PGMS: Pseudo-optimal Greedy Media Streaming Algorithm for Heterogeneous Wireless Networks

Ahmed H. Zahran and Cormac J. Sreenan
Department of Computer Science
University College Cork, Ireland
[a.zahran, cjs]@cs.ucc.ie

Abstract

The integration of different wireless access technologies combined with the huge characteristic diversity of supported services in next-generation systems creates a real heterogeneous system. This heterogeneity opens new avenues for improving the system utility of both operators and users. In this paper, we propose a Pseudo-optimal greedy media streaming (PGMS) algorithm to reduce the cost of streaming sessions using an optimization framework that considers the real-time requirements of the vertical handoff decision. The proposed algorithm results in noticeable reductions in session cost, signaling load, and blocking probability in comparison to previously proposed heuristics. More importantly, the results show the adaptability of PGMS to different operating scenarios including different mobility patterns and service cost profiles.

1. Introduction

The future of wireless networking is envisioned as an integrated system of wireless radio access technologies with heterogeneous features, such as coverage, cost, and resources [1]. This integration is propelled by the interest in introducing new ubiquitous services such as gaming, conferencing, and media streaming. These applications span a wide spectrum of characteristics and quality of service (QoS) requirements. Hence, combining such applications and access technologies creates a real heterogeneous wireless system.

In such systems, roaming across heterogeneous domains, commonly known as vertical handoff (VHO), is a unique process that introduces several challenges to different system design aspects such as system architecture and handoff decision due to the embedded system heterogeneity. In the

mean time, this heterogeneity also brings several promising opportunities to improve the system utility for both operators and users. For example, Liang et al. [2] and Lau and Liang [3] employ document prefetching in cheap networks using a probabilistic framework to improve the user system utility. In this work, we focus on improving the user utility for streaming applications by benefiting from the cost and resource variations in different domains.

Streaming applications have a special buffering capability by which the application can save parts of the stream for future playout. It is worth mentioning that stream buffering is originally proposed to enable the application to avoid service interruption resulting from transmission errors, which are more probable in wireless systems due to possible channel deterioration. Combining this buffering capability with the cost and resource variation of wireless access technologies in integrated heterogeneous systems, one can clearly anticipate a possible opportunity for cost savings by buffering the stream in cheaper resource-rich intermittent technologies, e.g. WLANs in a 3G-WLAN integrated system.

In heterogeneous systems, greedy media streaming (GMS) evolves as a natural stream management algorithm for heterogeneous systems. In this context, GMS strategy entails downloading the stream at the maximum allowable rate in the cheaper network and avoid the utilization of the expensive network. However, we show in [4] that further cost savings can be attained using *heuristics*, whose design is inspired by the optimal streaming policy for a theoretical infinite streaming session. In this work, we propose a pseudo-optimal GMS (PGMS) streaming policy using an analytical framework to optimize the design of PGMS such that the stochastic expected streaming cost is minimized. The simulation results show that PGMS realizes further cost savings in comparison to previously proposed heuristics. More importantly, the results show that PGMS successfully adapts to different operating scenarios such as variable mobility patterns and cost profiles.

The rest of this paper is organized as follows. Section 2 introduces background and related work. In Section 3, we

present the system model and problem formulation. PGMS is presented in Section 4 followed by the simulation results in Section 5. Finally, conclusion and future work are presented in Section 6.

2. Background and Related Work

Audio and video streaming popularity is significantly growing in both the wired and wireless domains. Many techniques at different system levels have been proposed in the literature [5] to improve the user streaming experience. Initial playout latency has been proposed to avoid playout interruption due to transmission errors. Adaptive media playout is also proposed to decrease initial latency and accommodate transmission delay jitter [6]. Layered video compression is another technique that enables the user to have different levels of stream QoS.

Stream rate control is another adaptive technique that benefits from video compression to improve the user streaming experience. Practically, stream rate control can be performed using any session management protocol such as real-time streaming protocol (RTSP) [7]. In heterogeneous systems, stream rate control is of great importance due to the significant resource variations that accompany VHOs. These variations may lead to stream degradation on moving to a resource-limited access technology. Hence, adopting a proactive VHO algorithm, e.g. [8], combined with the media independent handover framework IEEE 802.21 [9] is highly recommended to avoid such degradation. Using this framework, the application can track the received signal strength of the current base station to predict the network transition instant and act proactively to avoid any possible stream QoS deterioration.

Few papers have addressed media streaming in heterogeneous systems. Many of these papers [10, 11] focus on improving the session quality through adjusting the stream coding rate according to the available network resources. In [12], using an experimental testbed, the authors demonstrate the possibility of performing seamless policy-triggered vertical handoff using Mobile IPv4 while running video sessions. These papers mainly focus on benefiting from bandwidth variations during VHOs ignoring other aspects of system characteristic diversity.

In [4], we initiate a new research direction for stream management in heterogeneous networks by considering both resource and cost changes that naturally accompany VHOs. We propose two *heuristic* stream management algorithms for two-tier heterogeneous systems, namely pseudo-optimal and greedy-optimal media streaming algorithms. The design of these algorithms is based on the analysis of a theoretical infinite session. This analysis suggests that the applications should buffer its expected application consumption during its residence in the expensive network. Ad-

ditionally, it proposes that the application should stream at a rate lower than the average stream rate in the expensive network if the buffering target is not reached. Among the previously proposed heuristics, the greedy optimal media streaming (GOMS) algorithm shows noticeable cost savings when compared to GMS streaming strategy at the cost of an increased signaling load and session blocking probability. This improvement is attained due to the optimal streaming policy enforcement by buffering at the maximum allowable rate in the intermittent network until the buffering target is attained. In this work, we develop PGMS as a novel streaming algorithm using an *analytical optimization framework*.

3. System Model and Problem Formulation

3.1. System Model

In our model, we assume a two-tier integrated wireless system composed of networks N_i for $i \in \{1, 2\}$, where 1 and 2 correspond to the technologies that provide universal and intermittent coverage respectively. Each network has a non-decreasing rate-dependent cost profile, denoted as $\chi_i(r_i)$ where r_i represents the data service rate in network N_i. Typically, r_i is non-negative and upper-bounded by a maximum service rate of r_{imax} in network N_i, i.e. $r_i \in [0, r_{imax}]$. Generally, network operators are expected to adopt monotonic pricing strategies as a congestion avoidance mechanism, especially with the increasing interest in an improved QoS support [13]. Note that the design of the network cost profiles in heterogeneous integrated systems is a non-trivial task as it represents one of the most important factors in VHO decision. We additionally assume that short range technologies will implement QoS management and admission control techniques to be capable of competing with other technologies in supporting real-time service requirements [14].

Figure 1 shows a typical scenario for a streaming session in a two-tier heterogeneous system. Generally, we assume that the session duration follows a generic heavy tailed distribution [15]. Clearly, at specific instants t_i during the session lifetime, the user encounters different events including session start, technological transitions, and session end. The durations spent by the user in different coverage combinations, denoted as τ_i, are assumed to have generic probability and cumulative density functions, denoted as $f_i(\tau_i)$ and $F_i(\tau_i)$ respectively. The parameters of $f_i(\tau_i)$ can be fitted to the collected measurement as presented in [16, 4]. Note that $\overline{\tau_i}$ represents the residual time distribution of τ_i.

The streaming application buffer status at any instant $t \in [t_k, t_{k+1})$, denoted as $x_{k+1}(t)$, is governed by the following differential equation

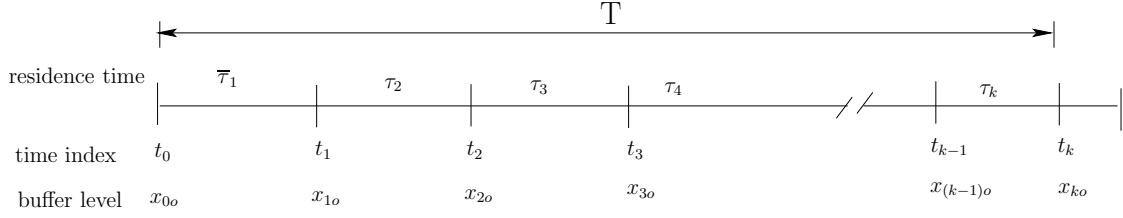$$x_{k+1}(t) = x_{ko} + (r_x - r_o)(t - t_k), \qquad (1)$$

Figure 1. Streaming Session

where x_{ko} represents the buffer state at the transition instant t_k and r_o represents the average playout data rate. That is to say that the buffer status at any instant equals the buffered information at the previous instant in addition to the difference between the downloaded and consumed data. To this end, it is worth pointing out that the presented problem ignores possible horizontal handoff (HHO) when the user roams between different cells of the universal overlay network. Generally, this process should not affect the streaming policy design. Though it may slightly impact the session blocking probability as the session may get dropped if the next-cell does not have enough resources.

3.2. Optimal Streaming Policy

Based on this model, our goal is to find the optimal streaming policy that minimizes the average session cost J_{av}. This policy defines a set of streaming rates r_x that minimize the average session cost, which can be expressed as

$$J_{av} = E\left[\frac{J}{T}\right] = E\left[\frac{\sum_{i=0}^{n} J_i}{\sum_{i=0}^{n} \tau_i}\right], \qquad (2)$$

where J represents the total session cost and J_i represents the session cost during the time interval $[t_{i+1}, t_i)$, i.e. τ_i. To this end, it is worth noting that there are practical bounds on the streaming rate in each stage. These bounds are an upper bound, r_{imax} determined by the network as the maximum average download rate per user and a lower bound r_{imin} determined to satisfy the playout quality by maintaining the buffer level above a pre-specified initial playout latency, ρ.

The presented optimal streaming policy falls under the umbrella of multistage stochastic sequential decision problems. Typically such problems take one of two forms: (i) multistage stochastic programs or (ii) stochastic dynamic programs [17]. The solution of these programs is computationally demanding because it usually involves generating and averaging a tremendous number of sample trajectories. Hence, using such approach is impractical due to the limited processing capacity of handheld devices and the VHO decision real-time requirement. Hence, our proposed policy is based on a suboptimal approach in which the cost optimization is performed on sub-durations of the streaming

sessions. Clearly, the streaming session is naturally subdivided into durations spent in different technologies, i.e., τ_i. Additionally, the cost saving is mainly attained by buffering in the cheap intermittent network and minimizing the utilization of the expensive one. Hence, we subdivide the streaming session into *optimization cycles*. Each cycle consists of two time components that correspond to the durations spent by the user under the coverage of intermittent and universal networks. In the following section, we present the proposed PGMS policy that is based on minimizing the streaming cost over these optimization cycles.

4. Pseudo-Optimal GMS (PGMS)

PGMS optimize the GMS heuristic by introducing a buffering threshold in the cheaper network. Hence, the application buffers the stream in WLAN at the maximum supporting streaming rate until the buffers hits this threshold, denoted as ρ_i. Typically, this buffer level is reached at a threshold crossing instant $\tau_w = \frac{\rho_i - x_o}{r_m - r_o}$, where x_o represents the buffer level at the beginning of the optimization cycle; i.e. at the beginning of visiting the cheaper intermittent network. At the threshold crossing instant, the application readjusts the streaming rate to the nominal rate. Consequently, the cost of the first stage, denoted as C_1, in the first part of the optimization cycle can be expressed as

$$C_1 = \begin{cases} \chi_1(r_m)\tau_1 & , \tau_1 < \tau_w \\ \chi_1(r_m)\tau_w + \chi_1(r_o)(\tau_1 - \tau_w) & , \tau_1 \geq \tau_w \end{cases}.$$

Similarly, the buffer level at the beginning of the second stage is expressed as

$$x_1 = \begin{cases} x_o + (r_m - r_o)\tau_1 & , \tau_1 < \tau_w \\ \rho_i & , \tau_1 \geq \tau_w \end{cases}.$$

On moving to the expensive network, the application pause the streaming process and the buffer is depleted at stream nominal rate; i.e. r_o. Hence, the buffer level may be depleted to the threshold value at time $\tau_{do} = \frac{x_1 - \rho}{r_o}$. Note that this depletion time is constant for $\tau_1 \geq \tau_w$ and is expressed as $\overline{\tau_{do}} = \frac{\rho_i - \rho}{r_o}$. At the depletion instance, the application continues streaming at the nominal rate as a recourse

action to maintain the streaming quality. Hence, the streaming cost in this stage, denoted as C_2, can be expressed as

$$
C_2 = \begin{cases}
0 & , \tau_1 < \tau_w, \ \tau_2 \le \tau_{do} \\
\chi_2(r_o)(\tau_2 - \tau_{do}) & , \tau_1 < \tau_w, \ \tau_2 > \tau_{do} \\
0 & , \tau_1 \ge \tau_w, \ , \tau_2 \le \overline{\tau_{do}} \\
\chi_2(r_o)(\tau_2 - \overline{\tau_{do}}) & , \tau_1 \ge \tau_w, \ \tau_2 > \overline{\tau_{do}}
\end{cases} .
$$

Our goal is to determine the optimal threshold value ρ_i^* that minimizes the average streaming cost over the optimization cycle. Hence, we are interested in the following problem

$$
\begin{aligned}
\min_{\rho_i} \int_0^{\tau_w} \chi_1(r_m)\tau_1 f_1(\tau_1) d\tau_1 + \\
\int_{\tau_w}^{\infty} (\chi_1(r_m)\tau_w + \chi_1(r_o)(\tau_1 - \tau_w)) f_1(\tau_1) d\tau_1 + \\
\int_0^{\tau_w} \int_{\tau_{do}}^{\infty} \chi_2(r_o)(\tau_2 - \tau_{do}) f_2(\tau_2) d\tau_2 f_1(\tau_1) d\tau_1 + \\
\int_{\tau_w}^{\infty} \int_{\overline{\tau_{do}}}^{\infty} \chi_2(r_o)(\tau_2 - \tau_{do}) f_2(\tau_2) d\tau_2 f_1(\tau_1) d\tau_1
\end{aligned} \quad . \quad (3)
$$

By some algebraic manipulation, the streaming cost of a complete optimization cycle can be rewritten as

$$
\begin{aligned}
\min_{\rho_i} \ \frac{\chi_1(r_o)}{\mu_1} + \frac{\chi_2(r_o)}{\mu_2} + \\
(\chi_1(r_m) - \chi_1(r_o)) \int_0^{\tau_w} (1 - F_1(\tau_1)) \, d\tau_1 - \\
\chi_2(r_o) \left[\int_0^{\tau_w} \int_0^{\tau_{do}} (1 - F_2(\tau_2)) \, d\tau_2 f_1(\tau_1) d\tau_1 \right] - \\
\chi_2(r_o) \left[\int_{\tau_w}^{\infty} \int_0^{\overline{\tau_{do}}} (1 - F_2(\tau_2)) \, d\tau_2 f_1(\tau_1) d\tau_1 \right]
\end{aligned} \quad . \quad (4)
$$

Note that the optimal threshold ρ_i^* is lower-bounded by the initial playout buffer level to satisfy a basic streaming QoS requirement. Additionally, the result of the integrals is generally nonlinear in the optimization parameter. Hence, the presented program is classified as a nonlinear optimization program with boundary constraints. The solution of this program requires two main design decisions

- determining generic mobility distributions for different mobility patterns, and

- choosing a solution approach for the resultant program.

In this work, we use phase-type distributions [18], which are chosen due to their flexibility and analytical tractability. More specifically, we fit collected measurements based on the coefficient of variation of the collected measurements, denoted as θ_x[1] to hyper-exponential, exponential, and hypo-exponential distributions for $\theta_x > 1$, $\theta_x = 1$, and $\theta_x < 1$ respectively. These distributions are chosen due to their simple structure and computationally inexpensive fitting procedures. Note that the simple structure of these distributions enables obtaining closed form for the integrals in eq(4). The details of the fitting process and the resultant expressions are omitted due to page limit.

---

[1]$\theta_x = \frac{\sigma_x}{\mu_x}$, where σ_x and μ_x represent the standard deviation and the mean of the corresponding measurements respectively.

Table 1. Simulation Parameters

Param.	Value	Param.	Value	Param.	Value
r_{dmax}	2MBps	r_{umax}	50 KBps	r_o	25 KBps
a_u	3.2e-9	b_u	40e-6	b_d	70e.6
μ_T	22min	θ_T	1.11	ρ	6 sec
P_v	0.01	P_h	0.01		

Table 2. Statistics of mobility patterns

	μ_c	σ_c	μ_1	σ_1	μ_2	σ_2
0	703.09	1097.8	122.63	415.70	55.97	143.84
0.25	563.63	828.71	103.95	312.93	47.75	107.55
0.5	461.39	625.64	89.07	228.61	40.45	77.23
0.75	343.66	418.05	76.73	148.91	34.92	52.48
1	248.70	264.04	64.69	81.42	29.53	36.56

The solution of nonlinear programs is generally realized using different numerical techniques. Practically, the adopted optimization technique should be simple, easy to implement, and applicable to any model without restrictions. All these features are satisfied by *direct search* methods [19]. More importantly, these methods are known for their speed and robustness when the application target is one or two digits of accuracy, which satisfies the requirements of the presented program.

5. Simulation Results

In order to assess the performance of the proposed algorithms, we simulate a 3G-WLAN integrated system using NS-2 [20]. Table 1 shows the default values of the system parameters. We adopt zone residence time model [16, 21] for mobility simulations. Additionally, we assume that $\chi_1(r) = a_1 r^2 + b_1 r$ and $\chi_2(r) = b_2 r$. P_v and P_h represent the blocking probabilities of vertical and horizontal handoffs respectively. The chosen maximum network service rates represent possible rate allocation strategy of CDMA2000 and 802.11g. The session has a hyperexponential distribution with mean μ_T and coefficient of variation θ_T. In the following figures, we compare the performance of the GMS, GOMS, and PGMS algorithms versus different mobility patterns and pricing strategies. Each point represents the mean of 1000 sessions with its corresponding 95% confidence interval. For the optimization problem solution, we use the parallel direct search method from the OPT++ optimization package [22].

5.1. Mobility Randomness Impact

In this subsection, we investigate the impact of user mobility randomness on different performance metrics. In the following figures, 0 and 1 in the abscissa correspond to
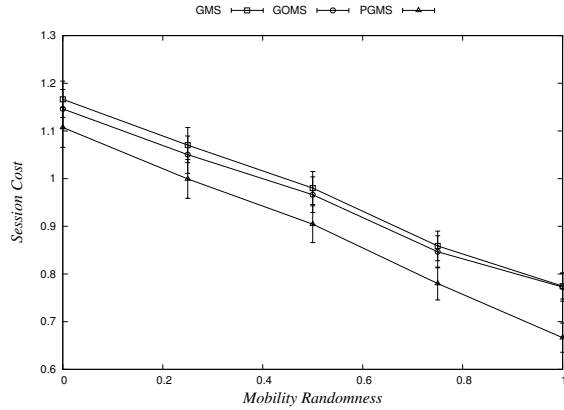
Figure 2. Session cost versus WLAN mobility randomness

Figure 3. Session HHO versus WLAN mobility randomness

random-walk and fluid-flow respectively. Hence, the abscissa scans the complete mobility spectrum. Table 2 shows the average and standard deviation of cell residence time, unique coverage zone, and dual coverage zone respectively denoted as μ_c, σ_c, μ_1, σ_1, μ_2, and σ_2. To this end, it is worth noting that as the mobility randomness decreases, the user tends to have shorter residence times in different technologies and more frequent transitions between zones as well as overlay cells. Hence, the mobility behavior has a significant impact on the obtained performance metrics as will be shown.

Figure 2 plots the average session cost versus different mobility patterns. Clearly, the figure shows a decreasing trend in the average session cost as the user mobility becomes more uniform. This drop is a natural consequence to frequent WLAN visits as the user mobility becomes more uniform. Additionally, the figure shows the superior performance of PGMS in comparison to GMS and GOMS. More importantly, it is worth noting that the cost saving gap between GOMS and GMS shrinks as the user mobility becomes less random (toward fluid-flow). On contrary, PGMS maintains a noticeable cost saving margin in comparison to GMS. This improvement in adaptation to mobility randomness is due to the dependence of optimal buffering threshold of PGMS on the residence time distribution, which provides complete information about the mobility pattern statistics, while GOMS only depends on the first order statistics.

Figure 3 plots the expected number of *executed* horizontal handoff (HHO) rate versus mobility randomness. In this context, an executed HHO corresponds to a situation in which the application is streaming while handing off from one cell to another in a unique coverage area. Note that
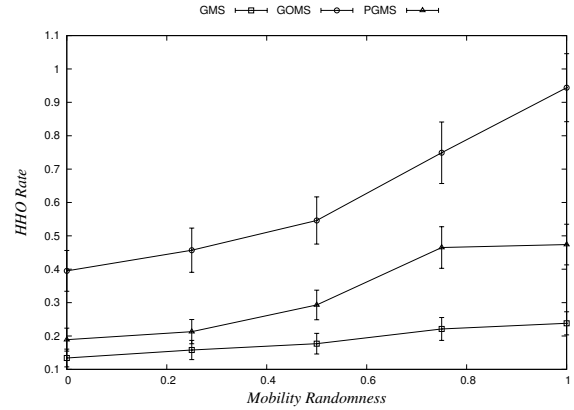
the user will not request resource allocation if it is crossing cells through a dual coverage zone or if it is relying on the buffered data. Hence, in the latter two cases, the application will not initiate any handoff procedure. Clearly, Figure 3 shows an increasing trend in signaling load as the mobility randomness decreases. This increase is due to the frequent transitions that take place as the mobility pattern shifts toward fluid-flow. Additionally, the figure shows that GMS results in the least HHO signaling load followed by PGMS and then GOMS. This result is a natural consequence for the persistent greedy buffering policy in WLANs and depending on this reserve in the cellular network. On contrary, PGMS adaptively chooses a buffering threshold to minimize the session cost. Similar results are obtained for VHO signaling; however, results are not shown due to page limit.

Figure 4 plots the RTSP signaling in 3G and WLAN networks for the presented algorithms versus different mobility patterns. The noticeable gap between the RTSP signaling in both networks is due to adopting a proactive VHO strategy. Hence, most of the RTSP messages are transmitted in the cheaper network just after moving into the dual zone or proactively before leaving the WLAN. Clearly, PGMS leads to a noticeable reduction in RTSP signaling load in the cellular network, though the reduction in RTSP WLAN is not significantly improved when compared to GOMS.

Figure 5 plots the forced termination probabilities due to both VHOs and HHOs for different algorithms versus different mobility patterns. As a natural consequence for the fewest executed handoffs performed by GMS, it also has the lowest forced termination probabilities. For similar reasons, PGMS has higher blocking probability, followed by GOMS. However, it is worth noting that increase in the
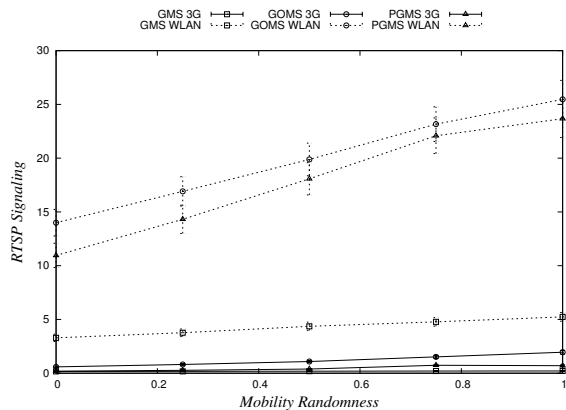
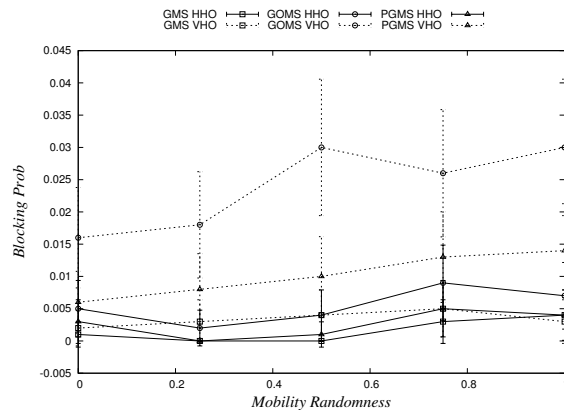**Figure 4. Session RTSP signaling versus WLAN mobility randomness**

Figure 5. Session blocking probabilities versus WLAN mobility randomness

blocking probability is insignificant.

5.2. Pricing Strategy Impact

Figure 6 plots the average session cost versus WLAN cost coefficient; i.e. b_2. Note that as b_2 decreases, the pricing strategy in the intermittent network becomes more flat. Clearly, the figure shows that using a static buffer threshold, as in GOMS, ignores possible further cost savings as the service price ratio drops. On the contrary, PGMS optimization framework successfully adapts to variable pricing strategies. Hence, it maintains its superiority as an optimal cost saving strategy under different operating scenarios including different mobility patterns and pricing strategies.

6. Conclusion

The service integration of wireless access technologies combined with the huge characteristic diversity of different applications creates a new heterogeneous networking paradigm that opens the door to improving the user system utility. In this paper, we propose a novel management policy for media streaming in heterogeneous systems. The proposed PGMS algorithm shows noticeable cost reductions in comparison to previously proposed heuristic-based algorithms. More importantly, the algorithm shows great adaptability to different operating scenarios including different mobility patterns and pricing strategies. Hence, we foresee the PGMS algorithm as an initial promising heuristic solution for reduced cost media streaming in next-generation heterogeneous wireless networks. As future work, we are interested in introducing new dimension to the user satis-

faction such as perceived stream quality. Additionally, we are interested in implementing the proposed solution in a real testbed.

References

[1] M. M. Buddhikot, G. Chandranmenon, S. Han, Y. W. Lee, and S. M. L. Salgarelli, "Integration of 802.11 and Third Generation Wireless Data Networks," in *Proc. of IEEE INFOCOM*, San Francisco, US, Apr. 2003, pp. 503 – 512.

[2] B. Liang, S. Drew, and D. Wang, "Performance of multiuser network-aware prefetching in heterogeneous wireless systems," *to appear in ACM-Springer Wireless Networks*.

[3] J. Lau and B. Liang, "Optimal pricing for selfish users and prefetching in heterogeneous wireless networks," in *Proc. of the IEEE International Conference on Communications (ICC 07)*, June 2007.

[4] A. H. Zahran and C. J. Sreenan, "Cost efficient media streaming algorithms for rate-dependent pricing strategies in heterogeneous wireless networks," submitted for publication.

[5] D. Wu, Y. Hou, W. Zhu, Y.-Q. Zhang, and J. Peha, "Streaming video over the Internet: approaches and directions," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 282–300, Mar 2001.
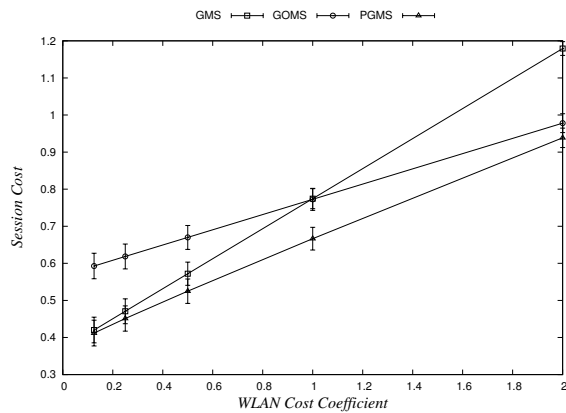
GMS ⊢—□—⊣   GOMS ⊢—◇—⊣   PGMS ⊢—△—⊣

Figure 6. Session cost versus WLAN cost co-efficient

[6] E. Steinbach, N. Farber, and B. Girod, "Adaptive play-out for low latency video streaming," in *Proc. International Conference on Image Processing*, vol. 1, 2001, pp. 962–965.

[7] H. Schulzrinne, A. Rao, and R. Lanphier, "Real Time Streaming Protocol," RFC 2326, April 1998.

[8] A. Zahran, B. Liang, and A. Saleh, "Signal Threshold Adaptation for Vertical Handoff in Heterogeneous Wireless Networks," *ACM/Spring Mobile Networks and Applications (MONET)*, vol. 11, no. 4, pp. 625 – 640, Aug 2006.

[9] IEEE, "Local and Metropolitan Area Networks: Media Independent Handover Services," Draft Standard, Feb. 2007.

[10] L.-J. Chen, T. G. Y. Sun, M. Sanadidi, and M. Gerla, "Adaptive video streaming in vertical handoff: a case study," in *Proc of The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services*, 2004, pp. 111 – 112.

[11] J.-W. Ding, C.-T. Lin, and K.-H. Huang, "ARS: an adaptive reception scheme for handheld devices supporting mobile video streaming services," in *Proc. International Conference on Consumer Electronics (ICCE '06)*, 2006, pp. 141– 142.

[12] A. Schorr, A. Kassler, and G. Petrovic, "Adaptive media streaming in heterogeneous wireless networks," in *Proc. of IEEE 6th Workshop on Multimedia Signal Processing*, 2004, pp. 506 – 509.

[13] L. Badia, S. Merlin, A. Zanella, and M. Zorzi, "Pricing VoWLAN services through a micro-economic framework," *IEEE Wireless Commun.*, vol. 13, no. 1, pp. 6–13, Feb. 2006.

[14] G. Boggia, P. Camarda, L. A. Grieco, and S. Mascolo, "Feedback-based control for providing real-time services with the 802.11e mac," *IEEE/ACM Trans. Netw.*, vol. 15, no. 2, pp. 323–333, 2007.

[15] M. Li, M. Claypool, R. Kinicki, and J. Nichols, "Characteristics of streaming media stored on the Web," *ACM Trans. Inter. Tech.*, vol. 5, no. 4, pp. 601–626, 2005.

[16] A. H. Zahran and B. Liang, "A Generic Framework for Mobility Modeling and Performance Analysis in Next-Generation Heterogeneous Wireless Networks," *IEEE Commun. Mag.*, vol. 45, no. 9, pp. 92–99, Sep. 2007.

[17] M. S. Casey and S. Sen, "The scenario generation algorithm for multistage stochastic linear programming," *Math. Oper. Res.*, vol. 30, no. 3, pp. 615–631, 2005.

[18] G. Latouche and V. Ramaswami, *Introduction to Matrix analytic Methods in Stochastic Modeling*, ser. ASA-SIAM series on Statistics and Applied Probability. SIAM, 1999.

[19] R. M. Lewis, V. Torczon, and M. W. Trosset, "Direct search methods: then and now," *J. Comput. Appl. Math.*, vol. 124, no. 1-2, pp. 191–207, 2000.

[20] "Network Simulator." [Online]. Available: http://www.isi.edu/nsnam/ns/

[21] A. H. Zahran, "Modeling and Design of Next-Generation Heterogeneous Wireless Networks," Ph.D. dissertation, University of Toronto, 2007.

[22] J. C. Meza, R. A. Oliva, P. D. Hough, and P. J. Williams, "OPT++: An object-oriented toolkit for nonlinear optimization," *ACM Trans. Math. Softw.*, vol. 33, no. 2, p. 12, June 2007.