Load Balancing for Cellular/WLAN Integrated Networks

Wei Song and Weihua Zhuang, University of Waterloo
Yu Cheng, Illinois Institute of Technology

Abstract

The interworking between heterogeneous third-generation cellular networks and wireless local area networks is one promising evolution approach to fourth-generation wireless networks, which can exploit the complementary advantages of the cellular network and WLANs. Resource management for the 4G-oriented cellular/WLAN integrated network is an important open issue that deserves more research efforts. In this article we present a policy framework for resource management in a loosely coupled cellular/WLAN integrated network, where load balancing policies are designed to efficiently utilize the pooled resources of the network. A two-phase control strategy is adopted in the load balancing policies, in which call assignment is used to provide a statistical quality of service guarantee during the admission phase, and dynamic vertical handoff during the traffic service phase is used to minimize the performance variations. Numerical results are presented to demonstrate that the proposed load balancing solution achieves significant performance improvement over two other reference schemes.

Motivated by the ever-increasing demand for wireless communications, the cellular network has evolved to the third generation (3G), for example, the Universal Mobile Telecommunication System (UMTS), which is specified by the Third Generation Partnership Project (3GPP) and is one of the most popular 3G systems nowadays. The 3G cellular network is capable of supporting quality of service (QoS) critical to multimedia services, but at the expense of high complexity and implementation cost. For example, four service classes are supported in the UMTS: conversational, streaming, interactive, and background services. However, the expensive radio spectrum for 3G cellular networks prohibits rapid deployment, and the low bandwidth restricts system capacity. The future fourth-generation (4G) wireless networks need to effectively address these existing constraints and problems. Heterogeneous networking is a promising approach to accelerate the technological evolution toward 4G wireless networks. In recent years, IEEE 802.11 wireless local area networks (WLANs) have proliferated due to a high performance-to-cost ratio. Usually operating at license-free frequency bands, WLANs can occupy a much wider spectrum than the cellular system, and provide data services using a simple medium access control (MAC) protocol. The complementary characteristics of the 3G cellular network and WLANs promote their interworking. Future mobile devices can be equipped with network interfaces to both the 3G network and WLANs at a reasonable price. The dual-mode mobile devices can then enjoy enhanced services in the cellular/WLAN integrated network.

The standardization for cellular/WLAN interworking is now underway in 3GPP from the cellular operator's perspective. Six interworking scenarios are defined in 3GPP TR 22.934 to implement 3GPP/WLAN interworking step by step. In the latest Release 6 of the 3GPP standard, the first three interworking levels are included, which support 3GPP-based access control and enable packet-switched services to integrated WLANs. In 3GPP TS 23.234 a reference model is specified for 3GPP/WLAN interworking and deals with many details in various security aspects. Less effort has been devoted to QoS provisioning except for 3GPP TR 23.836, which was drafted in November 2005 and limited to very high-level discussion. In the literature, there are many research works on vertical handoff for cellular/WLAN interworking. For example, many researchers study the application of cross-layer techniques (e.g., link layer triggering) and soft handoff to reduce handoff latency and packet loss during vertical handoff. Also, intelligent vertical handoff decisions are investigated to select a best handoff target from multiple heterogeneous candidate networks [1].

With cellular/WLAN interworking, the resources of the two networks can be viewed as a shared resource pool. Management of the pooled resources in the integrated network is another essential research issue for cellular/WLAN interworking. Given the two-tier overlaying structure and heterogeneous underlying mobility and QoS support, an important aspect of the resource management problem is how to properly distribute incoming traffic loads to the integrated systems so that the interworking is exploited to maximize the overall utilization. Optimal assignment strategies of selecting a target system for incoming calls are proposed in [2] only for data service based on user mobility and traffic characteristics. In [3], an optimal joint session admission control scheme based on a semi-Markov decision process is proposed for multiclass traffic, which maximizes overall network revenue with QoS constraints. Dynamic session transfer for hierarchical integrat-

0890-8044/06/$20.00 © 2006 IEEE

ed networks is studied in [4] as an analog to the task migration in distributed operating systems. In this study we aim to develop a more comprehensive solution to the resource management problem in order to balance multiservice traffic loads over integrated cellular and WLAN systems. We take into account more implementation-specific aspects, and more details of practical cellular and WLAN systems. A policy-based resource management framework is developed to implement the load balancing policies. First, multiservice incoming calls are properly distributed to the cellular network and WLAN by admission control. Further, the traffic loads are dynamically transferred between the two networks via vertical handoff when a user moves within the overlaying area served by both the cellular network and WLAN. The load sharing problem for cellular/WLAN interworking involves the dynamics of both integrated systems, which is very complex, particularly for a multiservice scenario. In [5] we analyzed the impact of various user mobility and traffic parameters on system utilization. Here, we further consider dynamic vertical handoffs within the WLAN, which are not only triggered by user mobility but also by network states.

The remainder of the article is organized as follows. The system model of the cellular/WLAN integrated network, which adopts a loosely coupled interworking architecture and a policy-based resource management framework, is presented. We elaborate the load balancing policies via admission control and vertical handoff, and then discuss the performance evaluation and numerical results. Conclusions and further work are given in the final section.

System Model

Cellular/WLAN interworking can take good advantage of their complementary characteristics and provide enhanced services to mobile users. According to the interdependence of the two networks, the integration can have a relatively tight or loose coupling architecture. The cellular/WLAN interworking can be very tight with integration at the access network (AN) or less tight with integration in the core networks (CNs). The interworking is even looser when the two networks are integrated beyond the core networks and usually through an external Internet Protocol (IP) network. The loosely coupled architecture enables independent deployment of the two networks and usually follows standard mechanisms in the IP community, such as the Mobile IP for mobility management and the authentication, authorization, and accounting (AAA) framework for user access control. However, the loosely coupled architecture is relatively inefficient because of the long signaling path, redundant processing in the two networks, and the large number of network elements involved for management operations. Extra mechanisms are needed to overcome this inefficiency, for example, by adopting cross-layer techniques and context management. On the other hand, tight integration induces higher implementation complexity, since the WLAN needs to have a compatible interface to the cellular core network and even to the cellular radio access network. Nonetheless, since cellular-like operations are followed in the tightly coupled network, efficient management is achievable.

In fact, interworking is only an interim solution. It is expected that the future wireless network would be a converged network, in which a common IP-based core network is shared by a variety of access networks [6]. The access technology-specific functions only propagate up to the gateway to the core network. As such, the access heterogeneity terminates within the access network, and homogeneous management is provided for mobility, security, and QoS. Although heterogeneous systems in the future converged network will share a common core network similar to the cellular/WLAN tight coupling, the shared core network is not necessarily the cellular core network as in the tightly coupled interworking architecture, but may be a separate IP backbone network. The evolution of both the cellular network and WLANs is consistent with the *all-IP* direction. Since the WLAN is designed to be a wireless extension of the wired Ethernet, only the physical-layer and link-layer specifications are defined in WLAN standards. It is a natural choice for the WLAN to adopt IP-based protocols for higher-layer operations. The 3G cellular network is also evolving toward the all-IP network. Many IP-based technologies such as the Mobile IP and AAA framework are introduced into the cellular core network, which functions more and more similarly to an IP backbone.

Loosely Coupled Interworking Architecture

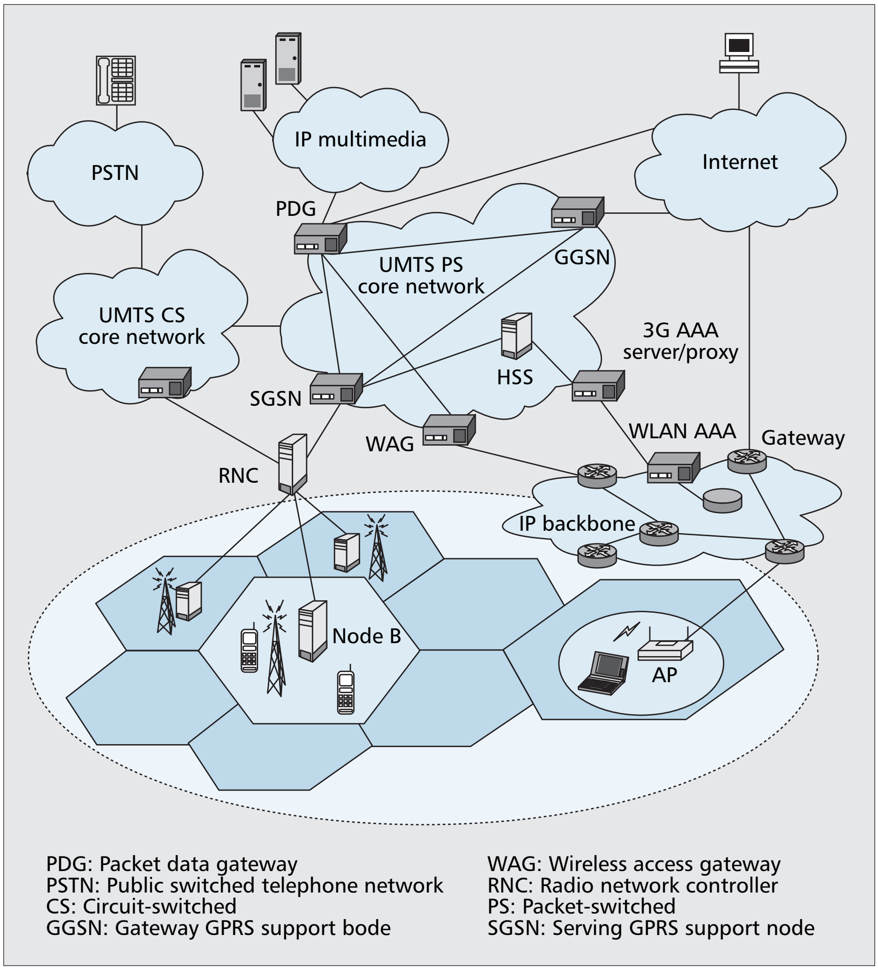
With the converging trend of wireless networks, the interworking system model should target future converged networks. Considering the advantages of independent deployment and implementation flexibility, loose coupling is adopted in this study to integrate the cellular network and WLANs via an external IP-based core network. As such, the resource management discussed in this article can be extended to future converged wireless networks with a common IP-based core network for heterogeneous access networks.

Based on the reference model in 3GPP TS 23.234 and the gateway approach in [7], we consider a cellular/WLAN interworking architecture as illustrated in Fig. 1. The UMTS system is taken as an example of the cellular network for the interworking. The mobile users roaming in the integrated network may subscribe to the cellular network, the WLAN, or both networks. According to the subscription, the mobile users can be authenticated and authorized by the AAA server in their home network before they can access the services. For example, for a 3G subscriber who roams into the WLAN, the authentication request can be relayed by the WLAN AAA server/proxy to its home network (the 3G network), which performs the authentication with the aid of the home subscriber server (HSS). On the other hand, mobility management between the cellular network and WLAN can be supported by introducing Mobile IP functionality into the cellular core network. As defined in 3GPP TR 22.934, for interworking level 3 and beyond, the 3G packet-switched services are accessible from WLANs. For example, by incorporating WLAN access gateway (WAG) and packet data gateway (PDG), WLAN users can access 3G-specific services in the IP multimedia subsystem (IMS). Other external IP networks such as the public Internet also become accessible through the 3G network similarly in addition to direct IP access from the WLAN. Therefore, the gateway routers, such as the gateway general packet radio service (GPRS) support node (GGSN) and PDG, should be equipped with home agent (HA) and foreign agent (FA) functions.

Policy-Based Resource Management

A policy-based approach is increasingly popular for network management with the merits of flexibility and scalability. The service QoS requirements and details of network configuration are decoupled to facilitate dynamic control. The Internet Engineering Task Force (IETF) has defined a policy-based management framework [8], which has been adapted and introduced in the QoS architecture of UMTS systems [9]. The adaptive and decoupling properties of the policy framework can properly address the challenges posed by the inherent heterogeneity of cellular/WLAN integrated networks.

In this study we consider a policy-based framework for resource management in the cellular/ WLAN integrated network, as shown in Fig. 2, which extends the policy architecture for call admission control and handoff management presented in [10]. Within the policy framework, the policy components can be implemented as logical entities at the mobile terminal (MT), the radio network controller (RNC) of the UMTS network, and the access point (AP) of the WLAN shown in Fig. 1. When a policy event is triggered, an entity called a policy decision point (PDP) retrieves corresponding policy rules from the *policy repository* to make a policy decision. The policy events registered at the MT and networks are different but correlated. Consistency among the corresponding policy repositories needs to be maintained via service level agreement (SLA) negotiation. Then the decision is transformed into configuration actions, which are sent to the entity called the policy execution point (PEP). The PEP executes these actions in response to the triggered policy event. Note that the policy-based management at both networks operates in a distributed manner. That is, the policy decision and execution are performed separately in each network, although network information is exchanged by proper signaling and taken into account when making a decision. Equipped with dual network interfaces, the MT can act as a relay node between the cellular network and WLAN. This management framework matches well the loosely coupled architecture in Fig. 1.



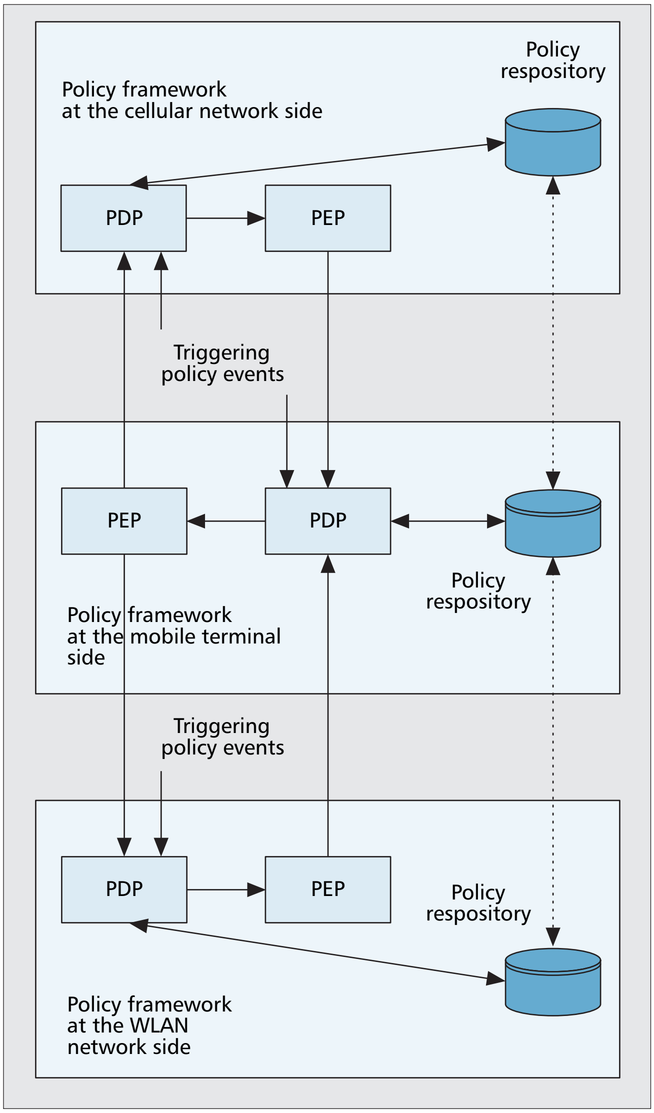
■ Figure 1. *The loosely coupled UMTS/WLAN interworking architecture.*

Load Balancing by Admission Control and Vertical Handoff

In the cellular/WLAN integrated network, the cellular network provides ubiquitous coverage, while the WLANs are deployed disjointly in hotspot areas. In the area with WLAN coverage (referred to as *double coverage area* below), access to both the cellular network and WLAN is available, while there are also service areas with only cellular access (referred to as *cellular only areas*). With this two-tier overlaying structure, a new call in the double coverage area can be admitted to either the cell or the WLAN. Moreover, ongoing calls can be dynamically transferred between the cell and the WLAN by vertical handoff, which may not be necessary to maintain a call, but is mainly for load balancing and/or QoS enhancement. The cellular network and WLANs are complementary in QoS provisioning for different services. In the cellular network, the base station controls access to the shared radio link and reserves resources for admitted calls in a cell. This centralized control and reservation-based resource allocation enables fine-grained QoS. On the other hand, the MAC protocol in WLANs is usually a contention-based random access protocol, for example, the distributed coordination function (DCF) of IEEE 802.11. Because of inevitable collisions and backoffs, it is difficult for this type of MAC to support services with strict QoS requirements such as real-time voice service, although it is efficient in serving bursty data traffic. With this complementary QoS support, when multiple services are considered, the traffic loads of different classes should be properly distributed to the cells and WLANs by admission control and vertical handoff.

As discussed above, the overlaying structure of the cellular/WLAN integrated network results in the network selection problem, which is very challenging and needs to take into account various factors, such as the characteristics of the heterogeneous underlying networks, network condition, service type, user mobility, user preference, and service cost [11]. By combining the factors in different manners, many selection algorithms have been proposed to select a best access network for an incoming service request so that predefined objectives are achieved, such as providing the best performance to the incoming traffic in an efficient and cost-effective manner. Some advanced techniques such as fuzzy logic, analytic hierarchy process (AHP), and grey relational analysis (GRA) [1] are employed to make a best decision. The operational complexity of the decision algorithms introduces some implementation problems and prevents their online application in a highly dynamic network.

To maximize the utilization of the pooled resources in the integrated network, it is necessary to adaptively balance the dynamic multiservice traffic loads over the two networks. A feasible solution to the load balancing problem should well address the traffic dynamics and may need to trade off performance for affordable complexity. Here, we propose a two-



■ Figure 2. *The policy-based resource management framework for the cellular/WLAN integrated network.*

phase control strategy for load balancing. First, call arrivals in the overlaying area are directed to a target cell or WLAN based on an assignment probability, which is adapted to traffic and network dynamics. With accurate traffic estimation and performance evaluation, the overall traffic loads can be properly balanced over the two networks, so as to achieve consistent performance across the two networks and ensure QoS satisfaction in a statistic sense. Second, to minimize the performance variations induced by network dynamics and traffic randomness, the complementary nature of the two interworked systems can be exploited by vertical handoff. For example, with observed congestion or severe performance degradation in one network, some ongoing sessions can be handed over to the other network by vertical handoff. In summary, by applying this two-phase strategy, we aim to achieve good load balancing and efficient utilization. Statistic QoS guarantee is provided by call assignment during the admission phase, and performance variations are reduced by dynamic vertical handoff during the traffic lifetime. In the following, we discuss the load balancing policies for voice and data services as an example. Depending on the triggering policy event, the PDP and PEP components interact and collaborate to respond to the incoming policy event, as illustrated in Fig. 3.

New Call Assignment Policy

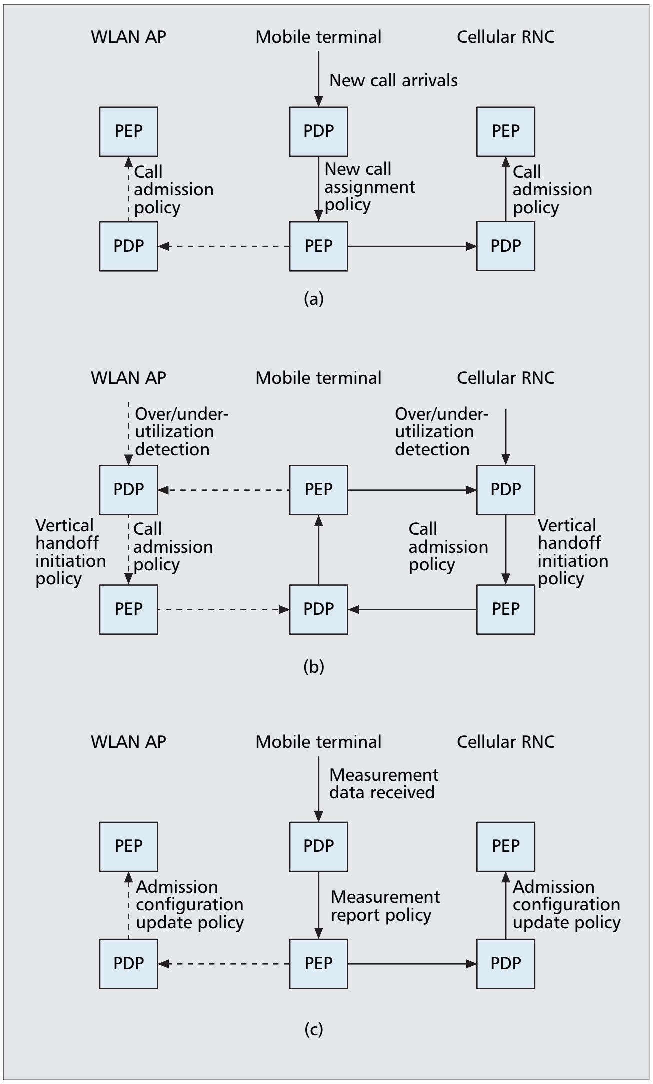
As shown in Fig. 3a, triggered by a new call arrival at an MT, the new call assignment policy is referred to for a policy decision on the target network to admit the new call. Depending on the service type of the incoming call, a voice (data) call is directed to the cell with a probability θ_v^c (θ_d^c) and to the WLAN with a probability $\theta_v^w = 1 - \theta_v^c$ ($\theta_d^w = 1 - \theta_d^c$). The parameters θ_v^c and θ_d^c (or θ_v^w and θ_d^w) are determined based on traffic load estimation from previous measurements and need to be dynamically adapted to load variations. The contention-based resource sharing in WLANs results in different QoS support for real-time voice service and delay-insensitive data service with different traffic characteristics. To ensure that the WLAN operates at its efficient states and complements the cellular network effectively, the resource should be properly shared between the voice and data traffic by controlling the admission. In [12] we propose a QoS evaluation approach for voice and data services in a cellular/WLAN interworking scenario, based on which the parameters θ_v^c and θ_d^c can be determined for an estimated load. Once the policy decision is made based on the assignment probabilities, the PEP at the MT sends admission request to the target network, which triggers the event involving the call admission policy at the network PDP.

Call Admission Policy

Triggered by the event of a call admission request, the call admission policy at the network PDP is referred to for a decision. In order to bound the delay for voice service and guarantee the throughput for data service, it is necessary to restrict the numbers of voice and data calls admitted in the cell and WLAN. There have been extensive works on admission control schemes for the cellular network and WLANs. The admission region is usually defined by a set of vectors (N_v, N_d), which respectively denote the maximum numbers of voice and data calls simultaneously carried by the network with QoS guarantee. Given the system bandwidth, the admission region can be derived for an offered traffic load to the network and needs to be dynamically adjusted with traffic load changes. In the cellular/WLAN interworking scenario, the vertical handoff traffic load also plays an important role in determining the admission region. New calls, horizontal handoff calls, and vertical handoff calls can be further differentiated in the admission policy such as in [12].

Vertical Handoff Initiation Policy

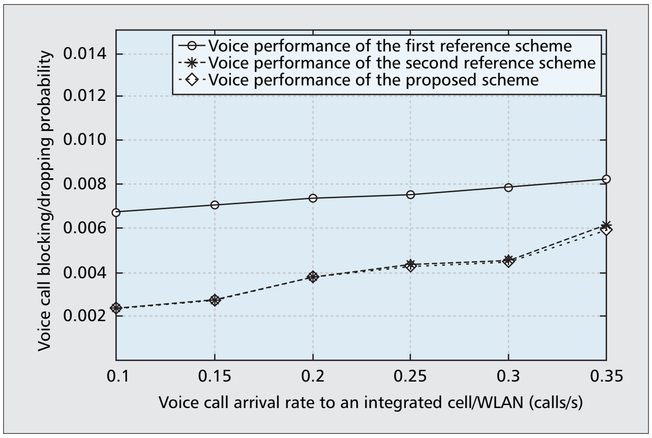
As shown in [13], an effective indicator for the channel utilization of the WLAN is the busyness ratio, which is defined as the ratio of the time that the channel is determined busy to the total time. The maximum WLAN channel utilization is found to exist in the unsaturated case instead of the saturated case (in which each admitted user always has packets in the queue and thus keeps contending for the channel). Consequently, it is reasonable to dynamically balance the traffic loads from and to the cellular network by vertical handoff, so as to make sure that the WLAN operates in its most efficient region. Figure 3b illustrates the interaction between policy entities to perform the load balancing by vertical handoff. First, over an observation window, if the busyness ratio detected at the AP is above a predefined threshold ε_w^u, the policy event involving the vertical handoff initiation policy is triggered at the PDP of the WLAN. The ongoing calls carried by the WLAN are prioritized based on their service type, remaining service time, subscription profile, and so forth. According to the priority order, the vertical handoff initiation policy decides to hand over certain calls from the WLAN to the cell. The voice calls from cellular subscribers are selected first for



■ Figure 3. *The interaction between load balancing policies: a) the interaction between new call assignment policy and call admission policy; b) the interaction between vertical handoff initiation policy and call admission policy; c) the interaction between measurement report policy and admission configuration update policy.*

vertical handoff, because the cellular network is efficient in supporting voice service and serves the calls from its subscribers with priority. The policy decision is prompted to the PDP of the MT carrying the selected call. Upon receiving the vertical handoff request, the PDP of the MT directs the PEP to send an admission request to the corresponding cell, which in turn triggers the policy event involving the call admission policy at the PDP of the cellular network.

Second, from the measurement report relayed by the MT, the information about the WLAN network condition is available to the base station of the cellular network. If the busyness ratio of the WLAN is observed to be consistently below a threshold ε_w^l for a period of time, it triggers the policy event, which invokes the vertical handoff initiation policy at the cellular network. The cellular PDP makes decisions on transferring a certain traffic load to the WLAN via vertical handoff. Data calls with a long remaining service time and from mobile users with the WLAN as their home network can be first directed to hand over to the WLAN. The rationale behind this principle is to exploit the large bandwidth and high efficiency of the WLAN in supporting data service when the WLAN is underutilized, so that the cellular network is prevented in advance from becoming a system bottleneck. The decision to initiate vertical handoff made by the cellular PDP is then transformed into signaling messages sent to the MT by the cellular PEP. Finally, the PDP of the MT directs its PEP to request admission to the WLAN for potential vertical handoff.



■ Figure 4. *The voice call blocking/dropping probabilities of the three schemes in comparison.*

Measurement Report Policy

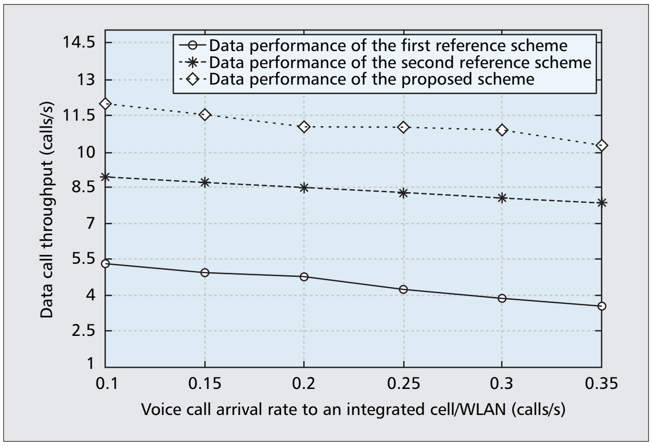
To perform resource management for the integrated network in a distributed manner without a central controller, it is necessary to exchange network information between the interworked systems so that the potential gain of joint management over independent control is well exploited. However, as shown in Fig. 1, the cellular network and WLANs are interworked through an external IP network. This loose coupling results in a large overhead and long latency for the signaling exchange between the two networks. A possible solution is to utilize the MT with dual network interfaces as a relay node. Since ubiquitous coverage is provided by the cellular network, the cellular network interface of the MT can be always enabled for controlling messages, even when the MT is connected to the WLAN for data transmission [7]. As a result, the MT can keep receiving periodic advertisement messages from the cellular network indicating its network condition such as link performance, channel utilization, traffic load, data throughput, call blocking/dropping rates, and so on. Then the MT assembles the cellular network information and relays it to the AP of the WLAN via its WLAN interface. Similarly, the network information of the WLAN can also be relayed to the base station of the cellular network. As illustrated in Fig. 3c, with the triggering event of measurement data received, the measurement report policy is referred to at the PDPs and determines whether it is the right time to update the network information database with the latest measurement report. The policy decision can be based on a control timer for periodic reports or registered criteria such as threshold violation for certain metrics. For example, when the call blocking rate is larger than a predefined threshold, it indicates that the system has more blocked calls than expected with a traffic load increase. In this case an updated measurement report may help the current network PDP decide whether to transfer some calls by vertical handoff to the other network if it is less congested.

Admission Configuration Update Policy

The admission parameters used in the new call assignment policy and call admission policy depend on the offered traffic loads to the networks. When it is observed from the latest

Parameter	Value
Cell bandwidth (Mb/s)	2.0
WLAN bandwidth (Mb/s)	11.0
Ratio of WLAN coverage area over that of the cellular cell	0.1
Average user residence time of cellular only area in a cell (s)	600
Average user residence time of double coverage area in a cell (s)	840
Average voice call duration (s)	140
Average data file size (kb)	512
Requirement of call blocking/dropping probability	0.01
Requirement of data transmission rate (kb/s)	128

■ Table 1. *System parameters for performance evaluation.*



■ Figure 5. *The data call throughputs of the three schemes in comparison.*

measurement report that changes of the offered traffic loads have exceeded an allowable range and/or have even caused QoS violation, it is necessary to properly adjust the admission configuration according to the traffic load changes. As shown in Fig. 3c, the admission configuration update policy is retrieved upon the triggering event of a measurement report. The decision results of the policy are directed to the network PEP, which updates the policy repository accordingly. To keep consistency, the updated configuration is also sent to the MTs connected to the current network.

Performance Evaluation

To validate the effectiveness of the load balancing scheme, in the following we evaluate and compare its performance with two other reference schemes for resource management. The first reference scheme uses random network selection, in which an incoming service request in the overlaying area is directed to the cellular network and WLAN with equal probabilities. No vertical handoff from the cell to the overlaying WLAN is considered, since this handoff is not necessary to maintain a connection. The second reference scheme uses a network selection mechanism similar to the new call assignment policy described above. The assignment probabilities are adapted to the traffic loads of different services. Vertical handoff between the cell and WLAN is taken into account, but only limited to the calls at the boundary of the WLAN coverage. The performance of the two reference schemes is evaluated by the analytical approach proposed in [12]. On the other hand, for the load balancing policies presented earlier, vertical handoff is also performed when a user moves within the WLAN coverage so as to relieve temporary congestion and maximize utilization. Computer simulation is used to evaluate its performance due to difficulties in analytical modeling. The system parameters given in Table 1 are used in the following performance evaluation.

Figures 4 and 5 show the voice call blocking/dropping probabilities and data call throughputs of the three schemes in comparison, respectively. It can be seen that significant performance improvement is achieved with the proposed load balancing scheme. The voice call blocking/dropping probabilities of the three schemes are well below 1 percent, which is the predefined QoS requirement. The voice call blocking/dropping probability of the proposed load balancing scheme is very close to that of the second reference scheme, but much lower than that of the first reference scheme with the random network selection. In terms of data call throughput, the second reference scheme has a performance gain of more than 50 percent over the first reference scheme. The achievable data call throughput of the proposed load balancing scheme is even 35 percent higher than that of the second reference scheme. The significant performance improvement comes from the fact that the proposed scheme not only balances the multiservice traffic loads over the two integrated networks by admission control, but also takes good advantage of dynamic vertical handoff to adapt to network states and traffic load variations.

Conclusions and Further Work

Cellular/WLAN interworking is an effective approach to facilitate evolution toward 4G wireless networks. With an all-IP-based heterogeneous platform, the pooled resources in the cellular/WLAN integrated network should be utilized efficiently for QoS provisioning. In this article we have discussed the load balancing problem for cellular/WLAN integrated networks, which takes good advantage of their unique characteristics such as the two-tier overlaying structure and complementary QoS support for multiple services. A policy-based resource management framework is presented for the loosely coupled cellular/WLAN integrated network. High utilization is achievable by dynamically balancing the offered traffic loads over the two networks via admission control and vertical handoff. Significant performance improvement is observed in comparison with two other reference schemes. For further work, we will develop an effective analytical model for the cellular/WLAN load balancing problem and investigate how to determine the policy parameters so as to maximize the system performance.

References

[1] Q. Song and A. Jamalipour, "Network Selection in an Integrated Wireless LAN and UMTS Environment using Mathematical Modeling and Computing Techniques," *IEEE Wireless Commun.* vol. 12, no. 3, June 2005, pp. 42–48.

[2] T. E. Klein and S.-J. Han, "Assignment Strategies for Mobile Data Users in Hierarchical Overlay Networks: Performance of Optimal and Adaptive Strategies," *IEEE JSAC*, vol. 22, no. 5, June 2004, pp. 849–61.

[3] F. Yu and V. Krishnamurthy, "Efficient Radio Resource Management in Integrated WLAN/CDMA Mobile Networks," *Springer Telecommun. Sys.*, vol. 30, no. 1–3, Nov. 2005, pp. 177–92.
[4] S. Lincke-Salecker, "Load Shared Integrated Networks," *Proc. 5th Euro. Personal Mobile Commun. Conf.*, Apr. 2003, pp. 225–29.
[5] W. Song *et al.*, "Improving Voice and Data Service Provisioning in Cellular/WLAN Integrated Networks by Admission Control," *Proc. IEEE GLOBECOM '06*, Nov.–Dec. 2006.
[6] K. Sabnani, "Converged Networks of the Future," *NSF Wireless/Mobile Planning Group Wksp.*, Aug. 2005.
[7] J.-C. Chen and H.-W. Lin, "A Gateway Approach to Mobility Integration of GPRS and Wireless LANs," *IEEE Wireless Commun.*, vol. 12, no. 2, Apr. 2005, pp. 86–95.
[8] R. Yavatkar, D. Pendarakis, and R. Guerin, "A Framework for Policy-Based Admission Control," IETF RFC 2753, Jan. 2000.
[9] W. Zhuang *et al.*, "Policy-based QoS Architecture in the IP Multimedia Subsystem of UMTS," *IEEE Network*, vol. 17, no. 3, May–June 2003, pp. 51–57.
[10] K. Murray, R. Mathur, and D. Pesch, "Intelligent Access and Mobility Management in Heterogeneous Wireless Networks Using Policy," *Proc. ACM 1st Int'l. Symp. Info. and Commun. Tech.*, 2003, pp. 181–86.
[11] W. Song *et al.*, "Resource Management for QoS Support in Cellular/WLAN Interworking," *IEEE Network*, vol. 19, no. 5, Sept.–Oct. 2005, pp. 12–18.
[12] W. Song, H. Jiang, and W. Zhuang, "Performance analysis of the WLAN-First Scheme in Cellular/WLAN interworking," to appear, *IEEE Trans. Wireless Commun.*
[13] H. Zhai, X. Chen, and Y. Fang, "How Well Can the IEEE 802.11 Wireless LAN Support Quality of Service" *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, Nov. 2005, pp. 3084–94.

Biographies

WEI SONG (wsong@bbcr.uwaterloo.ca) received a B.S. degree in electrical engineering from Hebei University, China, in 1998 and an M.S. degree in computer science from Beijing University of Posts and Telecommunications, China, in 2001. She is currently working toward a Ph.D. degree at the Department of Electrical and Computer Engineering, University of Waterloo, Canada. Her current research interests include resource allocation and QoS provisioning for the integrated cellular networks and WLANs.

WEIHUA ZHUANG [M'93, SM'01] (wzhuang@bbcr.uwaterloo.ca) received a Ph.D. degree from the University of New Brunswick, Canada, in electrical engineering. Since October 1993 she has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, where she is a professor. She is a co-author of the textbook *Wireless Communications and Networking* (Prentice Hall, 2003). Her current research interests include multimedia wireless communications, wireless networks, and radio positioning. She received the Outstanding Performance Award in 2005 and 2006 from the University of Waterloo and the Premier's Research Excellence Award (PREA) in 2001 from the Ontario Government for demonstrated excellence of scientific and academic contributions. She is the Editor of *IEEE Transactions on Vehicular Technology* and an Editor of *IEEE Transactions on Wireless Communications.*

YU CHENG [S'01, M'04] (cheng@iit.edu) received B.E. and M.E. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1995 and 1998, respectively, and a Ph.D. degree in electrical and computer engineering from the University of Waterloo, Canada, in 2003. From September 2004 to July 2006 he was a postdoctoral fellow in the Department of Electrical and Computer Engineering, University of Toronto, Canada. Since August 2006 he has been an assistant professor in the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago. His research interests include application-oriented networking, Internet performance analysis and resource allocation, wireless networks, and wireless/wireline interworking. He received a Postdoctoral Fellowship Award from the Natural Sciences and Engineering Research Council (NSERC) of Canada in 2004.