Load Balancing for Cellular/WLAN Integrated Networks

Wei Song [†], Weihua Zhuang [†], and Yu Cheng [‡]

[†]University of Waterloo, Canada, [‡]University of Toronto, Canada

Abstract – The interworking between the third-generation (3G) cellular network and wireless local area networks (WLANs) is one of the promising approaches to next-generation wireless networks, which can exploit the complementary advantages of the cellular network and the WLAN. Resource management for the 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 provision statistic quality of service (QoS) guarantee during the admission phase, and dynamic vertical handoff during the traffic serving phase is used to smooth the performance variance. Numerical results are presented to demonstrate that the proposed load balancing solution achieves significant performance improvement over two other reference schemes.

Index Terms – Loosely-coupled cellular/WLAN interworking, call admission control, load balancing, policy-based resource management, vertical handoff.

I. Cellular/WLAN Interworking

Motivated by the ever-increasing demand for wireless communications, the cellular network has evolved to the third generation (3G), e.g., the universal mobile telecommunication system (UMTS) specified by the third generation partner project (3GPP), which 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 system, i.e., the conversational, stream, interactive and background services. However, the expensive radio spectrum for 3G cellular networks prohibits the rapid deployment and the low bandwidth restricts the system capacity. In recent years, the 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, the WLAN can occupy a much wider spectrum than the cellular system, and provide data services under a simple medium access control (MAC) protocol. The complementary characteristics of the 3G cellular network and WLANs promote their interworking. In the near future, mobile devices can be equipped with network interfaces to both the 3G network and WLANs at a reasonable price [1]. In the cellular/WLAN integrated network, enhanced services are accessible to these dual-mode mobile devices.

The standardization for cellular/WLAN interworking is now under way in 3GPP from the cellular operator perspective. Six interworking scenarios are defined in 3GPP TR 22.934 to implement the 3GPP/WLAN interworking step by step. In the latest Release 6 of 3GPP standard, the first three interworking levels are included, which support 3GPP-based access control and enable packet-switched services available to integrated WLANs. A reference model is specified in 3GPP TS 23.234 for 3GPP/WLAN interworking. Many details in the standard deal with various security aspects. Less efforts have been devoted to QoS provisioning except for 3GPP TR 23.836, which was drafted in November 2005 and is limited to very high-level discussion. Also, in the literature, many researchers investigate vertical handoff between the cellular network and WLAN. For example, many studies investigate how to apply cross-layer techniques such as link-layer triggering and soft handoff to reduce handoff latency and packet loss during vertical handoffs, and how to make intelligent vertical handoff decisions in selecting a best handoff target network from multiple heterogeneous candidate networks [2, 3]. Although the cross-layer design and a proper network selection are constructive to improving user satisfaction and system utilization, further research is essential to effectively manage the resources pooled together in the cellular/WLAN interworking.

With the cellular/WLAN interworking, the resources of the two networks can be viewed as a resource sharing pool. The underlying heterogeneity poses many new challenges for the resource management. For example, the complementary mobility and QoS support capabilities of the two networks should be taken into account in the resource allocation. Moreover, the ownership of WLANs further complicates the interworking scenarios. The WLANs can be deployed by cellular operators, professional wireless Internet service providers (WISPs), specific organizations such as business corporations, real property management companies and airport authorities, etc. The specific interworking architecture may change with the service level agreement (SLA) between the interworked networks. The interworking architecture in turn affects the management of the pooled resources in the integrated network. There have been some initial works on this problem in the literature such as [4]. In this study, our objective is to develop a more comprehensive solution to the resource management problem and consider more implementation-specific issues, so that high-quality services are provided and the overall utilization is maximized. First, the traffic loads of multiple services should properly distributed to the cellular network and WLAN by admission control. Further, the traffic loads should be dynamically transferred between the two networks via vertical handoff when a user moves around within the overlaying area served by both the cellular network and WLAN. In this article, we investigate this load balancing problem with a policy-based resource management framework and aim at maximizing the interworking effectiveness and optimizing the QoS provisioning.

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

II. System Model

The cellular/WLAN interworking can take a good advantage of their complementary characteristics and provide enhanced services to mobile users. According to the inter-dependence of the two networks, the integration can have a relatively tight or loose coupling architecture. The cellular/WLAN interworking can be very tight with an integration at the access network (AN) level or less tight with the 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 the inefficiency, e.g., by adopting cross-layer techniques and context management. Loosely-coupled cellular/WLAN interworking has drawn much attention from the research community.

On the other hand, tight integration induces higher implementation complexity since the WLAN needs to provide a compatible interface to the cellular core network and even to the cellular radio access network. Cellular-like operations are followed in the integrated network, which enables efficient management. Joint resource allocation is possible with the tight coupling so as to maximize the overall system capacity [4]. It is likely that the tight coupling architecture will be the focus of further research on cellular/WLAN interworking due to its potential of enhanced performance [5].

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 till 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. The 3G cellular network is evolving toward the *all-IP* network [7]. On the other hand, as 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. Although heterogeneous access networks 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. To conclude, the converged network is the evolution direction, while the interworking is an interim solution.

A. Loosely-coupled Interworking Architecture

With the converging trend of wireless networks, the interworking system model should target at future converged networks. In this study, considering the advantages of independent deployment and implementation flexibility, loose coupling is adopted to integrate the cellular network and WLANs via an external IP-based core network. Motivated by the all-IP evolution trend, the cellular core network behaves more and more like an IP backbone by introducing more IP-based technologies such as the Mobile IP and AAA framework. As such, the resource management discussed in this article can be smoothly extended to tightly-coupled cellular/WLAN networks. The extension may be even easier for 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 [8], 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 general GPRS (general packet radio service) support node (GGSN) and PDG should be equipped with home agent (HA) and foreign agent (FA) functions.

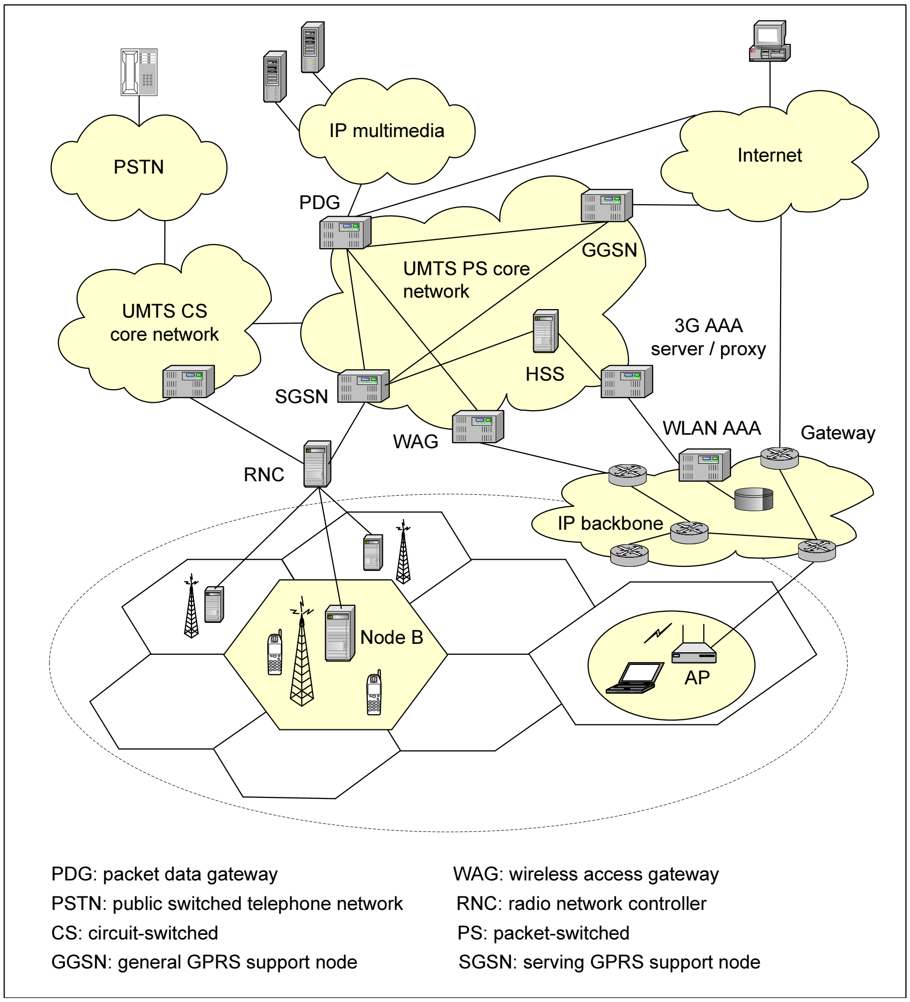


Fig. 1. The loosely-coupled UMTS/WLAN interworking architecture.

B. Policy-based Resource Management

Policy-based approach is increasingly popular for network management with the merits of flexibility and scalability. The service QoS requirements and the details of network configuration are decoupled to facilitate dynamic control. The Internet engineering task force (IETF) has defined a policy-based management framework [9]. When a policy event is triggered, an entity named policy decision point (PDP) retrieves corresponding policy rules from the *policy repository*. These policy rules are transformed into configuration actions, which are sent to the entity named policy execution point (PEP). The PEP executes these actions to respond to the triggered policy event. The IETF policy framework has been adapted and introduced in the QoS architecture of UMTS systems [10]. The adaptive and decoupling properties of the policy framework can properly address the challenges posed by the inherent heterogeneity of cellular/WLAN integrated networks. Works on policy-based QoS and handoff management for heterogeneous wireless networks have been reported in the literature, such as [11, 12].

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 two-level policy architecture for call admission control and handoff management in [12]. The policy components of PDP, PEP and policy repository are deployed at the mobile terminal (MT), the cellular network, and the WLAN. They are logical entities implemented at the mobile terminal, the RNC of the UMTS network, and the access point (AP) of the WLAN in Fig. 1, respectively. Consistency among their corresponding policy repositories needs to be maintained via SLA negotiation. 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 the 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.

III. 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 hot-spot areas. In the area with WLAN coverage (referred to as *double-coverage area* in the following), 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 area*). With this two-tier overlaying structure, a new call in the double-coverage area can be admitted either to the cell or to the WLAN. Moreover, on-going calls can be dynamically transferred between the cell and the WLAN by vertical handoffs, which may not be necessary to maintain a call, but is

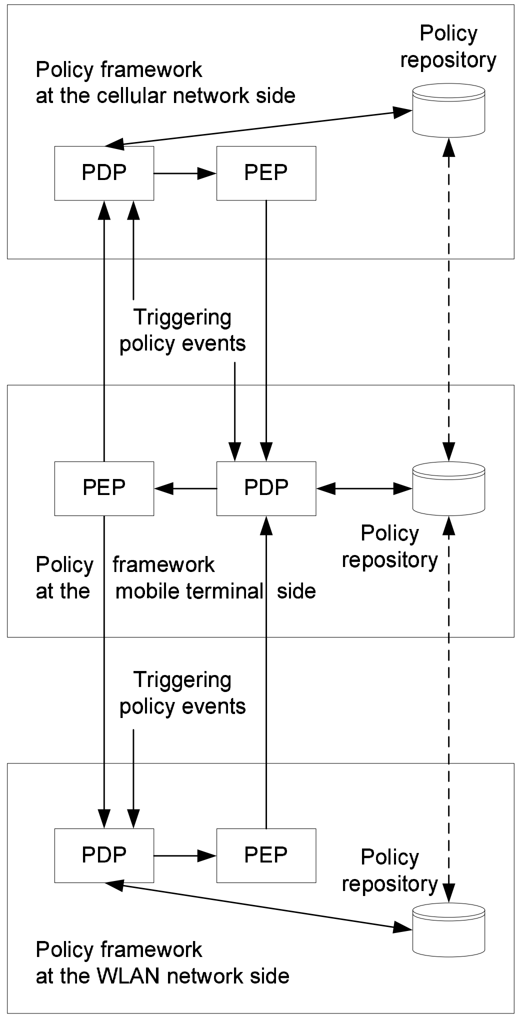


Fig. 2. The policy-based resource management framework for the cellular/WLAN integrated network.

mainly for load balancing and/or QoS improvement. The heterogeneous underlying QoS support of the integrated network can be exploited to maximize the overall resource utilization. In the cellular network, the base station controls access to the shared radio spectrum 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, e.g., the distributed coordination function (DCF) of IEEE 802.11. Because of inevitable collisions and backoffs, this type of MAC is difficult 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.

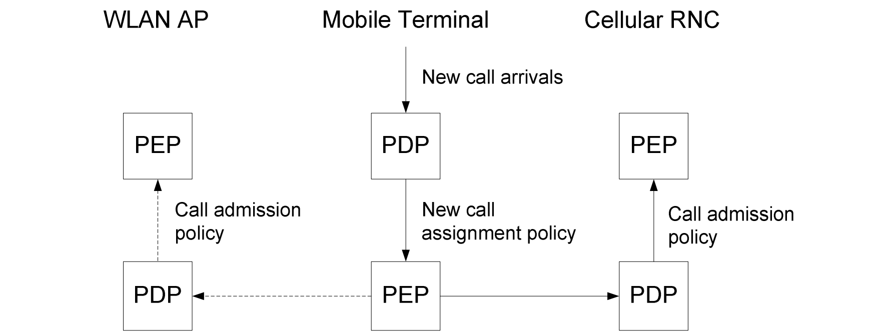
As discussed earlier, the overlaying structure of the cellular/WLAN integrated network results in the network selection problem, which is challenging due to the heterogeneity of the underlying networks in QoS provisioning, mobility support, service cost, etc. Various factors of the candidate networks need to be taken into account such as network characteristics, service type, user mobility, network condition, user preference and service cost [13]. Many selection algorithms have been proposed to select a best access network for an incoming service request [2, 3]. The specific decision criteria depend on predefined objectives such as providing the best performance to the incoming traffic in an efficient and cost-effective manner. Some advanced techniques such as fuzzy logic [2], analytic hierarchy process (AHP) and grey relational analysis (GRA) [3] are employed to make a decision. The operational complexity of the decision algorithms introduces some implementation problems and prevents their on-line application for highly dynamic network condition. To maximize the utilization of the pooled resources in the integrated system, the multi-service traffic loads should be adaptively balanced over the two networks.

To achieve a good balance between the complexity and performance, we develop the following two-phase control strategies. First, call arrivals in the overlaying area are directed to a target cell or WLAN with a probability according to the service type. Based on traffic load prediction and QoS evaluation, the assignment probabilities are determined and adapted to traffic and network dynamics. As such, the overall traffic load can be properly balanced to the two networks and the achievable performance is kept consistent across the two networks. With accurate traffic prediction and QoS evaluation, the user QoS can be guaranteed in a statistic sense. Second, to minimize the performance variance induced by network dynamics and traffic randomness, the complementary nature of the two networks can be exploited by vertical handoff. For example, with observed congestion or severe performance degradation in one network, some on-going sessions can be handed over to the other network by vertical handoff. In summary, by applying this two-phase strategy, we aim at achieving good load balancing and efficient utilization. Statistic QoS guarantee is provided by call assignment during admission phase and performance variance is reduced by dynamic vertical handoff during the traffic lifetime.

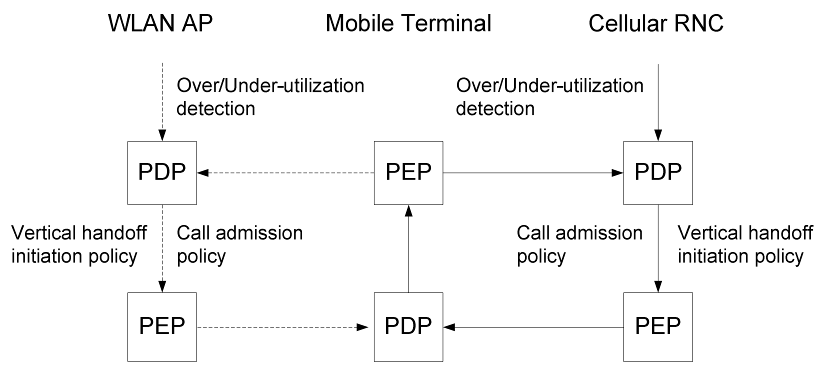
As an example, consider the load balancing problem with proper policies for voice and data services in the following. Within the policy framework, when a policy event is triggered, the PDP refers to the policy repository for the matched policy rules to make a policy decision. Then, the decision is transformed into configuration actions, which are sent to the PEP for execution. The policy events registered at the MT and networks are different but correlated. Depending on the triggering policy events, the following load balancing policies are developed. The PDP and PEP components interact and collaborate to respond to an incoming policy event as illustrated in Fig. 3.

A. *New call assignment policy*

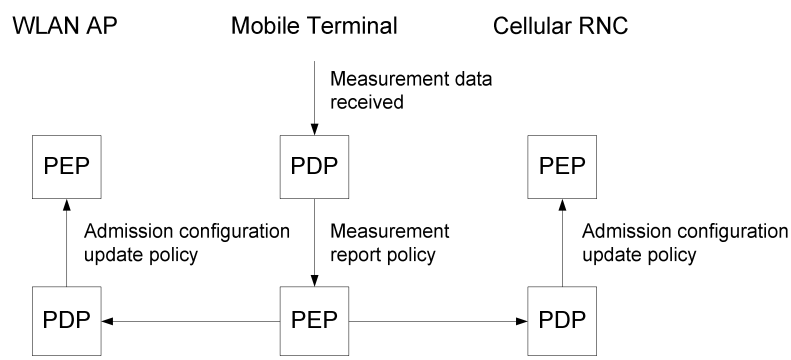
As shown in Fig. 3(a), triggered by a new call arrival at an MT, the new call assignment policy is referred to for a policy decision. Depending on the service type of the incoming call, the policy determines the target network to admit the new call. For a voice (data) call, the decision is directed to 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 the traffic load estimated from traffic measurements and need to be dynamically adapted to load changes. 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 [14], we proposed 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 corresponding selected network, which triggers the event involving the call admission policy at the network PDP.



(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.

Fig. 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.

B. 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 [15] and WLAN [16]. The admission region is usually defined by a set of vectors (N_v, N_d), which denote respectively 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. The configuration of admission region 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 [14].

C. Vertical handoff initiation policy

As shown in [16], the busyness ratio is an effective indicator for the channel utilization of the WLAN. The maximum utilization is found to exist in the unsaturated case instead of the saturated case. Consequently, it is reasonable to dynamically balance the traffic load from and to the cellular network by vertical handoff, so as to make sure that the WLAN operates in its most efficient region. With the policy framework in Fig. 2, this load balancing by vertical handoff can work as shown in Fig. 3(b). First, over an observation window, if the busyness ratio is detected at the AP to be above a predefined threshold ε_w^u, the policy event involving the vertical handoff initiation policy is triggered at the PDP of the WLAN. By prioritizing the on-going calls carried by the WLAN, certain calls are selected and handed out of the WLAN in order. The on-going calls are prioritized based on their service type, remaining service time, subscription profile, etc. The voice calls from cellular subscribers are selected first for vertical handoff, because the cellular network is good at supporting voice service and has a priority to serve the calls from its subscribers. 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 will 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, some basic information about the WLAN network condition is available to the base station of the cellular network. With an observation that the busyness ratio of the WLAN is below a threshold ε_w^l for a continuous period of time, the policy event is triggered and the vertical handoff initiation policy is invoked at the cellular network. Data calls with a longer remaining service time and from mobile users with the WLAN as their home network are 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 under-utilized, so that the cellular network is prevented in advance from becoming the system bottleneck. Similar to the first case, the vertical handoff decision made at the WLAN PDP is transformed into signaling messages sent to the MT by the WLAN PEP. Finally, the PDP of the MT directs its PEP to request admission to the WLAN for the potential vertical handoff calls.

D. Measurement report policy

To perform the 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 gain of joint management over independent control is well exploited [4]. 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. To address this problem, 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 attached with the WLAN for data transmission [8]. As a result, the MT can keep receiving periodic advertisement messages from the cellular network indicating its network condition such as link performance, network load, data throughput, call blocking/dropping rates, etc. 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. 3(c), with the triggering event of measurement data received, the measurement report policy is referred to determine whether it is the right time to update the network information data base at the PDPs with the latest measurement report. The policy decision can be based on a control timer for periodic report 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 the system has more calls blocked than expected with an increased traffic load. 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.

E. 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 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 changes. As shown in Fig. 3(c), 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 attached to the current network.

IV. Performance Evaluation

To validate the effectiveness of the load balancing scheme, we evaluate the performance of the resource management policies in the following. The performance is compared with that of two reference schemes. The first one applies a 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 in Section III-A. 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 boundary of the WLAN coverage. The performance of the two reference schemes is evaluated by the analytical approach proposed in [14]. On the other hand, for the load balancing policies given in Section III, 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 used in the following performance evaluation are given in Table I.

TABLE I

SYSTEM PARAMETERS FOR PERFORMANCE EVALUATION.

Parameter	**Value**
Cell bandwidth (Mb/s)	2
WLAN bandwidth (Mb/s)	11
Ratio of WLAN coverage area over that of the cellular cell	0.1
Average residence time of cellular-only area in a cell (s)	600
Average residence time of double-coverage area in a cell (s)	840
Average voice call duration (s)	140
Average data file size (bytes)	64 K
Requirement of call blocking/dropping probability	0.01
Requirement of data transmission rate (Kb/s)	128

Fig. 4 and Fig. 5 show the voice call blocking/dropping probability and data call throughput of the three schemes in comparison, respectively. It can be seen that significant performance improvement is achieved with the proposed load balancing scheme, which takes a good advantage of vertical handoff to address traffic and network dynamics. The voice call blocking/dropping probability of the three schemes is well below 1%, 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 random network

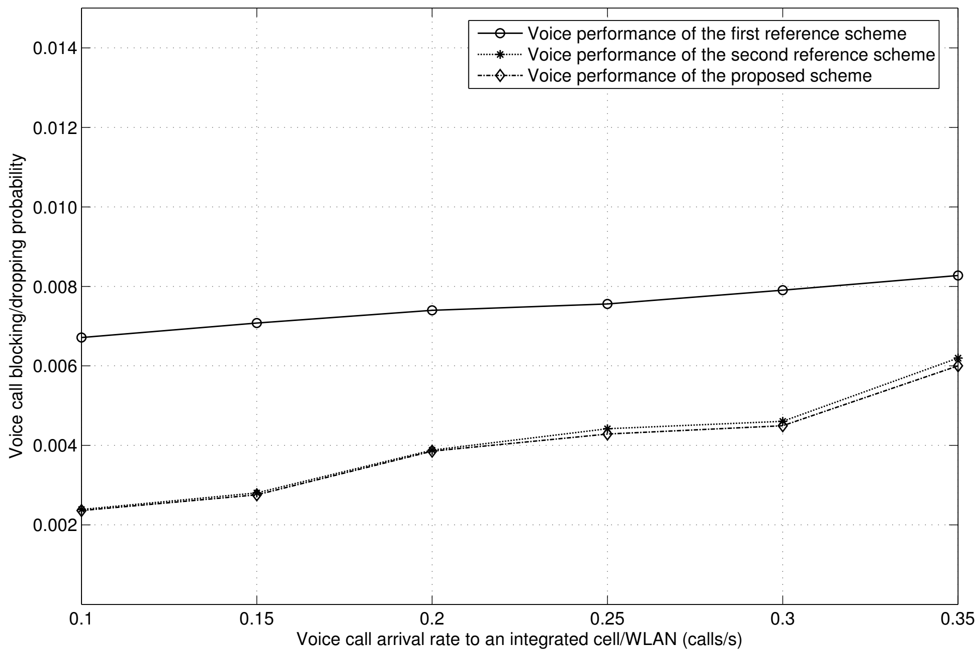


Fig. 4. The voice call blocking/dropping probability of the three schemes in comparison.

selection. In terms of data call throughput, the second reference scheme has a performance gain of more than 50% over the first reference scheme. The achievable data call throughput of the proposed load balancing scheme is even 35% higher than that of the second reference scheme. The significant performance improvement comes from the fact that the proposed scheme not only applies admission control adapted to multi-service traffic loads, but also balances network and traffic dynamics by vertical handoff.

V. Conclusion

To achieve effective interworking between the cellular network and WLANs, the pooled resources of the integrated network should be utilized efficiently for QoS provisioning. In this article, we discuss the load balancing problem for cellular/WLAN integrated networks, which takes a 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

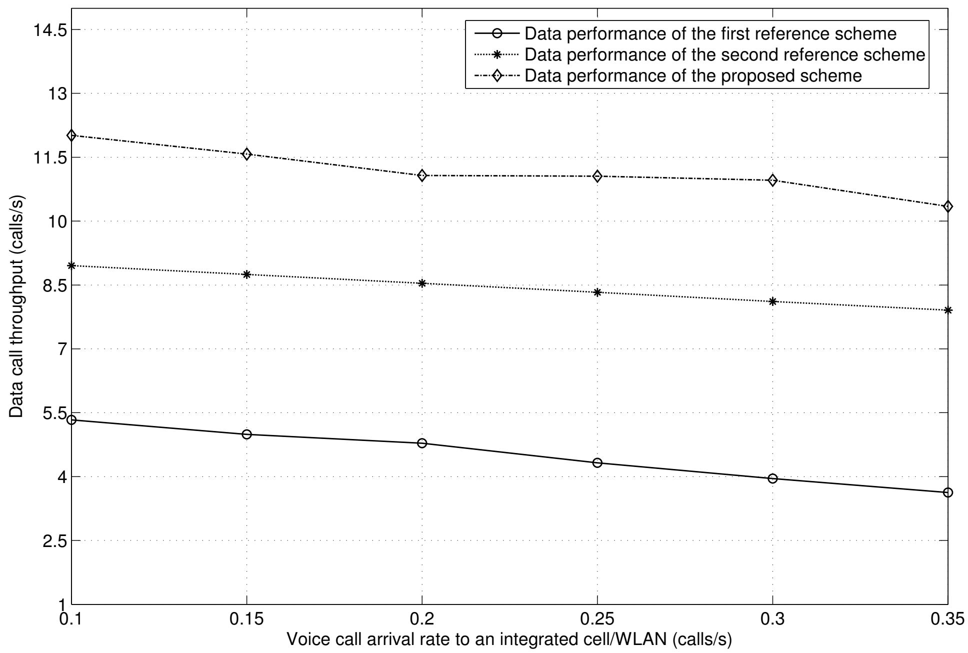


Fig. 5. The data call throughput of the three schemes in comparison.

the loosely-coupled cellular/WLAN integrated network. Various load balancing policies are designed with the two-level policy framework at both the mobile terminals and networks. High utilization is achievable by dynamically balancing the offered traffic load to 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 properly determine and adaptively adjust the policy parameters so as to maximize the performance.

REFERENCES

[1] D. I. Axiotis, T. Al-Gizawi, E. N. Protonotarios, F. I. Lazarakis, C. Papadias, and P. I. Philippopoulos, "Services in interworking 3G and WLAN environments," *IEEE Wireless Commun. Mag.*, vol. 11, no. 5, pp. 14–20, Oct. 2004.

[2] W. Zhang, "Handover decision using fuzzy MADM in heterogeneous networks," in *IEEE WCNC*, vol. 2, Mar. 2004, pp. 653–658.

[3] Q. Song and A. Jamalipour, "Network selection in an integrated wireless LAN and UMTS environment using mathematical modeling and computing techniques," *IEEE Wireless Commun. Mag.*, vol. 2, no. 3, pp. 42–48, June 2005.

[4] J. Luo, R. Mukerjee, M. Dillinger, E. Mohyeldin, and E. Schulz, "Investigation of radio resource scheduling in WLANs coupled with 3G cellular network," *IEEE Commun. Mag.*, vol. 41, no. 6, pp. 108–115, June 2003.

[5] G. Lampropoulos, N. Passas, L. Merakos, and A. Kaloxylos, "Handover management architectures in integrated WLAN/cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 7, no. 4, Fourth Quarter 2005, http://www.comsoc.org/livepubs/surveys.

[6] K. Sabnani, "Converged networks of the future," in *NSF Wireless/Mobile Planning Group Workshop*, Aug. 2005.

[7] 3GPP2, "cdma2000 wireless IP network standard," P.S0001-B v2.0, Sep. 2004.

[8] J.-C. Chen and H.-W. Lin, "A gateway approach to mobility integration of GPRS and wireless LANs," *IEEE Wireless Commun. Mag.*, vol. 12, no. 2, pp. 86–95, Apr. 2005.

[9] R. Yavatkar, D. Pendarakis, and R. Guerin, "A framework for policy-based admission control," IETF RFC 2753, Jan. 2000.

[10] W. Zhuang, Y. S. Gan, K. J. Loh, and K. C. Chua, "Policy-based QoS architecture in the IP multimedia subsystem of UMTS," *IEEE Network*, vol. 17, no. 3, pp. 51–57, May 2005.

[11] W. Zhuang, Y.-S. Gan, K.-J. Loh, and K.-C. Chua, "Policy-based QoS-management architecture in an integrated UMTS and WLAN environment," *IEEE Commun. Mag.*, vol. 41, no. 11, pp. 118–125, Nov. 2003.

[12] K. Murray, R. Mathur, and D. Pesch, "Intelligent access and mobility management in heterogeneous wireless networks using policy," in *ACM 1st Int'l. Wksp. Info. and Commun. Tech.*, 2003, pp. 181–186.

[13] W. Song, H. Jiang, W. Zhuang, and X. Shen, "Resource management for QoS support in cellular/WLAN interworking," *IEEE Network*, vol. 19, no. 5, pp. 12–18, Sep. 2005.

[14] W. Song, H. Jiang, and W. Zhuang, "Performance analysis of the WLAN-first scheme in cellular/WLAN interworking," *IEEE Trans. Wireless Commun.*, to appear.

[15] C. W. Leong, W. Zhuang, Y. Cheng, and L. Wang, "Call admission control for integrated on/off voice and best-effort data services in mobile cellular communications," *IEEE Trans. Commun.*, vol. 52, no. 5, pp. 778–790, May 2004.

[16] H. Zhai, X. Chen, and Y. Fang, "A call admission and rate control scheme for multimedia support over IEEE 802.11 wireless LANs," in *1st Conf. Quality of Service in Heterogeneous Wired/Wireless Networks*, Oct. 2004, pp. 76–83.