# Analytical Modeling for Mobility Signalling in Ultradense HetNets

Azar Taufique, Abdelrahim Mohamed, Hasan Farooq, Ali Imran, and Rahim Tafazolli

***Abstract*—Multiband and multitier network densification is being considered the most promising solution to overcome the capacity-crunch problem of cellular networks. In this direction, small cells are being deployed within the macro cell (MC) coverage, to off-load some of the users associated with the MCs. This deployment scenario raises several problems. Among others, signalling overhead and mobility management will become critical considerations. Frequent handovers (HOs) in ultradense small cell (SC) deployments could lead to a dramatic increase in signalling overhead. This suggests a paradigm shift toward a signalling conscious cellular architecture with smart mobility management. To this end, the control/data separation architecture (CDSA) with dual connectivity is being considered for the future radio access. Considering the CDSA as the radio access network architecture, we quantify the reduction in HO signalling w.r.t. the conventional approach. We develop analytical models which compare the signalling generated during various HO scenarios in the CDSA and conventionally deployed networks. New parameters are introduced, which can with optimum value significantly reduce the HO signalling load. The derived model includes HO success and HO failure scenarios along with specific derivations for continuous and noncontinuous mobility users. Numerical results show promising CDSA gains in terms of saving in HO signalling overhead.**

***Index Terms*—Cellular networks, control data separation architecture (CDSA), dual connectivity handover, signalling load, radio access networks (RAN).**

## I. Introduction

USER mobility has been the raison d'etre of wireless cellular systems. Studies project that by 2021, global mobile traffic will increase sevenfold [1] as compared with 2016. This mobile traffic can be analysed to gain deep insights into human behavior, transportation, networking etc. [2]. At the same time, this trend highlights the need for making mobility management in future cellular networks even more resource efficient and seamless than ever before. In addition, multi-band and multi-tier networks consisting of cells of varying sizes on conventional sub 6 GHz and above 20 GHz bands, referred to as mmWaves (mmW) henceforth, are being perceived as the panacea for the looming capacity crunch. Particularly mmW small cells are being considered [3], [4] essential for future ultra dense multi-tier multi-band networks vis-a-vis 5G and beyond. This is because harnessing the abundant and short range mmW spectrum has strong potential to solve the two long-standing and intertwined problems in cellular networks: spectrum scarcity and interference. However, it is worth noticing that while advent of ultra dense mobile networks may solve these two problems, it creates a new challenging problem i.e., how to manage user mobility in such a dense network consisting of cells of varying sizes on a wide range of frequency bands with entirely different propagation characteristics. The following challenges define the breadth and depth of this impending problem.

Mobility management in the current networks requires periodic signalling to support HO preparation, execution and completion phases. This conventional approach may not be suitable in ultra-dense networks because HO rates and the associated signalling overhead will become unacceptably high [5], [6]. In the long term evolution (LTE), HO failure rate is targeted below 5 percent [7]. However, recent third generation partnership project study [5] shows that adding only ten small cells per macro can push the HO failure rate to as high as 60 percent, indicating the breakdown of current mobility management mechanism in ultra dense networks. Given the much smaller average cell size and thus small user sojourn time, the time to complete a HO must be reduced significantly from the current LTE target of 65 ms [8], [9]. New agile HO signalling procedures are also needed to meet the ambitious low latency requirements of the 5G system.

According to [10], the range of mmW based cells is 100–200 m. With ultra-dense mmW based SC deployments, mobility management becomes complex because HOs will happen frequently even for slow mobility users. In the conventional RAN architecture, the HO procedure includes transferring all channels (i.e., control and data) from one base station to another with a significant core-network signalling load [11]. For instance, the results reported in [12]–[14] indicate high signalling overhead and call drop rates when the conventional HO mechanisms are applied in dense SC deployment scenarios. To solve this problem, a futuristic radio access network architecture with a logical separation between control plane and data plane has been proposed in research community [15].

In CDSA, MCs are chosen to be control base stations. These control base stations provide basic connectivity and control

Manuscript received October 11, 2017; revised February 20, 2018 and May 1, 2018; accepted June 11, 2018. Date of publication August 9, 2018; date of current version March 14, 2019. The work was supported by Computer and Network Systems Program of National Science Foundation under grants 1559483 and 1619346. The review of this paper was coordinated by Prof. D. B. da Costa. *(Corresponding author: Azar Taufique.)*

A. Taufique is with Amazon.com LLC and also with the TechTrained.com LLC, Richardson TX 75080 USA (e-mail: azar.taufique@techtrained.com).

A. Mohamed and R. Tafazolli are with theInstitute of Communication Systems, Home of 5GIC, University of Surrey, Guildford GU2 7XH, U.K. (e-mail: abdelrahim.mohamed@surrey.ac.uk; r.tafazolli@surrey.ac.uk).

H. Farooq and A. Imran are with the BSON LaboratoryUniversity of Oklahoma, Tulsa OK 74135 USA (e-mail: hasan.farooq@ou.edu; ali.imran@ou.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2018.2864627

signalling needed for services as evident in Fig. 2. Under the umbrella of control base station, high speed and on demand data services are provided through SCs which are termed as data base stations. As illustrated in Fig. 2, all user equipments (UEs) are connected to control base station and only active UEs are connected to both control base station and data base station via a dual connection [14]–[16]. This type of coverage and connection offers simple and secure HO because in this case radio resource control (RRC) connection is maintained by the CBS. In this way, UEs are connected to a CBS with large coverage area. The HO trigger takes place when UE moves towards the edge of serving CBS. As the signal strength/quality of the serving CBS lowers and the neighboring CBS signal strength/quality and coverage becomes greater than the sum of the signal strength/quality of the serving CBS and a hysteresis (defined by the network), the HO event is triggered. The UE sends measurement report to the serving CBS. After reading the measurement report, the serving CBS decides if HO trigger is necessary or not. If HO is needed, serving CBS sends HO request to the target (neighbor) CBS and once HO request is acknowledged, the serving CBS sends HO trigger command to the UE to continue moving towards destination CBS. The procedure of HO trigger and decision is aligned with 3GPP standard [17]. Intra CBS HOs are transparent to the core network (CN), because DBSs are under the coverage of CBS all the time. Similar architecture configuration has been proposed in [18], for heterogeneous wireless networks. These configurations reduces mobility signalling, facilitate HO in control-user plane split scenarios [19], [20] and reduces the associated overhead.

There has been existing research work on other areas of HetNets such as energy efficiency [21]–[24] etc. However, according to authors' knowledge very little work has been done on reducing mobility signalling in HetNets. Most of the existing work in the literature for HO modeling has been done considering general HO scenario in macro cells or HetNets with conventional architecture. This paper provides a model for HO failure in CDSA. To the best of our knowledge, this is the first scheme that models HO failure under a CDSA configuration. In addition, we have also included parameter setting in terms of shared coverage area to minimize signalling and reduce HO failure as part of modeling. Therefore, the importance of this work is twofold. 1) It proposes a HO model for CDSA in HetNets 2) It proposes an analytical model for active mobility signalling generated during HO for CDSA in HetNets.

The work presented in [25] focuses on optimization of HO procedure in HetNets by incorporating context information such as user speed, channel gains and traffic load in the cells. This work proposes a Markov chain-based framework to model the HO process for the mobile user and derives an optimal context-dependent HO criterion. This work clearly demonstrates that context-awareness can indeed improve the HO process and significantly increase the performance of mobile UEs in HetNets. In contrast to [25], this study aims to address the question of how much signalling is generated in case of HOs for CDSA based HetNets. On the other hand, the works presented in [26] does provide HO analysis for CDSA based HetNets. This work provides first tractable mobility aware model for a two-tier downlink cellular network with ultra-dense small cells and Control plane / User plane split architecture. The work performs in depth HO analysis and sheds light on HO costs in terms of number of HOs taking place per unit length. However, widely differing from the scope of our study, [26] does not compare quantitatively how much amount of signalling load is generated in case of HO success and HO failure in CDSA for HetNets. The work in [27] focuses on HO problem in two tier networks which arises in HetNets due to network densification. The solution to the problem is specified in terms of HO skipping based on velocity of the user, so that connection can be maintained for longer duration without causing any connection interruption. HO cost is defined based on the delay incurred on account of HO interruption which takes place during a HO. Even though [27] considers two tier HetNet model, unlike our study it does not consider a CDSA architecture specifically. In addition, it does not take into account for the mobility signalling load considerations.

According to Nokia Siemens networks, in current network deployments signalling is growing 50 percent faster than data traffic [28]. Previously published works [14], [29]–[31] on mobility signalling claims that mobility signalling is reduced as long as UE's mobility is within the coverage area of CBS. As a result signalling channel is not changed and mobility signalling is reduced. However, this is not the case when the UE moves from one CBS to another CBS. Other studies [32]–[34] analyse the dual connectivity and HO failure rate of the CDSA using simulations, without providing a concrete analytical framework. In order to evaluate the HO signalling cost, [35] and [36] propose HO management schemes and evaluate the signalling cost for femtocells. However, these analysis assume that the HO is successful for 100 percent of the time which is not the case in real networks.

In order to assess the mobility and signalling reduction benefits of CDSA, it requires a framework that quantifies the mobility-related signalling load. A first attempt towards this framework is reported in [11]. While this framework provides the foundation, it misses out two important facts.

1. It does not consider HO failures scenario. 2. It does not take into account quality of service (QoS) requirements such as HO time, shared coverage factor and HO duration related parameters which are essential for QoS requirements of time sensitive applications in ultra-dense HetNets.

Keeping the above limitations of [11] in mind, we make the following contributions in this paper.

1) We propose an analytical model which provides probability of HO failure signalling, probability of HO success signalling and probability of no HO signalling.
2) We propose an analytical model to quantify the mobility signalling generated in the core network for the case of HO success, and HO failure in both CDSA and conventional network.
3) The analytical model quantifies the expected mobility CN signalling generated: when either a non-continuous mobility or continuous mobility scenario takes place.
4) HO related parameters such as mobility time duration ($T_d$), time taken for a HO completion ($T_p$) and coverage factor (c) are introduced. These parameters when

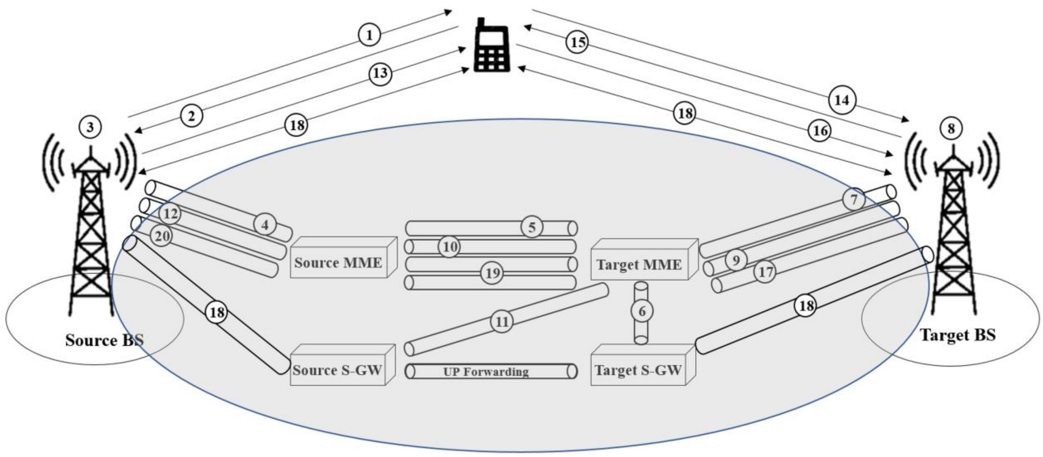



Fig. 1. Signalling messages exchanged with in a typical S1 handover scenario.

configured appropriately can reduce HO related CN signalling. Also, for a given topology these parameters can decide the success or failure of a HO.

5) Finally, this work informs the reader whether multi-tier multi band SCs ought to be deployed using current HetNets format or CDSA approach. The results confirm that for ultra-dense HetNets deployment CDSA ought to be used.

The remainder of the paper is organized as follows. In Section II we discuss the system model and assumptions, Section III presents the signalling probability model. The expected CN mobility signalling model is presented in Section IV. A special scenario for the case of continuous mobility signalling is presented in Section V. We present numerical results in Section VI followed by conclusion, in Section VII respectively.

## II. System Model and Assumptions

The analytical model for reducing HO signalling in this paper focuses only on the RRC related CN mobility signalling exchange which takes place during the HO procedure (active mode) in a cellular system. Fig. 1 indicates the amount of signalling exchange which is generated during a typical S1 HO scenario [37]. The actual sequence of HO messages are shown in Table I. It is evident that compared to RAN side majority of the mobility signalling messages are being exchanged with the CN as shown in the shaded region in Fig. 1. In the CDSA the CN signalling remains unchanged as long as UE mobility is within the same CBS because intra CBS HOs are transparent to the CN, as described in Introduction section and shown in Fig. 2.

The model is applicable for both intra-frequency and inter-frequency HO scenarios. In the CDSA network, the CBS and DBS are usually deployed in separate frequency bands to avoid inter-layer interference. Although this may complicate the UE radio frequency (RF) design, a separate frequency deployment is being considered in the new radio guidelines of the 3GPP [38]. In this direction HOs within the footprint of the same CBS require changing the DBS only, hence they are considered intra-frequency HOs. On the other hand, inter-CBS HOs, i.e., HOs between two different CBSs, require changing both the CBS and DBS, i.e., a two-link HO. Such a scenario may involve intra-

TABLE I
HO Messages

| Number | Description |
|---|---|
| 1 | RRC Connection Reconfiguration |
| 2 | RRC Measurement Report |
| 3 | HO Decision |
| 4 | S1 Handover Required |
| 5 | S10 forward Relocation Request |
| 6 | S11 Create Bearer Request/Response |
| 7 | S1 handover request |
| 8 | Admission Control |
| 9 | S1 Handover Request Acknowledge |
| 10 | S10 Forward Relocation Response |
| 11 | S11 Create Bearer Request/Response |
| 12 | S1 Handover Command |
| 13 | RRC Connection Reconfiguration |
| 14 | Random Access Preamble |
| 15 | Random Access Response (UL Allocation + TA) |
| 16 | RRC Connection Reconfiguration Complete |
| 17 | S1 Handover Notify |
| 18 | Data Transfer in Target |
| 19 | S10 Forward Relocation Complete/ACK for control base station |
| 20 | S1 UE Context Release Command |

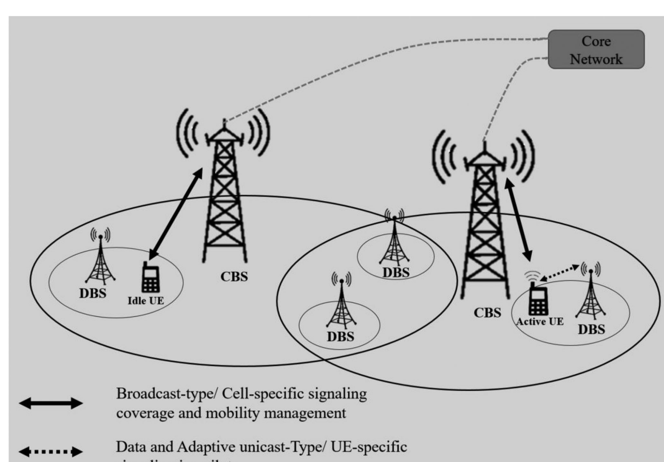



Fig. 2. Schematic of control data plane separated Architecture.

frequency and inter-frequency HOs. Consequently, a longer or shorter measurement gap may be required depending upon the cell deployment density. According to Mahbas *et al* [39], using smaller values of measurement gap, better system performance can be achieved in case of dense cell density and higher values of measurement gap in case of sparse cell density. This issue can be solved by limiting the number of CBSs / DBSs that are being monitored by the UE, e.g., the UE monitors the top-n cells per cell categorization to ensure that data transmission is balanced against the accurate measurement cycle. However, modeling this specific scenario is beyond the scope of this paper, and can be goal of future work.

In order to derive the analytical model for mobility signalling in case of CDSA. The following assumptions are made.

- The user remains in the system upon HO success or HO failure. In case of HO failure UE will remain connected to the CBS but will require RRC re-establishment with the DBS.
- Different amount of signalling is generated in CN in case of HO failure and HO success. Specifically, a HO failure event generates more signalling than a HO success event as shown in Fig. 4.
- A user can remain within the same CBS and not perform an inter-CBS HO with probability $P_{no}$. It can perform an inter-CBS HO from one CBS to another CBS. Inter CBS HO can be successful with probability $P_s$ or it could be a failure with probability $P_f$.
- An LTE system with equal number of low and high mobility distributed users are considered. The term sector is used with the same meanings as a cell.
- HO failure is considered on account of too late HO. These assumptions are valid for CDSA in ultra-dense networks, as with densification more HO failures may take place because of too late scenario if HO parameters are not tuned accordingly.
- HO failure can take place due to various other reasons other than too late HO, such as transport network reliability i.e., S1 interface is down, poor RF conditions, radio link failure and partial HO etc. Analytical model for HO failure due to reasons other than too late HO can be derived accordingly. For the case of HO failure caused by poor RF conditions, radio link failure is triggered when the downlink signal to interference noise ratio (SINR) is below a certain threshold (Qout $= -8$ dB) and stays below 6 dB for at least 1 sec [40]. Using this approach, probability of SINR greater than the threshold can be computed. If SINR threshold is less than the threshold for a given time, it will be a HO failure and vice versa. Similarly, HO failure caused by partial HO can be characterized by calculating the probability that whether all bearers get transferred completely or not. By computing the probability of all bearers transferred or not, we can compute probability of HO success or HO failure respectively. For a more detailed discussion on possible HO failure scenario, reader is referred to [40].
- Regardless of the HO failure reasons all factors contribute to the same amount of CN signalling load.

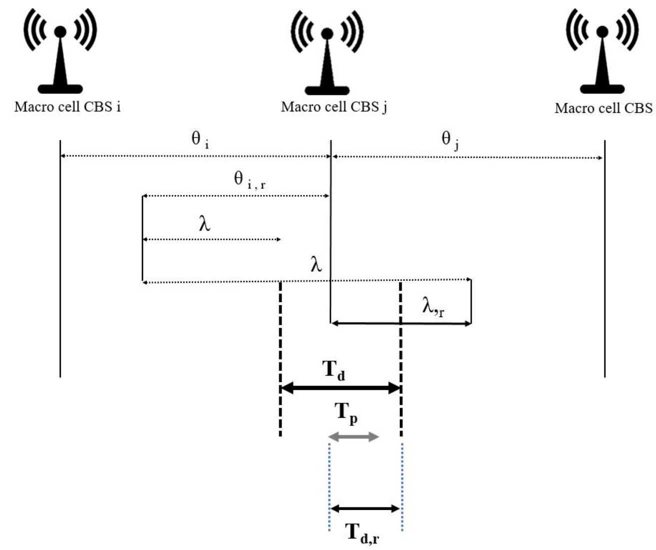



Fig. 3. Timing diagram of handover model parameters.

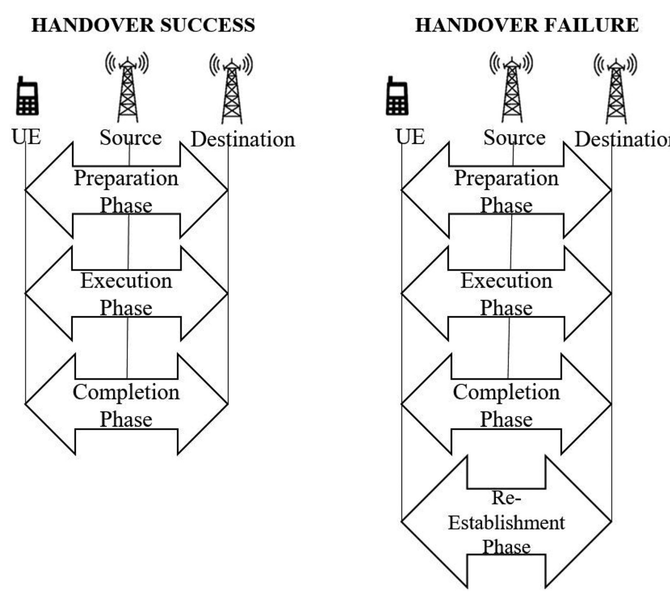



Fig. 4. Comparison of Handover Success and Handover Failure Procedures.

- For the evaluation and comparison purposes the modeling approach proposed in Sections III to V can be adapted to model the conventional network mobility signalling in order to assess the CDSA gains as proposed in [11].

## III. Signalling Probability Model

In order to evaluate the CN signalling load as a result of HO, the probabilities of failure ($P_f$), success ($P_s$) and no HO ($P_{no}$) needs to be modelled. To compute these values, we need to model the HO procedure between two CBS in terms of a timing diagram as shown in Fig. 3. From the timing diagram in Fig. 3 the abbreviations for various terms used and defined are provided in Table II.

TABLE II
SYMBOL AND ACRONYMS DESCRIPTION

| Symbol | Description |
|---|---|
| $S_f$ | Normalized CN mobility signalling load on account of HO failure |
| $S_s$ | Normalized CN mobility signalling load on account of HO success |
| $\lambda$ | Session Duration |
| $\lambda_{,r}$ | Residual Session Duration |
| $\theta$ | Cell residence time |
| $\theta_{,r}$ | Residual Cell residence time |
| $\theta_{i,r}$ | Residual Cell residence time of control base station i |
| $\theta_{,2}$ | Residual Cell residence time of data base station |
| $T_d$ | Mobility time duration during which HO takes place |
| $T_{d,r}$ | Residual Mobility time duration during which HO takes place |
| $T_p$ | Time taken for handover completion |
| $\rho_1$ | Cell density of control base station |
| $\rho_2$ | Cell density of data base station |
| $v$ | Average velocity |
| $L_1$ | Length of perimeter of the control base station |
| $S_1$ | Area of control base station |
| $c$ | Coverage factor |
| $LTE$ | Long Term Evolution |
| $mmW$ | Millimeter Waves |
| $MME$ | Mobility Management Entity |
| $3GPP$ | Third Generation Partnership Project |
| $MME$ | Mobility Management Entity |
| $CN$ | Core network |
| $HO$ | Handover |
| $RRC$ | Radio Resource Control |
| $BS$ | Base Station |
| $CDSA$ | Control Data Separation Architecture |
| $CP$ | Control Plane |
| $DP$ | Data Plane |
| $CBS$ | Control Base Station |
| $DBS$ | Data Base Station |
| $TA$ | Time Alignment |
| $UE$ | User Equipment |
| $SC$ | Small Cell |

### A. *System Model*

Consider a CDSA cellular network where the CBSs are modelled as a Poisson Point Process (PPP) with density $\rho_1$, while DBSs are modelled as another PPP with density $\rho_2$, where $\rho_2 \geq \rho_1$.

Assume a session duration λ with probability density function (pdf) $f_S(\lambda)$ and mean E[λ]. The CBS residence time is modelled as a random variable $\theta_1$ with pdf $f_{R_1}(\theta_1)$ and mean E[$\theta_1$], while the DBS cell residence time is modelled as a random variable $\theta_2$ with pdf $f_{R_2}(\theta_2)$ and mean E[$\theta_2$].

Fig. 3 provides a timing diagram that illustrates the definition of all the parameters. Without loss of generalization, we follow [11] and assume that users move at random directions with a random velocity. Under this assumption, E[$\theta_1$] can be approximated by the ratio between the number of UEs in a CBS and the number of UEs leaving a CBS per unit time [41]. According to [41]

$$\mathrm{E}[\theta_1] = \frac{\text{Number of UEs in a CBS}}{\text{Number of UEs leaving a CBS}} \tag{1}$$

Following derivations in [41], E[$\theta_1$] can be approximated as:

$$E[\theta_1] = \frac{\pi * S_1}{E[v]L_1} \tag{2}$$

where, the symbols $S_1$ in equation (2) indicates area of the CBS and $L_1$ represents length of the perimeter of CBS as given in Table II. As we are considering a PPP model, according to [42]

$$S_1 = \frac{1}{\rho_1} \tag{3}$$

and

$$L_1 = \frac{4}{\sqrt{\rho_1}} \tag{4}$$

Substituting (3) and (4) into (2). Equation (2) can be re-written as

$$E[\theta_1] \approx \frac{\pi}{4E[v]\sqrt{\rho_1}} \tag{5}$$

Similarly, the expected cell residence time for DBS can be listed as

$$E[\theta_2] \approx \frac{\pi}{4E[v]\sqrt{\rho_2}} \tag{6}$$

### B. *Mobility Time Duration ($T_d$)*

If a user is in moving state, on account of mobility it eventually moves from one CBS to another CBS. In order for a user to perform a HO successfully the HO must be completed within a required time duration. Otherwise, UE may move out of coverage of the serving CBS and HO failure occurs. As the user approaches the edge of CBS, it starts receiving signal coverage from a neighboring CBS. Ideally, the HO ought to take place when a user is receiving signal from both neighboring and serving CBS, while it is moving in the direction of the neighboring CBS. This duration while UE is moving in the direction of neighboring CBS and receiving coverage from both serving and neighboring CBS is termed as mobility time duration and abbreviated as $T_d$.

In order to define $T_d$ mathematically. We proceed as follows; a UE stays in a cell for a given time equal to average cell residence time ($\theta$). $T_d$ is a function of average cell residence time. Cell residence time is dependent upon cell density and user velocity. Therefore, in order to derive a relation between mean cell residence time and mobility time duration (HO duration), we model it as:

$$T_d = E[\theta] * c * 10^{-3} \tag{7}$$

Where $T_d$ is the HO time duration (HO time) in the above equation. HO time depends upon the coverage parameter *c*. For larger shared coverage area, HO duration is longer because it takes longer for a user to traverse the intercell coverage area and for small coverage area it is shorter accordingly. Therefore, the coverage parameter *c* ranges between 0.1 and 0.9. The coverage factor *c* is dimensionless in our model. As mobility time duration is taken in milliseconds, whereas cell residence time is in seconds (depending upon the cell radius) therefore a factor of 1/1000 is added for conversion.

### C. Time Taken For a Handover Completion ($T_p$)

The HO procedure starts from the instant measurement report is sent by the UE to the source CBS, and concludes once UE receives RRC connection reconfiguration message from the target CBS. The HO procedure consists of three phases: preparation, execution and completion phase as shown in Fig. 4. For a successful HO all three phases need to be completed successfully. The time taken to complete all the phases of HO successfully is termed as $T_p$. In [43], authors have studied the HO failure rate and delay of the HO as well. Their result include overall HO duration which is around 83–95 ms. HOs can take place sooner than this duration as well. In order to compute the effect of signalling load in case of both HO success and failure scenario we use the value of $T_p$ as 100 ms, an upper bound to meet the HO delay requirements in this work.

### D. Probability of Handover Failure

For the CDSA system model, it is known that CN signalling is generated in inter CBS HOs only [15]. Expressed differently all the DBS HOs do not generate CN signalling as long as the CBS anchor point remains the same. The definition of $P_f$ is equivalent to the UE attempting to change the serving CBS, while doing so it is not successful. With reference to Fig. 3, $P_f$ is equivalent to the probability that $T_p$ occurs beyond residual mobility time duration $T_{d,r}$. The session started when UE was associated with CBS$_i$ and failed to finish and drops the session when the UE tries to attempt a HO in order to associate with CBS$_j$, where j $>$ i and $T_p > T_{d,r}$. Considering Fig. 3, we can write $P_f$ as:

$$P_f = \text{Prob.}(T_p > T_{d,r}) * \text{Prob.}(\lambda > \theta_{1,r}) \tag{8}$$

where Prob.( ) means probability of an event and $\theta_{1,r}$ is the residual cell residence time of a CBS. The probability that session duration ($\lambda$) is greater than the residual cell residence time, it is computed as:

$$\text{Prob.}(\lambda > \theta_{1,r}) = 1 - \text{Prob.}(\lambda < \theta_{1,r})$$

When session duration is less than residual cell residence time it is computed as:

$$\text{Prob.}(\lambda < \theta_{1,r}) = \int_{x=0}^{\infty} f_{\theta_{1,r}}(x) \int_{y=0}^{x} f_\lambda(y)\, dy\, dx \tag{9}$$

$$\text{Prob.}(\lambda > \theta_{1,r}) = 1 - \int_{x=0}^{\infty} f_{\theta_{1,r}}(x) \int_{y=0}^{x} f_\lambda(y)\, dy\, dx \tag{10}$$

Similarly, $T_{d,r}$ is the residual mobility time duration during which HO takes place as shown in Fig. 3. The probability that time taken for an inter-CBS HO completion ($T_p$) is greater than residual mobility time duration is computed as:

$$\text{Prob.}(T_p > T_{d,r}) = 1 - \text{Prob.}(T_p < T_{d,r})$$

When time taken for HO completion is less than residual mobility time duration, it is computed as:

$$\text{Prob.}(T_p < T_{d,r}) = \int_{z=0}^{\infty} f_{T_{d,r}}(z) \int_{v=0}^{z} f_{T_p}(v)\, dv\, dz \tag{11}$$

$$\text{Prob.}(T_p > T_{d,r}) = 1 - \int_{z=0}^{\infty} f_{T_{d,r}}(z) \int_{v=0}^{z} f_{T_p}(v)\, dv\, dz \tag{12}$$

Plugging the values from equations (12) and (10) in equation (8), we get probability of failure as:

$$P_f = \left(1 - \int_{z=0}^{\infty} f_{T_{d,r}}(z) \int_{v=0}^{z} f_{T_p}(v)\, dv\, dz\right) * \left(1 - \int_{x=0}^{\infty} f_{\theta_{1,r}}(x) \int_{y=0}^{x} f_\lambda(y)\, dy\, dx\right) \tag{13}$$

### E. Probability of Handover Success

The definition of $P_s$ is equivalent to the UE attempting to change the serving CBS, and it is successful in doing so. With reference to Fig. 3, $P_s$ is equivalent to the probability that $T_p$ instant occurs before $T_{d,r}$ duration. In other words, the session started when UE was associated with $CBS_i$ and finished successfully in the next CBS when the UE tried to attempt a HO in order to associate with $CBS_j$, where j $>$ i and $T_p < T_{d,r}$. Considering Fig. 3, we can write $P_s$ as:

$$P_s = \text{Prob.}(T_p < T_{d,r}) * \text{Prob.}(\lambda > \theta_1, r) \tag{14}$$

Plugging values from equations (10) and (11) into equation (13), we get:

$$P_s = \left(\int_{z=0}^{\infty} f_{T_{d,r}}(z) \int_{v=0}^{z} f_{T_p}(v)\, dv\, dz\right) * \left(1 - \int_{x=0}^{\infty} f_{\theta_{1,r}}(x) \int_{y=0}^{x} f_\lambda(y)\, dy\, dx\right) \tag{15}$$

### F. Probability of No Handover

The definition of $P_{no}$ is the probability that the UE does not attempt to change the serving CBS. With reference to Fig. 3, $P_{no}$ is equivalent to the probability that the session started when UE was associated with $CBS_i$ and finished successfully in the same CBS and UE did not try to attempt a HO in order to associate with $CBS_j$, where j $>$ i. Considering Fig. 3, we can write $P_{no}$ as:

$$P_{no} = \text{Prob.}(\lambda < \theta_1, r) \tag{16}$$

Plugging value from equation (9) into equation (16), it becomes:

$$P_{no} = \int_{x=0}^{\infty} f_{\theta_{1,r}}(x) \int_{y=0}^{x} f_\lambda(y) dy dx \tag{17}$$

The probabilities of HO failure, HO success and no HO signalling considering general distributions are shown in equations (13), (15) and (17) respectively. In order to have closed form expression for these probabilities we consider exponential distribution as follows.

### G. Exponential Distribution for Session Duration, Mobility Time Duration, and Cell Residence Time

The expressions for $P_f$, $P_{no}$ and $P_s$ computed earlier in equations (13), (15) and (17) are given for general distribution. In order to have a closed form solution, we consider the scenario where the session duration, cell residence time and mobility time duration are exponentially distributed. Exponential distribution has been considered in this paper as it represents the worst-case scenario from signalling load perspective. The model(s) in [11] show that the HO-related signalling load is memoryless under exponential distribution and the signalling probability is independent of the previous case. A HO success or failure at one CBS does not mean it will result in HO upon the next consecutive CBS. Consequently, we consider the exponential distribution to model the worst-case scenario in both the CDSA and the conventional architecture and to evaluate the upper bound of the signalling load that corresponds to insights into the worst case scenario.

According to [11] when session duration and the cell residence time are exponentially distributed, the residual session duration and the residual cell residence time are also exponentially distributed such that

$$f_\lambda(t) = f_{\lambda,r}(t) = \frac{e^{-\frac{t}{E[\lambda]}}}{E[\lambda]} \tag{18}$$

$$f_{\theta_1}(t) = f_{\theta_{1,r}}(t) = \frac{e^{-\frac{t}{E[\theta_1]}}}{E[\theta_1]} \tag{19}$$

The mobility time duration is derived from cell residence time. Therefore, if cell residence time is considered exponential. Hence, probability density function (pdf) of mobility time duration is given as:

$$f_{T_d}(x) = \frac{e^{\frac{-x}{E[T_d]}}}{E[T_d]}$$

Similarly, using Lemma 1 in [11] the pdf of residual mobility time duration in case of exponential distribution is given as:

$$f_{T_{d,r}}(t) = f_{T_d}(t) = \frac{e^{\frac{-t}{E[T_d]}}}{E[T_d]} \tag{20}$$

Substituting (18), (19) and (20) into equations (13), (15) and (17) respectively. After mathematical simplification, we get $P_f, P_s$ and $P_{no}$ closed form expressions to be as follows:

$$P_f = \left(\frac{E[T_p]}{E[T_{d,r}] + E[T_p]}\right) * \left(\frac{4E[\lambda]E[v]\sqrt{\rho}_1}{\pi + 4E[\lambda]E[v]\sqrt{\rho}_1}\right) \tag{21}$$

$$P_s = \left(\frac{E[T_{d,r}]}{E[T_{d,r}] + E[T_p]}\right) * \left(\frac{4E[\lambda]E[v]\sqrt{\rho}_1}{\pi + 4E[\lambda]E[v]\sqrt{\rho}_1}\right) \tag{22}$$

$$P_{no} = \frac{\pi}{\pi + 4E[\lambda]E[v]\sqrt{\rho}_1} \tag{23}$$

The closed form expressions for $P_f$ and $P_s$ indicate that they depend upon cell density, user velocity, session duration, mobility time duration and time taken for a HO completion. Mobility time duration in turn depends upon cell residence time and cell coverage factor. Therefore, from a design perspective larger value of coverage factor and high cell residence time result in better values of successful HO probability. This insight can help cellular network designers to plan better ultra-dense networks which can result in more successful HOs and less amount of mobility signalling compared to conventional networks.

## IV. Mobility Signalling Model

The total CN mobility signalling load generated during a HO depends upon a number of factors such as:

- UE speed and mobility
- Base station (BS) density
- Session duration
- Transport network reliability (stability)
- Coverage factor
- Miscellaneous

The user(s) is assumed to be RRC connected and active in the network. We model the HO scenario and CN signalling generated as a result of probability of HO failure, success and no HO signalling using Markov chain as shown in Fig. 5. In the CDSA approach shown in Fig. 2, each inter-CDSA HO success or HO failure generates CN signalling, we will denote this CN signalling as $S_{i,j}$, while intra-CBS HOs i.e., DBSs do not generate CN signalling. The coefficients $\alpha, \beta$ and $\gamma$ in Fig. 5 are HO coefficients. In a cellular network, the probability to HO from one cell to another is not the same for all the sectors. This difference is on account of various factors described above. These HO coefficient values represent the difference in probabilities for HO coefficients from one cell to another cell. Using [11] we can model the CN signalling on amount of HO success, failure and no HO as shown in Fig. 5.

where,

- $P_f$ = Probability that signalling is generated as a result of HO failure
- $P_s$ = Probability that signalling is generated as a result of HO success
- $P_{no}$ = Probability that HO attempt will not be made
- $S_{i,j}$ = CN mobility signalling load generated on account of *i* handover failure(s) and *j* handover success(s)

The goal is to find out the average or expected amount of mobility CN signalling which is generated in case of HO success(s) and HO failure(s) including how no HO attempts will influence the aggregate signalling. From Fig. 5 it is clear that user will always generate mobility signalling starting from state $S_{0,0}$ and will not stay in that state.

The expected value of RRC CN mobility signalling load E[$S_{i,j}$] generated by a UE in the CDSA can be calculated as:

$$E[S_{i,j}] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} S_{i,j} * \text{Prob.}(S_{i,j}) \tag{24}$$

The Prob. ($S_{i,j}$) can be calculated by solving the Markov chain shown in the Fig. 5. Since the amount of signalling generated by the user(s) movement increases with time, a transition from CN signalling state $CS_{i,j}$ to $CS_{m,n}$ has zero probability when $i, j > m, n$. Based on this Markov chain, Prob. ($S_{i,j}$) can be

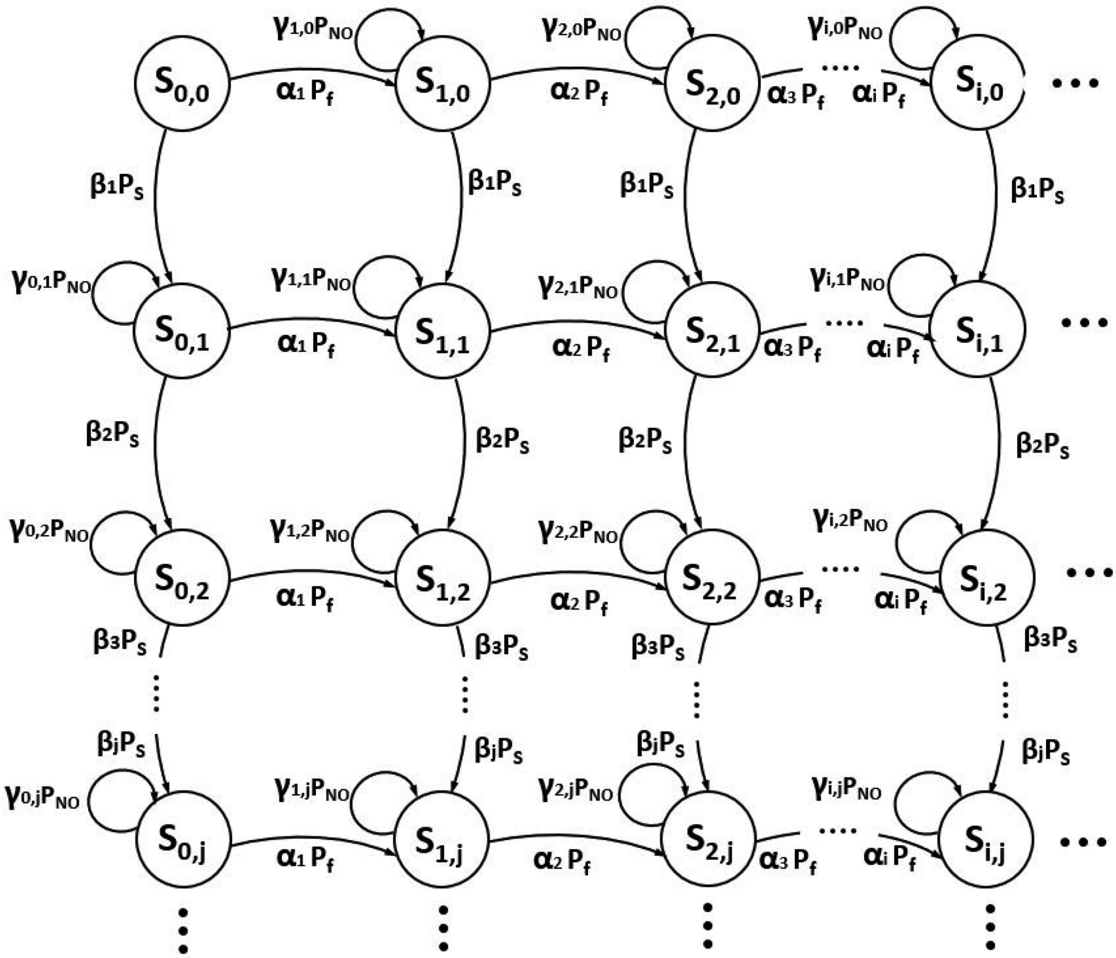



Fig. 5. Markov chain modeling of no-handover, handover failure and handover success related core-network mobility signalling.

formulated as:

$$\text{Prob.}(S_{i,j}) = \begin{cases} P(S_{0,0}) & ,\text{for } i=0,\ j=0 \\ \dfrac{\alpha_i P_f^i}{(1-P_{no})^i} * P(S_{0,0}) & ,\text{for } i>0,\ j=0 \\ \dfrac{\beta_j P_s^j}{(1-P_{no})^j} * P(S_{0,0}) & ,\text{for } i=0,\ j>0 \\ \dfrac{(i+j)\alpha_i P_f^i \beta_j P_s^j}{(1-P_{no})^{i+j}} * P(S_{0,0}) & ,\text{for } i>0,\ j>0 \end{cases} \tag{25}$$

From the Markov chain in Fig. 5 the values of HO coefficients $\alpha, \beta$ and $\gamma$ are such that the following conditions are true.

$$\beta_1 P_s + \alpha_1 P_f = 1, \text{for } i=0,\ j=0$$
$$\beta_1 P_s + \alpha_{i+1} P_f + \gamma_{i,0} P_{no} = 1, \text{for } i>0,\ j=0$$
$$\beta_{j+1} P_s + \alpha_1 P_f + \gamma_{0,j} P_{no} = 1, \text{for } i=0,\ j>0$$
$$\beta_{j+1} P_s + \alpha_{i+1} P_f + \gamma_{i,j} P_{no} = 1, \text{for } i>0,\ j>0$$

For a cellular network the values of $\alpha, \beta$ and $\gamma$ are arranged in the following order:

$$\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \alpha_4 \geq \cdots \geq \alpha_i$$
$$\beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4 \geq \cdots \geq \beta_j$$
$$\gamma_{1,0} \leq \gamma_{2,0} \leq \gamma_{3,0} \leq \gamma_{4,0} \leq \cdots \leq \gamma_{i,0}$$
$$\gamma_{0,1} \leq \gamma_{0,2} \leq \gamma_{0,3} \leq \gamma_{0,4} \leq \cdots \leq \gamma_{0,j}$$
$$\gamma_{1,1} \leq \gamma_{2,1} \leq \gamma_{3,1} \leq \gamma_{4,1} \leq \cdots \leq \gamma_{i,1}$$
$$\gamma_{1,2} \leq \gamma_{2,2} \leq \gamma_{3,2} \leq \gamma_{4,2} \leq \cdots \leq \gamma_{i,2}$$

Similarly,

$$\gamma_{1,j} \leq \gamma_{2,j} \leq \gamma_{3,j} \leq \gamma_{4,j} \leq \cdots \leq \gamma_{i,j}$$

*Lemma 1:* For exponential distribution of cell residence time and session duration. The values of $\alpha, \beta$ and $\gamma$ are:

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 =, \ldots, \alpha_i = 1$$
$$\beta_1 = \beta_2 = \beta_3 = \beta_4 =, \ldots, \beta_j = 1$$
$$\gamma_{1,0} = \gamma_{2,0} = \gamma_{3,0} = \gamma_{4,0} =, \ldots, \gamma_{i,0} = 1$$

$$\gamma_{0,1} = \gamma_{0,2} = \gamma_{0,3} = \gamma_{0,4} =, \ldots, \gamma_{0,j} = 1$$
$$\gamma_{1,1} = \gamma_{2,1} = \gamma_{3,1} = \gamma_{4,1} =, \ldots, \gamma_{i,j} = 1$$

*Proof. Preliminary:* Given the session duration and the CBS residence time are exponentially distributed, the residual session duration and the residual CBS residence time will also be exponentially distributed [11]. Consequently, the probability of not generating signalling is memoryless and independent of the state i.e., independent of the state whether signalling has been generated previously or not. ■

This implies that $P_{no} = \gamma_{1,0}P_{no} = \gamma_{2,0}P_{no} \cdots = \gamma_{i,0}P_{no}$ resulting in $\gamma_{1,0} = \gamma_{2,0} \cdots = \gamma_{i,0} = 1$. Similarly, for other $\gamma_{0,j} = \gamma_{i,j} = 1$. Since at any given state, $\alpha_i P_f + \beta_j P_s + \gamma_{i,j} P_{no} = 1$. When $\gamma_{i,j} = 1$ for all states, then the term $\alpha_i P_f + \beta_j P_s$ remains the same in all the states. As the residual session duration and the residual cell residence time have exactly the same distribution as the session duration and the cell residence time, respectively, $\alpha_i P_f$ and $\beta_j P_s$ remain constant in all states. This condition can only be satisfied when $\alpha_1 = \alpha_2 = \cdots = \alpha_i = 1$ and $\beta_1 = \beta_2 = \cdots = \beta_j = 1$.

As the probabilities of the signalling states in Markov chain for Fig. 5 are shown in equation (25). These probabilities depend on state $S_{0,0}$ i.e., $P(S_{0,0})$. Once we compute the probability of this state, we can compute probabilities for other states as well. For a Markov chain we know that.

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \text{Prob.}(S_{i,j}) = 1$$

Using Fig. 5 we can sum up all the signalling states such that:

$$\text{Prob.}(S_{0,0}) + \sum_{i=1}^{\infty} \text{Prob.}(S_{i,0}) + \sum_{j=1}^{\infty} \text{Prob.}(S_{0,j}) + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \text{Prob.}(S_{i,j}) = 1$$

After simplifying the equation above, we can write the Prob. $(S_{0,0}) = P(S_{0,0})$ as:

$$P(S_{0,0}) = \frac{1}{1 + \sum_{i=1}^{\infty} \text{Prob.}(S_{i,0}) + \sum_{j=1}^{\infty} \text{Prob.}(S_{0,j}) + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \text{Prob.}(S_{i,j})} \tag{26}$$

After mathematical procedures and solving equation (26). We get

$$P(S_{0,0}) = \frac{P_s P_f}{P_s P_f + P_f^2 + P_s^2 + P_f(1-P_{no}) + P_s(1-P_{no})} \tag{27}$$

The probability P($S_{0,0}$) is computed in equation (27). After plugging it in equation (24). We can find the expected CN mobility signalling.

$$E[S_{i,j}] = S_{0,0} * \text{Prob.}(S_{0,0}) + \sum_{i=1}^{\infty} S_{i,0} * \text{Prob.}(S_{i,0}) + \sum_{j=1}^{\infty} S_{0,j} * \text{Prob.}(S_{0,j}) + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} S_{i,j} * \text{Prob.}(S_{i,j}) \tag{28}$$

where,

$$S_{0,0} = [0 * S_f] + [0 * S_s]$$
$$S_{i,0} = [i * S_f] + [0 * S_s]$$
$$S_{0,j} = [0 * S_f] + [j * S_s]$$
$$S_{i,j} = [i * S_f] + [j * S_s]$$

After plugging the values from equations (25) and (27) in equation (28) and after mathematical simplification results in expected signalling. The expected signalling as a result of HO failures and HO success can be computed as:

$$E[S_{i,j}] = \left(\frac{P_f(1-P_{no})}{P_s^2} + \frac{(P_f - P_{no} + 1)(1-P_{no})}{P_s^2} + \frac{(1-P_{no})^2}{P_s P_f}\right) S_f P(S_{0,0}) + \left(\frac{P_s(1-P_{no})}{P_f^2} + \frac{(P_s - P_{no} + 1)(1-P_{no})}{P_f^2} + \frac{(1-P_{no})^2}{P_s P_f}\right) S_s P(S_{0,0}) \tag{29}$$

Equation (29) can be used to quantify the RRC CN mobility signalling load for a mobile user. The expected signalling load can be computed by substituting the values of $P_f$, $P_s$ and $P_{no}$.

## V. Continuous Mobility Signalling Model

During a mobility HO scenario, one of the two cases can happen. Either the HO is successful or the HO is not successful. User remains in the system even in case of HO failure. For the case of continuous mobility, user generates mobility signalling as a result of HO success and HO failures while the session duration is continuous. It is assumed that the user remains RRC connected with the CBS in the system even in case of HO failure and gets connected back to DBS through RRC connection reestablishment. As the session is continuous, the probability of no HO signalling is zero in this case. The probability of failure and probability of success are complement of each other in this case. The Markov chain for this special scenario is shown in Fig. 6. Looking at the 2D Markov chain in Fig. 6. It can be inferred.

1) $P_f$ = Probability that CN mobility signalling will be generated as a result of HO failure
2) $P_h$ = Probability that CN mobility signalling will be generated as a result of HO success in case of continuous mobility
3) $CS_{i,j}$= Aggregate CN mobility signalling load on account of *i* HO failures and *j* HO successes in case of continuous mobility.

The goal is to find out the average or expected amount of CN signalling which is generated in case of continuous mobility as a result of continuous HO success and failures respectively. This probability P($CS_{i,j}$) can be calculated by solving Markov chain shown in Fig. 6. Since the amount of signalling generated by the users movement (HO failures and success) increase with time, a transition from CN signalling state $CS_{m,n}$ to $CS_{i,j}$ has zero probability when $m, n > i, j$.

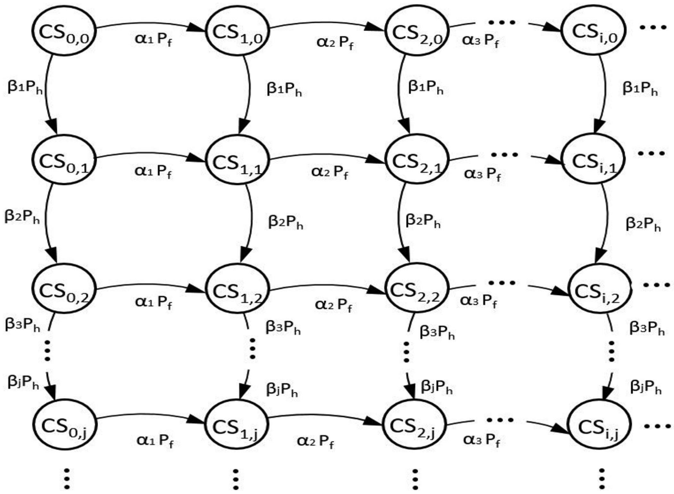



Fig. 6. Markov Chain representing continuous handover success and failure signalling scenarios.

Based on this model, Prob $(CS_{i,j})$ can be formulated as:

$$\text{Prob.}(CS_{i,j}) = \begin{cases} P(CS_{0,0}) & i=0,\ j=0 \\ \alpha_i P_f^i P(CS_{0,0}) & i>0,\ j=0 \\ \beta_j P_h^j P(CS_{0,0}) & i=0,\ j>0 \\ (i+j)\alpha_i P_f^i \beta_j P_h^j P(CS_{0,0}), & i>0,\ j>0 \end{cases} \tag{30}$$

In Lemma 1 of section IV it is already shown, for exponential cell residence and session duration the values of $\alpha$ and $\beta$ are:

$$\beta_1 = \beta_2 = \beta_3 = \beta_4 =, \ldots, = \beta_j = 1$$
$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 =, \ldots, = \alpha_i = 1$$

### A. Computation of Continuous Mobility Probability of Success ($P_h$)

The probability of failure in case of continuous mobility is the same as computed in Section III equation (21) earlier for non-continuous scenario.

$$P_f = \left(\frac{E[T_p]}{E[T_{d,r}] + E[T_p]}\right) * \left(\frac{4E[\lambda]E[v]\sqrt{\rho_1}}{\pi + 4E[\lambda]E[v]\sqrt{\rho_1}}\right)$$

Considering a continuous mobility scenario. The probability of success is the complement of probability of failure. If HO failure will not take place then it will be probability of success. The probability of HO success is given as:

$P_h = 1 - P_f$

The probability of HO success is computed as:

$$P_h = 1 - \left[\left(\frac{E[T_p]}{E[T_{d,r}] + E[T_p]}\right) * \left(\frac{4E[\lambda]E[v]\sqrt{\rho_1}}{\pi + 4E[\lambda]E[v]\sqrt{\rho_1}}\right)\right] \tag{31}$$

### B. Handover Signalling for Continuous Mobility Users

In case of continuous mobility scenario large number of HOs take place. When we consider, there are a lot of HOs successes and failures happening consistently and users have a high mobility. It requires us to compute another expression for CN signalling generated as a result of continuous HOs scenario. Let the number of HO failures is denoted by $\boldsymbol{i}$ and number of HO successes is denoted by $\boldsymbol{j}$. The expected CN signalling is given as.

$$E[CS_{i,j}] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} CS_{i,j} * \text{Prob.}(CS_{i,j}) \tag{32}$$

where,

$$CS_{0,0} = [0 * S_f] + [0 * S_s]$$
$$CS_{i,0} = [i * S_f] + [0 * S_s]$$
$$CS_{0,j} = [0 * S_f] + [j * S_s]$$
$$CS_{i,j} = [i * S_f] + [j * S_s]$$

To compute the expected signalling in case of continuous mobility, we need to find out the probability of state $CS_{0,0}$ i.e., $P(CS_{0,0})$. Looking at Fig. 6 and we know that for a Markov chain:

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \text{Prob.}(CS_{i,j}) = 1$$

Expanding the expression using Fig. 6

$$\text{Prob.}(CS_{0,0}) + \sum_{i=1}^{\infty} \text{Prob.}(CS_{i,0}) + \sum_{j=1}^{\infty} \text{Prob.}(CS_{0,j}) + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \text{Prob.}(CS_{i,j}) = 1$$

Resolving the mathematical expression to compute the value of $CS_{0,0}$

$$P(CS_{0,0}) = \frac{1}{1 + \sum_{i=1}^{\infty} \text{Prob.}(CS_{i,0}) + \sum_{j=1}^{\infty} \text{Prob.}(CS_{0,j}) + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \text{Prob.}(CS_{i,j})} \tag{33}$$

Simplifying the mathematical procedures of equation (33) above. The probability of state $CS_{0,0}$ is given as:

$$P(CS_{0,0}) = \frac{P_h P_f}{P_h P_f + P_f^2 + P_h^2 + P_f + P_h} \tag{34}$$

Now in order to compute the expected mobility signalling in case of continuous mobility we plug values from equation (34) into equation (32) and solve:

$$E[CS_{i,j}] = CS_{0,0} * \text{Prob.}(CS_{0,0}) + \sum_{i=1}^{\infty} CS_{i,0} * \text{Prob.}(CS_{i,0}) + \sum_{j=1}^{\infty} CS_{0,j} * \text{Prob.}(CS_{0,j}) + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} CS_{i,j} * Prob.(CS_{i,j})$$

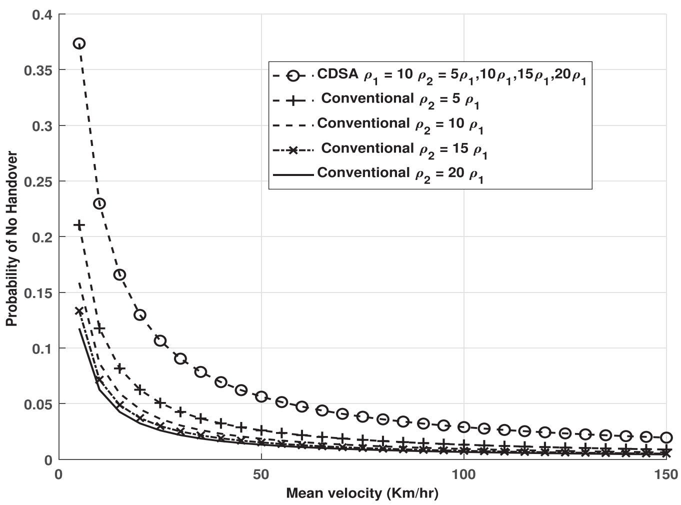



Fig. 7. Probability of no handover signalling vs. mean velocity for CDSA and different cell densities of conventional network with E[λ] = 5 mins.

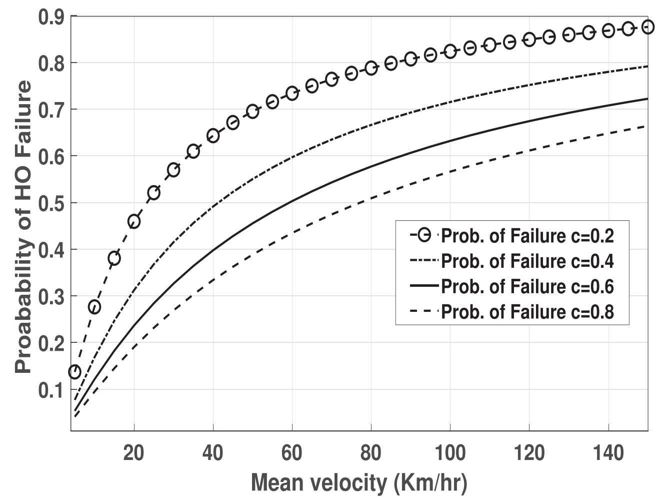



Fig. 8. Probability of HO failure vs. mean velocity for different values of coverage factor. E[λ] = 5 mins and E[$\rho$] = 10.

After solving the CN mobility signalling for continuous mobility users in case of HO successes and failures turns out to be:

$$E[CS_{i,j}] = \left(\frac{P_f}{P_h^2} + \frac{(P_f+1)}{P_h^2} + \frac{1}{P_h P_f}\right) S_f P(CS_{0,0}) + \left(\frac{P_h}{P_f^2} + \frac{(P_h+1)}{P_f^2} + \frac{1}{P_h P_f}\right) S_s P(CS_{0,0}) \quad (35)$$

The CN signalling for high mobility users depends upon $P_f$ and $P_h$. After substituting in equation (35) the values of $P_f$, $P_h$ and $P(CS_{0,0})$ from equations (21), (31) and (34) respectively, mobility signalling for continuous mobility users can be evaluated.

After comparing the two analytical equations (29) and (35) respectively (normal and continuous mobility) it is evident that in case of continuous mobility scenario the probability of no HO is zero ($P_{no} = 0$). If we substitute the value of $P_{no} = 0$ in normal mobility expected signalling equation, the two equations apparently become equal. Even though the two equations look equal for $P_{no} = 0$, it is not true mathematically as the values of $P_f$, $P_h$ and $P_s$ are different including the values of $S_{i,j}$ and $CS_{i,j}$. We perform the analysis of the normal and continuous mobility scenario in Subsection C of Section VI.

## VI. Numerical Results

### A. *Probability of Handover Signalling and Coverage Factor*

This subsection evaluates the probability of signalling in case of HO success, failure and no HO versus velocity for different values of coverage factor *(c)*. The evaluation is based on exponential distribution for session duration, cell residence time and mobility time duration. The evaluation is based on normalized densities w.r.t the CBS density. The value of *c* influences overall mobility signalling. As in Fig. 8, for low value of *c*, probability of failure signalling is high and decreases with increase in coverage factor. For low values of *c*, the HO boundary shrinks resulting in smaller values of $T_d$ as a result probability of failure signalling increases whereas for high value of *c*, the HO boundary region is suitable for HO success, therefore the probability of HO success signalling increases with coverage factor as shown in Fig. 9. Also worthy to note, for reasonable coverage factor values, at very low speeds, probability of success increases with gradual increase in speed. However with increase in mobility at higher speeds, probability of failure increase while probability of success starts to decrease as evident in Fig. 9 The probability of no HO signalling is the same for all coverage factor values and does not depend on the value of *c* but changes with velocity and cell density. In order to observe $P_{no}$ for different cell densities, the value of $P_{no}$ is shown in Fig. 7. Probability of no HO signalling has highest value for CDSA and it decreases with increase in cell density for conventional architecture. Probability of failure is lowest for CDSA versus conventional architecture while probability of success starts low for CDSA but with increase in mean velocity it is higher than conventional

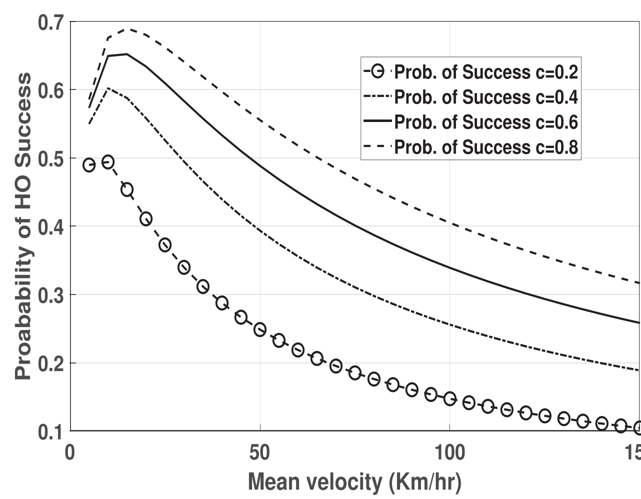



Fig. 9. Probability of HO success vs. mean velocity for different values of coverage factor. E[λ] = 5 mins and E[$\rho$] = 10.

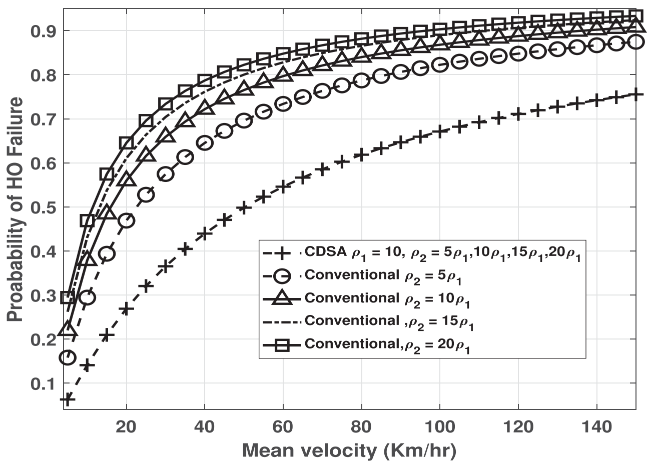



Fig. 10. Probability of HO failure vs. mean velocity for different values of cell density. E[λ] = 5 mins and c = 0.5.

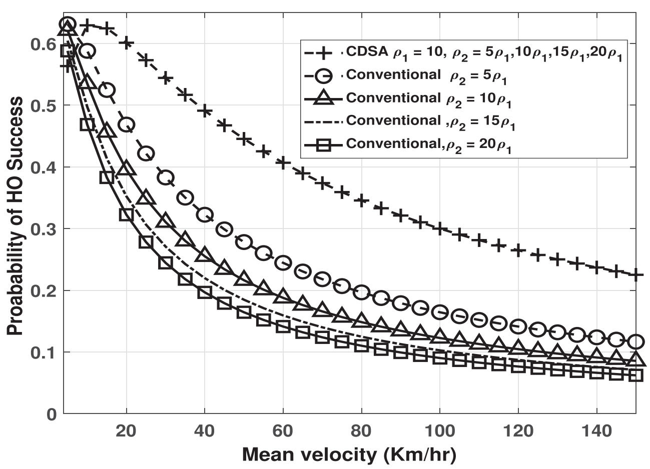



Fig. 11. Probability of HO success vs. mean velocity for different values of cell density. E[λ] = 5 mins and c = 0.5.

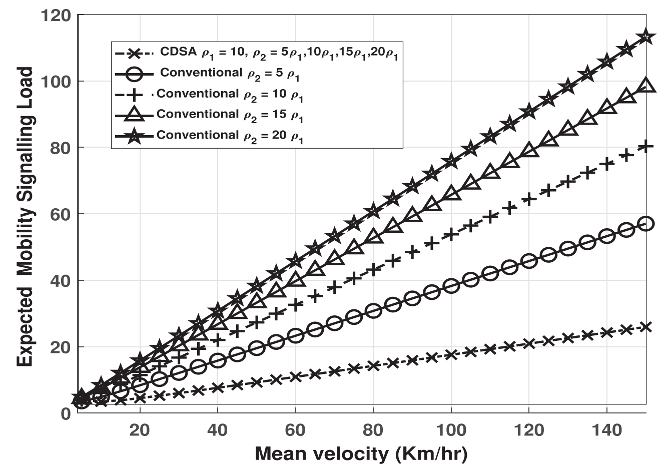



Fig. 12. Normalized expected signalling load vs. mean velocity for c = 0.1 and E[λ] = 5 mins.

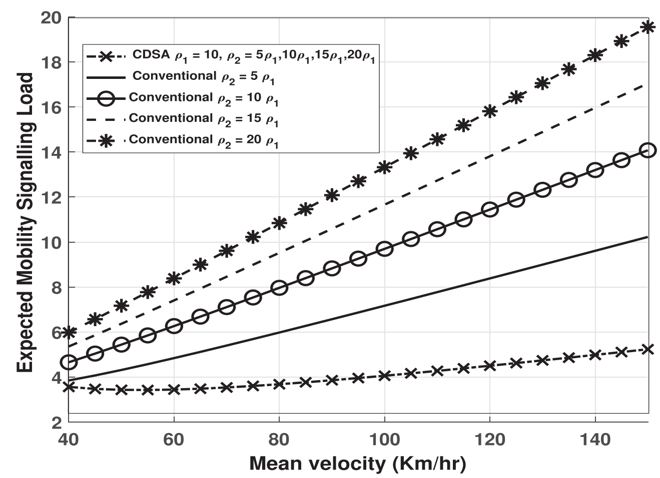



Fig. 13. Normalized expected signalling load vs. mean velocity for c = 0.6 and E[λ] = 5 mins at high speeds.

architecture as shown in Fig. 10 and Fig. 11 respectively for a value of $c = 0.5$.

## *B. Signalling in CDSA Versus Conventional Networks*

In this subsection we evaluate how much expected CN mobility signalling is generated in case of CDSA versus conventional networks as proposed in analytical model of Section IV. We consider exponential distribution for session duration, cell residence and mobility duration time. The evaluation is based on normalized densities w.r.t CBS density. In addition the RRC signalling load (in terms of expected value) is normalized with S ( more specifically $S_f$ for HO failures and $S_s$ for HO success). Fig. 12 shows the normalized expected mobility signalling load vs. mean velocity for coverage factor $c = 0.1$ while Fig. 13 provides this information for coverage factor value of 0.6. With low values of coverage factor there is a high probability of HO failure and increase CN mobility signalling load, even at slow speeds. Fig. 12 indicates CDSA generates $\{31, 54, 72, 87\}$*S times less signalling load compared to conventional network with different cell densities. For a high coverage factor expected mobility signalling load reduces. With increase in medium and high mobility speeds, probability of failure increase, so expected mobility signalling is supposed to increase with increase in mobility. Fig. 13 shows that CDSA results in $\{5, 9, 12, 14\}$*S times less signalling load vs conventional networks even at high velocity and coverage factor respectively. CDSA is a clear winner for generating less mobility signalling load. These plots suggests that CDSA performs equally better at greater mobility and high speed scenarios. This proves our initial hypothesis that in case of ultra dense networks CDSA deployment is beneficial whereas conventional networks results in excessive mobility signalling load.

In CDSA a continuous and reliable coverage layer is provided by CBS, where the large footprint ensures robust connectivity and mobility. Whereas the data plane is supported by flexible, adaptive, high capacity and energy efficient DBSs, that provide data transmission along with the necessary signalling as shown in Fig. 2 in Section I. Whereas in conventional network, network

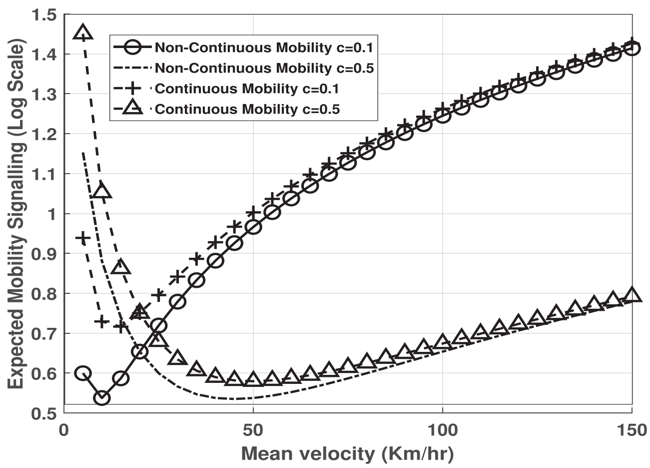



Fig. 14. Expected Mobility Signalling vs. mean velocity in case of normal and continuous mobility for c = 0.1 and c = 0.5.

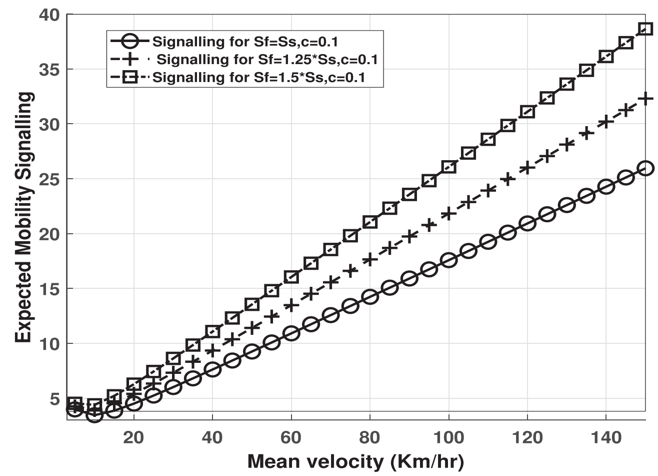



Fig. 15. Expected mobility signalling comparison of handover failure and success for c = 0.1. E[$\rho$] = 10 indicating how handover failure results in more mobility signalling.

remains on all the time and signalling and data connectivity operations are controlled by the eNodeB alone. Every single HO generates CN signalling which adds load on the network elements and increases delay. In case of CDSA for any HO between DBS to DBS is transparent to the CN and does not generate any CN signalling. This saves a lot of capacity and resources of the CN. The only time when CN mobility signalling is generated in CDSA is when a HO takes place between CBS to CBS.

### *C. Continuous Mobility Versus Noncontinuous Mobility*

In this subsection we compare the expected non-continuous CN mobility signalling versus continuous CN mobility signalling as derived analytically in Section IV and V. Fig. 14 indicates non-continuous and continuous mobility signalling for $c = 0.1$ and $c = 0.5$ respectively. In both scenarios, continuous mobility signalling is much higher compared to non-continuous mobility at low speeds. For non-continuous mobility, as velocity increases probability of no HO signalling approaches zero.

From the numerical comparison of expected normal and continuous mobility signalling in the Fig. 14 it is evident that continuous mobility signalling provides the upper bound for expected signalling generated as a result of HO. $P_{no}$ is zero at all the times for continuous scenario. For normal scenario, at low speeds $P_{no}$ is not equal to zero. However, with increase in velocity $P_{no}$ starts approaching zero. This is evident from expected signalling generated at high velocities is the same both in case of normal and continuous mobility scenarios. This confirms in order to compute upper limit for mobility signalling in any case, continuous mobility scenario can be used.

### *D. Quantification of Handover Failure Signalling*

A typical HO procedure consists of three phases preparation, execution and completion phases [9]. In this study for the current system model, the user gets connected back to the system through RRC connection re-establishment in case of HO failure. It must be kept in mind, HO failure can take place at either of the preparation, execution and completion phase(s). In case when a HO failure takes place. Then UE has to go through connection re-establishment procedure once again in order to get connected with a DBS. Numerical computation of HO signalling considering each message and processing at different nodes is computed in [35], [36], [44] and [45]. In order to approximate, how much additional signalling is generated in case of HO failure. Consider, if HO failure takes place during HO completion phase, then RRC re-establishment will take place to keep the user in the system after HO failure. This procedure results in more signalling messages compared to HO success alone as shown in Fig. 4. HO failure signalling is normalized with $S_f$ and HO success signalling is normalized with $S_s$. The total HO signalling (failure and success signalling) normalized by S is given as follows:

$$S = S_f + S_s \tag{36}$$

Looking at Fig. 4, we can write $S_f$ in terms of $S_s$

$$S_f = S_s + 0.25 * S_s$$
$$S_f = 1.25 * Ss$$

With reference to Fig. 4, therefore total signalling is:

$$S = 1.25 * S_s + S_s$$
$$S = 2.25 * S_s \tag{37}$$

This indicates that HO failure signalling load has different quantitative evaluation than HO success signalling load. Differences in HO failure signalling load depend upon the scenario(s) considered. Fig. 15 provides information about increase in expected signalling load for different $S_f$ values for a coverage factor $c = 0.1$. It shows that for the given scenario considered, HO failure signalling load results in almost 1.6 times more normalized expected signalling load.

## VII. CONCLUSION

In this work, we developed an analytical model to quantify the expected mobility signalling load generated in cellular networks as a result of HO success and HO failures. Closed form expressions were developed for probability of HO failure, HO success and no HO signalling. We also identified the analytical evaluation of overall CN mobility signalling load for various HO scenarios. The analytical framework presented was used to assess the advantage of CDSA over conventional network architecture using exponential cell residence time, exponential session duration time and exponential mobility time duration respectively. Analytical evaluation is presented for continuous and non-continuous mobility scenarios. We introduced new mobility parameter(s) which affect mobility signalling. With proper settings it can help in reducing mobility signalling load in ultra-dense networks. Results indicate that coverage factor *(c)*, and velocity directly affect the mobility signalling load. From the results, it is clear that CDSA is a clear winner when it comes to ultra dense HetNet deployment.

## Acknowledgment

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation.

## References

[1] C. Inc., "Cisco visual netnetwork index: Global mobile data traffic forecast update, 2016–2021," Tech. Rep., 2017.

[2] Y. Qiao, Y. Cheng, J. Yang, J. Liu, and N. Kato, "A mobility analytical framework for big mobile data in densely populated area," *IEEE Trans. Veh Technol.*, vol. 66, no. 2, pp. 1443–1455, Feb. 2017.

[3] Y. Niu *et al.* "Energy-efficient scheduling for mmwave backhauling of small cells in heterogeneous cellular networks," *IEEE Trans. Veh Technol.*, vol. 66, no. 3, pp. 2674–2687, Mar. 2017.

[4] Y. Niu, Y. Li, D. Jin, L. Su, and D. Wu, "Blockage robust and efficient scheduling for directional mmwave wpans," *IEEE Trans. Veh. Technol.*, vol. 64, no. 2, pp. 728–742, Feb. 2015.

[5] 3GPP, "Mobility Enhancements in Heterogeneous Networks," 3GPP TR 36.839, 2012.

[6] Study on Small Cell Enhancements for E-UTRA and E-UTRAN: Higher Layer Aspects, 3GPP TR 36.842, 2015.

[7] M. B. A. Anpalagan and R. Vanithamby, *Design and Deployment of Small Cell Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[8] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA): Requirements for Support of Radio Resource Management," 3GPP TS 36.133, 2015.

[9] A. Helenius, "Performance of handover in long term evolution," MS thesis, School Elect. Eng., Aalto University, Helsinki, Finland, 2011.

[10] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[11] A. Mohamed, O. Onireti, M. A. Imran, A. Imran, and R. Tafazolli, "Predictive and core-network efficient RRC signalling for active state handover in RANs with control/data separation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1423–1436, Mar. 2017.

[12] 3GPP, "Study on Small Cell Enhancements for E-UTRA and E-UTRAN: Higher Layer Aspects," 3GPP TR 36.842, 2013. [Online]. Available: http://www.3gpp.org/DynaReport/36842.htm

[13] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA): Mobility Enhancements in Heterogeneous Networks," 3GPP TR 36.839, Dec. 2012. [Online]. Available: http://www.3gpp.org/DynaReport/36839.htm

[14] Y. K. H. Ishii and H. Takahashi, "A novel architecture for lte-b c-plane/u-plane split and phantom cell concept," *IEEE Globecom Workshops*, 2012, pp. 624–630.

[15] A. Mohamed, O. Onireti, M. A. Imran, A. Imran, and R. Tafazolli, "Control-data separation architecture for cellular radio access networks: A survey and outlook," *IEEE Commun. Surveys Tut.*, vol. 18, no. 1, pp. 446–465, Jan.–Mar. 2016.

[16] I. G. *et al.*, "Earth project deliverable d3.3: Final report on green network technologies," 2012. [Online]. Available: http://bscw.ict-earth.eu/pub/bscw.cgi/d70472/EARTH WP3 D3.3.pdf

[17] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); and Evovled Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description; Stage 2," 3GPP TS 36.300, 2014.

[18] L. Yan, X. Fang, and Y. Fang, "Control and data signaling decoupled architecture for railway wireless networks," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 103–111, Feb. 2015.

[19] Y. Li, F. Xuming, and Y. Fang, "A novel network architecture for C/U-plane staggered handover in 5G decoupled heterogeneous railway wireless systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 12, pp. 3350–3362, Feb. 2017.

[20] H. Song, X. Fang, and L. Yan, "Handover scheme for 5g C/U plane split heterogeneous network in high-speed railway," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4633–4646, Nov. 2014.

[21] X. Ge, B. Du, Q. Li, and D. S. Michalopoulos, "Energy efficiency of multiuser multiantenna random cellular networks with minimum distance constraints," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1696–1708, Feb. 2017.

[22] R. Balakrishnan and B. Canberk, "Traffic-aware QoS provisioning and admission control in OFDMA hybrid small cells," *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 802–810, Feb. 2014.

[23] S. Bu, F. R. Yu, and H. Yanikomeroglu, "Interference-aware energy-efficient resource allocation for OFDMA-based heterogeneous networks with incomplete channel state information," *IEEE Trans. Veh. Technol.*, vol. 64, no. 3, pp. 1036–1050, Mar. 2015.

[24] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, "Channel-access-aware user association with interference coordination in two-tier downlink cellular networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5579–5594, Jul. 2016.

[25] F. Guidolin, I. Pappalardo, A. Zanella, and M. Zorzi, "Context-aware handover policies in HetNets," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1895–1906, Mar. 2016.

[26] H. Ibrahim, H. ElSawy, U. T. Nguyen, and M. S. Alouini, "Mobility-aware modeling and analysis of dense cellular networks with $c$-plane/ $u$-plane split architecture," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4879–4894, Nov. 2016.

[27] R. Arshad, H. ElSawy, S. Sorour, T. Y. Al-Naffouri, and M. S. Alouini, "Velocity-aware handover management in two-tier cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1851–1867, Mar. 2017.

[28] "Signaling is growing 50 percent faster than data traffic," *Nokia Siemens Netw.*, 2012. [Online]. Available: https://docplayer.net/6278117-Signaling-is-growing-50-faster-than-data-traffic.html.

[29] X. Xu, G. He, S. Zhang, Y. Chen, and S. Xu, "On functionality separation for green mobile networks: concept study over LTE," *IEEE Commun. Mag.*, vol. 51, no. 5, pp. 82–90, May 2013.

[30] C. H. K. S. Liu, J. Wu and V. Lau, "A 25 Gb/s(/km2) urban wireless network beyond IMT-advanced," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 122–129, Feb. 2011.

[31] A. Capone, A. F. dos Santos, I. Filippini, and B. Gloss, "Looking beyond green cellular networks," in *Proc. 9th Annu. Conf. Wireless On-Demand Netw. Syst. Serv.*, Jan. 2012, pp. 127–130.

[32] J. Zhang, J. Feng, C. Liu, X. Hong, X. Zhang, and W. Wang, "Mobility enhancement and performance evaluation for 5G ultra dense networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2015, pp. 1793–1798.

[33] 3GPP Nokia Siemens Networks, "Mobility Statistics for Macro and Small Cell Dual-Connectivity Cases," 3GPP Nokia Siemens Networks, Chicago, IL, USA, techreport, Apr. 2013. [Online]. Available: http://www.3gpp.org/DynaReport/TDocExMtgR2-81b30048.htm

[34] 3GPP Huawei, "Feasible Scenarios and Benefits of Dual-Connectivity in Small Cell Deployment," 3GPP TSG-RAN WG2 Meeting 81, 3GPP Huawei, HiSilicon, St. Julian's, Malta, Feb. 2013. [Online]. Available: Available:http://www.3gpp.org/DynaReport/TDocExMtgR2-8130047.htm

[35] L. Wang, Y. Zhang, and Z. Wei, "Mobility management schemes at radio network layer for LTE femtocells," in *Proc. VTC Spring IEEE 69th Veh. Technol. Conf.*, Apr. 2009, pp. 1–5.

[36] H. Zhang, W. Ma, W. Li, W. Zheng, X. Wen, and C. Jiang, "Signalling cost evaluation of handover management schemes in LTE-advanced femtocell," in *Proc. IEEE 73rd Veh. Technol. Conf.*, May 2011, pp. 1–5.

[37] 3GPP, "General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access," 3GPP, TR 23.401, 2017.

[38] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and NR; Multi-Connectivity; Stage 2 (release 15)," 3GPP TS 37.340, 2017.

[39] A. Mahbas, H. Zhu, and J. Wang, "The role of inter-frequency measurement in offloading traffic to small cells," in *Proc. IEEE 85th Veh. Technol. Conf.*, Jun. 2017, pp. 1–4.

[40] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); Mobility Enhancements in Heterogeneous Networks," 3GPP TR 36.839, Dec 2012.
[41] H. Xie, S. K. Xie, and S. Kuek, "Priority handoff analysis," in *Proc. IEEE Veh. Technol. Conf.*, 1993, pp. 855–858.
[42] V. Luarini, "From symmetry breaking to Poisson point process in 2D Voronoi tessellations: The generic nature of hexagons," *Statist. Phys.*, vol. 130, no. 6, pp. 1047–1062, 2008.
[43] K. Dimou *et al.*, "Handover within 3GPP LTE: Design principles and performance," in *Proc. IEEE 70th Veh. Technol. Conf. Fall*, Sep. 2009, pp. 1–5.
[44] 3GPP, "Feasibility Study for Evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN)," 3GPP, TR 25.912, 2017.
[45] Z. Li and W. Mick, "User plane and control plane separation framework for home base stations," *Fujitsu Scientific Tech. J.*, vol. 46, no. 1, pp. 79–86, 2010.

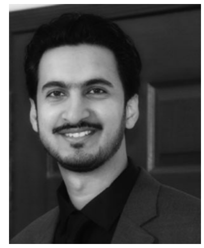
**Azar Taufique** received the B.Sc. (Hons.) degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan in 2007, the masters' degree in telecommunications engineering from the University of Texas at Dallas, Richardson, TX, USA in 2010, and the Ph.D. degree in electrical and computer engineering from the University of Oklahoma, Norman, OK, USA, in 2018.

He was a Telecommunications Technical Trainer teaching 4G LTE courses to more than 2500 professionals with Verizon, T-Mobile, Sprint, Ericsson, AT&T, etc. in North America. His telecommunications blog is read by hundreds of ICT professionals per week. He is currently a Technical Training Program Manager with Amazon Inc. and a Technical Trainer with TechTrained.com LLC. His research interests include mobility management and new architectures for mobility management support in 5G.

**Abdelrahim Mohamed** (S'15–M'16) received the B.Sc. degree (first class) in electrical and electronics engineering from the University of Khartoum, Khartoum, Sudan, in 2011, the M.Sc. degree (distinction) in mobile and satellite communications and the Ph.D. degree in electronics engineering from the University of Surrey, Guildford, U.K., in 2013 and 2016, respectively.

He is currently a Postdoctoral Research Fellow with ICS/5GIC, the University of Surrey. He is currently involved in the RRM, MAC, and RAN management work area, and the new physical layer work area at the 5GIC, University of Surrey. He is involved in the Energy Proportional eNodeB for LTE-Advanced and Beyond Project, the Planning Tool for 5G Network in mm-wave band project, in parallel to working on the development of the 5G system level simulator. His research contributed to the QSON project, and the FP7 CoRaSat Project. His main areas of research interest include radio access network design, system-level analysis, mobility management, energy efficiency, and cognitive radio. He secured first place and top ranked in the Electrical and Electronic Engineering Department, University of Surrey during his M.Sc. studies.

Dr. Mohamed was a recipient of the Sentinels of Science 2016 Award.



**Hasan Farooq** received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2009, and the M.Sc. by Research degree in information technology from Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia, in 2014 wherein his research focused on developing *ad hoc* routing protocols for smart grids. He is currently working toward the Ph.D. degree in electrical and computer engineering at the University of Oklahoma, Norman, OK, USA.

His research interests include big data empowered proactive self-organizing cellular networks focusing on intelligent proactive self-optimization and self-healing in HetNets utilizing dexterous combination of machine learning tools, classical optimization techniques, stochastic analysis, and data analytics. He has been involved in multinational QSON project on self organizing cellular networks (SON) and is currently contributing to two NSF funded projects on 5G SON.

Mr. Farooq is the recipient of Internet Society First Time Fellowship Award toward IETF 86th meeting held in the USA in 2013.

**Ali Imran** (M'05) received the M.Sc. and Ph.D. from University of Surrey, Guildford, U.K. He is the Founding Director of Big Data Enabled Self-Organizing Networks Research Lab, University of Oklahoma, Norman, OK, USA. His current research interests include self-organizing wireless Networks (SON) functions design for enabling 5G, big data enabled SON, artificial intelligence enabled wireless networks, new RAN architectures for enabling low-cost human-to-human as well as IoT and D2D communications. On these topics, he has authored or co-authored more than 65 refereed journal and conference papers. He has given tutorials on these topics at several international conferences including IEEE ICC, WF-IoT, PIMRC, WCNC, CAMAD, and European Wireless and Crowncom. He is currently Principle Investigator on several multinational research projects focused on next generation wireless networks, for which he has secured research grants of more than 3 million USD. He is an Associate Fellow of Higher Education Academy, U.K., President of ComSoc Tulsa Chapter, Member of Advisory Board for Special Technical Community on Big Data at IEEE Computer Society; Board Member of ITERA, and Associate Editor of IEEE ACCESS special section on heterogeneous networks.

**Rahim Tafazolli** (SM'09) received the Ph.D. from University of Surrey, Guildford, U.K. He is a Regius Professor and the Director with the Institute for Communication Systems and the 5G Innovation Centre, University of Surrey, Guildford U.K. He has authored or coauthored more than 500 research papers in refereed journals, international conferences and has been an Invited Speaker as well. He is the Editor for two books on technologies for wireless future published by Wiley's vol. 1 in 2004 and vol. 2 2006. He was appointed the Fellow of Wireless World Research Forum in April 2011, in recognition of his personal contribution to the wireless world, as well as heading one of Europe's leading research groups.