
End-to-End Reliability and Optimization of Intra and Inter-Domain IMS-based Communication Networks

Chayapol Kamyod¹, Rasmus Hjorth Nielsen², Neeli Rashmi Prasad³,
Ramjee Prasad⁴ and Nattapol Aunsri⁵

¹*Future Technologies for Business Ecosystem Innovation (FT4BI)
in the Department of Business Development and Technology,
Aarhus University, Herning, Denmark*

²*Movimento, USA*

³*International Technological University, San Jose, USA*

⁴*Professor of Future Technologies for Business Ecosystem Innovation (FT4BI)
in the Department of Business Development and Technology,
Aarhus University, Herning, Denmark*

⁵*School of Information Technology, Mae Fah Luang University,
Chiang Rai, Thailand*

Corresponding Author: in_ck@es.aau.dk

Received 11 July 2016; Accepted 3 July 2017;
Publication 26 July 2017

Abstract

End-to-end system reliability and availability of different IMS-based network communication scenarios have been simulated and analyzed by using the combination of effective stochastic processes and models. Besides, both system reliability and availability characteristics of Intra and Inter IP multimedia subsystem (IMS) domain communication networks have been simulated and compared with different redundancy conditions. Moreover, parallel redundancies of the core IMS units were optimized at different availability or reliability standard conditions. The results demonstrate interesting system reliability characteristics of different communication scenarios and the results reveal possibility to create high reliability and availability system for the Next Generation Networks.

Journal of Cyber Security, Vol. 5_3, 233–256.

doi: 10.13052/jcsm2245-1439.533

© 2017 River Publishers. All rights reserved.

Keywords: Reliability, Redundancy, IMS, Optimization.

1 Introduction

Reliability of a system or service generally refers to the consistency of providing its required function within a given condition. Nowadays, it is one of the most important quality terms for all services and systems, especially considering the fact that everything is available through online services all the time. Besides, another important quality term called resilience has gained interest as the term refers to the system's ability to recover from a failure to some specific working conditions [1, 2]. Therefore, availability and stability or connectivity of the services and systems are very important qualities for the Next Generation Networks (NGNs) and services. In particular, the IP multimedia subsystem (IMS) has been proposed as a next-generation platform to provide real-time services and end-to-end quality of service (QoS) [3]; however, guaranteeing end-to-end QoS has many challenges such as interoperability and management among various networking technologies and vendors [4, 5]. In addition, the next generation services especially online services are required to be available anytime and anywhere. Therefore, availability, reliability and resilience are very important qualities for implementing and developing IMS toward the NGNs.

System availability is the probability that the system is in a ready state to perform its functions; the system reliability is the probability that the system can function without failure; besides, system resilience represents how well the system can overcome the failure. Thus, availability is a function of the reliability and resilience represents the reliability degree of the system. Therefore, these quantities are strongly related to each other and can indicate the performance of the system.

As mentioned, availability is a function of reliability. The measurement of system availability alone represents no significant details about the system such as a number of component failures or a number of replaced components. Therefore, both reliability and availability analysis of the system are needed for representing comprehensive reliability quality of the system. In order to evaluate the system reliability, overall system characteristics need to be considered. Moreover, a system is an integration of components or subsystem to perform its functions. Therefore, a connection topology of the system components or subsystems and the system configurations will directly affect overall system reliability. Accordingly, the reliability-wise configuration of the components is needed to be considered. Besides, measuring overall qualities

of the system needs to take into account an end-to-end quality evaluation framework. The term “end-to-end” refers to the connection between user’s end devices (UEs) across the IMS-based network. As mentioned above, it is difficult to guarantee end-to-end quality over an IMS system due to various communication scenarios between end users and the complexity of the networks and services. However, model-based reliability evaluations have provided useful information for designing and managing the IMS system [6, 7]. Nevertheless, end-to-end reliability and availability analysis of the IMS systems have not been fully studied. As mentioned above, availability and stability of the system are the key qualities for NGNs. Therefore, both availability and reliability of the system need to be analyzed together to achieve useful reliability or resilience of the system.

The main contributions of this paper include the analysis of end-to-end reliability and availability for both intra and inter-domain IMS-based reference network by using the proposed three and five-state Continuous Time Markov Chain (CTMC) model and Reliability Block Diagram (RBD). The proposed model is proven to outperform the state-of-the-art models [8]. Both end-to-end availability and reliability characteristics are compared and analyzed with and without redundancy cases. Moreover, the fault tolerance system model and redundancy optimization at different availability values are presented. The paper is organized as follows. Section 2 presents related works. Section 3 defines the IMS-based reference network and the proposed end-to-end availability and reliability analysis models respectively. The simulation and comparison results between intra and inter-domain communication at different conditions and the redundancy optimization results are given in Section 4. The conclusions are given in the final section.

2 Overview

Over the past decade, various reliability evaluation techniques have been proposed for the IMS system. For example, state-space methods such as Queuing Network Model (QNM) [9–11], Queuing Petri Nets (QPN) [12], and the Markov model [8, 13–17] were applied for performance and reliability evaluation of the IMS network. The QNM is normally applied to analyze the processing delay. The QPN was shown to combine some advantages of both Stochastic Petri Nets (SPN) and queuing networks models [12]. The Markov model can take into account some detailed working states and system parameters such as failover and recovery rates. Besides, redundancy effect can also be included into the model for an evaluation [8, 15–17].

In addition, reliability evaluation at different failover success rates was performed by using three-state Markov model [13]. The coverage factor and the failover period were proven to influence the system downtime with redundancy [14]. Further, system reliability and availability with internal and external redundancy were evaluated through the Markov model. The system availability was shown to improve for all redundancy mechanisms [15]. Moreover, an end-to-end reliability and resilience properties of the IMS system were evaluated by using the Markov Rewards Model and Reliability Block Diagrams (RBD) [16]. The 1:1 redundancy effect at different end-to-end communication scenarios were simulated and shown that total reliability will be highly affected by the individual reliability of the system components. Further, the end-to-end availability analysis of the intra and inter-domain IMS-based communication network had been proposed by using the five-state Markov model and RBD [8]. Besides, the proposed model was compared with state-of-the-art three-state Markov models and proved to provide better system behaviours of the simplex and redundant systems. The simulation results of end-to-end availability were shown to significantly improve when having a redundancy of the S-CSCF unit, especially for long distance communication. Besides, performance optimization of the IMS core network with parallel redundancy was evaluated by using the CTMC and Universal Generating Function (UGF) [17].

Furthermore, an improvement of IMS system reliability with a single redundancy of the S-CSCF unit was shown through the network simulation software called OpenIMS [18]. Later, end-to-end reliability and resilience characteristics of the IMS-based communication within similar and across registered domain were simulated and analyzed through the well-known OPNET software [19, 20]. In particular, end-to-end resilience behavior of communication across multiple IMS domains were demonstrated and compared. The results showed various resilience behavior of long distance communications and the effect when adding redundancy.

From above, a combination of models and simulation can be applied for reliability and availability evaluation of the IMS-based network [8]. The proposed availability model is chosen due to its advantages for both single and redundancy model of the components. However, an overall evaluation of the IMS system has not been fully studied. This paper focuses on an optimization of end-to-end reliability and availability of the IMS-based network at different communication scenarios by using a combination of three and five-state CTMC models and RBD models. Moreover, optimization of different parallel redundancies of the IMS core unit is simulated and compared with different communication scenarios and reliability standard requirements.

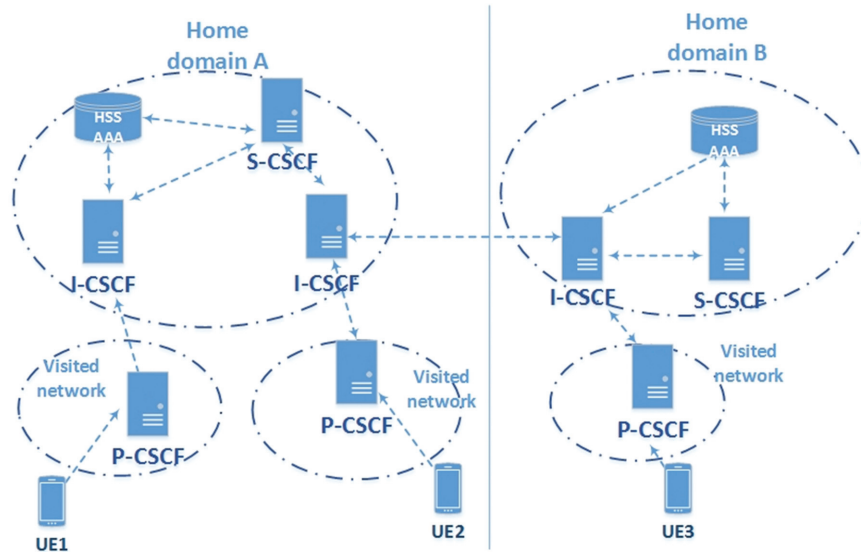


Figure 1 The IMS reference network.

For covering major communication scenarios of the IMS-based network, the communication network reference architecture per Figure 1 is used to evaluate end-to-end availability and reliability. The reference network can represent feasible real-world communication scenarios for both intra and inter-communication between UEs.

3 End-to-End Availability and Reliability Analysis Model

For covering all failures and recovery characteristics in the model, the coverage factor is considered for representation of the proportion of failures which can be recovered. Besides, for simplicity of an analysis, each of the components is assumed to have independent properties and the failure and recovery rates are assumed to be exponentially distributed. In addition, the communication between UEs, based on the network reference architecture of Figure 1, is used to evaluate the end-to-end availability and reliability of different communication scenarios.

3.1 End-to-End Availability Analysis Model

The five-state CTMC model [8] is applied for an analysis with two main failure types- Soft Failure (SF) and Hard Failure (HF) as shown in Figure 2 for both

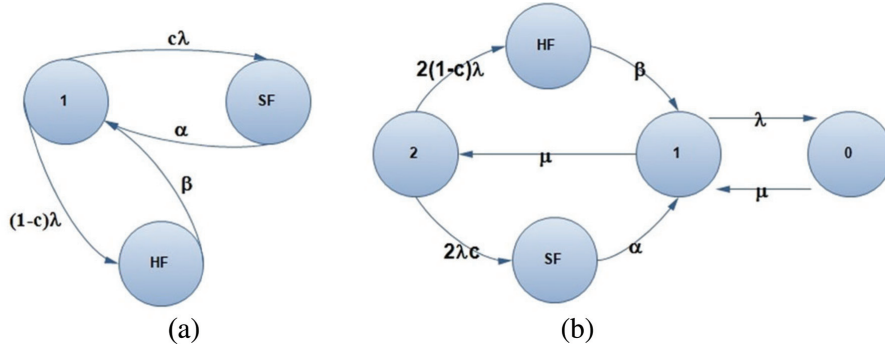


Figure 2 The three-state and five-state CTMC model: (a) simplex unit, (b) redundant unit.

single and redundant unit. The SF is defined to cover degradation and ordinary failure types (for both hardware and software) that can be automatically or manually recovered. The HF can be defined to cover instant or severe failure types that need longer recovery time than SF for instance manual repaired processes with many hours of recovery time. The coverage factor, c , is the conditional probability that a failure will be automatically recovered. Improper failure treatment will lead to the HF state instead of the SF state. Parameters α and β are the recovery rates for SF and HF respectively. These rates, where in time $1/\beta$ is longer than $1/\alpha$, refer to the recovery period for both failure types back to a state that can be further recovered to the working condition. The recovery and failure rates are represented by μ and λ respectively. These parameters are proven to efficiently provide more details on availability behaviours than previous models for both simplex and redundant models as presented in [8]. Therefore, the set of state-space of the model is given by $S = \{0, 1, 2, SF, HF\}$. State “2” and “1” represent the working state conditions of two and one component respectively. Accordingly, state “0” represents the complete failure state. The system state probabilities of the proposed model can be written as

$$\begin{cases} d\pi_2(t)/dt = -2\lambda\pi_2(t) + \mu\pi_1(t) \\ d\pi_{HF}(t)/dt = 2(1-c)\lambda\pi_2(t) - \beta\pi_{HF}(t) \\ d\pi_{SF}(t)/dt = 2\lambda c\pi_2(t) - \alpha\pi_{SF}(t) \\ d\pi_1(t)/dt = \beta\pi_{HF}(t) + \alpha\pi_{SF}(t) - (\lambda + \mu)\pi_1(t) + \mu\pi_0(t) \\ d\pi_0(t)/dt = \lambda\pi_1(t) - \mu\pi_0(t). \end{cases} \quad (1)$$

Assuming an initial working condition where $\pi_2(0) = 1, \pi_{HF}(0) = \pi_{SF}(0) = \pi_1(0) = \pi_0(0) = 0$ and $\pi_2(t) + \pi_{HF}(t) + \pi_{SF}(t) + \pi_1(t) + \pi_0(t) = 1$. Hence, the transient and steady state availability of all available states, where the states occupy the probability of 1, can be evaluated. Let $A_{5R}(t)$ be availability of the proposed model; then the steady state availability is given by

$$\lim_{t \rightarrow \infty} A_{5R}(t) = \frac{\alpha\beta\mu^2 + 2\alpha\beta\lambda\mu}{(2\alpha\lambda\mu^2 + 2\alpha\beta\lambda^2 + \alpha\beta\mu^2 + 2\alpha\beta\lambda\mu + 2\beta c\lambda\mu^2 - 2\alpha c\lambda\mu^2)}. \quad (2)$$

The model for the simplex system is proposed and shown in Figure 2a where two detailed failures are used instead of the normal failure state. “1” represents the working state and “0_{*i*}” represents the down state with the state description *i*. Figure 2a has three state-spaces, $S = \{1, HF, SF\}$, where “HF” and “SF” represent hard and soft failures respectively. In other words, both failures represent down state or state “0” for a normal two-state CTMC. The system is assumed to be initially up: $\pi_1(t) = 1$. Thus, the system state probabilities of the simplex model can be written as

$$\begin{cases} d\pi_1(t)/dt = -\lambda\pi_1(t) + \alpha\pi_{SF}(t) + \beta\pi_{HF}(t) \\ d\pi_{SF}(t)/dt = c\lambda\pi_1(t) - \alpha\pi_{SF}(t) \\ d\pi_{HF}(t)/dt = (1-c)\lambda\pi_1(t) - \beta\pi_{HF}(t) \end{cases} \quad (3)$$

Thus, the steady state availability for the simplex model per Figure 2a is given by A_s , where the subscripts *s* denote the proposed simplex.

$$\lim_{t \rightarrow \infty} A_s(t) = \frac{\alpha \times \beta}{\langle \alpha \times (\beta + \lambda) \rangle - \langle c \times \lambda \times (\alpha - \beta) \rangle} \quad (4)$$

The RBD is employed to estimate the total availability of the communication path between UEs. The IMS core components can be represented as serial or parallel connections depending on the analyzed network architecture, which can combine parallel redundancy into the system. The communication scenarios between two UEs at similar (UE1 and UE2) and different communication domains (UE1 and UE3) are employed to determine the end-to-end availability. The RBDs of the call setup path for both scenarios are shown by Figure 3a and 3b respectively. In the figures, the parallel connection illustrates the redundancy of the core CSCF unit. For the connection scenarios without redundancy, the serial connection of RBDs will be considered.

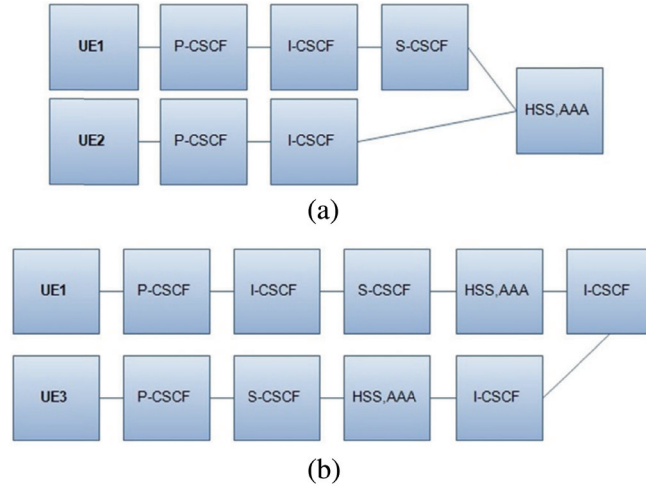


Figure 3 RBD of a communication network scenario: (a) similar home domain (UE1 & UE2), (b) different home domain (UE1 & UE3).

3.2 End-to-End Reliability Analysis Model

In the same way, total reliability of different end-to-end communication scenarios can be evaluated using the RBD. For the calling scenarios as given by Figure 3 and Figure 4, assuming all communication links are reliable at the beginning of an operating period. The reliability function at a time, $[0, t]$ between UEs can be represented by each reliability of the block diagram.

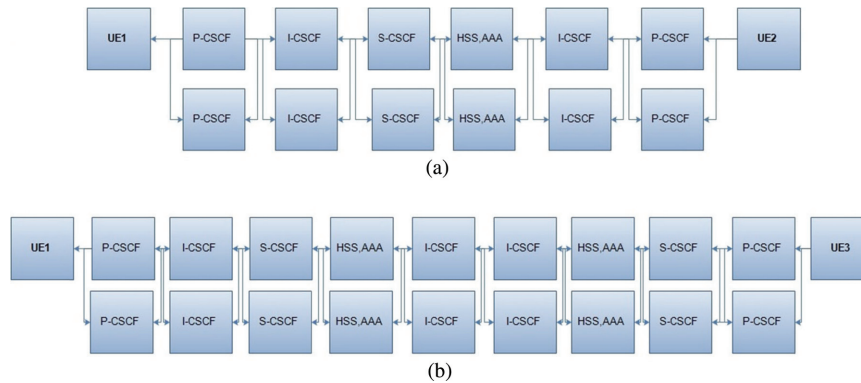


Figure 4 RBD of a communication network scenario with parallel redundancy: (a) similar home domain (UE1 & UE2), (b) different home domain (UE1 & UE3).

Let $R_U(t)$, $R_P(t)$, $R_I(t)$, $R_S(t)$, $R_A(t)$ represent the reliability of the each component, where UE , P , I , S , and A denote the first letter of the user terminal and the SIP servers. Therefore, for series and parallel systems the total reliability can be given by R_{T_S} and R_{T_P} .

$$R_{T_S}(t) = \prod_{i=1}^n \pi_i(t) = \prod_{i=1}^n R_i(t) \quad (5)$$

$$R_{T_P}(t) = 1 - \prod_{i=1}^n [1 - \pi_i(t)] = 1 - \prod_{i=1}^n [1 - R_i(t)] \quad (6)$$

Where R_{T_S} and R_{T_P} represent the total reliability of the serial and parallel connection, $R_i(t)$ is the reliability of unit i , and $\pi_i(t)$ is the probability of the unit being operational at time t . Therefore, the total reliability for each communication scenario per Figure 3 and Figure 4 can be given by

$$R_{T_{3a}}(t) = \left\{ [R_{UE}]^2 \times [R_P]^2 \times [R_I]^2 \times R_S \times R_H \right\} \quad (7)$$

$$R_{T_{3b}}(t) = \left\{ [R_{UE}]^2 \times [R_P]^2 \times [R_S]^2 \times [R_I]^3 \times [R_H]^2 \right\} \quad (8)$$

$$R_{T_{4a}}(t) = [R_{UE}(t)]^2 \times [1 - \langle 1 - R_P(t) \rangle^2]^2 \times [1 - \langle 1 - R_I(t) \rangle^2]^2 \\ \times [1 - \langle 1 - R_S(t) \rangle^2] \times [1 - \langle 1 - R_H(t) \rangle^2] \quad (9)$$

$$R_{T_{4b}}(t) = [R_{UE}(t)]^2 \times [1 - \langle 1 - R_P(t) \rangle^2]^2 \times [1 - \langle 1 - R_I(t) \rangle^2]^3 \\ \times [1 - \langle 1 - R_S(t) \rangle^2]^2 \times [1 - \langle 1 - R_H(t) \rangle^2]^2 \quad (10)$$

where $R_{T_{3a}}(t)$, $R_{T_{3b}}(t)$, $R_{T_{4a}}(t)$, and $R_{T_{4b}}(t)$ are the total reliability of the communication scenario per Figures 3a, 3b, 4a, and 4b respectively. Then, the exponential total reliability for each scenario can be given by

$$R_{T_{3a}}(t) = \left\{ \left[e^{-2\lambda_{UE}(t)} \right] \times \left[e^{-2\lambda_P(t)} \right] \times \left[e^{-2\lambda_I(t)} \right] \right. \\ \left. \times \left[e^{-2\lambda_S(t)} \right] \times \left[e^{-2\lambda_H(t)} \right] \right\} \quad (11)$$

$$R_{T_{3b}}(t) = \left\{ \left[e^{-2\lambda_{UE}(t)} \right] \times \left[e^{-2\lambda_P(t)} \right] \times \left[e^{-2\lambda_S(t)} \right] \right. \\ \left. \times \left[e^{-3\lambda_I(t)} \right] \times \left[e^{-2\lambda_H(t)} \right] \right\} \quad (12)$$

$$R_{T_{4a}}(t) = \left\{ \begin{array}{l} [e^{-2\lambda_{UE}(t)}] \times [1 - \langle 1 - e^{-\lambda_P(t)} \rangle^2]^2 \times \\ [1 - \langle 1 - e^{-\lambda_I(t)} \rangle^2]^2 \times [1 - \langle 1 - e^{-\lambda_S(t)} \rangle^2] \times \\ [1 - \langle 1 - e^{-\lambda_H(t)} \rangle^2]^2 \end{array} \right\} \quad (13)$$

$$R_{T_{4b}}(t) = \left\{ \begin{array}{l} [e^{-2\lambda_{UE}(t)}] \times [1 - \langle 1 - e^{-\lambda_P(t)} \rangle^2]^2 \times \\ [1 - \langle 1 - e^{-\lambda_I(t)} \rangle^2]^3 \times [1 - \langle 1 - e^{-\lambda_S(t)} \rangle^2]^2 \times \\ [1 - \langle 1 - e^{-\lambda_H(t)} \rangle^2]^2 \end{array} \right\} \quad (14)$$

where $\lambda_{UE}, \lambda_P, \lambda_I, \lambda_S,$ and λ_A are the failure rate of the UE, P-CSCF, I-CSCF, S-CSCF and AAA respectively.

3.3 The Fault-Tolerant System Models: The M-out-of-N Reliability Model and Optimization

One way to increase fault-tolerance of the system is to increase the number of parallel server units for each subunit. This will definitely increase availability and reliability of the subunit and also the system. Figure 5 represents the M-out-of-N model for each SIP server unit. In this case, each node has N parallel redundancy. The N-component system will fail if at least M of the N unit fails. The model can also represent series and parallel connection when M is equal to 1 and N respectively. Moreover, the model is a very popular type of redundancy in fault-tolerant systems. With independent properties assumption, the reliability for each subunit can be given and evaluated by using binomial distribution as

$$R_S(M, N, R) = \sum_{r=M}^n \binom{N}{r} R^r (1 - R)^{N-r} \quad (15)$$

where R_S the transient total reliability of subunit, N is the total number of each parallel subunit, M is the minimum number of the unit required for subunit to function, and R is the transient reliability of each unit. Therefore, the total reliability of each scenario can be given by substituting R_S into Equation (5) and (6) for similar and different home domain communication scenarios respectively. Then, the reliability for the fault tolerant model of the communication scenario per Figure 5 can be given by

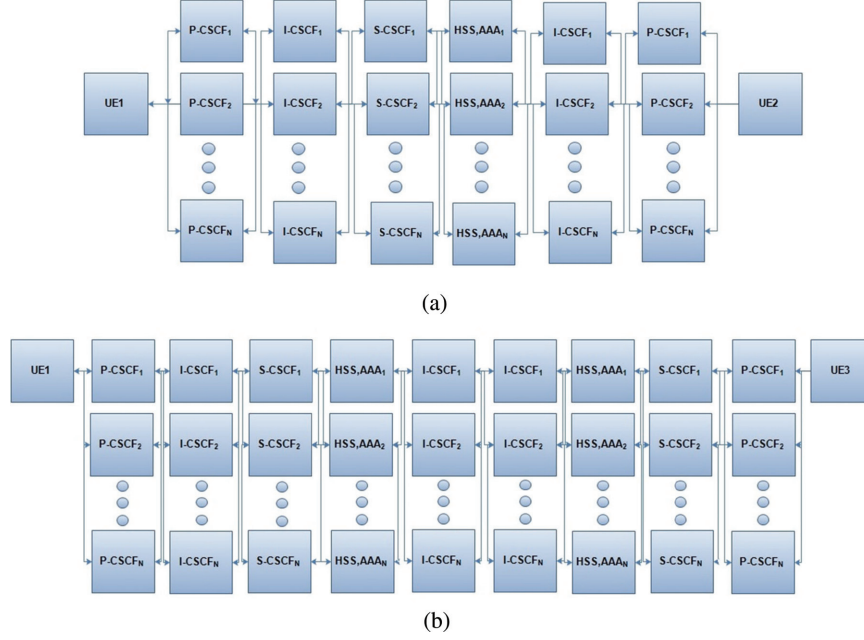


Figure 5 RBD of a communication network scenario with N-parallel redundancy: (a) similar home domain (UE1 & UE2) (b) different home domain (UE1 & UE3).

$$R_{T_{5a}}(t) = \left\{ \left[\sum_{r=M}^n \binom{N}{r} R^r (1-R)^{N-r} \right]_{UE}^2 \times \left[\sum_{r=M}^n \binom{N}{r} R^r (1-R)^{N-r} \right]_P^2 \times \left[\sum_{r=M}^n \binom{N}{r} R^r (1-R)^{N-r} \right]_I^2 \times \left[\sum_{r=M}^n \binom{N}{r} R^r (1-R)^{N-r} \right]_S \times \left[\sum_{r=M}^n \binom{N}{r} R^r (1-R)^{N-r} \right]_H \right\} \quad (16)$$

$$\begin{aligned}
R_{T_{5b}}(t) = & \left\{ \left[\sum_{r=M}^n \binom{N}{r} R^r (1-R)^{N-r} \right]_{UE}^2 \right. \\
& \times \left[\sum_{r=M}^n \binom{N}{r} R^r (1-R)^{N-r} \right]_P^2 \\
& \times \left[\sum_{r=M}^n \binom{N}{r} R^r (1-R)^{N-r} \right]_S^2 \\
& \times \left[\sum_{r=M}^n \binom{N}{r} R^r (1-R)^{N-r} \right]_I^3 \\
& \left. \times \left[\sum_{r=M}^n \binom{N}{r} R^r (1-R)^{N-r} \right]_H^2 \right\} \quad (17)
\end{aligned}$$

From Equations (13)–(17), in order to achieve maximum system availability and reliability, one simple solution is by adding a number of redundancies into the system. However, as a result, this would also increase system cost, complexity, maintenance, and management. Therefore, the parameters of the system according to the Equations (13)–(17) can be optimized as reported in the simulation results to gain benefit from the reliability, availability, and cost.

4 Simulation Results

From Section 4, due to the fact that a failure rate value is less than one and is significantly lower than the recovery rate ($\lambda \ll \mu$), therefore, from the communication topology and Equations (4)–(16), we can simply analyse that the communication scenario (a) has higher reliability than the communication scenario (b) for both with and without redundancy cases. This clearly exposed that less complex communication scenario would have less failure probability or high reliability. However, simulation results are still needed for an insight system reliability of the communication network.

We applied the effective five-state redundancy model and the three-state model per Figure 2a and 2b respectively to simulate end-to-end availability and reliability of different IMS-based communication scenarios. The simulation is different from [8], where the five state CTMC model was applied only for

the redundancy model. For practical and ease of the analysis, the failure and recovery rates have approximated based on statistical values of failure and recovery of realistic optical and IP networks [21]; the rates are assumed for all IMS components. In this case, failure is assumed to occur eight times per year, $\lambda = 2.4897 \times 10^{-7}$ sec⁻¹, $\mu = 1.3889 \times 10^{-4}$ sec⁻¹ (corresponding to the average time to repair of 120 min). The assumed average recovering time of SF and HF are 15 min and 300 min respectively, $\alpha = 1.1000 \times 10^{-3}$ sec⁻¹ and $\beta = 5.5556 \times 10^{-5}$.

4.1 Intra-Domain Communication: Simplex and Redundancy Models

The simulation results of the transient end-to-end availability and reliability of intra-domain communications versus time are presented by Figure 6 and Figure 7, where the scenario-(a) and scenario-(a)-redundant legend represent the intra-domain communication scenarios between UE1 and UE2 without and with single redundancy respectively.

With the assumption of ready initial condition at the beginning of an operation, the overall availability is equal to 1 or 100%. Then with the given failure and recovery factors, end-to-end availability is decreased versus time and moving toward steady state liked region at some operating period.

The end-to-end availability is increased when adding just a single redundancy into the system. The percentage change or increasing percentage when having redundancy at the steady state is about 0.26%. This percentage change level is quite high when considering in term of an overall availability or percentage of uptime per year. Moreover, overall end-to-end availability characteristics tend to move toward steady state liked period faster when having redundancy.

Accordingly, decreasing of end-to-end reliability is directly proportional to an operating period. Unlike availability characteristics, reliability has no steady state liked region due to it represent a no failure probability. Besides, overall reliability results are not improved with redundancy.

The falling gradient of the end-to-end reliability of redundancy case is larger than without redundancy case. This implies that involving more equipment into the system will decrease overall reliability of the system, even though an overall availability is improved from the beginning of an operation when adding redundancy.

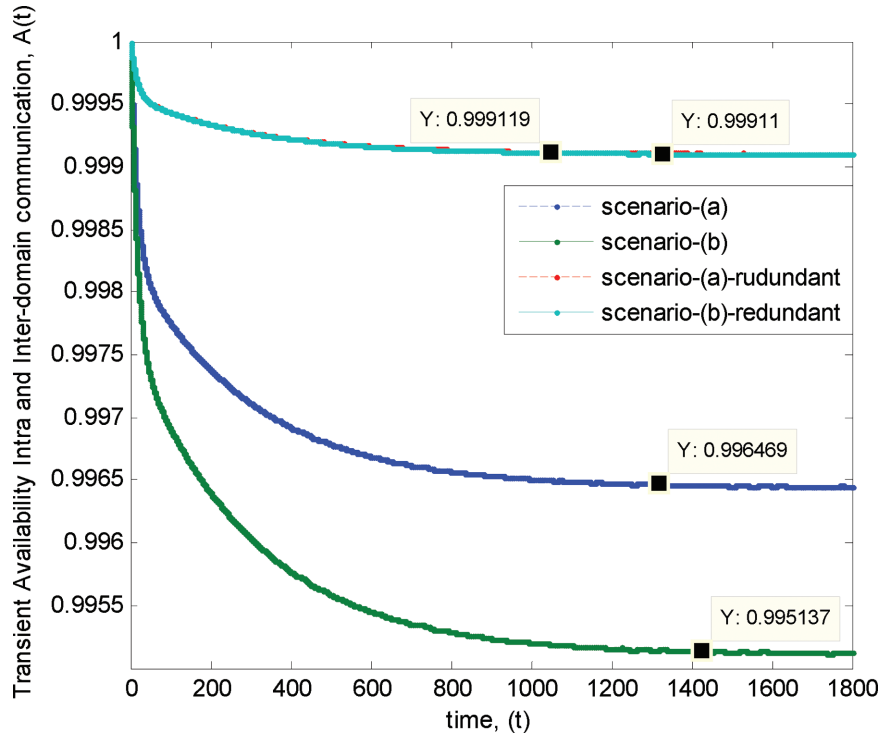


Figure 6 End-to-end availability results of intra and inter-domain communications.

4.2 Inter-Domain Communication: Simplex and Redundancy Models

Figure 6 and Figure 7 represent end-to-end availability and reliability of the inter-communication scenarios versus time, where legend scenario-(b) and scenario-(b)-redundant represent intra-domain communication scenarios between UE1 and UE3 of without and with redundancy respectively.

Similarly to the intra-domain communication scenarios, overall availability characteristics are decreased at the long operating period and moving toward the steady state liked region at some period, where the intra-domain communication scenarios with redundancy will reach steady state liked region faster than without redundancy case. The percentage change of intra-domain communication with redundancy at the steady state liked region is $\approx 0.40\%$. The percentage change is quite high and is about two times higher than the intra-domain communication case. Thus, redundancy has high impact to an

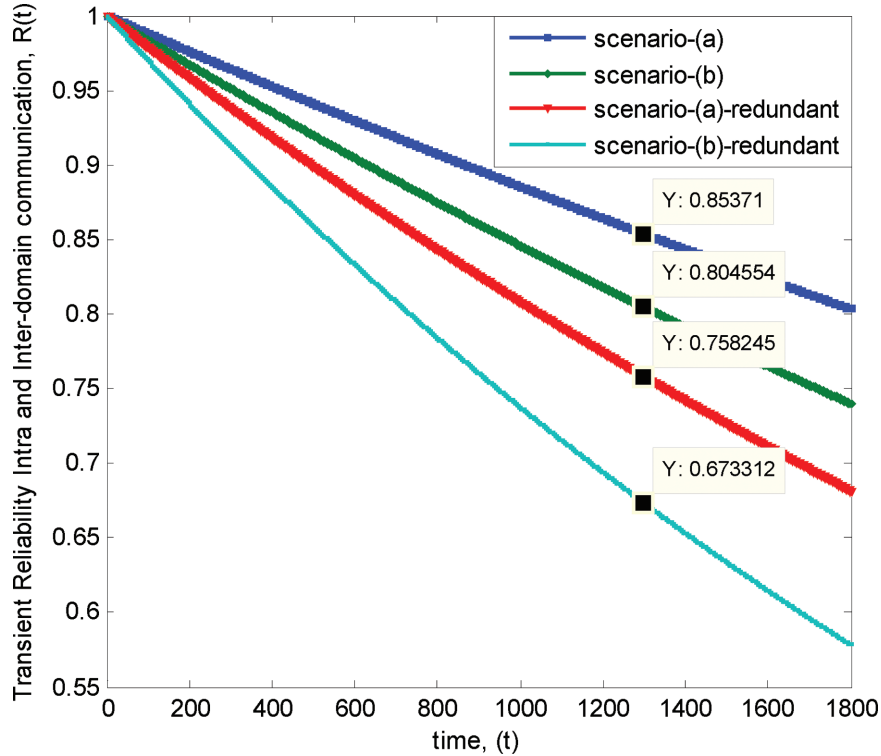


Figure 7 End-to-end reliability results of intra and inter-domain communications.

overall availability characteristic of the inter-domain communication or long communication setup path.

Consequently, the end-to-end reliability of intra-domain communication is decreased when increasing operating period. Besides, a decreasing of end-to-end reliability can be clearly observed with redundancy. Moreover, the falling slopes of the reliability results are steeper than end-to-end reliability results of the intra-domain communication cases. In other words, an increasing amount of communication equipment highly affects overall reliability of the system regardless of the number of redundancies. This effect can be clearly observed at the beginning of the operating period. Moreover, end-to-end quality of both reliability and availability are lower than the intra-domain communication scenarios.

4.3 Comparison of Intra and Inter-Domain Communications

The transient characteristics of the end-to-end availability and reliability of all communication scenarios with and without redundancy cases can be compared based on Figure 6 and Figure 7.

4.3.1 End-to-end availability

The results show that intra-domain communication scenario or communication within the similar home domain has higher availability than communication across the different home domain. The significant improvement of the availability can be observed when adding just one redundancy for both inter and intra-domain communication scenarios. In particular, the redundancy highly affects inter-domain communication scenarios, where the end-to-end availability values almost reach the similar level with the intra-domain communication cases.

The percentage difference of the availability gap between without and with redundancy case of intra-domain communication is $\approx 0.26\%$. The availability gap between without and with the redundancy of inter-communication domain is $\approx 0.40\%$. While, the percentage difference of the availability gap between intra and inter-communication of no redundancy case is $\approx 0.14\%$ and the percentage difference of the gap in case of redundancy is $\approx 9 \times 10^{-4}\%$. Then, the percentage difference of the availability gap between intra and inter-communication domain is reduced by 10^4 times. Therefore, for inter-communication domain or long communication setup paths, redundancy is needed to improve the end-to-end availability of the system.

4.3.2 End-to-end reliability

The transient characteristic of end-to-end reliability tends to decrease when the communication scenarios involve many signalling setup paths. Moreover, end-to-end reliability is reduced when adding redundancy or more equipment into the system. In particular, the overall reliability of inter-domain communication with redundancy is less than the reliability of the intra-domain communication without redundancy case.

From Figure 7, the falling percentage or percentage change of the reliability gap between without and with redundancy case of intra-domain communication is $\approx 11.18\%$, and the percentage change of the gap between without and with the redundancy of inter-domain communication is $\approx 16.31\%$. While, the percentage difference of the reliability gap between intra and inter-domain communication of no redundancy case is $\approx 5.93\%$ and the percentage

difference of the reliability gap in case of redundancy is $\approx 11.87\%$. Therefore, having redundancy can increase the complexity of the system and can diminish end-to-end reliability quality of the system.

4.4 The Fault-tolerant System Models and Optimization

According to the high availability system where the system aims to ensure some operational performance level and the performance is represented by an availability value such as 3-nines system availability (99.9%) means that the system has only 8.76 hours downtime per year. In this paper, for supporting NGNs, we are focusing on the desired 5-nines (99.999%) and 6-nines (99.9999%) system availability which corresponding to the system downtime of 5 minutes per year and 31 seconds per year respectively for our optimization sample.

From the simulation results, increasing redundancy into the system may end up with increasing the complexity and decreasing overall reliability of the system. However, adding redundancy has significant effect to an overall availability of the system especially for intra-domain or long distance communication scenarios. Therefore, high availability does not refer to high reliability. On the other hand, high reliability would refer to high availability. Then, from the high availability concept and from (15) and (16), with initial redundancy or $M = 2$, and with similar failure rate assumption for comparison purpose, $\lambda = 2.4897 \times 10^{-7} \text{ sec}^{-1}$, the minimum redundancy number to achieve different end-to-end reliability as relating to high availability concept can be calculated and given in Table 1.

From the optimization results, two nines and three nines system reliability require at least four parallel redundancy of each core IMS unit for both intra and inter-domain communication cases. This implies that in order to achieve up to three nines end-to-end reliability or the maximum system downtime per year

Table 1 Minimum redundancy unit at different end-to-end reliability requirement of intra and inter-domain communication scenarios

End-to-End Reliability	Inter-Domain (Min No. of Redundancies)	Inter-Domain (Min No. of Redundancies)	% Increasing of N
0.99%	4	4	0%
0.999%	4	4	0%
0.9999%	5	5	25%
0.99999%	6	6	20%
0.999999%	6	6	0%

equal to 8.76 hours for both intra and inter-domain communication scenarios, we need at least four parallel redundant units for each core IMS unit and for each communication scenarios. In addition, four parallel redundancy means that we need a minimum total number of 18 core IMS units for intra-domain communication and 27 core IMS units for inter-domain communication case. So the percentage difference of the total unit between the intra and inter-domain system is 40%. Therefore, the required total unit of the inter-domain communication is less than twice of the intra-domain communication case.

Moreover, from Table 1, we could achieve up to five and six nines system reliability or the system downtime per year equal to 31.5 seconds by having at least six parallel redundancies for each core IMS units. The different percentage of the increasing number of redundancy is 25% from the three nines condition and 20% from the four nine condition. In addition, with six redundant units, the system could support up to six nines system reliability which is equivalent to the percentage change of 0.1%. This changing amount may not seem to represent a significant change for a normal statistical value. However, for this stochastic process, the value represents a very significant change of the system reliability. In this case, the average system downtime per year was shown to improve from 8.76 hours to be 31.5 seconds which is almost a thousand times better. For five parallel redundancies, we need a total number of 30 core IMS units for intra-domain communication and we need a total number of 45 core IMS units for inter-domain communication. In particular, the percentage difference of the total number of the IMS core unit between intra and inter-domain communication is 40% and is similar to the three nines condition. Therefore, with the similar increasing ratio of the total unit, six nines reliability condition could be achieved.

These optimization results imply that a certain or limited amount of parallel redundant unit could be evaluated and optimized for inter-domain communication scenarios or a complex communication system to improve both system availability and reliability with a desired reliability and availability standard.

5 Conclusions

The paper focuses on end-to-end reliability and availability evaluation of the IMS-based communication scenarios by using a combination of stochastic processes and models. The simulation results show that the IMS-based system availability can be significantly improved by adding redundancy unit onto the system especially for inter-domain or long distance communication scenarios. However, adding redundancy unit could end up decreasing system reliability

and increasing system complexity. Moreover, an optimization of the number of parallel redundant unit that corresponding to the different standard of the IMS-based system reliability requirement is presented. The results demonstrate an interesting fact that high system availability and reliability can be achieved with a suitable amount of the IMS core redundant unit. Therefore, for the NGNs where a high-end computer can be created at suitable costs, network or service providers can build practical high reliability or availability system to support high-quality applications.

References

- [1] Queiroz, C., Garg, S.K., and Tari, Z. (2013). A probabilistic model for quantifying the resilience of networked systems. *IBM J. Res. Dev.* 57, 3–1.
- [2] Autenrieth, A., and Kirstädter, A. (2002). Engineering end-to-end IP resilience using resilience-differentiated QoS. *Commun. Mag. IEEE* 40, 50–57.
- [3] 3GPP Technical Specification (2014). TS 23.207 version 12.0.0: End-to-end Quality of Service (QoS) concept and architecture.
- [4] Zhang, J., and Ansari, N. (2011). On assuring end-to-end QoE in next generation networks: challenges and a possible solution. *Commun. Mag. IEEE*, 2011, 49, 185–191.
- [5] Shakir, M. (2010). Challenging issues in NGN implementation and regulation. *Proceedings of the 6th International Conference, Wireless Communications Networking and Mobile Computing (WiCOM)*, Sep. 2010, pp. 1–4.
- [6] O'Connor, P., and Kleyner, A. (2011). *Practical Reliability Engineering*. Hoboken, NJ: John Wiley & Sons.
- [7] Modarres, M., Kaminskiy, M. P., and Krivtsov, V. (2009). *Reliability Engineering And Risk Analysis: A Practical Guide*. Boca Raton, FL: CRC press.
- [8] Kamyod, C., Nielsen, R.H., and Prasad, N.R. (2013). “End-to-end availability analysis of IMS-based networks: Simplex and redundant system,” in *Proceedings of the 2013 IEEE International Conference on Wireless Communications and Networking Conference (WCNC)*, Shanghai, 1103–1108.
- [9] Nagy, L., Tombal, J., and Novotny, V. (2013). “Proposal of a queueing model for simulation of advanced telecommunication services over IMS architecture,” in *Proceedings of the 36th International Conference Telecommunications and Signal Processing (TSP)*, Rome, 326–330.

- [10] Mkwawa, I. M., and Kouvatso, D. D. (2008). "Performance modelling and evaluation of handover mechanism in IP multimedia subsystems," in *Proceedings of the 3rd International Conference Systems and Networks Communications*, Athens, 223–228.
- [11] Jianhui, W., Hao, J., and Wenguang, W. (2009). "A novel queuing model for IMS-based IPTV system," in *Proceedings of the 2nd IEEE International Conference Broadband Network & Multimedia Technology*, Beijing, 560–564.
- [12] Luo, A.A., Lin, C., and Wang, K. (2009). Performance modeling and evaluation using Queuing Petri Nets in IMS. *Proceedings of the 4th International Conference Communications and Networking*, Beijing, 1–5.
- [13] Mendiratta, V. B., and Pant, H. (2007). "Reliability of IMS Architecture," in *Proceedings of the International Conference Telecommunication Networks and Applications*, Dunedin, 1–6.
- [14] Arnold, T. F. (1973). The concept of coverage and its effect on the reliability model of a repairable system. *IEEE Trans. Comput.* 100, 251–254.
- [15] Zhang, X., Pham, H., and Johnson, C. R. (2010). Reliability models for systems with internal and external redundancy. *Int. J. Syst. Assur. Eng. Manag.* 1, 362–369.
- [16] Kamyod, C., Nielsen, R. H., and Prasad, N. R. (2012). "Resilience in IMS: end-to-end reliability analysis via Markov reward models," in *Proceedings of the 15th International Conference Wireless Personal Multimedia Communications (WPMC)*, Taipei, 564–568.
- [17] Guida, M., Longo, M., Postiglione, F. (2010). "Performance evaluation of IMS-based core networks in presence of failures," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM 2010)*, Miami, FL, 1–5.
- [18] Liu, Y., Liu, Y., and Wang, X. (2011). "Reliability Mechanism for the Core Control Network-Element S-CSCF in IMS," in *Proceedings of the International Conference Network Computing and Information Security (NCIS)*, Guilin, 57–61.
- [19] Kamyod, C., Nielsen, R. H., and Prasad, N. R. (2013). IMS intra-and inter domain end-to-end resilience analysis. In *Wireless Communications, Proc. of the 3rd Int. Conf. Vehicular Technology, Information Theory and Aerospace & Electronic Systems (VITAE)*, June 2013, pp. 1–5.
- [20] Kamyod, C., Nielsen, R. H., Prasad, N. R. (2014). "Resilience of the IMS system: the resilience effect of inter-domain communications," in *Proceedings of the 4th International Conference Wireless Communications*,

Vehicular Technology, Information Theory and Aerospace & Electronic Systems (VITAE), Aalborg, 1–4.

- [21] Verbrugge, S., Colle, D., and Pickavet, M. (2006). Methodology and input availability parameters for calculating OpEx and CapEx costs for realistic network scenarios. *J. Opt. Network.* 5, 509–520.

Biographies



Chayapol Kamyod is a lecturer in Computer Engineering program at School of Information Technology, Mae Fah Luang University, Chiang Rai, Thailand. He is also a postdoc researcher with Prof. Ramjee Prasad at Aarhus University, Herning, Denmark. He received his Ph.D. in Wireless Communication from the Center of TeleInFrastruktur (CTIF) at Aalborg University (AAU), Denmark. He received B.Eng. in Telecommunication Engineering and M.Sc. in Laser Technology and Photonics from Suranaree University of Technology, Nakhon Ratchasima, Thailand. In addition, he received M.Eng. in Electrical Engineering from The City College of New York, New York, USA. His research interests are resilience and reliability of communication and computer network, Internet of Things (IoT) and embedded technology.



Rasmus Hjorth Nielsen is currently heading product management as senior director at Movimento for over-the-air software updates in automotive. He has more than ten years of R&D and business experience in industry and academia. He is the founder and partner of R&D intensive network planning

companies and is working in the field of Communications and Networks focusing on Secure Network Architectures, Wireless Sensor Networks, Internet of Things, Optical Network Planning and Optimization, Virtualization and Management. From 2010, he was in academia as Postdoc and from 2011 to 2012 as Assistant Professor. He has been a technical manager on several industrial and European Commission (EC) projects. He has extensive project management and write-up experience for several project proposals. He was formerly principal architect in New Ventures as part of IOTG, Cisco Systems, driving IoT platform, edge computing and deterministic Ethernet. Rasmus has broad R&D knowledge in architectures, operational research and optimization. He has expertise in applied R&D for market and industrial needs and focus on Internet of Things (Wireless Sensor Networks, RFID, etc.), planning, monitoring, optimization, security and virtualization of next generation networks/future internet (NGN/FI) and architectures.



Neeli Rashmi Prasad is a security, mobile, wireless, networking and Internet of Things (IoT) strategist. She has throughout her career been driving business and technology innovation, from incubation to prototyping to validation and currently an entrepreneur and consultant in Silicon Valley. She has made her way up the waves of secure communication technology by contributing to the most groundbreaking and commercial inventions. She has general management, leadership and technology skills, having worked for service providers and technology companies in various key leadership roles. Prasad has lead a global team of more than 20 researchers across multiple technical areas and projects in Japan, India, throughout Europe and USA.

Her notable accomplishments include enhancing the technology of multi-nationals including CISCO, HUAWAI, NIKSUN, Nokia-Siemens and NICT, defining the reference framework for Future Internet Assembly and being

one of the early key contributors to Internet of Things. She is also an expert member of governmental working groups and cross-continental forums.

Previously, she has served as chief system/network architect on large-scale projects at both network operators and vendors looking across the entire product and solution portfolio covering security, wireless, mobility and cloud in domains such as IoT/M2M, eHealth and smart cities. She was one of the key contributors to the commercialization of WLAN for which she has published two books and she has a broad range of publications in internationally acclaimed journals and conference publications.



Ramjee Prasad, Fellow IEEE, IET, IETE, and WWRF, is a Professor of Future Technologies for Business Ecosystem Innovation (FT4BI) in the Department of Business Development and Technology, Aarhus University, Herning, Denmark. He is the Founder President of the CTIF Global Capsule (CGC). He is also the Founder Chairman of the Global ICT Standardisation Forum for India, established in 2009. GISFI has the purpose of increasing of the collaboration between European, Indian, Japanese, North-American and other worldwide standardization activities in the area of Information and Communication Technology (ICT) and related application areas.

He has been honored by the University of Rome “Tor Vergata”, Italy as a Distinguished Professor of the Department of Clinical Sciences and Translational Medicine on March 15, 2016. He is Honorary Professor of University of Cape Town, South Africa, and University of KwaZulu-Natal, South Africa.

He has received Ridderkorset af Dannebrogordenen (Knight of the Dannebrog) in 2010 from the Danish Queen for the internationalization of top-class telecommunication research and education.

He has received several international awards such as: IEEE Communications Society Wireless Communications Technical Committee Recognition Award in 2003 for making contribution in the field of “Personal, Wireless

and Mobile Systems and Networks”, Telenor’s Research Award in 2005 for impressive merits, both academic and organizational within the field of wireless and personal communication, 2014 IEEE AESS Outstanding Organizational Leadership Award for: “Organizational Leadership in developing and globalizing the CTIF (Center for TeleInFrastruktur) Research Network”, and so on.

He has been Project Coordinator of several EC projects namely, MAGNET, MAGNET Beyond, eWALL and so on.

He has published more than 40 books, 1000 plus journal and conference publications, more than 15 patents, over 100 Ph.D. Graduates and larger number of Masters (over 250). Several of his students are today worldwide telecommunication leaders themselves.



Nattapol Aunsri received the B.Eng. degree and M.Eng. degree in Electrical Engineering from Khon Kaen University and Chulalongkorn University, Thailand in 1999 and 2003, respectively. He obtained M.Sc. degree in Applied Mathematics and Ph.D. degree in Mathematical Sciences from New Jersey Institute of Technology, Newark, NJ, in 2008 and 2014, respectively. Currently, he is a Lecturer at the School of Information Technology, Mae Fah Luang University, Chiang Rai, Thailand. His research interests include ocean acoustics, Bayesian filtering, signal processing, and mathematical and statistical modeling.