PerformabilityEvaluation of Voice Services in ConvergedNetworks

Avaliação de Performabilidade de Serviços de Voz em Redes Convergentes

Almir Pereira Guimarães1*, Dênis Rodrigues1, Bruno Costa e Silva Nogueira1

Abstract: In the last years, the transmission of voice services in converged networks has experienced a huge growth. However, there are still some questions considering the ability of these networks to deliver voice services with acceptable quality. In this paper, we applied analytical modeling and simulation to analyze the quality of voice services using a new index, called $MOS_a$, which considers jointly the MOS index and the availability of the subjacent infrastructure. We consider the influence of different CODECs (G.711 and G.729), queuing policies (Priority Queuing and Custom Queuing), and the warm standby redundancy mechanism. Our goal is to analyze the quality of these services by taking into account overloading conditions in different architectures/scenarios. These scenarios were constructed using the modeling mechanisms Reliability Block Diagram and Stochastic Petri Nets in addition to a discrete event simulator. Experimental results indicate that the G.711 CODEC has a higher sensitivity both in terms of data traffic volume and allocated network resources in relation to the G.729 CODEC.

Keywords: VoIP — MOS — Availability — Reliability Block Diagram — Stochastic Petri Nets

1. Introduction

Over the last few years, the use of voice transmission services in converged networks has significantly increased. This considerable growth is somewhat related to interoperability offered by voice and data transmission and its low cost. In many companies, VoIP (Voice over Internet Protocol) services are strategic and their degradation may lead to considerable losses. They should be continuously provided even when events like congestion, link malfunctioning, routing instabilities, sabotage, natural disasters, hardware or software failures happen [1]. Hence, either in normal operation or in case of network failure, VoIP services should be preserved.

An important factor that is used for the performance analysis of VoIP services is the MOS (Mean Opinion Score) [2] index. This index is directly impacted by performance factors and it can be calculated based on the voice packet loss rate [3, 4] of a voice connection. For a better perception of the voice services quality, it is important to consider both the performance aspects, reflected by the MOS index, as the
availability aspects of these services. Our work analyzes the behavior of voice services using a new index, called MOS_p, which considers jointly the MOS index and the availability [5] of the subjacent infrastructure. Several scenarios and architectures were constructed through different CODECs (Coder-Decoder), such as G.711 [6] and G.729 [7]; queuing policies, such as Custom queuing [8] and Priority Queuing [9]; and the warm standby redundancy mechanism [10]. Due to the system inexistence and costs of settings, analytical and simulation models are adopted as a strategy for a quantitative analysis.

The proposed scenarios and architectures were constructed using Graphical Network Simulator-3 (GNS-3) [11] and Mercury [12] tools. For measuring performance, simulation environments using GNS-3 tool are utilized. GNS-3 is a software emulator for complex networks, launched in 2008. In turn, for dependability evaluation we used analytical models built with the Mercury tool. Mercury has been conceived to allow dependability/sustainability evaluation of general systems. The corresponding dependability models were created using the Reliability Block Diagram (RBD) [10] and Stochastic Petri Nets (SPN) [13] modeling mechanisms following a hierarchical approach.

The rest of this paper is organized as follows. Section 2 presents the state of the art in this area. Section 3 describes the network infrastructure models. Section 4 explains the adopted methodology in order to construct and analyze different scenarios. Section 5 presents the evaluation of the performance and dependability scenarios, concerning different CODECs, queuing policies and the warm standby redundancy mechanism. Finally, Section 6 concludes the paper and introduces ideas for future research works.

2. Related Work

In the last few years, much has been done to deal with issues relating to performance or dependability of real time applications, such as VoIP services [14, 15, 16, 17, 18, 19, 20, 21, 22]. Initially, Reference [14] provides an insight into the impact of TCP segmentation in VoIP monitoring and the solution that has been applied to face it. This study focuses on voice signaling messages which include relevant data during calls negotiation. The importance of such messages is justified because if they are lost, monitoring tools may produce erroneous statistics and lose the corresponding multimedia transmissions.

In turn, the work presented in [15] simulates the VoIP over IEEE 802.11b and 802.11e wireless local area networks (WLAN) with different CODECs by using the NS-2 tool (Network Simulator-2). The main objectives were to analyze the Mean Opinion Score (MOS), throughput, and packet drop rate of VoIP traffic. It was found that the CODEC G.711 provides the best quality of VoIP calls for the network scenarios analyzed.

In [16], a systematic approach for quantifying the reliability of enterprise VoIP networks is presented. It identifies two key challenges for designing enterprise VoIP service infrastructure: i) there are no universally accepted objectives or standards for the reliability of the enterprise VoIP services; ii) there is no well known common set of reliability metrics to be used in enterprise VoIP service planning. [16] presents an enhanced method and procedure for reliability calculation, using a network matrix representation and RBD.

The authors of [17] focus on network impairments that can seriously degrade VoIP performance, based on different queuing mechanisms. Three different mechanisms, DropTail (FIFO), RED and DiffServ, and their effects are analyzed. Measurements of jitter, end to end delay and packet loss are studied. Experimentation has proven that under burst traffic conditions up to a congestion level, DiffServ seems to perform better in all three categories examined.

[18] evaluates, with measurements, the effect of the voice CODECs, G.711, G.729, and G.723.1, on the quality of VoIP calls over a MANET in a real indoor environment which includes a corridor within a university campus.

[19] analyzes and plans the capacity of a real enterprise-class voice gateway system, which supports the voice communication of six university campuses. It was proposed the use of computer modeling and simulation for planning the critical resources of the investigated gateway system. Different scenarios are simulated to assess the current and future capacity of the system under study.

In turn, [20] investigates and compares the performance of the G.711, G.723 and G.729 CODECs on SIP architectures. It analyzes parameters such as jitter, throughput and MOS. The Openet tool was used in order to compare the different CODECs.

The work in [21] utilizes simulation to determine the optimal values of Transmission Opportunity (TXOP) and Frame Aggregation (FA) that maximizes VoIP capacity. TXOP and FA are two important Medium Access Control (MAC) Layer enhancements provided by IEEE 802.11n standard. The voice

![Figure 1. Basic Structures](image-url)
capacity is calculated when optimal values of TXOP and FA are simultaneously used. Additionally, the VoIP capacity over User Datagram Protocol (UDP) and TCP Friendly Rate Control (TFRC) protocol in the presence of Transmission Control Protocol (TCP) traffic is determined. Finally, Reference [22] analyses the Quality of Service (QoS) of VoIP in 802.11ac networks under varying networking conditions. It measured various QoS metrics such as throughput, jitter, latency, loss ratio and Mean Opinion Score (MOS) using an Extended Service Set (ESS) testbed with IPv4 and IPv6 as network layer protocols. The result shows a degradation in VoIP quality when background traffic increases. IPv6 VoIP calls had a better Quality compared to IPv4 calls.

Researchers have used different approaches, however note that none of them deal jointly with issues of performance and dependability of VoIP services. This research extends our preliminary works in [23, 24, 25] merging performance and dependability issues and proposing the MOS \(_n\) index that correlates MOS index and the availability of the subjacent infrastructure in different scenarios and architectures. The proposed index reflects, in a more realistic way, the effects of availability on the voice services quality. In addition, hierarchical models, which consider the main components of the infrastructure, are proposed to calculate its availability.

3. Network Infrastructure Models

This section presents the base models for quantifying the availability metric of converged network infrastructures. The proposed models are generic enough to represent a wide variety of mechanisms found in real converged networks infrastructures.

3.1 RBD Models

The most common structures on converged network topologies are serial, parallel and both [26]. Figure 1 depicts three examples where the blocks (\(C_1\), \(C_2\), \(C_3\), \(C_4\) and \(C_5\) components) are arranged through series (Figure 1(a)), parallel (Figure 1(b)) or series-parallel (Figure 1(c)) compositions. A series structure on a set of components means that for the whole subsystem to work, every component has to be functioning. Assuming a structure with \(n\) components in series, the availability (reliability) \([10]\) is obtained by:

\[
P_{(s)}(t) = \Pi_{i=1}^{n} P_i(t)
\]

where \(P_i(t)\) is the availability or the reliability of block \(i\). On the other hand, a parallel structure means that the whole subsystem can function if at least one (or more) of the components is working. Taking into account \(n\) components in a parallel structure, the system availability (reliability) is:

\[
P_{(p)}(t) = 1 - \Pi_{i=1}^{n} (1 - P_i(t))
\]

In order to calculate the availability (reliability) of a series-parallel structure, the serial results must be combined and placed into parallel equations. For other examples and closed-form equations, the reader should refer to [10].

3.2 SPN Models

This section briefly presents the SPN building blocks [27] for obtaining dependability metrics.

3.2.1 Simple Component

Figure 2(a) shows the dependability model of a generic simple component. This component is characterized by the absence of redundancy. Places \(X_{ON}\) and \(X_{OFF}\) are the model component’s activity and inactivity states.

The \(MTTF\) (Mean Time To Failure) \(^1\) and \(MTTR\) (Mean Time To Repair) \(^2\) parameters represent the delay associated with the transitions \(X_{MTTF}\) (in our study, label “X” must be applied according to the component name) and \(X_{MTTR}\) respectively.

3.2.2 Warm Standby Model

Figure 2(b) shows the dependability model of the warm standby [10] redundancy approach. Places \(X_{ON}\), \(X_{OFF}\) represent the activity and inactivity states of the components. Place \(X_{Sp_{-}ON}\) represents the spare (Sp) component of \(X\) in non-operational state. In turn, place \(X_{OSp_{-}ON}\) represents the spare component of \(X\) in operational (OSp) state. Initially, the spare component is in the non-operational state. As the main component fails, the transition \(X_{ACTSp}\) is enabled. Its

\(^1\)This parameter is supplied by the manufacturer.

\(^2\)This parameter is closely related to the maintenance policy adopted by organizations, which depend on equipment and site criticality as well as on the associated delays needed to reach remote locations.
firing represents the start of operation of the spare component. This time period is named Mean Time To Activate ($MTTA_w$). Immediate transition $X_{DCTSp}$ represents the return to initial condition.

The $MTTR$ parameter represents the delay associated with the $X_{MTTR}$, $X_{Sp-MTTR}$ and $X_{OSP-MTTR}$ transitions. The $MTTF$ parameter of each component represents the delay associated to the $X_{MTTF}$, $X_{Sp-MTTF}$ and $X_{OSP-MTTF}$ transitions. In this work, it is considered that the transition $X_{Sp-MTTF}$ has a delay 50% higher than transition $X_{OSP-MTTF}$. The timed transitions are exponentially distributed and have a single server concurrency policy.

4. Evaluation Methodology

This section presents the proposed methodology that aims the analysis of performance and dependability issues of converged networks. It consists of seven macro activities: Problems and Components Definition, Obtaining Information, Model Generation and Metrics Mapping, Validation of the Model, Calculating the MOS index, Evaluation of the Scenarios and Results Analysis (see Figure 3).

The first activity determines the problem together with its scope. At the end of this activity, two artifacts are obtained, namely: Problem Diagram and the considered Metrics. Problem Diagram is a document that contains aspects of the infrastructure under analysis such as: its topology, interconnection and component dependencies, component definitions and hardware settings. The second artifact, the set of metrics, defines the indicators used to analyze the dependability and performance aspects of the system. The main metrics are the voice packet loss rate, that it is used to calculate the MOS index of a voice connection [3], and the steady state availability of the subjacent infrastructure.

Information about the components is obtained in the Obtaining Information activity. For dependability models, information is related to the following parameters: MTTF and MTTR. In turn, the parameters of the performance models components are configured in accordance with each defined scenario and refers to different queuing policies, CODECs etc.

In the Models Generation and Metrics Mapping activity, the artifacts generated are the models (performance and dependability) and their metrics. It is accomplished by modeling each system component and composing them through composition rules. This activity also takes into account the metrics mapping that represent the metrics through expressions depicted by elements of each model. Regarding dependability models, this work adopts a hybrid modeling strategy [28] that considers the advantages of both RBD and SPN. Such a hierarchical approach is adopted to mitigate the complexity for representing large systems that can generate the state space explosion problem [29] for some state based models.

The activity Validation of the Model analyzes and carries out the adjustments when necessary. It concerns the evaluation of metrics considering performance and dependability scenarios. The checkpoint means that the models are mature enough to be evaluated according to established criteria. The end of this activity is reached when one or more models are obtained to solve the problems described.

After Validation of the Model activity, the Calculating the MOS$_i$ index activity can be conducted. From the obtained metrics, it is calculated the MOS$_i$ index (see Equation 3) that relates the MOS index of a scenario $i$, obtained through the voice packet loss rate, with the steady state availability ($A$) of the subjacent infrastructure.

$$MOS_i = MOS_i \times A \quad (3)$$

Then, the Evaluation of the Scenarios activity can be conducted. For evaluation purposes of the MOS$_i$ index, the parameters used for the MOS index will be considered. [30] says that a MOS rating above 4.0 matches the level of quality available in the current Public Switched Telephone Network (PSTN), a rating above 4.3 corresponds to the best quality, where users are very satisfied, and a rating between 4.0 and 4.3 corresponds to a high quality level, where users are satisfied. A MOS rating between 3.6 and 4.0 corresponds to a medium quality level, in which some users are dissatisfied. A MOS rating in the range between 3.1 and 3.6 corresponds to a low level of quality, where many users are dissatisfied. At a MOS rating in the range between 2.6 and 3.1, the level of quality is poor, and nearly all users are dissatisfied. Finally, a MOS below 2.6 is not recommended. Regarding availability, [31] states that a system with an availability of 99.999% is classified as high availability system. On the other hand, an availability of 90% and 99.0% classifies the system as unmanaged and managed respectively.

Finally, in the Results Analysis activity, the results obtained are interpreted and explained by adopting an appropriate vocabulary to the customer background and objectives.

5. Case Study

In this section our goal is a analysis of the voice services considering the influence of different CODECs, queuing policies and redundancy mechanisms on the their quality through different architectures and scenarios that were built to obtain performance and dependability results. The evaluation scenarios are similar to those obtained in small and medium enterprise networks.

Our work is based on two architectures (see Figure 4) [27]. Figure 4(a) shows the simulation environment of the first architecture (architecture $A_1$). If a component (Router $R_1$, Link $L_1$ or Router $R_2$) fails, the system goes down. Figure 5(a) shows the respective dependability model that adopts a hierarchical approach. In the upper layer, each component is modeled as a block in a series structure. The lower level models represent the routers (Router $R_1$ or Router $R_2$) in a series structure.

In turn, Figure 4(b) shows the simulation environment of the Architecture $A_2$. This architecture is similar to $A_1$.
and was obtained through the insertion of a redundant link (Link $L_2$) that is configured as a spare that adopts the warm standby redundancy mechanism, such that when the primary link (Link $L_1$) fails, the spare link assumes the role of the primary one. After recovery, the system returns to the initial state. If a router (Router $R_1$ or Router $R_2$) or both links fail, the system goes down. The dependability model adopts a hierarchical approach (see Figure 5(b)). The upper level model is a RBD and the lower level models are RBD and SPN. The SPN model represents the warm standby redundancy approach between the primary and spare links (see Figure 2(b)). Table 1 shows the parameters used in the dependability models of the architectures $A_1$ and $A_2$.

In the simulation environments, the IP phones are responsible for the voice connections and the machines generate the data traffic. The IP phones generate voice traffic with different CODECs (G.711 and G.729). Table 2 shows that G.711, titled Pulse Code Modulation (PCM), uses the most bandwidth, whereas it introduces the least amount of latency. These factors have made PCM inexpensive to deploy and highly effective in transporting voice traffic over long distance. By contrast, G.729, described as coding of speech at 8kbit/s using code-excited linear prediction speech code (CS-ACELP), provides significant bandwidth savings at the expense of increased latency.

It will be used six, eight, ten or twelve machines in different scenarios in order to analyze the behavior of voice traffic with the increase of data traffic. Each machine, connected on router $R_1$, generates ICMP packets with 1432 bytes at a rate of one packet per second.

The serial interfaces 1 (Architectures $A_1$ and $A_2$) and 3 (Architecture $A_2$) were chosen for the configuration of the queuing policies custom queuing ($CQ$) and priority queuing ($PQ$) due to the fact that they are the critical points where the voice and data traffic will dispute the access to the network resources. The analysis of each performance scenario will be done in these interfaces considering the resulting MOS index from the voice packet loss ratio in a time of one minute. This index is calculated considering different CODECs (G.711 and G.729), number of machines generating data traffic and network resources configured.

The different network resources were attributed through the queuing policies $CQ$ and $PQ$. $CQ$ allows a custom configuration of resources in an interface to different applications. It allows configuration of a specified number of bytes to be forwarded from a queue each time that queue has been serviced. It has sixteen possible queues. A maximum number of packets per queue can be specified [8]. In our work, $CQ$ was configured with two queues, one queue for voice and one queue for data and attributed different network resources for each one.

$PQ$ provides static allocation of network resources in an interface. It has a set of queues and a scheduler that empties them in priority sequence. When asked for a packet, the scheduler examines the highest priority queue and, if there is data present, returns a packet from that queue [9]. There are four output queues: High, Medium, Normal and Low. The network component places traffic in these queues, based on predefined filters. $PQ$ was configured with High and Medium queues being used.

Figure 6 shows the simulation results for the $PQ$ queuing policy. This figure depicts the influence of the number of machines (6, 8, 10 and 12) and the influence of the queue (High or Medium) where voice traffic is configured on the $MOS_v$ index that considers the performance and dependability aspects.

In the Figure 6(a), the G.711 CODEC was used. When the voice traffic is configured to the high queue, the $MOS_v$ index is not influenced by the number of machines in both $A_1$ and $A_2$ architectures. This is due to the fact that the high queue has the highest priority. In turn, when the voice service is configured to the medium queue, the $MOS_v$ index is sensitive to the increase in the number of machines. In both cases, there is a small variation of this index due to the effect of the availability of each architecture.

Figure 6(b) shows the behavior of the G.729 CODEC. This CODEC provides a stability of the $MOS_v$ index values for both high and medium queues. In turn, this index presents a small variation resulting from the effect of the availability of each architecture.

Considering $CQ$ queuing policy, Figure 7 shows both the influence of the allocated network resources (25.00%, 50.00% and 75.00%) to the voice traffic and the influence of the number of machines (6, 8, 10 and 12) on the $MOS_v$ index.
Performability Evaluation of Voice Services in Converged Networks

In the Figure 7(a), the G.711 CODEC was used. The voice traffic has a greater sensitivity to the number of machines generating data traffic when it has 25% or 50% of the allocated network resources. On the other hand, this traffic is less sensitive when it has 75% of the allocated resources. This is demonstrated by the MOS index variation. In both cases, the availability of each architecture causes a small variation of this index.

Figure 7(b) shows the behavior of voice traffic when the G.729 CODEC was used. It is interesting to note that the voice traffic is less sensitive to variations in both allocated network resources and data traffic. The availability of each architecture also causes a small variation of this index.

### Table 1. Resulting Availability and Dependability Parameters of the Architectures A1 and A2

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Availability</th>
<th>MTTF(R1)</th>
<th>MTTR(R1)</th>
<th>MTTF(R2)</th>
<th>MTTR(R2)</th>
<th>MTTF(L1)</th>
<th>MTTR(L1)</th>
<th>MTTF(L2)</th>
<th>MTTR(L2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0.9608</td>
<td>105,000</td>
<td>12</td>
<td>105,000</td>
<td>12</td>
<td>296</td>
<td>12</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>A2</td>
<td>0.9985</td>
<td>105,000</td>
<td>12</td>
<td>105,000</td>
<td>12</td>
<td>296</td>
<td>12</td>
<td>296</td>
<td>12</td>
</tr>
</tbody>
</table>

### Table 2. Encoding and MOS

<table>
<thead>
<tr>
<th>Coding</th>
<th>ITU-T Standard</th>
<th>Bit Rate (kbps)</th>
<th>Coding Delay (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCM</td>
<td>G.711</td>
<td>64</td>
<td>0.125</td>
</tr>
<tr>
<td>CS-ACELP</td>
<td>G.729</td>
<td>8</td>
<td>20</td>
</tr>
</tbody>
</table>

### 6. Conclusion

This work proposed simulation and analytical models to evaluate performance and dependability aspects of voice services. Several architectures and scenarios were constructed combining different number of machines, queuing policies, CODECs and a redundancy mechanism.

The merging of performance and dependability issues into the same problem is one of our contributions. Another relevant contribution is the proposed $MOS_a$ index that aims to quantify the effects of the infrastructure availability on the MOS index, associated with each voice connection. This index showed that the G.711 CODEC has a higher sensitivity both in terms of data traffic and allocated network resources.

For future work, we plan to extend these models to include the effects of different maintenance policies, multiple repairing units as well as taking into account different recovery strategies.

### Author contributions

The authors contributed equally to this work.
Performability Evaluation of Voice Services in Converged Networks

![Hierarchical Models - $A_1$ and $A_2$]

**Figure 5.** Hierarchical Models - $A_1$ and $A_2$

**References**


Performability Evaluation of Voice Services in Converged Networks

Figure 6. $MOS_a$ index in accordance with the availability of the subjacent infrastructures (architectures $A_1$ and $A_2$)


