An Optimal methodology for efficient resource distribution among contending Sims in a single RF subsystem multi-sim UE

Ruchi Kansara[*1], Jitendra Otwani[†2], Niladri Shekhar Paria[‡3]and Sajal Das[§4]

Intel Technologies, Bangalore, 560103, India

Abstract

With the evolution of wireless broadband technology and increasing demand of multi-sim User Equipment (UE), new challenges of resource sharing arise where conventional methods of creating small gaps in resource usage pattern of one SIM do not suffice. The advent of Voice over LTE (VoLTE) services accompanied by the immense mobile broadband demand of the users, require continuous resource availability on both the SIMs even in a single RF subsystem. Dual SIM Dual Active (DSDA) architecture that can meet the above requirements is not popular due to higher associated cost. In this paper, we consider the fundamental problem of resource sharing across SIMs in multi-sim architecture especially Single Receive Dual SIM Dual Standby (SR-DSDS) and Dual Receive Dual SIM Dual Standby (DR-DSDS). We formulate resource sharing as an optimization problem to maximize the ratio of resource allocation fairly for each contending SIM considering several important factors like current buffer occupancy, average time criticality of the buffer content and channel quality of respective SIM. We solve the formulated optimization problem using Karush Kuhn Tucker (KKT) conditions to derive the closed-form expressions for optimal resource allocation. We also compute the number of transitions possible with the derived optimal fair allocation in a practical multi-sim architecture. Additionally, we present the analytical results to depict the efficacy of the algorithm.

1. Introduction

With the increased penetration of smartphone and wireless technology advancement, there is a tremendous increase in the demand of user data and voice traffic. Also, the improvement in capability of UE to house two subscriber identity modules (SIMs) give rise to a new challenge where modem designers have to design algorithms to allocate fair share of resources to both the contending SIMs in a multi-sim architecture. Impact of MSIM architecture on a radio access network is discussed in [1].

Today there are various multi-sim architecture variants available in market. Among these, SR-DSDS is the most popular one as it has only 1 RF transceiver which makes it the most cost effective architecture. The conventional SR-DSDS architecture solves the challenge of RF sharing, when both the contending SIMs request for RF resource, by static priority based scheduling algorithms. This scheduling algorithm essentially creates a small gap in RF usage pattern for one SIM, when the other contending SIM resource request priority is higher. In [2], authors present a dynamic scheduling algorithm that weighs the procedures dynamically to resolve the resource conflicts.

Static and dynamic scheduling algorithms are effective only when one of the contending SIM is in idle discontinuous reception (DRX) mode i.e. need RF resource only for a short interval of time. Therefore, current DSDS equipment reduces the capability of UE by data preference on only one of the SIM and hence

---

[*]R. Kansara is currently associated with Samsung Research India, Bangalore (Email: kruchi19@gmail.com)

[†]J. Otwani is currently associated with Intel Technologies, Bangalore (Email: gotwani@gmail.com)

[‡]N. Paria is currently associated with NXP Semiconductors, Austria (Email: niltanu@gmail.com)

[§]S. Das is currently associated with Intel Technologies, Bangalore (Email: sajal_das@yahoo.com)

the other SIM operates in idle mode. When the idle SIM receives a call, data traffic on the data preferred SIM observes large interruptions which diminishes the end user experience. This challenge was addressed partially in [3] by re-establishing the mobile data connection on the idle SIM when it receives a voice call on idle SIM.

With the advancement in technology, various new challenges are arising where both the contending SIMs need RF resource continuously. One such example is when a VoLTE call is established on one SIM and a video download is ongoing on the other SIM. An optimal solution for such a challenge is not addressed in literature yet. In this paper, we try to address this problem by developing a generic scheduling algorithm which provides optimal fair share among the contending SIMs instead of static or dynamic priority based scheduling which are sub-optimal in nature. The proposed algorithm takes various quality of service parameters as an input and generates the optimal allocation percentage for each contending SIM.

This paper is organized as follows. In section 2, system model is presented. In section 3, we present the proposed unique RF resource sharing algorithm along with mathematical analysis for obtaining the closed form expression for optimal allocation. In section 4, the feasibility of the theoretical results are derived by introducing the time latency involved in each RF transition between the contending SIMs. Section 5, shows analytical results of proposed resource sharing algorithm and Section 6 contains concluding remarks.

2. System Model

In this paper we propose a unique resource sharing algorithm for DSDS UE, which effectively share the available RF resource that improves UE performance and end-user experience. The proposed algorithm shares either the single transmission path in DR-DSDS device or the RF resource in SR-DSDS device considering several factors which affects the end user experience. These quality of service parameters serve as an input for the scheduling algorithm to generate the optimal fair share among contenders. Following quality of service parameters are considered,

(i) Data Buffer status of the corresponding SIM,

(ii) Delay or Latency requirements associated with procedure,

(iii) Link quality between UE and network

Fig 1 shows closed loop system of proposed MSIM architecture along with resource sharing algorithm. Based on above quality of service factors as an input, the scheduling algorithm computes optimal resource allocation which is further discussed in Section 3. The optimal resource allocation ratios are then fed to

feasibility controller that considers practical latency constraints required in RF switching to compute maximum supported number of transitions. Further details about the mathematical model used to compute the maximum possible transitions is explained in detail in Section 4. Then the results from scheduling algorithm and feasibility controller are fed to slot mask mapper which generates the bit mask that depicts the time allocation to utilize the common RF resource between the contending SIMs.

3. Scheduling Algorithm

Let us assume, n is the total number of SIMs supported by UE and the subscript $i \in \{1, 2, \ldots, n\}$ represents i^{th} SIM. B_i is the current buffer occupancy corresponding to i^{th} SIM and $B_i \in \{0, 1, \ldots, B_{max}\}$. T_i is the average time critical factor associated with the data present in buffer corresponding to i^{th} SIM where $T_i \in \{1, 2, \ldots, T_{max}\}$. CQI_i is the channel quality indicator corresponding to i^{th} SIM such that $CQI_i \in \{1, 2, \ldots, CQI_{max}\}$. $0 \leq W_i \leq 1$ is the optimal percentage of the frame allocation to i^{th} SIM. $Mask_i$ is the optimal sub-frame mask generated based on W_i and N, where N is maximum number of supported transitions computed by feasibility controller.

As the algorithm is executed at the UE, CQI is readily available and defines the link quality between UE and network. T represents the average latency requirements of the data traffic. This can be derived at UE based on the content generating application running at UE. It naturally motivates to allocate more resource to time critical applications as it leads to better end user experience. B stands for the amount of buffer occupied at the transmitter. In the downlink, buffer occupancy is linearly proportional to the amount of downlink shared channel (Physical Downlink Shared Channel, PDSCH) allocation by network for the UE. While in uplink direction, this is readily available at UE's Medium Access Control (MAC) layer. Intuitively more data traffic should lead to more resource allocations and hence B is considered as one of the input parameter to the scheduling algorithm.

Scheduling algorithm aims to identify optimal W_i that provides optimal fair share of the RF resource among contending SIMs. For the sake of simplicity, let us limit the mathematical model to $n = 2$ i.e. SR-DSDS assumption although the results obtained are equally applicable to any generic value of n. Optimal allocation W_i such that $i \in [1, 2]$ can be computed by solving the below convex problem [4] if $B_i \neq 0 \ \forall i$ (as contention resolution should happen only if both SIM buffers are occupied),
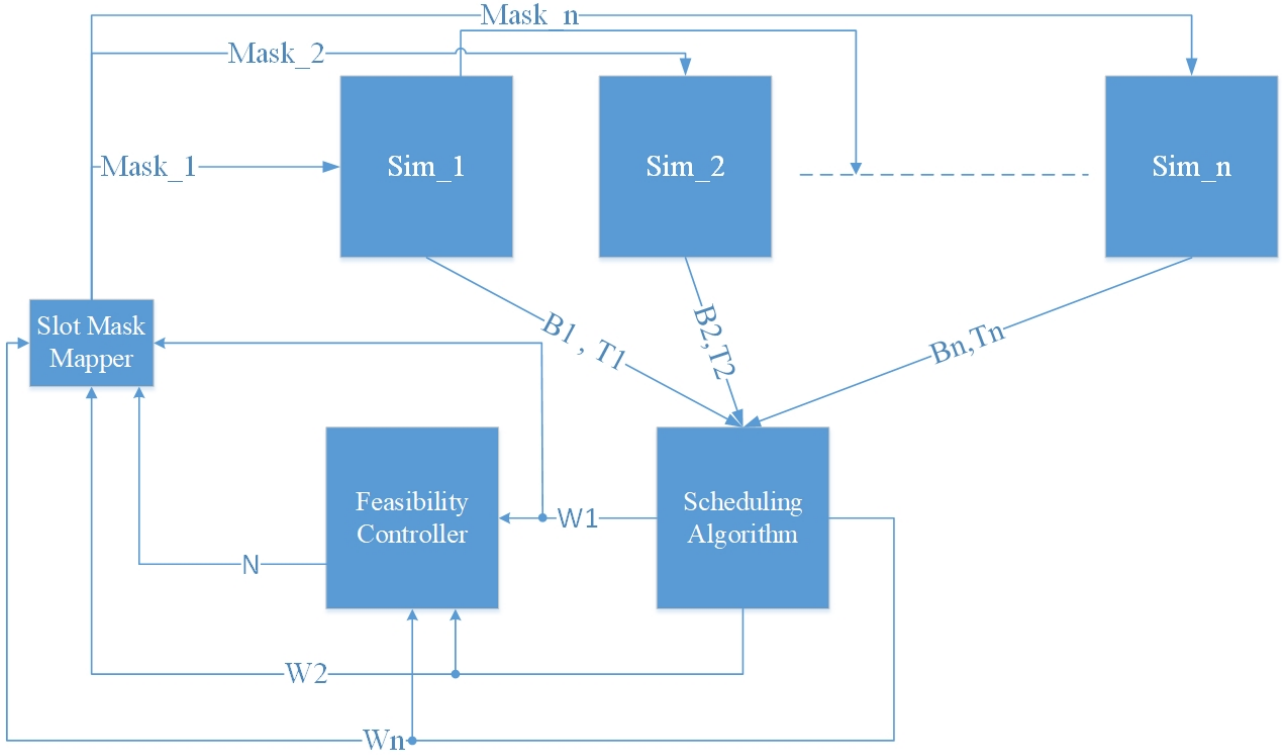
Figure 1. Proposed MSIM Architecture Model

maximize $\quad W_1 + W_2 \qquad\qquad$ (1a)
W_1, W_2

subject to

$(C1): \dfrac{W_1}{CQI_1 B_1 T_1} - \dfrac{W_2}{CQI_2 B_2 T_2} = 0, \qquad$ (1b)

$(C2): W_1 + W_2 \leq 1, \qquad\qquad$ (1c)

$(C3): W_1 \geq 0, \qquad\qquad$ (1d)

$(C4): W_2 \geq 0., \qquad\qquad$ (1e)

$(C5): B_1, B_2 \neq 0. \qquad\qquad$ (1f)

The granularity over which this problem needs to be solved is the maximum of the Transmission Time Interval (TTI) of the associated procedure executed on the contending SIMs. To generalize, we consider this granularity to be of unit time. Objective function in (1) tries to maximize the sum resource allocation. As higher values of CQI, B and T indicate more occupied time critical buffer with good link, equation (1b) schedule larger proportion of frame W_i for the SIM with more occupied time critical buffer. Larger $CQI_i B_i T_i$ value will lead to smaller $\left[\frac{1}{CQI_i B_i T_i}\right]$ which eventually results in higher W_i. Further, equation (1c) makes sure atmost unit time is allocated collectively among the contending SIMs. Moreover, lower bound on the resource allocation for any contender can be 0 which is captured in

equations (1d) and (1e). Also, we should perform contention resolution only if both the contending SIMs have content in buffer which is incorporated in equation (1f). To solve the optimization problem, we use Karush Kuhn Tucker (KKT) conditions to compute optimal dynamic allocation ratio. Lagrangian for the optimization problem in (1) can be given as,

$$\mathcal{L}(W_1, W_2; \mu, \lambda) = W_1 + W_2 - \lambda(W_1 + W_2 - 1)$$
$$- \mu\left(\frac{W_1}{CQI_1 B_1 T_1} - \frac{W_2}{CQI_2 B_2 T_2}\right). \qquad (2)$$

For optimality we set $\nabla\mathcal{L}(W_1, W_2; \mu, \lambda) = 0$. Hence, optimality conditions can be further solved as in (3) and (4),

$$\frac{\partial\mathcal{L}(W_1, W_2; \mu, \lambda)}{\partial W_1} = 0$$
$$\Rightarrow 1 - \lambda - \frac{\mu}{CQI_1 B_1 T_1} = 0. \qquad (3)$$

$$\frac{\partial\mathcal{L}(W_1, W_2; \mu, \lambda)}{\partial W_2} = 0.$$
$$\Rightarrow 1 - \lambda - \frac{\mu}{CQI_2 B_2 T_2} = 0. \qquad (4)$$

Moreover, complementary slackness condition can be given as,

$$\lambda(W_1 + W_2 - 1) = 0. \qquad (5)$$

We obtain closed form expression of optimal resource allocation W_i for all possible cases of μ and λ,

Case 1: $\mu = 0$, $\lambda = 0$, Substituting this in equation (3) and (4) yields impossible conditions and hence is an infeasible case.

Case 2: $\mu = 0$, $\lambda \neq 0$, Incorporating this in equation (3) and (4) yields $\lambda = 1$. Using this result with equation (5) leads to following,

$$W_1 = 1 - W_2 \qquad (6)$$

From (1b) and (6)

$$W_1^{\star} = \frac{CQI_1 B_1 T_1}{CQI_1 B_1 T_1 + CQI_2 B_2 T_2} \qquad (7)$$

$$W_2^{\star} = \frac{CQI_2 B_2 T_2}{CQI_1 B_1 T_1 + CQI_2 B_2 T_2} \qquad (8)$$

Case 3: $\mu \neq 0$, $\lambda = 0$, Substituting this in equation (3) and (4) we get,

$$\mu = CQI_1 B_1 T_1 = CQI_2 B_2 T_2 \qquad (9)$$

From (1b) and (9)

$$W_1^{\star} = W_2^{\star} \qquad (10)$$

Case 4: $\mu \neq 0$, $\lambda \neq 0$, Incorporating this in equation (3) we get,

$$\lambda = 1 - \frac{\mu}{CQI_1 B_1 T_1} \qquad (11)$$

Using (11) and (4) along with $\mu \neq 0$, we get,

$$CQI_1 B_1 T_1 = CQI_2 B_2 T_2 \qquad (12)$$

From (12) and (1b), it can be deduced that,

$$W_1^{\star} = W_2^{\star} \qquad (13)$$

Additionally, using equations (13), (5) and $\lambda \neq 0$ we get,

$$W_1^{\star} = W_2^{\star} = 0.5 \qquad (14)$$

Case 1 results in invalid solution, Case 3 and 4 yields a deterministic fair allocation among SIMs that is independent of CQI, Buffer Occupancy and Time criticality parameters. Only Case 2 yields an optimal dynamic allocation that takes into account various important quality parameters like CQI, Buffer occupancy and Time criticality of the buffer content. Eventually all three cases yields the maximization of objective function but there is only one result which is dynamically choosing allocation based on various factors of interest. Hence, we consider case 2 results as the closed form optimal allocation expression for $W_1^{\star}$ and $W_2^{\star}$. The closed form expression dynamically generates larger allocation to the more occupied time critical buffer with better link quality SIM contending for the resource.

Now, we need to figure out whether contiguous sub-frames allocation will be optimal or a certain pattern will generate optimal outcome. $W_i^{\star}$ does not consider the practical RF outage latencies involved in switching between the contending SIMs. Hence, in the subsequent section we try to address this and compute the maximum transitions possible considering RF switch time latencies.

4. Feasibility Controller

For each transition of operation between contending SIMs, RF needs to re-tune which results into Tx/Rx blanking period. During this period, I/Q samples become invalid resulting in only fraction of total time utilized for actual transmission and reception. As the number of transitions increase, UE resource utilization decreases, thus it should be considered as a factor for generating optimal mapping pattern $Mask_i$ by slot mask mapper. In this section, we try to compute the maximum possible transitions N given $W_i^{\star}$ is known.

Let us assume,

(i) τ is the transition delay while switching between the contending SIMs, it is system specific and is fixed based on the UE design.

(ii) F is the total time duration over which resource allocation W_i is computed (in section 3 unit time is considered for the sake of simplicity). In other words, F is a number multiple of the longest of TTI associated with contending SIMs.

(iii) F_i is the time allocation corresponding to W_i associated with SIM i such that $i \in \{1, 2, \ldots, n\}$ and $F_i = [1, F)$.

(iv) θ is the threshold indicating minimum fraction of time F which makes resource allocation feasible for a practical system. θ is also a system specific fixed parameter depending on usecase, e.g. F is set such that $F \times \theta$ is multiples of TTI.

(v) N is maximum possible number of transitions.

(vi) N_i is the number of transitions corresponding to SIM i such that $i \in \{1, 2, \ldots, n\}$ and for a practical system $N_i \geq 1$.

In this section we identify maximum possible transitions N and provide to the slot mask mapper along with frame allocation W_i, as depicted in Fig. 1. F_i for a system can be given by,

$$F_i = F W_i - N_i \tau \qquad (15)$$

For a practical system F_i is bounded as,

$$F_i > \theta F \quad (16)$$

Substituting F_i from (15) into (16) gives an upper bound on N_i as follows,

$$N_i < \frac{FW_i - \theta F}{\tau} \quad (17)$$

Equation (15) depicts F_i for SIM i reduces from FW_i due to transition delay τ for each transitions. To determine feasibility of the system F_i must be greater than certain threshold θ of total time F as given in (16).

Maximum possible number of transitions N is limited by the minimum of N_i as i^{th} SIM could only support atmost N_i transitions, given in (18).

$$N = min(N_1, N_2, \ldots, N_n) \quad (18)$$

From (17), it can be deduced that $N_i \propto W_i$, hence

$$N \propto min(W_1, W_2, \ldots, W_n) \quad (19)$$

Finally, the computed N is informed to the slot mask mapper along with W_i to generate the optimal feasible allocation pattern for the contending SIMs. As we are already aware that the Uplink HARQ messages receive its acknowledgment with a pre-defined delay. Hence, the mapper block will take the optimal W_i and N as an input and generate the optimal pattern of resource allocation for SIM i catering the transmission to acknowledgment delay. Fig. 1 depicts the block diagram of the proposed MSIM scheduling architecture along with slot mask mapper that maps W_i and N to $Mask_i$. Hence, with the proposed architecture, we are able to distribute optimally the RF resource between the contending SIMs for a practical system.

5. Analytical Results

We considered the SR-DSDS i.e. $n = 2$ case for obtaining the optimal frame allocation ratio (W) and maximum possible transition count (N) computed in section 3 and 4 respectively. This section studies and demonstrates the impact of varying the quality of service parameters CQI, B and T on the optimal frame allocation W and maximum possible transition count N. As various 3GPP standards specify, CQI in the range of $[1, 30]$, we set $CQI_{max} = 30$. Generally, a finite buffer length will be available at the transmitter, hence we set $B_{max} = 10$. Average time criticality can be quantified in the range of $[0, 1]$, but for ease of numerical computation we scale T_{max} to 10. The associated quality parameters used for obtaining the analytical results of the mathematical modeling are provided in figure's caption.

As conventional scheduling algorithms can only generate binary decision about resource usage based on static or dynamic priority, the proposed architecture
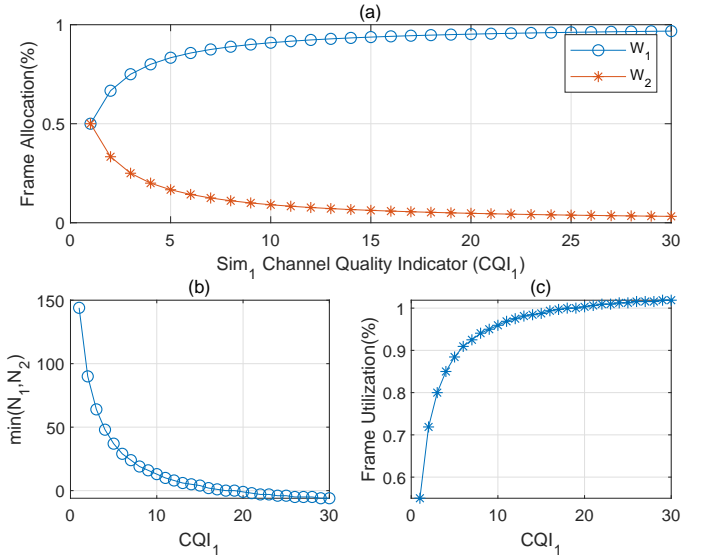

Figure 2. Frame Allocation and max transitions vs CQI_1 for $(CQI_2, B_1, B_2, T_1, T_2) = (1,1,1,1,1)$ respectively

cannot be directly compared for efficacy measurements. Hence, we simulate the proposed mathematical model and discuss inferences drawn from the analytical results. Each analytical result plot is further sub-divided into three subplots that depicts,

(i) The impact of variation of CQI or B or T on frame allocation W_1 and W_2,

(ii) The maximum possible transition count N for optimal allocation W_i and fixed $F = 640$ ms, $\tau = 2$ ms and $\theta = 5\%$,

(iii) The lower bound on frame utilization period for Rx/Tx considering same fixed parameter setting as in (b).

5.1. Impact of CQI variation on Scheduling Algorithm

Here we study the variation of CQI associated with SIM 1 from 1 to 30 while all other input parameters are kept fixed. In Fig. 2(a), buffer occupied and time criticality of both the SIMs is equal. CQI observed by SIM 2 is fixed at 1. The algorithm initially allocates equal resource to both SIMs at $CQI_1 = 1$ and then dynamically allocates more resources to SIM 1 as CQI_1 increases. In Fig. 2(b), peak of max possible transitions N is observed when $CQI_1 = 1$, i.e., when equal resources are allocated to both SIMs. Later as CQI_1 increases W_2 reduces and hence the curve of N follows W_2. This inference follows from (19). It can be observed in Fig. 2(c) that as N decreases, actual frame utilized for Rx/Tx increases. It follows naturally as total transition delay will be $N \times \tau$ where τ is fixed system driven parameter. This is the simplest permutation case to depict effect of varying CQI on resource optimal allocation.
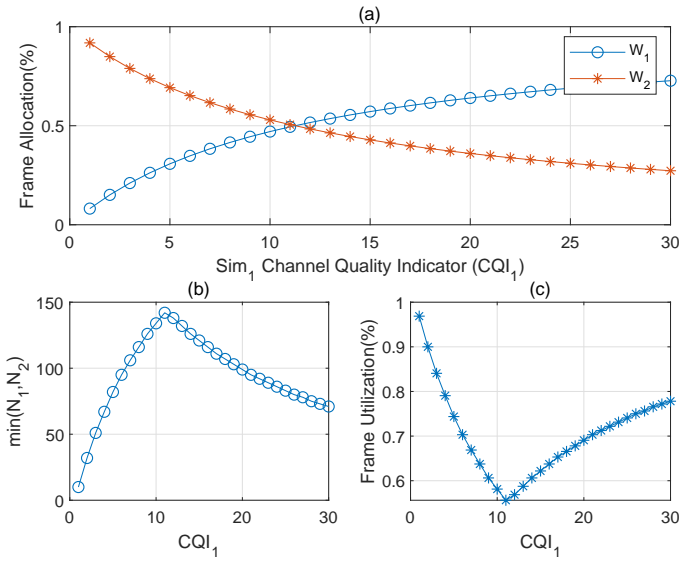
Figure 3. Frame Allocation and max transitions vs CQI_1 for $(CQI_2, B_1, B_2, T_1, T_2) = (15,7,3,4,7)$ respectively

Figure 4. Frame Allocation and max transitions vs Buffer Occupancy B_1 for $(CQI_1, CQI_2, B_2, T_1, T_2) = (30,30,7,1,1)$ respectively

In Fig.3(a), SIM 1 buffer is occupied more as compared to that of SIM 2 whereas the buffered content is more time critical in SIM 2. CQI observed by SIM 2 is set to 15. The scheduling algorithm dynamically chose to allocate more allocation to time critical buffer till CQI quality of SIM 1 is inferior. At $CQI_1 = 11$ even though the CQI_1 is inferior to CQI_2, the algorithm dynamically allocates more resources to SIM 1 as buffered content quantity in SIM 1 is more compared to that of SIM 2. Therefore, this plot depicts the efficacy of the mathematical model presented in section 3. Fig. 3(b) curve follows $\min(W_1, W_2)$ similar to the inference made while discussing Fig. 2(b). The peak for max possible transition count N appears at $CQI_1 = 11$ when W_1 and W_2 are equal. Fig. 3(c) is inferred to be a complement of Fig. 3(b). Similar inference can be made in all the subsequent analytical results and hence this redundant description of sub-plots (b) and (c) is further omitted for the sake of reader's interest.

5.2. Impact of variation of Buffer occupancy (B)

In Fig.4(a) the impact of varying SIM 1 buffer occupancy (B_1) on resource allocation W_1 and W_2 is depicted. CQI and T are kept fixed and same for both the SIMs. SIM 2 Buffer occupancy B_2 is set to 7 and B_1 is varied from 1 to B_{max}. It can be observed that the algorithm dynamically chose to allocate more resource to more buffered SIM i.e. SIM 2 until $B_1 = B_2 = 7$. Beyond this inflection point, resource allocation crossover is observed where algorithm chose to assign more resources to more buffered SIM 1 as compared to that of SIM 2.
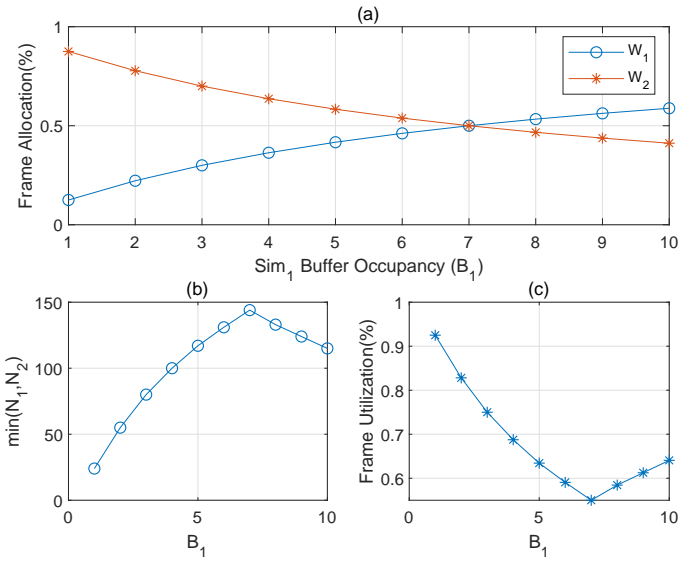
5.3. Impact of variation of Time criticality (T)

In Fig.5(a), the impact of varying the average time criticality parameter (T) on resource allocation W_1 and W_2 is demonstrated. CQI and buffer occupancy B are kept as similar constants for both the SIMs. Average time criticality, T_1 of SIM 1 buffered content is fixed at 4 whereas T_2 associated with SIM 2 is varied from 1 to T_{max}. Similar inferences can be made about inflection point and resource allocation crossover as that of Section 5.2. $T_1 = T_2 = 4$ is the inflection point where resource allocation crossover is observed.

5.4. Impact of utilization threshold (θ) and transition delay (τ) on Feasibility Controller

Table 1. Impact of variation of θ on N with $(F, \tau) = (640ms, 2ms)$ and $(CQI_2, B_1, B_2, T_1, T_2) = (15,7,3,4,7)$

$N(CQI_1)$	$\theta=5\%$	$\theta=10\%$	$\theta=20\%$
$N(1)$	51	-6(Not realizable)	-38(Not realizable)
$N(3)$	51	35	3
$N(11)$	142	126	94
$N(30)$	71	55	23

Table 1, shows impact of frame utilization threshold θ on maximum possible transitions N. It is observed that as θ increases, N decreases. It is evident from the results that it is good to allocate resource completely to one SIM during some cases, for instance when $CQI_1 = 1$ and $\theta=10\%$ or 20%, N turns out to be negative. For
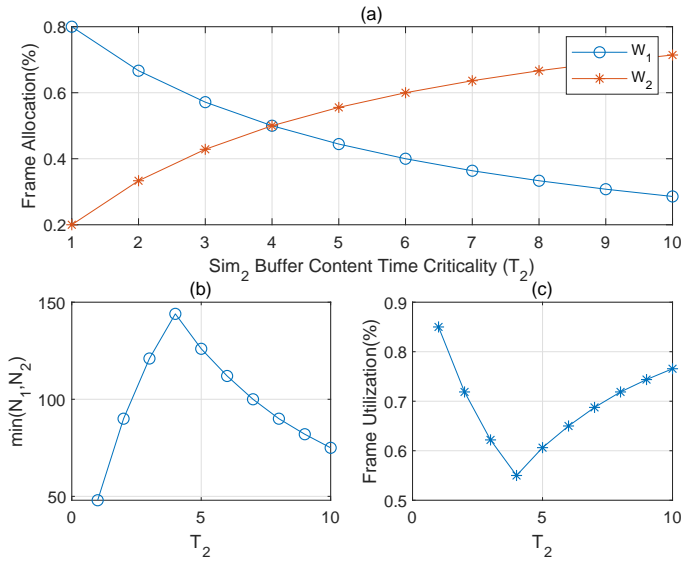
Figure 5. Frame Allocation and max transitions vs Buffer Content Time Criticality T_2 for $(CQI_1, CQI_2, B_1, B_2, T_1) = (30,30,10,10,4)$ respectively

Table 2. Impact of variation of τ on N with $(F, \theta) = (640ms, 5\%)$ and $(CQI_2, B_1, B_2, T_1, T_2) = (15,7,3,4,7)$

$N(CQI_1)$	τ=2ms	τ=10ms	τ=20ms
$N(1)$	51	2	1
$N(3)$	51	10	5
$N(11)$	142	28	14
$N(30)$	71	14	7

such cases, feasibility controller will decide not to apply dynamic resource sharing.

Table 2, shows impact of transition delay τ on maximum possible transition count N. As τ increases, total duration over which resources could be utilized reduces and thus N decreases. For instance, when CQI_1=1 varying τ from 2 ms to 20 ms, N reduces from 51 to 1.

For the sake of completeness, let us also compute the lower bound of threshold θ for which the feasibility condition satisfies. Considering fixed system parameters as $(F, \tau) = (640ms, 2ms)$ and $(CQI_1, CQI_2, B_1, B_2, T_1, T_2) = (1,15,7,3,4,7)$, W_1 and W_2 from equations (7) and (8), gives $W_1 = 0.0816$ and $W_2 = 0.9184$ and to satisfy feasibility constraint, N must be at least set to 1. Substituting W_1, W_2 and N in (17), gives $\theta = 7.8\%$ as the upper bound. A similar computation can be done for each case, which however is omitted to maintain the readability of the paper.

6. Conclusion

In this work, we considered the fundamental problem of resource sharing, inherently associated with multi-sim single RF subsystem UE. We proposed a multi-sim architecture with a scheduling algorithm and a feasibility controller that optimizes the shared resource across the contending SIMs. We formulated an optimization problem that optimally allocates the shared resource while dynamically choosing the allocation based on various important quality of service parameters like CQI, buffer occupancy and time criticality of buffered content. Further, extensive mathematical analysis is done using KKT conditions to compute the closed form optimal frame allocation ratio. Subsequently, we incorporated the RF blackout periods involved in switching between the SIMs for a practical multi-sim architecture, to compute the maximum number of transitions possible. Subsequently, analytical results presented, emphasize the effectiveness of proposed architecture. We also tabulate the cases where feasibility conditions fail, hence render scheduling algorithm ineffective. Therefore, the proposed scheme brings significant advantage over the conventional RF resource sharing method especially in SR-DSDS and DR-DSDS architecture and hence helps to design better multi-sim capable modem.

7. Copyright statement

Presented at WinTechCon-2019, organized by IEEE CAS Bangalore Chapter, IEEE Bangalore Section, and IEEE WiE Council Bangalore.

7.1. Copyright

The Copyright licensed to EAI.

References

[1] Ericsson, "Dual-sim dual-standby ues and their impact on the ran." 3GPP Technical Report RP-111637, 2011.

[2] J. L. Buthler and T. Sorensen, "Dynamic multi-sim gap creating procedure," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*. IEEE, 2016, pp. 1–5.

[3] L. Hang and C.-P. Hsu, "Simultaneous voice and data for dual-sim-dual-standby (dsds) wireless device," Jun. 23 2015, uS Patent 9,066,330.

[4] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.