附件一

行政院國家科學委員會補助專題研究計畫 □ **成果報告** ☑ **期中進度報告**

具高效能及服務品質之合作式網路設計

子計畫二：合作式網路之頻寬管理機制

計畫類別：□ 個別型計畫　　☑ 整合型計畫
計畫編號：NSC 98－2221－E－009－059－MY2
執行期間：　98 年　8 月　1 日至　100 年　7 月　31 日

計畫主持人：林一平（國立交通大學 資訊工程學系(所）教授）
共同主持人：楊舜仁（國立清華大學 資訊工程學系(所）助理教授）
計畫參與人員：邱冠龍、高健淇、李易達、林沂珊、王佳韞、施宗欽、莊雅茹、吳家瑋、邱靖捷、謝萬瀚

成果報告類型(依經費核定清單規定繳交)：☑精簡報告　□完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
☑出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年☑二年後可公開查詢

執行單位：國立交通大學、國立清華大學

中　華　民　國　　99　年　　5　月　　30　日

（一） 計畫中文摘要。（五百字以內）

隨著使用者對於高速無線網際網路的需求與期待日益增加，未來的行動通訊系統被期待能提供更高的資料傳輸率、更廣的訊號涵蓋範圍及更可靠的移動性。為了滿足無線網際網路高頻寬的需求，基於基礎建設網路（Infrastructure-based network）之多躍中繼（Multihop Relay）技術於近年內被提出，並迅速成為被廣泛討論的研究議題。多躍中繼網路（Multihop Relay Network）是一種合作式通訊網路（Cooperative Network），其主要概念是增設中繼站（Relay Station，RS）作為 BS 與 MS 之間的轉傳（relay）媒介。只要將 RS 佈放在適當的位置，訊號將可避開不理想的傳播路徑並減少訊號衰減。由於在轉傳訊號時可再度提高其傳輸功率及使用直視傳輸（Line-Of-Sight，LOS），因此 MS 所接收到的訊號品質可以大幅改善。在以上這些優點同時作用下，網路系統最終能達到提升傳輸速率（throughput enhancement）、增加系統容量（capacity enhancement）、延伸服務範圍（coverage extension）、解決遮蔽衰落效應影響等目的。此外，由於 RS 與 BS 之間採用無線傳輸，因此可以節省後端固接網路（wire-line backhaul）的建置成本。

本計畫將針對 IEEE 802.16j Multihop Relay（MR）WiMAX 系統，深入研究多躍中繼網路之 RS 分群演算法及空間再利用無線電頻寬排程演算法。利用多中繼網路技術，IEEE 802.16j 標準已被制定用以提升現有 IEEE 802.16e 網路的系統效能。然而，多躍中繼技術同時也會造成頻繁換手的問題，進而降低 IEEE 802.16j 的系統效能。RS 分群機制是用以解決此類問題的一項技術。RS 分群的概念是將相鄰的 RS 聚集成為一個 RS 群組(邏輯上可視為一個覆蓋大範圍的 RS)。在這個計畫中，我們以系統流量(throughput)和換手頻率(handoff rate)做為研究 RS 分群制制效能的依據。我們提出一個貪婪 RS 分群演算法用以最小化換手頻率。在此演算法中，我們以相鄰 RS 間的換手頻率為依據，換手頻率較高的 RS 配對即較優先被選為同一群組。我們的網路模擬結果顯示，我們提出的 RS 分群演算法可大量地減少換手頻率。此外，我們提出兩種集中式空間再利用排程演算法。第一種為流量優先策略，用以最大化系統總流量。第二種為延遲優先策略，用以最小化封包延遲時間。模擬結果指出延遲優先策略不僅能最小化平均封包延遲時間，同時也能提供良好的公平性給不同流量負載的使用者。

關鍵詞：多躍中繼網路、合作式通訊網路、RS、IEEE 802.16j 網路、RS 分群演算法、空間再利用

（二） 計畫英文摘要。（五百字以內）

As the user demand for high-speed wireless communication service increases gradually, the future wireless mobile systems are expected to provide higher data transmission rate, wider signal coverage and more reliable mobility. In order to meet the high-bandwidth requirement, infrastructure-based multihop relay technologies have been proposed in the recent years. A multihop relay network is a type of cooperative networks. The main concept of multihop relay networks is to establish relay stations (RSs) as relay media between base stations (BSs) and mobile stations (MSs). If the RSs are deployed in proper positions, the situation of choosing unsatisfactory signal routes can be avoided and the signal attenuation can be reduced. In addition, while relaying signals the transmission power can be increased and the line-of-sight transmission may be used. In this case, the signal quality which MSs receive can be improved significantly. From the above advantages, multihop relay networks can achieve the goals of throughput enhancement, capacity enhancement and coverage extension. Moreover, based on the wireless transmission between RSs and BSs, the installation cost of the wire-line backhaul can be completely eliminated as well.

The IEEE 802.16j standard has been developed to provide performance enhancement to the existing IEEE 802.16e network by incorporating the multihop relay (MR) technology. However, frequent handoffs and low spectrum-utilization issues that were not encountered in IEEE 802.16e may be incurred in IEEE 802.16j. The relay station (RS) grouping is one optional mechanism in the IEEE 802.16j MR standard to overcome these problems. The concept of RS grouping is to group neighboring RSs together to form an RS group, which can be regarded as a logical RS with larger coverage. In this project, we investigate RS grouping performance enhancement in terms of throughput and handoff frequency. We design an RS grouping algorithm to minimize handoffs by utilizing a greedy grouping policy: RS pairs with higher handoff rates will have higher priority for selection. The simulation results show that the handoff frequency of the considered MR network can significantly be reduced, and suitable RS grouping patterns can be derived using our grouping algorithm. In addition, we propose two centralized scheduling policies, i.e., the throughput-first (TF) policy to maximize the system throughput and the delay-first (DF) policy to minimize the average packet delay. By integrating our RS grouping algorithm and centralized scheduling algorithms, the simulation results indicate that, for the case of fixed users, groupings with smaller group sizes can result in better throughput performance. However, when user mobility is considered, the throughput value increases as the group size increases. Furthermore, we also show that the DF policy can both minimize the average packet delay and provide the fairness property among users with different traffic loads.

Keywords: Grouping algorithm, IEEE 802.16j, multihop relay (MR), scheduling policy, WiMAX.

（三）研究計畫之背景及目的

本研究計畫之背景

Incorporating the *multihop relay* (MR) technology [25], the IEEE 802.16j MR standard [14] has been developed to provide throughput improvement, coverage extension, and capacity enhancement to the existing IEEE 802.16e protocol [1]. By deploying relay stations (RSs), the end-to-end communication quality between base stations (BSs) and mobile stations (MSs) can be improved without high infrastructure deployment costs. In addition, *spatial reuse* [19] is another promising approach that can be employed in IEEE 802.16j MR networks to improve spectral efficiency. Based on the centralized scheduling, spatial diversity gain can be achieved if multiple simultaneous transmissions using the same bandwidth resources are realized within the same BS cell.Although IEEE 802.16j has the potential to provide substantial performance enhancements, several issues that were not addressed in IEEE 802.16e may be encountered. For example, frequent handoffs may occur during the movements of MSs since the RS cells are smaller than the BS cells. To avoid the consequent performance degradation, *RS grouping* has been identified as an optional mechanism in the IEEE 802.16j standard. The main idea of RS grouping is to put adjacent RSs together to form an *RS group*, wherein the RS members are required to simultaneously receive and transmit the same data. From the MSs' viewpoint, the RS group can be regarded as a logical RS with larger coverage. Therefore, the handoff frequency can be reduced since no handoff procedure would be triggered, even when an RS crossing event within the RS group occurs.

本研究計畫之目的

To the best of our knowledge, no RS grouping strategy has been proposed or discussed in the literature. We argue that different grouping criteria may lead to various performance results. Specifically, utilizing a smaller RS group size is advantageous to spatial reuse, because more RS groups can perform simultaneous transmissions at the same time and frequency. Thus, improved average system throughput can be expected. However, a smaller RS group size may also lead to higher packet loss rate due to more frequent handoffs between RS groups. In conclusion, when implementing the RS grouping mechanism in IEEE 802.16j MR networks, the performance tradeoff between throughput and handoff frequency should seriously be considered. In this project, we analyze grouping strategies for RS grouping-enabled IEEE 802.16j MR networks and propose an efficient RS grouping algorithm to minimize the handoff frequency. As we have pointed out, RS grouping strategies will also influence the throughput performance. To investigate the impacts of RS grouping on the IEEE 802.16j system throughput, we design two centralized downlink (DL) scheduling policies for RS-grouping-enabled IEEE 802.16j MR networks. One of these two scheduling policies aims to maximize the system throughput, and the other is to minimize the average packet delay.

（四）研究計畫之方法與成果

(1) 設計最少換手次數之 RS 分群演算法

我們針對 IEEE 802.16j 標準中的 RS 群組化機制，透過分析 RS 分群的一些特性，設計出理想的 RS 分群演算法。我們發現只要控制合理的群組大小，且能找到群組內各 RS 間之相交邊數最多的群組排列，便能使 IEEE 802.16j 網路的 MS 換手次數有效降低，且同時達到有效無線頻寬分配及傳輸速率提高等優點，這些特性都是我們設計 RS 分群演算法時的重要考量因素，我們稱之為最少換手次數之 RS 分群演算法(Handover-Minimizing RS Grouping Algorithm)。

(2) 設計使用空間再利用的集中式下傳頻寬排程演算法

為了在 IEEE 802.16j 網路中能更有效地實現空間再利用及 RS 分群的好處，我們設計結合 RS 分群及空間再利用特性之集中式下傳頻寬排程演算法。其中我們採用 throughput optimal policy 的排程演算法，使用夏農定理(Shannon’s formula)來模擬傳輸速率的變化，透過 selective relaying 的概念來模擬 RS 群組的合作式傳送 (cooperative transmission)，並利用 Rayleigh fading 的數學模型來模擬 RS 與 MS 之間的非直視無線傳輸(Non-Line-Of-Sight, NLOS)。我們設計公式(1)

$$D_g(t;\phi) = R_g(t;\phi) \max_{j\in\{1,\dots,K\}} \left[X_j^g(t)\right] \quad , g = 1,\dots, G \qquad (1)$$

其中 $D_g(t;\phi)$代表群組 g 在第 t 個時槽內所需要的傳輸量比重，此比重值愈大表示該群組會被挑選作為同時傳輸的機率愈高。最後，我們再利用公式(2)

$$\hat{\phi}(t) = \arg \max_{\phi\in\Omega} \left(\sum_{g\in\phi} D_g(t;\phi) \right) \qquad (2)$$

決定出在第 t 個時槽內，在所有 activation set ϕ 可能組合的總集合 Ω 中最適合被安排同時傳輸資料的 RS 群組組合 $\phi(t)$。根據此演算法所得之結果，可以作出第 t 個時槽的頻寬分配。

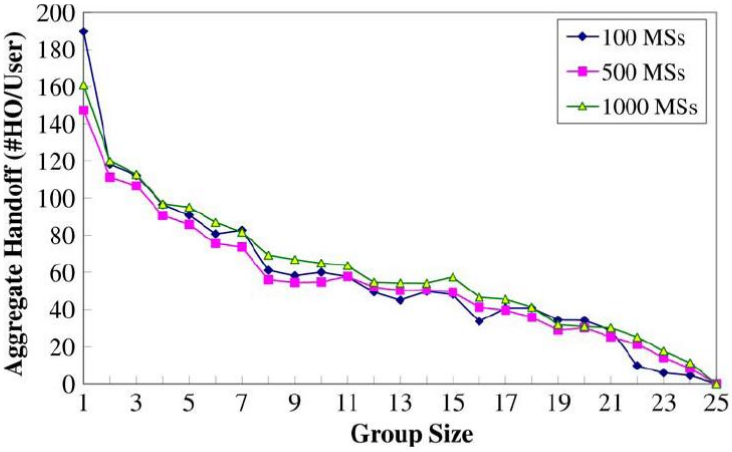
(3) 實作群組化 IEEE 802.16j 多躍中繼網路系統之模擬環境

為了要評估上述RS分群演算法及頻寬排程演算法對IEEE 802.16j多躍中繼網路的效能影響，我們實作出此 IEEE 802.16j 系統的模擬環境。透過模擬所得的結果，我們可以計算系統傳輸量(system throughput)、總換手次數(the number of handovers)、封包遺失率(packet loss rate)、封包傳送延遲(packet transmission delay)等系統效能指標。利用這些數據我們可以分析出 RS 分群演算法及排程演算法在哪些條件下會讓系統達到最佳狀態，以驗證我們所提出的演算法的確比原方法有更好的效率。

Simulation Results and Discussion

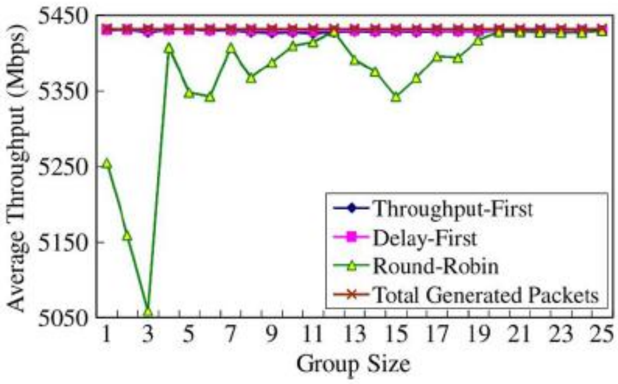
A. Effects of RS Group Sizes on Handoff Frequency

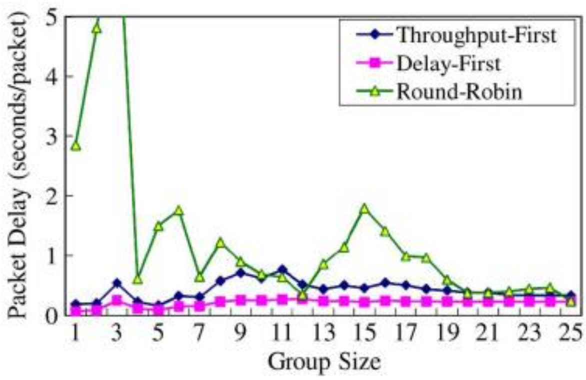
The following figure individually evaluates the handoff frequency of each grouping under different user densities (specifically, 100, 500, and 1000 users in our experiments). The results indicate that, using the greedy grouping policy of our RS grouping algorithm, the handoff frequencies of the groupings gradually decline, regardless of the user density, as the group size increases. The main reason is that larger group sizes generally lead to lower handoff probabilities. It could be concluded that it is reasonable to choose larger group sizes if higher user mobility speeds are observed.

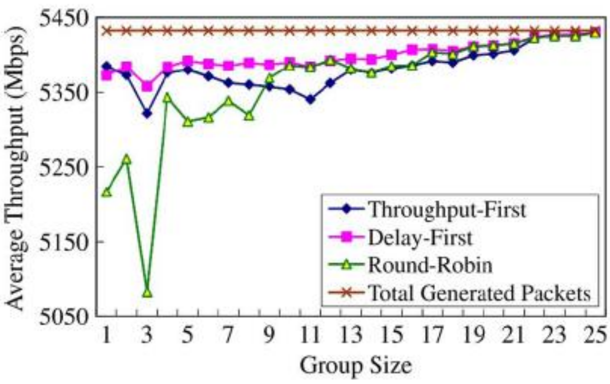


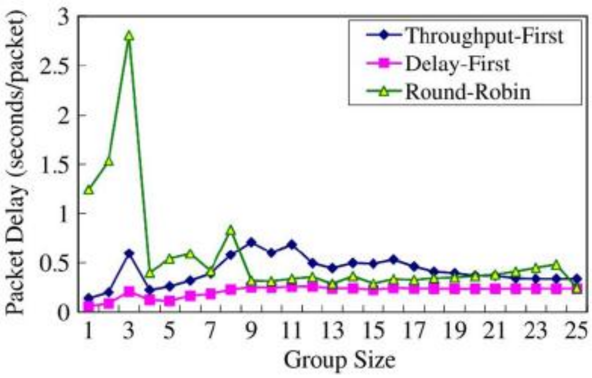
B. Performance Analysis of Different Scheduling Policies

Based on the groupings, we estimate the throughput and delay performances of the TF, DF, and Round-Robin centralized scheduling policies as functions of the group size. The following figures present the average throughput and delay of the network under the situations with fixed and mobile users, respectively. Note that the average throughput is estimated as the average amount of packets actually received by the users within the simulation time. Moreover, the packet delay indicates the time difference between when a packet arrives at the BS and when it is received by the corresponding user. The figures show the average network throughput under the three scheduling policies. For the case where users are stationary, the TF and DF policies achieve similar throughput performance. However, for the case with user mobility, the average throughput under the DF policy is higher than those under the other two scheduling policies. This phenomenon implies that the mobility behavior of the MSs is a major factor affecting the throughput performance under the TF policy, whereas the DF policy can provide relatively stable throughput, regardless of the user-mobility effect. Specifically, since the DF policy is able to balance the queue length of each user while the TF policy only prioritizes the user with the longest queue length, the average buffered packet amounts for all users under the DF policy are relatively small. Therefore, the number of dropped packets during the handoffs in DF can reasonably be reduced, and the throughput can further be improved. The Round-Robin policy inefficiently performs in both cases since its scheduling decisions are sequentially made, despite the system state information. In addition, we also observe that, compared with groupings with larger group sizes, groupings with smaller group sizes result in better throughput performance for the case of fixed users. However, when user mobility is considered, the throughput value increases as the group size increases. This is because, when users are stationary, the throughput performance is mainly dominated by the spatial diversity gain. Under such circumstances, smaller RS group sizes lead to higher throughput values. On the other hand, when the user mobility speed is high enough, the larger packet loss rates due to more frequent handoffs of smaller group sizes have more significant impact on the throughput performance. Consequently, larger RS group sizes are more advantageous. The figures also show the packet delay performance for the two scenarios without and with user mobility. We notice that groupings with larger group sizes result in higher average packet delay. This is owing to the lower spectrum reuse of larger group sizes, and the packets in the system may suffer from longer queueing delay. Moreover, as expected, the DF policy can provide the lowest average delay for all group size cases, compared with the TF and Round-Robin policies. The TF policy causes higher average delay since it considers only the queue length states but no packet waiting time information. The Round-Robin policy, which depends on neither the queue length nor the waiting time information, incurs the worst average delay performance through almost all group size cases.









（五）結論

The IEEE 802.16j MR standard has been developed to provide performance enhancement to the existing IEEE 802.16e network. However, issues such as frequent handoffs and low spectrum utilization, which were not encountered in IEEE 802.16e, may occur in IEEE 802.16j. The RS grouping is one optional mechanism in the IEEE 802.16j MR standard to overcome these problems. This project has investigated the RS grouping performance enhancement in terms of throughput and handoff frequency. An RS grouping algorithm was designed by utilizing a greedy grouping policy: RS pairs with higher handoff rates will have higher priority to be selected. The simulation results have shown that the handoff frequency of the considered MR network can significantly be reduced, and suitable RS grouping patterns with determined activation set assignments can be derived using our grouping algorithm. In addition, we have proposed the TF and DF centralized scheduling policies to maximize the system throughput and to minimize the average packet delay, respectively. By integrating our RS grouping algorithm and centralized-scheduling algorithms, the simulation results have indicated that, for the case of fixed users, groupings with smaller group sizes can result in better throughput performance. However, when user mobility is considered, the throughput value increases as the group size increases. Furthermore, we have also shown that the DF policy cannot only minimize the average packet delay but can also provide fairness among different traffic-load users.

(六)成果自評

The main contribution of this project is to propose the first integrated algorithmic framework that can be utilized to investigate the performance interaction between RS grouping and resource scheduling for IEEE 802.16j MR networks. The results of this work have been published in the international journal (IEEE Transactions on Vehicular Technology, 2010) [A1]. In addition, this project also supports us to attend the conference in Korea for presenting our another work [A2].

1.完成之工作項目

- 設計 IEEE 802.16j 最少換手次數之 RS 分群演算法
- 設計 IEEE 802.16j 使用空間再利用的集中式下傳無線電頻寬排程演算法
- 實作 IEEE 802.16j 群組化多躍中繼網路系統之模擬環境

2. 對於學術研究、國家發展及其他應用方面之貢獻

- 本計畫的研究成果已發表於國際期刊[A1]
- 提出電腦模擬模型以研究“合作式通訊網路之系統效能”
- 這個計畫的結果可以提供工業界做為參考，我們很樂意將研發的技術技轉給工業界

3.對於參與之工作人員，獲得以下訓練

- 在研究的過程中瞭解並學習 Cooperative Communication Networks 的相關知識和運作方式，及其設計的內含意義，並熟識 IEEE 802.16e、IEEE 802.16j 等相關業界標準
- 學習“Discrete-Event System Simulation”之電腦模擬模型技巧

Appendix:

[A1] Shun-Ren Yang, Chien-Chi Kao, Wai-Chi Kan, and Tzung-Chin Shih, “Handoff Minimization Through a Relay Station Grouping Algorithm With Efficient Radio-Resource Scheduling Policies for IEEE 802.16j Multihop Relay Networks,” IEEE Transactions on Vehicular Technology, Feb. 2010.

[A2] Shin-Hua Yang, Shun-Ren Yang, and Chien-Chi Kao, “Analyzing VoIP Capacity with Delay Guarantee for Integrated HSPA Networks,” International Conference on Future Generation Communication and Networking **(**FGCN), 2009.

（七）參考文獻

[1] Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, IEEE Std. 802.16e-2005, Feb. 2006.

[2] C. Castelluccia, "Extending mobile IP with adaptive individual paging: A performance analysis," in *Proc. IEEE ISCC*, 2000, pp. 113–118. 910

[3] C.-M. Weng and P.-W. Huang, "Modified group method for mobility management," *Comput. Commun.*, vol. 23, no. 2, pp. 115–122, Jan. 2000.

[4] C.-Y. Hong and A.-C. Pang, "Link scheduling with QoS guarantee for wireless relay networks," in *Proc. IEEE 28th INFOCOM*, Apr. 2009, pp. 2806–2810.

[5] D. Ghosh, A. Gupta, and P. Mohapatra, "Admission control and interference-aware scheduling in multi-hop WiMAX networks," in *Proc. IEEE MASS*, Oct. 2007, pp. 1–9.

[6] D. J. A. Welsh and M. B. Powell, "An upper bound for the chromatic number of a graph and its application to timetabling problems," *Comput. J.*, vol. 10, no. 1, pp. 85–86, 1967.

[7] D. Schultz, R. Pabst, and T. Irnich, "Multi-hop based radio network deployment for efficient broadband radio coverage," in *Proc. WPMC*, Yokosuka, Japan, 2003, pp. 377–381.

[8] G. Varsamopoulos and S. K. S. Gupta, "Dynamically adapting registration areas to user mobility and call patterns for efficient location management in PCS networks," *IEEE/ACM Trans. Netw.*, vol. 12, no. 5, pp. 837–850, Oct. 2004.

[9] G. Varsamopoulos and S. K. S. Gupta, "Optimal offline and online registration techniques for location management with overlapping registration areas," *IEEE Trans. Mobile Comput.*, vol. 4, no. 5, pp. 474–488, Sep./Oct. 2005.

[10] G. Wan and E. Lin, "A dynamic paging scheme for wireless communication systems," in *Proc. ACM/IEEE MobiCom*, 1997, pp. 195–203.

[11] H. Viswanathan and S. Mukherjee, "Performance of cellular networks with relays and centralized scheduling," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2318–2328, Sep. 2005.

[12] H. Xie, S. Tabbane, and D. J. Goodman, "Dynamic location area management and performance analysis," in *Proc. IEEE Veh. Technol. Conf.*, May 1993, pp. 536–539.

[13] Ranging Process for IEEE 802.16j, IEEE 802.16j-06/172, Nov. 2006.

[14] Part 16: Air Interface for Broadband Wireless Access Systems Amendment 1: Multiple Relay Specification, IEEE Std. 802.16j-2009, Dec. 2009.

[15] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[16] J. Nuevo, Mobility Generator for NS-22002. [Online]. Available: http://www.math.keio.ac.jp/matumoto/emt.html

[17] J. Tang, G. Xue, and C. Chandler, “Interference-aware routing and bandwidth allocation for QoS provisioning in multihop wireless networks,” *Wireless Commun. Mobile Comput. J.*, vol. 5, no. 8, pp. 933–943, Dec. 2005.

[18] L. Kleinrock, *Queueing Systems*. Hoboken, NJ: Wiley, 1975.

[19] L. Kleinrock and J. Silvester, “Spatial reuse in multihop packet radio networks,” *Proc. IEEE*, vol. 75, no. 1, pp. 156–167, Jan. 1987.

[20] L. Wang, Y. Ji, and F. Liu, “A novel centralized resource scheduling scheme in OFDMA-based two-hop relay-enhanced cellular systems,” in *Proc. IEEE WIMOB*, Oct. 2008, pp. 113–118. 959

[21] M. K. Awad and X. Shen, “OFDMA based two-hop cooperative relay network resources allocation,” in *Proc. IEEE ICC*, May 2008, pp. 4414–4418.

[22] M. Yang and P. H. J. Chong, “Time slot allocation schemes for multi964 hop TDD-CDMA cellular system,” in *Proc. IEEE WCNC*, Mar. 2007, pp. 3099–3104.

[23] P. Djukic and S. Valaee, “Link scheduling for minimum delay in spatial re-use TDMA,” in *Proc. IEEE 26th INFOCOM*, May 2007, pp. 28–36.

[24] R. Jain, W. Hawe, and D. Chiu, “A quantitative measure of fairness and discrimination for resource allocation in shared computer systems,” Digital Equipment Corp., Maynard, MA, Tech. Rep. DEC-TR-301, Sep. 1984.

[25] R. Pabst, B. H. Walke, and D. C. Schultz, “Relay-based deployment concepts for wireless and mobile broadband radio,” *IEEE Commun. Mag.*, vol. 42, no. 9, pp. 80–89, Sep. 2004.

[26] T.-W. Kim, T.-Y. Min, and C.-G. Kang, “Opportunistic packet scheduling algorithm for load balancing in a multi-hop relay-enhanced cellular OFDMA-TDD system,” in *Proc. 14th APCC*, Oct. 2008, pp. 1–5.

[27] T. Wu, G. Li, Y. Wang, J. Huang, X. Yu, and H. Tian, “Fairness-oriented scheduling with equilibrium for multihop relaying networks based on OFDMA,” in *Proc. IEEE 68th VTC—Fall*, Sep. 2008, pp. 1–5.

[28] W.-C. Kan and S.-R. Yang, The Supplements to “Handoff minimization through a relay station grouping algorithm with efficient radio resource scheduling policies for IEEE 802.16j multihop relay networks,” Nat. Tsing Hua Univ., Hsinchu, Taiwan. [Online]. Available: http://www.cs nthu.edu.tw/~sryang/submission/TechHMRSGA.pdf

[29] X. Guo, W. Ma, Z. Guo, X. Shen, and Z. Hou, “Adaptive resource reuse scheduling for multihop relay wireless network based on multicoloring,” *IEEE Commun. Lett.*, vol. 12, no. 3, pp. 176–178, Mar. 2008.

Handoff Minimization through a Relay Station Grouping Algorithm with Efficient Radio Resource Scheduling Policies for IEEE 802.16j Multihop Relay Networks

Shun-Ren Yang, *Member, IEEE*, Chien-Chi Kao, *Student Member, IEEE*, Wai-Chi Kan, and Tzung-Chin Shih

***Abstract*—The IEEE 802.16j standard has been developed to provide performance enhancement to the existing IEEE 802.16e network by incorporating the multihop relay (MR) technology. However, frequent handoffs and low spectrum utilization issues that were not encountered in IEEE 802.16e may be incurred in IEEE 802.16j. The relay station (RS) grouping is one optional mechanism in the IEEE 802.16j MR standard to overcome these problems. The concept of RS grouping is to group neighboring RSs together to form an RS group, which can be regarded as a logical RS with larger coverage. In the paper, we investigate the RS grouping performance enhancement in terms of throughput and handoff frequency. This paper designs an RS grouping algorithm to minimize handoffs by utilizing a greedy grouping policy: RS pairs with higher handoff rates will have higher priority to be selected. The simulation results show that the handoff frequency of the considered MR network can be significantly reduced and suitable RS grouping patterns can be derived using our grouping algorithm. In addition, we propose two centralized scheduling policies, the throughput-first (TF) policy to maximize the system throughput and the delay-first (DF) policy to minimize the average packet delay. By integrating our RS grouping algorithm and centralized scheduling algorithms, the simulation results indicate that for the case of fixed users, the groupings with smaller group sizes can result in better throughput performance. However, when user mobility is considered, the throughput value increases as the group size increases. Furthermore, we also show that the DF policy can both minimize the average packet delay, and provide the fairness property among users with different traffic loads.**

***Index Terms*—Grouping algorithm, IEEE 802.16j, multihop relay, scheduling policy, WiMAX.**

Manuscript received March 19, 2009; revised August 18, 2009 and November 16, 2009. This work was supported in part by the National Science Council (NSC) of Taiwan under Contracts 96-2752-E-007-003-PAE, 96-2221-E-007-025-, 96-2221-E-007-027-, 96-2219-E-007-012- and 96-2219-E-007-011-, and Chunghwa Telecom. The review of this paper was coordinated by Prof. Vincent Wong.

S.-R. Yang is with the Department of Computer Science and Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan 300, R.O.C. (e-mail: sryang@cs.nthu.edu.tw).

C.-C. Kao is with the Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan 300, R.O.C. (e-mail: mickey@wmnet.cs.nthu.edu.tw).

W.-C. Kan was with the Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan 300, R.O.C. He is now with the Synology Incorporated, Taipei, Taiwan 10351, R.O.C. (e-mail: waichik@gmail.com).

T.-C. Shih is with the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan 300, R.O.C. (e-mail: stonc@wmnet.cs.nthu.edu.tw).

I. INTRODUCTION

Incorporating the *multihop relay* (MR) technology [25], the IEEE 802.16j MR standard [14] has been developed to provide throughput improvement, coverage extension and capacity enhancement to the existing IEEE 802.16e protocol [1]. By deploying relay stations (RSs), the end-to-end communication quality between base stations (BSs) and mobile stations (MSs) can be improved without high infrastructure deployment costs. In particular, it becomes possible to forward data to an MS using a high transmission rate in Line-Of-Sight (LOS) conditions through an MR path to avoid the Non-Line-Of-Sight (NLOS) direct (i.e., single-hop) transmission with bad channel quality. In addition, *spatial reuse* [19] is another promising approach that can be employed in IEEE 802.16j MR networks to improve spectral efficiency. Based on the centralized scheduling, spatial diversity gain can be achieved if multiple simultaneous transmissions using the same bandwidth resources are realized within the same BS-cell.

Although IEEE 802.16j has the potential to provide substantial performance enhancements, several issues that were not addressed in IEEE 802.16e may be encountered. For example, frequent handoffs may occur during the movements of MSs since RS-cells are smaller than BS-cells. Moreover, the unbalanced resource allocation of different RSs may result in inefficient spectrum utilization. To avoid the consequent performance degradation, *RS grouping* has been identified as an optional mechanism in the IEEE 802.16j standard. The main idea of RS grouping is to put adjacent RSs together to form an *RS group*, wherein the RS members are required to receive and transmit the same data simultaneously. From MSs' viewpoint, the RS group can be regarded as a logical RS with larger coverage. Therefore, the handoff frequency can be reduced since no handoff procedure would be triggered even when an RS crossing event within the RS group occurs. In addition, since the radio resources of the RS members are aggregated together, the resources can then be allocated more flexibly and efficiently to achieve higher spectrum utilization. However, to the best of our knowledge, no RS grouping strategy has been proposed or discussed in the literature. We argue that different grouping criteria may lead to various performance results. Specifically, utilizing a smaller RS group size is advantageous to spatial reuse because more RS groups can perform simultaneous transmissions at the same time and frequency. Thus, improved average system throughput can be expected. However, a smaller RS group size may also lead to higher packet loss rate due to more frequent handoffs between RS groups. In conclusion, when implementing the RS grouping mechanism in IEEE 802.16j MR networks, the performance tradeoff between throughput and handoff frequency should be seriously considered.

In this paper, we analyze grouping strategies for RS grouping-enabled IEEE 802.16j MR networks and propose an efficient RS grouping algorithm to minimize the handoff frequency. As we have pointed out, RS grouping strategies will also influence the throughput performance. To investigate the impacts of RS grouping on the IEEE 802.16j system throughput, we design two centralized downlink scheduling policies for RS grouping-enabled IEEE 802.16j MR networks. One of these two scheduling policies aims to maximize the system throughput and the other is to minimize average packet delay. The throughput estimation results under different grouping configurations can assist network service providers to choose the most appropriate settings of grouping factors (e.g., group size). Notice that the spatial reuse concept is considered during the scheduling procedure to improve the spectral efficiency. By integrating our proposed RS grouping algorithm and centralized scheduling policies, the simulation results show that the throughput and delay performance can be improved in addition to the significantly reduced handoff frequency. The main contribution of this paper is to propose the first integrated algorithmic framework that can be utilized to investigate the performance interaction between RS grouping and resource scheduling for IEEE 802.16j MR networks.

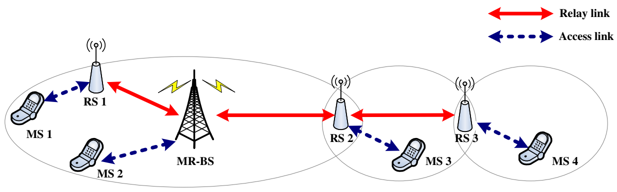


Fig. 1. Sample topology of IEEE 802.16j MR networks

The remaining parts of this paper are organized as follows. In sections II and III, we give an overview of the IEEE 802.16j standard and a discussion of the IEEE 802.16j RS grouping mechanism, respectively. Section IV presents an RS grouping strategy analysis and details our proposed RS grouping algorithm. In section V, two efficient centralized downlink scheduling policies for IEEE 802.16j MR networks are addressed. The performance evaluations of our grouping algorithm and scheduling policies are presented in section VI. Section VII summarizes the related work. Finally, we conclude the paper in section VIII.

II. IEEE 802.16J MULTIHOP RELAY NETWORKS

A. Network architecture

The IEEE 802.16j standard is expected to enhance the system performance of IEEE 802.16-based networks through multihop relaying technologies. A typical topology example of IEEE 802.16j MR networks is illustrated in Fig. 1. In this network, an MS can access the MR-BS either through a multihop relaying path (e.g., MS1, MS3 and MS4) or directly (e.g., MS2). In addition, a station (BS or RS) is called an *access station* if it provides network attachment functionality to a given MS or RS. On the other hand, an RS is a *subordinate* RS of another station if that station serves as the access station for that RS. For instance, RS2 is the access station of RS3, and RS3 is a subordinate RS of RS2. The wireless links that directly connect access stations with their respective subordinate RSs are called *relay links*, while the links between MSs and their corresponding access stations are known as *access links*. Since an RS can only be subordinated to one station, the MR-BS and the RSs in this MR network form a tree-based multihop relay topology. Note that it has been shown that the throughput decreases as the number of hop-counts increases [7], and thus we only investigate two-hop IEEE 802.16j networks in this paper.

B. Frame structure

The frame structure of IEEE 802.16j MR systems is extended from that of IEEE 802.16e networks, which also adopt orthogonal frequency division multiple access (OFDMA) as the primary channel access mechanism for NLOS communication. The basic unit of resource for allocation in OFDMA is a *slot*, which is comprised of a number of symbols in the time domain, and one sub-channel in the frequency domain. The timeline is divided into contiguous frames, each of which further consists of a downlink (DL) and a uplink (UL) subframes. In IEEE 802.16j, the DL and UL subframes shall include one *access zone* for MR-BS ↔ RS and MR-BS ↔ MS transmissions and may include one *relay zone* for RS ↔ subordinate MS transmissions, respectively.

III. IEEE 802.16j RS Grouping Mechanism

Although deploying RSs in IEEE 802.16 networks can provide significant throughput or coverage enhancements, several issues regarding the relaying architecture of IEEE 802.16j should be addressed. These issues include frequent handoffs, redundant control overhead and low spectral efficiency. It is perceived that these issues will result in unpredictable performance reduction for IEEE 802.16j MR networks. Therefore, the IEEE 802.16j standard provides the optional RS grouping mechanism to reduce the impacts of these issues. The concept of an RS grouping is that adjacent RSs could be grouped together as an RS group which acts as a virtually regular RS to its associated MSs. The grouping criteria are decided by the controlling MR-BS, based on the targeted performance requirement. Note that the coverage of an RS group is larger than that of its regular member RSs, and no handoff event would be triggered even though an RS-cell crossing event within the same RS group occurs. Consequently, the MSs under the RS grouping mechanism will experience lower handoff probability. On the other hand, MR-BSs can manage RS groups using only one set of control header, and hence the control signal overheads are reduced. Finally, the radio

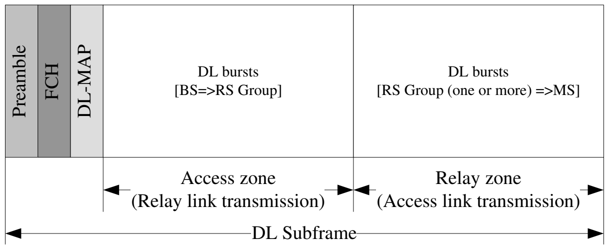


Fig. 2. Modified IEEE 802.16j downlink subframe structure for supporting RS grouping

resource (i.e., the relay zone) of each member RS can be aggregated and shared by all the MSs under the corresponding RS group, so that the spectral efficiency is improved.

For the downlink operation, the member RSs of a group should be configured to transmit equivalent data signals to the same MS. Thus the subordinated MSs can receive the best quality signal within the group, no matter where they are located. This operation is also called cooperative transmission because the member RSs will form a virtual antenna array to exploit macro-diversity. On the other hand, the diversity combining of the information received by the member RSs of an RS group can be performed in the uplink situation. Both the downlink and uplink diversity gains can be achieved under a centralized scheduling scheme by keeping the MS list of each RS group at the respective MR-BS. Therefore, the RS grouping mechanism is reasonable to improve the data transmission rate.

To support an RS grouping, the original IEEE 802.16j frame structure should be modified (see Fig. 2 for the modified DL subframe). Specifically, access zones should handle the transmissions between BSs and RS groups, while relay zones should handle the transmissions between RS groups and subordinate MSs. From the mobility management point of view, we note that implementing RS grouping will not incur extra costs to IEEE 802.16j. As aforementioned, for movement between RS-cells of an RS group, an MS will not initiate the handoff procedure. The MS CDMA periodic ranging process with aggregated ranging sub-channel allocation [13] can be employed to handle the RS reselection during intra-RS-group movement. On the other hand, when the MS roams from an RS group to another RS group, since an RS group can be seen as a legacy BS, the conventional MAC layer handoff procedure is applied directly for this scenario.

IV. Our Proposed RS Grouping Algorithm for IEEE 802.16j MR Networks

This section proposes an IEEE 802.16j RS grouping algorithm to reduce the handoff frequency of mobile users under prescribed IEEE 802.16j MR network performance requirements. The system assumptions and the concept of our proposed algorithm are described in the first subsection. Then, the proposed algorithm is discussed as a three-phase procedure in the following subsections.

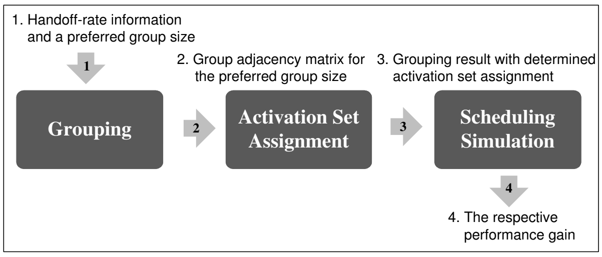


Fig. 3. The concept of the proposed RS grouping algorithm

A. System Assumptions and the Concept of Our RS Grouping Algorithm

To accommodate general scenarios, our algorithm does not make any assumptions of the underlying IEEE 802.16j MR network topology, the user mobility behavior, and/or the packet traffic pattern. Specifically, within a considered BS, the IEEE 802.16j RSs can be deployed arbitrarily and the coverage area of each RS can be irregular. In addition, the MSs within the considered BS can move randomly. In such an arbitrary environment, we only require that the handoff-rate information between each two RS-cells should be available. The handoff rate between two RS-cells, RS-cell 1 and RS-cell 2, is the total rate that the resident MSs hand off from RS-cell 1 to RS-cell 2 or from RS-cell 2 to RS-cell 1. Note that the handoff-rate information can be simply derived from the statistical data that are collected by the network service providers.

To design the RS grouping algorithm in our considered environment, we first specify the factors that may affect the grouping result. First, the group size is a factor which has the potential to influence the spatial diversity gain and the handoff frequency. It can be observed that utilizing a smaller group size will result in more RS groups. Such a grouping policy is beneficial to spatial reuse, though it causes higher handoff frequency. On the other hand, a larger group size has reverse effects on the spatial reuse and the handoff frequency, respectively. In addition to group size, the selection order of group members is also significant to the grouping strategy design. Even with the same group size, different grouping orders may lead to different numbers of handoff events. Once an RS grouping layout is provided by applying the most desired group size and grouping order, determining which set of RS groups to transmit data simultaneously further impacts the spectrum efficiency critically. In this paper, we define an *activation set* as a particular set of RS groups which can transmit data to their respective resident MSs at the same time and frequency without interference. Clearly, the assignment of activation sets is another important factor that should be considered in the design of the RS grouping algorithm.

Taking all the above factors into account, we propose our RS grouping algorithm, which contains three phases as illustrated in Fig. 3: *grouping phase*, *activation-set-assignment phase* and *scheduling-simulation phase*. Based on the handoff-rate information, the grouping phase first constructs the handoff-minimizing RS grouping result for a preferred group size. Then, the grouping phase generates the adjacency matrix of the constructed RS groups. Based on the adjacency matrix, the activation-set-assignment phase assigns each RS group to an activation set in order to enhance the spatial diversity gain. Afterward, in the scheduling-simulation phase, the downlink transmission simulation is executed for the given grouping result with determined activation set assignment. By comparing the performance gains of the scheduling results from different preferred group sizes, network service providers can finally choose the most favorable grouping result with the associated activation set assignment as their IEEE 802.16j MR network configuration.

The operation of our RS grouping algorithm is depicted in Algorithm 1, where two input parameters, the set α of all the RSs within a considered BS and the handoff-rate matrix H of the RSs in α, should first be provided. In H, the entry $H[i][j]$ represents the total handoff rate between RS-cell i and RS-cell j. After variable initialization, the three algorithm phases at lines 14-16 are repeated for each preferred group size. Then, at line 17, the weighted function $\omega(S_i)$ calculates the weighted performance gain for a simulation result S_i. The highest performance gain is recorded in key at line 18 and the best grouping result with determined activation set assignment is stored in Ω at line 19.

Algorithm 1 THE PROPOSED RS GROUPING ALGORITHM

Input:
α (the set of RSs within a considered BS), H (the handoff-rate matrix of the RSs in α).

Output:
Ω (the desired grouping result with determined activation set assignment).

Initialization:
$\Omega \leftarrow NIL$;
$key \leftarrow 0$;
$M \leftarrow rows[H]$;

Procedure:
for $i = 1$ to M **do**
  $\Gamma_i, A_i \leftarrow Grouping(\alpha, H, i)$;
  $\Omega_i \leftarrow ActivationSet(A_i)$;
  $S_i \leftarrow Scheduling(\Gamma_i, \Omega_i)$;
  if $\omega(S_i) > key$ **then**
    $key \leftarrow \omega(S_i)$;
    $\Omega \leftarrow \Omega_i$;
  end if
end for

When Algorithm 1 terminates, the desired grouping result can thus be obtained. In the following subsections, we present the details of each phase individually.

B. Grouping Phase

The operation of the grouping phase is shown in Algorithm 2 and is explained as follows. According to the functionalities, the RS grouping procedure is partitioned into four main portions:

Input parameters: At the beginning of the procedure (line 2), three parameters, α, H and X, are required as the inputs, where X denotes the preferred group size.

Output results: The outputs of the procedure (line 5) are the final grouping result $\Gamma = \{g_1, g_2, ..., g_n, ...\}$ and the adjacency matrix A of the constructed RS groups, where g_n denotes the nth RS group.

Initialization stage: In this stage (lines 8 to 10), some variables must be initialized before executing the grouping procedure. We use $Temp_\alpha$ to denote the set of the ungrouped RSs within the considered BS. $Temp_\alpha$ is initialized as α. Then, we use M to denote the total number of RSs in α. Moreover, the variable n is initialized as 0.

Procedure: After initialization, the main grouping procedure is started from line 13 to line 47. Since the purpose of our algorithm is to minimize the handoff probability, we introduce a greedy grouping policy, under which RS pairs with higher handoff rates (provided from H) will have higher priority to be selected. A detailed description of the steps is as follows.

- Group-construction loop for non-single-RS groups (lines 14 to 27): This loop constructs RS groups from the ungrouped adjacent RS pairs. The ungrouped adjacent RS pair with the highest handoff rate is first considered (line 16). Of this pair, the RS that has a higher total handoff rate to/from other RSs in $Temp_\alpha$ is selected to be the starting member of the nth group g_n (lines 17 to 21). Then, the ungrouped neighboring RSs of g_n are selected to join g_n in descending order of the total handoff rate to/from g_n. This join procedure (lines 22 to 25) is repeated until the group size of g_n is equal to the preferred group size X or g_n has no ungrouped neighboring RS. This group-construction loop is iterated to form RS groups

Algorithm 2 $Grouping(\alpha, H, X)$

1: ***Input***:
2: α (the set of RSs within a considered BS), H (the handoff-rate matrix of the RSs in α), X (the preferred group size).
3:
4: ***Output***:
5: Γ (a set of constructed RS groups $\{g_1, g_2, ..., g_n, ...\}$), A (the adjacency matrix of the constructed RS groups in Γ).
6:
7: ***Initialization***:
8: $Temp_\alpha \leftarrow \alpha$;
9: $M \leftarrow rows[H]$;
10: $n \leftarrow 0$;
11:
12: ***Procedure***:
13: /* Each RS group is constructed based on the handoff-rate matrix H. */
14: **while** there exist pairs of ungrouped adjacent RSs in $Temp_\alpha$ **do**
15:   $n \leftarrow n + 1$;
16:   Find the ungrouped adjacent RS pair (a, b) that has the highest handoff rate $H[a][b]$;
17:   **if** $a.TotHR(Temp_\alpha) \geq b.TotHR(Temp_\alpha)$ **then**
18:     Remove a from $Temp_\alpha$ and add it into the nth group g_n;
19:   **else**
20:     Remove b from $Temp_\alpha$ and add it into the nth group g_n;
21:   **end if**
22:   **while** $|g_n| < X$ and group g_n has ungrouped neighboring RSs **do**
23:     Find the neighboring RS c with the highest total handoff rate to/from group g_n;
24:     Remove c from $Temp_\alpha$ and add it into g_n;
25:   **end while**
26:   Add g_n into Γ;
27: **end while**
28: /* An isolated RS forms an RS group itself. */
29: **while** there exists an isolated ungrouped RS d in $Temp_\alpha$ **do**
30:   $n \leftarrow n + 1$;
31:   Remove d from $Temp_\alpha$ and add it into the nth group g_n;
32:   Add g_n into Γ;
33: **end while**
34: /* Initialize the RS-group adjacency matrix A. */
35: **for** $i = 1$ to n **do**
36:   **for** $j = 1$ to n **do**
37:     $A[i][j] \leftarrow 0$;
38:   **end for**
39: **end for**
40: /* Compute the RS-group adjacency matrix A. */
41: **for** $k = 1$ to M **do**
42:   **for** $l = 1$ to M **do**
43:     **if** the handoff rate $H[k][l] > 0$, RS $k \in g_i$ and RS $l \in g_j$ **then**
44:       $A[i][j] \leftarrow 1$;
45:     **end if**
46:   **end for**
47: **end for**

until no pair of ungrouped adjacent RSs exists.

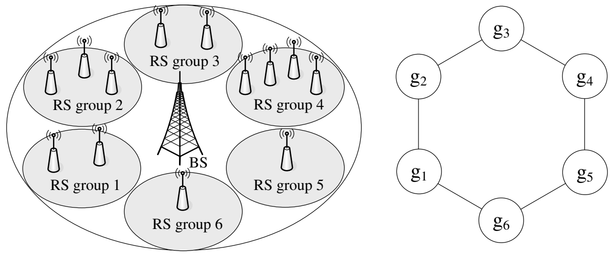
- Group-construction loop for isolated RSs (lines 29 to 33): This loop constructs RS groups for isolated ungrouped RSs. Each isolated ungrouped RS forms an RS group itself (line 31).
- Group adjacency matrix initialization (lines 35 to 39): Each entry of the group adjacency matrix A is initialized as 0.
- Group adjacency matrix computation (lines 41 to 47): After initialization, the adjacency matrix A is computed by examining whether a member of a group g_i is adjacent to a member of another group g_j (line 43). If yes, the entry $A[i][j]$ is set to 1 (line 44). After examining all the M RSs, the adjacency matrix A of the RS groups in Γ can be derived.

After the above grouping procedure, the grouping result Γ and the group adjacency matrix A can be obtained for use in the subsequent activation-set-assignment phase. We note that the time complexity of the grouping phase is mainly dominated by the two nested loops in lines 14-27 and in lines 41-47. For the nested loop in lines 14-27, the numbers of iterations for both the outer and inner loops are no larger than $|\alpha|$. On the other hand, for the nested loop in lines 41-47, the numbers of iterations for the outer and inner loops are both equal to $|\alpha|$. Therefore, the overall time complexity of the grouping phase is $O(|\alpha|^2)$.

C. Activation-Set-Assignment Phase

Before elaborating on how our activation-set-assignment phase can generate appropriate assignment results, we first point out that the number of activation sets significantly affects the data transmission performance. Specifically, the fewer the activation sets, the more the RS groups that can transmit data simultaneously. Consider the RS-group layout in Fig. 4(a), where its RS-group adjacency graph is shown in Fig. 4(b). Clearly, the assignment $\{\{g_1, g_3, g_5\}, \{g_2, g_4, g_6\}\}$ is better than another assignment $\{\{g_1, g_4\}, \{g_2, g_5\}, \{g_3, g_6\}\}$ since the former assignment contains fewer activation sets and thus has improved spatial diversity gain.

Given an RS grouping result Γ and the corresponding adjacency matrix A from Algorithm 2, we first derive the adjacency graph $G(V, E)$ where V represents the set of the RS groups and E represents the interference relation among the RS groups. To minimize the number of activation sets, we model the activation-set-assignment



(a) The RS-group layout (b) The RS-group adjacency graph

Fig. 4. An activation-set-assignment example

problem as the minimum coloring problem of G. The rule to color all the vertices of G is that no two adjacent vertices can share the same color. Based on this rule, each set of the vertices that are assigned the same color is equivalent to an activation set. However, to color G with the minimum number of colors is known as an NP-hard problem. To achieve a polynomial computation time, the well-known greedy coloring algorithm, Welsh Powell algorithm [6], is adopted in this paper. For the self-containedness purpose, we briefly summarize the key concept of the Welsh Powell algorithm below. First, all possible colors are numbered. Then following the descending order of vertex degree, each vertex in G is sequentially selected to be assigned a color. When a vertex v attempts to be colored, the first color is examined to check if it has already been occupied by any of v's neighbors. If no, the first color is assigned to v. Otherwise, the next unoccupied color (by v's neighbors) is used to color v. This procedure is repeated for the subsequent vertices until the vertices in G have all been processed. The time complexity of the Welsh Powell algorithm is proven to be $O(|V|^2)$. This algorithm guarantees that the number of required colors is at most one more than the maximum degree $\Delta(G)$ of G. That is, the Welsh Powell algorithm determines at most $\Delta(G) + 1$ activation sets. Our activation-set-assignment phase applies the Welsh Powell algorithm and therefore can assign each constructed RS group (from the grouping phase) to an appropriate activation set.

D. Scheduling-Simulation Phase

In this subsection, we characterize the concept of the scheduling simulation. By adopting a specific scheduling policy, scheduling simulations are conducted for the different grouping results with associated activation set assignments provided from the grouping and activation-set-assignment phases. From these simulations, several output measures (e.g., system throughput and packet delay time) can be derived. The network service providers can evaluate the performance gains of these grouping results by assigning weighted values to the respective output measures following their preference rules. Then, by comparing the weighted indices from the different grouping results, the most desirable grouping result can be determined. In the following sections, the scheduling policies for RS grouping-enabled IEEE 802.16j MR networks will be elaborated on in section V and the representative simulation results will be presented in section VI.

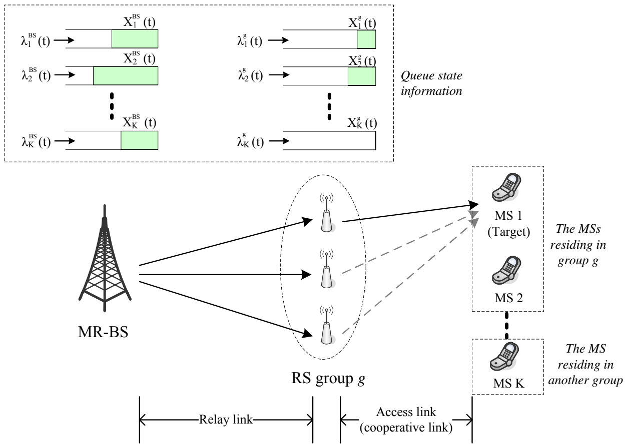


Fig. 5. System model of RS grouping-enabled IEEE 802.16j MR networks

V. Multihop Centralized Downlink Scheduling Policies

In the scheduling-simulation phase of our RS grouping algorithm described in section IV, centralized downlink scheduling policies are required to evaluate the performance gain of a given RS grouping layout. In this section, we first define the scheduling problem for RS grouping-enabled IEEE 802.16j MR networks. Then the system description of our considered multihop network is addressed. Finally, we propose two centralized downlink scheduling policies for IEEE 802.16j MR networks under RS grouping and spatial reuse assumptions, with the objectives of maximizing the system throughput and minimizing the downlink traffic delay, respectively.

A. Scheduling problem for RS grouping-enabled IEEE 802.16j MR networks

Consider the downlink transmission of an IEEE 802.16j MR network with a BS and a set of preconfigured RS groups, as shown in Fig. 5. Assume that a number of packets from the external networks are desired to be delivered to MS 1 residing in RS group g. The BS, which acts as the network gateway for its served MSs, will first buffer these packets for MS 1 in the corresponding packet queue. In an appropriate time frame, the BS will transmit these packets to RS group g through the relay link transmission. These packets received by RS group g are also buffered in the packet queue corresponding to MS 1 until the access link transmission between RS group g and MS 1 is scheduled in another following frame.

To realize this two-hop relaying operation, the scheduling procedure should be performed by the BS at the beginning of each frame. Specifically, the scheduling procedure consists of two parts: *relay link scheduling* and *access link scheduling*. In the relay link scheduling, the BS selects an RS group. The relay link between the BS and this RS group will be activated during the access zone for transmitting packets destined to the MSs residing in this RS group. These packets will be queued in the RS group for future relaying. On the other hand, in the access link scheduling, the BS selects an appropriate activation set of RS groups. The access links between these RS groups and their respective resident MSs will be activated during the relay zone for packet relaying. Note that since the BS has only one radio transmitter, only one relay link between the BS and a particular RS group can be activated at a time in the access zone of each time frame. However, based on the concept of spatial reuse, more than one access link transmissions can be activated simultaneously in the relay zone.

To resolve the scheduling problem for RS grouping-enabled IEEE 802.16j MR networks, scheduling policies should be employed to decide suitable relay link transmissions in access zones and access link transmissions in relay zones. Different scheduling policies may be designed based on different criteria to meet different system performance requirements.

B. Radio Resource Scheduling Policies

1) System model: Consider again the single-BS IEEE 802.16j MR network in Fig. 5. Using our RS grouping algorithm, the RSs in this network are partitioned into a set Γ of RS groups, and the RS groups are classified into a set Ω of activation sets. The RS groups within each activation set can be activated simultaneously.

For both the relay link and access link transmissions in the considered IEEE 802.16j MR network, the Rayleigh fading channel model is assumed to be applied in the physical layer. Letting $src(l)$ be the transmitter and $dst(l)$ the receiver of a radio link l, we can compute the transmission rate over the radio link l as

$$R_l = B \cdot \log_2(1 + |h_{dst(l)}^{src(l)}|^2 \cdot PW_{src(l)}/\sigma_N^2). \quad (1)$$

In (1), B denotes the bandwidth (in Hz) of the MR network spectrum, and $PW_{src(l)}$ and σ_N^2 represent the transmission power of $src(l)$ and the variance of signal noise, respectively. Moreover, $h_{dst(l)}^{src(l)}$ in (1) is a circularly symmetric complex Gaussian random variable. Readers interested in the derivation of (1) are referred to [28] for a detailed explanation.

During the relay link transmission, the BS adopts the minimum transmission rate among the relay links between the BS and the selected RS group as the real transmission rate. This is due to the fact that the RSs within the selected RS group must correctly receive and decode the data from the BS at the same time and frequency. In this case, the effective transmission rate is dominated by the relay link with the minimum rate. Thus the transmission rate of the relay link between the BS and an RS group g can be computed based on (1) as

$$R_{BS,g} = B \cdot \log_2(1 + \min_{\forall r \in g} |h_r^{BS}|^2 \cdot PW_{BS}/\sigma_N^2). \quad (2)$$

For the access link transmission, although the RS members in the same RS group should serve a particular MS at the same time, this signal-combining cooperative transmission with multiple sources is difficult to implement. Therefore, we adopt a compromise cooperation scheme called *selection relaying* [15] to realize the access link transmission, whereby only the RS member with the highest signal quality is selected for data transmission towards a resident MS. Therefore, the transmission rate between the RS group g and a resident MS j can be derived similar to (2) as

$$R_{g,j} = B \cdot \log_2(1 + \max_{\forall r \in g} |h_j^r|^2 \cdot PW_r/\sigma_N^2). \quad (3)$$

In the two proposed scheduling policies that will be described later, the network queue state is a common parameter utilized to make the scheduling decision. We denote $X_j^{BS}(t)$ as the queue length of the queue corresponding to MS j in the BS at the start of time frame t. In the RS group g of MS j, every RS member will also maintain a queue for MS j. However, as mentioned above, all the RS members serve MS j at the same time. Thus from MS j's viewpoint, the queues maintained in the RSs for MS j can be regarded as a logical single queue. We denote $X_j^g(t)$ as the length of the logical queue for MS j in RS group g at the start of time frame t.

Note that the RS buffer synchronization for selection relaying can be easily achieved by utilizing the standard IEEE 802.16j ARQ mechanism and by arranging a new multicast information element in the DL-MAP. More specifically, when the BS receives an ACK from an MS indicating the sequence number of the latest correctly decoded packet, it will, in the next time frame, announce an additional DL-MAP multicast information element containing this sequence number to the corresponding RS group. The obsolete packets at each individual RS buffer can then be removed accordingly.

2) General scheduling procedure of the RS grouping-enabled IEEE 802.16j MR network: Recall that the scheduling procedure should make the decision of the relay link and access link transmissions of a frame in the relay link scheduling and access link scheduling, respectively. Although the selection criteria of these two parts are actually the same under the given scheduling policy, some detailed processes are different due to the special characteristics of RS grouping such as minimum transmission rate constraints and selection relaying cooperative transmissions. Therefore, we detail the procedures of these two scheduling parts individually as follows.

Part 1: Relay link scheduling. In this part, the objective of the scheduler is to activate the most favorable relay link between the BS and an RS group in Γ. For comparison, we claim that the major task of a scheduling policy is to assign a weighted index $D_g^{RL}(t)$ for each RS group g to represent the priority of RS group g to be selected in the relay link scheduling of time frame t. Note that the implementation of $D_g^{RL}(t)$ is dependent on the concerned performance metrics (e.g., throughput or delay time) of the scheduling policy. In order to make the scheduling decision in a channel-aware manner, the scheduling policy may require the transmission rate $R_{BS,g}$ to decide the $D_g^{RL}(t)$ of RS group g. Two different approaches are given later in this section. After both the transmission rate and the weighted index computations of each RS group, the final target RS group $\hat{g}(t)$ for time frame t is selected as follows:

$$\hat{g}(t) = \arg\max_{g\in\Gamma}\{D_g^{RL}(t)\}. \quad (4)$$

In other words, the RS group with the maximum weighted index will be selected as the target RS group. During the target RS group selection, the set $\hat{I}_g(t)$ of MSs whose data bursts will be delivered to the chosen RS group $\hat{g}(t)$ should also be identified. This MS identification is based on the underlying performance metrics as well. Following the scheduling result, the data bursts belonging to these MSs in $\hat{I}_g(t)$ are scheduled in the access zone of the time frame t for transmission to $\hat{g}(t)$. All RS members of $\hat{g}(t)$ should receive and buffer the data for the purpose of second-hop transmission.

Part 2: Access link scheduling. To achieve simultaneous access link transmissions in each time frame, the scheduler is responsible for selecting an appropriate activation set and the corresponding target MSs in the access link scheduling. Likewise, a scheduling policy is also needed to assign a weighted index $D_g^{AL}(t)$ for each RS group g to denote the priority of RS group g to be selected in the access link scheduling of time frame t. The transmission rate $R_{g,j}$ may also be required to decide the $D_g^{AL}(t)$ of RS group g by the scheduling policy. Since the activation set assignment is provided a priori, the activation set $\hat{\phi}(t)$ with the maximum total RS group weighted indices can be selected from Ω for transmission at time frame t, namely

$$\hat{\phi}(t) = \arg\max_{\phi\in\Omega}\{\sum_{g\in\phi} D_g^{AL}(t)\}. \quad (5)$$

Regarding the choice of MSs to be served in the access link scheduling, we denote $\hat{j}_g(t)$ as the MS residing in RS group $g \in \hat{\phi}(t)$ whose packets will be transmitted over the selected cooperative access link at time frame t. In the following two proposed scheduling policies, the respective $\hat{j}_g(t)$ are provided in (8) and (14).

3) Throughput-First Scheduling Policy: Different scheduling policies may be applied in the scheduling procedure to provide the different definitions of the weighted indices $D_g^{RL}(t)$ and $D_g^{AL}(t)$ and the MS selections $\hat{I}_g(t)$ and $\hat{j}_g(t)$ based on their respective performance criteria. Similar to the throughput-optimal scheduling policy mentioned in [11], we introduce a throughput-first (TF) scheduling policy which is compatible to the scheduling procedure described in subsection V-B2. The objective of the TF policy is to achieve the maximum system throughput for our considered IEEE 802.16j MR network.

Consider the scheduling procedure using the TF scheduling policy. Let Λ_g be the set of MSs residing in the same group g. In the relay link scheduling of time frame t, the weighted index $D_g^{RL}(t)$ of a group $g \in \Gamma$ is defined as

$$D_g^{RL}(t) = R_{BS,g}\sum_{j\in\Lambda_g} X_j^{BS}(t), \quad (6)$$

where $R_{BS,g}$ is computed from (2). In the access link scheduling, the weighted index $D_g^{AL}(t)$ is defined as

$$D_g^{AL}(t) = \max_{j\in\Lambda_g}[R_{g,j}\cdot X_j^g(t)], \quad (7)$$

where $R_{g,j}$ is computed from (3). By (6) and (7), the transmission links with larger queue length and higher transmission rate are desired to be served. For the demonstration purpose, the simple round-robin approach is applied to determine the $\hat{I}_g(t)$ set during the relay link scheduling. On the other hand, during the access link scheduling, the MS selection has already been done in the computation of $D_g^{AL}(t)$. That is, if a group $g \in \hat{\phi}(t)$ is selected to be active at time frame t, the MS whose packets will be transmitted by the group g is

$$\hat{j}_g(t) = \arg\max_{j\in\Lambda_g}[R_{g,j}\cdot X_j^g(t)]. \quad (8)$$

4) Delay-First Scheduling Policy: Since delay-sensitive applications (e.g., Voice over Internet Protocol call) are widely used, a delay-oriented scheduling policy for IEEE 802.16j MR networks is desired. With compatibility to the scheduling procedure described in subsection V-B2, we propose a delay-first (DF) scheduling policy which is aimed to minimize the average packet delay time. The concept of the DF scheduling policy is to first serve the user queue with the maximum predicted mean packet waiting time.

We assume that there is an observation time window with length T_w. The average packet arrival rate $\lambda_j^i(t)$ of MS j on a transmitter i (the BS or an RS group) of the T_w frames before time frame t can be estimated with an iterative approach as follows

$$\lambda_j^i(t) = (1-\frac{1}{T_w})\lambda_j^i(t-1) + \frac{1}{T_w}a_j^i(t), \quad (9)$$

where $a_j^i(t)$ denotes the packet arrival rate of MS j on the transmitter i in time frame t. Note that the $\lambda_j^i(t)$ in

(9) is updated at the end of each frame transmission on all transmitters. Similarly, the mean queue length $\bar{X}_j^i(t)$ of MS j on a transmitter i of the T_w frames before time frame t can be estimated as

$$\bar{X}_j^i(t) = (1 - \frac{1}{T_w})\bar{X}_j^i(t-1) + \frac{1}{T_w}X_j^i(t). \quad (10)$$

Therefore, by Little's formula [18], the predicted mean packet waiting time of MS j on the transmitter i in time frame t can be written as

$$W_j^i(t) = \frac{\bar{X}_j^i(t)}{\lambda_j^i(t)}. \quad (11)$$

Let $W_j^{BS}(t)$ and $W_j^g(t)$ be the predicted mean waiting times for the packets of MS j at the BS and an RS group $g \in \Gamma$, respectively. In the relay link scheduling, the weighted index $D_g^{RL}(t)$ of RS group g is defined as

$$D_g^{RL}(t) = \sum_{j \in \Lambda_g} W_j^{BS}(t). \quad (12)$$

Moreover, the weighted index $D_g^{AL}(t)$ in the access link scheduling is formulated as

$$D_g^{AL}(t) = \max_{j \in \Lambda_g} W_j^g(t). \quad (13)$$

Similarly, for the demonstration purpose, the $\hat{I}_g(t)$ set for relay link transmission is also determined in a round-robin manner. Finally, if a group $g \in \hat{\phi}(t)$ is selected to be active in time frame t by the access link scheduling, the MS whose packets will be transmitted by the group g is

$$\hat{j}_g(t) = \arg\max_{j \in \Lambda_g} W_j^g(t). \quad (14)$$

VI. Simulation Results and Discussion

In this section, we present the performance evaluations of the RS grouping algorithm and centralized downlink scheduling policies. We consider an IEEE 802.16j MR network with 10-MHz system bandwidth and 2.5-GHz carrier frequency. Note that only the downlink transmission is discussed in this paper, and therefore the MAC frame structure in our simulation experiments is simplified to comprise only the DL subframe in each time frame. The MAC frame length is set to 10 ms. Since we only focus on the performance of resource allocation, the controlling signal overheads are not considered in the simulation, and the durations of the access zone and the relay zone within each MAC frame are set to 0.5 ms and 9.5 ms, respectively. Moreover, we assume that the relay link (BS-to-RS) is operated under the LOS channel

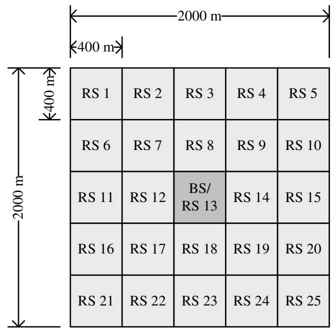


Fig. 6. Simulation topology: A single BS-cell system with a 25-RS-cell grid layout

condition, and thus the target throughput rate R_{TH} is assumed to be 60 Mbps. On the other hand, for the access link (RS-to-MS) transmission, the NLOS channel condition with $R_{TH} = 40$ Mbps is considered.

To reduce our simulation complexity, the square-based cellular topology depicted in Fig. 6 is adopted in our simulation environment. We consider a single square-shaped BS-cell topology where the BS is at the center and 25 square-shaped RS-cells are regularly organized to form a 5×5 grid layout. The widths of the BS-cell and RS-cell are 2000 and 400 meters, respectively. We also consider that a number of MSs are uniformly distributed in the BS-cell. Note that we adopt the uniform Random-Walk mobility model [16] to realize the MS movement behavior. Based on the handoff-rate information from this mobility model, our RS grouping algorithm generates the RS grouping layouts with determined activation set assignments as shown in Fig. 7 (among the 25 RS grouping layouts, only the layouts for the preferred group sizes 1-12 are demonstrated). The packet arrivals of each user at the BS follow a Poisson process with the mean inter-arrival time setting to 100 ms and the packet length is assumed to be exponentially distributed. Moreover, we equally divide all the users into five classes which are offered with different downlink traffic loads. Except for the experiment of Fig. 9 (which investigates the effects of traffic loads and RS group sizes on system stability), the mean packet length of class i users is set to $i \times 1000$ bits (e.g., the mean packet lengths of class 1 and class 5 users are 1000 and 5000 bits, respectively).

A. Effects of RS group sizes on handoff frequency

Fig. 8 evaluates the handoff frequency of each grouping in Fig. 7 individually under different user densities (specifically, 100, 500 and 1000 users in our experiments). The results indicate that using the greedy grouping policy of our RS grouping algorithm, the handoff frequencies of the groupings gradually decline regardless of the user density as the group size increases. The main reason is that larger group sizes generally lead to lower handoff probabilities. It could be concluded that it is reasonable to choose larger group sizes if higher user mobility speeds are observed.

B. Effects of RS group sizes on system stability

Fig. 9 shows the mean packet delay under the grouping scenarios with group sizes 2, 6 and 10 as illustrated in Fig. 7. The delay performance is expressed as a function of the average packet traffic load to 100 MSs. For this experiment, we apply the TF policy during the scheduling procedure. Under the grouping scenario with the largest group size (i.e., group size of 10), the system becomes unstable (i.e., the users' packet delays go to infinity) when the average packet traffic load is larger than 1000 Kbps. Nevertheless, decreasing the group size can gradually extend the stable region. For example, in the grouping scenarios with group sizes 6 and 2, the stable regions are enhanced to 2000 and 4000 Kbps, respectively. These results are primarily due to the spatial diversity gain discrepancies under different grouping scenarios. Specifically, the groupings with smaller group sizes may have more opportunities to perform simultaneous transmissions, which means the system capacities can thus be reasonably enlarged. Note that smaller group sizes may, however, lead to more handoff events.

C. Performance analysis of different scheduling policies

Based on the groupings illustrated in Fig. 7, we estimate the throughput and delay performances of the TF, DF and Round-Robin centralized scheduling policies as functions of the group size. Figs. 10 and 11 present the average throughput and delay of the network under the situations with fixed and mobile users, respectively. Note that the average throughput is estimated as the average amount of packets actually received by the users within the simulation time. Moreover, the packet delay indicates the time difference between when a packet arrives at the BS and when it is received by the corresponding user.

Figs. 10(a) and 11(a) show the average network throughput under the three scheduling policies. For the case where users are stationary, the TF and DF policies

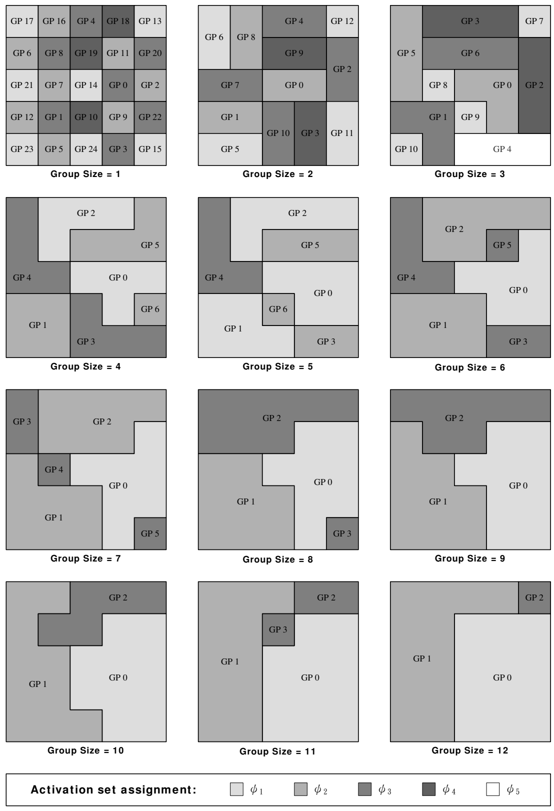


Fig. 7. Generated grouping results with activation set assignments

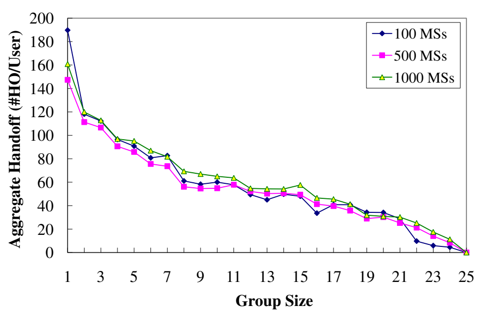


Fig. 8. Handoff comparison between different group sizes and different user densities

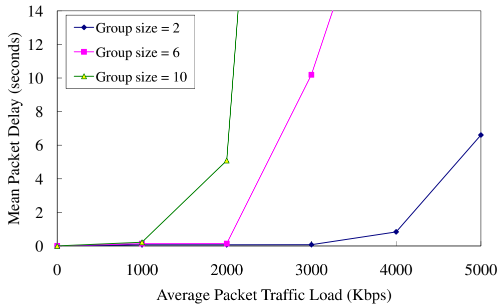


Fig. 9. Mean packet delay vs. average packet traffic load under the groupings with group sizes 2, 6 and 10

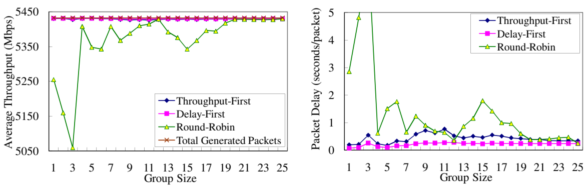


Fig. 10. Performance comparison between different group sizes in terms of (a) throughput and (b) delay without user mobility

achieve similar throughput performance. However, for the case with user mobility, the average throughput under the DF policy is higher than those under the other two scheduling policies. This phenomenon implies that the mobility behavior of the MSs is a major factor affecting the throughput performance under the TF policy, whereas the DF policy can provide relatively stable throughput regardless of the user mobility effect. Specifically, since the DF policy is able to balance the queue length of each user while the TF policy only prioritizes the user with the longest queue length, the average buffered packet

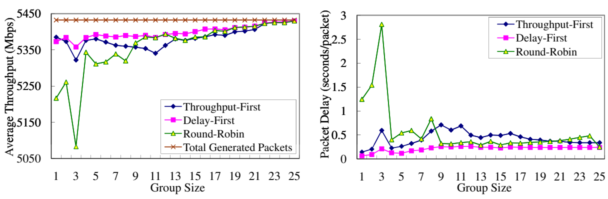


Fig. 11. Performance comparison between different group sizes in terms of (a) throughput and (b) delay with user mobility

amounts for all users under the DF policy are relatively small. Therefore, the number of dropped packets during the handoffs in DF can be reasonably reduced and the throughput can be further improved. The Round-Robin policy performs inefficiently in both cases since its scheduling decisions are made sequentially despite the system state information. In addition, we also observe that compared to the groupings with larger group sizes, the groupings with smaller group sizes result in better throughput performance for the case of fixed users. However, when user mobility is considered, the throughput value increases as the group size increases. This is because when users are stationary, the throughput performance is mainly dominated by the spatial diversity gain. Under such circumstances, smaller RS group sizes lead to higher throughput values. On the other hand, when the user mobility speed is high enough, the larger packet loss rates due to more frequent handoffs of smaller group sizes have more significant impact on the throughput performance. Consequently, larger RS group sizes are more advantageous.

Figs. 10(b) and 11(b) show the packet delay performance for the two scenarios without and with user mobility. We notice that the groupings with larger group sizes result in higher average packet delay. This is owing to the lower spectrum reuse of larger group sizes, and the packets in the system may suffer from longer queueing delay. Moreover, as expected, the DF policy can provide the lowest average delay for all group size cases compared to the TF and Round-Robin policies. The TF policy causes higher average delay since it considers only the queue length states but no packet waiting time information. The Round-Robin policy, which depends neither on the queue length nor on the waiting time information, incurs the worst average delay performance through almost all group size cases.

D. Effects of RS grouping patterns

In Figs. 10 and 11, we also observe that the performance under the case of group size 3 is worse than those under the other cases. Two primary reasons for this phenomenon are given below. First, the activation sets used in this case are more than those used in all the other cases (see Fig. 7). As discussed in subsection IV-C, the more the number of activation sets, the less the spatial diversity gain. Therefore, poorer throughput and delay performance results are expected for group size 3. Second, since the group size in this case is small, the packet losses from handoff events would also influence

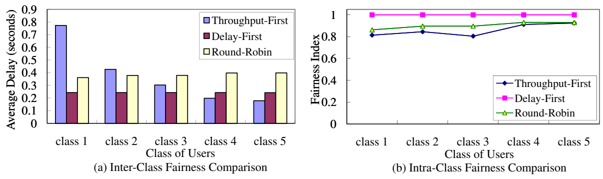


Fig. 12. Fairness comparison in terms of packet delay

the throughput performance. Under the joint impacts of low spatial diversity gain and high packet loss rate, the worst performances are observed for the case of group size 3. This special case implies that if the RS grouping patterns can not be appropriately determined, the system performance would be significantly affected. The simulation results shown in Figs. 10 and 11 demonstrate that our proposed RS grouping algorithm can derive suitable RS grouping patterns in most cases.

E. Fairness analysis

The fairness properties of the TF, DF and Round-Robin scheduling policies are illustrated in Fig. 12, where the 100 users are classified into five classes. The scenario with group size 20 is considered in the experiments. We use average delay for the inter-class fairness comparison. On the other hand, for the intra-class fairness comparison, we use the per user packet delay to compute the Jain Fairness Index [24] as follows

$$Fairness\,Index = \frac{(\sum_1^n x_i)^2}{n\sum_1^n x_i^2}, \qquad (15)$$

where x_i is the average packet delay of user i. A scheduling policy achieves the optimal fairness if this index is 1 and is completely unfair if it is $\frac{1}{n}$.

It can be observed that fairness among users is not satisfied under the TF scheduling policy. Since the high-load users have the high service priority, the average delays of these users are significantly low. However, the users who have lighter offered loads tend to suffer from higher packet delay. On the other hand, both the DF and Round-Robin scheduling policies balance each user's average delay, which implies that fairness among users (no matter inter-class or intra-class) is guaranteed in terms of the average delay under these two policies. Moreover, the DF scheduling policy provides lower average delay than the Round-Robin policy for all the five classes of users because the maximum mean waiting time of all users' queues is minimized under the DF policy. These observations lead to the conclusion that our DF scheduling policy can provide not only the minimization of the average delay, but also the fairness property among different traffic-load users.

VII. Related Work

In the literature, registration area (RA) grouping schemes and radio resource scheduling policies have been intensively investigated for mobile telecommunications networks and for MR networks, respectively. However, we note that in these studies, either scheduling issues were not addressed in the RA grouping schemes, or grouping concepts were not considered in the MR scheduling policies. Grouping mechanisms and scheduling algorithms have never been jointly discussed. In the following, we summarize some representative RA grouping schemes and MR scheduling policies separately.

- **RA grouping schemes:** In mobile telecommunications networks, several cells constitute an RA. How to group cells to form RAs is an important cellular network planning issue. In [12], Xie et al. first proposed a per-user adaptive RA scheme. By taking paging rates, location-update rates and the respective costs into account, the proposed scheme decides the appropriate size of an RA to minimize the overall cost. Then, Castelluccia addressed the same grouping problem in [2] by considering a more general network configuration that supports overlapping RAs. Grouping users into different velocity classes according to their velocities upon location update instants, the scheme in [10] reduces the paging cost by properly decreasing RA sizes. To identify the optimal RA size and to cope with the location tracking problem, a dual group system was proposed in [3]. The dual group method enhances the utilization of network resources because it eliminates unnecessary location updates due to its fewer group boundary passes. Dynamically adaptive grouping schemes were also discussed in [8], [9]. In [8], the proposed scheme dynamically adjusts overlapping RAs according to user mobility and call patterns. In this work, the effects of cell inclusions/exclusions into/from an RA were analytically characterized. By periodically determining the proper number of cells within an RA, the proposed scheme greatly reduces the average signaling cost and database access load on location registers. In [9], the authors further demonstrated how the grouping of an RA affects the number of subsequent registrations in the overlapping RA environments.

- **MR scheduling policies:** Various scheduling policies have been proposed to achieve different performance objectives, such as throughput maximization [11], [29], [22], delay minimization [23], proportional fairness [26], [27], [20] and QoS guarantee [4], [5], [21], [17]. According to specific objectives, the resource scheduling problem can be formulated to different optimization problems. For instance, the scheduling problem was modeled as a link selection problem in [11] to maximize the signal to interference plus noise ratio (SINR) of the links and to prevent all the user queues from exploding, a boundary selection problem in [26] to balance the traffic loads among relays, a graph multicoloring problem in [29] to maximize the sizes of non-adjacent-relay sets, a user required-rate problem in [21] to satisfy each user's required rate, a network flow problem in [23] to select an appropriate transmission order, and an interference minimization problem in [17] to find out a less-interference path for each connection. To resolve these problems in polynomial time, heuristic scheduling policies were proposed in the respective researches.

VIII. CONCLUSIONS

The IEEE 802.16j MR standard has been developed to provide performance enhancement to the existing IEEE 802.16e network. However, issues such as frequent handoffs and low spectrum utilization which were not encountered in IEEE 802.16e may occur in IEEE 802.16j. The RS grouping is one optional mechanism in the IEEE 802.16j MR standard to overcome these problems. This paper investigated the RS grouping performance enhancement in terms of throughput and handoff frequency. An RS grouping algorithm was designed utilizing a greedy grouping policy: RS pairs with higher handoff rates will have higher priority to be selected. The simulation results showed that the handoff frequency of the considered MR network can be significantly reduced and suitable RS grouping patterns with determined activation set assignments can be derived using our grouping algorithm. In addition, we proposed the TF and DF centralized scheduling policies to maximize the system throughput and to minimize the average packet delay, respectively. By integrating our RS grouping algorithm and centralized scheduling algorithms, the simulation results indicated that for the case of fixed users, the groupings with smaller group sizes can result in better throughput performance. However, when user mobility is considered, the throughput value increases as the group size increases. Furthermore, we also showed that the DF policy can not only minimize the average packet delay, but also provide fairness among different traffic-load users.

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers. Their valuable comments have significantly enhanced the quality of this paper.

REFERENCES

[1] Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands. *IEEE Computer Society and the IEEE Microwave Theory and Techniques Society*, February 2006.

[2] C. Castelluccia. Extending Mobile IP with Adaptive Individual Paging: a Performance Analysis. *IEEE Symposium on Computers and Communications (ISCC)*, pages 113–118, 2000.

[3] C.-M. Weng and P.-W. Huang. Modified Group Method for Mobility Management. *Computer Communications*, 23:115–122, 2000.

[4] C.-Y. Hong and A.-C. Pang. Link Scheduling with QoS Guarantee for Wireless Relay Networks. *IEEE 28th International Conference on Computer Communications, INFOCOM 2009*, pages 2806–2810, April 2009.

[5] D. Ghosh, A. Gupta, and P. Mohapatra. Admission Control and Interference-Aware Scheduling in Multi-hop WiMAX Networks. *IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems, MASS 2007*, pages 1–9, Oct. 2007.

[6] D. J. A. Welsh and M. B. Powell. An Upper Bound for the Chromatic Number of a Graph and Its Application to Timetabling Problems. *The Computer Journal*, 10(1):85–86, 1967.

[7] D. Schultz, R. Pabst, and T. Irnich. Multi-Hop Based Radio Network Deployment for Efficient Broadband Radio Coverage. *Proc. WPMC*, Yokosuka, Japan, 2003.

[8] G. Varsamopoulos and S.K.S. Gupta. Dynamically Adapting Registration Areas to User Mobility and Call Patterns for Efficient Location Management in PCS Networks. *IEEE/ACM Transactions on Networking*, 12(5): 837–850, Oct. 2004.

[9] G. Varsamopoulos and S.K.S. Gupta. Optimal Offline and Online Registration Techniques for Location Management with Overlapping Registration Areas. *IEEE Transactions on Mobile Computing*, 4(5): 474–488, Sept.-Oct. 2005.

[10] G. Wan and E. Lin. A Dynamic Paging Scheme for Wireless Communication Systems. *ACM/IEEE MobiCom97*, pages 195–203, 1997.

[11] H. Viswanathan and S. Mukherjee. Performance of Cellular Networks with Relays and Centralized Scheduling. *IEEE Trans. Wireless Commun.*, 4(5): 2318–2328, Sept. 2005.

[12] H. Xie, S. Tabbane, and D.J. Goodman. Dynamic Location Area Management and Performance Analysis. *IEEE Vehicular Technology Conference*, pages 536–539, May 1993.

[13] IEEE 802.16j-06/172. Ranging Process for IEEE 802.16j. November 2006.

[14] IEEE 802.16j-2009 Standard. Part 16: Air Interface for Broadband Wireless Access Systems Amendment 1: Multiple Relay Specification. December 2009.

[15] J. Laneman, D. Tse, and G. Wornell. Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior. *IEEE Trans. Inform. Theory, vol. 50, no. 12*, pages 3062–3080, April 2004.

[16] J. Nuevo. Mobility Generator for NS-2. 2002. http://www.math.keio.ac.jp/matumoto/emt.html.

[17] J. Tang, G. Xue, and C. Chandler. Interference-Aware Routing and Bandwidth Allocation for QoS Provisioning in Multihop Wireless Networks. *Wireless Communications and Mobile Computing Journal*, 5(8):933–943, 2005.

[18] L. Kleinrock. *Queueing Systems*, volume 1. John Wiley and Sons, 1975.

[19] L. Kleinrock and J. Silvester. Spatial Reuse in Multihop Packet Radio Networks. *Proceedings of the IEEE*, 75(1):156–167, January 1987.

[20] L. Wang, Y. Ji, and F. Liu. A Novel Centralized Resource Scheduling Scheme in OFDMA-Based Two-Hop Relay-Enhanced Cellular Systems. *IEEE International Conference on Wireless and Mobile Computing, WIMOB 2008*, pages 113–118, Oct. 2008.

[21] M. K. Awad and X. Shen. OFDMA Based Two-Hop Cooperative Relay Network Resources Allocation. *IEEE International Conference on Communications, ICC 2008*, pages 4414–4418, May 2008.

[22] M. Yang and P. H. J. Chong. Time Slot Allocation Schemes for Multihop TDD-CDMA Cellular System. *IEEE Wireless Communications and Networking Conference, WCNC 2007*, pages 3099–3104, March 2007.

[23] P. Djukic and S. Valaee. Link Scheduling for Minimum Delay in Spatial Re-Use TDMA. *IEEE 26th International Conference on Computer Communications, INFOCOM 2007*, pages 28–36, May 2007.

[24] R. Jain, W. Hawe, and D. Chiu. A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems. *Technical Report DEC-TR-301*, Sept. 1984.

[25] R. Pabst, B.H. Walke, and D.C. Schultz. Relay-based Deployment Concepts for Wireless and Mobile Broadband Radio. *IEEE Communications Magazine*, pages 80–89, September 2004.

[26] T.-W. Kim, T.-Y. Min, and C.-G. Kang. Opportunistic Packet Scheduling Algorithm for Load Balancing in a Multi-Hop Relay-Enhanced Cellular OFDMA-TDD System. *14th Asia-Pacific Conference on Communications, APCC 2008*, pages 1–5, Oct. 2008.

[27] T. Wu, G. Li, Y. Wang, J. Huang, X. Yu, and H. Tian. Fairness-Oriented Scheduling with Equilibrium for Multihop Relaying Networks Based on OFDMA. *IEEE 68th Vehicular Technology Conference, VTC 2008-Fall*, pages 1–5, Sept. 2008.

[28] W.-C. Kan and S.-R. Yang. The Supplements to "Handoff Minimization through a Relay Station Grouping Algorithm with Efficient Radio Resource Scheduling Policies for IEEE 802.16j Multihop Relay Networks". *Technical Report, Nat'l Tsing Hua Univ.*, 2008. http://www.cs.nthu.edu.tw/∼sryang/submission/TechHMRSGA.pdf.

[29] X. Guo, W. Ma, Z. Guo, X. Shen, and Z. Hou. Adaptive Resource Reuse Scheduling for Multihop Relay Wireless Network Based on Multicoloring. *IEEE Communications Letters*, 12(3):176–178, March 2008.

PLACE PHOTO HERE

Shun-Ren Yang (M'08) received the BSCSIE, MSCSIE, and Ph.D. degrees from National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1998, 1999, and 2004, respectively.

From April 1, 2004 to July 31, 2004, he was appointed as a Research Assistant in the Department of Information Engineering, the Chinese University of Hong Kong. Since August 2004, he has been with the Department of Computer Science and Institute of Communications Engineering, National Tsing Hua University, Taiwan, where he is an Assistant Professor. His current research interests include design and analysis of personal communications services networks, computer telephony integration, mobile computing, and performance modeling.

PLACE PHOTO HERE

Chien-Chi Kao (S'09) received the BSCSIE degree from National Chung Cheng University, Chiayi, Taiwan, in 2006, and the MSCOM degree from National Tsing Hua University, Hsinchu, Taiwan, in 2008. He is now a Ph.D. candidate in the Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan.

In 2006, he joined the Wireless and Mobile Networking Laboratory, National Tsing Hua University, Hsinchu, Taiwan. He is a student member of the IEEE. His current research interests include design and analysis of mobile telecommunications networks and performance modeling.

PLACE PHOTO HERE

Wai-Chi Kan received the BSCSIE degree from National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 2006, and the MSCS degree from National Tsing Hua University, Hsinchu, Taiwan, R.O.C., in 2008.

In 2006, he joined the Wireless and Mobile Networking Laboratory, National Tsing Hua University, Hsinchu, Taiwan, R.O.C. Since 2008, he has been with the Synology Incorporated, Taipei, Taiwan, R.O.C. His current research interests include design and analysis of mobile telecommunications networks and performance modeling.

PLACE PHOTO HERE

Tzung-Chin Shih received the BSCOM degree from National Chiao Tung University, Hsinchu, Taiwan, in 2008. He is now a master student in the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan.

In 2008, he joined the Wireless and Mobile Networking Laboratory, National Tsing Hua University, Hsinchu, Taiwan. His current research interests include design and analysis of mobile telecommunications networks and performance modeling.

Analyzing VoIP Capacity with Delay Guarantee for Integrated HSPA Networks

Shin-Hua Yang, Shun-Ren Yang, and Chien-Chi Kao

Dept. of CS & Inst. of COM, NTHU, Hsinchu, Taiwan, R.O.C.
{kim@wmnet,sryang@,mickey@wmnet.}cs.nthu.edu.tw

Abstract. Voice over IP (VoIP) is a key driver in the evolution of voice communications, and the high transmission rate property of High Speed Packet Access (HSPA) is expected to satisfy the strict delay requirements of VoIP. Therefore, the aim of this paper is to evaluate the performance of VoIP service in HSPA network. This paper presents a mathematical model for VoIP capacity in HSPA under the constraints of delay threshold and voice quality requirements. This study also analyzes the impact of scheduling schemes, the user's channel quality, and variations in packet bundle size on VoIP performance. These results are derived from simulation results, which also validate the correctness of the proposed analysis model and show that VoIP performance is limited by uplink transmission technology. Based on the E-model, this study concentrates on each VoIP connection's quality in HSUPA network.

Keywords: Delay, E-model, HSPA, VoIP capacity.

1 Introduction

Voice over IP (VoIP) is becoming a key driver in the evolution of voice communications. Compared to traditional Circuit Switched (CS) voice networks, the main advantages of VoIP are reduced operating costs and improved user flexibility. However, due to the restrictions of spectral efficiency and delay requirements, VoIP cannot exceed CS voice performance. Hence, 3GPP Release 5 and Release 6 introduced High Speed Downlink Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA), respectively, to provide a higher transmission rate and improve system capacity. High Speed Packet Access (HSPA) is a collection of two mobile telephony protocols HSDPA and HSUPA. The evolution of HSPA was originally designed to increase data transmission rates and achieve higher capacities for high-performance applications based on innovative techniques such as shortened Transmission Time Interval (TTI) and fast packet scheduling controlled by Node B. HSPA increases the uplink and downlink peak transmission rate up to 5.76 Mbit/s and 14.4 Mbit/s, respectively. For VoIP, the end-to-end delay is the most important factor determining connection quality. The high transmission rate of HSPA should satisfy the strict delay requirements of VoIP. The maximum number of VoIP connections that HSPA systems can support is a critical performance indicator of significant interest to mobile network operators.

D. Ślęzak et al. (Eds.): FGCN/ACN 2009, CCIS 56, pp. 316–323, 2009.

The evolution of VoIP has resulted in numerous capacity analysis models for wireless networks. To maximize the resource utilization, researches in [2], [3], and [8] presented capacity analysis models that concentrate on resource allocation for each VoIP connection. Optimal resource utilization leads to the greatest VoIP performance. Nevertheless, this type of capacity analysis model is not able to predict the maximum number of supported VoIP connections because it cannot guarantee voice quality or delay. Considering voice quality, [1], [4], and [6] present VoIP capacity analysis models for HSPA. These analytical studies provide a suitable approximation for VoIP capacity, and ensure satisfactory connection quality for each user. However, this analytical model type does not discuss the effects of delay threshold on VoIP performance. Taking delay restrictions and the voice quality certification into account, this study builds a mathematical model for analyzing VoIP capacity for both HSDPA and HSUPA. This approach employs a queueing model to approximate the VoIP capacity under delay threshold which includes the packet waiting time and the packet transmission time between UE and Node B. For uplink transmission, the interference caused by other UEs affects the voice quality heavily. Hence, we take the connection quality into account when deriving the maximum number of VoIP connections. Thus, the proposed analysis model defines VoIP capacity with the guarantee of transmission time and voice quality. Results show the improved VoIP performance in an HSPA system and then the bottleneck of VoIP capacity will be HSUPA.

2 Capacity Analysis

The high transmission rate property of HSPA is expected to satisfy strict delay requirements of real time services. In this section, we build a mathematical model for analyzing the VoIP capacity when both HSDPA and HSUPA are developed. The end-to-end packet transmission delay, T_{end}, from a sending UE in BTS_1 to its receiving UE in BTS_2 consists of three parts: T_{up}, T_{cn}, and T_{down}. T_{cn} is the packet transmission time between BTS_1 and BTS_2 via the core network. Compared with the wireless HSPA transmissions of lower bit rate and higher error rate, the core network provides relatively sufficient and stable bandwidth support. In this case, T_{cn} is more steady than T_{up} and T_{down}. Therefore, we neglect the effects of T_{cn} on T_{end} and simply concentrate on T_{up} and T_{down} to derive the maximum number of VoIP connections. Let N_c be the VoIP capacity of the HSPA systems, i.e., the maximum number of supported VoIP sessions under the constraint that the average end-to-end delay of voice packets comply with the QoS requirements. Under the aforementioned assumptions, we first calculate T_{up} and T_{down}. Based on the obtained T_{up} and T_{down} values, we next derive the maximum number of VoIP connections N_{up} and N_{down} in HSUPA and HSDPA, respectively. Because of the symmetry property of voice connections, the maximum number of voice connections N_c in HSPA system is expressed as

$$N_c = Min\{N_{down}, N_{up}\}. \tag{1}$$

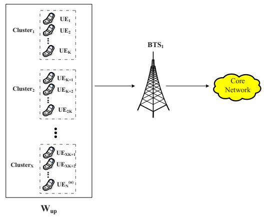


Fig. 1. Packet transmission process in the HSUPA system

2.1 Calculation of T_{up}

Every HSUPA TTI, BTS_1 schedules UEs and the scheduled UEs can transmit voice packets buffered in their corresponding queues to BTS_1. We derive the uplink transmission delay, T_{up} as follows. T_{up} is composed of two parts: the time period W_{up} during which a voice packet waits to be served in its corresponding UE queue and the transmission time T_{utx} of this uplink packet, when scheduled, from its UE to BTS_1. To simplify our analysis, Round Robin (RR) scheduling scheme is applied for HSUPA in BTS_1. Fig. 1 depicts the detailed packet transmission process at BTS_1 with the $N^{(u)}$ number of VoIP connections. In HSUPA, several UEs can transmit packets simultaneously during the same TTI. Accordingly, our model assumes that UEs under BTS_1 are grouped into X clusters, where each cluster can maximum consist of K in-service VoIP UEs. The voice packet generation process of a VoIP connection within each cluster is assumed to follow a Poisson process with arrival rate λ, whose value depends on the adopted voice codec. In this analysis, we model the BTS_1 HSUPA transmission process as a polling queueing system [5] with certain modifications. To apply the result of the polling queueing model, we assume that in the HSUPA transmission process, the UE queues within one cluster are virtually merged as a single queue. The incoming packets of each UE in $Cluster_1$ are placed in the virtual queue and transmitted to BTS_1 sequentially. We also assume that in our polling model, only one packet instead of a bundled packet in the virtual queue can be served upon every polling visit. However, the mean service time of our polling model should be modified. Let $\bar{N}_{pdu}^{(u)}$ represent the mean number of served bundled packets during a TTI for the original HSUPA transmission process. That is, the total number of transmitted packets during one TTI is $\bar{N}_{pdu}^{(u)} * N_b$. Then, under the constraint that $E[T_{up}]$ should be less than a pre-defined transmission delay budget during T_{delay} for a transmission between a UE and BTS_1, the maximum number of voice connections N_{up} in HSUPA can be derived by the following equation:

$$N_{up} = Max\{N^{(u)} : \frac{TTI^{(u)}}{\bar{N}_{pdu}^{(u)} * N_b} + \frac{N^{(u)}\lambda E[T_{utx}^2]}{2(1-\rho^{(u)})} < T_{delay}\}. \quad (2)$$

2.2 Calculation of T_{down}

On receiving a packet, BTS_2 distributes this packet to its corresponding waiting queue. Within every TTI, BTS_2 schedules and transmits packets in the waiting queues to multiple receiving UEs using HSDPA technology. We next derive the transmission delay, T_{down}. T_{down} is also composed of two parts. The first part is W_{down}, the time period during which a voice packet waits to be served in BTS_2. The second part is T_{dtx}, the transmission time of a scheduled downlink packet from BTS_2 to the destined UE. For the tractability of our analysis model, the RR scheduling policy is adopted for HSDPA in BTS_2 as well. We assume that the packet arrivals of each VoIP connection to BTS_2 follow a Poisson process with rate, λ. Similarly, our we analysis models the BTS_2 HSDPA transmission process as a polling queueing system. Then, under the constraint that $E[T_{down}]$ should be less than a pre-defined transmission delay budget during T_{delay} for a transmission between a UE and BTS_2, the maximum number of voice connections N_{down} in HSDPA can be derived by the following equation:

$$N_{down} = Max\{N^{(d)} : \frac{TTI^{(d)}}{\bar{N}^{(d)}_{pdu}} + \frac{N^{(d)}\lambda E[T^2_{dtx}]}{2(1-\rho^{(d)})} < T_{delay}\}. \tag{3}$$

Due to the space limitation, the details of the calculations are provided in our technical report [7].

3 Performance Evaluation

3.1 Simulation Validation

The analytical models have been validated by the simulation assumptions. Based on the system model for HSDPA and the simulation parameters shown in Table 1, Table 2 shows the analytical and simulation results. For VoIP over HSDPA, the simulations yield capabilities of 41 and 51 simultaneous connections for delay budgets 80 and 150 ms, respectively. Clearly, the simulation results match the analytical results well. For HSUPA, we take packet bundling into consideration. By applying the detailed simulation settings shown in Table 1 into our event-driven simulation model, we validate the correctness of our uplink analysis model. Table 3 compares the simulation results with those of our HSUPA polling model

Table 1. Simulation Parameters For HSDPA/HSUPA System

Parameter for HSDPA	Value	Parameter for HSUPA	Value
TTI size	2ms	TTI size	10ms
Downlink data rate	3.6Mbps	Uplink data rate	5.76Mbps
VoIP codes type	GSM6.10	VoIP codes type	GSM6.10
UE profile	Pedestrian A	UE Maximum Transmit power	21dBm
UE speed	3kmph	$(\beta_{ec}/\beta_c)^2$	3dB
UE receiver structure	Single-antenna Rake	$(\beta_{ed}/\beta_c)^2$	8dB

Table 2. Comparison Between The Analytical And Simulation Results On HSDPA

	T_{delay}=80 ms	T_{delay}=150 ms
N_{down}(Analytical)	39	49
N_{down}(Simulation)	41	51

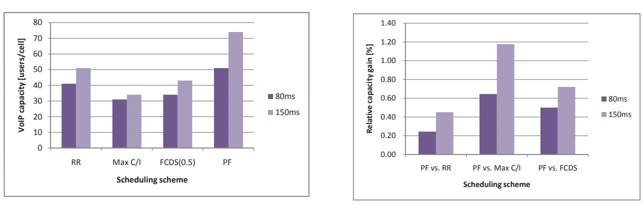
Table 3. Comparison Between The Analytical And Simulation Results On HSUPA

$T_{delay} = 80ms$	N_b=1	N_b=2	N_b=3	$T_{delay} = 150ms$	N_b=1	N_b=2	N_b=3
N_{up}(Analytical)	63	80	53	N_{up}(Analytical)	103	127	94
N_{up}(Simulation)	62	77	51	N_{up}(Simulation)	99	123	90

under different delay budgets 80 ms and 150 ms and different packet bundle sizes. It is clear that the analytical analysis is consistent with the simulation results.

3.2 Simulation Results and Analysis

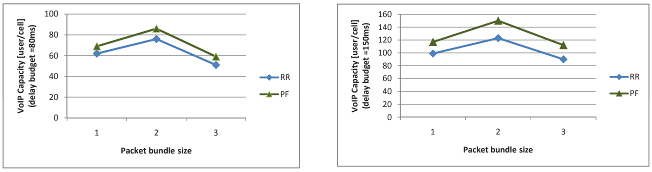
Based on the HSDPA and HSUPA simulation models validated against the analytic analysis, we design different simulation scenarios to study and compare the VoIP performance over HSDPA and HSUPA technologies. First, we concentrate on HSDPA. For HSDPA, we have provided the analytical and simulation VoIP capacity results based on RR scheduling in the last subsection. When other factors such as channel quality and throughput are taken into account for scheduling, the VoIP capacity may be influenced by the adopted scheduling algorithms. Fig. 2 shows the VoIP capacity performance over HSDPA. Fig. 2(a) represents the VoIP capacities when applying RR, Max C/I, FCDS and PF scheduling schemes under the delay budgets 80 and 150 ms. The specific relative capacity gains from each scheduling scheme compared to PF are shown in Fig. 2(b). PF provides better VoIP capacity performance than RR because it is on average able to schedule users at better channel conditions. Compared to Max C/I, PF is more fair because of taking UE's average throughput into consideration. The results verify that packet scheduling plays a key role in achieving good



(a) VoIP cell capacities with different scheduler.

(b) Relative capacity gain of PF over the other algorithms.

Fig. 2. VoIP capacities (a) different scheduling schemes (b) relative capacity gain



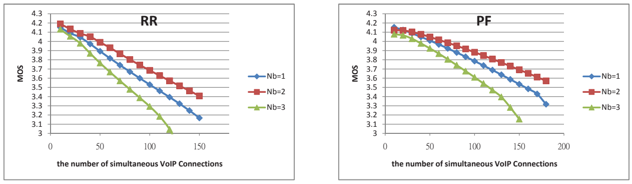
(a) Delay Budget = 80 ms (b) Delay Budget = 150 ms

Fig. 3. Number of Supported VoIP Connections on (a) Delay Budget = 80 ms, (b) Delay Budget = 150 ms

VoIP capacity in HSDPA. From Fig. 2(a), we also note that the delay budget is another factor that affects the VoIP capacity. The longer the delay budget, the more the number of supported voice connections.

After discussing the VoIP capacity performance over HSDPA, we study the VoIP capacity over HSUPA. Considering RR, we have already provided the analytical and simulation VoIP capacity results over HSUPA. We also take the PF scheduling algorithm into account for observing the HSUPA VoIP performance. Fig. 3 shows the VoIP capacity applying the RR and PF scheduling algorithms with different packet bundle sizes under 80 and 150 ms delay budgets. From Fig. 3(a) and Fig. 3(b), PF scheduling provides better VoIP capacity than RR scheduling regardless of the delay budget or the packet bundle size. Moreover, the packet bundle size is the other factor that influences VoIP performance. Packet bundle size two provides the better capacity performance than that without packet bundling. The phenomenon is explained as follows. The bundled packet simply contains one header and then the header overhead will be shared among the two packets. Hence, packet bundling decreases the header overhead and provides better bandwidth efficiency. We also observe that applying packet bundling should waste time on waiting enough packets to be bundled and transmitted. However, when increasing the bundle size from two to three, the VoIP capacity decreases. This is because the more the packet bundle size, the longer the waiting time for bundling packets. Hence, the VoIP capacity will decrease when the packet bundling delay is intolerable. From Fig. 3, we know that there is a trade-off between the packet bundle size and the improvement of VoIP capacity. As a result, bundling two packets into one packet is the best option to enhance the VoIP performance over HSUPA in our simulation parameter settings.

After discussing the VoIP capacity over HSDPA and HSUPA, we know that the VoIP performance is influenced by the adopted scheduling scheme, the delay budget, and the packet bundle size. In the above experiments, we obtain the capacity based on the delay budget, one of the QoS requirements. From our simulation results, under the 80 ms delay budget, we obtain that the optimal capacity achieved with PF scheduling over HSDPA is 51 users. Due to the restrictions on the UE downlink data reception transmission rate, the peak data rate in our NS2 simulation scenarios is 3.6 Mbit/s while the ideal transmission rate defined in UMTS Release 5 is 14.4 Mbit/s. If the downlink transmission rate can be



(a) RR algorithm. (b) PF algorithm.

Fig. 4. The number of simultaneous VoIP connections under the different packet bundle sizes and the corresponding MOS values when applying (a) RR (b) PF

increased up to 14.4 Mbit/s, fourfold number of voice packets can be transmitted during each TTI. Hence, the VoIP capacity over HSDPA may be improved at least two or three times and then the maximum supported number of VoIP connections may be 102 or 153. Similarly, under the 80 ms delay budget and applying the PF scheduling, the optimal capacity achieved with packet bundling size two under HSUPA is 86 users. Thus, when both uplink and downlink packets can be transmitted with the ideal peak date rate, the VoIP performance is bounded by the uplink transmission performance. In the following, we focus on the VoIP performance under HSUPA and obtain the VoIP capacity according to E-model. Fig. 4 shows the number of simultaneous VoIP connections under the different packet bundle sizes and the corresponding MOS values. Clearly, the more restricted the demanding connection quality, the less the number of supported VoIP connections. Compared to the three curves in Fig. 4(a), those in Fig. 4(b) degrade less sharply. This means when applying PF scheduling, the increase in the number of simultaneous VoIP connections leads to less decrease in connection quality. Hence, PF scheduling algorithm provides better VoIP performance than RR scheduling algorithm. Based on E-model, we know that the connection quality of our analytical VoIP capacity is above 3.6 MOS value regardless of the adopted scheduling scheme and packet bundle size. This means that our analysis model approximates the VoIP capacity with guaranteed connection quality with which most of the users satisfy.

4 Conclusion

This paper focused on the VoIP performance in HSPA networks and builded a mathematical VoIP capacity analysis model to obtain the maximum number of supported VoIP connections that still meet transmission delay and voice quality requirements. Using simulation results, this study also validated the correctness of the proposed analytical method and discussed other factors affecting VoIP performance. Experimental results indicated that PF is the most appropriate algorithm for VoIP service, even in a mixed traffic environment or when the UE has poor channel quality. For packet bundling, there is a tradeoff between the packet bundle size and the bundle delay. In the simulated scenario in this study, a packet bundle size of two causes less bundle delay and decreases packet overhead.

Hence, bundling two packets together improves VoIP capacity tremendously. Our precise analytical model showed that VoIP performance is bounded by uplink VoIP performance. Following to the E-model, this study also examined the effects of other QoS requirements on uplink VoIP capacity and verified that our analysis model approximates the VoIP capacity with guaranteed connection quality with which most of the users satisfy.

Acknowledgment

We would like to thank Chunghwa Telecom. This work was accomplished under close discussions with the researchers of Chunghwa Telecom. This work was supported in part by the National Science Council (NSC) of Taiwan under Contracts 96-2752-E-007-003-PAE, 96-2221-E-007-025-, 96-2221-E-007-027-, 96-2219-E-007-012- and 96-2219-E-007-011-, and Chunghwa Telecom.

References

1. Chen, T., Kuusela, M., Malkamaki, E.: Uplink capacity of voip on hsupa. In: IEEE 63rd Vehicular Technology Conference, 2006. VTC 2006-Spring, May 2006, vol. 1, pp. 451–455 (2006)
2. Lee, H., Kwon, T., Cho, D.-H.: An efficient uplink scheduling algorithm for voip services. IEEE 802.16 bwa systems 5 (September 2004)
3. Lee, H., Kwon, T., Cho, D.-H.: Extended-rtps algorithm for voip services. IEEE 802.16 systems 5, 2060–2065 (2006)
4. Qualcomm White paper. Air interface cell capacity of wcdma systems (May 2007)
5. Takagi, H.: Queuing analysis of polling models. ACM Comput. Surv. (1988)
6. Wanstedt, S., Ericson, M., Hevizi, L., Pettersson, J., Barta, J.: The effect of f-dpch on voip over hsdpa capacity. In: IEEE 63rd Vehicular Technology Conference, 2006. VTC 2006-Spring, May 2006, vol. 1, pp. 410–414 (2006)
7. Yang, S.-R., Yang, S.-H.: Analyzing voip capacity with delay guarantee for integrated hspa networks. Technical Report, Natl. Tsing Hua Univ. (2009)
8. Zhao, D., Shen, X.: Performance of packet voice transmission using IEEE 802.16 protocol. IEEE Wireless Communications 14(1), 44–51 (2007)