 2, ... M-1 (1)

The aggregate repair rate can be expressed as follows, once it is dependent on the total number of technicians available, which is limited to *N* technicians:

$\mu_k, = k\mu$, while there is an available technician, when k = 0, 1, 2, ... N (2)

or,

$\mu_k, = N\mu$, when all technicians are busy, when k = N+1, ... M

The solution of this model can be obtained using the theorem of the generalized birth-death processes - GBD (Menascé et al., 2004), resulting in the establishment, in the steady state, of the following probability p_k to find the system in the state *k*, which is equivalent to the probability to have *k* machines in failure:

$$p_k = p_0 \cdot \prod_{i=0}^{k-1} \lambda_i / \mu_{i+1} \quad , \quad \text{for k = 0, 1, 2, 3, ...} \quad (3)$$

Hence:

$p_k = p_0 \cdot [M\lambda/\mu \cdot (M-1).\lambda/2\mu \cdot (M-2) \cdot \lambda/3\mu \ldots (M-k+1).\lambda/k\mu)]$, for k = 1, 2, ... N (4)

or,

$p_k = p_0 \cdot [M\lambda/\mu \cdot (M-1).\lambda/2\mu \ldots (M-N+1) \cdot \lambda/N\mu] \cdot [(M-N).\lambda/N\mu \cdot (M-N-1) \cdot \lambda/N\mu \ldots (M-k+1)\lambda/N\mu]$, for k = N+1, ... M

Considering that the sum of the probabilities to find the system in any of the M states is equal to 1 (from k=0 to k=M failed machines), then:

$$\sum_{k=0}^{M} p_k = 1 \quad (5)$$

Therefore, the value of p0 can be obtained from the known parameters of the model, as follows:

$$p_0 = 1 \Big/ \left[\sum_{k=0}^{N} \left(\frac{\lambda}{\mu}\right)^k \cdot \binom{M}{k} + \sum_{k=N+1}^{M} \left(\frac{\lambda}{\mu}\right)^k \cdot \binom{M}{k} \cdot \frac{N^{N-k}.k!}{N!} \right] \quad (6)$$

Once p_0 is calculated in Equation 6, any value p_k can be obtained by using Equation 4. Therefore, an interactive process can be organized, for instance, by elaborating a spreadsheet, in which the all probabilities p_k can be easily obtained.

2.3. Example of the queuing network solution

Once basic reliability parameters of a system are known, which can be obtained by theoretical surveys or by field data collection, the questions mentioned in item 2.1 can be answered using the model described in 2.2. For example, in the case study cited (Menascé et al., 2004) it was considered a data center with 120 machines (*M*=120), with mean time between failures of 500 minutes (MTTF=500 min., that is, λ=0,002 failures per minute) and an average time to repair of 20 minutes

(MTTR=20 min., or μ=0,05 repairs per minute). The solution for the queuing model enables to answer question I, as follows:

I. The probability that exactly *j* machines are operational at any given time is the probability p_j that *M-j* machines are failed, which can be calculated using Equation 4, resulting the graphic example shown in Figure 4, as follows:

- If only two technicians are available (N = 2), P_j is negligible for $j < 34$ or $j > 67$ and the peak of the distribution occurs for about 50 machines, with $P_{50} = 5{,}6\%$;
- If the number of technicians increases to five (N = 5), the situation improves dramatically, resulting that the relevant values for P_j are concentrated between 92 and 120 machines, with the peak at 116 machines, where $P_{116} = 10\%$;
- If N = 10, a better situation can be observed, with relevant values for P_j concentrated between 108 and 120 machines, with the peak at 116 machines and $P_{116} = 19\%$.

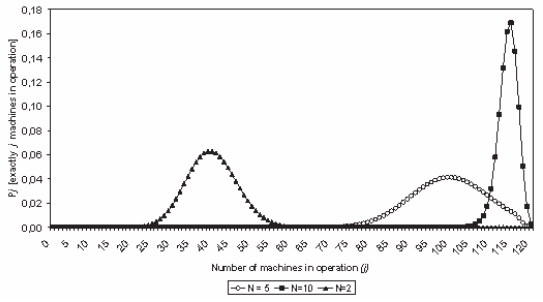


Figure 4: Probabilities (Pj) to have exactly j machines in operation (Menascé et al., 2004)

3. APPLYING THE QUEUING MODEL TO THE AIRSPACE CONTROL SYSTEMS

The automation systems used in airspace control centers have their basic functional structures defined by international organization standards as ICAO and FAA, mentioned previously, as well as the European Organization for the Safety of Air Navigation (EUROCONTROL) and, in Brazil, some equivalent government organizations, as the Department of the Airspace Control (DECEA) at the Brazilian Air Force Command and the National Agency of Civil Aviation (ANAC).

3.1. Estrutura das organizações de controle do tráfego aéreo

Air traffic control services (ICAO, 1996; DEPV, 1999b) are performed by operational structures whose hierarchies are established, basically, with the four levels of control illustrated by Figure 5, and described as follows:

a) Tower Control level (TWR), where local

management of landings and take-offs are performed in an aerodrome;

b) Terminal Area level (APP), where the air traffic control of a terminal area takes place, managing approach procedures of the aircrafts for landing, as well as their departures, from take-offs to en-route flights;

c) Area Control Center level (ACC), responsible for the control functions of the aircrafts flying through the airways, both for the national and international routes;

d) Air Traffic Flow Management level (AFTM), equivalent to the strategic level analysis executed by the Brazilian organization Center for the Management of the Air Traffic (CGNA), responsible for the optimization of operational flows, also involving the long term planning issues, relative to the future flight demands.

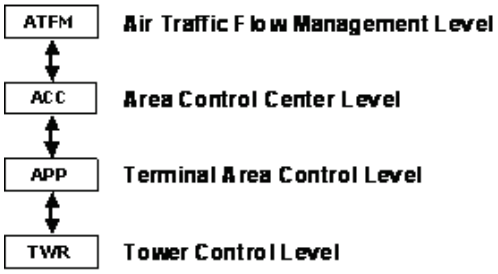


Figure 5: Hierarchical levels of the air traffic control organizations

To each described level there is a correspondent time scale, in which are included the process controls in real time, ranging from the sponsoring and monitoring processes involved in decisions made in seconds or minutes, at the Tower and at APP levels; through the control of en-route operations, also involving real time operations of some hours, at the ACC level; and at the ATFM strategic level, considered the operational record, the elaboration of statistical data and activity planning for days or even for months.

Likewise, military systems responsible for the air defense and airspace vigilance are organized in similar hierarchical levels, from the command and control operations at the central strategic level, to the local systems, intended to perform control and support of some aerial military operations (FAA, 2004).

3.2. Exemplifying some Modes of Operation upon an APP

International normative organizations establish the recommended practices and procedures for providing the services of air traffic control (ICAO, 1996) as well as for coordination of military special operations (FAA, 2004), in order to guarantee safety and continuity of these services, even under any degraded situation, during eventual failures or transient unavailability of some support services

(DEPV, 1999a).

As example of the foreseen degraded operational situations, an eventual unavailability of the data links between an ACC and an APP, when the APP would start to operate in its called autonomous mode, or degraded mode. In the normal situation case, when all the communication systems are available, characterizes the operation of an APP in its normal mode. Such situations, for instance, affect the way that flight plans are entered into the system, assuring operational continuity in the occurrence of any mentioned unavailability.

While in the normal mode, the usual method to input flight plans into the system is the automatic one. Once in the autonomous mode, there is no automatic creation of flight plans, and this function is started manually. The transition from the normal mode to the autonomous mode demands the adoption of some operational measures, to assure the continuity of the services by the responsible organization (APP), as follows:

- Manual flight plans input into the system, performed by a flight plan operator;
- Automatic extraction of repetitive flight plans, from a computer file (RPL). The operations supervisor verifies the convenience of using this file, completely or partially;
- Operational coordination with the corresponding ACC starts to be carried on through another communication line.

The situation previously described exemplifies a type of function that is usually performed by an automatic process into the computational system, while operating in its normal mode. When the system enters in a degraded operational mode, the same exemplified function would start to demand some additional human intervention from the controllers.

Therefore, using the same model first presented, in which it was considered only the availability of each machine – and once any of them entered into a failure a new maintenance service was required – analogously, the model could also consider an equivalent situation in which each machine performs an automatic function that, if it happens to be temporarily unavailable, this machine demands a new human intervention with appropriate skills for that function.

3.3. Application example for the queuing model in a Terminal Area Center (APP)

From the same model used in the example of the data center mentioned previously, it can be established an equivalent queuing model, as illustrated by Figure 6, in which is considered a

total of M machines in normal (or automatic) operation, where a subset of these machines, when in a degraded (or manual) operation state, requires the intervention of D operators. These operators are the ones necessary to perform some kind of extra manual action, in substitution of any automatic system function temporarily unavailable, as previously exemplified, regarding the case of manual introductions of flight plans into an APP operation.

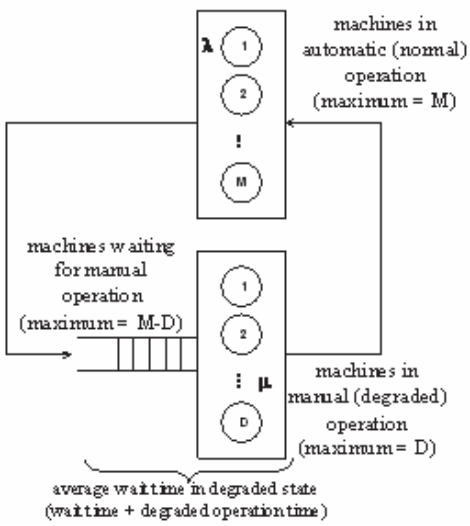


Figure 6: Queuing model for a human based operation (example: a Terminal Area Center)

As example, in this case that a hypothetical control center with 120 machines was considered (M=120), representing the equivalent amount of the positions with access to the automatic functions. In this example, they admit: a) a mean time between failures of 80 hours (λ=0,0125 failures/hour), to characterize the rate at which any automated function is lost; and b) an average time to return to the normal operation is considered to be 4 hours, that is, the average total time spent to wait the operator plus the entire length of a manual operation, which would be equivalent to the additional occupation time for an extra operator (μ=0,25 would represent the recovery rate from the degraded or manual operation).

Then, the solution of the queuing model can be obtained in the same way presented in item 2.3, also answering the questions II and III, as follows:

II. The probability that at least j machines are in normal (automatic) operation can be calculated with the accumulated values, in an analogous solution of the one in question I, as illustrated by Figure 7;

III. The number D of additional operators, necessary to guarantee that the manual operations are executed (degraded situation), causing that at least j machines to return to normal operation (automatic), can also be verified on the basis of the accumulated probabilities data, as illustrated by Figure 7, where the following relevant aspects are observed:

- For *D*=2, the probability of having at least 30 operational machines is of 96%, whereas to have at least 40 machines in the operational status this confidence falls to 53% (as shown by the black arrows in Figure 7);
- If a confidence level of 90% is required, the conditions to assure the number of machines in normal (or automatic) operation are verified as follows: a) from an extra team of 4 operators (*D*=4), it is observed that only 69 machines could be guaranteed to be in the normal state; b) for *D*=5, 88 machines are normal with the confidence of 90%; and c) 112 machines are in the normal status, with 90% confidence, for a staff of 10 additional operators (as shown by the gray arrows in Figure 7).

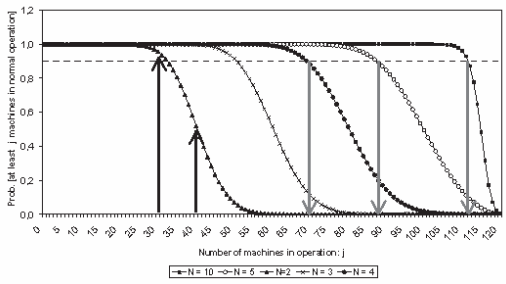


Figura 7: Probabilities (Pj) to have at least j machines in normal operation

3.4 – Proposed extended model for availability analysis of a Terminal Area Center

Considering the previous model and the possibility of degraded operation existing in the real computational system in air traffic control systems, an extended model can be established, integrating the two queuing nets, in order to represent the transition among the three states: normal operation, degraded operation, and failed machine.

Therefore, this queuing net model can be extended to the configuration illustrated in Figure 8. Thus, the model considers not only the effect of the size N of the maintenance staff, but also the effect of the size D, regarding the team of extra available operators, who must be prepared for the execution of any manual operation necessary, attending degraded situations, when some of the automatic processes happen to be temporarily unavailable.

In this model, the rate λ represents the flow of machines that leave the normal operation status, corresponding to the addition of the flows $\lambda.p_1$ and $\lambda.p_2$, referring to the transition from the normal (automatic) status to the failed one, with probability p_1, or from de normal status to the degraded (manual) operational situation, with probability p_2. To

return from the maintenance status to normal operation occurs with the repair rate μ_1, while μ_2 represents the rate that machines leave the degraded operation status, passing from the manual operation status to the normal automatic condition, with probabilities p_3 and p_4 respectively, thus composing the flows $\mu_2.p_3$ and $\mu_2.p_4$.

Based on this new model, whose solution is analogous to the one presented in item 2 of this work, a Mean Value Analysis (MVA) can be carried through (Menascé et al., 2004), establishing the influences of the operation and the maintenance teams in the system availability, in this case considering more than one of the degraded status.

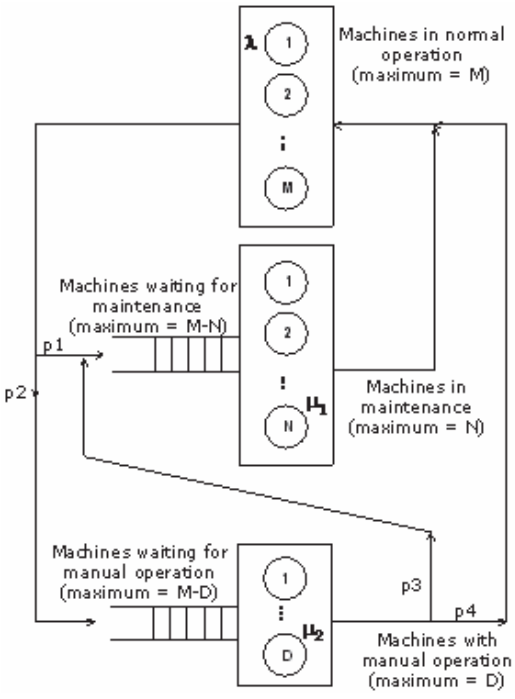


Figure 8: Availability Model of a Terminal Area Center (considering degraded operation)

The importance to consider the possibility of degraded operations in this type of system can be justified by the risk classification imposed by the software usage, which is defined as a function of the control level executed by this software in the specific process. In this case, regarding a computer system operating into an airspace control center, some descriptions are found in the families of standards related to safety management in defense applications (Herrmann, 1999). Thus, an air traffic control system could be classified in the III-b category of control, as mentioned in MIL-STD-882 standard practice (DoD, 1998), where "software generates information of a safety-critical nature, used to make safety-critical decisions. There are several redundant independent safety measures for each hazardous event".

4. CONCLUDING REMARKS

From the concepts and the availability model for computer centers presented in this work, based on the use of analytical methods from the queuing theory, numerical applications can be developed to solve important questions of interest for the management of these centers.

The intended contribution of this work is to present a simple tool for availability analysis. From field data collection, concerning the

measurements of real repair and failure rates in the operational airspace control centers, this technique can be adopted in the early stages of the systems engineering and reliability requirements designs, during the specification and the acquisition phases of the computational systems used in these centers.

Therefore, this work becomes an initial tool applicable, for example, in the development of studies for operational and maintenance staff sizing of airspace control centers, whose parameters orient the reliability requirements, including the technical characteristics of support systems (hardware, software applications, characteristics of the commercial infrastructure used, etc.), and the operational aspects, regarding the number of necessary control positions (operational consoles).

Similarly, other applications for specific cases could be considered, where eventual limitations from physical, technical, or economic nature determine small increments on the reliability indexes. In these cases, the model could also be applied to determine the ideal size of the operational and maintenance staff for this type of computational systems, in which human operation are mandatory.

ACKNOWLEDGMENTS

The authors thank the professionals and organizations of DECEA at Brazilian Air Force, specially to Lieutenant Colonel Engineer José Antonio da Motta Matinha (CINDACTA III), for the understanding contributions on specific operational and technical topics, as well as for the incentive to the accomplishment of this work.

REFERENCES

DEPV (1999a) *DMA 63-1 - Telecomunicações Aeronáuticas e Controle do Tráfego Aéreo: Diretrizes Básicas para Situações de Degradação*. Diretriz do Ministério da Aeronáutica, Diretoria de Eletrônica e Proteção ao Vôo, Rio de Janeiro.

DEPV (1999b) *IMA 100-12 - Tráfego Aéreo: Regras do Ar e Serviços de Tráfego Aéreo*. Instrução do Ministério da Aeronáutica, Diretoria de Eletrônica e Proteção ao Vôo, Rio de Janeiro.

DoD (1998) *MIL-STD-882D - Mishap Risk Management (System Safety)*, U.S. Department of Defense Standard Practice, Washington, DC, USA.

FAA (2004) *Order* 7610.4K - *Special Military Operations* – U.S Department of Transportation, Federal Aviation Administration, Washington, DC, USA.

FAA (2006) *Order* 7210.3U - *Facility Operation And Administration* – U.S Department of Transportation, Federal Aviation Administration, Washington, DC, USA.

Herrmann, D. (1999) - *Software Safety and Reliability* (Chapter 5 – *Defense Industry - MIL-STD-882D / DEF STAN 00-55*) – IEEE Computer Society Press, Los Alamitos, CA, USA.

ICAO (1996) *Doc 4444-RAC/501 – Rules Of The Air And Air Traffic Services – Procedures For Air Navigation Services.* International Civil Aviation Organization, Montreal.

Menascé, D., Almeida, V. e Dowdy, L. (2004) *Performance by Design* (Chapter 7 – *Case Study III: A Data Center;* Chapter 10 – *Markov Models / Generalized Birth-Death Models; Chapter 12 – Single Class MVA; Chapter 13 – Queuing Models with Multiple Classes*). Prentice Hall PTR, NJ, USA.

Shooman, M. L. (2002) *Reliability of Computer Systems and Networks* (Appendix B6 – Markov Reliability and Availability Models). John Wiley and Sons Inc., New York, USA.