Semi-Markov Models for Performance Evaluation of Telecommunication Networks in the Presence of Failures

Maurizio Guida[a], **Maurizio Longo**[a], **Fabio Postiglione**[a*], **Kishor S. Trivedi**[b], **Xiaoyan Yin**[b]
[a]Department of Electronic and Computer Engineering, University of Salerno, Fisciano (SA), Italy
[b]Department of Electrical and Computer Engineering, Duke University, Durham, USA

Abstract: Planning and deployment of next generation telecommunication networks based on the Third Generation Partnership Project IP Multimedia Subsystem (IMS) must take into account the occurrence of random failures causing performance degradations, in order to assess and maintain the Quality of Service offered by telecommunication service providers to their subscribers. In particular, core network signalling servers of IMS can be modelled as repairable multi-state elements where server states correspond to different performance levels. In this paper, we evaluate IMS signalling network performance in terms of the number of sessions handled by the network per time unit, by adopting a semi-Markov model for the IMS servers, which allows as well for non-exponential probability distributions of sojourn times, as often observed in practical network scenarios. Furthermore, a redundancy optimisation problem is solved in an IMS-based realistic scenario, to the aim of minimizing the deployment cost of a telecommunication network with a given availability requirement.

Keywords: Availability, IP multimedia subsystem, Multi-state system, Non-exponential model, Performance evaluation.

1. INTRODUCTION

A Next Generation telecommunication Network (NGN) is aimed to offer seamless services to users connected to both wired and wireless infrastructures, with a high level of end-to-end Quality of Service (QoS) [1]. To this aim, telecommunication industry must cope with network/service failures affecting performance. Thus, comprehensive approaches to assess performance in the presence of failures occurring in the network nodes are necessary [2, 3], also known as *performability* analysis [4]. In this paper, which is an extension of [5], we focus on telecommunication networks based on the IP Multimedia Subsystem (IMS), proposed by the Third Generation Partnership Project (3GPP) [6]. In particular, we model core network signalling servers, namely the Call Session Control Functions (CSCFs), as multi-state systems where each state corresponds to a different performance level and whose dynamical behaviour can be described by non-exponential models, such as semi-Markov processes [7, 8]. This analysis is stimulated by evidences in real networks, where sojourn times in the states are often distributed according to a probability distribution different from the exponential (memoryless) one. For every telecommunication system, one of the performance metrics related to the signalling server functionalities is the number of call set-up sessions in the time unit handled by a single CSCF server. Given a constant demand profile, we discuss the problem of meeting typical requirements for an IMS-based NGN in terms of signalling service availability, which is defined as the probability that the multi-state system is in one of the states where performance is not less than the demand.

The paper is organized as follows. In Section 2, we introduce the IMS system and discuss its main performability issues. In Section 3, we propose for the IMS core network a multi-state performance model in the presence of failures based on semi-Markov processes. In Section 4, we illustrate some techniques to compute IMS signalling system availability and to meet availability requirements while minimizing the cost. In Section 5, the proposed methodology is applied to an explanatory network scenario. Finally, we provide some concluding remarks.

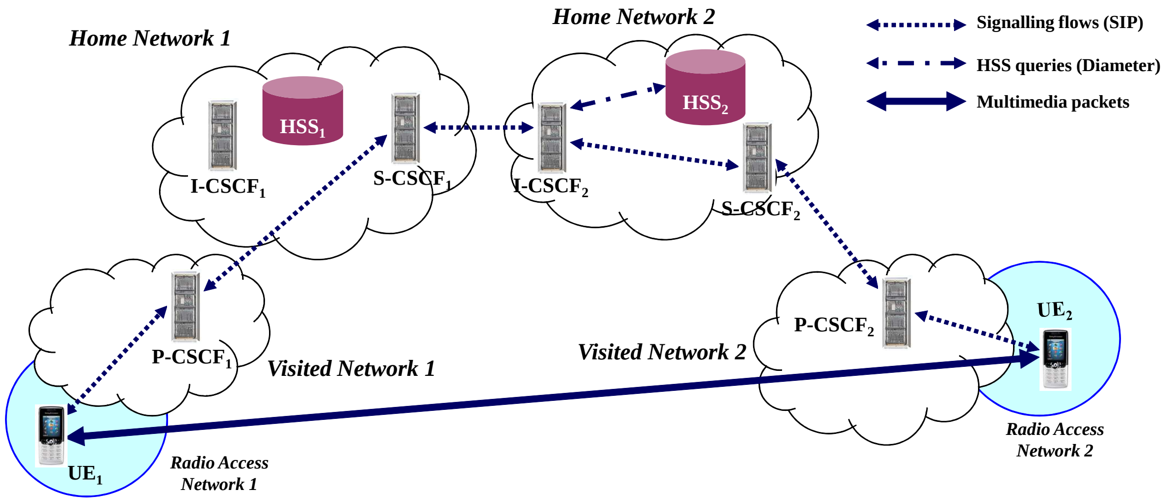


Figure 1: A typical call set-up scenario involving two roaming mobile phones (UE) attached to different radio access networks. Users are subscribed to different telecom operators (Home Networks).

2. AN OVERVIEW OF THE IP MULTIMEDIA SUBSYSTEM

2.1. System Description

The IMS proposed by 3GPP [6] is the core of next generation mobile networks: it enables both real-time audio/video communications and data transport services, and can be connected to both the traditional telephone network and Internet by means of specific gateways. The wireless connections between a User Equipment (UE), representing the mobile terminal, and the core network are provided by Radio Access Networks (RANs). The 3GPP adopts Session Initiation Protocol (SIP) [9] as the signalling protocol for handling real-time audio/video sessions. To this aim, some new nodes are presented in an IMS-based core network: those managing signalling flows, i.e., Call Session Control Function (CSCF) nodes, and those handling multimedia flows, which typically traverse different paths in the network with respect to signalling packets. The CSCF functionalities are assigned to three different servers: *(i)* the Proxy CSCF (P-CSCF), usually located in the Visited Network, is the first contact point to the IMS network for the UE, and forwards the SIP signalling to the subscribers' Home Network (HN), typically the operator to which the user is subscribed; *(ii)* the Serving CSCF (S-CSCF) is located in the HN and manages multimedia sessions; *(iii)* the Interrogating CSCF (I-CSCF), located in the HN, is responsible for selecting the appropriate S-CSCF for all the requests entering the HN. Another important signalling server in the Home Network is the Home Subscriber Server (HSS), a database containing users' details retrieved by S-CSCFs, during the user registration phase, or I-CSCFs, during the call set-up procedures. The HSS databases are interrogated by another protocol (Diameter). The interaction among the three CSCF and the HSS nodes allows complete management of multimedia sessions.

A typical network scenario, shown in Fig. 1, outlines the call set-up of a real-time session between two roaming UEs including: signalling messages exchanged among UEs and IMS nodes; two Home Networks reproducing network domains of telecom operators to which users are subscribed; two Visited Networks that represent service providers networks with two different RANs that the roaming subscribers are attached to. It is assumed that the UEs have already notified their actual locations to their home networks (registration phase).

2.2. Performability Issues

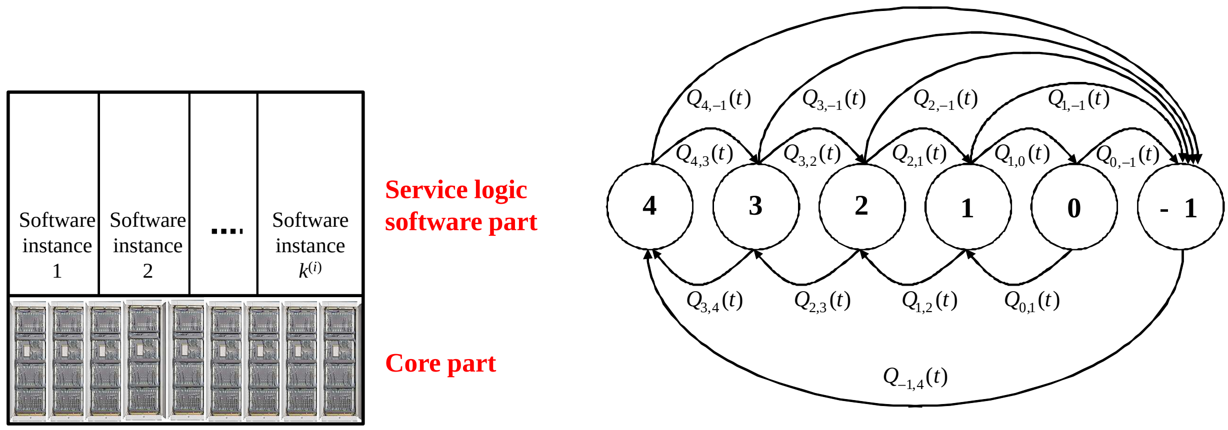
A real-time session can be established between two registered UEs if and only if all signalling nodes are available during call set-up procedure, and hence, in the availability perspective, the signalling network constitutes a *series system*. Since all servers involved in the call set-up session are failure-prone, availability methodologies are necessary to evaluate the overall IMS-based system availability, which, in line of principle, should be comparable to that of the traditional telephony. In addition, IMS nodes have typically different performance levels and several failure and repair modes with various effects on the entire signalling network performance. Hence, an availability/performance (i.e., performability) model should consider, on one side, this distributed nature of call control functionalities influenced by every single node behavior, and, on the other side, the variety of failures, ranging from accidental failures of software modules and/or hardware equipments to man-made faults due to the activity of malicious users [10].

3. A SEMI-MARKOV PERFORMABILITY MODEL FOR THE IMS SERVERS

In line with [5], in order to evaluate the performance of an IMS-based next generation network in the presence of failures, we choose as a performance metric the number of call set-up sessions that an IMS core network is in charge of. Obviously, this metric is a function of the number of sessions that every single node, involved in the signalling procedures, can manage. We assume that only IMS servers behavior can influence the performability evaluation of the network, while all switches/routers and (wired/wireless) links of the network are in fully operational state and do not limit system performance. Thus, some parallel redundancy might be necessary to respect performance/availability constraints, and hence more than one node would be deployed for the same IMS server i. Henceforth, we suppose those parallel nodes have the same software/hardware characteristics[1]. Thus, each IMS node, whose architecture is represented in Fig. 2(a), can be considered as composed by a service logic software part, consisting of a given number of parallel instances of the same CSCF server, and a core part, consisting of the hardware setting (processors, memories, power supply, etc.) and the node operating system [11]. Let $n^{(i)}$ be the number of call set-up sessions that each server instance can serve, and let $k^{(i)}$ be the number of those instances, where the superscript $i \in \{P1, S1, I2, H2, S2, P2\}$ indicates the P-CSCF$_1$, S-CSCF$_1$, I-CSCF$_2$, HSS$_2$, S-CSCF$_2$, and P-CSCF$_1$ servers, respectively.

In [5], a Multi-State System (MSS) performance model for an IMS signalling server is provided, where transitions between states are exponentially distributed, resulting in a Continuous-Time Markov Chain (CTMC). However, CMTC models have restrictions in terms of applicability to some real world problems, where sojourn times in the states do not fit an exponential distribution. In this paper, we adopt a model containing the same states and transitions of that proposed in [5], but, unlike it, sojourn intervals could be non-exponentially distributed. The resulting model is not a CTMC anymore, and it can be studied by non-exponential performability models, such as those based on semi-Markov processes [12], a subclass of the more general Markov Regenerative processes [13]. The MSS semi-Markov model of a generic server i is the one reported in Fig. 2(b), where: the states $j = 0, ..., k^{(i)}$ indicate that j-out-of $k^{(i)}$ software instances are operational in the node, given that the core part of the element is working; state $j = -1$ denotes the out-of-order condition of the core part; in the set $j = 0, ..., k^{(i)}$, transitions may happen between states differing by one active sub-module or toward the state $j = -1$ (core failure), while in state $j = -1$ a single transition to state $j = k^{(i)}$ is possible, because it is assumed that the repair action on the node site is concluded by a complete recovery of the hardware/software functionalities.

[1] This assumption can be easily generalised allowing nodes with different characteristics.



(a) (b)

Figure 2: An IMS node running $k^{(i)}$ software instances: (a) the architecture, (b) an MSS semi-Markov model with $k^{(i)} = 4$ parallel software instances implementing the service logic.

3.1. The Proposed MSS Semi-Markov Model

Let $g_j^{(i),l}$ be the performance level (the maximum number of call set-up sessions that can be handled) of l-th node of the server i in the state $j \in \{-1, ..., k^{(i)}\}$. From the semi-Markov model described in Section 3, the performance levels are

$$g_j^{(i),l} = \max\{jn^{(i)}, 0\}, \tag{1}$$

and hence the set containing these values is

$$\mathbf{g}^{(\mathbf{i}),\mathbf{l}} = \{g_{-1}^{(i),l}, g_0^{(i),l}, ..., g_{k^{(i)}}^{(i),l}\}. \tag{2}$$

Let the stochastic process $G^{(i),l}(t) \in \mathbf{g}^{(i),l}$ represent the performance level of the l-th node of the server i at any instant $t \geq 0$, given an initial state $G^{(i),l}(0) = g_r^{(i),l}$ at $t = 0$.

In this paper, we study this process by means of the kernel matrix $\mathbf{Q}^{(i),l}(t)$, whose elements $Q_{r,s}^{(i),l}(t)$ determine the probability that a transition from state r to state s occurs in $[0, t]$. These elements are computed for the pairs (r, s) connected by an arc and are related to the Cumulative Distribution Functions (CDFs) $F_{r,s}^{(i),l}(t)$ of the transition times $T_{r,s}^{(i),l}$ from state r to state s by:

$$Q_{r,s}^{(i),l}(t) = \Pr\left\{\left(T_{r,s}^{(i),l} \leq t\right) \bigcap_{v \neq s} \left(T_{r,v}^{(i),l} > T_{r,s}^{(i),l}\right)\right\} = \int_0^t \mathrm{d}F_{r,s}^{(i),l}(u) \prod_{v \neq s} \left[1 - F_{r,v}^{(i),l}(u)\right], \tag{3}$$

where the intersection (and hence the product) regards the states v that can be reached from r in one step, according to the state space representation of the model (see Fig. 2). This formulation of a semi-Markov model is also referred to as *Competing Process Model* [8]. The kernel matrix allows to derive the one-step transition probabilities of the *embedded Markov chain*

$$\pi_{r,s}^{(i),l} = \lim_{t \to \infty} Q_{r,s}^{(i),l}(t), \tag{4}$$

and the distribution of the *conditional sojourn time* $\tau_{r,s}^{(i),l}$ of the process in state r if the next state is s, with mean $\overline{\tau}_{r,s}^{(i),l}$ and CDF

$$H_{r,s}^{(i),l}(t) = \frac{Q_{r,s}^{(i),l}(t)}{\pi_{r,s}^{(i),l}}. \tag{5}$$

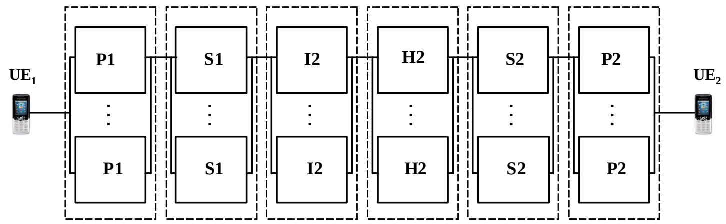


Figure 3: A series-parallel reliability block description of a redundant IMS signalling network.

Furthermore, let $\tau_r^{(i),l}$ be the *unconditional sojourn time* in state r, whose CDF is

$$H_r^{(i),l}(t) = \sum_s Q_{r,s}^{(i),l}(t), \tag{6}$$

while its expected value is

$$\overline{\tau}_r^{(i),l} = \sum_s \pi_{r,s}^{(i),l} \overline{\tau}_{r,s}^{(i),l}, \tag{7}$$

where the summation is extended to the states s reachable in one step from r, and let $\theta_{r,s}^{(i),l}(t)$ be the probability that the process $G^{(i),l}(t)$ starting in state r is in state s at time t, and $\theta_s^{(i),l} = \lim_{t\to\infty} \theta_{r,s}^{(i),l}(t)$ be the steady-state probability that the node is in state s, which by ergodicity does not depend on the initial state. Then the steady-state probabilities can be computed by

$$\theta_s^{(i),l} = \frac{p_s^{(i),l} \overline{\tau}_s^{(i),l}}{\sum_{s=-1}^{k^{(i)}} p_s^{(i),l} \overline{\tau}_s^{(i),l}}, \tag{8}$$

where $p_s^{(i),l}$ are the steady-state probabilities of the embedded Markov chain, that are the solutions of the system

$$\begin{cases} p_s^{(i),l} = \displaystyle\sum_{r=-1}^{k^{(i)}} p_r^{(i),l} \pi_{r,s}^{(i),l}, \quad \text{for } s = -1, ..., k^{(i)} \\ \displaystyle\sum_{s=-1}^{k^{(i)}} p_s^{(i),l} = 1. \end{cases} \tag{9}$$

Summing up, the key point is determining the kernel matrix elements $Q_{r,s}^{(i),l}(t)$ in (3); afterwards, it is possible to obtain $\pi_{r,s}^{(i),l}$ from (4) and $\overline{\tau}_r^{(i),l}$ from (7), and then $\theta_s^{(i),l}$ by using (8) and (9).

4. IMS SIGNALLING NETWORK AVAILABILITY EVALUATION

An IMS signalling network is an interconnection of different MSS (i.e., IMS servers), being in turn an MSS itself, and it is considered successful when it is able to meet a required performance level, or *demand*.

As pointed out in Section 2, two terminals can establish a communication if *all* the network elements managing the signalling messages are actually able to handle call set-up requests and answers, namely if the end-to-end signalling service is available. In most practical cases, some redundancy may be necessary to meet a demand $W(t)$. In this paper, we consider parallel redundancy of each signalling network element, as depicted in Fig. 3 by Reliability Block Description (RBD).

The system to be evaluated is a series-parallel one with *flow dispersion* [8] and the stochastic process describing its performance levels is

$$G(t) = \min_i \sum_{l=1}^{m^{(i)}} G^{(i),l}(t), \tag{10}$$

where $m^{(i)}$ is the number of parallel elements of the network subsystem i, and $i \in \{P1, S1, I2, H2, S2, P2\}$. Let $\mathbf{m}$ be a vector that contains all said numbers. The process $G(t)$ and therefore the variable G, representing its asymptotic behaviour as $t \longrightarrow \infty$, have values in the set

$$\{g_1, ..., g_K\}, \tag{11}$$

where $K = \prod_i \left(k^{(i)} + 2\right)^{m^{(i)}}$ is the number of different states of the overall MSS system. The steady-state probability vector is

$$\{p_1, ..., p_K\}, \tag{12}$$

where $p_k = \lim_{t \longrightarrow \infty} \Pr\{G(t) = g_k\} = \Pr\{G = g_k\}$, which is related to the values of G in (11).

The collection of pairs (g_k, p_k), $k \in \{1, ..., K\}$, completely determines the steady-state output performance distribution of the IMS-based signalling system, whose *instantaneous availability* $A(t)$ is [8]

$$A(t) = \Pr\{G(t) - W(t) \geq 0\}, \tag{13}$$

i.e., the probability that the MSS at $t > 0$ is in one of the *acceptable states*, where performance is not less than demand $W(t)$. Of course, the MSS initial state has vanishing practical influence on the availability as t grows. Therefore, given a constant demand level $W(t) = w$, the stationary signalling network availability $A_{IMS}(w)$ can be derived from the steady-state output performance distribution, viz:

$$A_{IMS}(w) = \sum_{k=1}^{K} p_k \mathbf{1}(g_k \geq w) = \sum_{g_k \geq w} p_k, \tag{14}$$

where $\mathbf{1}(True) = 1$, and $\mathbf{1}(False) = 0$.

4.1. A Universal Generating Function Approach for Availability Evaluation

A convenient approach to compute the system availability is based on the Universal Generating Function (UGF), also known as u-transform [14], [5]. The u-transform of the steady-state output performance of the IMS-based signalling system G is defined as a polynomial-shape function

$$u_{IMS}(z) = \sum_{k=1}^{K} p_k z^{g_k}, \tag{15}$$

where g_k and p_k are elements of the sets (11) and (12), respectively.

Two fundamental operators can be effectively introduced in order to determine the function $u_{IMS}(z)$ and all terms p_k and g_k appearing in (15). This computation relies on the steady-state behavior of all elements composing the system, that can be determined as described in Subsection 3.1.

In case of multi-state components with flow dispersion [8], the output performance of system composed by n elements connected in *parallel* is given by the sum of the capacities of each element, and the corresponding u-function can be calculated by using the following π operator:

$$u_p(z) = \pi\left(u_1(z), u_2(z), ..., u_n(z)\right) = \prod_{l=1}^{n} \left(\sum_{j_l=1}^{k_l} p_{j_l}^{l} z^{g_{j_l}^{l}} \right). \tag{16}$$

Thus, $u_p^{(i)}(z)$, $i \in \{P1, S1, I2, H2, S2, P2\}$, the u-function of the i-th subsystem of the IMS signalling system in Fig. 3 can be written as

$$u_p^{(i)}(z) = \pi\left(u_1(z), ..., u_{m^{(i)}}(z)\right) = \sum_{j_1=-1}^{k^{(i)}} \sum_{j_2=-1}^{k^{(i)}} \cdots \sum_{j_{m^{(i)}}=-1}^{k^{(i)}} \left(\prod_{l=1}^{m^{(i)}} \theta_{j_l}^{(i),l} z^{\sum_{l=1}^{m^{(i)}} g_{j_l}^{(i),l}} \right), \tag{17}$$

where $g_{j_l}^{(i),l}$ is the performance level of l-th parallel element of the network subsystem i in the state j_l in (2), and $\theta_{j_l}^{(i),l}$ is the corresponding steady-state probability, computed by (8).

On the other hand, when the components are connected in *series*, the overall output performance is simply the minimum of individual components performance levels. Then, the UGF for m series components can be computed by the following σ operator:

$$u_s(z) = \sigma\left(u_1(z), u_2(z), ..., u_m(z)\right), \tag{18}$$

where, for a pair of components connected in series,

$$\sigma\left(u_1(z), u_2(z)\right) = \sigma\left(\sum_{i=1}^{I} \epsilon_i z^{e_i}, \sum_{j=1}^{J} \varphi_j z^{f_j}\right) = \sum_{i=1}^{I}\sum_{j=1}^{J} \epsilon_i \varphi_j z^{\min\{e_i, f_j\}}, \tag{19}$$

and where the parameters e_i and f_j are the performance levels of the two components, while ϵ_i and φ_j are the corresponding steady-state probabilities, $\forall i = 1, .., I$ and $\forall j = 1, .., J$. The UGF in (18) can be computed by applying recursively (19) from 2 to m series components.

Finally, the u-function of the IMS signalling system in (15) can be obtained by putting (17) $\forall i \in \{P1, S1, I2, H2, S2, P2\}$ into (18), viz:

$$u_{IMS}(z) = \sum_{k=1}^{K} p_k z^{g_k} = \sigma\left(u_p^{(P1)}(z), u_p^{(S1)}(z), u_p^{(I2)}(z), u_p^{(H2)}(z), u_p^{(S2)}(z), u_p^{(P2)}(z)\right). \tag{20}$$

The polynomial-shape function thus obtained provides immediately the signalling system performance levels, which are the exponents of z, and their steady-state probabilities, which are the corresponding coefficients; in turn they provide the stationary IMS signalling system availability A_{IMS} according to (14).

4.2. A Redundancy Optimization Problem

Letting $C^{(i)}$ be the cost of a network element i, the overall cost of the signalling system is

$$C_{IMS}(\mathbf{m}) = \sum_i m^{(i)} C^{(i)}. \tag{21}$$

The problem is to find the minimal cost system configuration $\mathbf{m}^*$ that provides a required level of stationary availability A_0 for the IMS system, viz.

$$\mathbf{m}^* = \arg\{C_{IMS}(\mathbf{m}) \longrightarrow \min | A_{IMS}(w, \mathbf{m}) \geq A_0\}. \tag{22}$$

This combinatorial optimization problem is typically referred to as *redundancy optimization problem* [15]. Typical approaches to solve the problem (22) are based on some heuristic techniques [8].

5. A NUMERICAL EXAMPLE

A numerical example of availability analysis of an IMS-based signalling network is now presented, and a minimal deployment cost network configuration is computed by solving a redundancy optimisation problem with a given availability requirement, as described in Sections 3 and 4. For comparative purposes, numerical example data are selected to be the same of those in [5]. In line with [5], we assume that all nodes in the IMS-based signalling network have one and the same cost, that we suppose, with no lack of generality, equal to 1, i.e. $C^{(P1)} = C^{(P2)} = C^{(S1)} = C^{(S2)} = C^{(I2)} = C^{(H2)} = 1$. Furthermore, it is assumed that all nodes in the signalling network have the same characteristics, such as the number of concurrent server instances and their serving capacity. Thus, we have $k^{(i)} = k = 4$ and $n^{(i)} = n = 1000$ sessions per second, for each $i \in \{P1, S1, I2, H2, S2, P2\}$.

Regarding node transition times distributions, we assume that they are the same with the corresponding software/hardware failure transitions of the CTMC presented in [5]. In particular (see Fig. 2(b)),

Table 1: Numerical Example Data.

Parameter	**Value**
k	4
n	1000 sessions/sec
λ_s	7.716×10^{-7} sec^{-1}
λ_c	6.342×10^{-8} sec^{-1}
μ_s	5.556×10^{-3} sec^{-1}
w	3800 sessions/sec
a	3
b	8063 sec

$T_{j,j-1}^{(i)}$ from the state j to the state $j-1$ ($j = 0, 1, .., k$) and $T_{j,-1}^{(i)}$ (core failure) are exponentially distributed, with parameters $j \cdot \lambda_s^{(i)} = j \cdot \lambda_s$ and $\lambda_c^{(i)} = \lambda_c$, respectively. Hence, the corresponding transition time CDFs are the following: $F_{4,3}(t) = 1 - \exp\{-4\lambda_s t\}$, $F_{3,2}(t) = 1 - \exp\{-3\lambda_s t\}$, $F_{2,1}(t) = 1 - \exp\{-2\lambda_s t\}$, $F_{1,0}(t) = 1 - \exp\{-\lambda_s t\}$, $F_{r,-1}(t) = 1 - \exp\{-\lambda_c t\}$, $r = 0, ..., 4$, defined for $t \geq 0$. The parameters are related to the failure rates $\lambda_s = 7.716 \times 10^{-7}$ sec^{-1} (corresponding to 2 software instance failures per month) and $\lambda_c = 6.342 \times 10^{-8}$ sec^{-1} (corresponding to 2 core failures per year) of the single server instance and the core part of node, respectively, of the CTMC in [5].

A similar behaviour is assumed for software instances repair transitions. Indeed, $T_{j,j+1}^{(i)}$ from the state j to the state $j+1$ ($j = 0, 1, .., k-1$) are exponentially distributed, with parameters $(k-j) \cdot \mu_s^{(i)} = (k-j) \cdot \mu_s$. Hence, software repair time CDFs are: $F_{0,1}(t) = 1 - \exp\{-4\mu_s t\}$, $F_{1,2}(t) = 1 - \exp\{-3\mu_s t\}$, $F_{2,3}(t) = 1 - \exp\{-2\mu_s t\}$, $F_{3,4}(t) = 1 - \exp\{-\mu_s t\}$, defined for $t \geq 0$. The parameters are related to the software instance repair rate $\mu_s = 5.556 \times 10^{-3}$ sec^{-1}, corresponding to a mean repair time of 3 min, which is in keeping with a typical software instance repair action consisting of killing/restarting a process.

On the other hand, unlike [5], we assume that the transition times related to repair actions on the node core part can be distributed according to a non-exponential model, because operations of this kind are often devoted to solve hardware problems and typically require completely different procedures performed on-site by engineers. Hence, the Weibull distribution has been chosen in keeping with flexible and popular models describing failure data in technological systems affected by aging. Furthermore, it encompasses the exponential model as a special case. However, other distribution models could be adopted for this repair time, such as the Erlang or the lognormal distribution [16], or any other fitting data collected in real IMS systems.

In detail, the time $T_{-1,4}^{(i)}$ of the transition in Fig. 2(b) from state $r = -1$ to state $s = 4$ (core repair) has CDF

$$F_{-1,4}^{(i)}(x) = 1 - \exp\left\{-\left(x/b^{(i)}\right)^{a^{(i)}}\right\}, \quad x \geq 0,$$

whose shape parameter $a^{(i)} = a$ is (arbitrarily) fixed, while the scale parameters $b^{(i)} = b$ is selected in order to provide the same mean value δ of the corresponding transition time of the CTMC model in [5]. Indeed, the scale parameters can be computed by $b = \delta/\Gamma(1 + 1/a)$. Assuming $a = 3$, we obtain $b = 8062.9$ sec, since $\delta_{-1,4} = 7200$ sec. Finally, the service demand level is $w = 3800$ sessions per sec, chosen to be close to the highest performance level of a single node $k \cdot n = 4000$.

Table 1 summarizes data given in this numerical example.

The steady-state probabilities of a single node are computed by (8). By assuming the required availability equal to $A_0 = 1 - 10^{-6}$, we solve numerically the problem (22) by means of (16)-(20). Six equivalent solutions have been found by the numeric algorithm, corresponding to the system configurations $\mathbf{m}^*$ consisting of the combinations of five nodes with 3 parallel elements and the remaining one with 2 parallel elements. The output performance levels of the resulting signalling systems and their

Table 2: Steady-State Probabilities and Performance Levels for Configuration $\mathbf{m}^*$.

Probability	**Performance** (sessions/sec)
2.088×10^{-7}	0
9.806×10^{-15}	1000
1.059×10^{-10}	2000
5.083×10^{-7}	3000
9.150×10^{-4}	4000
1.508×10^{-10}	5000
5.427×10^{-7}	6000
1.116×10^{-3}	7000
0.997967	8000

steady-state probabilities are reported in Table 2, and are the same as those reported in [5] related to a Continuous Time Markov Chain. The availability of IMS system obtained by a semi-Markov model is indeed $A_{IMS}(w, \mathbf{m}^*) = 1 - 7.172 \times 10^{-7}$, and the cost is the same: $C_{IMS}(\mathbf{m}^*) = 17$. This is due to the fact that the semi-Markov process asymptotic behavior is governed by conditional mean sojourn times, which do not not vary if a Weibull transition time for core part repair activities is adopted instead of an exponential one with the same mean value, because this transition is not concurrent with any other one.

6. CONCLUSION

An availability analysis has been presented for an IMS-based signalling network of a next generation mobile system, useful to design a signalling network able to maintain high availability levels and to minimize the deployment cost. Every signalling network node has been modelled as a semi-Markov process with different states corresponding to different performance levels: this model is more flexible than Markov models to meet real operation data. The performance metric under study is the number of call set-up requests handled by the signalling system. A parallel redundancy optimization problem has been solved numerically by computing the u-function of the signalling system, given a constant service demand level and a high system availability target, and a signalling network configuration has been found that minimizes the deployment cost.

However, no difference between steady-state behaviours of Markov and a semi-Markov model with one non-exponential transition (core part repair) arises. This might be expected since Weibull transition times are not concurrent with any other time in the proposed MSS semi-Markov model, and hence system performance depends only on core repair mean time and not on its distribution.

Further research activities are planned toward more realistic and sophisticated models that can cope with non-exponential repair and/or time to failures, additional performance metrics, such as the call set-up delay and the call blocking probability, and to the validation of Markov or non-exponential multi-state models in real scenarios of maintenance operations of a 3G network, inferring model parameters from actual failure/repair actions time series.

REFERENCES

[1] ITU-T. ITU-T next generation networks global standards initiative, 2009. http://www.itu.int/ITU-T/ngn/.

[2] Guida M, Longo M, and Postiglione F. Reliability and survivability methodologies for next generation networks. In *Proc. MoMM/BWWCA 2008*, pages 335–340, Linz (Austria), November 24-26, 2008.

[3] Liu Y and Trivedi K S. A general framework for network survivability quantification. In *12th GI/ITG Conference on Measuring, Modeling and Evaluation of Computer and Communication Systems (MMB) and 3rd Polish-German Telegraphic Symposium (PGTS)*, pages 369–378, Dresden, Germany, 2004.

[4] Meyer J F. On evaluating the performability of degradable computing systems. *IEEE Transactions on Computers*, 28(8):720–731, 1980.

[5] Guida M, Longo M, and Postiglione F. Reliability analysis of next generation mobile networks. In Briš, Guedes Soares, and Martorell, editors, *Reliability, Risk and Safety: Theory and Applications (ESREL'09)*, volume 3, pages 1999–2006. Taylor & Francis Group, London, 2010.

[6] 3GPP. Tech. spec. TS 23.228 version 8.5.0, IP multimedia subsystem; Stage 2 (Release 8), 2008. http://www.3gpp.org/ftp/Specs/html-info/23228.htm.

[7] Ross S M. *Stochastic processes*. Wiley, New York, 1996.

[8] Lisniansky A and Levitin G. *Multi-state system reliability*. World Scientific, Singapore, 2003.

[9] Rosenberg J D et al. Session Initiation Protocol (SIP), 2002. IETF RFC 3261, http://www.ietf.org/rfc/rfc3261.txt.

[10] Nicol D M, Sanders W H, and Trivedi K S. Model-based evaluation: From dependability to security. *IEEE Transactions on Dependable and Secure Computing*, 1(1):48–65, 2004.

[11] Pant H, Chu C-H K, Richman S H, Jrad A, and O'Reilly G P. Reliability of next-generation networks with a focus on IMS architecture. *Bell Labs Tech. J.*, 12(4):109–126, 2008.

[12] Ciardo G, Marie R A, Sericola B, and Trivedi K S. Performability analysis using semi-Markov reward processes. *IEEE Transactions on Computers*, 39(10):1251–1264, 1990.

[13] Logothetis D, Trivedi K S, and Puliafito A. Markov regenerative models. In *Proc. IPDS'95*, pages 134–142, Erlangen, Germany, 24 - 26 April, 1995.

[14] Ushakov I A. A universal generating function. *Soviet Journal of Computing System Science*, 24(5):37–49, 1986.

[15] Ushakov I A. Optimal standby problems and a universal generating function. *Soviet Journal of Computing System Science*, 25(4):79–82, 1987.

[16] Gnedenko B and Ushakov I. *Probabilistic reliability engineering*. Wiley, New York, 1995.