Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Performability Evaluation of EFT Systems using Expolinomial Stochastic Models

Erica Sousa, Paulo Maciel, Carlos Araújo
Center of Informatics
Federal University of Pernambuco
Recife, PE, Brazil
etgs,prmm,cjma@cin.ufpe.br

***Abstract*—The performance evaluation of EFT Systems has paramount importance for Electronic Transactions providers, since the computing resources must be efficiently used in order to attain requirements defined in Service Level Agreements (SLA). Among such requirements, we may stress agreements on availability, reliability, scalability and security. In the EFT system, faults can cause severe degradation on system performance, and as a result modeling can lead to incomplete or inaccurate results ignoring the dependability effects on performance. This paper presents an expolinomial stochastic model for evaluating the performance of the EFT system processing and storage infrastructure considering a load variation range. This work also presents a model for evaluating the dependability of the EFT system infrastructures and combines both models (dependability and performance) for evaluating the impact of dependability issues on the system performance. The performability analysis employs a hierarchical decomposition method aiming at reducing the computational effort and avoiding stiffness problems.**

***Index Terms*—Performability Analysis, Expolinomial Stochastic Model, Performance Model, Dependability Model, Electronic Funds Transfer System**

I. INTRODUCTION

The increasing computational capacity and integration of payment services as well as the advances of new technologies have fostered the growth and complexity of electronic transactions. Organizations promoting Electronic Funds Transfer (EFT) must not only supply correct services, but also meet the customers' performance expectations.

Over the last decade, the EFT market has been massively expanding, thus demanding companies to offer reliable services, high availability and security at affordable costs. Furthermore, performance and dependability evaluations have also been widely adopted as an essential activity for improving the quality of services provided, infrastructure planning, and for tuning components of systems in order to improve the overall performance and reducing service costs. In addition, modeling based techniques may represent systems at several abstraction levels, each of them more suitable for solving specific problems [7], [15].

In systems as EFT systems the fault and recovering actions of a specific component do not necessarily affect the overall service provided by the system, but do affect the performance. Hence, in such systems the performance analysis considering faults is of major importance. The integrated modeling of performance and dependability aspects of the systems' computational resources known as performability modeling permits performance evaluation considering degraded levels of services due to faults [6].

Performance and dependability are often modeled separately based on the assumption that the individual component or subsystem faults do not necessarily affect performance. This assumption is invalid for many systems that have either recovering or fault-tolerant mechanisms. Such mechanism aims at continuously providing the specified service even though performance levels might be reduced. Ignoring fault effects on degradable systems may lead either to incomplete or inaccurate performance results. A performability model might be represented by a dependability model, a performance model, and a method of combining the respective results [2].

Different works have been proposed for performance and dependability analysis, but, to the best of our present knowledge, none of them focus on performance degradation effects of EFT systems represented by expolinomial stochastic models [9]. Sesmun [4] et al propose a technique that uses performability in designing communication networks aiming at deriving a design methodology for fault-tolerant networks. The results show that using performability for supporting decisions on topology at the end of each iteration definitely generates a more reliable network. Hellerstein [11] proposes a performance and availability model for application providers using a server cluster in which any of the servers it owns can respond to an application request; a business-oriented cost model for allocating application resources providers, and a closed form approximation for servers allocation based on the proposed performance model.

The performability evaluation strategy conducted is based on a hierarchical modeling approach that combines results from dependability and performance models for supporting capacity planning and SLA assurance. The proposed models focus on storage and processing resources which are represented by Stochastic Petri Nets (SPN) [8]. The obtained results allow evaluation of performance figures, such as throughput and utilization degree for capacity specification of EFT infrastructure. The adopted approach takes into account the points of sales number and transaction frequency besides considering the effects on performance related fault occurrences.

This paper is structured as follows: Section 2 presents few fundamental basic concepts. Section 3 presents the EFT System Performance Model. The EFT System Dependability

978-1-4244-2794-9/09/$25.00

Model is presented in Section 4. Section 5 presents the EFT System Performability Evaluation. Experimental Results are given in Section 6. Section 7 presents the concluding remarks.

II. PRELIMINARIES

Generalized Stochastic Petri Nets are high-level formalism widely applied for performance and dependability evaluation. GSPN has been extensively applied for both automatic generation of Markov Chains and Stochastic Simulation. GSPNs are derived from standard Petri nets [17] by partitioning the set of transitions into two subsets comprising of timed and immediate transitions. An exponentially distributed random firing time is associated with each timed transition, whereas immediate transitions fire in zero time. It is shown that GSPNs are equivalent to continuous-time stochastic processes, and solution methods for deriving the steady state of probability distributions are presented in [8], [14].

GSPNs models consider only exponentially distributed timed transitions and immediate transitions. The immediate and timed transition model exponential and immediate actions, activities and events.

Phase approximation technique can be applied for modeling non-exponential activities. A variety of performance and dependability activities can be constructed in GSPN models by using throughput subnets. These throughput subnets represent expolinomial functions, such as the Erlang and Hypoexponential distributions [1].

Measured figures (empirical distribution) from a system with an average μ_D and a standard deviation σ_D must adjust their stochastic behavior through the phase approximation technique. The inverse of the coefficient of variation of the measured figure (Equation (1)) allows the selection of which distribution matches it best. For a deeper understanding on moment matching and expolinomial distributions, the reader is referred to [9] [16]. In this work, the adopted distributions for moment matching are: Erlang and Hypoexponential.

$$\frac{1}{CV} = \frac{\mu_D}{\sigma_D} \tag{1}$$

When the inverse of the coefficient of variation is a whole number and different from one, the empirical figure should be characterized by an Erlang distribution, represented in GSPN by a sequence of exponential transitions whose length is calculated by Equation (2). The rate of each exponential transition is calculated by Equation (3). The Petri Net model depicted in Figure 1 represents an Erlang distribution.

$$\gamma = (\frac{\mu}{\sigma})^2 \tag{2}$$

$$\lambda = \frac{\gamma}{\mu} \tag{3}$$

When the inverse of the coefficient of variation is a number greater than one (but not an integer), the empirical figure is represented by a hypoexponential distribution which is represented by a GSPN composed of a sequence whose length is calculated by Equation (4). The transition rates of exponential transitions are calculated by Equations (5) and (6) where the respective average time (expected value) assigned to the exponential transitions are calculated by the equations (7) and (8). The model presented in Figure 2 is a net that depicts a hypoexponential distribution.

$$(\frac{\mu}{\sigma})^2 - 1 \leq \gamma < (\frac{\mu}{\sigma})^2 \tag{4}$$

$$\lambda_1 = \frac{1}{\mu_1} \tag{5}$$

$$\lambda_2 = \frac{1}{\mu_2} \tag{6}$$

$$\mu_1 = \mu \mp \frac{\sqrt{\gamma(\gamma+1)\sigma^2 - \gamma\mu^2}}{\gamma+1} \tag{7}$$

$$\mu_2 = \gamma\mu \pm \frac{\sqrt{\gamma(\gamma+1)\sigma^2 - \gamma\mu^2}}{\gamma+1} \tag{8}$$

III. EFT SYSTEM PERFORMANCE MODEL

The electronic funds transfer system is composed of client and management applications. The client application is responsible for interfacing points of sales and management application. The management application controls the whole transaction process. The transaction process phases are message displaying, reading the magnetic card, password collections and coupon printing. In addition, the management application gets all requests generated by the points of sales including all parameters needed for the construction of messages to be sent by the terminals and then forwards the transaction.

This section presents the GSPN model conceived for EFT system performance evaluation (See Figure 3). The proposed EFT model is described via its "sub-models" (subnets that describes system's components). Client subnets represent EFT client applications in different companies, such as drugstores, supermarkets etc or even workload related to a particular period or season. These subnets could be refined to represent a vast range of traffic. These models might represent a large number of points of sale forwarding trade transactions with distinct occurrence frequencies. Hence, service demands can be represented by varying the transfer transaction frequencies and other parameters. The markings N_i assigned to places drugstores, supermarkets etc on client subnets specify the number of points of sale of respective particular type, and the generic stochastic transitions represent the delay distribution between transactions. The place buffer (Buffer subnet) represents the temporarily hold transactions waiting to be served. Its dual place marking $(M(P10))$ represents the buffer storage capacity.

The management model represents the server processing and storage infrastructure where the management applications are configured. The management model is composed by the Processing Transaction and Disk Transaction subnets. The Processing Transaction subnet represents the transaction's processing and the Disk Transaction subnet represents the disk reading

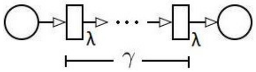


Fig. 1. Erlang Distribution

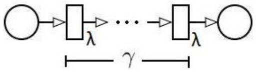


Fig. 2. Hypoexponential Distribution

and writing operations related to transactions. The place Processor marking Np denotes the processing capacity, that is the number of concurrent transactions supported by the processing resource. The place Disk marking Nd denotes the number of concurrent disk transactions supported by the storage resource. The stochastic transitions T_ip and T_id (throughput subnets) firing represent the transaction processing time and the storage (reading and writing related to a transaction) operation time (See Figure 3).

This model supports performance analysis and planning of EFT systems by evaluating processing and storage infrastructure utilization levels considering a given load range. Such an evaluation aims to provide means for deciding about suitable configurations taking into account further customers demands, fluctuations and attaining assured service levels.

A. EFT System Performance Model Validation

This section presents a case study conducted for validating the proposed model. The adopted system, called SCOPE, is a system that manages the whole EFT process, which is accomplished by means of transactions between the points of sale and credit and debit cards authorizers. In addition, the SCOPE system manages the operations of subsidiaries, controls the status of shops, points of sale and of message exchanges between the points of sale and the authorizers. The EFT system is composed of client and management applications [10].

In order to validate the EFT System Performance Model proposed, experiments were conducted in an environment set in the CIn-Itautec Performance Evaluation Laboratory considering real transaction traces. The client applications were installed in a range of MX201 servers. These client applications forward real transaction traces collected from EFT system users. These traces were carefully analyzed and chosen to represent workloads of particular interest. The management application was configured in a MX221 server.

The client applications were configured to represent 3345 points of sale distributed in a Shopping Center. The points of sale register the transactions through client applications. The evaluated scenarios describe a Shopping Center with points of sale where the demand register occurred at 7 different rates, which are 100, 200, 300, 400, 500, 600 and 700 tpms (transactions per minute). This range corresponds to the boundary of interest related to the business scenario of study.

The measures obtained on the server through the Windows Performance Monitor (Perfmon) are: processor and disk idle time ratios, disk transfers per second and disk average transferring time [13]. Some metrics as service time for disk and processor are indirectly estimated [12], [15].

The measured data was analyzed for deciding which expolinomial distribution best fits the processing and storage operations (represented by transitions Tip and Tid). The respective processing and storage time means (μ_D) and standard deviations (σ_D) were calculated and the distributions chosen, according the process described in Section 2. These transitions were refined according the results presented in Table I.

TABLE I
AVERAGE AND STANDARD DEVIATION

Service Time	μ_D (s)	σ_D (s)	Suitable Distribution
Processor	0.001311	0.000508	Hypoexponential
Disk	0.002756	0.000353	Hypoexponential

After defining which distribution is suitable for representing the measured figure, the related distribution parameters have to be calculated. Since the hypoexponential model was chosen for refining both Tip and Tid, μ_1, μ_2 and γ should be computed. These values are calculated using Equations 4, 7 and 8. Table II shows the respective values of μ_1, μ_2 and γ for the models that refine Tip and Tid. Then, a refined version of the EFT System Performance Model is generated.

TABLE II
DISTRIBUTION PARAMETERS

Transition	μ_1 (s)	μ_2 (s)	γ
T_ip	0.000080	0.00054	6
T_id	0.000001	0.00005	61

The processor and disk utilization levels are obtained using the following expressions: $UProc = P\{\#Processor = 0\}$ and $UDisk = P\{\#Disk = 0\}$ [5], respectively. Figure 4 shows a visual comparison between the results obtained by stationary analysis and measurements, considering a set of different trade transaction frequencies: 100, 200, 300, 400, 500, 600 and 700 tpms (transactions per minute). The Figure 4(a) shows a visual comparison between measured processor utilization levels and values obtained through the model by the metric $UProc = P\{\#Processor = Np\}$. The result presents a maximum relative error of 13.60%.

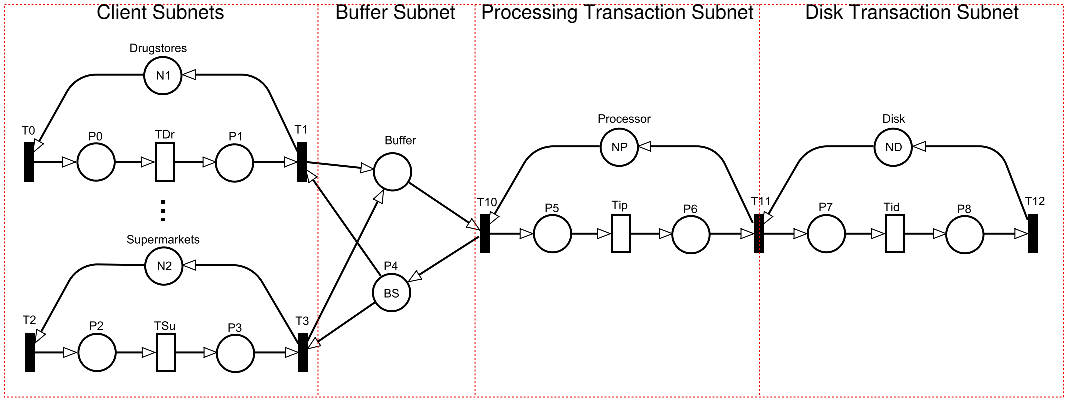


Fig. 3. EFT System Performance Model

The Figure 4(b) depicts similar comparison related to disk utilization levels. The values presented are those obtained by measurements and by the evaluation of the model through the metric $UDisk = P\{\#Disk = Nd\}$. The result presents a maximum relative error of 11.65%.

The validation process is conducted by comparing the metric results obtained through the EFT System Performance Model against measurements taken on the Scope EFT System. After validating the EFT System Performance Model, it is ready for estimating the performance indices of a EFT system under several conditions and settings [7].

IV. EFT SYSTEM DEPENDABILITY MODEL

This section presents the GSPN model conceived for dependability evaluation of the EFT system. The proposed EFT model is described via its "sub-models" (subnets that describe EFT system infrastructure). The Processing Transaction and Disk Transaction subnets represent faults and repairing activities related to processors and disks in the EFT systems. The Figure 5 depicts the EFT System Dependability Model.

The place Processor marking Np denotes an operational processor and a token in the place $PFailed$ indicates that the processor has failed. The generic stochastic transition $PFailure$ represents a fault and its delay represents the MTTF (mean time to failure). Likewise, transition $PRepair$ represents the repairing activity, and its delay the related to MTTR (mean time to repair). The place Disk marking Nd denotes the operational disk and one token in the place $DFailed$ indicates a failure.

The place Team marking (Nt) represents the maintenance team specific type. The number of tokens in the place Team directly affects the concurrency degree related to repairing process. If distinct teams with distinct skills are considered, those teams are represented by specific places, and their initial markings depict the number of available teams of each particular expertise or proficiency. The Periodic Maintenance subnet represents the intervals between maintenance. The place Maintenance indicates the maintenance number before a device replacement. The generic stochastic transitions $MTBM$ and $MTBR$ represent the mean time between maintenance and the mean time between replacements, respectively. The maintenance policy may give higher priority to recover processing infrastructure than storage subsystems, since storage infrastructure usually have a higher number of storage devices than process units, hence allowing longer maintenance delays.

The processing and disk operational states are depicted through the following expressions: $UPProc = P\{\#Processor = 1\}$ and $UPDisk = P\{\#Disk = 1\}$ [5].

This EFT System Dependability Model aims evaluating device and system failures; the respective effects on the overall EFT system as well as planning the maintenance team dimension and policies for assuring contracted availability levels.

V. EFT SYSTEM PERFORMABILITY EVALUATION

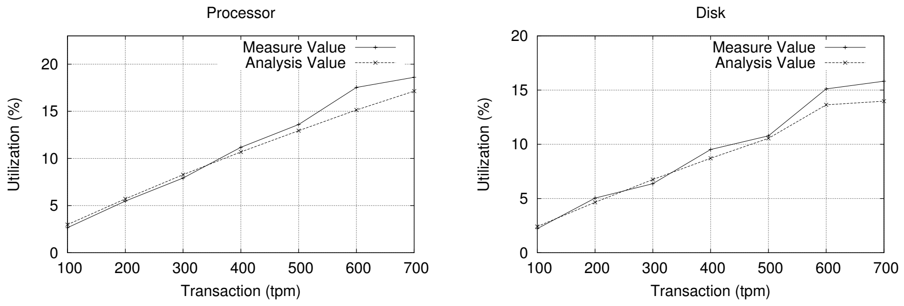
The performability evaluation of the EFT system is conducted through the EFT system dependability and performance models. The decomposition approach, which splits the model into two separate models [6], aims at handling largeness and stiffness problems [9].

The largeness problem is a consequence of the state space size of the model and the stiffness problem is related to the different orders of magnitude between the occurrence rates of performance-related actions, fault events and repairing activities. The decomposition technique combines a higher level EFT system dependability model, which takes faults and component repairing process into account, and a set of a lower-level EFT system performance models, one for each state of the EFT system dependability model [3] [16].

The performability evaluation of the EFT system infrastructure depicts the effect of dependability on degradable-system performance through the dependability and performance metrics. Intermediate metrics are independently calculated from the dependability and performance models, then combined to show the effect of dependability on performance.

VI. EXPERIMENTAL RESULTS

The system described in Section 3 illustrates the approach studied in this paper. The performance experiments describe a Shopping Center containing various sale points that register transactions. The transaction rates adopted in this scenario



(a) Processor Utilization

(b) Disk Utilization

Fig. 4. EFT System Performance Model Validation

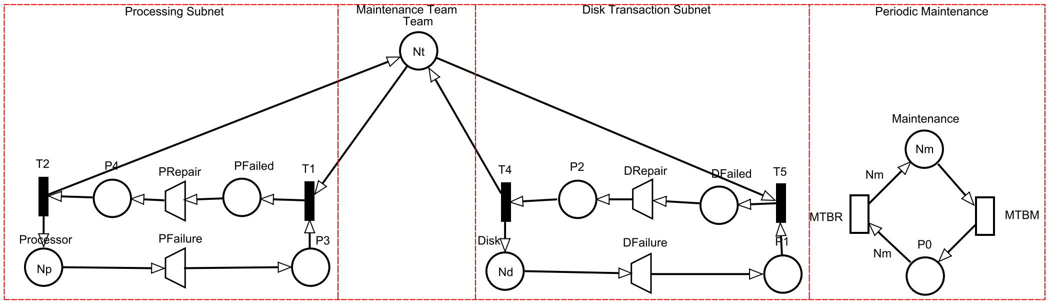


Fig. 5. EFT System Dependability Model

were $3,500$, $4,500$, $4,900$, $5,600$, $6,300$, $7,200$, $8,100$, $9,801$, $14,850$ and $16,830$ tpms (transaction per minute). Figure 6(a) depicts the processor and disk utilization. These results show the effect of workload fluctuations on processor and disk utilization considering each trade transaction frequencies. The processor and disk utilization are obtained using the following reward expressions: $UProc = (P\{\#Processor = 0\})$ and $UDisk = (P\{\#Disk = 0\})$ [5], respectively.

The dependability experiments consider that the mean time to failure (MTTF) of application processing server is $17,520$ hours and its mean time to repair (MTTR) is 16 hours. Likewise, the MTTF of storage server is $43,800$ hours and the MTTR is 18 hours. The mean time between maintenances (MTBM) is $4,380$ hours and the mean time between replacements (MTBR) is $43,800$ hours.

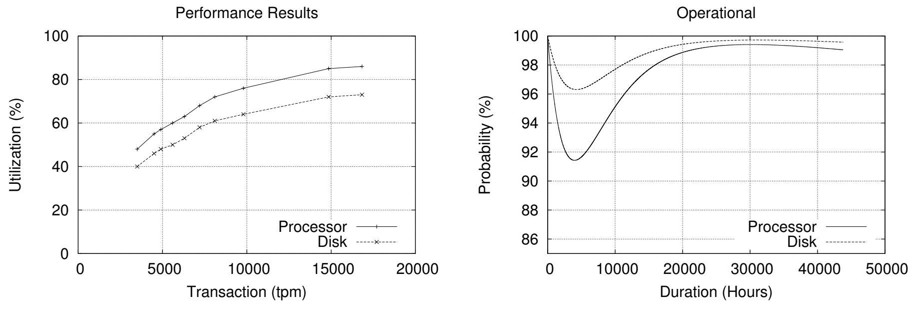
The Figure 6(b) shows the probability of the processing and storage units being in operational states if one maintenance team is allocated for carrying out repairing activities over a period of $43,800$ hours. These results are obtained through reward expressions $UPProc = (P\{\#Processor = 1\})$ and $UPDisk = (P\{\#Disk = 1\})$, respectively. The dependability results show the impact of the fault occurrences and maintenance activities on EFT system availability. The availability results take into account the team that is performing maintenance, the interval between maintenance and the interval between replacement of a piece of equipment.

Faults in the processing and storage infrastructure may affect EFT system performance. The performability results are obtained by combining the results related to the processing and storage units availability and the respective resource utilization, considering a set of different trade transaction frequencies: 3500, 4500, 4900, 5600, 6300, 7200, 8100, 9801, 14850 and 16830 tpms. These results present the processing and storage infrastructure performance degradation. Figures 7(a) and 7(b) show the impact of fault occurrences and maintenance activities on the processor utilization and disk utilization through the combination of dependability results and a set of performance results, considering a set of different frequencies of commercial transactions. These results can be adopted for estimating the performance degradation, avoiding the decrease of the service quality.

VII. CONCLUSIONS

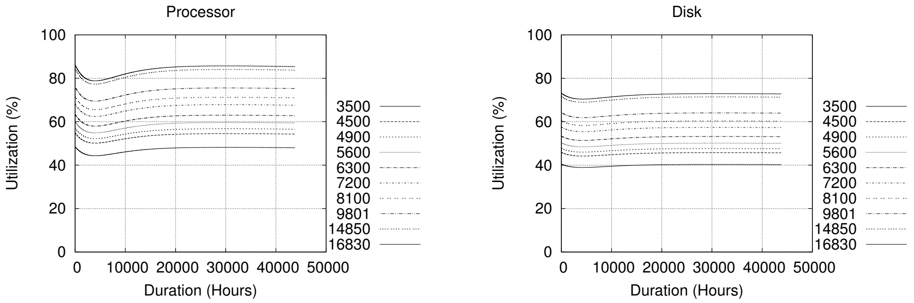
This paper proposes a performability model and employs a hierarchical method for performability analysis of EFT systems. The adopted method aims at evaluating EFT systems performance considering fault events and recovering activities; and also aims at tackling largeness and stiffness problems. The adopted process takes into account a set of significant structural EFT system states, each state corresponding to a configuration that results into a particular EFT system performance.

The performability analysis is carried out through a dependability and a performance model. The EFT system performance model allows the analysis for sustaining the service quality and preventing performance degradation related to the workload fluctuations. The dependability model allows the evaluation of structural variations of EFT systems.



(a) Performance Results (b) Dependability Results

Fig. 6. Performance and Dependability Results



(a) Processor Utilization (b) Disk Utilization

Fig. 7. Performability Results

The adopted performability analysis and the proposed EFT System expolinomial stochastic models were applied to an EFT system named SCOPE. The experiments were conducted considering real transactions traces collected from EFT systems users. The performability evaluation results combined probability resources (processing and storage units) in operational states, their utilization, considering a set of different trade transaction frequencies. It is important to stress that evaluating these experiments without models is a complex and expensive task.

As future works, we intend to detail the training process and duration of the diagnosis process. We also intend to adjust the mean time to repair, the maintenance team sizing, geographical distribution and the related costs in order to ensure the SLA's clauses.

REFERENCES

[1] A. Desrochers and R. Al-Jaar. Applications of Petri Nets in Manufacturing Systems: Modeling, Control, and Performance Analysis. IEEE Press, 1995.

[2] A. L. Reibman. Modeling the Effect of Reliability on Performance. In SIGMETRICS, page 213, 1989.

[3] A. Puliafito, S. Riccobene, and M. Scarpa. Evaluation of Performability Parameters in Client-server Environments. Comput. J., 39(8):647662, 1996.

[4] A. Sesmun and L. Turner. Using Performability in the Design of Communication Networks. In Computers and Digital Techniques, IEE Proceedings, Volume 147, pages 355363, 2000.

[5] A. Zimmermann and M. Knoke. A Software Tool the Performability Evaluation with Stochastic and Colored Petri Nets. Technische Universitt Berlin. Real-Time Systems and Robotics Group, 2007.

[6] B. Haverkort, R. Marie, G. Rubino, and K. S. Trivedi. Performability Modelling : Techniques and Tools. 2001.

[7] D. Menascé and V. Almeida. Performance by Design: Computer Capacity Planning by Example. Prentice Hall PTR, 2005.

[8] G. Balbo. Introduction to Stochastic Petri Nets. Lectures on Formal Methods and Performance Analysis: First EEF/Euro Summer School on Trends in Computer Science, Berg en Dal, The Netherlands, July 3-7, 2000: Revised Lectures, 2001.

[9] G. Bloch, S. Greiner, H. de Meer, and S. Trivedi. Queueing networks and markov chains: Modeling and performance evaluation with computer science applications, 1998.

[10] Itautec. Reference Guide to the Scope - Electronic Payment System. 1999.

[11] J. Hellerstein, K. Katircioglu, M. Surendra, I. Center, and N. Yorktown Heights. An On-line, Business-oriented Optimization of Performance and Availability for Utility Computing. Selected Areas in Communications, IEEE Journal on, 23(10):20132021, 2005.

[12] K. Trivedi. Probability and Statistics with Reliability, Queuing, and Computer Science Applications. John Wiley and Sons, 2nd Edition, 2006.

[13] M. Friedman. Microsoft Windows Server 2003 Performance Guide. Microsoft Press Redmond, WA, USA, 2005.

[14] M. Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis. Modelling with Generalized Stochastic Petri Nets. ACM SIGMETRICS Performance Evaluation Review, 26(2), 1998.

[15] R. Jain. The Art of Computer Systems Performance Analysis: Techniques For Experimental Design, Measurement, Simulation, and Modeling. John Wiley and Sons, New York, Chichester, 1991.

[16] R. Sahner, K. Trivedi, and A. Puliafito. Performance and Reliability Analysis of Computer Systems: An Example-Based Approach Using the SHARPE Software Package. Kluwer Academic Publishers, 1996.

[17] T. Murata. Petri nets: Properties, Analysis and Applications. Proceedings of the IEEE, volume 77, number 4, pages 541, 1989.