due to PU arrivals or channel failures, instead of being terminated are allowed to queue into the orbit for later retrial transmission. This results in higher service retainability. Moreover, the larger the retrial probability is, the lesser likely is the interrupted ESU to be terminated forcibly and thus significantly improves the retainability level further.

C. Channel Availability

From the retainability point of view, we observe that the system achieves significantly improved performance with reserved channels. However, the price the system pays for channel reservation is the reduced capacity as well as the decreased channel availability. Since service interruptions lasting over long periods is not desirable, thus to maintain a higher level of service retainability is considered as more fundamental. Thereby, the reservation of channels is recommended up to a certain extent although the channel availability of new users decreases as depicted in Figs. 5.10 and 5.11.

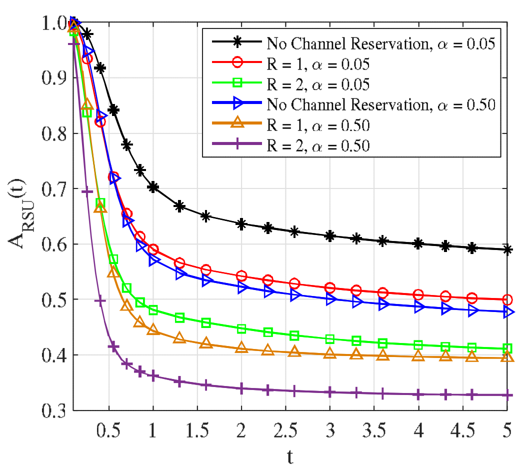

Figure 5.11: Channel Availability for RSUs as a function of time when channel reservation is applied.

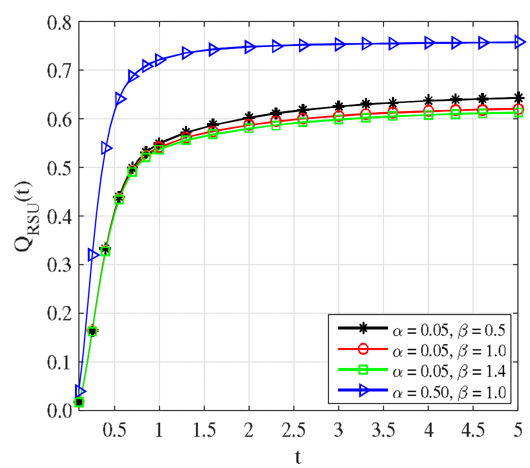

Figure 5.12: Network unserviceable probability of RSUs as a function of time with different failure and repair rates.

As time grows, the channel availability of PUs and RSUs decreases due to the commencement of more users in the system. It is clear from these figures that the channel availability of PU and RSU services get decreased because of both channel failures as well as channel reservations. In the case of no channel reservation, all channels would

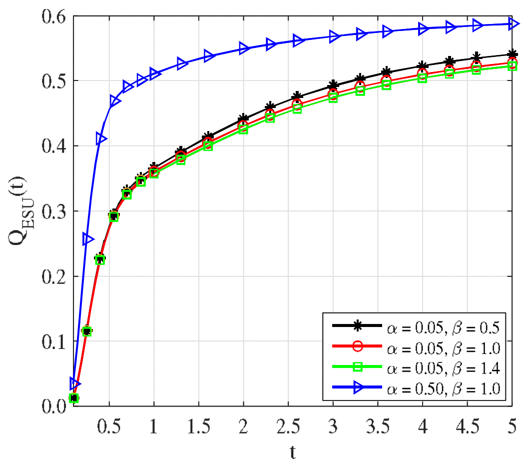

Figure 5.13: Network unserviceable probability of ESUs as a function of time with different failure and repair rates.

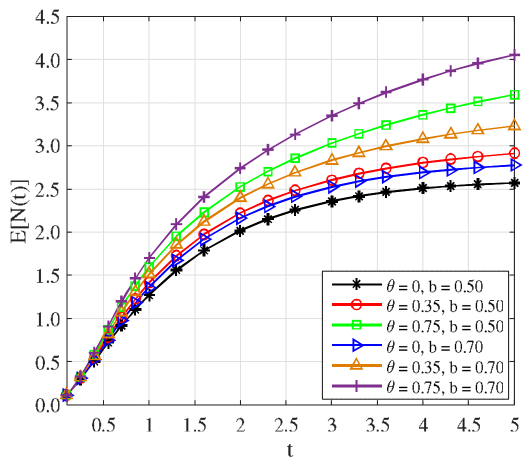

Figure 5.14: Mean orbit size as a function of time with different balking and retrial probabilities.

be available for new users and therefore the scheme with no channel reserved always results in the highest channel availability. Moreover, when a fewer number of channels is reserved i.e. at smaller values of R, more channel access opportunities are given to the new users and consequently, the channel availability becomes higher. These figures further illustrate that the PUs' channel availability is higher than that of the RSUs due to its priority on accessing the channels. In conclusion, this result reveals a need for an assessment of the trade-off between the ongoing services' retainability and the new users' channel availability.

D. Network Unserviceable Probability (NUP)

When evaluating the NUP, the effects of both channel availability and service retainability are considered. Thus, deriving and plotting such metric helps us assessing the overall performance of the proposed scheme. In Figs. 5.12 and 5.13, we evaluate the NUP of RSU and ESU services respectively considering different configurations. As expected, NUP of both services first increases with an increase in time and then shows the steady-state behavior, however, the curve for ESUs converges to the steady-state NUP at comparatively larger t. The results in these figures confirm that NUP increases with a higher channel failure rate. On the other hand, with a higher channel repair rate, the NUP improves to some extent. This is because, with higher β, the failed channels are repaired shortly and become idle again, leading to increased channel opportunities for users. However, as can be observed, for smaller values of t, the increase of repair rate may not help to improve the NUP.

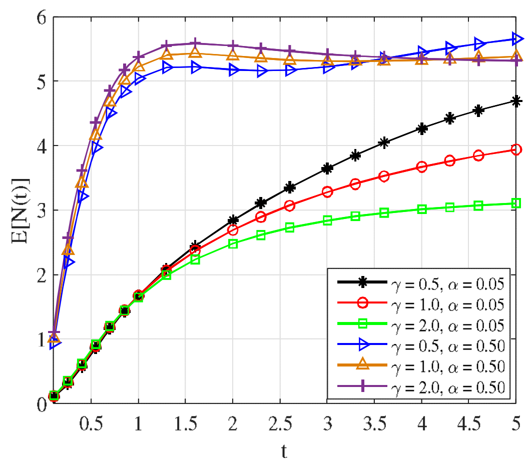

Figure 5.15: Mean orbit size as a function of time with different retrial and failure rates.

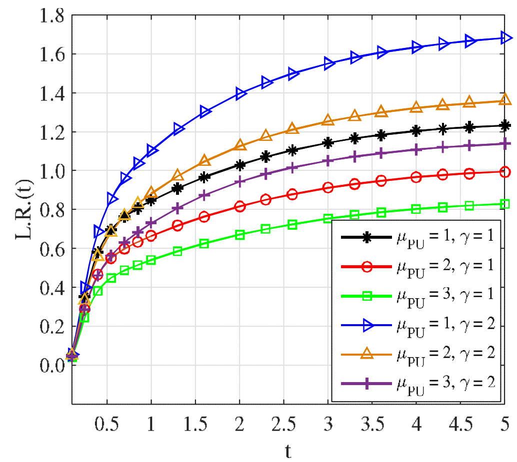

Figure 5.16: Average loss rate due to impatience as a function of time with different retrial and service rates.

E. Mean Orbit Size

Figs. 5.14 and 5.15 depict the expected size of the orbit as a function of time under various configurations. We notice that as time passes the mean orbit size increases as it should be. Fig. 5.14 confirms that since employing the access retrial to ESUs ($\theta > 0$), allow the would be terminated users to join the orbit, therefore, with higher retrial probability, a greater number of interrupted ESUs join the orbit for later transmission and consequently increases the mean orbit size. Also, it is observed that smaller the balking probability $(1-b)$ is, an arriving ESU is routed to enter the orbit with a bigger probability (b), resulting in larger orbit size, which agrees with our expectations.

Furthermore, Fig. 5.15 illustrates that the mean orbit size is significantly influenced by the channel failure rate. With higher α, an ongoing session of ESU service is more likely to be interrupted and accordingly employing the retrial policy leads to a bigger orbit size. On the other hand, with a higher retrial rate, the size of the orbit decreases. However, under certain time values, the situation changes for higher values of α. With a larger α, the number of idle channels in the network diminishes. Thus, as the retrial rate increases, a greater number of blocked ESUs rejoin the orbit due to retrial congestions and consequently the mean orbit size increases. This indicates that the mean orbit size is sensitive to the value of t as well as α.

F. Loss Rate due to Impatience

Finally, in Fig. 5.16 the average loss rate of ESU services due to impatience at different time points is plotted. We observe that the loss rate increases gradually and then attains the steady-state with the passage of time. Further, as shown in the figure, diminishing the PU service duration, frees up its channels faster, enabling less balking and reneging of ESUs. This leads to an improvement in ESUs' performance in terms of their loss rate. On the other hand, with a higher retrial rate, a greater number of ESUs are blocked due to retrial congestion, thereby increases the loss rate. Consequently, in brief, these results enable a trade-off between the PU service rate and the retrial rate that can be tuned according to the QoS requirements of the secondary traffic.

5.8 Chapter Summary

The transient analysis of CRNs from the perspective of dependability theory remains largely uncharted despite extensive research efforts in the past decade. In this chapter, a DSA scheme with channel reservation and retrial policy is proposed which enhances the heterogeneous SUs' performance in CRNs prone to random channel failures, in terms of service retainability and network unserviceable probability. The mathematical expressions of transient performance metrics for channel access in CRNs are derived utilizing the uniformization tool. Through CTMC modeling, the performance of the proposed scheme has been evaluated. According to the numerical results, although channel reservation may not be required at very small time points and very low channel failure rates, but the proposed reservation scheme is recommended for QoS provisioning. Furthermore, results have also shown that employing the retrial phenomenon and impatient behavior subject to balking and reneging allows investigating the SU performance more accurately. Overall, the scheme presented in this chapter provides a systematic approach for time domain reliability analysis of channel access in multichannel CRNs.

Chapter 6

Reliability and Availability Analysis in CRNs with Multi-Level Channel Reservation

Ability is important in our quest for success, but dependability is critical.
— Zig Ziglar

6.1 Introduction

Along the lines of Chapter 5, this work also examines availability and reliability analysis of cognitive radio networks (CRNs) in time domain from the perspectives of dependability theory. However, this chapter encompasses a different range of important features, allowing the analysis of their joint effects in more complex scenarios. Herein, the dynamic spectrum access (DSA) scheme includes multi-level channel reservation technique to facilitate heterogeneous secondary users (SUs). To capture the diversity of channel failures, the distinct channels with different failure rates and repair rates are investigated. Moreover, while modeling the CRN using a continuous-time Markov chain (CTMC) based analytical model, a state of the CTMC is represented by different channel operation modes.

In the CRN context, since the primary users (PUs) have absolute priority over the SUs, secondary transmissions collide with primary transmissions. In view of this, several authors have explored a priority mechanism called channel reservation (see e.g., El Azaly et al. [260]), but in a limited manner. A channel reservation policy was developed by Chakraborty and Misra [246] where a certain number of channels are reserved for PUs, but by employing this policy, forced terminations of SUs may not be decreased substantially. On the other hand, Ding and Zhao [261] proposed a scheme that exclusively reserves a number of channels to the SUs, not allowing PUs to access these. However, none of these works have dealt with the heterogeneity of SUs, while it is a valuable feature since each arriving SU carries a distinct job and may differ in sensitivity to delay. For instance, the real-time applications such as e-commerce transactions cannot tolerate delay whereas the non-real time applications such as email can tolerate acceptable delay. The works of Chu et al. [249] and Falcão et al. [250] differ in that sense, as the analysis therein is presented by considering a heterogeneous secondary system. Chu et al. [249] provided channel reservation with two SU types however, not distinguishing in number of reserved channels for each SU type. With such single level reservation, if the higher priority SUs' access is limited it necessarily impacts the lower priority SUs as well and accordingly may degrade the network performance. Later, Falcão et al. [250] proposed a DSA scheme with multi-level channel reservation by allowing specific spectrum restriction for each SU type. Nevertheless, the effect of unreliable channels on CRN is not addressed. Note additionally that none of the above preceding papers is targeted at performing reliability or availability analysis from a dependability perspective.

Motivated by the aforementioned observations, this chapter targets the lack of a more complete model. The present work admits multi-level channel reservation which allows limiting a specific number of channels for each SU type i.e., real-time SU (RSU) and elastic SU (ESU). Consequently, the PUs' highest access priority is ensured by providing access to any channel including those reserved for the SUs, thus reflecting the opportunistic spectrum access (OSA) concept. We investigate the effects of multi-level channel reservation and distinct failure and repair rates on the error-prone CRN performance. In brief, different from those existing works as indicated in Table 5.1, the scheme proposed in this chapter is designed by jointly considering four main features, i.e., multi-channel CRNs, hybrid SU traffic, channel failures and multi-level channel reservation, and is then examined in time domain from the dependability theory's perspective.

The rest of the chapter is organized in the following manner. In Section 6.2, we describe the system model together with the assumptions. Section 6.3 presents the proposed DSA scheme. An extensive analytical model based on multi-dimensional CTMC is given in Section 6.4. Following this, in Section 6.5, we obtain the solution for the transient probabilities and derive some performance indices of the system. In Section 6.6, we present and discuss the main numerical results and thereafter conclude the chapter in Section 6.7.

6.2 System Model and Assumptions

A centralized CRN architecture that allocates spectrum resources for both PUs and SUs is considered. The licensed spectrum band consists of $N \in \mathbb{Z}^+$ number of channels for PUs, where $\mathbb{Z}^+$ denotes the set of positive integers. PUs have the full privilege to utilize the spectrum whereas SUs can opportunistically access channels that are not occupied by the PUs. With regards to the priority mechanism, the RSU services have higher priority than the ESU services in case of preemption of a SU service upon PU arrival i.e. the ongoing ESU services are subject to termination prior to the RSU services. Additionally, an individual number of channels are restricted from access to RSUs and ESUs owing to multi-level channel reservation. Moreover, all of the communication channels in the considered CRN are subject to failures caused by various reasons like fading, shadowing or interference, with distinct failure rates.

Further, the following assumptions are made to develop our analytical model.

- The user arrivals follow Poisson processes with rates λ_{PU}, λ_{SE} and λ_{SR} and the service times are exponentially distributed with rates μ_{PU}, μ_{SE} and μ_{SR} for PU, ESU and RSU services, respectively.

- The channel failure is generated by an exponential distribution with rate α_i for the i^{th} channel, where $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_i \leq \alpha_{i+1} \leq \cdots \leq \alpha_N$. Also, it is assumed that both idle and occupied channels are prone to failures.

- As soon as a channel fails, the failed channel undergoes for repair immediately. The repair time of the i^{th} failed channel is assumed to have an exponential distribution with a repair rate β_i.

- The channel reservation allows restricting R_1 and R_2 amount of spectrum resources (in channel units) with higher failure rates from RSUs and ESUs, respectively. In brief, during operation an RSU can access $N - R_1$ channels, with $N \geq R_1$ and similarly, an ESU can access $N - R_2$, with $N \geq R_2$ and ($R_1 + R_2 \leq N$), whereas the PU services can utilize all channels.

- The inter-arrival times, service times, and repair times are assumed to be independent of each other.

- The spectrum adaptation and sensing latency are much shorter in contrast with the duration of service events. Herein, the events indicate the arrival and departure of both PU and SU services and the failure and recovery of the channels.

6.3 The Proposed Dynamic Spectrum Access Scheme

This section provides details of the considered five channel status modes and the channel access procedure adopted upon the occurrence of service arrivals, departures, channel failures and repairs, respectively.

6.3.1 Channel Status with Five Operation Modes

Considering the occupancy and failures of channels, the status of each channel in the considered CRN could reside in one of the following modes.

- *Mode1*: the idle state of a channel. An idle channel can be allocated for a new request or interrupted service.

- *Mode2*: the ESU occupied state of a channel. In this state, a channel cannot be allocated for a service until the completion of an ongoing session.
- *Mode3*: the RSU occupied state of a channel. Again, in this state, a channel is not available for allocation until an ongoing session has finished.
- *Mode4*: the PU occupied state of a channel. Likewise, in this state, a channel being busy cannot be allocated for a service before the completion of an ongoing session.
- *Mode5*: the failed state of a channel. Both occupied and idle channel may fail.

Different failure and repair rates may apply to distinct channels since network parameters like frequency and PHY/MAC parameters are not identical, as explored by Balapuwaduge and Li [262]. Re-ordering the channels in the CRN lexicographically and identifying them by their order number i.e. $C_1, C_2, \cdots, C_N$, which hereafter is referred to as the channel index, as shown in Fig. 6.1.

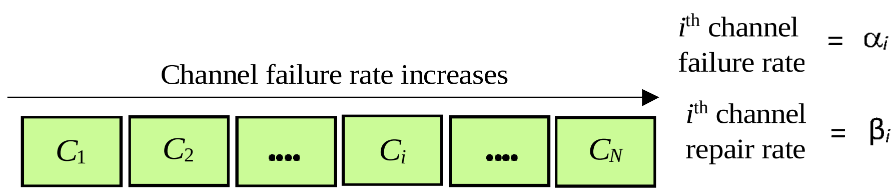

Figure 6.1: Heterogeneous channels with their failure and repair rates.

6.3.2 Channel Allocation Scheme

In this section, we provide details of the channel access procedure adopted upon the occurrence of an event. Also, the scheme is partially illustrated in flow charts in Figs. 6.2 and 6.3. The states while modeling the system are represented by the status of the channels. State transitions are triggered by one of the following events.

1) PU Arrival: Upon the arrival of a new PU service, the available idle channel with the lowest index i.e. having the lowest failure rate is allocated to it. In the case where no idle channel is available while at least one channel is currently occupied by an ESU, then

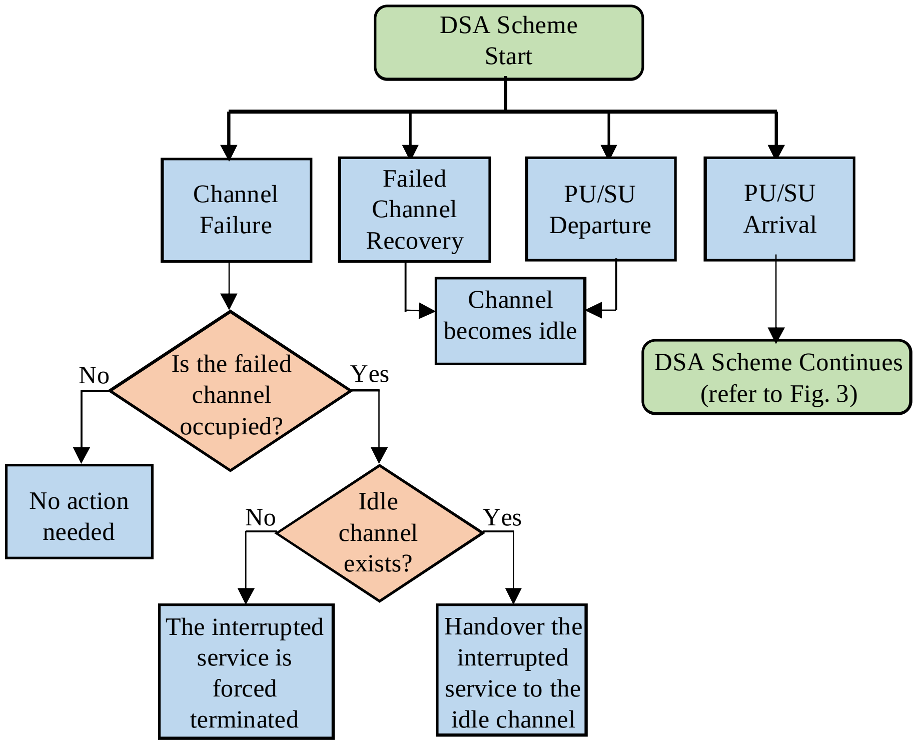

Figure 6.2: Flow chart illustration of the proposed DSA scheme in part.

one of those ESUs which has occupied the lowest index channel has to vacate the channel immediately by re-allocating it to the newly arriving PU. The interrupted ESU service performs handover to another accessible idle channel with the lowest channel index, if available, otherwise, it is forced to terminate. On the other hand, if the arriving PU finds neither an idle channel nor an ESU occupied channel available while at least one channel is currently occupied by an RSU, then one of the RSUs which has occupied the lowest index channel gets interrupted and is transferred to another accessible vacant channel with the lowest index. If no vacant channel is found, the interrupted RSU service is forced to terminate. As mentioned earlier, the reason for the above mentioned priority mechanism is to preserve ongoing RSUs' transmissions prior to that of the ESUs. However, if all the channels are occupied by PUs and/or have failed, then a new PU request gets blocked.

2) SU Arrival: When a new SU service (either ESU or RSU) arrives, the system allocates an idle channel (based on their accessibility) with the lowest index to the newly

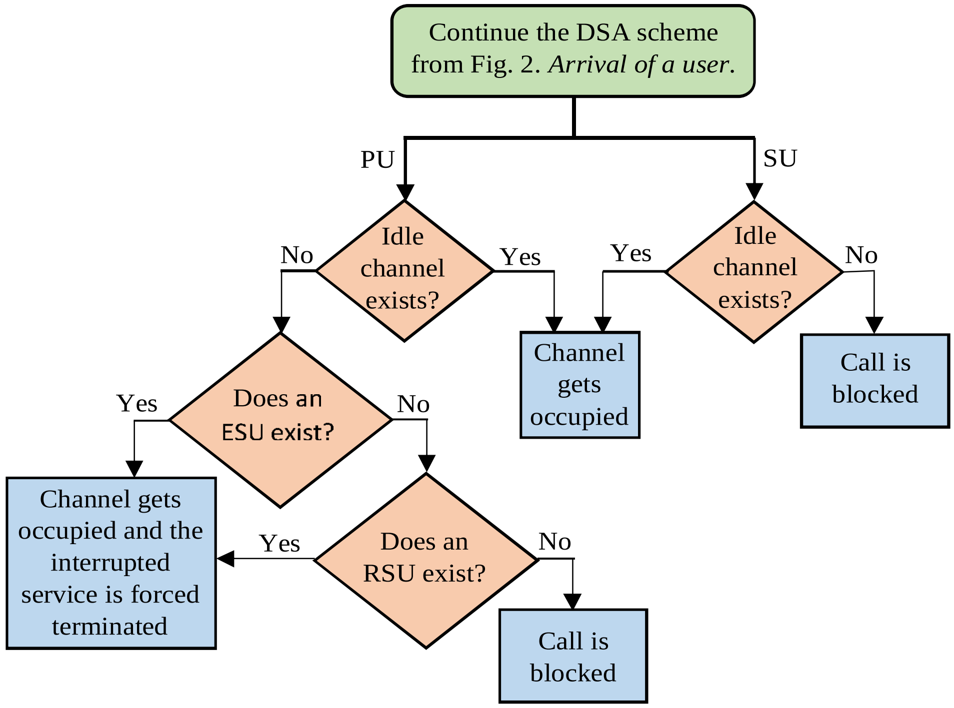

Figure 6.3: Flow chart illustration of the proposed DSA scheme in part when a user arrives.

arriving SU, if it exists. Otherwise, the newly requested service is simply blocked.

3) PU/SU Departure: As a result of a departure event of a PU or a SU service, the corresponding channel becomes idle and is considered as available for allocation to a new or interrupted service.

4) Channel Failure: When a vacant channel fails, it becomes unavailable for the next request. On the other hand, if an occupied channel fails, the interrupted user may perform a spectrum handover to another vacant channel with the lowest channel index. However, if no vacant channel is found, the interrupted user service is forcibly terminated.

5) Channel Repair: A channel is said to be repaired when a failed channel is re-established to its normal conditions and ready to serve the users. However, no spectrum handover is performed when a failed channel is repaired.

6.4 Markov Chain Modeling

An analytical model utilizing a CTMC is developed for the proposed dynamic channel access scheme. Let $\boldsymbol{z} = (z_1, z_2, \cdots, z_N)$ be the general state representation of the system where z_i denotes the status of the i^{th} channel. Based on the operation modes presented in Section 6.3.1, the status of the i^{th} channel is represented as follows:

$$z_i = \begin{cases} 0; & \text{if the } i^{th} \text{ channel is in } \textit{Mode 1}, \\ 1; & \text{if the } i^{th} \text{ channel is in } \textit{Mode 2}, \\ 2; & \text{if the } i^{th} \text{ channel is in } \textit{Mode 3}, \\ 3; & \text{if the } i^{th} \text{ channel is in } \textit{Mode 4}, \\ 4; & \text{if the } i^{th} \text{ channel is in } \textit{Mode 5}. \end{cases} \tag{6.1}$$

Accordingly, the corresponding set of feasible states of the system is denoted as

$$\begin{aligned} S = \{\boldsymbol{z} | 0 \leq z_1, z_2, \cdots, z_N \leq 4; z_{N-R_1+1}, \cdots, z_{N-1}, \\ z_N \neq 2, N - R_1 + 1 \leq N; z_{N-R_2+1}, \cdots, z_{N-1}, \\ z_N \neq 1, N - R_2 + 1 \leq N\}. \end{aligned} \tag{6.2}$$

Table 6.1 lists the state transition rates associated with different events under certain conditions. Therein, notations AR and DP indicate an arrival and a departure event, respectively. Also, e_k denotes an N element vector whose k^{th} element is equal to 1 while the other elements are equal to 0. Furthermore, the transitions and the transition rates mentioned in this table are utilized to formulate the transition rate matrix Q of the CTMC.

6.5 Performance Analysis

In this section, we perform a transient analysis of the proposed scheme and compute the performance metrics of interest.

6.5.1 Transient Analysis: Uniformization

Let $\pi_{\boldsymbol{z}}(t)$ be the probability that the system is in state $\boldsymbol{z}$ at time t. Using the generator matrix Q of the CTMC under study, the transient probability vector solution, $\Pi(t)$ constituted by the probabilities $\pi_{\boldsymbol{z}}(t)$ is again evaluated utilizing the uniformization method, as given in equation (5.3).

Table 6.1: Transitions form a generic state $\boldsymbol{z} = (z_1, z_2, \cdots, z_N)$ of the proposed DSA scheme upon PU/RSU/ESU events and channel failures and repairs.

S.No.	Activity	Dest. State	Trans. rate	Conditions
1.	PU AR. An idle channel C_i is allocated.	$\boldsymbol{z} + 3\boldsymbol{e}_i$	λ_{PU}	$\exists k : z_k = 0, k \leq N, i = min\{j \vert z_j = 0, j \leq N\}$.
2.	PU AR. No idle channels exist. An ESU service on C_i is forced terminated.	$\boldsymbol{z} + 2\boldsymbol{e}_i$	λ_{PU}	$\exists k : z_k = 1, k \leq N - R_2, z_l > 0 \ \forall \ l < N, i = min\{j \vert z_j = 1, j \leq N - R_2\}$.
3.	PU AR. No idle channels exist. An RSU service on C_i is forced terminated.	$\boldsymbol{z} + \boldsymbol{e}_i$	λ_{PU}	$\exists k : z_k = 2, k \leq N - R_1, z_l > 0 \ \forall \ l < N, z_m > 1 \ \forall \ m \leq N - R_2, i = min\{j \vert z_j = 2, j \leq N - R_1\}$.
4.	PU DP from C_i.	$\boldsymbol{z} - 3\boldsymbol{e}_i$	μ_{PU}	$z_i = 3, i = 1, 2, \cdots, N$.
5.	ESU AR. An idle channel C_i is allocated.	$\boldsymbol{z} + \boldsymbol{e}_i$	λ_{SE}	$\exists k : z_k = 0, k \leq N - R_2, i = min\{j \vert z_j = 0, j \leq N - R_2\}$.
6.	RSU AR. An idle channel C_i is allocated.	$\boldsymbol{z} + 2\boldsymbol{e}_i$	λ_{SR}	$\exists k : z_k = 0, k \leq N - R_1, i = min\{j \vert z_j = 0, j \leq N - R_1\}$.
7.	ESU DP from C_i.	$\boldsymbol{z} - \boldsymbol{e}_i$	μ_{SE}	$z_i = 1, i = 1, 2, \cdots, N - R_2$.
8.	RSU DP from C_i.	$\boldsymbol{z} - 2\boldsymbol{e}_i$	μ_{SR}	$z_i = 2, i = 1, 2, \cdots, N - R_1$.
9.	Idle channel C_i failed.	$\boldsymbol{z} + 4\boldsymbol{e}_i$	α_i	$z_i = 0, i = 1, 2, \cdots, N$.

10.	A PU occupied channel C_i failed. The interrupted PU service performs handover to another idle channel.	$\boldsymbol{z} + \boldsymbol{e}_i + 3\boldsymbol{e}_k$	α_i	$z_i = 3, i = 1, 2, \cdots, N, k = min\{j \vert z_j = 0, j < N\}$.
11.	A PU occupied channel C_i failed. The interrupted PU service is forced terminated.	$\boldsymbol{z} + \boldsymbol{e}_i$	α_i	$z_i = 3, i = 1, 2, \cdots, N, z_k > 0 \, \forall \, k < N$.
12.	An ESU occupied channel C_i failed. The interrupted ESU service performs handover to another idle channel.	$\boldsymbol{z}+3\boldsymbol{e}_i+\boldsymbol{e}_k$	α_i	$z_i = 1, i = 1, 2, \cdots, N - R_2, k = min\{j \vert z_j = 0, j < N - R_2\}$.
13.	An ESU occupied channel C_i failed. The interrupted ESU service is forced terminated.	$\boldsymbol{z} + 3\boldsymbol{e}_i$	α_i	$z_i = 1, i = 1, 2, \cdots, N-R_2, z_k > 0 \, \forall \, k < N - R_2$.
14.	An RSU occupied channel C_i failed. The interrupted RSU service performs handover to another idle channel.	$\boldsymbol{z}+2\boldsymbol{e}_i+2\boldsymbol{e}_k$	α_i	$z_i = 2, i = 1, 2, \cdots, N - R_1, k = min\{j \vert z_j = 0, j < N - R_1\}$.
15.	An ESU occupied channel C_i failed. The interrupted ESU service is forced terminated.	$\boldsymbol{z} + 2\boldsymbol{e}_i$	α_i	$z_i = 2, i = 1, 2, \cdots, N-R_1, z_k > 0 \, \forall \, k < N - R_1$.
16.	Failed channel C_i is repaired.	$\boldsymbol{z} - 4\boldsymbol{e}_i$	β_i	$z_i = 4, i = 1, 2, \cdots, N$.

6.5.2 Transient Performance Metrics

Following the definitions given in Section 5.6.2, we derive the mathematical expressions of transient performance metrics for the current model:

A. Capacity at time t

Let S_P be the set of channel occupied indexes by PU services and it is expressed as

$S_P = \{j|z_j = 3, j = 1, 2, \cdots, N\}$. Similarly, the set of RSU and ESU occupied channel indexes can be expressed as $S_R = \{j|z_j = 2, j = 1, 2, \cdots, N - R_1\}$ and $S_E = \{j|z_j = 1, j = 1, 2, \cdots, N - R_2\}$, respectively. We let $\rho_{PU}(t)$, $\rho_{RSU}(t)$, and $\rho_{ESU}(t)$ be the capacity of PU, RSU and ESU services respectively, at time t. Then

$$\rho_{PU}(t) = \sum_{\boldsymbol{z} \in S_P} n(S_P)\mu_{PU}\pi_{\boldsymbol{z}}(t), \tag{6.3}$$

$$\rho_{RSU}(t) = \sum_{\boldsymbol{z} \in S_R} n(S_R)\mu_{SR}\pi_{\boldsymbol{z}}(t), \quad \text{and} \tag{6.4}$$

$$\rho_{ESU}(t) = \sum_{\boldsymbol{z} \in S_E} n(S_E)\mu_{SE}\pi_{\boldsymbol{z}}(t), \tag{6.5}$$

where $n(A)$ denotes the cardinality of set A.

B. Channel availability at time t

Let S_1 be the set of channel available states for newly arriving PU services and it is expressed as $S_1 = S_A \cup S_B \cup S_C$, where $S_A = \{\boldsymbol{z}|\exists j : z_j = 0, j = 1, 2, \cdots, N\}$, $S_B = \{\boldsymbol{z}|\exists j : z_j = 1, j = 1, 2, \cdots, N - R_2\}$, and $S_C = \{\boldsymbol{z}|\exists j : z_j = 2, j = 1, 2, \cdots, N - R_1\}$. Similarly, the set of channel available states for newly arriving RSU and ESU services can be expressed as $S_2 = \{\boldsymbol{z}|\exists j : z_j = 0, j = 1, 2, \cdots, N - R_2\}$ and $S_3 = \{\boldsymbol{z}|\exists j : z_j = 0, j = 1, 2, \cdots, N - R_1\}$, respectively. Therefore, following the definition, the channel availability of PUs, $A_{PU}(t)$, RSUs, $A_{RSU}(t)$, and ESUs, $A_{ESU}(t)$ at time t, respectively is obtained as

$$A_{PU}(t) = \sum_{\boldsymbol{z} \in S_1} \pi_{\boldsymbol{z}}(t), \tag{6.6}$$

$$A_{RSU}(t) = \sum_{\boldsymbol{z} \in S_2} \pi_{\boldsymbol{z}}(t), \quad \text{and} \tag{6.7}$$

$$A_{ESU}(t) = \sum_{\boldsymbol{z} \in S_3} \pi_{\boldsymbol{z}}(t). \tag{6.8}$$

Accordingly, the blocking probabilities of PU, RSU, and ESU services at time t, denoted as $P^B_{PU}(t)$, $P^B_{RSU}(t)$, and $P^B_{ESU}(t)$ respectively, are expressed as

$$P^B_{PU}(t) = 1 - A_{PU}(t), \tag{6.9}$$

$$P^B_{RSU}(t) = 1 - A_{RSU}(t), \quad \text{and} \tag{6.10}$$

$$P^B_{ESU}(t) = 1 - A_{ESU}(t). \tag{6.11}$$

C. Retainability at time t

The retainability of service at time t, $\theta(t)$, is obtained as

$$\theta(t) = 1 - P_F(t), \tag{6.12}$$

where $P_F(t)$ is the forced termination probability of that service at time t.

The transient forced termination probability of RSUs, $P_F^{RSU}(t)$, can be expressed as the ratio of mean forced termination rate of RSUs at time t to the mean admitted RSU rate at time t, $\Lambda_{RSU}(t) = A_{SU}(t)\lambda_{SR}$. Correspondingly, the forced termination probability of RSUs at time t is given by

$$P_F^{RSU}(t) = \frac{\left(R_{RSU}(t) + R'_{RSU}(t)\right)}{\Lambda_{RSU}(t)}, \tag{6.13}$$

where $R_{RSU}(t)$ and $R'_{RSU}(t)$ denote the forced termination rate of SUs at time t because of PU arrivals and channel failures, respectively. Then, we have

$$R_{RSU}(t) = \lambda_{PU} \sum_{\boldsymbol{z} \in S_A} \pi_{\boldsymbol{z}}(t), \quad \text{where}$$

$$S_A = \{\boldsymbol{z} | \exists j : z_j = 2, j \leq N - R_1, z_k > 0 \,\forall\, k < N, z_l > 1 \,\forall\, l \leq N - R_2\}, \quad \text{and} \tag{6.14}$$

$$R'_{RSU}(t) = \sum_{i=1}^{N-R_1} \sum_{\boldsymbol{z} \in S_B} \alpha_i \pi_{\boldsymbol{z}}(t), \quad \text{where}$$
$$S_B = \{\boldsymbol{z} | \exists j : z_j = 2, j \leq N - R_1, z_k > 0 \,\forall\, k < N - R_1\}. \tag{6.15}$$

Likewise, the transient forced termination probability of ESUs, $P_F^{ESU}(t)$, is obtained as

$$P_F^{ESU}(t) = \frac{\Big(R_{ESU}(t) + R'_{ESU}(t)\Big)}{\Lambda_{ESU}(t)}, \tag{6.16}$$

where $\Lambda_{ESU}(t) = A_{SU}(t)\lambda_{SE}$, and $R_{ESU}(t)$ and $R'_{ESU}(t)$, respectively are computed as

$$\begin{aligned} R_{ESU}(t) &= \lambda_{PU} \sum_{\boldsymbol{z} \in S_C} \pi_{\boldsymbol{z}}(t), \quad \text{where} \\ S_C &= \{\boldsymbol{z} | \exists j : z_j = 1, j \leq N - R_2, z_k > 0 \, \forall \, k < N\}, \end{aligned} \tag{6.17}$$

$$\begin{aligned} R'_{ESU}(t) &= \sum_{i=1}^{N-R_2} \sum_{\boldsymbol{z} \in S_D} \alpha_i \pi_{\boldsymbol{z}}(t), \quad \text{where} \\ S_D &= \{\boldsymbol{z} | \exists j : z_j = 1, j \leq N - R_2, z_k > 0 \, \forall \, k < N - R_2\}. \end{aligned} \tag{6.18}$$

Along the similar lines, the transient forced termination probability of PUs due to channel failures, $P_F^{PU}(t)$, can be calculated as

$$P_F^{PU}(t) = R'_{PU}(t)/\Lambda_{PU}(t), \tag{6.19}$$

where $R'_{PU}(t)$ and $\Lambda_{PU}(t)$ are given by

$$\begin{aligned} R'_{PU}(t) &= \sum_{i=1}^{N} \sum_{\boldsymbol{z} \in S_E} \alpha_i \pi_{\boldsymbol{z}}(t), \\ S_E &= \{\boldsymbol{z} | \exists j : z_j = 3, j \leq N, z_k > 0 \, \forall \, k < N\}, \end{aligned} \tag{6.20}$$

and $\Lambda_{PU}(t) = A_{PU}(t)\lambda_{PU}$, respectively.

D. Network Unserviceable Probability at time t

The transient NUP for RSU services, $Q_{RSU}(t)$, is given by

$$\begin{aligned} Q_{RSU}(t) &= 1 - (\text{Prob. of completion of an RSU service}), \\ &= 1 - \frac{\lambda_{SR}\big(1 - P_{SU}^B(t)\big)\big(1 - P_{RSU}^F(t)\big)}{\lambda_{SR}}, \\ &= P_{SU}^B(t) + P_{RSU}^F(t) - P_{SU}^B(t) P_{RSU}^F(t). \end{aligned} \tag{6.21}$$

Similarly, the NUP for ESU and PU services at time t, denoted as $Q_{ESU}(t)$ and $Q_{PU}(t)$, respectively, is given by

$$Q_{ESU}(t) = P^B_{SU}(t) + P^F_{ESU}(t) - P^B_{SU}(t)P^F_{ESU}(t), \tag{6.22}$$

$$Q_{PU}(t) = P^B_{PU}(t) + P^F_{PU}(t) - P^B_{PU}(t)P^F_{PU}(t). \tag{6.23}$$

Table 6.2: Summary of symbols in the CTMC and value configuration

Symbol	Description	Default Value
N	Total number of channels in CRN	6
R_1	Number of channels restricted from RSU	1
R_2	Number of channels restricted from ESU	2
λ_{PU}	PU arrival rate	5 services per unit of time
$\lambda_{SR}, \lambda_{SE}$	RSU and ESU arrival rates	2.5 services per unit of time
μ_{PU}	PU service rate per channel	2 services per unit of time
μ_{SR}, μ_{SE}	RSU and ESU service rates per channel	1 service per unit of time
$\Pi(0)$	Initial Condition	$\{1, 0, 0, \cdots, 0\}$
$\alpha_i, i = 1, 2, \cdots, 6$	Failure rate of i^{th} channel	$0.10 + 0.02 \times (i-1)$ failures per unit of time
$\beta_i, i = 1, 2, \cdots, 6$	Repair rate per channel	1.0 repairs per unit of time

6.6 Numerical Results and Discussion

In this section, we report numerical results conducted to evaluate the performance of the proposed scheme. We have analyzed the main performance measures as a function of the time elapsed (in time units) since the origin. The values of the network configuration are listed in Table 6.2 unless otherwise stated. The first state of set S is assumed as $\boldsymbol{z} = (0, 0, 0, 0, 0, 0)$, i.e., a state in which all channels are available. Accordingly, $\Pi(0) = \{1, 0, 0, \cdots, 0\}$ in Table 6.2 indicates that the probability of being in state $\boldsymbol{z} = (0, 0, 0, 0, 0, 0)$ at $t = 0$ is 1. Note that two types of channel failures are considered. For the first type, FT-I, the channel failure rates are exactly the same as mentioned in Table 6.2 (α_i). For the second type, FT-II, the failure rate of each channel is twice as the corresponding rate of the FT-I, i.e. $2 \times \alpha_i$ (see, Balapuwaduge and Li [262]). Moreover, being R_1 and R_2 the number of reserved channels from each SU type (RSU and ESU), each SU type may access only $N - R_1$ and $N - R_2$ channels. Four different input configurations are computed: $R_1 = R_2 = 0$ (configuration 1), $R_1 = 0$, $R_2 = 1$ (configuration 2), $R_1 = 1$, $R_2 = 0$ (configuration 3) and $R_1 = 2$, $R_2 = 2$ (configuration 4). Note that the curves representing *No Channel Reservation* are obtained by configuring $R_1 = R_2 = 0$.

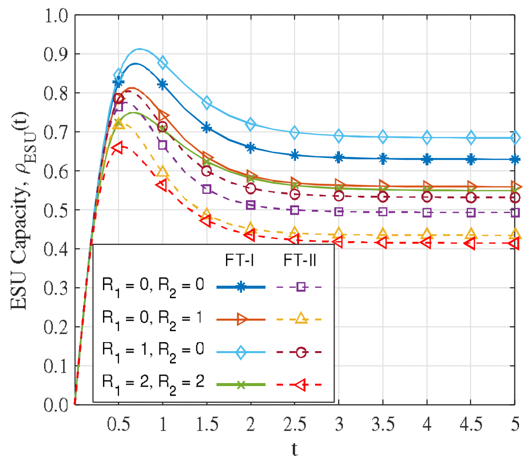

Figure 6.4: Capacity achieved by ESUs as a function of time when channel reservation is applied.

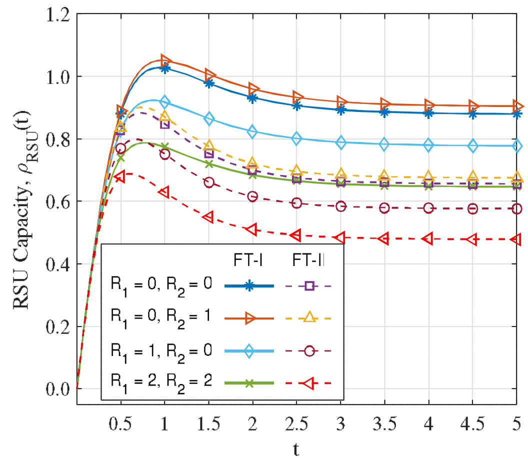

Figure 6.5: Capacity achieved by RSUs as a function of time when channel reservation is applied.

A. Capacity

Figs. 6.4 and 6.5 represent the capacity of ESU and RSU services as a function of time respectively, under different values of R_1, R_2 and α_i. From these figures, for all scenarios, we notice that the capacity of SUs initially increases with time, but eventually decreases and shows a steady behavior. As is to be expected, with higher channel failure rates, the capacity loss increases because of the lack of idle channels and the effect is noticeable for reasonably larger values of t.

Furthermore, the channel reservation scheme witnesses better performance compared with the CRN without channel reservation in terms of ESU capacity for configuration three, and in terms of RSU capacity for configuration two. This is because configuration three (two) provides fewer channels for RSU (ESU) admission, but gives full resources for the ESUs (RSUs). However, it is observed that configuration four degrades the SU

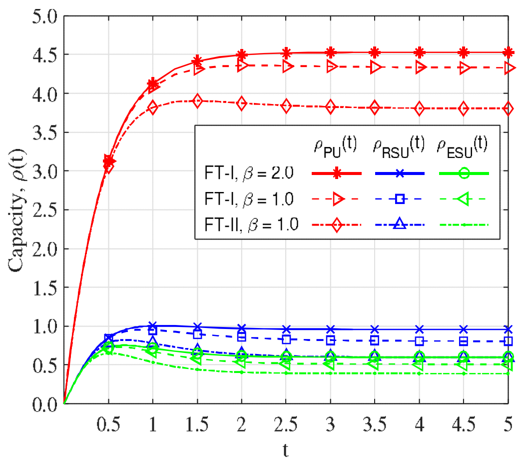

Figure 6.6: Capacity of PUs and SUs as a function of time with different failure and repair rates.

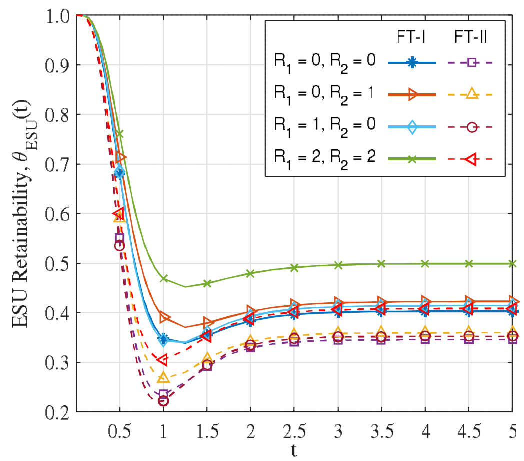

Figure 6.7: Retainability of ESUs as a function of time when channel reservation is applied.

service capacity and the degradation is proportional to the number of reserved channels since the number of channels available for new service diminishes.

Let us now observe the capacity of PU and SU services as a function of time for different values of failure and repair rates, presented in Fig. 6.6. The capacity of each service decreases with higher failure rates and lower repair rates due to decreased channel opportunities. This is because, with large α_i, channels are more likely to fail and with lower β, the failed channels remain unavailable for a longer duration of time, resulting in capacity reduction. However, the PUs' capacity is higher than that of SUs' owing to its priority privileges for channel access. Consequently, the number of idle channels in the network diminishes and thus the capacity degrades. However, the capacity of PU services is always higher than that of SU services owing to its channel access priority privileges.

B. Retainability

Let us now observe the achieved retainability by SU and PU services as the time varies, plotted in Figs. 6.7 to 6.9. As expected, these figures confirm that the achieved retainability level for each service decreases with time due to more active users in the system and then attains the steady-state. It is noted that when channel reservation is triggered, configuration four has the highest retainability level followed by configuration two (Fig. 6.7)

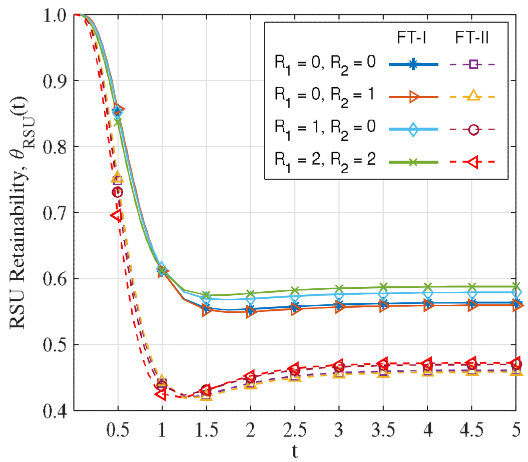

Figure 6.8: Retainability of RSUs as a function of time when channel reservation is applied.

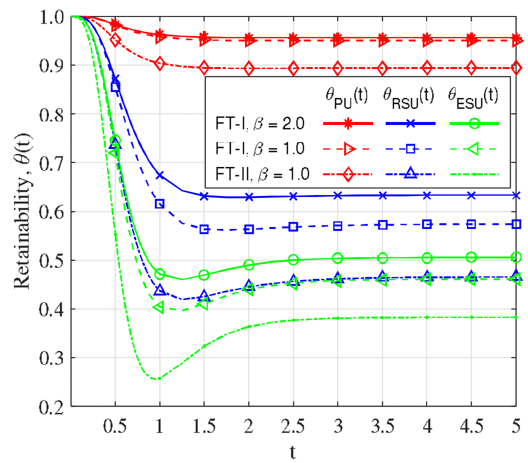

Figure 6.9: Retainability of PUs and SUs as a function of time with different failure and repair rates.

and configuration three (Fig. 6.8). The reason being that the reservation enables fewer services admitted in the CRN, which lessens their probability of being forcibly terminated. Moreover, we also note that the improvement in retainability level in contrast with the CRN without channel reservation is more prominent for ESUs since the RSU being prioritized also influences the ESU behavior.

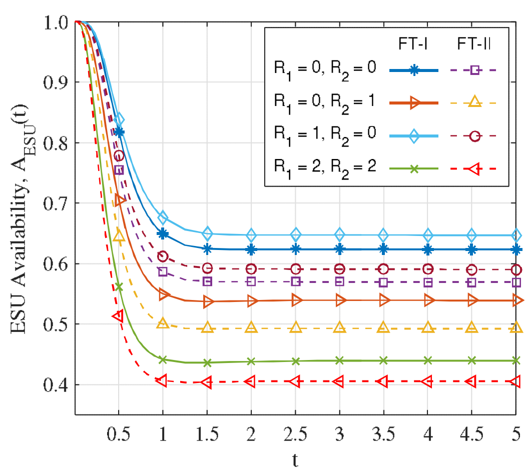

Figure 6.10: Channel availability for ESUs as a function of time when channel reservation is applied.

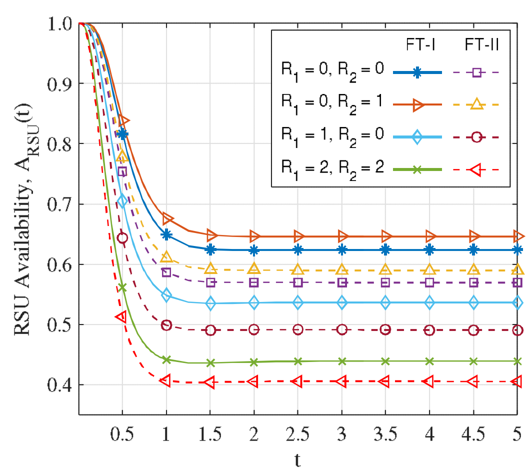

Figure 6.11: Channel availability for RSUs as a function of time when channel reservation is applied.

Furthermore, Fig. 6.9 depicts that a higher channel repair rate results in increased channel access opportunities, which in turn lessens the PU and SU services to be terminated and thereby increases the retainability level. Note that the increment is more significant for SUs as compared to PUs. On the other hand, an increase in channel failure rate degrades the retainability level for each service due to a lack of idle channels. Also, it is worth mentioning that the retainability level of PU services is always higher than that of SU services and moreover, that of RSUs outperforms ESUs' due to its priority on accessing channels.

C. Channel Availability

From the afore-discussion, we observe that the system exhibits significantly improved performance via channel reservation in terms of achieved retainability level. However, the price the system pays for channel reservation is the reduced capacity as well as the decreased channel availability. Since service interruptions lasting over long periods is not desirable, thus to maintain a higher level of service retainability is considered as more fundamental. Moreover, allowing channel reservation enables a tradeoff between the new users' channel availability and the ongoing services' retainability as illustrated in Figs. 6.10 and 6.11. For instance, it is observed that configuration three provides the third worst performance in retainability level (Fig. 6.7) which might seem to contradict at first sight. However, briefly for this experiment, configuration three seems to have the best compromise between the adopted metrics since it has the best channel availability (Fig. 6.10) and a reasonable retainability level along with the lowest NUP for ESUs, as can be further observed (Fig. 6.13). On the other hand, again for very small values of t, channel reservation seems not to be worthwhile.

Fig. 6.12 confirms that the channel availability for each service decreases with higher channel failure rates and lower channel repair rates, which agrees with our expectations. In conclusion, these results reveal a need for an assessment of the tradeoff that can be tuned according to the QoS requirements of the secondary traffic.

D. Network Unserviceable Probability (NUP)

In Figs. 6.13 and 6.14, we evaluate the NUP of ESU and RSU services respectively con-

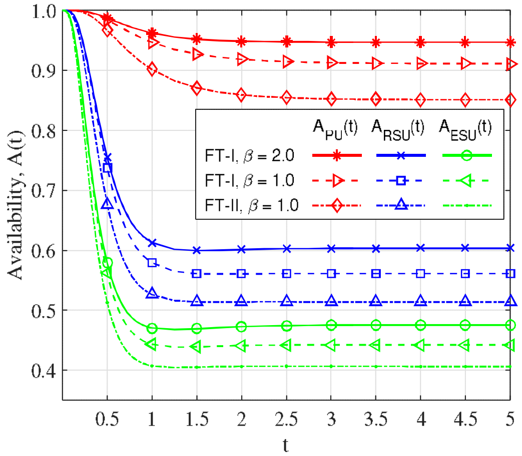

Figure 6.12: Channel availability for PUs and SUs as a function of time with different failure and repair rates.

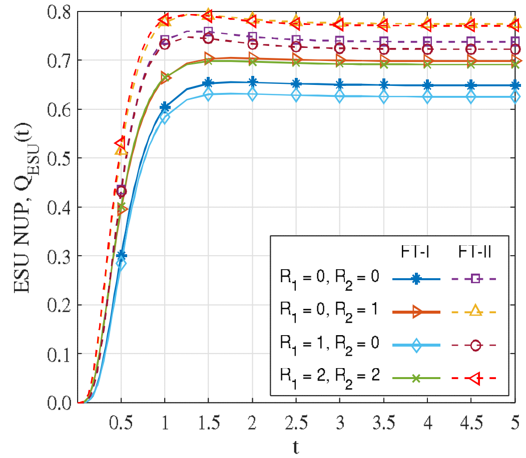

Figure 6.13: Network unserviceable probability of ESUs as a function of time when channel reservation is applied.

sidering different configurations. As expected, NUP of both services first increases with an increase in time and then shows the steady-state behavior. The results presented here indicate that, under higher values of t, for a particular channel failure type, channel reservation with configuration three (two) provides the best performance for ESUs (RSUs) in contrast with the CRN without channel reservations and is more prominent for ESU services. However, as can be observed, for smaller time points, allowing channel reservation may not help to improve the NUP. In brief, it can be concluded that the proposed reservation scheme helps to reduce the NUP in a CRN when t is higher.

Another important observation regarding the NUP performance of PUs and SUs can be found in Fig. 6.15. It can be seen that the NUP of each service is significantly influenced by the channel repair rate and highly depends on the failure type. With a higher failure rate, the number of available channels in the network decreases, leading to low NUP. Contrarily, the figure depicts that the NUP can be improved with a higher β. When channel repair rate is raised, the failed channels are repaired shortly and thereby the services can find more channel access opportunities, resulting in higher NUP. Fig. 6.15 illustrates further that the PUs' NUP is lowest followed by that of the SUs' and this is due to the channel access priority privileges of each user type.

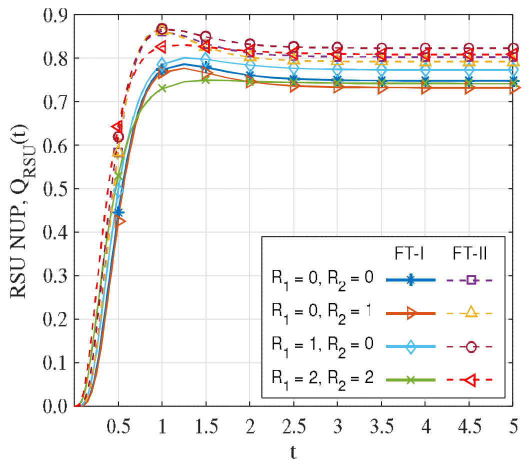

Figure 6.14: Network unserviceable probability of RSUs as a function of time when channel reservation is applied.

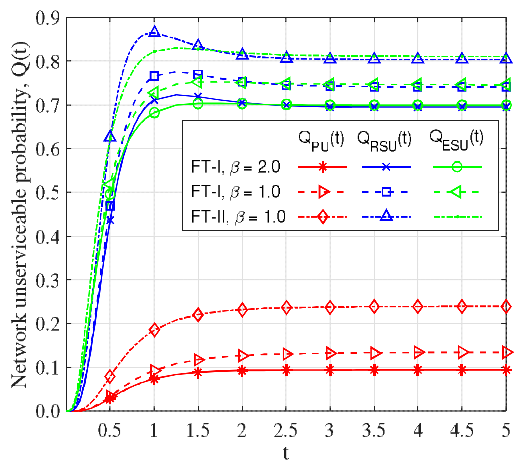

Figure 6.15: Network unserviceable probability of PUs and SUs as a function of time with different failure and repair rates.

6.7 Chapter Summary

This chapter have performed the transient analysis of CRNs from the perspectives of dependability theory addressing how to achieve ultra-reliable communication in 5G and beyond networks. We have proposed and analyzed a DSA scheme with multi-level channel reservation which enhances the heterogeneous SUs' performance in CRNs prone to random channel failures, that can help to overcome various challenges such as network dimensioning and secondary QoS guarantees. Furthermore, the mathematical expressions of transient performance metrics for channel access in CRNs are derived and a CTMC model has been developed to evaluate the performance of the proposed scheme. The obtained numerical results have revealed that the channel reservation not always provide reasonable performance trade-offs as most works suggested, contrarily, its success highly depends on the network's state, i.e. the PU/SU traffic load and time points. In addition, the channel failure rate and recovery rate also have a significant influence on the availability and reliability provided to end users.

Chapter 7

A $MAP/PH/1$ Queue with Working Breakdowns and Working Vacations

"Don't read success stories, you will only get a message. Read failure stories, you will get some ideas to get success."
— A. P. J. Abdul Kalam

7.1 Introduction

In this chapter, we relax the assumption of exponentially distributed inter-arrival times and service times, and deal with a very generic $MAP/PH/1$ queueing system with working breakdowns and working vacations. Such system can be used for modeling wireless cellular networks. Although these much more general assumptions with respect to the nature of the input flow and service time distribution complicate the study of system, they allow us to capture correlation and explosive nature of traffic in 5G and beyond networks. Herein, the server (base station) (i) takes vacations (whenever the system becomes empty), (ii) breaks down or fails (due to external shocks) and (iii) gets repaired.

This work has appeared in **Srinivas R. Chakravarthy, Shruti and R. Kulshrestha, Operations Research Perspectives, Elsevier, 7, 100131 (2020) 1-13.** (https://doi.org/10.1016/j.orp.2019.100131)

The backup server takes over serving the customers (calls) at a reduced rate whenever the main server is either on vacation or under repair. Upon the availability of the main server (either through returning from a vacation or a repair completion), the existing service rate will be set back to the normal rate. This model generalizes the recently studied one by Ye and Liu [157], wherein $MAP/M/1$ queue is analyzed with working breakdowns. The use of phase type (PH$-$) distributions enables one to model real service processes well with a variety of distributions for the services. For example, as discussed in Chapter 2, some classical distributions such as generalized Erlang and finite mixtures of exponentials among others are very special cases of PH$-$distributions. The durations of vacations are modeled with exponential distribution. It is possible for the (main) server to take multiple vacations due to no customers seen waiting at the time of the server returning from a vacation. The effect of the variability of service times in the context of vacation/working vacation models has not been fully explored and we focus on this aspect also in this chapter. Thus, the model studied in our work is very generic and can be applied in communication networks.

Our working-vacation-breakdown-repair queueing model in this chapter is applicable to cellular networks, as we know that in the cellular network each cell has a base station that controls the call admissions and the quality of service of the network. If we want to model the base station properly and adequately, then we should consider the possibility of many users (customers) accessing the Internet on their mobiles at the same time. This creates traffic in the network to be bursty at times. Self-similar and bursty nature of the incoming traffic causes correlation in inter-arrival times of the incoming traffic. Therefore, as discussed earlier, the use of MAP to model such arrivals makes sense as it allows for the correlation. The services provided by the base station controller can be modeled as $PH-$services. This choice is due to the fact that when the caller requests for a connection, the base station controller needs to (a) verify the caller; (b) check the current location of the caller; (c) search for a route to be assigned; and after these phases, the connection is made. Like any other electronic component, the base station controller is also exposed to risks due to external shocks, and therefore subject to breakdowns.

But at the same time, the services to mobile users are very important. Hence, the service providers cannot afford full interruptions in their services leading to backup servers being relied upon to provide services at reduced rates whenever the main sever is under repair. This is referred as working breakdowns. The base stations consume a large amount of energy for serving their customers. In order to save a significant amount of energy, a base station can go on a vacation (i.e, a sleep mode of the base station) when there is no call waiting in the network. In such situations, a backup server needs to be initiated (possibly serving at reduced rates) to serve arriving new calls and this is known as working vacations. While there are many variants of the arrival/service systems studied thus far, to the best of our knowledge, models using the Markovian arrivals and phase type services with working vacations and working breakdowns are studied for the first time in this thesis. Accordingly, the proposed model generalizes some of the previously published ones on working-vacation-breakdown-repair queues.

The layout of the chapter is described now. The model considered here is described in full detail in Section 7.2. The analysis in steady-state of the proposed model is presented in Section 7.3. The effective service time of a customer is obtained in Section 7.4. A discussion of a few other performance measures is presented in Section 7.5 and Section 7.6 deals with some special cases. Section 7.7 is devoted to decomposition results for the case of the $M/M/1$ model and are shown to be in agreement with the previously published ones. Some illustrative examples to point out the effects of having a backup server are presented in Section 7.8 and a few concluding remarks are given in Section 7.9.

For the rest of the chapter, we adopt the following standard notation.

- The symbol $'$ stands for the transpose notation.
- As discussed in Chapter 2, the notations $\otimes$ and $\oplus$ stand for the Kronecker product and Kronecker sum, respectively. For more details on these, one may refer to the standard books on matrix algebra by Marcus and Minc [263], and Steeb and Hardy [264].
- $\boldsymbol{e}' = (1, 1, \cdots, 1)$, whose dimension should be clear in the context. Where more

clarity is needed, the dimension will be mentioned, e.g., $\boldsymbol{e}(m)$ is a column vector of 1's of dimension m.

- $\boldsymbol{e}_i^{'} = (0, 0, \cdots, 1, 0, \cdots, 0)$, where 1 is in the i^{th} position.

- I denotes an identity matrix, whose dimension is dictated by the context.

7.2 Model Description

The model considered in this chapter is outlined in this section. The customers enter singly into the system according to Neuts' [176] versatile point process, namely MAP, which is described by two matrices, D_0 and D_1. Thus, (D_0, D_1) of order m represents the MAP to describe the arrivals. While the matrix D_0 is to model the transitions pointing to no arrivals, the transitions within D_1 point to the arrivals. Assume $D = D_0 + D_1$ to be an irreducible matrix with an stationary vector $\boldsymbol{\delta}$. That is,

$$\boldsymbol{\delta} D = \boldsymbol{0}, \boldsymbol{\delta}\boldsymbol{e} = 1. \tag{7.1}$$

The average rate of arrivals is $\lambda = \boldsymbol{\delta} D_1 \boldsymbol{e}$. The system has a single server, which will be referred to as the main server. In the sequel, we interchangeably use the terms server and main server to refer to this server. The server is exposed to external shocks that occur according to a Poisson process with rate γ. These shocks will have an effect on the server only when the server is busy serving a customer. When the shock has an effect, the server breaks down and is sent for repair immediately, and at that moment the backup server instantaneously takes over serving the customer but at a reduced rate. The repair times are exponentially distributed with rate ξ. Upon finding the system to be empty, the main server goes on a vacation. It is possible for the server to go on multiple vacations. The duration of each vacation is modeled with an exponential distribution with rate η.

Note that the backup server is not subject to breakdowns. The backup server is available on a standby basis to take over serving the customers whenever the main server is on a vacation or under repair. Upon completion of a repair or a vacation, the main server either takes over the service from the backup server instantaneously or goes on a vacation due to the system being empty.

The service times of the main server as well as the backup server are modeled using PH−distributions with representations, respectively, $(\boldsymbol{\beta}, S)$ and $(\boldsymbol{\beta}, \theta S)$, of order n, where $0 < \theta < 1$. Note that the service rates of the main and the backup servers are given by μ and $\theta\mu$, respectively, where $\mu = [\boldsymbol{\beta}(-S)^{-1}\boldsymbol{e}]^{-1}$. Moreover, we denote $\boldsymbol{S}^0$ as the column vector satisfying $S\boldsymbol{e} + \boldsymbol{S}^0 = \mathbf{0}$. The underlying random variables governing the durations of the inter-arrivals, the services, the vacations, the repairs, and the shocks are all assumed to be mutually independent of each other.

7.2.1 The Markov Process

A number of variables needed to describe the model under study are defined now. Let

- $N(t)$: number of customers in the system;
- $J_1(t)$: server status;
- $J_2(t)$: service phase, if any;
- $J_3(t)$: arrival phase;

at time t, where

$$J_1(t) = \begin{cases} 0; & \text{if main server is busy,} \\ 1; & \text{if main server is vacationing,} \\ 2; & \text{if main server is under repair.} \end{cases}$$

It is easy to verify that the process $\{N(t), J_1(t), J_2(t), J_3(t) \ : \ t \ \geq \ 0\}$ is a Markov process possessing the QBD-structure on the state space $\Omega = \bigcup\limits_{i=0}^{\infty} r(i)$, where $r(0) = \{(0, j_1, j_3), 1 \leq j_1 \leq 2, 1 \leq j_3 \leq m\}$ and $r(i) = \{(i, j_1, j_2, j_3) : j_1 = 0, 1 \text{ or } 2, 1 \leq j_2 \leq n, 1 \leq j_3 \leq m\}, \ i \geq 1$. The model described above is studied as a quasi-birth-and-death process (QBD).

By level $\underline{\boldsymbol{i}}$, for $i > 0$, we denote the set of $3mn$ states given by $\underline{\boldsymbol{i}} = \{(i, j_1, j_2, j_3) : j_1 = 0, 1 \text{ or } 2, 1 \leq j_2 \leq n, 1 \leq j_3 \leq m\}$. Note that when the system is empty (i.e., when

$N(t) = 0$), $J_2(t)$ is undefined. A pictorial description of the transition diagram of the system is shown in Fig. 7.1.

The $QBD-$process has the (infinitesimal) generator, Q, given by

$$Q = \begin{pmatrix} B_1 & B_0 & & & & \\ B_2 & A_1 & A_0 & & & \\ & A_2 & A_1 & A_0 & & \\ & & A_2 & A_1 & A_0 & \\ & & & \ddots & \ddots & \ddots \end{pmatrix}, \tag{7.2}$$

where the (block) entries in Q are as follows.

$$B_1 = \begin{bmatrix} D_0 & 0 \\ \xi I_m & D_0 - \xi I_m \end{bmatrix}, \quad B_0 = \begin{bmatrix} 0 & \boldsymbol{\beta} \otimes D_1 & 0 \\ 0 & 0 & \boldsymbol{\beta} \otimes D_1 \end{bmatrix}, \tag{7.3}$$

$$B_2 = \begin{bmatrix} \boldsymbol{S}^0 \otimes I_m & 0 \\ \theta \boldsymbol{S}^0 \otimes I_m & 0 \\ 0 & \theta \boldsymbol{S}^0 \otimes I_m \end{bmatrix}, \quad A_2 = \begin{bmatrix} \boldsymbol{S}^0 \boldsymbol{\beta} \otimes I_m & 0 & 0 \\ 0 & \theta \boldsymbol{S}^0 \boldsymbol{\beta} \otimes I_m & 0 \\ 0 & 0 & \theta \boldsymbol{S}^0 \boldsymbol{\beta} \otimes I_m \end{bmatrix}, \tag{7.4}$$

$$A_1 = \begin{bmatrix} S \oplus D_0 - \gamma I_{mn} & 0 & \gamma I_{mn} \\ \eta I_{mn} & \theta S \oplus D_0 - \eta I_{mn} & 0 \\ \xi I_{mn} & 0 & \theta S \oplus D_0 - \xi I_{mn} \end{bmatrix}, \quad A_0 = I_{3n} \otimes D_1, \tag{7.5}$$

where B_1 is a square matrix of dimension $2m$; B_0 and B_2 are, respectively, rectangular matrices of dimension $2m \times 3mn$ and $3mn \times 2m$; A_0, A_1 and A_2 have dimension $3mn$.

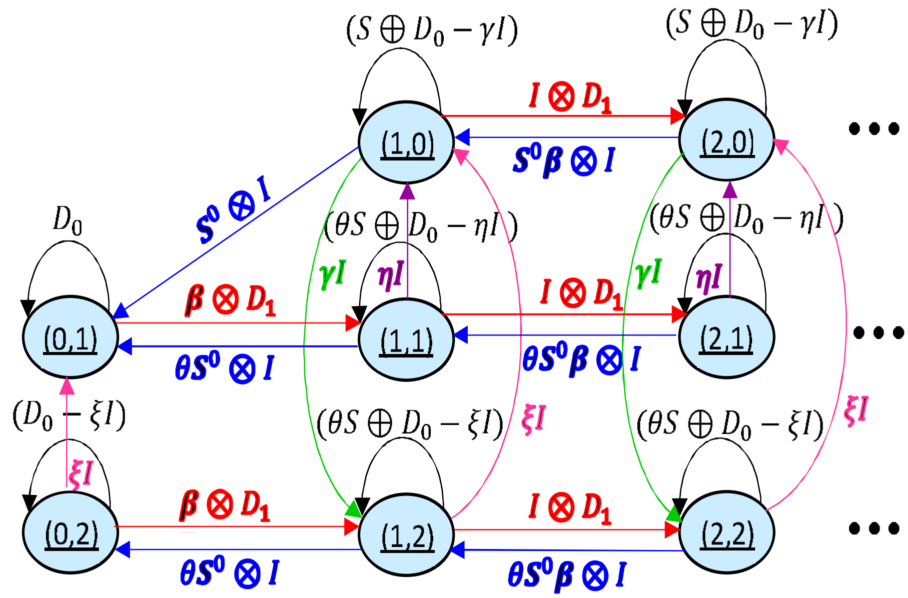

Figure 7.1: State transition diagram of the system.

Lemma 1: The stationary probability vector, $\boldsymbol{\pi}$ partitioned as $\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2)$, of the (reducible) generator $A = A_0 + A_1 + A_2$ is given by

$$\boldsymbol{\pi}_0 = \frac{\xi\mu}{\xi+\gamma}(\boldsymbol{\beta}(-S)^{-1} \otimes \boldsymbol{\delta}), \quad \boldsymbol{\pi}_1 = 0, \quad \boldsymbol{\pi}_2 = \frac{\gamma\mu}{\xi+\gamma}(\boldsymbol{\beta}(-S)^{-1} \otimes \boldsymbol{\delta}), \tag{7.6}$$

where $\boldsymbol{\delta}$ is the stationary vector of D given in (7.1).

Proof: First note that the matrix A is given by

$$A = \begin{bmatrix} (S + \boldsymbol{S}^0\boldsymbol{\beta}) \oplus D - \gamma I_{mn} & 0 & \gamma I_{mn} \\ \eta I_{mn} & (\theta[S + \boldsymbol{S}^0\boldsymbol{\beta}]) \oplus D - \eta I_{mn} & 0 \\ \xi I_{mn} & 0 & (\theta[S + \boldsymbol{S}^0\boldsymbol{\beta}]) \oplus D - \xi I_{mn} \end{bmatrix}. \tag{7.7}$$