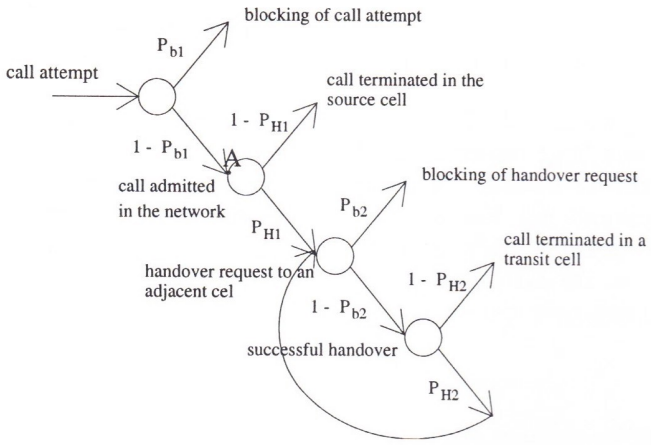

Figure 3. The handover generation process during call lifetime in the case that mobile sojourn times in subsequent cells are iid random variables.

respect to the cell. In our case we have $E[x_2] = 3\sqrt{3}R_1/4$, for hexagonal cells, and $E[x_2] = R_2$, for square cells. Finally, the following results are obtained for n_{h0}:

$$n_{h0} = \begin{cases} \dfrac{T_m V_{trk}}{R_1} \dfrac{4}{3\sqrt{3}} & \text{for hexagonal cells,} \\ \dfrac{T_m V_{trk}}{R_2} & \text{for square cells} \end{cases} \tag{19}$$

(handovers/call). If we assume equal areas for both cell shapes (this is true if $R_2 = \sqrt{3\sqrt{3}/2}R_1$), the hexagonal cell entails a 24% increase in the number of handover requests per call with respect to the square cell.

4.7. Average number of handover requests per call in the presence of blocking

Let us assume iid mobile sojourn times in subsequent cells and a homogeneous system (see section 3). In this case, the handover generation process during call lifetime is memoryless. Figure 3 shows the handover generation diagram for a call in the presence of blocking (i.e., $P_{b1} > 0$ and $P_{b2} > 0$). The distribution of the number of handover requests per call attempt is obtained as explained below.

- A call does not generate handover requests provided that either it is initially blocked or it ends in its source cell. This occurs with probability

$$Q_0 = P_{b1} + (1 - P_{b1})(1 - P_{H1}) = 1 - P_{H1}(1 - P_{b1}).$$

- A call generates one handover request if it is not initially blocked, and if only one handover is required (i.e., either this handover is unsuccessful or the call terminates in the first transit cell). This occurs with probability

$$\begin{aligned} Q_1 &= (1 - P_{b1})P_{H1}\left[P_{b2} + (1 - P_{b2})(1 - P_{H2})\right] \\ &= (1 - P_{b1})P_{H1}\left[1 - P_{H2}(1 - P_{b2})\right]. \end{aligned}$$

- A call generates k handover requests (with $k \geqslant 1$) if it is not initially blocked, if $k - 1$ handovers are successfully accomplished, if a further handover is requested and if no other handover is performed. This occurs with probability

$$\begin{aligned} Q_k &= (1 - P_{b1})P_{H1}(1 - P_{b2})\left[P_{H2}(1 - P_{b2})\right]^{k-2} \\ &\quad \times P_{H2}\left[P_{b2} + (1 - P_{b2})(1 - P_{H2})\right] \\ &= (1 - P_{b1})P_{H1}\left[P_{H2}(1 - P_{b2})\right]^{k-1} \\ &\quad \times \left[1 - P_{H2}(1 - P_{b2})\right]. \end{aligned}$$

On the basis of the distribution Q_k we obtain the average number of handover requests per call attempt, n_h, as follows:

$$\begin{aligned} n_h &= \sum_{k=1}^{\infty} kQ_k \\ &= \frac{P_{H1}(1 - P_{b1})}{1 - (1 - P_{b2})P_{H2}} \frac{\text{handover requests}}{\text{call attempt}}. \end{aligned} \tag{20}$$

It is straightforward to verify that by setting $P_{b1} = P_{b2} = 0$ in (20) we obtain (18). The effect of the blocking (i.e., $P_{b1} > 0$ and $P_{b2} > 0$) is that the average number of handover requests per call, n_h, decreases with respect to n_{h0}.

If we remove the assumption of iid mobile sojourn times in subsequent cells, we must evaluate n_h conditioned on a given mobile user trajectory and then remove the conditioning by using the trajectory probability distribution[4].

4.8. Relationship between the average handover rate and the average new call arrival rate for a cell

According to the basic assumptions made in section 3, we have a uniform traffic: λ denotes the average arrival rate for new call attempts in a given cell y, whereas λ_h is the mean arrival rate for handed-over calls in cell y. We consider that an equilibrium exists (figure 4), in any time interval, between the average number of calls that enter cell y and the average number of calls that leave cell y towards adjacent cells (*flow conservation*) [7,8]. The mean handover rate due to calls which leave a given cell y is obtained as the sum of the two following contributions:

- $\lambda(1 - P_{b1})P_{H1}$, which represents the mean rate of calls originated in cell y that leave cell y towards an adjacent cell;

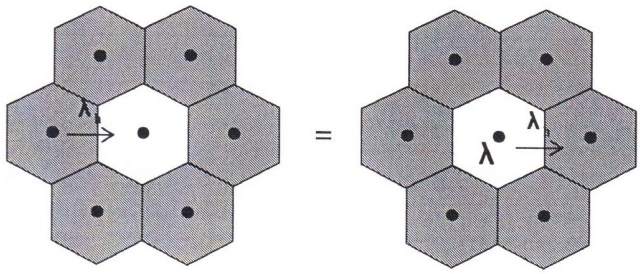

Figure 4. The equilibrium between calls that enter a cell and calls that go out from a cell.

[4] Under the assumptions made in section 3, all user trajectories are parallel and equally likely.

- $\lambda_h(1 - P_{b2})P_{H2}$, which represents the mean rate of calls arrived in cell y by handovers that leave cell y towards an adjacent cell.

The equilibrium condition requires that the sum of these two contributions is equal to the mean rate of handed-over calls towards cell y, that is λ_h. Hence, we have

$$\lambda_h(1 - P_{b2})P_{H2} + \lambda(1 - P_{b1})P_{H1} = \lambda_h. \tag{21}$$

From (21), we obtain λ_h/λ as

$$\frac{\lambda_h}{\lambda} = \frac{(1 - P_{b1})P_{H1}}{1 - (1 - P_{b2})P_{H2}}. \tag{22}$$

According to (22), the average rate of handover requests towards a cell, λ_h, depends on both the mean rate of new call attempts, λ, the handover probabilities P_{H1} and P_{H2}, and the blocking probabilities P_{b1} and P_{b2}. This entails a feedback in the loss queuing system used to model the behavior of a generic cell with the BCC policy: the blocking probabilities P_{b1} and P_{b2} depend on the total arrival process in a cell (i.e., handover requests plus new call attempts) and this arrival process depends, in turn, on the blocking probabilities. Hence, analytical derivations of the blocking probabilities need a recursive approach [7,8,11,39]. An example is shown in section 7.

If we make the additional assumption of iid mobile sojourn times in subsequent cells, from (20) and (22) we obtain the following relationship between n_h and λ_h/λ:

$$n_h = \frac{\lambda_h}{\lambda}. \tag{23}$$

Equation (23) represents a sort of *ergodicity condition* for the handover generation process: on the left side we have a parameter related to a generic call that is equal on the right side to a quantity related to a generic cell. This formula must not surprise, since it has been derived under the assumption of a homogeneous system and memoryless mobility conditions (i.e., all the cells have the same traffic, the same shape and size, the same mobility characteristics).

If we remove the assumption of iid mobile sojourn times in subsequent cells, we can still use (23) as a first approximation. It is important to note that there is a case where (23) is exact even if mobile sojourn times in subsequent cells are not independent: $P_{b1} = P_{b2} = 0$ (in such a case (17) and (22) become equal).

4.9. Call dropping probability

We consider the basic assumptions made in section 3 and the additional assumption of iid mobile sojourn times in subsequent cells. To study the call dropping event, the handover diagram shown in figure 3 must be considered starting from point A (i.e., a call accepted into the network). A call in progress is dropped at the kth handover request if the two following independent events occur [7]:

- a call lasts so as to produce at least k handover requests ($k = 1, 2, \ldots$); this occurs with probability $P_{H1}P_{H2}^{k-1}$;
- a call accepted into the system is dropped with probability $P_{b2}(1 - P_{b2})^{k-1}$ at the kth handover.

The call dropping probability P_{drop} is obtained as the sum of the probabilities that a call is dropped at the kth handover for k from 1 to infinity:

$$P_{drop} = \sum_{k=1}^{\infty} P_{H1}P_{H2}^{k-1}P_{b2}(1 - P_{b2})^{k-1} = \frac{P_{H1}P_{b2}}{1 - P_{H2}(1 - P_{b2})}. \tag{24}$$

If we remove the assumption of iid mobile sojourn times in subsequent cells, (24) is not generally applicable. In such a case, we must evaluate P_{drop} conditioned on a given mobile user trajectory and then remove the conditioning by using the trajectory probability distribution.

4.10. Grade Of Service

Several ways are possible to define the *Grade Of Service* (GOS). A first possibility, denoted by GOS_1, is to consider probabilities P_{b1} and P_{b2} weighed by the relative percentages of arrivals:

$$GOS_1 \triangleq \frac{\lambda}{\lambda + \lambda_h}P_{b1} + \frac{\lambda_h}{\lambda + \lambda_h}P_{b2}. \tag{25}$$

Parameter GOS_1 (see also [18,22] for a similar definition) will be used in the theoretical study made in section 7, since GOS_1 is the *call congestion* for the loss queuing system which models a cell with heterogeneous input traffic (due to new call attempts and handover requests). Under the assumption of iid mobile sojourn times in subsequent cells, we can use (23) in (25).

Another GOS definition could be to consider that the dropping of a call is a more frustrating event than the blocking of a new call attempt. Accordingly, we consider GOS_2 as follows:

$$GOS_2 \triangleq P_{b1} + 10P_{drop}. \tag{26}$$

GOS_2 weights P_{drop} 10 times more than P_{b1}: this parameter is not a probability, but may express the QoS perceived by users [17]. Of course, the higher the GOS_2 value the poorer the QoS provided to users. We will use the GOS_2 parameter to evaluate the impact of different user mobility conditions on the performance of a channel allocation scheme.

5. Performance evaluation

In this section, we derive the performance of a LEO-MSS with a specific cell shape. In addition to the assumptions made in section 3, we consider (figure 5):

- The cells (i.e., footprints of the antenna spot-beams from satellites) of the network are disposed on the Earth according to a hexagonal layout and have a square shape (with side $2R$). The use of this cell shape entails simpler

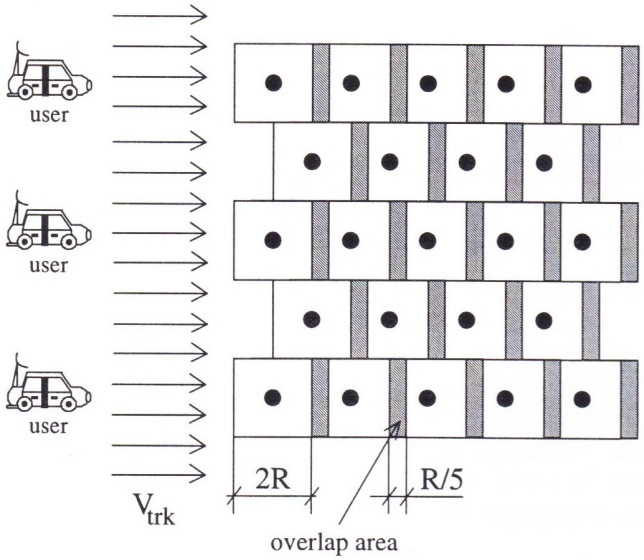

Figure 5. Illustration of the mobility model for LEO-MSSs.

analytical derivations to characterize the user mobility parameters (e.g., P_{H1}, P_{H2}, etc.).

- Mobile users cross the cellular network with a relative velocity V_{trk}, disposed with respect to cell sides, as shown in figure 5. Hence, denoting by $d_c(y)$ the distance crossed in cell y by a mobile user from the arrival instant of its call in cell y (this cell can be either the source cell or a transit one), we have[5]:
  - $d_c(y)$ is uniformly distributed between 0 and $2R$, if cell y is the source cell of the call;
  - $d_c(y)$ is deterministically equal to $2R$, if cell y is a transit cell for the call.
- The time to cross the overlap area, t_{wmax}, has a deterministic value, equal for any handover request, that is obtained according to the following formula:

$$t_{wmax} = \frac{R}{5V_{trk}}. \tag{27}$$

  According to (1), S is equal to 0.1 (conservative assumption).
- New calls originated in an overlap area between adjacent cells are immediately addressed to the destination cell in order to avoid that they immediately need to be handed-over.

This model has been adopted in the SAINT project (SATellite INTegration in the Future Mobile Network) [3] within the framework of RACE II, a European Commission financed research program. This is a one-dimensional mobility model that is also suitable for linear cellular networks used for highways and railways. Parameter α is obtained as the ratio between R and $T_m V_{trk}$. We have the following distributions for t_{mc1} and t_{mc2}:

$$f_{t_{mc1}}(t) = \frac{V_{trk}}{2R}\left[u(t) - u\left(t - \frac{2R}{V_{trk}}\right)\right], \tag{28}$$

$$f_{t_{mc2}}(t) = \delta\left(t - \frac{2R}{V_{trk}}\right). \tag{29}$$

These distributions represent a very special case which fulfills the excess life theorem (8). In addition to this, the mobility assumptions guarantee that the mobile sojourn time in a cell does not depend on that of the previous cells, i.e., mobile sojourn times in subsequent cells are iid. Then, the handover generation process is memoryless. The analytical derivations obtained in section 4 can be applied to this mobility case. In particular, on the basis of (10), (28) and (29) we obtain the following expressions for P_{H1} and P_{H2}:

$$P_{H1}(\alpha) = \frac{1 - e^{-2\alpha}}{2\alpha}, \qquad P_{H2}(\alpha) = e^{-2\alpha}. \tag{30}$$

It is worth noting that P_{H1} and P_{H2} are functions of the mobility parameter α: as α decreases to 0 (or increases to ∞), that is the mobility increases (decreases), both P_{H1} and P_{H2} approach 1 (0). Hence, from (19) and (30), n_{h0} results in

$$n_{h0} = \frac{1}{2\alpha}\,\frac{\text{handovers}}{\text{call}}. \tag{31}$$

By assuming a fixed value for T_m, the user mobility increases (i.e., n_{h0} increases) if V_{trk} increases and/or R decreases. Then, in general, we can consider that the mobility increases if the satellite altitude decreases; correspondingly, the number of satellites of the LEO constellation increases as well [24].

For LEO-MSSs, α values less than unity are expected. In particular, for the IRIDIUM system [7,9,10], we may consider V_{trk} = 26,600 km/h and R = 212.5 km; then, $\alpha \approx 0.16$ if T_m = 3 min. Correspondingly, $P_{H1} \approx 85\%$ and $P_{H2} \approx 72\%$. From (31), about 3.125 spot-beam handovers are required, on average, during call lifetime. Finally, on the basis of (27), t_{wmax} is about equal to 0.1 min.

In figure 6, the behaviors of P_{b1}, P_{drop} and GOS_2 have been shown as functions of parameter α. The results given in this figure have been obtained for a Fixed Channel Allocation scheme with the Queuing of the Handovers that cannot be immediately served (FCA-QH) [7,34] by simulating a parallelogram shaped cellular network folded onto itself as shown in [9,10]. Simulations have been carried out under the conditions listed below:

- the average call duration T_m is equal to 3 min;
- the maximum queuing time for handover requests t_{wmax} is obtained from (27);
- new call attempts arrive at a cell with a mean rate λ equal to 1.67 calls/min;

[5] The distance $d_a(y)$, crossed by a user in cell y while its call is active (irrespective of y being the source cell or a transit one), is equal to $d_c(y)$, except when y is the termination cell of the call (i.e., the cell where either the call naturally ends or it is dropped at cell boundaries for an unsuccessful handover towards an adjacent cell), because $d_a(y) \leqslant d_c(y)$, if the call naturally ends in cell y. In general we have: $d_a(y) = \min[d_c(y),\ t_d V_{trk}]$; this formula can be used to obtain the distribution of $d_a(y)$ from those of $d_c(y)$ and t_d, by following a similar approach to (12).

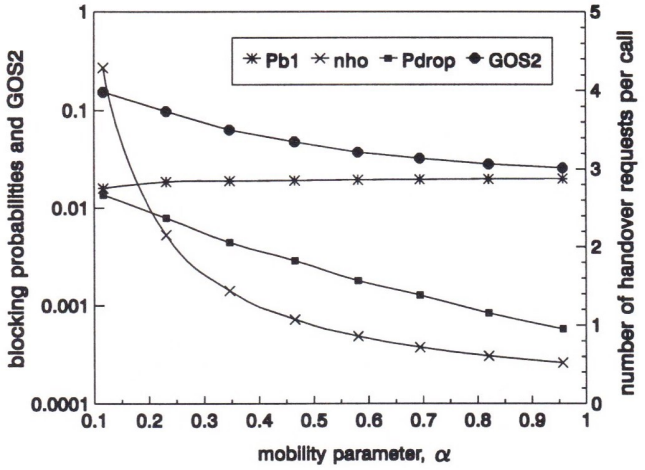

Figure 6. Behaviors of P_{b1}, P_{drop}, GOS_2 and n_{h0} for the FCA-QH technique as functions of parameter α in LEO-MSSs.

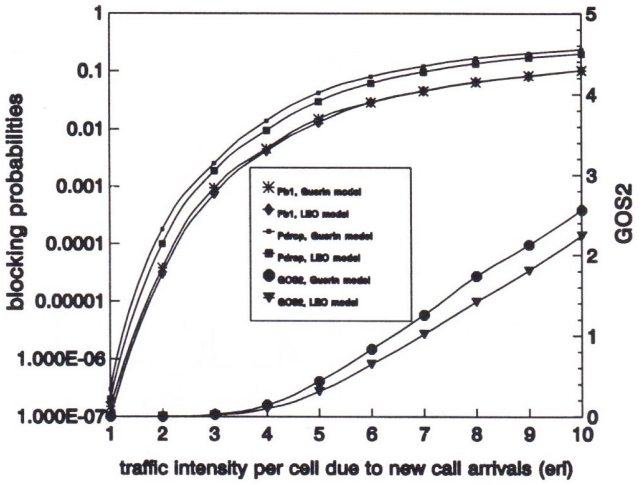

Figure 7. Comparison between the Guérin's mobility model and the LEO one at a parity of n_{h0}, for FCA.

- a cluster of 7 cells is considered [23];
- a parallelogram shaped cellular network folded onto itself with 7 cells per side is used;
- 70 channels are available to the system; then, according to the selected cluster size, each cell has permanently allocated 10 channels;
- a First Input First Output (FIFO) queuing discipline has been considered for handovers which do not immediately obtain service;
- each cell has 10 rooms for queuing handover requests;
- HPBW of the satellite antenna spot-beams has been kept fixed and equal to 0.27 radiants (this value gives a cell radius about equal to 212.5 km for a satellite altitude of 780 km). We have considered that the LEO satellite altitude increases from 500 to 2,000 km; hence, on the basis of HPBW = 0.27 radiants and T_m = 3 min, α ranges from 0.1 to 0.7, because V_{trk} diminishes [14] and R increases.

In figure 6, we have also shown the behavior of n_{h0} from (31), for an easy understanding of the mobility conditions that correspond to each α value. Figure 6 highlights that both P_{drop} and GOS_2 increase and P_{b1} decreases, if the user mobility increases (i.e., α decreases). The behavior of P_{b1} can be justified by taking into account that the mobility increase reduces the mean channel holding time in a cell with respect to the average unencumbered call duration; hence, the total traffic intensity in a cell decreases. However, if the user mobility increases, a call crosses a greater number of cells during its lifetime and at each cell change it may be dropped with probability P_{b2} due to the handover failure. This fact causes a significant increase in P_{drop}. A similar behavior (except for a scale factor) is obtained for GOS_2. A further validation of these results can be found in [12], where (under different mobility assumptions) the authors prove that user mobility entails a capacity increase in order to guarantee the same blocking requirements of an hypothetical cellular network with fixed users and the same offered traffic.

Under the same assumptions made for the results of figure 6, we have quantified by simulations the impact of the LEO mobility on the FCA-QH performance. We have evaluated the maximum traffic intensity per cell, ρ_{max}, that fulfills the ITU-T requirements shown in section 2 with 10 channels/cell. For instance, we have $\rho_{max} \approx 3$ erl/cell for $\alpha = 0.31$ and $\rho_{max} \approx 2.6$ erl/cell for $\alpha = 0.16$: if the number of handover requests per call doubles, there is about a 13% capacity reduction. A similar trend has been verified for the α values within the LEO range and for different numbers of system channels.

The impact of different mobility assumptions on the performance of a channel allocation technique can be highlighted on the basis of figure 7, which compares the performance of Fixed Channel Allocation (FCA) [10] for the LEO mobility model presented in this section and the Guérin's mobility model[6] that has been shown in [16]. In both cases, we have assumed the same mean mobile sojourn time in a cell and T_m = 3 min (i.e., the same n_{h0} value). In the LEO case, we have selected IRIDIUM-like mobility data (i.e., V_{trk} = 26,600 km/h, R = 212.5 km) and we have obtained n_{h0} equal to 3.125 handovers/call. While, in the Guérin's mobility case, we have considered $E[\nu]$ = 87 km/h, R = 1 km and, according to [16], we have still obtained n_{h0} equal to 3.125 handovers/call. As in the previous graph, we have simulated a parallelogram shaped cellular network with 7 cells per side, 7 cell reuse cluster and 70 system channels. The results in figure 7

[6] In [16] a mobility model suitable for terrestrial cellular systems is presented. The assumptions of this model are: a homogeneous cellular layout; iid and exponentially distributed mobile sojourn times in subsequent cells; uniform traffic. The excess life theorem (8) can be used to relate the distributions of t_{mc1} and t_{mc2}, and we find that now they are equal. The handover generation process is memoryless. Moreover, equations (9)–(18) and (20)–(26) of section 4 can be applied. Finally, in [16] it is shown that $n_{h0} \approx 0.7178/\alpha$ handovers/call, where α is given by $R/\{E[\nu]T_m\}$.

show that P_{b1} values are almost the same in both cases, whereas the Guérin's model entails higher P_{drop} and GOS_2 values than the LEO one. This difference can be justified if we consider that in the LEO mobility model users are more synchronized in their motion[7] and this aspect may favor the management of handovers.

6. Remarks on the handover traffic

A traffic is characterized by both the arrival process and the service time distribution. Under the basic assumptions made in section 3, analytic derivations of the blocking probability usually consider [7,8,17,35,39]: (i) a Poisson arrival process for handovers towards a cell with rate λ_h related to λ on the basis of (22); (ii) a handover arrival process cell-to-cell independent; (iii) an exponentially distributed channel holding time in a cell. However, this handover traffic characterization is approximated: generally, channel holding times are not exponentially distributed, as proved by (12); the handover arrival process is not Poisson, as discussed later in this section; handover arrival processes in adjacent cells are correlated, since a user may cross several cells during a call. This section presents some qualitative and quantitative considerations that are useful to characterize the handover arrival process and to evaluate its impact on the performance of channel allocation schemes.

In general, the handover arrival process towards a cell is the aggregation of several contributions coming from adjacent cells. However, in the LEO model of the previous section, the handover arrival process in a cell is derived from the output process of only one adjacent cell.

An interesting parameter to characterize an arrival process is the Index of Dispersion for Counts (IDC) [12]: IDC_t at time t is the variance of the number of arrivals in an interval of length t divided by the mean number of arrivals in t. For Poisson processes, $IDC_t \equiv 1$, $\forall t$. In general, IDC values greater than 1 highlight a more bursty arrival process than a Poisson one. Whereas, arrival processes with a lower variability than a Poisson one have $IDC < 1$. The limiting case is a deterministic arrival process, where $IDC_t \equiv 0$, $\forall t$. For a given distribution of the channel holding time and a given value of the average arrival rate, the blocking probability of a loss queuing system decreases, if we consider arrival processes with lower IDC values.

We use IDC_t to study the handover arrival process offered to a cell. Of course, since the new call arrival process is Poisson, it is characterized by $IDC_t \equiv 1$, $\forall t$. Whereas, the IDC_t value for the handover arrival process to a cell depends on both the user mobility conditions, the channel allocation technique, the traffic intensity and the number of available channels.

The handover arrival process and the new call arrival process are merged so as to form the input process to the loss queuing system which models the behavior of a cell according to the BCC policy. Since, on the basis of (12), the distributions of the channel holding times for new call attempts and handed-over calls are quite similar[8], IDC differences between the new call arrival process and the handover arrival process will cause different values of the related blocking probabilities. We have evaluated $IDC_{t=4\ min}$ for the LEO case (IRIDIUM-like mobility, $\alpha = 0.16$) with FCA by using the same simulation model outlined in section 5. In particular, for a given cell, the number of arrivals have been counted in intervals of 4 minutes for both new call attempts and handover requests. Correspondingly, we have obtained two histograms, as shown in figure 8 in the case of a traffic intensity due to new call attempts ($= \lambda T_m$) equal to 8 erl/cell (we have considered here a heavy traffic case in order to emphasize the impact of the blocking on the arrival process characteristics). These histograms show a higher peak for new calls than for handover requests. This difference is due to the fact that there is a practical limit to the maximum number of handover arrivals in 4 min. In this case, we have obtained $IDC_{t=4\ min} \approx 0.98$ for the new call arrival process and $IDC_{t=4\ min} \approx 0.56$ for the handover arrival process.

Figure 9 shows $IDC_{t=4\ min}$ for the handover arrival process with both Dynamic Channel Allocation (DCA) and FCA for different traffic intensity values due to new call attempts ($= \lambda T_m$). In particular, DCA assigns channels to cells on demand on the basis of a cost-function, as described in [9,10]. We have that $IDC_{t=4\ min} < 1$ in all cases for the handover arrival process, so highlighting that the handover traffic has a lower variability than a Poisson one. Moreover, $IDC_{t=4\ min}$ decreases as the traffic increases, because we have a higher handover failure probability (i.e., P_{b2}) that produces a smoother handover traffic. Finally, DCA gives greater $IDC_{t=4\ min}$ values than FCA for the same traffic intensity values, because DCA allows lower P_{b2} values. We have also verified that IDC_t values slightly reduce as mobility increases (i.e, n_{h0} increases), because a call crosses more cells during its lifetime and at each cell passage the handover traffic is smoothed due to the loss queuing system behavior of a cell. Of course, the IDC values depend on mobility assumptions and channel allocation techniques, but the handover arrival process characteristics that have been outlined above on the basis of the mobility model for LEO-MSSs are generally applicable.

The differences between the handover arrival process and the new call arrival process have an impact on the blocking performance of channel allocation techniques. Since the handover arrival process has a lower variability than the new call arrival process, we expect that P_{b2} is lower than P_{b1} even without any prioritization strategy for handover requests. This interesting consideration has been confirmed by the simulation results shown in fig-

[7] In the LEO mobility model considered in this section, all the users have the same speed, the same motion direction, the same cell sojourn time.

[8] If we look at equation (12), we note that both $f_{tmc1}(t)$ and $f_{tmc2}(t)$ are weighted by the same exponential factor e^{-t/T_m} (which is due to the distribution of t_d) and we can consider that they are quite close to each other, in particular if compared on the basis of parameter G introduced in [17].