QoS Perceived by Users of Ubiquitous UMTS:Compositional Models and Thorough Analysis

QoS Perceived by Users of Ubiquitous UMTS: Compositional Models and Thorough Analysis

Andrea Bondavalli, Paolo Lollini, Leonardo Montecchi
Università degli Studi di Firenze,
Dipartimento di Sistemi e Informatica,
viale Morgagni 65, I-50134, Firenze, Italy,
Email: {bondavalli, lollini, lmontecchi}@unifi.it

Abstract— **This paper provides a QoS analysis of a dynamic, ubiquitous UMTS network scenario in the automotive context identified in the ongoing EC HIDENETS project. The scenario comprises different types of mobile users, applications, traffic conditions, and outage events reducing the available network resources. Adopting a compositional modeling approach based on Stochastic Activity Networks (SAN) formalism, we analyze the Quality of Service (QoS) both from the users' perspective and from the mobile operator's one. The classical QoS analysis is enhanced by taking into account the congestion both caused by the outage events and by the varying traffic conditions. The impact of users' mobility on the selected QoS indicators is further investigated combining the SAN modelling approach with an ad-hoc mobility simulator, which also allows to refine the model representing the UMTS network behavior.**

Index Terms— **QoS analysis, UMTS networks, partial outages, compositional modeling, stochastic activity networks, mobility simulator.**

I. Introduction

Ubiquitous infrastructures are typically composed by a high number of mobile devices that move within some physical areas, while being connected to networks by means of wireless links. The supported mobile-based applications should be capable to provide the expected services in a dependable way, and maintaining the required Quality of Service (QoS) levels. The QoS from the users' viewpoint mainly depends on the attributes of the communication services that, in turn, are directly influenced by traffic and mobility. Hence, there is a need to take into account explicitly these characteristics in the modeling and quantitative evaluation of such kind of dynamic systems. The paper takes as motivating example a use-case scenario defined in the ongoing EC HIDENETS project [1]. The analyzed system is characterized by a UMTS communication network composed by several partially overlapping cells, and by a set of users (i.e., cars and emergency vehicles, equipped with UMTS network devices) moving through the network and requiring different UMTS-based applications (e.g., voice call and entertainment). The user perceived QoS level should always be higher than a minimum level, and this aspect becomes particularly critical when emergency situations are considered, for example in the case of an ambulance that is using a streaming application to transmit the ECG traces of an injured person while moving to the hospital. Since the user perceived QoS level depends on the availability of network resources, base stations' faults are also considered. More in detail we allow the presence of partial outages that may affect the availability of the UMTS resources.

In this paper we propose a compositional modeling approach based on Stochastic Activity Networks (SAN) to assess the QoS provided over complex, ubiquitous and dynamic infrastructures. The modeling approach, preliminarily presented in [2], is here refined and extended in the following directions: i) Exploiting the modularity of the modeling framework, it defines the approach for integrating the output produced by an ad-hoc mobility simulator into the modeling process itself, thus allowing to capture more complex and detailed mobility dynamics that may heavily affect the analyzed QoS indicators; ii) It refines the way in which the interference produced by each user is calculated, thus obtaining a more detailed and faithful model of the UMTS network; iii) It presents further numerical evaluations showing the impact of the mobility and network model refinement on the selected QoS indicators.

The interaction between the ad-hoc mobility simulator and the SAN models also constitutes a concrete example of application of a holistic evaluation approach, where the synergies and the characteristics of different evaluation techniques (here mobility simulators and SAN models) are exploited to capture system characteristics at a more detailed level, thus enabling a more refined QoS analysis that could be hardly obtained using a single technique. In the literature, several attempts have been made to couple modeling and simulation activities, most of them using low-level simulations (or even experimental approaches) to provide more accurate specific parameters to be used by higher-level analytic models (e.g., [3]). More general approaches have been developed in DBench [4] and HIDENETS, the former focusing on the interactions between modeling and experimentation for dependability benchmarking, the latter proposing an evaluation workflow for distribute mobile applications which integrates several tools and model transformation steps (see [5] and [6]). Nevertheless, to the best of our knowledge, there are not previous works fully combining and automatically integrating stochastic models and detailed simulations in the solution process itself, as done in this paper.

doi:10.4304/jsw.4.7.675-685

The rest of this paper is organized as follows. Section II provides the description of the analyzed system and it outlines the corresponding QoS measures of interest. The main UMTS aspects influencing the QoS analysis are discussed in Section III, and the modeling process is detailed in Section IV. Section V presents the basic models and the related numerical evaluations, while the models refinement and its impact on the selected QoS indicators are presented and discussed in Section VI. Finally, the conclusions are drawn in Section VII.

II. System Context and QoS Indicators

HIDENETS [1] is an ongoing EC project addressing the provision of available and resilient distributed applications and mobile services with critical requirements on highly dynamic and possibly unreliable open communication infrastructures. A set of representative use-case scenarios has been identified, each one composed by different applications (mostly selected from the field of car-to-car and car-to-infrastructure communications), different network domains (ad-hoc/wireless multi-hop domains, infrastructure network domains), different actors (end users, servers, routers, gateways), and characterized by different failure modes and challenges. In the following we give a brief description of the "car accident" scenario, which is analyzed in this paper and used as motivating example to describe the modeling process.

A. Definition of the "Car Accident" Use-case Scenario

The "car accident" use-case scenario evolves around a scene with an accident on a road, involving cars. The use-case covers mainly what happens after the accident but also involves some issues directly before and during the accident. The analyzed network scenario is composed by a set of overlapping UMTS cells covering a highway, and a set of mobile network devices (embedded or inside cars and emergency vehicles) moving in the highway and requiring different UMTS class of services (e.g., conversational, interactive, and background).

Immediately before the accident, several applications are used by the different mobile users, like entertainment and voice call. Right after the accident, many people may try to call the emergency services, call home, and send text and multimedia messages. Some time after the accident, an ambulance is approaching. Arriving at the place of the accident, and heading back to the hospital with the injured, there will be a need to transmit information on the positioning of the ambulance to communicate that it is approaching the hospital and at the same time maintain a multimedia connection with the medical expertise by use of voice, video and data transmission ("access to medical expertise" application).

The concrete UMTS scenario under analysis is depicted in Figure 1. Four base stations are considered: A, B, C and D. The base stations are subject to faults, which may reduce their available network resources. The users are moving in two different road lanes: part in the left to right lane (from A to D) and part in the right to left one (from D to A). We assume that the accident occurs in the C zone, in the left to right lane, forcing other users approaching that area to stop until the ambulance arrives, the crash site is cleaned and the normal traffic flow restored. The emergency vehicle heads back to the hospital towards the A zone where we suppose the hospital is located.

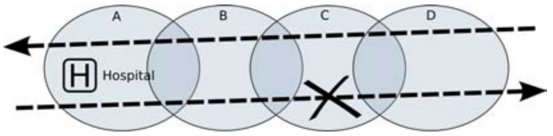

Figure 1. The analyzed scenario

Concerning the available UMTS services, we suppose that a generic user can use three different services (Telephony, Web Browsing and File Transfer), while the ambulance uses the "access to medical expertise" application that consists of two simultaneously running services (Emergency Streaming to transmit the ECG traces, and Emergency Video-conference to fully interact with the hospital), having higher requirements in term of signal to interference ratio with respect to the non-emergency services. The services mainly differ for the activity factor, the uplink and downlink throughput and the required signal-to-interference ratio. These parameters are summarized into a single value that represents the load factor increment they produce on the network (δ_{ul} and δ_{dl} parameters of Equation (1), Section III).

B. QoS Indicators

The measures of interest concern the QoS levels both from the users' perspective and from a mobile operator's point of view. The QoS level perceived by users (both normal cars and emergency vehicles) depends on their capability to successfully use the network services when required and for the time required. The users involved in the traffic-jam should be capable to call home, while the ambulance should be capable to maintain the multimedia connections while moving towards the hospital. Typical user-oriented QoS indicators are the following:

- The probability that a service request is successfully completed (P_{succ});
- The probability that a service request is blocked (P_{block}) or dropped (P_{drop}).

The network load factor is another system aspect that deserves special attention. Right after the accident, the behavior of the users involved in the consequent traffic-jam changes from normal to emergency, for example intensifying the service requests and trying to call the emergency services and call home, and this may cause congestion in the radio access network. In this context, typical mobile operator-oriented indicators are the following:

- The load factor, both in uplink (η_{ul}) and downlink (η_{dl});

- The number of allocated traffic channels, which corresponds to the average number of served users.

III. Communication Level Aspects Influencing the QoS Analysis

In this section we focus on the communication level aspects related to the "car accident" use-case, and in particular on three UMTS characteristics having important effects on the QoS: the *random-access procedure*, the *admission control* strategy and the *soft handover* mechanism. These characteristics mainly influence the so called "connection level" QoS, which are the quality indicators related to the connectivity properties of the network, like the call blocking or dropping probability.

When a user needs a service from the UMTS network, its User Equipment (UE) sends a channel request to the network through the Physical Random Access CHannel (PRACH), a specific channel dedicated to the uplink transmission of channel request. The access method, based on a *random-access procedure* (RACH), may cause collisions among requests by different UEs, thus worsening the expected QoS (e.g., see [7] for more details on this aspect).

The *admission control* strategy is needed to decide whether a new service request can be accepted based on the available network "capacity". Once the network receives the channel request, it performs the admission control procedure to decide if a traffic channel can be allocated to this new request. The goal is, in general, to ensure that the interference created after adding a new call does not exceed a pre-specified threshold, thus preventing the QoS to degrade below a certain level. There are several types of admission control algorithms studied in the literature, each one having different properties and aiming at optimizing different network parameters (e.g., [8]). Here we consider an admission control algorithm based on the load factor of the UMTS cell: a new call is accepted if the load factor level reached after adding the call does not exceed a pre-specified threshold, both in uplink and in downlink. That is:

$$\begin{aligned} \eta_{ul} + \delta_{ul} &\leq \eta_{ul\_threshold}, \\ \eta_{dl} + \delta_{dl} &\leq \eta_{dl\_threshold}, \end{aligned} \tag{1}$$

where η_{ul}, δ_{ul} and $\eta_{ul\_threshold}$ (or η_{dl}, δ_{dl} and $\eta_{dl\_threshold}$) are, respectively, the cell load factor before the admission of the new call, the load factor increment due to the admission of the new call and the pre-specified threshold level in uplink (or downlink). According to well-known UMTS equations (e.g., see [9], [10]), the load factor increment in downlink and uplink for a given user and service can be computed as:

$$\delta_{dl} = \frac{(E_b/N_0) \cdot R \cdot \nu}{W} \cdot \big((1-\alpha) + i\big), \tag{2}$$

$$\delta_{ul} = \frac{1}{1 + \dfrac{W}{(E_b/N_0) \cdot R \cdot \nu}} \cdot (1 + i), \tag{3}$$

where $\big(E_b/N_0\big)$ is the required service quality, R the service data rate, ν the service activity factor, α the average orthogonality factor, W is the chip rate (the total bandwidth, equal to 3.84 Mcps for UMTS) and i is the other-to-own interference ratio at current user position. While most of the parameters in Equations (2) and (3) are directly obtained from the service class or other network variables, the i factor depends on the power of received signals, which in turn depends on the user position in the scenario.

Another key aspect to be addressed is *soft handover*, a feature of the 3rd generation mobile networks, where a User Equipment can have two or more simultaneous connections with different cells (or cell sectors) and receive from them the same information signal. The signal received from different sources is then combined using rake receivers and under certain conditions this results in a amplified signal and better link quality. Beside providing better link quality, soft handover is also a key point in maintaining an ongoing service call, since it provides seamless switching between base stations.

IV. Modelling Process

In such ubiquitous landscape, system complexity comes out to be a paramount challenge to address from a number of different view points, including dependability and QoS evaluation. Complexity can be mastered by a modelling methodology able to detail only the relevant system aspects, allowing numerical results to be effectively computable. The complexity of models depends on the dependability measures to be evaluated, the modelling level of detail, and the stochastic dependencies among the components. Several works have been presented in the literature trying to cope with the complexity problem (e.g., see [11] for a survey). Some techniques try to circumvent the generation of large models using, for example, state truncation methods, lumping techniques and model decomposition techniques. Other techniques support the practical generation of large state-space models by structured composition and hierarchical modeling approaches, whose basic idea is to build the system model from the *composition* of submodels describing system components and their interactions.

In this paper we adopt a compositional modeling approach based on Stochastic Activity Networks (SAN [12]), which are stochastic extensions to Petri Nets. The composition operators available for SAN are the *join* and *replicate* operators: the first is used to compose different system models sharing some places, while the second is used to combine multiple identical copies of a submodel, which are called replicas. Another key point of the modeling approach is the "model parametrization". Following the object oriented philosophy, we develop some "template" SAN models describing the general behavior of the main system components. The overall model results from the composition of some "instances" of such classes, where an instance is a specification of a template model with a proper parameters' setting. For example, the parametric interface of the base station template model consists of a set of (common or

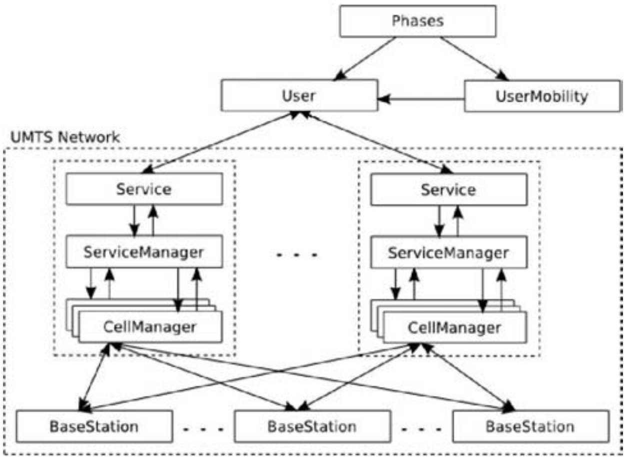

Figure 2. Atomic models and their interactions

extended) places representing the outage severity, the outage duration and the maximum load factor, which are properly initialized with different values to build the different base station instances. Using this approach we avoid duplicating the code and the structure of similar models, which is a very time-consuming and error-prone process; as a consequence, the overall model is easier to be modified and it can be more easily adapted to represent different scenarios.

In Figure 2 we have depicted the main basic SAN models (called "atomic" models in the SAN language) with their dependency relations (the arrows). An arrow from model X to Y means that model X can influence the stochastic behavior of model Y or, equivalently, that the state of Y may depend on the state of model X. In the following we outline the main system aspects captured by the different models.

- **Phases** atomic model. It represents the sequence of periods (phases) composing the system lifetime, each one characterized by diverse applications running, diverse types of users's behavior (normal behavior, before the car accident, or emergency behavior, right after the car accident) and different dependability properties to be ensured.
- **User** atomic model. It describes the user's behavior mainly in terms of services requested, duration of the services and idle periods.
- **UserMobility** atomic model. It represents the user movement across the UMTS network scenario.
- The "UMTS Network" model consists of several instances of the **BaseStation** atomic model and a number of models representing the available services. A network service is represented using three kinds of atomic models: **Service**, **ServiceManager** and **CellManager**. The Service atomic model represents the upper network layers and it is directly connected with the User model. When the user requests a network service, the User model interacts with the respective Service model which serves as interface between the user and the network. The Service model then asks for the needed resources to the ServiceManager atomic model, which handles the soft handover mechanisms allowing user to be served by multiple base stations. This is achieved using several instances of the CellManager atomic model, which serve as interfaces between the ServiceManager atomic model and each BaseStation model. Finally the BaseStation model represent a UMTS base station, with failure and repair activities, and holds the current base station state, like its current load factor and the number of allocated channels. In case of outage events this model also implements the congestion control algorithm, which drops a certain number of connections if the current load factor exceeds the remaining available system resources.

Once such basic template models have been developed, several different scenarios can be easily obtained resembling different network topologies, users' behaviors, users' mobility patterns and available applications. Therefore, the modularity of the modeling framework improves both the readability and the maintenance of the models, as well as their reusability.

V. Basic Models and Numerical Evaluations

In this section we detail the actual implementation of two SAN models representing, respectively, the user mobility aspects (Section V-B) and the overall model for the "car accident" use-case scenario (Section V-C). An exhaustive and detailed description of all the other models can be found as a technical report in [13] and it is not reported here for the sake of brevity. Then we will define the setting of the numerical parameters and present the numerical evaluations (Section V-D). First of all, in the following Section V-A we will detail the two main adopted modeling assumptions concerning user mobility and load factor estimation, which will be refined in the last part of the paper.

A. Modeling Assumptions concerning User Mobility and Load Factor Estimation

The UserMobility model defines how a user moves across the network. We assume that the network can be split into several "zones" characterized by a given active set of available base stations, and the user moves through these zones. When a user enters in a new zone, its active set is then updated. The time needed to cross a zone is uniformly distributed in the interval $[0.8t; 1.2t]$, where t is the mean time to cross the cell and its value is computed from the cell size and the average speed of the user (see Section V-D for the actual setup).

To perform the admission control procedure, the load factor increments detailed in Equations (2) and (3) have to be computed for each user and each network service. We assume that users which are using the same service in the same zone generate the same (fixed) amount of load (and interference) on the involved base station(s). In other words, the i parameter (i.e., the other-to-own interference ratio) has been set to an average value not depending

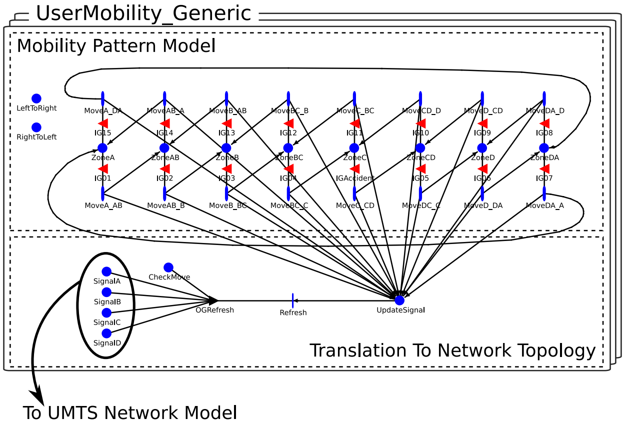

Figure 3. Example of UserMobility SAN atomic model and its submodels

on the current user position. In a zone covered by two or more base stations an user can take advantage of soft handover and connect to two or more base stations. When this happens, the load generated on each base station is lower (but still fixed) than the load that would be generated having a single connection (see Section V-D for the actual setup).

B. The UserMobility Model

In this section we give a brief description on how the mobility is handled in the modeling framework. The mobility pattern of every single user included in the scenario is modeled by an instance of the UserMobility atomic model. UserMobility atomic models belonging to different users may have different implementations, leading to different mobility patterns, or may be instances of the same "template" model, leading to users with the same mobility behavior. In the case of the car accident scenario we distinguished between the mobility pattern of the emergency vehicle, captured by the UserMobility_Ambulance model, and the behavior of the other "normal" users (cars), captured by a set of replicas of the UserMobility_Generic model.

Figure 3 shows the UserMobility_Generic model built for the car accident scenario. It can be split in two logically distinct submodels. The "Mobility Pattern" part (upper part of the picture) implements the user mobility rules, updating the user position in the scenario when required. The mobility scenario is represented through different zones, covered by one or more base stations. There is a place for each of them and uniformly distributed activities model the movement between these zones (as detailed in Section V-A). When one of these activities fires a token is added in place UpdateSignal and the "Translation" part (depicted in the lower part of the picture) is triggered. This part performs a mapping between the user position in the scenario and the network topology. For each UMTS base station X this model contains a place named SignalX, which is shared with the UMTS network model and contains a token if the user is capable to receive the signal of base station X. The translation part simply fills these places with the correct values, based on the current user position (i.e., the zone where the user is located).

C. The Composed Model for the Overall Scenario

As described in Section II-A, the analyzed scenario consists of four partially overlapping base stations and five services: services 1 (Telephony), 2 (Web Browsing) and 3 (File Transfer) are services for the "normal" users, while 4 (Emergency Streaming) and 5 (Emergency Videoconference) are services used by the emergency vehicle and will have higher requirements in term of signal to interference ratio. Figure 4 depicts the corresponding composed model. The composition involves three join levels. Starting from the lower level (the boxed part of the figure), joins relative to different services are shown, each one formed by four CellManager models (one for each base station), a Service and a ServiceManager model, as sketched in Figure 2. In the second level the services are joined with the respective user models, so services 1-3 are composed with User_Generic and UserMobility_Generic (on left part of the figure), while services 4-5 are composed with User_Ambulance and UserMobility_Ambulance (on right part). The generic user is then replicated as needed and added to the top-level join, which also includes the ambulance join, the four BaseStation models, the Phases model and the Startup model (used to initialize the multiple instances of the other atomic models with the proper values).

The advantage of using model parametrization is evident considering the effort required for the model construction process. To build the composed model shown in Figure 4, 40 atomic models are needed (exactly 20xCellManager, 5xServiceManager, 5xService, 4xBaseStation, 1xStartup, 1xPhases, 1xUser_Generic, 1xUserMobility_Generic, 1xUser_Ambulance and 1xUserMobility_Ambulance). Using model parametrization we need to create the basic template atomic models only, one for each type. For this scenario only 10 atomic models have been built, and those depicted in Figure 4 are just instances of these basic 10 models. Once the basic template models have been defined, we can easily build and analyze different scenarios with a very small effort. For example, deleting the JoinAmbulance composed model in Figure 4 we can limit the analysis to normal (not emergency) services, while adding another base station (thus obtaining a different network topology) would simply consist in adding another CellManager atomic model to each JoinSV composed model, and another BaseStation atomic model (BaseStationE) to the CarAccident composed model.

D. Settings and Results

We describe now some of the results obtained through the solution of the overall model previously described. A transient analysis has been performed, using the simulator provided by the Möbius tool [14]. Each point of the graphs has been computed as a mean of at least 1000

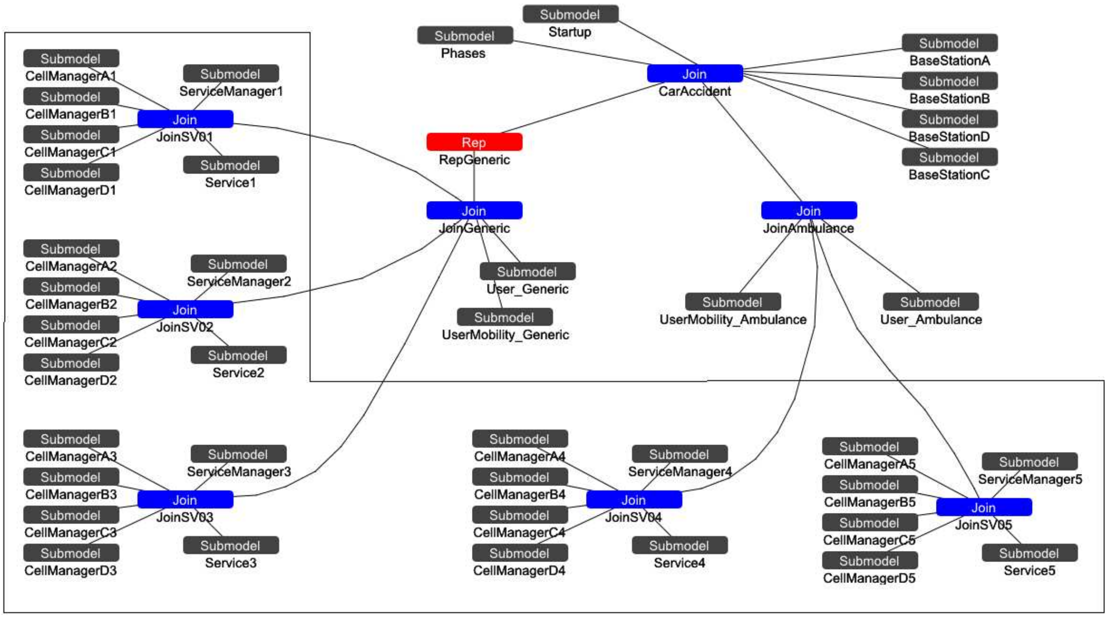

Figure 4. The overall composed model corresponding to the analyzed use-case scenario

		Service1	Service2	Service3	Service4	Service5
Bitrate (Kbps)	Downlink	12.2	57.6	384	0	64
	Uplink	12.2	57.6	64	64	64
Activity Factor	ν	0.67	1	1	1	1
E_b/N_0 (dB)	Downlink	7.5	5.5	3	8	7.5
	Uplink	5	4	2.5	5	4.5
Soft Handover Gain (dB)	Downlink	2	1.5	1	1.5	2
	Uplink	2	1.5	1	1.5	2
Load factor increment (no SHO)	Downlink (δ_{dl})	0.01357	0.07012	0.25500	0	0.10625
	Uplink (δ_{ul})	0.01632	0.08774	0.06200	0.11923	0.10814
Load factor increment (with SHO)	Downlink (δ_{dl})	0.00995	0.05100	0.17000	0	0.07792
	Uplink (δ_{ul})	0.00984	0.05602	0.03781	0.08543	0.06200

TABLE I.
MAIN SERVICE PARAMETERS AND CORRESPONDING LOAD FACTOR INCREMENTS, WITH OR WITHOUT SOFT-HANDOVER

simulation batches, converging within 95% probability in a 0.1 relative interval, run on a AMD Athlon XP 2500+ PC (2Gb RAM). The setting of the model's parameters has been mainly derived from [15] and adapted to the analyzed scenario. According to the highway scenario, common values experienced in suburban macro-cells have been used; with reference to Equations (2) and (3), the orthogonality factor α has been set to 0.7, while the average other-to-own interference ratio (i factor) has been set to 0.55. For each service, the other parameters used to compute δ_{ul} and δ_{dl} are shown in Table I, together with the obtained δ_{ul} and δ_{dl} values. The maximum load factor in uplink ($\eta_{ul\_threshold}$) and downlink ($\eta_{dl\_threshold}$) has been set, respectively, to 0.65 and 0.8. Each base-station has a coverage area of 2 Km, and 25% of the cell radius is overlapping with the adjacent cell. Whenever not differently specified, we consider a total of 50 cars moving in the scenario (and 1 ambulance), with an average speed of 90 Km/h (120 Km/h for the ambulance) when not involved in the traffic jam caused by the car accident. Moreover, we suppose that the car accident happens in the C cell at time t=10500 sec., and that the ambulance stays 600 sec. at the crash site before heading back to the hospital. The complete set of model's parameters and their setting can be found in [13], and it is not reported here for the sake of brevity.

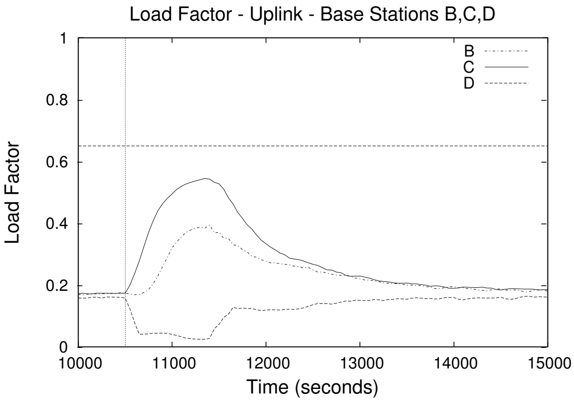

Figure 5. Load Factor of different base stations (uplink)

The uplink load factors of base stations near the accident zone are depicted in Figure 5. The vertical line represents the instant of time when the accident occurs, while the horizontal ones represent the maximum

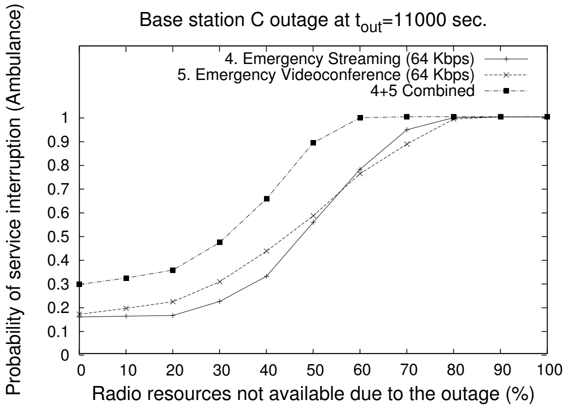

Figure 6. Probability of interruption of emergency services, with C cell outage

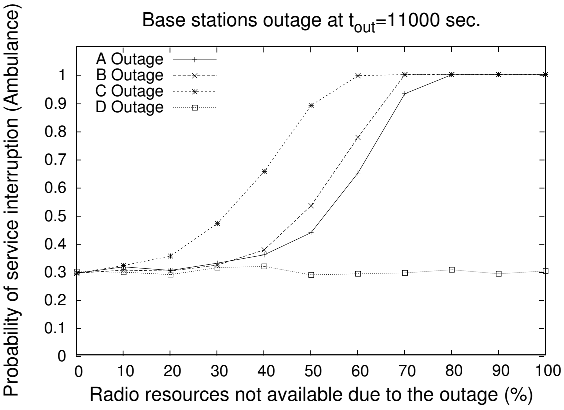

Figure 8. Probability of interruption of ambulance services (4+5 combined) when a single base station fails

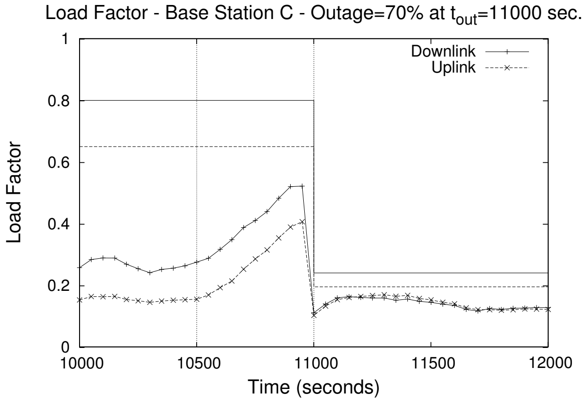

Figure 7. Load Factor of the base station C, with outage=70%

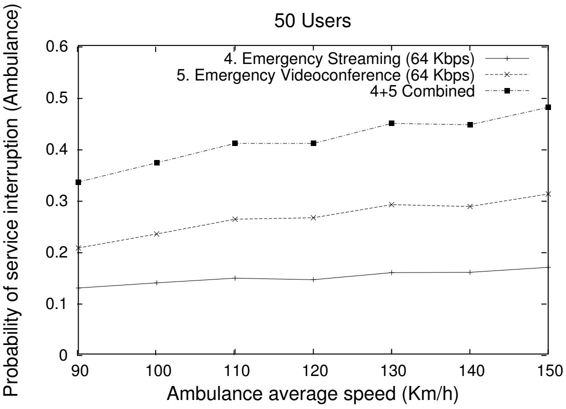

Figure 9. "Access to medical expertise" interruption probability

allowed load factor in uplink ($\eta_{ul\_threshold}$). Right after the accident the C cell becomes rapidly congested due to the users that are stopped in that area (due to the consequent traffic-jam). The congestion phenomenon is also exacerbated considering that the users' behavior changes during emergency conditions, in particular reducing the idle time between two consecutive service requests. After a certain delay a congestion is also produced on the base station B and this is due to the traffic-jam reaching its coverage area. On the contrary, the load factor of D rapidly decreases right after the accident, since the cars are blocked in the preceding cells. When the crash site is cleared and user (cars) are capable to move again all the load factors slowly return to the level they had before the car accident. The downlink load factor follows the same trend, and it is not reported here for the sake of brevity.

Figure 6 shows the impact of an outage affecting the base station C at time t=11000 sec. on the "access to medical expertise" application used by the ambulance (plot '4+5 Combined'). The probability of service interruption rapidly increases considering higher percentage of resources unavailability, reaching its maximum for values greater than 60%. Analyzing the single services forming the application (plots 'Emergency Streaming' and 'Emergency Video-conference'), initially the probability is lower for 'Emergency Streaming', but when limited resources are available 'Emergency Video-conference' has a lower probability of interruption. This happens because we have assumed that 'Emergency Streaming' requires more uplink resources than 'Emergency Video-conference' (see Table I), and after the outage the available uplink resources are lower than the downlink ones. For a better understanding, in Figure 7 we depict the uplink and downlink load factor of the base station C considering an outage equal to 70% (i.e., 70% of the cell resources becomes unavailable), at varying of time.

The load factor (both in uplink and downlink) increases after the car accident (at time t=10500 sec.), and then rapidly decreases after the outage event at time t=11000 sec. (due to the dropped services). After the outage the uplink load factor is near to its maximum allowed value and then the services requiring higher uplink resources will be probably not satisfied (due to the selected admission control algorithm, see Equation (1)).

Figure 8 shows the impact of the outage on the probability that the "access to medical expertise" application is interrupted, at varying of the outage severity (percentage of unavailable cell resources) and at varying of the base station affected by the outage. As expected, base station C is the most critical one, since it is the cell where the car accident occurs and the traffic is blocked, thus determining a high network congestion. Cell D does

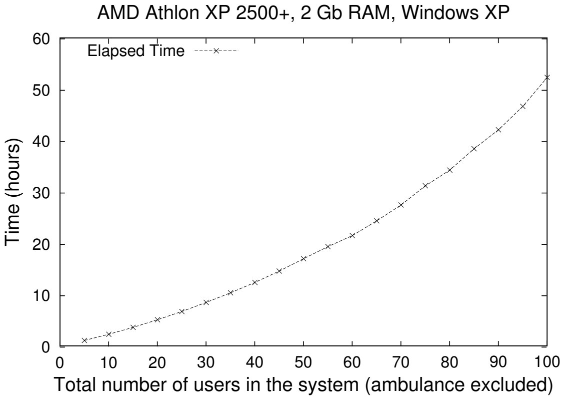

Figure 10. Average time to produce Figure 9, at varying of the total number of users in the system

not influence the ambulance connection at all, since the ambulance doesn't even enter the D zone.

Finally, Figure 9 shows the probability that the multimedia connections between the ambulance and the hospital are interrupted while the ambulance is going back to the hospital, varying the vehicle's average speed (no outages considered). Individual probabilities for the emergency services 'Emergency Streaming' and 'Emergency Video-conference' are shown, as well as the overall probability that corresponds to the "access to medical expertise" application. The probability of interruption is lower for service 'Emergency Streaming' because we assumed that it only uses uplink bandwidth and then it requires less network resources than service 'Emergency Video-conference'. Results also show that the probability increases when vehicle speed increases and this effect is in part caused by the RACH procedure delay.

1) On the Effectiveness of the Modeling Approach: In Section V-C we have shown the effectiveness of the modeling approach in facilitating the construction of the overall model, which can be obtained as composition of 40 models derived from a set of 10 basic template models only. Anyway, the modeling approach is really effective only if the computational cost required to solve the overall model is still manageable. In Figure 10 we present the average time (in hours) needed to produce all the plots of Figure 9, at varying of the total number of users in the system. The values in Figure 9 have been obtained performing 7 simulations, one for each considered 'ambulance average speed' value. As we can see, the computational time increases almost linearly for a low number of users, and the rate of grow slightly increases considering more than 60 users. Nevertheless, in the worst case (i.e., for number of users equal to 100) the whole set of simulations completed in less than 56 hours (therefore, less then 8 hours for each simulation).

VI. Models Refinement and Thorough Analysis

We now exploit the modularity of the proposed modeling approach to refine i) the model representing the mobility of the users, combining it with the output produced by an ad-hoc mobility simulator (Section VI-A), and ii) the model representing the UMTS network behavior, enabling a more refined load factor estimation (Section VI-B). The resulting numerical evaluations are finally presented in Section VI-C.

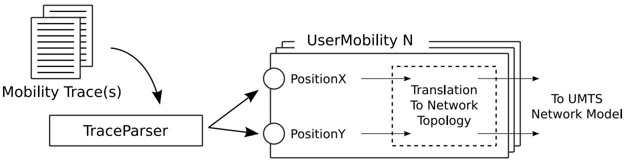

Figure 11. Integration between the SAN model and a mobility simulator, using traces

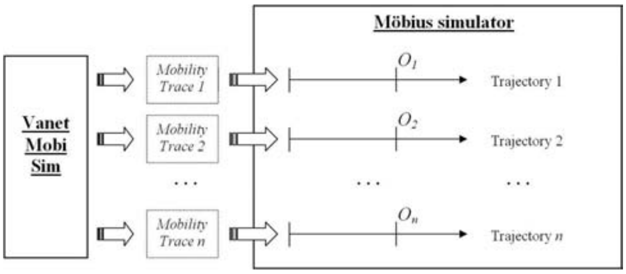

Figure 12. VanetMobiSim - Möbius interactions

A. Refinement of the Mobility Aspects

The focus is on the inter-operation of the UserMobility SAN model with a mobility simulator. A more refined representation of the user mobility aspects is achieved by allowing the SAN model to read and use the detailed mobility traces generated by a mobility simulator. Thanks to the modularity of the model and the clear separation between the two roles of the UserMobiliy model, the goal is achieved with few modifications to the original model. As first step, the "Mobility Pattern" part of the UserMobility models (see Figure 3) is replaced by some interface places, which hold the current user position and serve as input to the "Translation" part. These interface places will get their values from the mobility trace. An additional atomic model, the **TraceParser** model, is then introduced to read the trace file and fill the proper values in the interface places. This integration approach is depicted in Figure 11. All the other atomic models remain unchanged, and we just need to add the new TraceParser model to the top level join of the composed model of Figure 4.

In our current implementation the mobility traces have been generated using VanetMobiSim [16], a java-based mobility simulator which reads the scenario definition from an XML file and generates a trace file with the following format: `NodeID`, `Time`, `XPosition`, `YPosition`. In order to provide a stochastic mobility scenario, the mobility simulator interacts with the Möbius simulator as depicted in Figure 12. Before the execution of each Möbius simulation batch (trajectory i, with $i = 1, \ldots, n$, where n is the number of batches) we need to update the mobility trace file with a new one, stochastically generated by the mobility simulator. Given a (transient) measure of interest M, for each trajectory i an observation O_i is computed, which corresponds to a sample point of M; then, for example, its mean can be computed as $\bar{M} = \sum_{i=1}^{n} O_i/n$, and confidence intervals can be generated.

The mobility trace needed for a single simulation batch is generated and then read by the Möbius simulator itself through a fully automatic process, without additional libraries or dependencies. The whole process is handled by the additional TraceParser atomic model, which invokes the mobility simulator at the beginning of the simulation and then reads the generated trace at periodic intervals (matching the values of the `Time` field), until the Möbius simulation ends. The mobility simulator is executed using a blocking system call, therefore the Möbius simulation is suspended until the mobility trace has been generated.

Following this approach any mobility simulator or even experimental measurement tools could be used, provided that they are capable to generate mobility traces as output.

B. Refinement of the UMTS Network Model

At this point we can use this additional information, i.e., the exact position of the users in the network topology, to refine the UMTS network model through a more refined load factor estimation. As detailed in Section V-A, the basic UMTS model made the assumption that the load increment factors in uplink and downlink were fixed for each class of users and did not depend on the positioning of the users inside each zone. Since the refined mobility model provides the exact users position, we can now refine the load factor estimation of the UMTS network model. The concept of *path loss* describes the signal propagation in the modeled environment. Among other factors, it is a function of the distance between nodes and its calculation varies based on the selected propagation model. For simplicity we will consider the free-space path loss (see [15]), which is given by:

$$L_{\mathrm{dB}} = 32.44 + 20\log(f) + 20\log(d), \qquad (4)$$

where f is the operating frequency in Mhz and d is the distance in Km. Assuming that the total transmitted power of all base stations in the area is the same, the i factor for a given user can be computed as a function of the path losses between the user and the base stations [17]. The i factor for a mobile m served by cell j can thus be computed as: $i(m) = \sum_{k \neq j} L_{km}/L_{jm}$, where L_{km} is the path loss between base station k and mobile m.

C. Thorough Numerical Evaluations

The goal of this section is to evaluate and compare the QoS indicators obtained solving the basic and the refined models. The analyzed scenario is a simplified instance of the one described in Section II-A. We consider the same highway topology, the same network services and parameters, while we do not consider the car accident event (generic cars only). We also consider the outage of base station C (at time $t = 4001$) as well as its repair (at time $t = 5001$). As mentioned before, to compute the i factor in the trace-enhanced version the free-space path loss formula is used.

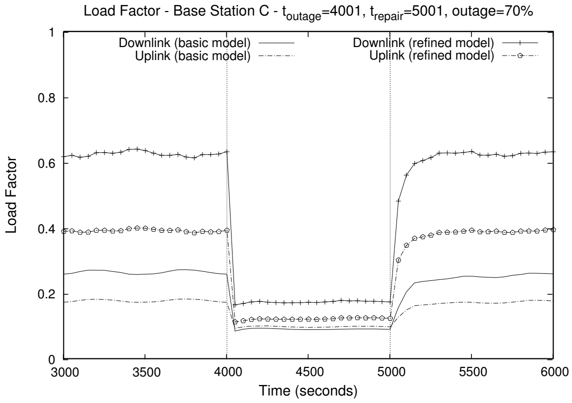

Figure 13. Load factor of base station C

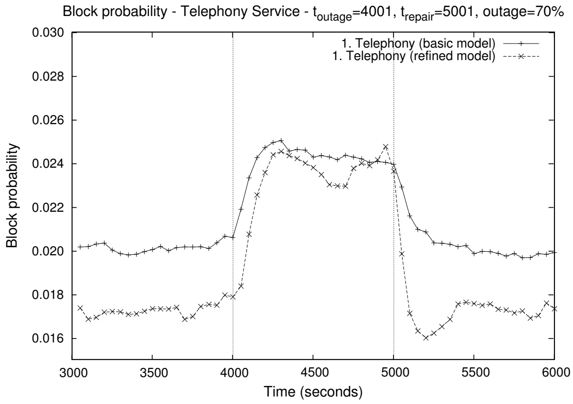

Figure 14. Probability that a service request is blocked

In Figure 13 we compare the load factor of base station C obtained using the basic and the refined models. The differences between the two set of plots (both in uplink and in downlink) are really significant. The refined estimation shows that the same number of users camped in the cell are actually producing a higher load factor, both in downlink and in uplink, and this is true both in nominal network conditions (i.e., before and after the outage period), and during the outage period (although the gap is less significant). This means that, with the adopted setting, the non-refined evaluation process underestimates the load factor of the cell.

In Figure 14 we compare the probability that a phone service request is blocked as time elapses. We note that the values obtained through the solution of the refined model are slightly lower when all the network resources are available, while they become almost the same during the outage period. The high variability of the refined plot is due to the high variability of the mobility traces generated by VanetMobiSim.

Finally, in Figure 15 we compare the probability that a phone service request is dropped as time elapses. In nominal network conditions the two plots have the same

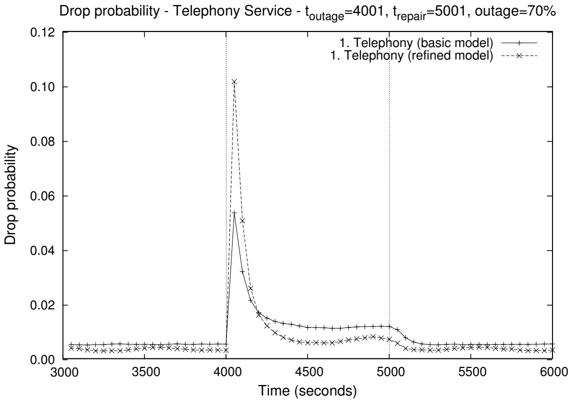

Figure 15. Probability that a service request is dropped

trend, and they both have a peak when the outage occurs due to the unavailability of a percentage of the network resources. However, the peak obtained from the refined analysis is initially much higher than the other, and it becomes slightly lower some time after. The higher peak is caused by users which are far from the center of the cell: these users experience a higher other-to-own interference ratio and thus are more likely to be dropped when the outage occurs.

Solving the refined model clearly needs a higher amount of time with respect to the basic model. However, we found that the greatest part of the simulation time is spent in running the mobility simulator: when considering 50 users, for example, the generation of a single trace takes about 1 minute, and then the related Möbius batch execution completes in few seconds only. Using a faster mobility simulator would thus greatly improve the overall performance.

VII. Conclusions

In this paper we have performed a QoS analysis of a dynamic, ubiquitous UMTS network scenario identified in the ongoing EC HIDENETS project. To do this, we have proposed a modular, hierarchical modeling approach based on composition, replication and parametrization, which facilitates the model construction process as well as the model reusability. The produced numerical results provide a useful insight in the relationships between the selected QoS measures, the users' behavior and the users' mobility, and they show the effectiveness of the adopted modeling approach. Then we have modified and extended the modeling process so to enable a more refined representation of the user mobility and UMTS network aspects. Following a holistic evaluation approach we have combined the SAN models with an ad-hoc mobility simulator, thus allowing a more detailed estimation of the load factor generated by the users. A through analysis has been performed showing the impact of the models refinement on some selected QoS indicators.

Acknowledgment

This work has been partially supported by the EC IST Project HIDENETS [1].

References

[1] European Project HIDENETS, "contract n. 26979," http://www.hidenets.aau.dk.

[2] A. Bondavalli, P. Lollini, and L. Montecchi, "Analysis of User Perceived QoS in Ubiquitous UMTS Environments Subject to Faults," in *Software Technologies for Embedded and Ubiquitous Systems*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2008, vol. 5287, pp. 186–197.

[3] A. Klemm, C. Lindemann, and M. Lohmann, "Traffic modeling and characterization for UMTS networks," *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, vol. 3, pp. 1741–1746 vol.3, 2001.

[4] DBench - Project Reports section, project short final report – http://www.laas.fr/DBench.

[5] HIDENETS D4.2.2 - Application of the evaluation framework to the complete scenario – http://www.hidenets.aau.dk/Public+Deliverables.

[6] M. Kovcs, P. Lollini, I. Majzik, and A. Bondavalli, "An integrated framework for the dependability evaluation of distributed mobile applications," in *Proc. of the RISE/EFTS Joint International Workshop on Software Engineering for Resilient Systems (SERENE 2008)*, November 2008, pp. 29–38.

[7] P. Lollini, A. Bondavalli, and F. Di Giandomenico, "QoS Analysis of a UMTS cell with different Service Classes," in *CSN-2005 The Fourth IASTED International Conference on Communication Systems and Networks*, September 12-14 2005.

[8] M. Andersin, Z. Rosberg, and J. Zander, "Soft and safe admission control in cellular networks," *IEEE/ACM Transaction on Networking*, vol. 5, no. 2, 1997.

[9] M. Nawrocki, H. Aghvami, and M. Dohler, *Understanding UMTS Radio Network Modelling, Planning and Automated Optimisation: Theory and Practice*. John Wiley & Sons, 2006.

[10] K. Sipilä, Z.-C. Honkasalo, J. Laiho-Steffens, and A. Wacker, "Estimation of Capacity and Required Transmission Power of WCDMA Downlink Based on a Downlink Pole Equation," in *Vehicular Technology Conference Proceedings, 2000*, vol. 2, Spring 2000, pp. 1002–1005.

[11] M. Kaaniche, P. Lollini, A. Bondavalli, and K. Kanoun, "Modeling the resilience of large and evolving systems," *International Journal of Performability Engineering*, vol. 4, no. 2, pp. 153–168, April 2008.

[12] W. H. Sanders and J. F. Meyer, "Stochastic activity networks: formal definitions and concepts," pp. 315–343, 2002.

[13] P. Lollini, L. Montecchi, and A. Bondavalli, "On the evaluation of hidenets use-cases having phased behavior," University of Florence, Dip. Sistemi Informatica, RCL group, Tech. Rep. rcl071201, December 2007, http://dcl.isti.cnr.it/Documentation/Papers/Techreports.html.

[14] D. Daly, D. D. Deavours, J. M. Doyle, P. G. Webster, and W. H. Sanders, "Möbius: An extensible tool for performance and dependability modeling," in *11th International Conference, TOOLS 2000*, ser. LNCS. Schaumnurg, IL: B. R. Haverkort, H. C. Bohnenkamp, and C. U. Smith (Eds.), 2000, pp. 332–336.

[15] M. Nawrocki, H. Aghvami, and M. Dohler, *Understanding UMTS Radio Network Modelling, Planning and Automated Optimisation: Theory and Practice*. John Wiley & Sons, 2006.

[16] VanetMobiSim – http://vanet.eurecom.fr/.

[17] H. Holma and A. Toskala, Eds., *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*. New York, NY, USA: John Wiley & Sons, Inc., 2001.

Andrea Bondavalli is a Professor in the Department of Systems and Informatics at the University of Florence. Previously, he was a researcher at the Italian National Research Council, working at the CNUCE Institute in Pisa, where he was responsible for the Dependable Computing Group. His research activity and interest are focused on the design and validation of critical systems and infrastructures, in particular, the design of fault tolerant architectures, mechanisms, and protocols; and their evaluation in terms of dependability attributes such as reliability, availability, and performability. He is an author of more than 110 refereed publications in international journals and conferences. He has been PI in many projects funded by the European Community, currently IST-FP7-216295 AMBER, acted in several occasions as an expert for the European community, and served as program chair of the most important conferences in the area, including DCC-DSN 05, IEEE SRDS 2000, EDCC-4, and LADC-3. Professor Bondavalli is a member of the IEEE, the IFIP W.G. 10.4 on "Dependable Computing and Fault-Tolerance," ENCRESS Club Italy, and the AICA Working Group on Dependability in Computer Systems.

Paolo Lollini received the laurea degree, and the PhD degree in Computer Science from the University of Florence, Italy, in 2001, and 2005, respectively. Since June 2006, he has been a research associate at the Computer Science Dept. of the same University. He is presently participating in several European funded projects including IST-FP7-216295 AMBER and ARTEMIS-2008-1-100022 CHESS. He is a member of the program committee for the 2009 IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2009 - PDS track), for the 2009 Latin-American Symposium on Dependable Computing (LADC 2009), and for the 2009 IEEE Symposium on Reliable Distributed Systems (SRDS 2009). His current research interests include the modeling and evaluation of performability attributes, with reference to a variety of application fields including railway, mobile telecommunications, and electric power systems.

Leonardo Montecchi received his Bachelor's degree in Computer Science at the University of Florence, Italy, in 2007 and he is now working towards the Master degree in the same University. He has been involved in the EU-funded projects IST-FP6-26979 HIDENETS and IST-FP6-031413 SAFEDMI, and he is presently participating in the IST-FP7-216295 AMBER project. His research interests focus on the modeling and evaluation of dependable systems.