

Document downloaded from:

<http://hdl.handle.net/10251/37938>

This paper must be cited as:

Bernal Mor, E.; Martínez Bauset, J.; Pla, V. (2012). Handover Performance for Elastic Flows in Mobile Cellular Networks. *IEEE Communications Letters*. 61(6):2686-2697. doi:10.1109/LCOMM.2012.081612.121366.



The final publication is available at

<http://dx.doi.org/10.1109/LCOMM.2012.081612.121366>

Copyright Institute of Electrical and Electronics Engineers (IEEE)

Handover Performance for Elastic Flows in Mobile Cellular Networks

Elena Bernal-Mor, *Student Member, IEEE*, Vicent Pla, *Member, IEEE*, and Jorge Martinez-Bauset

Abstract—The characterization of flows duration and their handover metrics is of paramount importance for the design and resource management of mobile networks. Elastic flows duration characterization is more involved than for streaming sessions, as it depends on the network load. In this work, the distribution of flows duration is characterized by a phase type distribution. The model is then applied to determine the handover probability under different cell residence time and flow size distributions. Simulation results confirm that the proposed model yields very accurate results under realistic assumptions.

Index Terms—Handover probability, elastic traffic, elastic flow duration.

I. INTRODUCTION

MODERN mobile networks support a variety of service classes (SCs) and carry different types of traffic. The teletraffic analysis of mobile networks, as well as the design of efficient mobility models, play an important role in the network dimensioning and resource planning [1]. In this context, knowing the probability of handover or the number of handovers that a session will execute are useful for resource dimensioning [2].

Internet traffic can be broadly categorized as streaming or elastic [3]. Streaming traffic requires a minimum transfer rate as well as bounds for the packet delay and jitter. It is generated by real-time services that support voice or video. Elastic traffic can adapt its transfer rate to the available free resources. It is generated by services that support the transfer of data like emails or web browsing.

For voice traffic, the session (call) duration and the cell residence (dwell) time distributions have been well studied [4], [5]. Also, the handover probability was studied in [1]. However, despite the enormous increase in the volume and economical relevance of mobile data traffic caused by the introduction of smartphones [6], to the best of our knowledge, such type of studies have not been carried out for elastic flows so far. Modeling the elastic flow duration and handover related metrics is qualitatively different and more complex than modeling their streaming traffic counterparts, as the duration of an elastic flow is heavily dependent on the network load.

In this paper, we obtain the distribution of both the flow duration and the number of handovers it has to go through. To make the model more realistic and of broader application, we consider impatient elastic users and a generic admission

policy. We provide exact results under the assumption that both flow sizes (in bits) and cell residence times are exponentially distributed. Then, the model is extended by introducing an approximate technique to deal with more realistic distributions of the cell residence time and the flow size.

In the next section, we define the Continuous-Time Markov Chain (CMTC) that models the system under study. In Section III we derive the distribution of the duration of elastic flows and handover related metrics. In Section IV we validate the analytical results by simulation. Finally, Section V concludes the paper.

II. SYSTEM MODEL

We consider the homogeneous case where all cells are statistically identical and independent. Each cell has a total of C resource units, each of them with a capacity of R bits per second. The system offers N_s different SCs that carry streaming traffic and N_e SCs that carry elastic traffic. Thus, the total number of SCs is $N = N_s + N_e$, where by $i = 1, \dots, N_s$ we refer to streaming SCs and by $i = N_s + 1, \dots, N$ to elastic SCs. Elastic flows use the capacity not occupied by streaming traffic. To avoid starvation, the system reserves 1 resource unit for elastic traffic.

For the sake of mathematical tractability we make the common assumptions of Poisson arrival processes for all SCs and exponentially distributed session durations and cell residence times for the streaming SCs. Let λ_i^n (λ_i^h) be the arrival rate for new (handover) streaming sessions or elastic flows of the i th SC. Let $\mu_i^{d,s}$, $\mu_i^{r,s}$ and $\mu_i^{d,s} + \mu_i^{r,s}$, $1 \leq i \leq N_s$, be the session duration, cell residence time and channel holding time rates, respectively. Also, let b_i , $b_i \in \mathbb{N}$, the number of resource units consumed by each request of the i th streaming SC. The system state is described by the N -tuple $\mathbf{x} = (x_1, \dots, x_N)$, where x_i denotes the number of ongoing streaming sessions or elastic flows of the i th SC, regardless they were initiated as new or handover arrivals. The amount of resources occupied at state \mathbf{x} by streaming traffic is denoted by $b(\mathbf{x}) = \sum_{i=1}^{N_s} x_i b_i$.

We model elastic traffic at the flow level and ignore interactions at the packet level (scheduling, buffer management, etc.). The flow content is then viewed as a fluid that is transmitted as a continuous stream with rate changes occurring only at streaming sessions or elastic flow arrivals and departures. An elastic flow of the i th SC is assumed to be rate-limited, either by terminal capabilities or because it is bottlenecked at the radio link. We denote its maximum rate by r_i^M . Also, elastic flows require a minimum bandwidth denoted by r_i^m . This can be guaranteed by an appropriate admission control (AC)

This work was supported by the Spanish Government through projects TIN2008-06739-C04-02 and TIN2010-21378-C02-02.

The authors are with the Department of Communications, Universitat Politècnica de València (UPV), ETSIT, Camino de Vera s/n, 46022 Valencia, Spain. Email: {jmartinez,vpla}@dcom.upv.es, elbermo@upvnet.upv.es

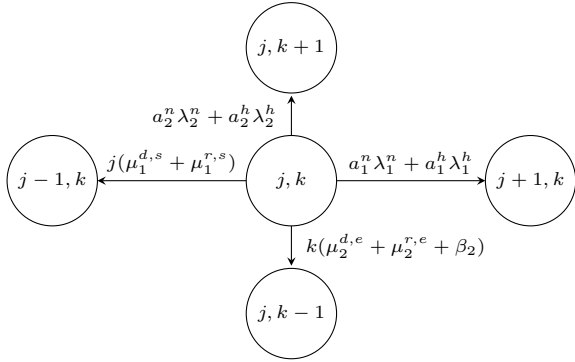


Fig. 1. Transition diagram of the CTMC with SC=2. The state (j, k) obeys $0 \leq j \leq \lfloor (C-1)/b_1 \rfloor$ and $0 \leq k \leq \lfloor CR/r_2^m \rfloor$

policy. We assume that flow sizes are exponentially distributed with mean L (bits). Let $\mu_i^M = r_i^M/L$ and $\mu_i^m = r_i^m/L$ (in flow/s) be the maximum and minimum service rates for the i th SC flow. For elastic flows, the cell residence time is assumed to be exponentially distributed with rate $\mu_i^{r,e}$. However, we also study other distributions in Section IV.

Without loss of generality, we consider that flows are ordered in increasing value of their rate limits r_i^M . The number of elastic flows in the system at state \mathbf{x} is denoted by $c(\mathbf{x}) = \sum_{i=N_s+1}^N x_i$. Let $\mu_i^{d,e}(\mathbf{x})$ be the service rate of the i th SC flow at state \mathbf{x} and $V(\mathbf{x}) = (C-b(\mathbf{x}))R/L$ the aggregated flow service rate at state \mathbf{x} . Flows share the available resource fairly and according to the following rule. For $i = N_s + 1$,

$$\mu_i^{d,e}(\mathbf{x}) = \min\left(\mu_i^M, \frac{V(\mathbf{x})}{c(\mathbf{x})}\right), \quad (1)$$

while for $i > N_s + 1$,

$$\mu_i^{d,e}(\mathbf{x}) = \min\left(\mu_i^M, \frac{V(\mathbf{x}) - \sum_{j=N_s+1}^{i-1} x_j \mu_j^{d,e}(\mathbf{x})}{\sum_{j=i}^N x_j}\right). \quad (2)$$

Clearly, in (2) the bandwidth that a SC cannot use is used by other SCs with higher rate limit. Flows become impatient and might leave the system due to a very low throughput. The patience time is modeled by an exponential distribution, which rate $\beta_i(\mathbf{x})$ at state \mathbf{x} is

$$\beta_i(\mathbf{x}) = \alpha_i \left(\frac{\mu_i^M}{\mu_i^{d,e}(\mathbf{x})} - 1 \right), \quad (3)$$

where α_i is a scaling factor that relates the throughput degradation and the patience rate. In (3), the patience rate increases when the service rate of the i th SC decreases. We consider a non preemptive AC policy. A streaming session or elastic flow are accepted if there are enough free resources to support it and if, after acceptance, all ongoing elastic flows obtain a service rate equal or bigger than their minimum μ_i^m . We denote by $a_i^n(\mathbf{x})$ ($a_i^h(\mathbf{x})$) the probability of accepting a new (handover) arrival of the i th SC in state \mathbf{x} . Then, the system can be modeled as a multidimensional birth and death process with state space,

$$\mathcal{W} := \left\{ \mathbf{x} : x_i \in \mathbb{N}; \sum_{i=1}^{N_s} x_i b_i \leq C - 1; \mu_i^{d,e}(\mathbf{x}) \geq \mu_i^m \right\}.$$

As an example, Fig. 1 shows the transition diagram of the CTMC that models a system where $N_s = 1$, $N_e = 1$. For clarity, the notation has been simplified writing a_i^n , a_i^h , β_2 and $\mu_2^{d,e}$ instead of $a_i^n(\mathbf{x})$, $a_i^h(\mathbf{x})$, $\beta_2(\mathbf{x})$ and $\mu_2^{d,e}(\mathbf{x})$, respectively. The stationary distribution $\{\pi(\mathbf{x})\}$ of this CTMC can be obtained by solving the global balance equations. From now on we will refer to it as the original CTMC.

III. DURATION OF ELASTIC FLOWS

In the system under study, the duration of an elastic flow is composed of a number of exponentially distributed phases with different rates, i.e. it follows a phase type distribution. A phase type distribution defines the time until absorption in an Absorbing Markov Process (AMP) [7]. It is commonly represented by a pair (α, \mathbf{S}) , where matrix \mathbf{S} defines the transition rates between the transient states, and vector α the probabilities that the process is started at any of the states. The transition rates from the transient states to the absorbing state is defined by vector τ , that satisfies $\tau = -\mathbf{S}e$, where e is a column vector of 1s.

An AMP is defined for each elastic SC i . It contains the states that might be visited by a tagged elastic flow of the i th SC until it abandons or terminates successfully. The AMP must consider that flows might be handed over multiple times to adjacent cells. We assume that the tagged flow is never blocked. The state space of the AMP for the i th SC flow is defined by $\mathcal{W}'_i := \{\mathbf{x} \in \mathcal{W}, x_i > 0\}$. The initiation vector α_i is derived from the stationary probabilities of the original CTMC $\pi(\mathbf{x})$ considering that the state space is restricted to \mathcal{W}'_i . As an example, Fig. 2 shows the AMP associated to the elastic flows of a system with $N_s = 1$, $N_e = 1$. State A represents the absorbing state and is visited when the labeled flow abandons or ends. (p, m) represents a set of states composed by all the feasible states that can be reached immediately after the tagged flow is handed over to an adjacent cell. The probabilities $\pi'(p, m-1)$ are obtained from the stationary distribution of the original CTMC $\{\pi(\mathbf{x})\}$ after removing the blocking states and then re-normalizing.

Let T_d and T_r be random variables that denote the duration and cell residence time of an elastic flow, respectively. At this point no assumptions are made regarding the distributions of these random variables. A flow is handed over when T_d is longer than T_r . When a flow enters a cell as a handover request, the probability that is handed over again is determined by

$$\begin{aligned} P^h &= 1 - P(T_d < T_r) = 1 - \int_0^\infty P(t_d < T_r) f_d(t_d) dt_d \\ &= 1 - \int_0^\infty (1 - F_r(t_d)) f_d(t_d) dt_d, \end{aligned} \quad (4)$$

where F_r is the distribution function of T_r , and f_d is the probability density function of T_d .

Let \hat{T}_r be the residual life of a flow cell residence time, i.e. the time elapsed since a flow is initiated until the terminal leaves the cell [8]. When a flow is initiated as a new request, it is handed over when T_d is longer than \hat{T}_r , and this occurs with probability P_1^h . Then, P_1^h , has the same expression as

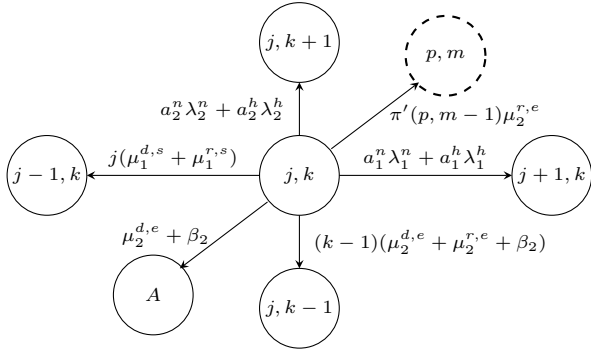


Fig. 2. Transition diagram of the AMP for SC=2. A is the absorbing state, $0 \leq j, p \leq \lfloor (C-1)/b_1 \rfloor$, $1 \leq k, m \leq \lfloor CR/r_2^m \rfloor$ and $(p, m) \neq (j, k)$.

in (4), but substituting F_r by the distribution function of the residual cell residence time, \hat{F}_r , which is given by:

$$\hat{F}_r(t) = \int_0^\infty \frac{1}{E[T_r]} (1 - F_r(t)) dt. \quad (5)$$

In the system under study, the flow duration follows a phase type distribution and therefore, $f_d(t) = \alpha e^{tS} \tau$. As a particular case, if we consider exponentially distributed cell residence times, $\hat{F}_r(t) = F_r(t) = 1 - e^{-\mu^{r,e} t}$. Therefore,

$$P_1^h = P^h = 1 - \int_0^\infty e^{-\mu^{r,e} t} \alpha e^{tS} \tau dt = 1 - \alpha (\mu^{r,e} I - S)^{-1} \tau. \quad (6)$$

Let N be the random variable number of handovers executed by an elastic flow during its lifetime. Its distribution is given by:

$$P[N = n] = \begin{cases} P_1^h (1 - P^h) (P^h)^{n-1} & n \neq 0 \\ 1 - P_1^h & n = 0. \end{cases} \quad (7)$$

Then, N is geometrically distributed with mean $P_1^h / (1 - P^h)$.

IV. NUMERICAL EVALUATION

In this section the analytical and simulation results are compared to validate the theoretical model. We consider a system with $N_s = 2$ streaming SCs and $N_e = 1$ elastic SC. The system parameter values, unless otherwise specified, are: $C = 10$, $R = 100$, $b = [1, 2]$, $\lambda^n = [0.008, 0.012, 5]$, $\mu^{d,s} = [0.008, 0.01]$, $\mu^{r,s} = [0.004, 0.006]$, $L = 200$, $r_3^m = 50$, $r_3^M = 500$ and $\alpha_3 = 1$.

Figures 3 and 4 show the evolution of the handover probability, both P^h and P_1^h , with $\mu^{r,e}$. Curves labeled with ‘Exp’ correspond to analytic results considering exponentially distributed cell residence times and obtained using (6). Curves labeled with ‘Logn P^h ’ and ‘Logn P_1^h ’ are obtained considering exponentially distributed cell residence times to determine (α, S) , and then using lognormal distributions and their residual distributions to model cell residence times when computing P^h and P_1^h from (4), as explained in Section III. The lognormal distribution is chosen as it models the cell residence time more realistically [9]. Its mean is set to $1/\mu^{r,e}$, while its Coefficient of Variation (CV) is set to 1 in Fig. 3 and to 0.5 and 2 in Fig. 4. Curves corresponding

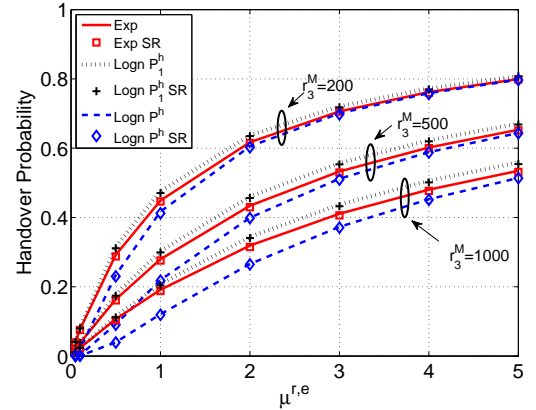


Fig. 3. P^h and P_1^h as a function of $\mu^{r,e}$ ($CV = 1$).

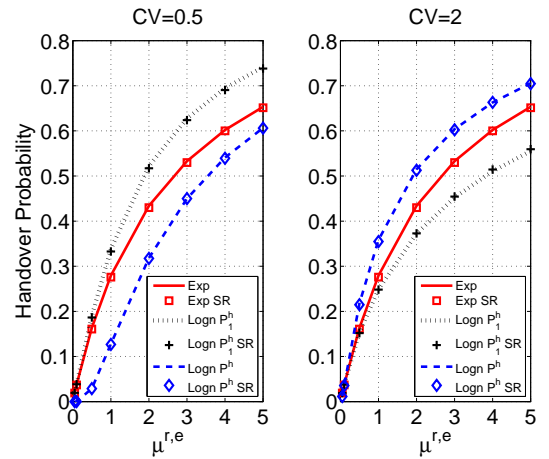


Fig. 4. P^h and P_1^h as a function of $\mu^{r,e}$. Lognormally distributed cell residence times with $CV = 0.5$ and $CV = 2$.

to simulation results are labeled with ‘SR’. They are obtained by considering a multi-cell scenario with a central cell and two outer rings of cells, which make a total 19 cells. Upon cell residence time termination, terminals select a neighbor cell with equal probability. We consider wraparound to avoid abnormal terminations at the edges. Observe the excellent agreement between the analytical and simulation results, i.e. they practically overlap. We conclude that the phase type distribution models appropriately the flow duration, even when the cell residence times are lognormally distributed.

In both Fig. 3 and 4, P^h and P_1^h increase with $\mu^{r,e}$ as expected, because the cell residence time decreases as the rate increases. In Fig. 3, also as expected, P^h and P_1^h decrease as r_3^M increases. This is because the higher r_3^M is, the earlier the flow transfer will terminate.

Figure 5 displays the evolution of the distribution of N (number of handovers) with $\mu^{r,e}$ for lognormally distributed cell residence times with $CV = 1$. Analytical results, obtained using (7), are represented using lines, while simulation results are represented only by markers at the evaluated points. Clearly, when the mobility increases, i.e. $\mu^{r,e}$ increases, $P[N = n]$, $n \geq 1$, increases, while $P[N = 0]$ decreases. That is, as the mobility increases it is more probable that a flow

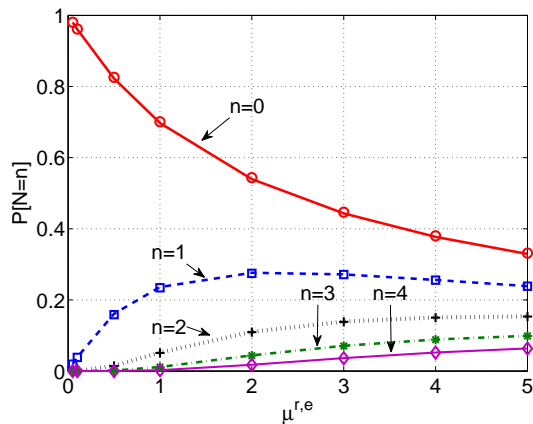


Fig. 5. Distribution of N (number of handovers) as a function of $\mu^{r,e}$.

executes 1 or more handovers, while it is less probable that executes 0 handovers. Note that the curve for $n = 1$, first increases with $\mu^{r,e}$ and then decreases. This is because, the higher $\mu^{r,e}$ is, the less likely that a flow executes exactly 1 handover and the more likely that it executes more than 1.

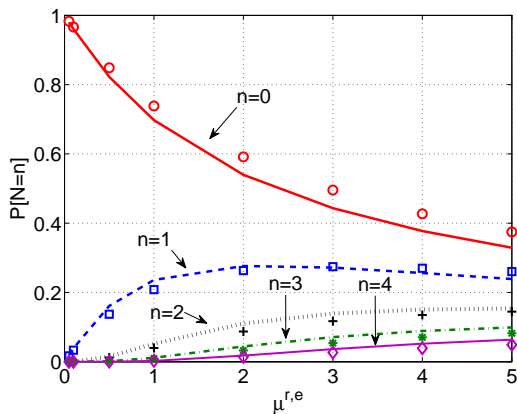


Fig. 6. Distribution of N (number of handovers) as a function of $\mu^{r,e}$. Exponential and lognormal flow size distributions.

Figure 6 shows the impact of the flow size distribution on the distribution of N for lognormal cell residence times with $CV = 1$. Analytical results, represented by lines, are obtained considering an exponential flow size distribution and using (7). Simulation results, represented only by markers at the evaluated points, are obtained considering a lognormal flow size distribution with $CV = 2$, as it has been considered more realistic [10]. Note that when the flow size is exponentially distributed, the memoryless property for the flow size holds and, after a handover, the residual flow size maintains the original distribution. However, when the flow size distribution is lognormal, the residual flow size does not maintain the original distribution after a handover, i.e. handover probabilities depend on the number of previous handovers. This fact has been taken into account in the simulation model.

Results in Fig. 6 show an excellent agreement between the analytical and simulation results. As expected, the agreement

is closer for low mobility scenarios (low $\mu^{r,e}$). That is, when the number of handovers experienced is 0 or close to 0, the difference in the residual distributions of the flow size has no impact. However, in high mobility scenarios, the proposed model is still able to capture with high precision the handover performance of the elastic flows.

V. CONCLUSIONS

We studied the flow duration distribution for elastic traffic in cellular networks and obtained handover related metrics, such as the distribution of the number of handovers a flow has to go through. To make the model more realistic and of broader application, we considered impatient elastic users, a generic admission policy that guarantees a minimum rate for elastic flows, and scenarios where cell residence times and flow sizes were modeled by different distributions.

The flow duration was characterized by a phase type distribution and the probabilities of the first and successive handovers were determined. We provide exact results under the assumption that both flow sizes and cell residence times are exponentially distributed. Then, the model was extended by introducing an approximate technique to deal with more realistic distributions. We used lognormal distributions because, as suggested in the literature, they model realistically both cell residence times and flow sizes. Note the the modeling approach can be used with any other distribution. Results obtained by the proposed technique were validated by simulation and very close agreement was found.

We conclude that the proposed method models appropriately the flow duration and handover performance under rather general assumptions.

REFERENCES

- [1] Y. Fang and I. Chlamtac, "Analytical generalized results for handoff probability in wireless networks," *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 396–399, Mar. 2002.
- [2] Y. Zhang, "Handoff performance in wireless mobile networks with unreliable fading channel," *IEEE Transactions on Mobile Computing*, vol. 9, no. 2, pp. 188–200, Feb. 2010.
- [3] T. Bonald and J.W. Roberts, "Congestion at flow level and the impact of user behaviour," *Computer Networks*, vol. 42, pp. 521–536, 2003.
- [4] F. Barceló and J. Jordán, "Channel holding time distribution in public telephony systems (PAMR and PCS)," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 5, pp. 1615–1625, Sep. 2000.
- [5] T. K. Christensen, B.F. Nielsen and V. B. Iversen, "Phase-type models on channel-holding times in cellular communication systems," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 3, pp. 725–733, May. 2004.
- [6] Y. Choi, H. W. Ji, J.-Y. Park, H.-C. Kim and J.A. Silvester, "A 3W network strategy for mobile data traffic offloading," *IEEE Communications Magazine*, vol. 49, no. 10, pp. 118–123, Oct. 2011.
- [7] M. Neuts, "Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach," The Johns Hopkins University Press, 1981.
- [8] H.-N. Hung, P.-C. Lee and Y.-B. Lin, "Random number generation for residual life of mobile phone movement," in *Proceedings of the IEEE ICNSC*, Taipei, Taiwan, Mar. 2004, pp. 30–33.
- [9] E. Zola and F. Barcelo-Arroyo, "Impact of mobility models on the cell residence time in WLAN networks," in *Proceedings of the IEEE SARNOFF'09*, New Jersey, USA, Mar. 2009, pp. 1–5.
- [10] G.-F. Zhao, Q. Shan, S. Xiao and C. Xu, "Modeling web browsing on mobile internet," *IEEE Communications Letters*, vol. 15, no. 10, pp. 1081–1083, Oct. 2011