MAINTENANCE POLICY AND ITS IMPACT ON THE PERFORMABILITY EVALUATION OF EFT SYSTEMS

Erica Sousa[1,2], Paulo Maciel[1], Fernando Lins[1,2] and Marcelo Marinho[1,2]

[1]Federal University of Pernambuco, Center of Informatics, Pernambuco, Brazil

www.cin.ufpe.br

[2]Federal Rural University of Pernambuco, Pernambuco, Brazil

http://www.ufrpe.br/

ABSTRACT

In the Electronic Funds Transfer (EFT) Systems, faults can cause severe degradation on the performance of this system. Thus, modelling the performance of EFT system without considering dependability aspects can cause inaccurate results. This paper presents a stochastic model for evaluating performance of processing and storage infrastructures of the EFT system. This work also presents a model for evaluating the effects of the proposed preventive maintenance policy and different service level agreements (SLA) on the dependability of the EFT system infrastructure. Then, this paper combines both models (dependability and performance) for evaluating the impact of dependability issues on the performance of the EFT system. Finally, case studies considering EFT system infrastructures are provided to demonstrate the applicability of the adopted approach. Moreover, the results of these case studies are depicted, stressing important aspects of dependability and performance for EFT system planning.

KEYWORDS

Performability Evaluation, Maintenance Policy, Service Level Agreements, Expolinomial Stochastic Model, Performance Model, Dependability Model

1. 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) are not only required to supply correct services, but also meet the performance expectations of customers. Over the last decade, the EFT market has been massively expanding, thus demanding companies to offer reliable services, high availability, scalability 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 the components of the system in order to improve the overall performance and reducing the services cost [1,2].In EFT systems, fault events and recovering actions of a specific component will surely affect its performance. In this context, maintenance activities play a very important role in sustaining performance and availability levels needed to guarantee the approved service level. Hence, maintenance strategies have an imperative impact on the availability of the systems, fault prevention, and on the overall cost of the services provided by the system. Thus, the performance evaluation of these systems

DOI : 10.5121/ijcsea.2012.2208

considering faults, failures [3] and maintenance strategies is an issue of major importance for attaining contracted service level agreements (SLA) [4,5].

The integrated modeling of the performance and dependability aspects of the systems' computational resources known as performability modeling permits performance evaluation considering degraded levels of services due to faults over a specified period of time [6]. Performance and dependability are often modeled separately based on the assumption that the faults of the individual component or subsystem do not necessarily affect performance. This assumption is invalid for many systems that have either recovering or fault tolerant mechanisms. Such a mechanism aims at continuously provide 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 [7]. The performance evaluation is an essential activity to promote improvements in the quality of service provided, for planning the system infrastructure and to reduce costs of service. Maintenance activities are important, since the constant operation of equipment has great influence on its condition, which results in a degradation process. In this context, coordination of preventive maintenance activities to a particular characteristics of each system and the desired goals (reduce costs of service and maximize the availability or downtime) is extremely important. This process is called the maintenance policy [8].

This paper proposes performance and dependability models focused on storage and processing resources of the EFT system infrastructure which are represented by Generalized Stochastic Petri Nets (GSPN) [9]. Then, this paper combines both models (performance and dependability) for evaluating the impact of dependability issues on the performance of the EFT system. This work also proposes a preventive maintenance policy and evaluates the impact of this maintenance policy on the perfomability of the EFT system infrastructure through expolynomial distribution models [10]. In addition, different service level agreements (SLA) [11] are considered to evaluate EFT system maintenance contracts. The performance and dependability models takes into account the number of points of sales and transaction frequency besides considering fault occurrences and the effects of maintenance policy on availability of the EFT systems.

This paper is structured as follows: Section 2 presents related works and few fundamental concepts are presented in Section 3. Section 4 presents the EFT System Performance Model while Section 5 presents the service level agreement. Section 6 presents the proposed EFT System Maintenance Policy. The EFT System Dependability Model is presented in Section 7 whereas Section 8 presents the EFT System Performability Evaluation. Finally, experimental results are given in Section 9 and Section 10 presents the concluding remarks.

2. RELATED WORKS

In the last years, some works have been developed to evaluate the performance of EFT system's infrastructure. Sousa [12] et al propose a stochastic model for evaluating the performance of the EFT system the processing and storage infrastructure considering a load variation range. Araujo [13] et al propose a stochastic model for evaluating the performance of EFT systems in the course of demonstrating that the effect of traffic variation mechanism in computational resources is magnified when the inter-arrivals are highly bursty and correlated.

Most works that evaluate the performance of systems do not consider the influence of dependability aspects in performance degradation of these systems. Performability evaluation has been widely adopted as an essential activity for improving the quality of services provided. Sousa [14,15] et al propose an expolynomial stochastic model for evaluating the performance of the

EFT system infrastructure. This work also presents a model for evaluating the dependability of the EFT system infrastructures and combines both models for evaluating the impact of dependability issues on the performance of the system. Sesmun [16] et al propose a technique that uses performability in designing communication networks aiming at deriving a design methodology for fault-tolerant networks.

Hellerstein [17] proposes a performance and dependability model for application providers using a cluster. Furthermore, this work proposes a performance and cost model for planning the allocation of servers that comprise this cluster.

Different from previous works, this paper proposes a preventive maintenance policy and and evaluates the impact of this maintenance policy on the perfomability of the EFT system infrastructure through expolynomial distribution models. This paper also considers different service level agreements to evaluate EFT system maintenance contracts.

3. BASIC CONCEPTS

This section presents a summary of the concepts needed for a better understanding of this work. An overview of the stochastic Petri nets and phase-type approximation technique are presented. Stochastic Petri Nets form a high-level formalism that is widely applied for performance and dependability evaluation. GSPNs have been extensively applied for both automatic generation of Markov Chains and Stochastic Simulation. GSPNs are derived from Place/Transition Petri nets [18] 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 [9,19].

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

Phase-type approximation technique can be applied for modeling non-exponential activities [10,20]. A variety of performance and dependability activities can be constructed in GSPN models by using throughput subnets and s-transitions as shown in Figures 1(a), 1(b), 1(c) and 1(d). This throughput subnets and s-transitions [20] represent expolynomial functions, such as the Erlang, Hypoexponential and Hyperexponential distributions [20].

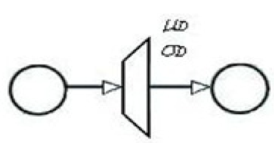

Figure 1(a): Empirical Distribution

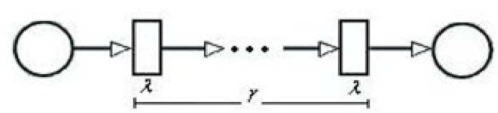

Figure 1(b):Erlang Distribution

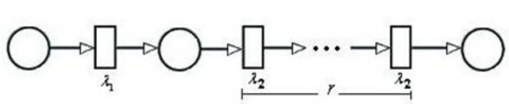

Figure 1(c): Hypoexponential Distribution

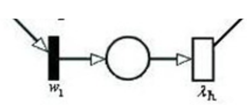

Figure 1(c): Hyperexponential Distribution

Figure 1. Expolinomial Distribution Nets

Measured figures from a system with an average μ_D and a standard deviation σ_D must adjust their stochastic behaviour through the phase-type approximation technique. The inverse of the coefficient of variation of the measured figure (Equation 1) allows the selection of which distribution matches it best, taking into account these two moments which are the average and standard deviation. For a deeper understanding on moment matching and expolynomial distributions, the reader is referred to [10,20]. In this work, the adopted distributions for moment matching are: Erlang, Hypoexponential, Hyperexponential and exponential distributions.

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

The Petri Net model depicted in Figure 1(a) represents an empirical distribution (a transition that represents an activity without a specified execution time distribution) by this trapezoidal shape transition.

When the inverse of the coefficient of variation is a whole number and different from one (if it is one, then the suitable distribution is the exponential), 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(b) represents an Erlang distribution.

Equation 2 $$\frac{1}{CV} = \frac{\mu_D}{\sigma_D}$$

Equation 3 $$\frac{1}{CV} = \frac{\mu_D}{\sigma_D}$$

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 1(c) is a net that depicts a hypoexponential distribution.

Equation 4 $$\left(\frac{\mu}{\sigma}\right)^2 - 1 \le \gamma < \left(\frac{\mu}{\sigma}\right)^2$$

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

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

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

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

When the inverse of the coefficient of variation is a number smaller than one, the empirical figure should be represented by a hyperexponential distribution. The exponential transition rate should be calculated by Equation 9 and the weights of immediate transitions are calculated by the Equations 10 and 11. The Petri Net model that represents this hyperexponential distribution is shown in Figure 1(d).

Equation 10 $$\lambda_h = \frac{2\mu}{\mu^2 + \sigma^2}$$

Equation 11 $$w_1 = \frac{2\mu^2}{\mu^2 + \sigma^2}$$

4. THE PERFORMANCE MODEL

This section presents the GSPN model for performance evaluation of the EFT system.
The electronic funds transfer system controls all commercial transaction process carried out among the points of sales and the credit and debit authorizers. This system is composed of client and management applications. Client applications are configured in points of sale terminals belonging to different companies and subsidiaries, where management applications are configured on the EFT server. The client application is responsible for interfacing points of sale 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 sale including all parameters needed for the construction of messages to be sent by the terminals and then forwards the transaction. Thus, the EFT system manages the commercial operations of companies and subsidiaries, and also controls the status of the points of sale including exchange of messages between the points of sale and the credit and debit authorizers.

The GSPN model depicted in Figure 2 presents the high-level model of the EFT system. The proposed EFT model is described via its “sub-models” (subnets that describe the system's components). Client subnets represent EFT client applications in different companies, such as drugstores, supermarkets, gas stations, shopping centre 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 credit and debit trade transactions with distinct occurrence frequencies. Hence, the demands of services can be represented by varying the transfer transaction frequencies and other parameters. The markings NI assigned to the places drugstores, supermarkets etc on client subnets specify the number of points of sale a particular type, and the generic stochastic transitions (s-transitions) represent the delay distribution between transactions. The place Buffer (Buffer subnet) represents the temporarily hold transactions waiting to be served. The dual place marking (M(P10)) represents the buffer storage capacity. This subnet permits the variation of the amount of commercial transactions that may be on the queue before being served, and thus a limited queue. Each place Buffer marking represents a commercial operation that will be processed by the EFT Server.

The management model represents the processing and storage infrastructure of the server where the management applications are configured. The management model is composed of the Processing Transaction and the Storage System Transaction subnets. The Processing Transaction subnet represents the processing of transaction and the Storage System Transaction subnet represents the disk reading and writing operations related to transactions. The place Processor marking NP denotes the processing capacity, which is the number of concurrent transactions supported by the processing resource (in other words, the concurrency degree). The place Storage System marking ND denotes the number of concurrent disk transactions supported by the storage resource (its concurrency degree). The stochastic transitions Tip and Tid (s-transitions) firing represent the transaction processing time and the storage (reading and writing related to a transaction) operation time (See Figure 2).

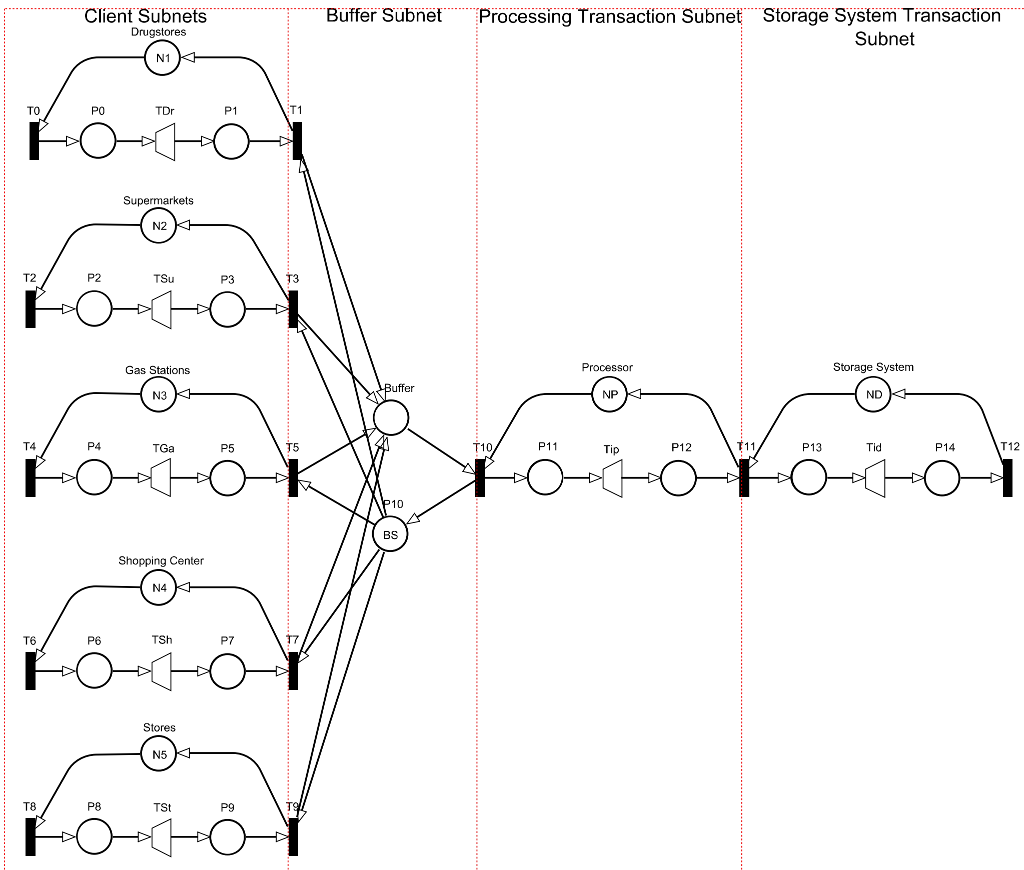

Figure 2: EFT System Performance Model

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

4.1. Model Refinement and Validation

This section presents a case study, which was used for validating the proposed model. The adopted system, called SCOPE [21], 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 exchange of messages between the points of sale and authorizers. The EFT system is composed of client (SCOPE Client) and management (SCOPE Server and SCOPE AUT) applications [21].

In order to validate the EFT System Performance Model proposed, experiments were conducted in an environment set in the CIn-Itautec Performance Evaluation Laboratory [22] considering real transaction traces [21]. The client applications were installed in a range of MX203 servers [21]. These client applications forward real transaction traces collected from EFT system users. The management application was configured in a MX223 server [21].

The client applications were configured to represent 3,345 points of sale distributed in a Shopping Centre. The points of sale register the credit and debit transactions through the SCOPE client applications. The evaluated scenarios describe a Shopping Centre with points of sale where the demand register occurred at 7 different rates, namely 100, 200, 300, 400, 500, 600 and 700 tpms (transactions per minute).

The performance 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 [23]. Some metrics such as service time and utilization for the storage system and processor are indirectly estimated [1,24].

After setting up and stabilizing the environment, measurements of performance metrics were initiated through the Windows Performance Monitor (Perfmon) [25]. During the measurements, processes that are not strictly necessary for the experiments were cut off so as to avoid interference in the collected data [26]. Measurements occurred for a period of 12 hours with a range of 1 minute between data collections. These data were stored in logs generated through the tool itself and used for statistical analysis. The collected data was stored on a disk partition isolated from the measuring environment in order to prevent the measured data from being affected. These measurements have been used for analyzing the impact of workload changes (commercial transactions), in the processing and storage infrastructure of the EFT server. The collected data was then statistically analyzed.

Among the evaluated performance measures, processor and storage system utilization, disk transfer per second (throughput) and processor and disk service time were chosen. The processor and the storage system service time were adopted to estimate the time for processing and the time for carrying out storage operations related to commercial transactions.

The measured data (empirical distribution) was analyzed for deciding which expolynomial distribution best fits the processing and storage operations (represented by the transitions Tip and Tid). The respective processing and storage time mean (μ_D) and standard deviations (σ_D) were calculated and the distributions were chosen according to the process described in Section 2. These transitions were refined according to the results presented in Table 1.

Table 1. Average and Standard Deviation.

Service Time	**μ_D (s)**	**σ_D (s)**	**Suitable Distribution**
Processor	0.001311	0.000508	Hypoexponential
Storage System	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 the Equations 4, 7 and 8. Table 2 shows the respective values of μ_1, μ_2 and γ for the models that refine Tip and Tid.

Table 2. Distribution Parameters.

Transition	**μ_D (s)**	**σ_D (s)**	**γ**
Tip	0.000080	0.00054	6
Tid	0.000001	0.00005	61

As a result, a refined version of the EFT System Performance Model is generated. Each refined model is depicted in Figure 3, where the complete refined model is not shown, since its graphic representation is directly obtained by substituting the s-transitions by each respective refined model (subnet). For validation purpose, the refined model is set according to specific - but representative - scenarios of interest. It should again be emphasized that for this particular study, the company interested in the study have provided the respective trace as a typical one to the study they were interested in. Nevertheless, it worth noting that changing the trace characteristics will also change the resulting model.

In refined model, the tokens in the place Shopping Centre represent points of sale forwarding credit and debit commercial transactions depicted through timed transition T1. The place P2 represents the buffer storage capacity. A large number of tokens were assigned as the buffer size. The reader should bear in mind that a large number is meant to be a buffer size that cases the storage of a much larger number of tokens (representing commercial transactions) than the SCOPE System may actually store on the real buffer, considering specific parameters of the other systems (processing time, storage system operation time, transaction frequencies). The numbers of tokens in the place Processor are adopted to define the concurrency degree related to the EFT system transaction executed by the processing components, that is, the number of transactions that could be simultaneously carried out by the management applications. Likewise, the marking

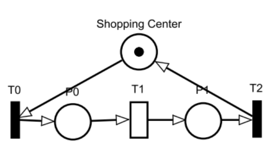

Figure 3(a): Client

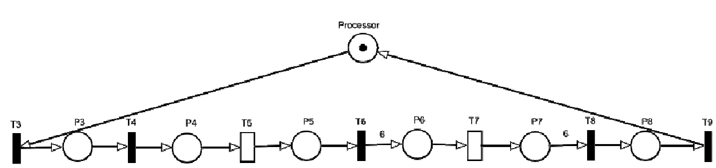

Figure 3(c): Processing Transaction

assigned to the place Storage System represents storage operations related to an EFT system transaction. The timed transitions T5 and T7 of the Processing Transaction subnet and timed transitions T11 and T13 of the Storage System Transaction subnet firing represent the processing time of the transaction and the storage (reading and writing related to a transaction) operation time.

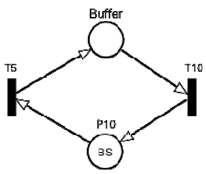

Figure 3(b): Buffer

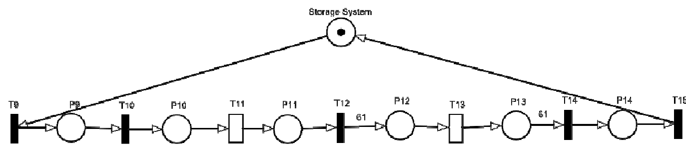

Figure 3(d): Storage System Transaction

The processor and storage system utilization levels as well as throughput are obtained using the following expressions: UProc=P{#Processor=0}, USS=P{#StorageSystem=0} and Throughput=P{#P12>0}/(μ_2) [29], respectively. Figures 4, 5 and 6 show a visual comparison between the results obtained by stationary analysis and measurements, considering a set of different commercial transaction frequencies: 100, 200, 300, 400, 500, 600 and 700 tpms (transactions per minute).

Figure 4 shows a comparison between the utilization levels of a measured processor and the values obtained through the model by the metric UProc=P{#Processor=0}. The result presents a maximum relative error of 13.60%.

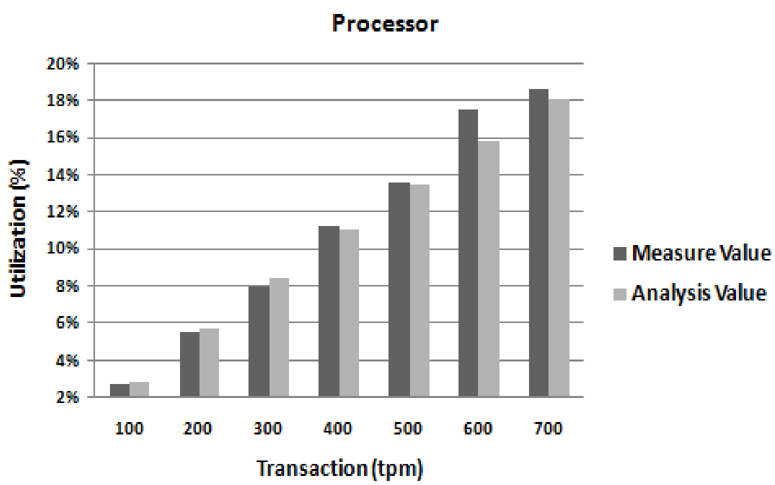

Figure 4: Processor Utilization

Figure 5 depicts similar comparison related to the storage system utilization levels. The values presented are those obtained by measurements and by the evaluation of the model through the metric USS=P{#StorageSystem=0}. The result presents a maximum relative error of 11.65%.

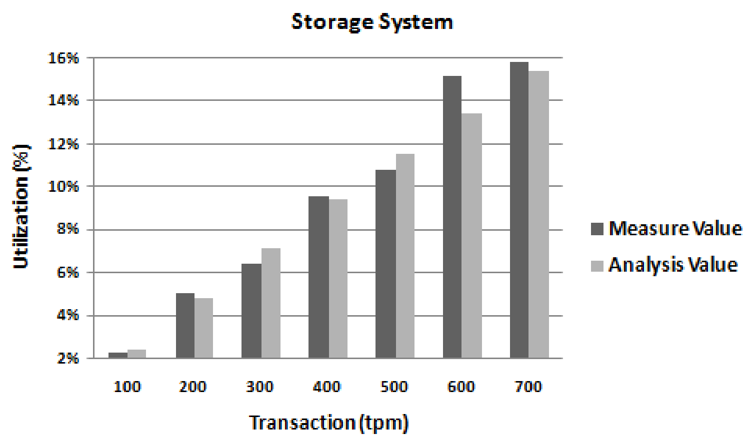

Figure 5: Storage System Utilization

Figure 6 presents a comparison between the measured disk throughput and values obtained through the model by the metric Throughput=P{#P12>0}/(μ_2). The result presents a maximum relative error of 13.67%. Table 3 sheds light on each performance expressions.

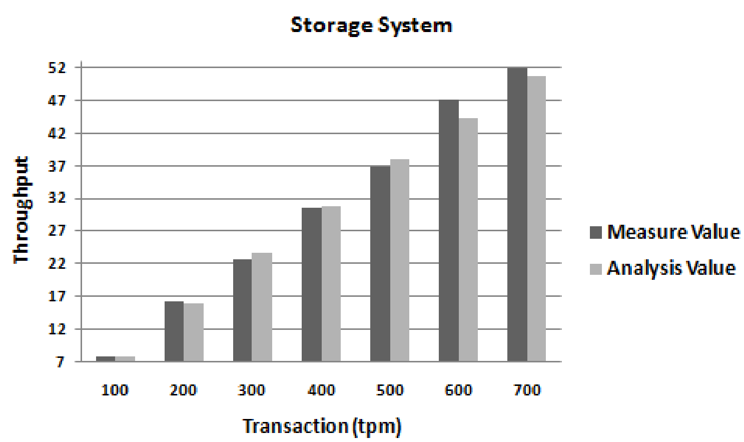

Figure 6: Throughput

The EFT System Performance Model is an abstraction of SCOPE EFT System. This model allows the understanding of system behaviour and estimate the system performance under different workloads or changes in infrastructure. The models can be expressed at different levels. The choice of granularity determines which analysis can be performed, depending on the details of the system components [2]. Already, the refined model corresponds to the abstract model and the statistics obtained through the measurement system. These statistics indicate the type of expolynomial distribution that best fits the empirical distribution (data collected). This adaptation is performed through the phase-type approximation technique [20]. After validation of the refined model, it is possible to perform analysis of different scenarios in order to find appropriate settings. This analysis includes the study of the system considering variations in demand and infrastructure systems.

5. SERVICE LEVEL AGREEMENT

Service Level Agreement is a document that defines a set of commitments between a company that provides services and a customer [11]. This document describes services to be contracted and the rates to be achieved for the fulfilment of all commitments agreed [27]. A service provider must offer service levels in accordance with performance and cost expectations of each customer [11].

Additionally, SLA contracts must ensure performance and penalties metrics in accordance with the services levels offered to customers [28]. Moreover, SLA contracts set an operational class that meets the availability expectations of customers and where fines are imposed when the levels availability is not achieved. The downtime is a metric used for imposing fines in the company that provides services adopted. In addition, the SLA contract analyzes the costs of clients as a result of the unexpected unavailability of a service and the probability that it occurs. The following items show the performance and penalty metrics.

- Availability: The service provider is responsible for ensuring the availability agreed in contract.
- Downtime: The outage duration should be maintained according to the service level offered by contract. The period of system in failure mode must be less than the downtime agreed in SLA contract.
- Penalty: Considering that the downtime period is greater than the one agreed in contract, the service provider must pay a fine. The fine is based on the hours of system outages and the costs of these hours are established according to the service level offered.

6. MAINTENANCE POLICY

This section presents the proposed preventive maintenance policy on an EFT system infrastructure and an equation for estimating the cost of maintenance activities. The Maintenance Policy (MP) describes the procedures to be adopted to minimize failure situations of the infrastructure of the EFT system ensuring the quality of the service offered.

In MP, corrective maintenance occurs at every failure and may result in repairing or replacing a piece of equipment. The preventive maintenance pursues the following procedures: the first maintenance occurs when the reliability indices of the systems reaches a critical specified level. After the first maintenance, the intervals between maintenance are fixed kT (k = 1,2 ,..., N). The replacement of equipment occurs when the number of intervals between sequential maintenance reaches kT (k = 1,2 ,..., N+1). In the first N maintenances, the piece of equipment is removed for the identification of defective devices and repairs avoid the failure occurrences. When the repairs activities reach N+1 times, the device should be substituted and the process restarts. T represents the period between each preventive maintenance and N, the number of preventive maintenance to be carried out before replacing the equipment. The local teams (LT's) and specialist teams (ST's) are allocated to carry out maintenance activities and for replacing a piece of equipment, and the maintenance planning team (MPT) is allocated to control maintenance activities.

The Local Teams (LT) conducts corrective and preventive maintenance in the company. These teams service a range of systems and, therefore, are not expert in any particular type of equipment. The LT carries out the routine checks, repairs and replacements. These technicians record all failures, repairs and the replacements.

Specialist Teams (ST) are called upon when the procedures to perform diagnosis, repairs or replacements require more expertise. The (ST) also visit companies in order to carry out more complex or critical periodic checks.

The Maintenance Planning Team (MPT) is responsible for managing information registered by the LT. Hence, the MPTs adjust the intervals between maintenance and maximum number of repairs before replacing each device. The result of this planning is to control the performance and dependability indices of the EFT systems.

The Mean Time to Repair (which could accomplishes the replacement of a device) (MTTR) of a piece of equipment depends on the team that is performing maintenance. MTTR measures the time between the interruption of the service and service restoration. The MTTR of Local Teams (LTMTTR) is composed of the Mean Time to Failure Perception and the Mean Time for Repairing. The MTTR of Specialist Teams (STMTTR) may be represented by the Mean Time to Failure Perception, the Mean Time for Repairing and the Mean Time for Team Arrival to the site. When maintenance is conducted by a ST, the MTTR may be dominated by the site location, which heavily affects the STMTTR, or by problem diagnosis. When changes are well documented and managed, the time required to successfully diagnose the problem might be intensely affected, and consequently, MTTR is lowered as well.

The maintenance costs are obtained by the Equation \ref{eq12}. N_i represents the number of maintenance teams performing a specific maintenance activity. $Team_i$ is a place that represents a specific type of maintenance team. C_i ($Team_i$) represents the maintenance cost per work hours of a specific type of maintenance team. Ra_j is the timed transition that represents maintenance activities as TP(Ra_j) represents the throughput of transition Ra_j. C_j(Ra_j) represents the annual cost of materials used in maintenance activities where T represents the period for calculating the maintenance cost.

Equation 12
$$C_a = \left[\left(\sum_{i=1}^{m} N_i \times C_i(Team_i) \right) + \left(\sum_{j=1}^{n} TP(RA_j) \times C_j(RA_j) \right) \times T \right]$$

7. THE DEPENDABILITY MODEL

This section presents the GSPN model conceived for dependability evaluation of the EFT system. The proposed the EFT model is described via its "sub-models" (subnets that describe EFT system infrastructure) (see Figure 7). The Processing Transaction and the Storage Transaction subnets represent faults and repairing activities related to the processors and storage system in the EFT systems.

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 transitions PFailure1 and PFailure2 (s-transitions) represent a fault and their delays represent MTTF (mean time to failure). Likewise, the transition PRepair represents the repairing activity, and its delay related to MTTR (mean time to repair). The place Storage System marking Nd denotes the operational storage system and one token in the place DFailed indicates a failure. The generic stochastic transitions DFailure1 and DFailure2 (s-transitions) represent a fault and their delays represent MTTF (mean time to failure). Similarly, transition DRepair represents the repairing activity, and its delay related to MTTR (mean time to repair).

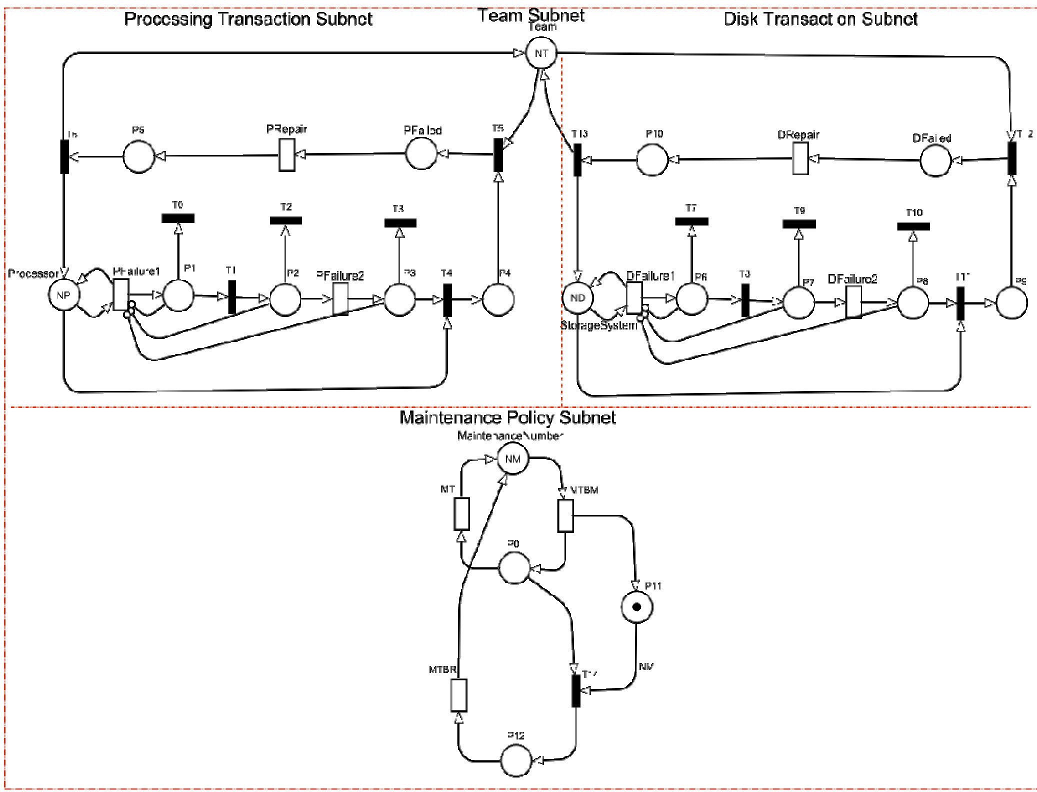

Figure 7: EFT System Dependability Model

The Maintenance Policy and Team subnets were designed according to the procedures of the Maintenance Policy (MP). The place Team marking (Nt) represents the maintenance team of a specific type. 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 number of tokens in the place Team directly affects the concurrency degree related to the repairing process. The MP enables corrective maintenances through the enabling function {#P4>0} in immediate transitions T5 and enabling function {#P9>0} in immediate transitions T12. Similarly, preventive maintenances are enabled by the enabling function {#P0>0} in immediate transitions T0, T2, T3, T5, T7, T9, T10 and T12. A possible maintenance policy may give a higher priority to recover the processing infrastructure than the storage subsystems, since the storage infrastructure usually has a higher number of storage devices than processing units, and hence allowing longer maintenance delays. To represent this priority in the repair process with respect to the processing infrastructure over the storage infrastructure, weights are assigned to immediate transitions T5 and T12 [29], where the weight with a higher value is attributed to the immediate transition T5.

The Maintenance Policy subnet (see Figure 7) represents the maintenance policy adopted. The place MaintenanceNumber marking Nm indicates the maintenance number before replacing an equipment. The generic stochastic transitions MTBM and MTBR represent the mean time between preventive maintenance and the mean time between replacements, respectively. The enabling function {#P0>0} [29] is assigned to immediate transitions T0, T2, T3, T5, T7, T9, T10 and T12 to represent the intervals between specified preventive maintenance in the maintenance policy adopted. Figure 7 depicts the EFT System Dependability Model.

The availability of both the processor and the storage system are represented by the evaluation of the expressions AProc=P{#Processor=1} and AStorageSystem=P{#StorageSystem=1} [29], meaning that the EFT system infrastructure will be in the operational mode when there is at least one token in the places Processor and StorageSystem.

This EFT System Dependability Model aims at 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.

8. PERFORMABILITY EVALUATION

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 focused on storage (physical disk) and processing resources which are represented by Generalized Stochastic Petri Nets (GSPN) [9]. The obtained results enable the evaluation of performance figures, such as throughput and utilization level for specifying the capacity of the EFT system infrastructure.

The adoption of two detached models [6], since the order to magnitude of failure and repairing rates, and performance figures are highly distinct aims at handling largeness and stiffness problems [10]. The composition technique leads to a natural hierarchy of models, but it is important to stress that the obtained results are approximations of exact solutions. The adopted method combines the dependability model, which takes into account failures and the repairing process and the system performance model [30,31]. The dependability and the performance models may also be hierarchically evaluated whenever the complexity of the problem or the size of the models is not reasonable.

The adopted approach takes into account the number of points of sales and transaction frequency besides considering fault occurrences and the effects of maintenance policy on availability of the EFT systems. The maintenance policy may prevent the occurrence of faults and failure through the adjustment of intervals between preventive maintenance, maximum number of repairs before replacement, the mean time to repair and the maintenance team sizing.

9. EXPERIMENTAL RESULTS

This section presents the results of the evaluation of the system described in Section 3. These results are related to a set of noteworthy scenarios for planning the EFT infrastructure for supporting the points of sales that process credit and debit transactions of a specific shopping centre.

The EFT system allows payments to companies through the exchange of electronic messages between the points of sales distributed in various businesses and companies. The popularity of electronic payment systems has required that these systems meet their customers with satisfactory performance levels. A few seconds of delay in payment of each customer can result in increased queues at cash registers and thus in customer dissatisfaction.

The transaction rates adopted in this scenario were 3,500, 4,500, 4,900, 5,600, 6,300, 7,200, 8,100, 9,801, 14,850 and 16,830 tpms (transactions per minute). These frequencies were adopted with the aim of finding the saturation of the processing and storage infrastructure. At a frequency of 16,830 tpms, the processor utilization was 86.19% and the storage system utilization was

73.10%. The processor's utilization is high hence this resource should be carefully evaluated, tuned, possibly updated or even replaced as a preventive measure for taking into account further workload demands [23] if that workload level is expected to be achieved in a specified time horizon. The same should be done with respect to the storage system. Figure 8 depicts the processor and storage system utilization. These results show the effect of workload fluctuations on the processor and storage system utilization considering each commercial transaction frequency. The processor and storage system utilization are obtained using the following reward expressions: UProc=(P{#Processor=0}) and USS=(P{#StorageSystem=0}) [29], respectively.

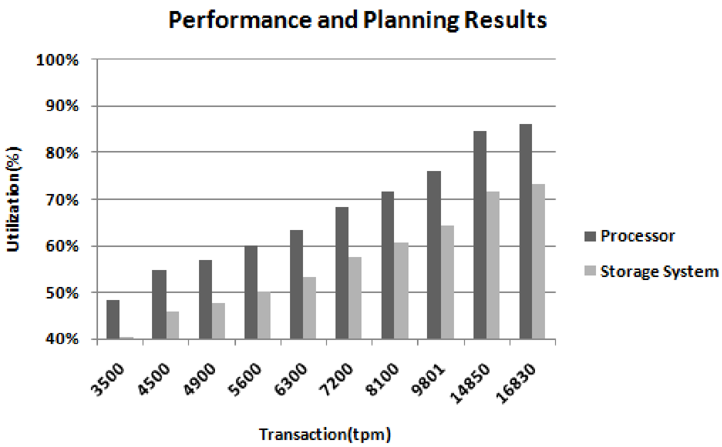

Figure 8: Processor and Storage System Utilization

To cope with the eventual workload peaks, fluctuations and to address bottlenecks, a new processing infrastructure is proposed to be added to the EFT System. On adding another equivalent processing infrastructure, the processor utilization was reduced to 42.67% when considering 16,830 tpms. Hence, two processing infrastructures drastically reduce its utilization and prevent any possible bottleneck that might occur due to workload oscillations. The two processing infrastructures represent the concurrency degree related to the transactions of the EFT system executed by the processing component, that is, the number of transactions that could be simultaneously processed by the management applications.

The following case study evaluates the EFT system considering the proposed preventive maintenance policy (MP) and SLA contracts provided by a maintenance company.

Different SLA contracts have been offered by the maintenance company as depicted in Table 5. For example, SLA I contract guarantees availability (Avail.) of 99.9%, which is the same as a maximum downtime (Down.) of 8.76 hours considering the period of one year. Furthermore, each contract has a price associated and a fine to be paid (per hour) in case the company cannot provide the values assigned in contract.

Table 3. SLA Contracts of Maintenance Company.

Scenarios	Avail.(%)	Nines	Down.(h)	Contract Price($)	Fine ($)
SLA I	99.90	3.00	8.76	32,153.78	3,000.00
SLA II	99.95	3.30	4.38	44,786.24	4,500.00
SLA III	99.99	4.00	0.88	60,537.04	6,000.00

Case study aims to assess the impact of the proposed maintenance policy (MP) on the availability of the EFT system processing and storage infrastructures. The adopted maintenance policy provides ten preventive maintenances at fixed intervals and a replacement of equipment after the occurrence of these maintenances. It is important to stress that the MTTFs of the processing and storage infrastructure used in this case study is based on the mean time between failures (MTBF) of these resources [21].

Tables 6 and 7 show the adopted metrics to quantify the impact of the preventive maintenance policy on the EFT system infrastructure. Furthermore, those tables also illustrate the three scenarios that have been adopted to quantify the impact of proposed maintenance policy through the different SLA contracts defined.

Table 4. Data for Processor.

Scenario	**MTBM (hr)**	**MTBR (hr)**	**MTTF (hr)**	**MTTR (hr)**
SLA I	4,380	43,800	17,520	16
SLA II	1,752	17,520	17,520	16
SLA III	876	8,760	17,520	16

Table 5. Data for Storage System.

Scenarios	**MTBM (hr)**	**MTBR (hr)**	**MTTF (hr)**	**MTTR (hr)**
SLA I	4,380	43,800	43,800	18
SLA II	1,752	17,520	43,800	18
SLA III	876	8,760	43,800	18

Table 8 shows the parameters adopted to obtain the annual maintenance cost for SLA I, II and III (see Table 5). The annual maintenance cost are based on the Equation 12 and the parameters $C_i(Equipe_{LT})$ and $C_j(RA)$. The loss associated to the downtime during the period of one year is US$ 50,000.00 per hour.

Table 6. Parameters of Maintenance Cost.

Parameters	**Maintenance Cost (US$)**
$C_i(Equipe_{LT})$	2,000.00
$C_j(RA)$	10,000.00

As aforementioned, this case study aims at quantifying the impacts of adopted maintenance policy on the availability of the EFT system. Regarding the obtained results, Tables 9 and 10 depict the availability (Avail.), the respective number of nines (Nines), downtime for one year (Down.), the maintenance cost associated (Maint.Cost) and the loss obtained by the evaluation of the EFT system infrastructure.

The results depict the availability of first scenario is smaller than, for example, the availability of the third scenario. Such behaviour had been expected, as the first scenario has the highest total maintenance time (see Tables 6 and 7). As a consequence, the first scenario also has the highest downtime that reflects the associated maintenance cost and financial loss.

Regarding the availability, it is important to highlight the significant increase from SLA I to SLA III. As expected, there was a sharp decrease in downtime from SLA I to SLA III. Consequently, an increment occurred the in the context of annual profit. To conclude although the SLA III contract price is the most expensive, the results obtained demonstrate a increase on the availability and on the annual profit of the company. This behaviour occurs due to the huge decrease in downtime, which improves the correspondent financial losses.

Table 7. Processor Results.

Scenarios	Avail.(%)	Nines	Down.(h)	Maint.Cost(US$)	Loss($)
SLA I	99.9842%	3.80	1.38	8,574.34	69,204.00
SLA II	99.9921%	4.10	0.69	10,565.23	34,602.00
SLA III	99.9957%	4.37	0.38	11,419.63	18,834.00

Table 8. Storage System Results.

Scenarios	Avail.(%)	Nines	Down.(h)	Maint.Cost(US$)	Loss($)
SLA I	99.9964%	4.44	0.32	8,574.34	15,768.00
SLA II	99.9984%	4.80	0.14	10,565.23	7,008.00
SLA III	99.9991%	5.05	0.08	11,419.63	3,942.00

Figure 9 shows a comparison between the availability, the respective downtime and the annual profit of the EFT system infrastructure for each SLA contract. Note that the baseline corresponds to the SLA I. Regarding the availability, it is important to highlight the significant increase from SLA I to SLA III (20%) which represents the effectiveness of each maintenance contract. As expected, there was also a sharp drop in the downtime from SLA I to SLA III (over 70%). Consequently, a relevant increment occurred in the context of annual profit from SLA I to SLA III (over 40%).

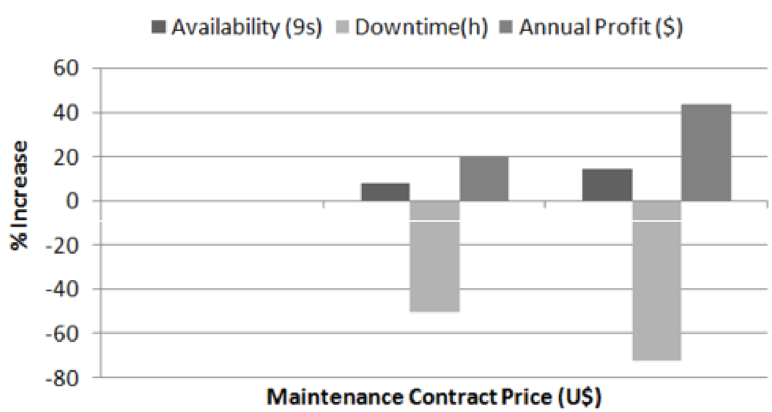

Figure 9: Comparing Availability, Downtime and Profit for the Storage System

The availability results of the processing and storage units of the EFT system infrastructure considering a maintenance team realizing maintenance and replacement activities according to MP for SLA I, II and III. The high levels of availability for SLA III compared to SLA I and II show that a shorter interval between preventive maintenance (MTBM) increases the availability of the EFT system's infrastructure. These results were obtained through the expressions AProc=(P{#Processor=1}) and AStorageSustem=(P{#StorageSystem=1}) [29], respectively.

The performability results are obtained by the combination of the dependability metrics with the performance metrics of the EFT system. These results are a combination of availability levels with the processor and storage system utilization considering the frequencies: 3500, 4500, 4900, 5600, 6300, 7200, 8100, 9801, 14850 and 16830 tpms (transactions per minute). These performability results present performance degradation of the EFT server infrastructure for SLA I, II and III.

Figure 10 shows the effects of failure, corrective and preventative maintenance activities in the utilization of processor and storage system, respectively, for SLA I, II and III considering a single maintenance team. This figure shows that a shorter interval between the preventive maintenances (MTBM) for SLA III compared to SLA I and II promotes a faster identification of defective equipment, avoiding failure events and reducing the performance degradation of the system's infrastructure. This result can be adopted in order to identify the causes of performance degradation of the system, and also to assure the quality of the offered service.

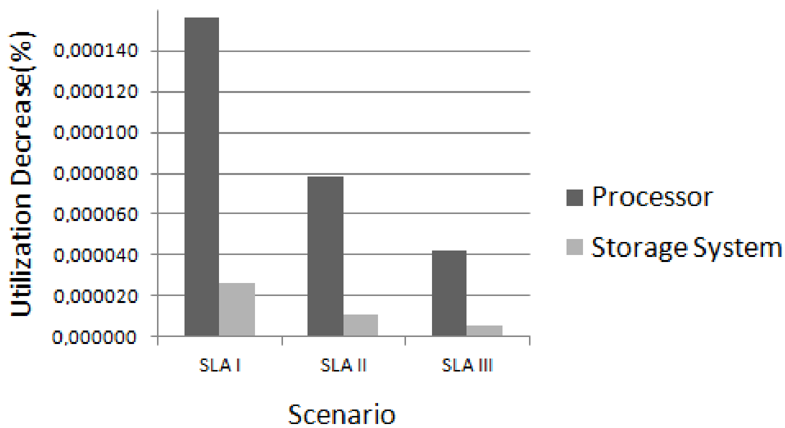

Figure 10: Utilization for the SLA I, II and III Considering the Availability Metric

This figure presents the percentage of the utilization decrease of processing and storage resources for each SLA contract. Regarding the utilization, it is important to highlight the significant decrease from SLA I to SLA III. As expected, there was a sharp decrease in downtime from SLA I to SLA III.

10. CONCLUSIONS

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

The performability analysis is carried out through dependability and a performance model. The EFT system performance model makes possible the analysis for sustaining quality of service and preventing performance degradation related to the workload fluctuations. In addition, the evaluation supports the search of suitable hardware and software configurations for ensuring performance agreements. The dependability model allows the evaluation of the structural variations of the EFT systems.

Additionally, the proposed model may also be adopted along side with the proposed Maintenance Policy and SLA commitments to demonstrate that a successful agreement can increase the annual profit of the company. Moreover, case studies presented the feasibility of the hierarchical method, in which the EFT system infrastructure have been analyzed and also its SLA maintenance contracts. The results provide an important impact associated to the selected maintenance policy, the availability of the whole system and, as a consequence, the annual profit. The adopted performability analysis and the proposed expolynomial stochastic models of the EFT System were applied to an EFT system named SCOPE. The experiments were conducted considering real transaction traces collected from users of the EFT systems. The results of performability evaluation combined probability resources (processing and storage units) in operational states over a period of 43,800 hours, their utilization, considering a set of different commercial transaction frequencies. It is important to stress that evaluating these experiments without models is a complex and expensive task.

The EFT systems have experienced a growth over the last decade. The provision of services and satisfying stringent requirements such as high availability at a proper cost are fundamental for organization's profit. The expolynomial stochastic models of the EFT system through the results presented show its suitability for being proper means for planning the EFT system infrastructure and ensuring high service levels.

As future works, we intend to detail the training process and the duration of the diagnosis process. We also intend to study other performance and dependability metrics considering a given load variation range.

REFERENCES

[1] Jain, R., (1991) The Art of Computer Systems Performance Analysis, John Wiley & Sons New York.
[2] Menascé, D. A. and Almeida, V. A. F. (2005) Performance by Design: Computer Capacity Planning by Example, Prentice Hall PTR.
[3] Laprie, J. C. C. and Avizienis, A. and Kopetz, H. (1992) Dependability: Basic Concepts and Terminology, Springer-Verlag New York, Inc. Secaucus, NJ, USA.
[4] Higgins, L. and Mobley, K. and Smith, R., (2002) Maintenance Engineering Handbook, McGraw-Hill Professional.
[5] Blischke, W. and Murthy, D. (2003) Case Studies in Reliability and Maintenance, John Wiley.
[6] Haverkort, B. R. and Marie, R. and Rubino, G. (2001) Performability modelling: techniques and tools, John Wiley & Sons Inc.
[7] Reibman, A. L. (1990) Modeling the effect of reliability on performance. Reliability, IEEE Transactions on. Volume 39. Number 3. Pages 314-320.
[8] Wang, H. and Pham, H. (2006) Reliability and optimal maintenance.Springer Verlag.
[9] Balbo, G. (2001) 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. Springer.
[10] Bolch, G. (2006) Queueing networks and Markov chains: modeling and performance evaluation with computer science applications. Wiley-Blackwell.
[11] Alpendre, M. (2006) Service Level Agreement: Um Conceito a Saber Usar. Universidade de Coimbra.
[12] Sousa, E. and Maciel, P. and Araujo, C. and Alves, G. and Chicout, F.(2009) Performance modeling for evaluation and planning of Electronic Funds Transfer Systems. Computers and Communications, 2009. ISCC 2009. IEEE Symposium. Pages 73-76.
[13] Araújo, C. and Sousa, E. and Maciel, P. and Chicout, F. and Andrade, E. (2009) Performance Modeling for Evaluation and Planning of Electronic Funds Transfer Systems with Bursty Arrival

Traffic.2009 First International Conference on Intensive Applications and Services, IEEE. Pages 65-70.
[14] Sousa, E. and Maciel, P. and Araújo, C. and Chicout, F. (2009) Performability Evaluation of EFT Systems for SLA Assurance. Parallel and Distributed Processing Symposium, International.IEEE Computer Society. Volume 0. Pages 1-8.
[15] Sousa, E. and Maciel, P. and Araújo, C. (2009) Performability Evaluation of EFT Systems using Expolinomial Stochastic Models. Proceeddings of ICDS.
[16] Sesmun, A. and Turner, L. F. (2000) Performability in the Design of Communication Networks. IEEE Proceedings-Computers and Digital Techniques. Volume 147. Number 5. Pages 355-363.
[17] Hellerstein, J. L. and Katircioglu, K. and Surendra, M. (2005) An on-line, business-oriented optimization of performance and availability for utility computing. Selected Areas in Communications, IEEE Journal on. Volume 23. Number 10. Pages 2013-2021.
[18] Murata, T. (1989) Petri Nets: Properties, Analysis and Applications. Proceedings of the IEEE. Volume 77. Number 4. Pages 541-580.
[19] Marsan, M. A. and Balbo, G. and Conte, G. and Donatelli, S. and Franceschinis, G. (1998) Modelling with Generalized Stochastic Petri Nets. ACM Press New York, NY, USA. Volume 26. Number 2.
[20] Desrochers, A. A. and Al-Jaar, R. Y. (1995) Applications of Petri Nets in Manufacturing Systems: Modeling, Control, and Performance Analysis. IEEE Press.
[21] Itautec (2009) http://www.itautec.com.br
[22] Itautec (2011) http://itautec.cin.ufpe.br
[23] Friedman, F. (2005) Microsoft Windows Server 2003 Performance Guide. Microsoft Press Redmond, WA, USA
[24] Trivedi, K. S. (2008) Probability & Statistics with Reliability, Queuing and Computer Science Applications. Wiley-India.
[25] Microsoft Corporation. (2003) Performance Tuning Guidelines for Windows Server 2003.
[26] Lilja, D. J. (2005) Measuring computer performance: a practitioner's guide. Cambridge Univ Pr.
[27] Sturm, R., Morris, W. and Jander, M. (2000) Foundations of Service Level Management.Sams Publishing. Pages 288.
[28] McConnell, J. and Siegel, E. (2004) Practical service level management: delivering high-quality web-based services. Cisco Press.
[29] Zimmermann, A. and Knoke, M. A. (2007) Software Tool the Performability Evaluation with Stochastic and Colored Petri Nets. Technische Universitt Berlin. Real-Time Systems and Robotics Group.
[30] Reibman, A. L. (1990) Modeling the effect of reliability on performance. Reliability, IEEE Transactions on. Volume 39. Number 3. Pages 314-320.
[31] Sahner, R. and Trivedi, K.S. and Puliafito, A. (1996) Performance and reliability analysis of computer systems: an example-based approach using the SHARPE software package. Kluwer Academic Pub.