Impact of Technology Overlap in Next-Generation Wireless Heterogeneous Systems

Ahmed Zahran[1], Ben Liang[1] and Aladdin Saleh[2]

[1] Department of Electrical and Computer Engineering, University of Toronto.
[2] Wireless Technology, Bell Canada.

Abstract. The integration of different wireless access technologies is propelled by the need to support new services and better resource utilization in next-generation wireless networks. This integration complicates the system design due to the interaction of different factors including network-oriented, application-oriented, and user-oriented system parameters. In this work, we present an analytical framework to estimate different session-level performance metrics in two-tier systems, using the 3G-WLAN integrated network as an example. We investigate the impact of the amount of coverage overlap and the topology of the underlay technology on different session performance metrics as well as the total session cost. The obtained results show that clustering can significantly reduce the vertical-handoff signaling load and the forced termination probability of different applications in comparison with a random topology. Additionally, the proposed cost analysis provides design guidelines for developing economical WLAN management mechanisms to maintain reduced session cost with extended WLAN coverage.

1 Introduction

Next-generation wireless network will converge the service of different pervasive access technologies while providing the user with a large set of novel revenue generating multimedia-based applications. Hence, heterogeneity will naturally become the main system feature due to the huge diversity in the characteristics of access technologies and applications. The 3G-WLAN integrated system is an example of this wireless heterogeneous networking paradigm that has received great support from industrial and standardization bodies [1–3]. Currently, both technologies are among the most pervasive wireless access approaches that complement each other. On the one hand, 3G networks provide an expensive universal coverage; on the other hand, WLANs provide ample networking resources for the users at a cheaper cost wherever available. Hence, users will generally enjoy the best of each access technology, and service providers will enjoy better utilization of their resources.

The integrated system is very rich in its parameter set, including technology-oriented parameters, application-oriented parameters, and user-oriented parameters. All these parameters are interwoven together in an inter-technology roaming enabled environment creating a real challenge for system analysis and performance evaluation. For example, in the 3G-WLAN integrated model, the bandwidth provided to the mobile terminal (MT) in both networks may vary by one order of magnitude after any change of network attachment point. Combining this fact with the bandwidth greediness of some applications due to their buffering or pre-fetching capabilities, one can directly conclude that inter-technology roaming, commonly known as vertical handoff (VHO) [4], will significantly influence next-generation session dynamics. This influence affects different system design aspects such as resource utilization, signaling, and quality-of-service (QoS). Consequently, investigating the impact of VHO on system performance under different scenarios is crucial for better design and performance evaluation of next-generation systems.

The performance evaluation of heterogeneous wireless network is a challenging task due to the system complexity and the large set of parameters that should be considered. To the best of our knowledge, very few papers [5,6] deal with mathematical modeling and performance analysis of heterogeneous wireless networks. In [5], the authors study the admission region of voice and data traffic within the integrated system. They derive expressions for the dropping and blocking probabilities of data and voice calls in 3G cell and WLAN. Their main result is that voice traffic in the double coverage area should be restricted from occupying all the WLAN bandwidth. In [6], we propose a novel mobility model for two-tier integrated system using Coxian phase-type structures to represent the cell residence time. Additionally, we develop a generic framework based on matrix-geometric theory and Markov reward models to estimate different *cell-level* metrics such as network utilization times and VHO rate.

In this paper, we present an analytical framework to estimate different *session-level* performance metrics. The obtained results are used to study the impact of WLAN coverage and topology variation on the session performance and total session cost. The analytical framework is validated through simulation. Additionally, the cost analysis shows that WLAN clustering in extended WLAN deployment can greatly reduce the session cost provided that the increased network management cost of extended WLAN deployment is eliminated. In Section 2, we present our 3G-WLAN integrated network model. Section 3 presents an analytical framework for estimating different session parameters and the session cost. We present the numerical results in Section 4 and conclude in Section 5.

2 Network Model

In this work, we assume that a cellular network provides universal coverage with WLANs overlapping at a portion P_{ow} of the total area. Consequently, the percentage of the unique cellular coverage area is $P_{oc} = 1 - P_{ow}$. Without loss of generality, we assume that active MTs prefer WLAN due to its larger band-

width and lower cost. Hence, any active MT will always handoff to WLAN when it is encountered. Additionally, we adopt the extended-Coxian model [6] for user mobility in this integrated network. This model is based on phase-type (PH) distributions [7]. Generally, a PH random variable is defined as the absorption time of an evanescent finite-state Markov process to a single absorbent state. Additionally, a PH random variable is conventionally denoted as $PH(\alpha, \mathbf{T})$, where α and $\mathbf{T}$ represent the initial transient state distribution and the transition rates among transient states, respectively. The proposed model employs the Coxian structure to represent cell residence time as a probabilistic sum of WLAN and cellular technology residence times (TRT), defined as the duration spent by the MT in a specific technology.

Figure 1 shows the extended-Coxian mobility model where each stage is la-
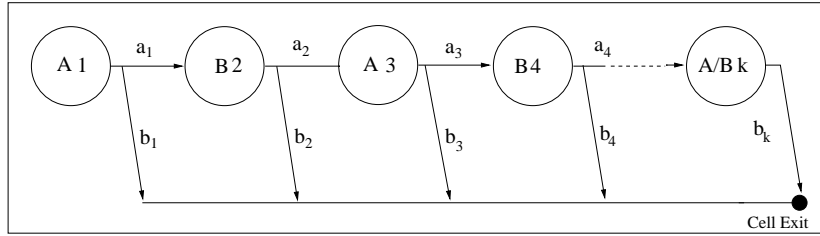

Fig. 1. Mobility Model

beled with a letter corresponding to the visited technology and a number representing the stage sequence within the intended cell. The model consists of consecutive stages such that upon exiting any stage i, the MT may proceed to another technology within the same cell with probability a_i or move to a neighbor cellular cell (i.e. get absorbed) with probability b_i. The TRT of stage i is PH distributed and denoted as $PH(\alpha_i, \mathbf{T}_i)$. Hence, the cellular residence time can be expressed as $PH(\alpha_m, \mathbf{T}_m)$

$$\mathbf{T}_m = \begin{pmatrix} \mathbf{T}_1 & a_1\mathbf{t}_1\alpha_2 & \mathbf{0}... & ..\mathbf{0} \\ \mathbf{0} & \mathbf{T}_2 & a_2\mathbf{t}_2\alpha_2 & ..\mathbf{0} \\ . & & . & . \\ \mathbf{0}.. & & \mathbf{0} & \mathbf{T}_k \end{pmatrix} \tag{1}$$

$$\alpha_m = [\alpha_1 \ \mathbf{0}] \tag{2}$$

where $\mathbf{t}_i = -\mathbf{T}_i\mathbf{e}$ [7].

Note that, in a two tier system, there are two mobility sub-models; each sub-model has a different initial phase technology. Hence, we use C-type and W-type cell to differentiate models with initial cellular and initial WLAN phases respectively. Furthermore, we assume that the MT application session has exponentially distributed holding times denoted as t_{ch} and t_{wh} with parameters λ_{ch} and λ_{wh} for cellular network and WLAN respectively. The exponential session

holding time assumption is common in wireless networks due to their time-based pricing strategies and the limited power resources of mobile devices.

3   Cost Analysis

In this section, we estimate a total session cost as a function of session resource utilization, session QoS, and network administrative cost. First we present our session model in Section 3.1. Then, we present a generic framework to estimate the total session network utilization and total session VHO load in Section 3.2. Then we present our proposed total session cost function in Section 3.3.

3.1   Session Modeling

In our session model, we combine the mobility model and application characteristics to represent both application activity and user mobility. Consequently, the main conceptual difference between the session and mobility models is the interpretation of absorption. On contrary to the mobility model in which absorption only occurs due to cell exit, the session may get absorbed for different reasons including normal session termination, denoted as the TERM state, successful handoff to a neighbor cell, denoted as the SHH state, session dropping during horizontal handoff, denoted as the HHFT state, or session dropping during vertical handoff, denoted as the VHFT state. Hence, the generator matrix of the Markovian session process will have the following structure:

$$\mathbf{Q}_S = \begin{pmatrix} \mathbf{Q}_{TT} & \mathbf{Q}_{Term} & \mathbf{Q}_{SHH} & \mathbf{Q}_{HHFT} & \mathbf{Q}_{VHFT} \\ \mathbf{0} & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{3}$$

where $\mathbf{Q}_{TT}$, $\mathbf{Q}_{Term}$, $\mathbf{Q}_{SHH}$, $\mathbf{Q}_{HHFT}$, $\mathbf{Q}_{VHFT}$ can be expressed as shown in [6].

Note that different cases are distinguished in the analysis according to different combinations of mobility sub-models and session types. We use A to denote mobility sub-model that is determined by the initial technology of the visited cell, where $A \in \{c, w\}$ for C-type and W-type cells respectively. Additionally, we use B to denote the session type, where $B \in \{n, h\}$ for new and handoff sessions respectively. As a notational remark, we use superscripts to define the session-type and mobility sub-model combination. Following this notation, we denote P_{TERM}^{AB}, P_{SHH}^{AB}, P_{VHFT}^{AB}, and P_{HHFT}^{AB} as the absorption probabilities to Term, SHH, VHFT, and HHFT states respectively for a B-type session starting in an A-type cell. These absorption probabilities can be estimated from the embedded Markov chain of the session model as shown in [8].

3.2   Session Performance Analysis

In this subsection, we present a generic analytical framework to estimate different *session-level* performance metrics from their corresponding *cell-level* metrics of different session types and mobility sub-models. The derivations of the cell-level

performance metrics are omitted here for brevity. However, interested readers are referred to [6] for the details. The following analytical framework is used to estimate session cellular network utilization, session WLAN utilization and session VHO rate, denoted as L_{sc}, L_{sw} and μ_{sv} respectively. Due to page limitation, in the following derivation, only μ_{sv} is shown for illustration purpose. For example, L_{sc} can be estimated by replacing μ_{cv}^{AB} and μ_{cv}^{h} with their corresponding metrics L_{cc}^{AB} and L_{cc}^{h} respectively.

The session VHO rate is defined as the expected total number of VHOs performed by the MT during an entire session. Given that the session starts in the W-type cell, a conditional session VHO rate , μ_{sv}^{w}, can be expressed as

$$\mu_{sv}^{w} = \mu_{cv}^{wn} + \sum_{k=1}^{\infty} P(n = k|W)k\mu_{cv}^{h}, \tag{4}$$

where μ_{cv}^{wn} is the VHO rate per cell of a new session starting in a W-type cell, $P(n = k|W)$ is the probability that a session starting in a WLAN will exactly visit k handoff cells before termination, and μ_{cv}^{h} is the VHO rate per cell of a handoff session and is expressed as $\mu_{cv}^{h} = P_{wh}\mu_{cv}^{wh} + P_{ch}\mu_{cv}^{ch}$, in which μ_{cv}^{wh} and μ_{cv}^{ch} represent the VHO rate per cell of a handoff session starting in a W-type and C-type cells respectively, and P_{wh} and P_{ch} represent the probability that the session will handoff to a W-type and C-type cells respectively. Note that P_{wh} and P_{ch} are topology dependent and equal the percentage of the cell edge with dual and unique coverage respectively. Clearly, μ_{sv}^{w} can be expressed as

$$\mu_{sv}^{w} = \mu_{cv}^{wn} + \mu_{hs}^{w}\mu_{cv}^{h}, \tag{5}$$

where μ_{hs}^{w} is the expected number of visited handoff cells per session given that the session starts in a W-type cell. In order to estimate μ_{hs}^{w}, we define the following probabilities

- P_{hh}: the probability that a handoff session will successfully perform a horizontal handoff to a neighbor cell. Hence $P_{hh} = P_{wh}P_{SHH}^{wh} + P_{ch}P_{SHH}^{ch}$.
- $\overline{P}_{hh}$: the probability that a handoff session will be terminated in the same cell either due to normal termination or due to forced termination during VHO.
- P_{s}: the probability that a handoff session will visit exactly one more cell. Hence, $P_{s} = P_{wh}P_{SHH}^{wh}\overline{P}_{hh} + P_{ch}P_{SHH}^{ch}\overline{P}_{hh}$.

Consequently, the marginal distribution function of the number of successful HHO given that the session starts in a W-type cell, N_{HH}^{w}, can be expressed as

$$P(N_{hs}^{w} = k) = \begin{cases} P_{Term}^{wn} + P_{VHFT}^{wn} + P_{HHFT}^{wn} \,, \, k = 0 \\ P_{SHH}^{wn}\overline{P}_{hh} \,, \, k = 1 \\ P_{SHH}^{wn}P_{hh}^{k-2}P_{s} \,, \forall k \geq 2 \end{cases}. \tag{6}$$

Hence, μ_{hs}^{w} is calculated as

$$\mu_{hs}^{w} = \sum_{k=0}^{\infty} kP(N_{hs}^{w} = k|W) \,. \tag{7}$$

Using the mathematical identity $\sum_{i=0}^{\infty} i c^i = \frac{c}{(1-c)^2}$, $|c| < 1$, μ_{hs}^w can be expressed as

$$\mu_{hs}^w = P_{SHH}^{wn} \left(\overline{P}_{hh} + P_s \left(\frac{2 - P_{hh}}{(1 - P_{hh})^2} \right) \right).$$ (8)

Similarly, the expected number of successful HHO for a session starting in the cellular network, μ_{hs}^c, can be expressed as

$$\mu_{hs}^c = P_{SHH}^{cn} \left(\overline{P}_{hh} + P_s \left(\frac{2 - P_{hh}}{(1 - P_{hh})^2} \right) \right).$$ (9)

Similar to (5), the expected total session VHO rate assuming the session starts in a C-type cell, μ_{sv}^c, can be expressed as

$$\mu_{sv}^c = \mu_{cv}^{cn} + \mu_{hs}^c \mu_{cv}^h.$$ (10)

Hence, using the total probability theorem, the total session VHO rate, μ_{sv}, is

$$\mu_{sv} = P_{wh} \mu_{sv}^w + P_{ch} \mu_{sv}^c.$$ (11)

3.3 Total Session Cost

The total session cost, χ, is calculated as a function of session performance metrics and system administrative cost. This cost function is expressed as

$$\chi = C_c L_{sc} + C_w L_{sw} + C_{vh} \mu_{vh} + C_{hh} \mu_{hh} + C_q P_{ft} + C_m * \eta^{P_{wo}},$$ (12)

where μ_{hh} represents the session horizontal handoff rate, P_{ft} represents the session forced termination probability, C_c and C_w represent cellular network and WLAN utilization cost coefficients respectively, C_{vh} and C_{hh} represent session vertical and horizontal handoff signaling cost coefficients respectively, C_q represents the session forced termination probability cost coefficient, C_m represents the management cost, and η is a WLAN area cost base. The derivation of μ_{hh} and P_{ft} is presented in [6]. Clearly, this cost function considers different aspects including resource utilization cost, signaling load cost, QoS factor, and administrative cost. The latter cost is considered due to its substantial increase in extended WLAN deployment [9]. The main source of the increase in management cost under extended WLAN deployment is administrative tasks such as RF monitoring and configuration, load balancing support, security issues, and station mobility support.

4 Numerical Results

In addition to the analysis, we have simulated an integrated heterogeneous system with square cells for simplicity of illustration. Each cell is sub-divided into N subdivisions, where WLANs are located. We consider two different WLAN topologies: random and clustered. In the random topology, WLANs are randomly

distributed in the cell and the cell encounter a new random topology as it moves to each new cell. In the clustered topology, all WLANs are aggregated in one hotspot that is randomly located within the cell. In order to emulate practical MT operation, a handoff area [10] of d_H seconds is assumed between overlay 3G cells. This delay corresponds to the hysteresis introduced in handoff algorithms to decrease the ping-ponging impact during horizontal handoff. Additionally, the MT 3G-WLAN handoff is delayed with d_s seconds as a typical delay required for WLAN discovery and handoff signaling. In the simulation, we adopt a two-dimensional Gauss-Markov movement model from [11] for MT mobility. In its discrete version, at time n, the MT velocity in each dimension, v_n, is given by

$$v_n = \alpha_v v_{n-1} + (1 - \alpha_v)\mu_v + \sqrt{1 - \alpha_v^2}\, x_{n-1}\,, \qquad (13)$$

where α_v represents a past velocity memory factor such that $0 \le \alpha_v \le 1$, μ_v is the asymptotic mean of v_n, and x_n is an independent and stationary Gaussian process with zero mean and standard deviation σ_v, where σ_v is the asymptotic standard deviation of v_n. The MT mobility parameters for the shown results are $\alpha = 0.9$, $\mu_v = 0$, and $\sigma_v = 2.5$. Table 1 lists the default values of analysis and simulation parameters.

Table 1. Analysis and simulation parameters

Parameter	Value	Parameter	Value	Parameter	Value
d_H(sec)	5	d_s(sec)	3	N	100
P_{vb}	0.01	P_{hb}	0.01	C_q	10000
C_c	20	C_w	2	η	10
C_{vh}	75	C_{hh}	75		

In this study, we categorize the applications as *symmetric* and *asymmetric*. The former preserves the same level of resource utilization independent of the available network resources, while the latter has a greedy nature and can consume as much bandwidth as the network can provide. Conversational applications, such as voice over IP and video conferencing, are examples of the former, while streaming applications with buffering capabilities, such as video on demand and radio on demand, are examples of the latter. Generally, asymmetric applications benefit from resource rich networks such as WLANs by increasing the download rate for smoother playback. In the simulation and analysis, we consider different combinations of session duration (long and short) and application type (symmetric and asymmetric) whose parameters are shown in Table 2. The following figures plot the analysis and simulation results (including 95% confidence intervals) for different session metrics. The obtained results show a good match between the analysis and simulation results. This match demonstrates the accuracy of the derived analytical framework in estimating different metrics.

Table 2. Application Parameters

| Unit Parameter	Symmetric Short	Symmetric Long	Asymmetric Short	Asymmetric Long
t_{ch} (minutes)	3	30	6	60
t_{wh} (minutes)	3	30	1	10

4.1 Session Performance

Figure 2(a) plots the session VHO rate versus WLAN coverage for different applications in random and clustered topologies. Clearly, the figure shows that WLAN aggregation significantly reduces VHO signaling load. This reduction reaches more than 70% at 50% coverage for all the simulated applications. Generally, the VHO rate is mainly driven by the circumference of WLAN hotspots. Hence, the VHO rate of symmetric applications always increases as WLAN coverage increases in the clustered topology, but it starts dropping as WLAN coverage increases beyond 50% in the random topology. Additionally, the VHO rate of asymmetric applications decreases as WLAN coverage increases due to the beneficial higher bandwidth of WLAN, which decreases the session duration of such applications. Figure 2(b) shows the session forced termination probability of different applications versus WLAN coverage in random and clustered topologies. Similar to VHO, clustering WLANs significantly reduces the forced termination probability of different applications compared to the random topology. For example, a 45% reduction is achieved at 35% coverage overlap for different applications.

Figure 2(c) shows HHO rate variations for different applications versus WLAN coverage for both random and clustered topologies. Generally, the HHO rate of symmetric applications is independent of WLAN coverage, while the HHO rates of asymmetric applications significantly decrease as WLAN coverage increases. Additionally, this figure shows that the applications perform more HHOs in the clustered topology compared to the random one. However, this slight increase is a logical consequence for extended application lifetime due to the reduced session forced termination probability in the former topology. Figures 2(d) and 2(e) plot the total-session cellular network and WLAN utilization, respectively, versus WLAN coverage for different applications in random and clustered topologies. Generally, a specific network utilization for a symmetric application is linearly proportional to its corresponding coverage independent of the WLAN topology. On contrary, as WLAN coverage is extended, the cellular utilization of asymmetric applications exponentially decreases with an insignificant increase in WLAN utilization. Additionally, the figure shows that the clustered cellular network and WLAN utilization are slightly larger when compared to the random topology. However, this is also due to the longer session lifetime in the clustered topology.
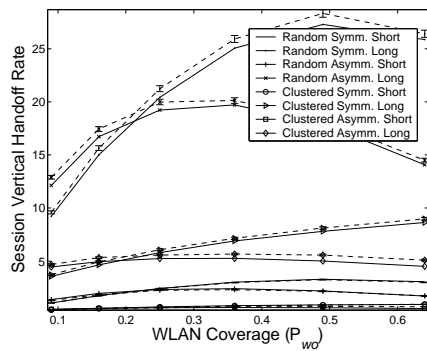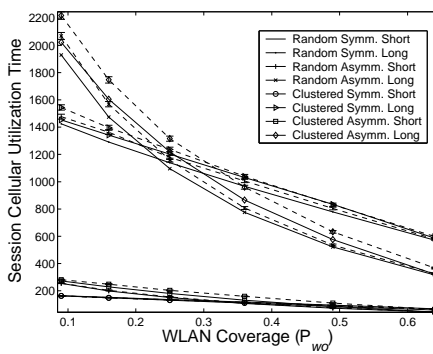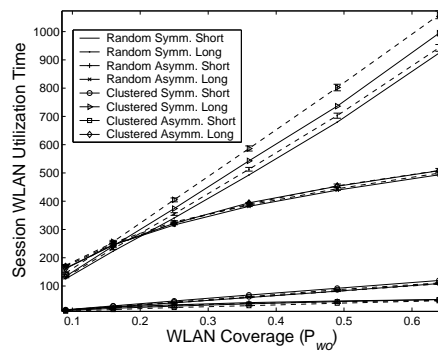
(a) Session VHO rate
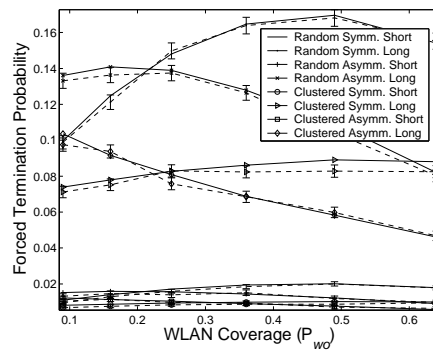
(b) Forced termination probability
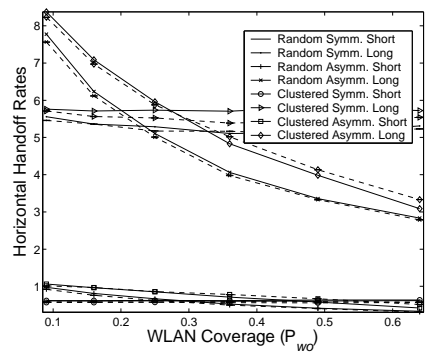
(c) HHO rate

(d) Session Cellular utilization

(e) Session WLAN utilization

Fig. 2. Session performance metrics versus WLAN coverage. The solid lines represent analysis, and the dashed lines represent simulation.

4.2 Session Cost

Figures 3(a) and 3(b) plot the total session cost versus the percentage of WLAN-cellular technology overlap for $C_m = 0$ and $C_m = 5000$ respectively. Figure 3(a) shows that increasing WLAN coverage reduces the total session cost in this scenario. This reduction may reach 50% and 22.5% for asymmetric and symmetric applications respectively as the WLAN coverage increases from 10% to 35%. However, the administrative costs of extended coverage approximately eliminates the integration cost benefits for the considered application set, except for asymmetric applications with large session duration, as shown in Figure 3(b). Even for this application, there exists an optimal WLAN coverage, e.g. approximately 50% overlap for the asymmetric application under consideration, beyond which the overall session cost is dominated by its administrative component. Hence, maintaining the cost reduction in an extended WLAN scenario mandates developing economical WLAN management schemes, such as remote management.

Figure 3(c) plots the total session cost versus the ratio between forced termination and cellular utilization costs for different applications at 36% coverage overlap. The figure suggests that WLAN clustering greatly reduces the total session cost as the application sensitivity to forced termination increases. Hence, clustering WLANs is highly recommended for systems dominated with this type of applications.

5 Conclusion

The inherent heterogeneity of next generation systems complicates system analysis and performance evaluation especially with the unavoidable interaction of different system parameters. A generic framework is presented to estimate different session-level performance metrics from their corresponding cell-level metrics. These metrics are used to study the impact of the amount of coverage overlap and the topology of the underlay technology on the total session cost. The obtained results show that WLAN clustering can reduce the VHO signaling load and the forced termination probability of different applications by up to 70% and 45%, respectively. Additionally, the cost analysis shows that increased administrative cost due to extended WLAN coverage can significantly reduce the integrated system cost benefits. Hence, developing economical WLAN management mechanisms is crucial to achieving real cost benefits from extended WLAN coverage.
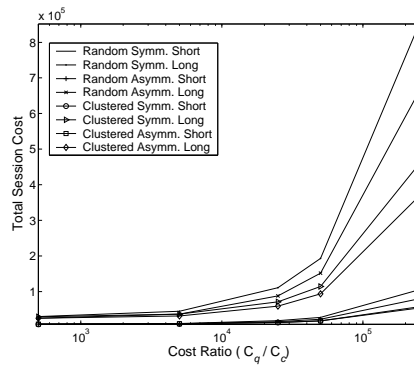
References

1. Buddhikot, M.M., Chandranmenon, G., Han, S., Lee, Y.W., Salgarelli, S.M.L. (2003): Integration of 802.11 and third generation wireless data networks. In Proc. of IEEE INFOCOM, San Francisco, US, 503-512.
2. 3GPP2 (2003): Feasibility study on 3GPP systems to. wireless local area network WLAN interworking. Technical report, 3GPP TR 22.934.
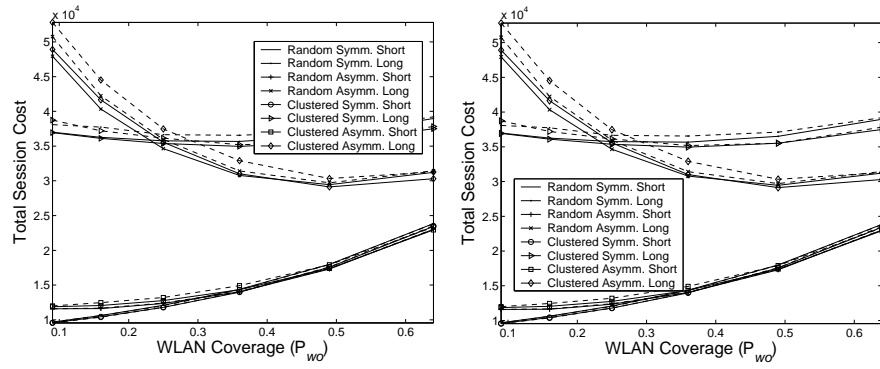
(a) $C_m = 0$

(b) $C_m = 5000$

(c) Network topology and application QoS sensitivity

Fig. 3. Total session cost. The solid lines represent analysis, and the dashed lines represent simulation.

3. ETSI (2001): Requirements and architectures for interworking between HIPER-LAN/3 and 3rd generation cellular systems. Technical report, ETSI TR 101 957.
4. Stemm, M., Katz, R.H. (1998): Vertical handoffs in wireless overlay networks. ACM Mobile Networks and Applications **3**(4) 335-350
5. Song, W., Jiang, H., Zhuang, W. (to appear): Performance analysis of WLAN-first scheme in Cellular/WLAN interworking. IEEE Trans. on Wireless Commun.
6. Zahran, A.H., Liang, B., Saleh, A. (2006): Modeling and performance analysis for beyond 3G integrated wireless networks. In Proc. of IEEE International Conference on Communications (ICC).
7. Latouche, G., Ramaswami, V. (1999): Intorduction to Matrix analytic Methods in Stochastic Modeling. ASA-SIAM series on Statistics and Applied Probability.
8. Papoulis, A., Pillai, S. (2002): Probability, Random Variables and Stochastic Processes. 4th edn. McGraw-Hill.
9. Hara, B.O., Calhoun, P., Kempf, J. (2005): Configuration and provisioning for wireless access points (CAPWAP) problem statement. RFC 3990.
10. Wang, J., Zeng, Q.A., Agrawal, D.P. (2003): Performance analysis of a preemptive and priority reservation handoff schemefor integrated service-based wireless mobile networks. IEEE Trans. Mob. Comput. **2**(1) 65-75.
11. Liang, B., Haas, Z.J. (2003): Predictive distance-based mobility management for multi-dimensional pcsnetworks. IEEE/ACM Trans. on Net. **11**(5) 718-732.