MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
IEEE COMMUNICATIONS SOCIETY
http://www.comsoc.org/~mmc

E-LETTER

Vol. 10, No. 1, January 2015

IEEE COMMUNICATIONS SOCIETY

CONTENTS

Message from MMTC Chair

Dear Fellow MMTC members,

It is really a great honor for me to serve as the Europe Vice-Chair for this ComSoc Technical Committee during 2014-2016. During the past few months, we have been working towards the setup of the new leadership teams for our fourteen Interest Groups (IGs) and we have started working at the promotion of journal special issues, as well as symposia and workshops within main conferences, together with the new IG leaders. Our IGs cover most of the key and emerging technical fields of multimedia communications and represent the core of our networking and scientific activities. You are very welcome to select and get involved in one or more IG(s) by contacting the Chair(s).

I am very happy to introduce the new IG leadership teams to you:

(QoEIG) **QoE for Multimedia Communications**
Chair: Hantao Liu, University of Hull, UK
Co-chairs: Martin Varela, VTT Technical Research Centre of Finland, Finland
Lea Skorin-Kapov, University of Zagreb, Croatia
Advisor: Alan Bovik, The University of Texas at Austin

(VAIG) **Visual Analysis and Content Management for Communications**
Chair: Rongrong Ji, Columbia University, USA
Co-chair: Chenwei Deng, Beijing Institute of Technology, China
Advisor: Jiebo Luo, Rochester University

(MPCIG) **Media Processing for Communications**
Chair: Sanjeev Mehrotra, Microsoft, USA
Co-chairs: Ce Zhu, University of Electronic Science and Technology of China, China
Carl J. Debono, University of Malta, Malta
Advisor: Ming-Ting Sun, University of Washington, USA

(MSIG) **Media Streaming**
Chair: Di Wu, Sun Yat-Sen University, China
Co-chairs: Alexander Raake, T-Labs, TU Berlin
Advisors: Keith W. Ross, New York University, USA
Oliver Dapeng Wu, University of Florida, USA

(SMPCIG) **Security in Media Processing and Communications**
Chair: Deepa Kundur, University of Toronto, Canada
Co-chairs: Rose Hu, Utah State University, USA
Anthony Ho, University of Surrey, UK
Advisor: Jiangtao Wen, Tsinghua University

(3DIG) **3D Rendering, Processing and Communications**
Chair: Tasos Dagiuklas, Hellenic Open University, Greece
Co-chairs: Aljoscha Smolic, Disney Research Zurich, Switzerland
Wanqing Li, University of Wollongong, Australia
Advisor: A. Murat Tekalp, Koc University, Turkey

(MNIG) **Multimedia Networking**
Chair: Ping Wang, Nanyang Technological University, Singapore
Co-chairs: Xinlin Huang, Tong Ji University, China
Zhenzhong Chen, Wuhan University, China
Advisor: Weihua Zhuang, University of Waterloo, Canada

(GMCIG) **Green Multimedia Communications**
Chair: Zuqing Zhu, USTC, China
Co-chairs: Changqiao Xu, BUPT, China
Advisor: Moshe Zukerman, City University of Hong Kong

(DSNIG) **Distributed and Sensor Networks for Mobile Media Computing and Applications**
Chair: Tay Wee Peng, NTU, Singapore
Co-chairs: Pedro Amado Assuncao, Intitute of Telecom, Portugal
Advisor: Dapeng WU, University of Florida

(MENIG) **Multimedia Services and Applications over Emerging Networks**
Chair: Zhou Su, Waseda University, Japan
Co-chairs: Suhua Tang, The University of Electro-Communications, Japan
Autur Lugmayr, Tampere University of Technology (TUT), Finland
Advisor: Hsiao-Hwa Chen, National University Taiwan

(MCDIG) **Multimedia Content Distribution: Infrastructure and Algorithms**
Chair: Xiaoqing Zhu, Cisco, China
Co-chairs: Harilaos Koumaras, National Centre of Scientific Research "Demokritos", Greece
David Hausheer, Technische Universität Darmstadt, Darmstadt
Advisor: Prof. Chang Wen Chen, University at Buffalo, The State University of New York

(MBDIG) **Multimedia Big Data**
Chair: Xiaokang Yang, Shanghai Jiao Tong University, China
Co-chairs: Zhengjun Zha, NUS, Singapore
Burak Kantarci, Clarkson University, USA
Advisor: Wenwu Zhu,Tsinghua University Beijing

(MCCIG) **Multimedia Cloud Computing**
Chair: Kuan-Ta Chen, Academia Sinica, Taiwan
Co-chairs: Chin-Feng Lai, Academia Sinica, Taiwan
Ali Begen, Cisco, USA
Advisor: Baochun Li, University of Toronto, Canada

(MCSIG) **Multimedia Communication Systems**
Chair: Yung-Hsiang Lu, Purdue University, USA
Co-chairs: Katarzyna Wac, University of Geneva, Switzerland
Pablo Cesar, Centrum Wiskunde & Informatica, Netherlands
Advisor: Mihaela van der Schaar, UCLA, USA

I would like to thank all the IG Chairs and co-Chairs for the work they have already done and they will be doing for the success of MMTC!

Best regards,
Maria

Maria Martini
Vice-Chair (Europe), IEEE ComSoc Multimedia Communications Technical Committee

EMERGING TOPICS: SPECIAL ISSUE ON NETWORKING FOR MULTIMEDIA

Guest Editors: Shaoen Wu[1] and Zhengang Pan[2]

[1]Department of Computer Science Ball State University, IN, USA, swu@bsu.edu
[2] China Mobile Research Institute, Beijing, China, panzhengang@chinamobile.com

Our world of multimedia calls for new networking technologies to support various types of multimedia applications. The challenges include network architectures and protocols, security, data management and localizations, etc.

This Special Issue of MMTC E-Letter focuses on the different research areas in networking that support multimedia for the future needs and requirements.

In the first article entitled, "A Case for Making Mobile Device Storage Accessible by an Operator", *Aaron Striegel* (from University of Notre Dame, USA) presents a case in real life how the mobile device can be enabled to support the access from an operator such as Verizon. The access of an operator to the mobile device storage can provide the operator the chance to mine the data across its network and enable the data centric networking that will drive the multimedia sharing in a highly efficient way. Meanwhile, various multimedia features, such as Quality of Experience (QoE), can be guaranteed.

In the second article, "Nash Bargaining Solution-based Datacenter Selection under Cloud Content Delivery Network Environments", *Heejae Kim, Yun-Gi Ha, and Chan-Hyun Youn* (from Korea Advanced Institute of Science and Technology, Korea) exploit the algorithm of Nash bargaining in selecting a proper datacenter for media content delivery. In the future, massive multimedia content will be archived in the cloud of numerous various datacenters for content storage. The way to find a proper datacenter to delivery the requested content will be a challenging task. The proposed solution examines a Nash bargaining solution to address this challenge.

The third article entitled, "Passive Vehicle Localization System based on DSRC Signals", by *Qing Yang* (from Montana State University, USA) presents a passive localization system in vehicular networking environments. Smart vehicles will provide travelers the access to tremendous entertainment content. The locations of smart vehicles will have a critical impact on car safety and application content delivery. This article proposes a novel solution that is based on Kalman Filter in order to estimate the locations of vehicles on the road. The proposed solution significantly improves the accuracy of location estimation.

The fourth article entitled "Mutual Information and Minimum Mean-Square Error in Multiuser Gaussian Channels" by *Samah Ghanem* (from Eurecom in France) studies the relationship between mutual information and the Minimum Mean-Square Error (MMSE) in a multiuser wireless communication. The investigation can unveil new unveiled relation, to allow the derivation of new closed form expressions of the mutual information for single user and multiuser channels driven by BPSK/QPSK inputs as well as to provide asymptotic expansions of the mutual information and the MMSE for multiuser setups. This work has the potential to improve the user experience of multimedia in wireless networking environments.

Finally, the fifth article, entitled "A Framework Towards SDAI", by *Qi Sun, Chih-Lin I, Shuangfeng Han, Zhengang* Pan (from China Mobile Research Institute, China) proposes a framework for the air interface of next generation (5G) wireless network. Facing the diverse requirements of 5G, the traditional network design paradigm needs to undergo a fundamental change. The 5G network needs to go soft, in terms of the network wide re-configurability of deployments, from the core network to the Radio Access Network (RAN), and the super adaptation of the communication protocols. This work bases on Software Defined Network (SDN) concept and proposes a software defined air interface. Such interface will be flexible in providing variable user experience in various applications, including multimedia.

The five articles cover different aspects of Multimedia and its future networking infrastructures. We hope they can inspire the audience with more research and interests in this area. Finally, we would like to thank all the authors for their great contribution and the E-Letter Board for making this Special Issue happen.

SHAOEN WU [M'09] (swu@bsu.edu) received a Ph.D. in computer science in 2008 from Auburn University. He is presently an assistant professor of computer science at Ball State University. He has held an assistant professor position in the School of Computing at the University of Southern Mississippi, a researcher scientist position at ADTRAN Inc., and a senior software engineer position at Bell Laboratories. His current research is in the areas of wireless and mobile networking, cyber security, cyber-physical systems, and cloud computing. He is a recipient of Best Paper Awards at IEEE ISCC 2008 and ANSS 2011. He has served as Chair, and on the committees and editorial boards of several international conferences and journals.

ZHENGANG PAN, IEEE Senior Member, principle staff of Green Communication Research Center (GCRC) of China Mobile Research Institute (CMRI), is now leading a team working on the key technologies of next generation (5G) wireless communication systems. Dr. Pan is also the vice-chair of technical WG of China IMT-2020 PG. Before join CMRI, Dr. Pan has been working with Hong Kong ASTRI for more than 6 years where he has been involved in multiple technical fields, from wireless communication (WiFi, WiMax, LTE), to mobile digital TV (T-DMB, DVB-T/H, CMMB), to wireline broadband access (HomePlug, MoCA), in both system/algorithm design and terminal SoC chip implementation. Dr. Pan has also been working with NTT DoCoMo Beijing Communication Labs Co. Ltd, on the frontier research for 4G wireless communication standards, including 802.11n, 802.16d/e, HSPA and LTE. Dr. Pan received his Ph.D degree in year 2004, from Department of Electrical and Electronic Engineering, the University of Hong Kong. Dr. Pan has expertise in many technical fields including time/frequency/sampling synchronization technology for single carrier/ multi-carrier (OFDM/A) based system, channel estimation, forward error correction coding, multiple antennas systems (MIMO) and space-time processing/coding, cross layer optimization and so on. Dr. Pan has published more than 60 papers in top journals and international conferences and filed more than 50 patents with at least 30 granted so far.

A Case for Making Mobile Device Storage Accessible by an Operator

Aaron Striegel, Xueheng Hu, Lixing Song

Department of Computer Science and Engineering, University of Notre Dame, USA
Email: {*striegel, xhu2, lsong2*}*@nd.edu*

1. Introduction

The past few years have seen a veritable explosion of wireless data consumption across a wide variety of wireless devices. The aptly dubbed *wireless data tsunami* finds its roots in growth trends predicting 1000x growth over a period of ten years [1]. With such massive demands on capacity, wireless carriers are forced to embrace an `all of the above' strategy embracing capacity gains or demand reductions wherever possible as no individual solution is likely to fully satisfy long-term needs.

In the near term, solutions such as WiFi offloading offer tremendous appeal with technologies such as ANDSF (Access Network Discovery and Selection Function) and Hotspot 2.0 poised to dramatically streamline seamless WiFi roaming [2]. Challenges emerge though with respect to how truly seamless such roaming will be and the ability to deliver consistent Quality of Experience (QoE) to end users by virtue of the fundamental nature of unlicensed spectrum.

From the cellular side, solutions such as small cells as supported by LTE-A bring the ability to augment capacity while preserving the seamless roaming and coverage afforded by cellular [3]. Whereas the cellular network can bring a more consistent QoE, challenges emerge though with respect to the expense, logistical mechanics, and newfound system complexity management. More dramatic gains can be found in various 5G efforts with significant potential viewed in the millimeter wave bands for dramatic increases in capacity. Unfortunately, significant challenges abound in such higher frequency bands with many research challenges yet to be solved [4].

While access-based solutions seek to increase the capacity to mobile users, an alternative technique is to change how and when mobile devices retrieve their information. Efforts have ranged from characterizing data and energy consumption by smartphone apps [5] to actively examining the efficiency of the data transfers themselves [6]. Broader efforts such as Information Centric Networks (ICN) and Content-Centric Networks (CCN) could be broadly viewed similarly though not necessarily particular to wireless. More recent research efforts have explored the extent to which D2D communications [7-9] might be leveraged to share cached information owing inspiration in part to prior concepts from Delay Tolerant Networks (DTN).

2. Time Shifting as a Foundational Service

We posit that the ability to time shift demand (ex. taking advantage of elasticity in the transfer time) will emerge as one of the most important mechanisms for satisfying user QoE in wireless networks. Time shifting can allow for flattening the demand curve and hence avoid catastrophic overages during peak demands which we believe is essential for user QoE.

Effective time shifting though requires two key components: (1) the availability of storage for time shifting and (2) steerable control of said storage towards network level objectives. With regards to the first component, storage remains a relatively inexpensive component for a mobile device. Our own studies have found an average of 25% to 50% of storage free on most user devices (4-8GB) [10] with user-modifiable storage able to easily add significant capacity. The second component though is considerably more difficult. Although ICN / CCN arrangements and D2D caching shift demand, such shifts may not be necessarily nor as controllable as needed to reasonably impact user QoE.

The contribution of this paper is to make the fairly radical argument operator-accessible mobile device storage should be a key feature of future mobile devices. We outline several mechanisms by which said storage could be achieved and then discuss two scenarios (Operator Push, User

Triggered Push) to demonstrate the mechanisms for such a system and to elicit further discussion from the community at large.

System Makeup

Consider one of the following two system variations for the purposes of illustration: a user-shared system variation and operator-dedicated system variation.

In the first system variation (user-shared), an underlying file system mechanism serves as an abstraction layer. To the user, the system behaves exactly as before with the user having both allocated storage (music, apps, movies, photos) and free storage. Writes and reads to the file system proceed as expected at the user level. From the perspective of the abstracted file system (less than best effort file system), the free space may be allocated as seen fit to the wireless operator. The 'free' space may be overwritten at any time by the user with a reasonable expectation that storage changes will occur gradually rather than rapidly (ex. a user may grab several hundred MBs but is unlikely to eliminate 4GB+ of storage quickly [10]). Prioritization of content for eviction would be provided via operator-defined policies.

In the second system variation (operator-dedicated), the storage is moved to be part of the network adapter(s) / chipset. No prioritization is needed for eviction as all space is exclusively accessible by the operator.

Storage Accessibility

With said storage in place, an API would be exposed for the operator to be able to write (or read if necessary) from the storage space. The API would need to be appropriately protected via appropriate security mechanisms (ex. secure connection and / or signed content). Connections would be initiated on the device side leaving open effectively a secure FTP channel equivalent for the carrier to drive content. Note that a secure device-initiated connection would also afford content steering even when the device is connected via WiFi. Support for pushing either via LTE Broadcast or multicast would also be possible.

Content would be stored locally in a block-wise manner with appropriate identifiers for access. Larger objects would likely be split into multiple blocks to accommodate cases where the device itself may not be able to cache the entirety of the larger object. An API would allow for localized trapping of content requests to the operator-managed storage as a function of the underlying operating system. Content could be sandboxed with respect to individual applications or shared across all applications.

3. Example Scenarios

Operator Push

In the first scenario, we assume that all content in the operator space must be pushed by the operator to the client. The operator could consider this content from one of two perspectives. In the first perspective, machine learning would be applied to pro-actively fetch content for a user. For instance, a user might access CNN or ESPN every morning. The connection is pre-fetched and saved in the local storage whereby the local storage is viewed as exclusively benefitting the QoE perception of the carrier (ex. carrier X's speed is acceptable). Alternative scenarios might involve LTE Broadcast and staging of larger system or application updates (ex. Angry Birds). Other alternative scenarios might involve the active involvement of on-line social networking sites effectively tagging shareable high volume content (ex. videos) for pro-active pushing as the networks allows.

From the second perspective, the storage is viewed as a potential mechanism for bringing in additional revenue. A content provider might pay a carrier for localized staging of content or even perhaps the operator might offer different levels of content distribution (localized in the geographic area, on the device, etc.). Pricing could be used to competitively prioritize cases where intense storage constraints were observed. Advertisements generated based on location might be pre-staged to avoid issues with connectivity ensuring smooth delivery (ex. payment for ad, payment for initial site content / information).

User Triggered Push

In the second scenario, we assume the presence of D2D caching whereby when mobile nodes come into range of one another, the nodes will share / exchange caches. Normally in such cases, information will propagate as mobile nodes `infect' one another with the data. In pure D2D schemes though, the only way to realize a beneficial time shift is if a mobile node comes in contact with another mobile node that already possesses the data prior to the mobile node needing the data. Effectively, the perfect arrangement of time, location, and access must occur for the full utility of D2D caching to be realized.

Although the Operator Push allows the operator to observe and potentially push popular content, the Operator Push model may miss content, particularly when content is received across multiple access mechanisms (ex. WiFi, cellular). The operator could observe checksums of similar content (via reads or simple relaying of the number of matches during a cache exchange) and use that early warning as a tripwire to push out content to specific users or even groups of users. With proper tuning, the operator could push out popular content much sooner and across broader time windows as the network allows. The key property is that by making the operator aware of such staging with the ability to control the staging, the operator can offer a better QoE and overall system optimization.

4. Conclusions

In this paper, we made the fairly radical argument that mobile device storage should be operator-accessible rather than exclusively owned and controlled by the user. Whether such storage is effectively hidden as a lower-level / less reliable storage or expressly hidden as part of the wireless chipset, the notion of a foundational network service for operator storage at the mobile device is the end goal. We believe that such storage opens up intriguing new opportunities for services whether those services are primarily in the benefit of the carrier or going even further viewed as a revenue stream for the operator. Finally, we note that there exist interesting constructs as well as applied to D2D caching and operator management of said services.

References

[1] Cisco, “Cisco VNI mobile data traffic forecast 2012-2017,” Feburary 2013.

[2] InterDigital, Cellular-WiFi-Integration, White Paper, 2011.

[3] 3GPP Release 10

[4] Andrews, J.G.; Buzzi, S.; Wan Choi; Hanly, S.V.; Lozano, A.; Soong, A.C.K.; Zhang, J.C., "What Will 5G Be?," Selected Areas in Communications, IEEE Journal on , vol.32, no.6, pp.1065,1082, June 2014

[5] N. Ding, D. Wagner, X. Chen, A. Pathak, Y. C. Hu, and A. Rice. Characterizing and modeling the impact of wireless signal strength on smartphone battery drain. In Proc. of SIGMETRICS, pages 29-40, New York, NY, USA, 2013. ACM.

[6] F. Qian, J. Huang, J. Erman, Z. M. Mao, S. Sen, and O. Spatscheck. How to reduce smartphone traffic volume by 30%?. In Proceedings of the 14th international conference on Passive and Active Measurement (PAM'13), Matthew Roughan and Rocky Chang (Eds.). 42-52, 2013.

[7] Al. Finamore, M. Mellia, Z. Gilani, K. Papagiannaki, V. Erramilli, and Y. Grunenberger. Is there a case for mobile phone content pre-staging?. In Proceedings of the ninth ACM conference on Emerging networking experiments and technologies (CoNEXT '13). ACM, New York, NY, USA, 321-326, 2013.

[8] A. Asadi, Q.Wang, and V. Mancuso, “A Survey on Device-to-Device Communication in Cellular Networks,” IEEE Communications Surveys

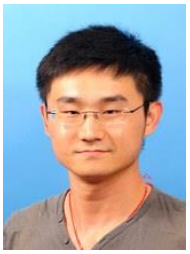
& Tutorials, vol. PP, no. 99, pp. 1–1, 2014.

[9] M. Ji, G. Caire, and A. F. Molisch, “Wireless Device-to-Device Caching Networks: Basic Principles and System Performance,” Computing

Research Repository, vol. abs/1305.5216, 2013.

[10] X. Hu, L. Meng, A. Striegel, Evaluating the Raw Potential for Device-to-Device Caching via Co-Location, in Proc. Of MobiSPC, Aug. 2014.

Aaron Striegel received his BS and Ph.D degrees from Iowa State University in Computer Engineering in 1998 and 2002, respectively. He is currently an associate professor and associate chair of Computer Science Engineering at the University of Notre Dame which he joined in 2003. He also serves on the Executive Committee of the Wireless Institute at Notre Dame. Prof. Striegel's research interests include large scale network instrumentation, wireless system dynamics, and content distribution across wireless and wired networks.



Xueheng Hu is a graduate student of Prof. Striegel in the Department of Computer Science and Engineering at the University of Notre Dame. He received his B. Eng. In Software Engineering from Beijing University in 2006 and his MS in Computer Science from Miami University of Ohio in 2011.

Lixing Song is a graduate student of Prof. Striegel in the Department of Computer Science and Engineering at the University of Notre Dame. He received his BS in Electrical Engineering from Wuhan University in 2011 and his MS in Computer Science from Ball State in 2014.

Nash Bargaining Solution-based Datacenter Selection under Cloud Content Delivery Network Environments

Heejae Kim, Yun-Gi Ha, and Chan-Hyun Youn

Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Korea
Email: {*kim881019, milmgas, chyoun*}*@kaist.ac.kr*

1. Introduction

Integrating Content Delivery Networks (CDNs) with cloud computing is an emerging issue in CDN technology recently. Amazon and Rackspace provides CloudFront [1] and Rackspace CDN [2] as the solutions for it using their clouds, and many CDN providers such as Akamai and Limelight have adopted cloud computing to improve performance of their CDNs [3,4]. Netflix started to provide their services using Amazon Web Service [5], and Google have boosted speed of YouTube using its cloud [6]. We refer to this kind of CDN combined with cloud computing as a cloud CDN [7]. Because there are usually many short busty periods in time series of end user demand of CDN [8], the use of cloud computing can give huge elasticity to CDN. Therefore, the cloud CDN can handle the dynamic demand adaptively as well as acquire enough resource even if the amount of resource which they have is relatively small. The cloud CDN providers provide their services using resources in cloud datacenter. Because quality of service (QoS) in the cloud CDN depends on the locations of the resources, datacenter selection is one of the important issues in the cloud CDN. Many researches have been studied with focusing on caching server placement [9, 10], data placement [11], domain name system (DNS) server placement [12], end user request placement [13,14,15].

In this paper, we present a novel datacenter selection algorithm to place caching server for the cloud CDN. The algorithm is based on Nash bargaining solution (NBS) [16,17] which is an attractive method for this problem guaranteeing pareto efficiency, symmetry, invariance to equivalent payoff representations, and independence of irrelevant alternatives. Also, the algorithm considers predicted end user demand and virtual machine (VM) reservation based on autoregressive integrated moving average (ARIMA) model and our previous work, C-VMR [18] respectively. In evaluation, we compare the algorithm with uniform and proportional fair sharing.

2. Datacenter Selection Algorithm

Figure 1 shows the cloud CDN environment considered in this paper. In the figure, a cloud CDN provider has several available geo-distributed datacenters. The cloud CDN provider leases VMs from the datacenters to build caching servers to handle caching server requests of the content providers. When the content providers request the use of caching servers to the cloud CDN provider, the cloud CDN provider places the caching servers to available datacenters by the datacenter selection algorithm. We suppose that it is predetermined that which datacenter handles content requests of end users in each region. Because the algorithm considers end user demand, the end user demand in every region of each content provider should be monitored and managed consistently. When end users send content requests to the DNS servers, the DNS servers determine a caching server and forward the packets to the caching server. Because QoS in content services such as online game and video streaming is highly sensitive to the factors such as network latency and throughput, the following datacenter selection algorithm is designed to handle caching server requests of the content providers effectively.

The datacenter selection algorithm is based on the NBS in consideration of end user demand prediction and VM reservation.

End User Demand Prediction

End user demand prediction is performed based on the ARIMA model similarly with [19, 20]. Before applying the model, time series of the demand should be preprocessed to make it stationary, and the order of the model should be determined.

VM Reservation

On-demand VMs (OVMs) and reserved VMs (RVMs) are two major VM types in the current cloud industry. These refer to VMs whose leasing time is relatively short such as an hour and long such as a month and respectively. Obviously, the price of RVMs in the unit time is set to be `lower than that of OVMs. Therefore, using the appropriate number of RVMs gives a benefit to the cloud CDN provider. For VM reservation in the datacenter selection algorithm, we use the C-VMR [18] to determine the appropriate number of RVMs to be leased. The mechanism of the C-VMR is as follows. The C-VMR is performed every the fixed period. At each epoch t, demand in period $[t,t + T_p]$ is predicted, and it is determined that how many RVMs are to be leased as depicted in (1) where $r_l(t)$ is the number of RVMs to be leased at epoch t, $r_e(t)$ is the number of

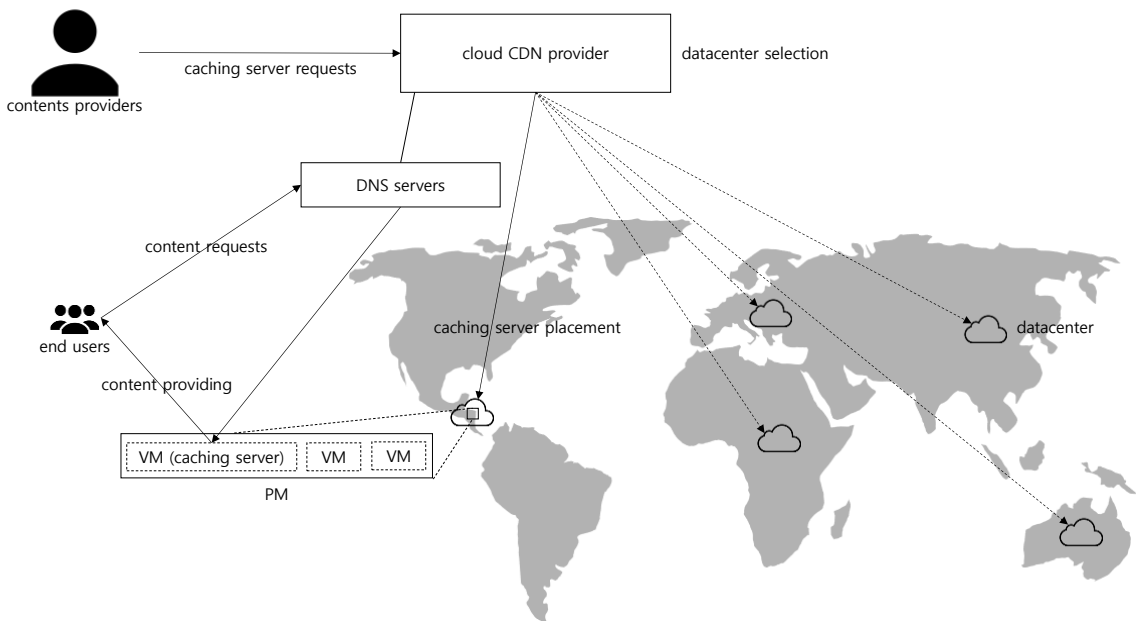


Figure 1 Cloud CDN environment.

RVMs which is available to the cloud CDN provider at the epoch *t*, and $d_p(t)$ is the predicted capable VM demand in period $[t,t+1]$. We note that the demand refers to the number of VMs which are capable to provide the caching server requests of the content providers.

$$r_l(t) = \left\lfloor \frac{1}{T_p} \sum_{k=t}^{t+T_p} d_p(k) - r_e(k) \right\rfloor. \quad (1)$$

Algorithm 1. C-VMR

Input time series of historical capable VM demand in period $[t - T_h, t - 1]$ where T_h is a period to be used as the historical demand in demand prediction

1: predict capable VM demand in period $[t, t + T_p]$

2: Obtain $r_l(t)$

3: lease additional RVMs as much as $r_l(t)$

NBS-based Datacenter Selection

We present a scheme for the NBS-based data selection in this section. The goal of the scheme is to determine the number of VMs which is to be created newly for each datacenter.

Based on the NBS, we formulate an optimization problem in consideration of end user demand and VM reservation as depicted in (2), (3), and (4).

$$\text{maximize} \sum_i \log\left(x_i - \frac{c \cdot \delta_i}{n} \right) \quad (2)$$

$$\text{subject to} \sum_i x_i - \left(s + \sum_i \rho_i \right) \quad (3)$$

$$x_i \geq \rho_i, \forall i. \quad (4)$$

$u(X)$ denotes an utility function where x_i in vector X is a decision variable and represents the number of VMs to be operated in datacenter $i \in I$. We note that I is set of available datacenters. Also, n is the number of capable end user requests in a VM, δ_i is the predicted end user demand of datacenter i, c is a control parameter. (3) is the constraint to guarantee that the sum of x_i is equal to the sum of the number of caching server requests s and the number of available RVMs where ρ_i denotes the number of available RVMs in datacenter i. (4) is the constraint to limit the minimal number of xi because x_i cannot be less than ρ_i. Finally $x_i - \rho_i$ represents the number of VMs which is to be created newly for datacenter i.

3. Evaluation

In this section, we evaluate the scheme for the NBS-based datacenter selection. For the evaluation, we generated a different end user demand for each datacenter arbitrarily and the demand is unit of the number of end user requests at an epoch. ASTSA package in R [21] is used for the demand prediction using the ARIMA model. The window size is set to 30. In the evaluation, we suppose that there are 3 available datacenters (Datacenter 1, Datacenter 2, and Datacenter 3), and the number of RVMs of each datacenter is 30, 80, and 60 respectively. Also, n is set to 10.

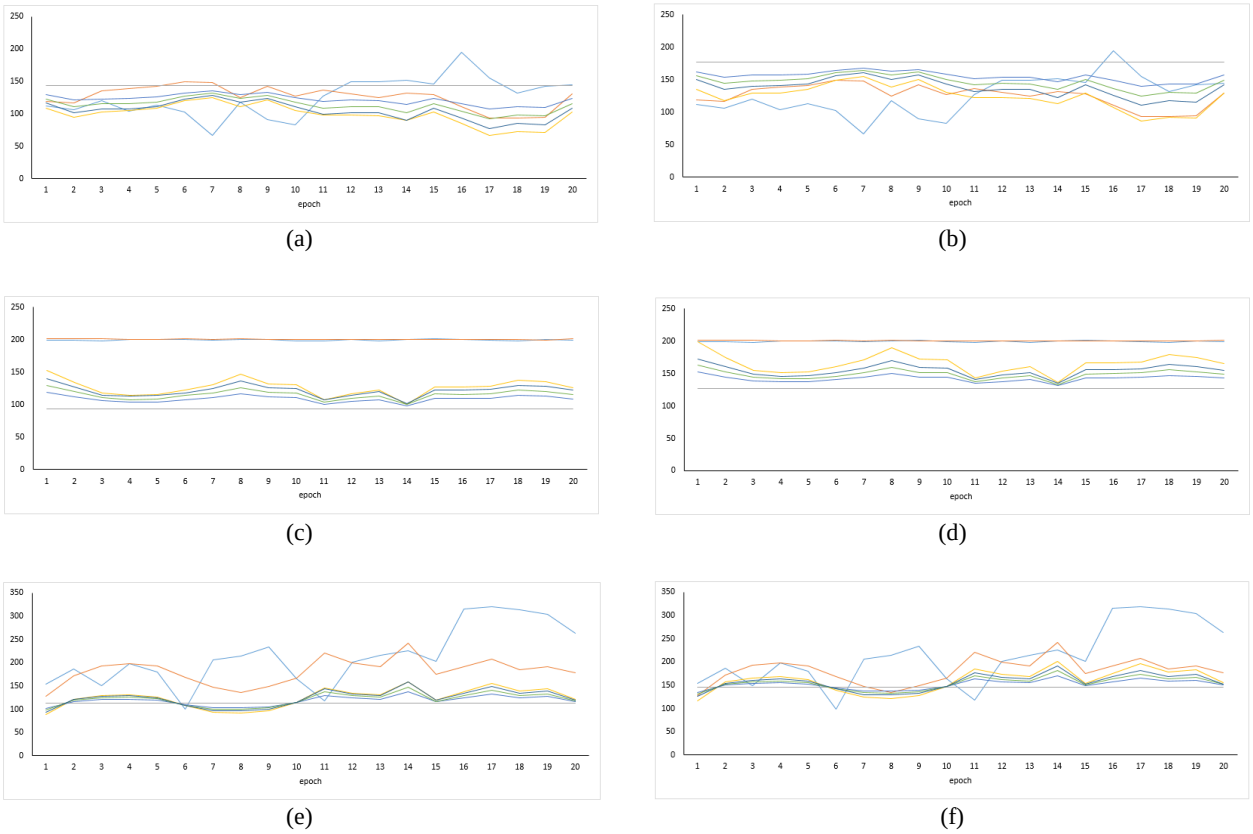


Figure 2 Evaluation results: Sky blue and orange represent actual end user demand / 10 and predicted end user demand / 10 respectively. Gray, yellow, blue, yellow green, and dark blue represent the datacenter selection result by uniform fair sharing, proportional fair sharing, NBS (c = 0.5), NBS (c = 0.7), and NBS (c = 0.9) respectively. (a) Datacenter 1, the number of caching server requests = 350. (b) Datacenter 1, the number of caching server requests = 450. (c) Datacenter 2, the number of caching server requests = 350. (d) Datacenter 2, the number of caching server requests = 450. (e) Datacenter 3, the number of caching server requests = 350. (f) Datacenter 3, the number of caching server requests = 450.

We compare the algorithm with uniform and proportional fair sharing. The uniform and proportional fair sharing are the approaches which divide the caching server requests to each datacenter uniformly and proportionally to end user demand respectively.
Figure 2 shows the result. As depicted in the figure, the trend is similar between Fig. 2(a), (c), (e) and (b), (d), (f) respectively. Demand prediction of Figure 2(c) and (d) are more accurate than Figure 2(a), (e) and (b), (f) respectively. It is because end user demands of Figure 2(c) and (d) are rarely fluctuated. Also, in all figures in Figure 2, the datacenter selection results of NBS are in between those of uniform and proportional fair sharing. Therefore, we can control *c* adaptively by system status, service level agreement (SLA), and so on. Therefore, cloud CDN providers can provide flexible services to contents providers.

4. Conclusions

In this paper, we presented the NBS-based datacenter selection for cloud CDN in geo-distributed clouds. To achieve effective datacenter selection for cloud CDN, we proposed the NBS-based datacenter selection algorithm in consideration of demand prediction and VM reservation. In evaluation, we compared the algorithm with uniform and proportional fair sharing. As on-going and future work, we are extending the algorithm to consider SLA such as latency, response time, throughput, etc.

Acknowledgement

This research was supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2014, and the MSIP under the ITRC (Information Technology Research Center) support program (NIPA-2014(H0301-14-1020)) supervised by the NIPA (National IT Industry Promotion Agency).

References

[1] CloudFront, available: http://aws.amazon.com/ko/cloudfront/.

[2] Rackspace CDN, available: http://www.rackspace.com/cloud/cdn-content-delivery-network/.

[3] Akamai, available: http://www.akamai.co.kr/.

[4] Limelight, available: http://www.limelight.com/.

[5] Google Cloud Boosting YouTube Upload Speeds, available: http://blogs.wsj.com/digits/2011/02/14/google-cloud-boosting-youtube-upload-speeds/.

[6] AWS Case Study: Netflix, http://aws.amazon.com/ko/solutions/case-studies/netflix/.

[7] F. Chen, K. Guo, J. Lin, and T. L. Porta, "Intra-cloud Lightning: Building CDNs in the Cloud," in INFOCOM, 2012.

[8] V. Aggarwal, V. Gopalakrishman, R. Jana, K. K. Ramakrishnan, and V. A. Vaishampayan, "Optimizing Cloud Resources for Delivering IPTV Services through Virtualization," in COMSNETS, 2012.

[9] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, "Cutting the Electric Bill for Internet-scale Systems," in SIGCOMM, 2009.

[10] Q. Zhang, Q. Zhu, M. F. Zhani, R. Boutaba, and J. L. Hellerstein, "Dynamic Service Placement in Geographically Distributed Clouds" in JSAC, vol. 31, no. 12, 2013.

[11] S. Agarwal, J. Dunagan, N. Jain, S. Saroju, A. Wolman, and H. Bhogan, "Volley: Automated Data Placement for Geo-distributed Cloud Services," in NSDI, 2010.

[12] H. Xu and B. Li, "A General and Practical Datacenter Selection Framework for Cloud Services," in CLOUD, 2012.

[13] H. Qian and Q. Wang, "Towards Proximity-aware Application Deployment in Geo-distributed Clouds," in ACSA, vol. 2, no. 3, 2013.

[14] P. Wendell, J. W. Jiang, M. J. Freedman, and J. Rexford, "DONAR: Decentralized Server Selection for Cloud Service," in SIGCOMM, 2010.

[15] Y. Wu, C. Wu, B. Li, L. Zhang, Z. Li, F. C. M. Lau, "Scaling Social Media Applications into Geo-distributed Clouds," in INFOCOM, 2012.

[16] J. F. Nash, Jr., "The Bargaining Problem," in Econometrica, vol. 18, no. 2, 1950.

[17] M. J. Osborne, An Introduction to Game Theory, Oxford University Press, 2009.

[18] H. Kim, Y. Ha, Y. Kim, and C. H. Youn, "A VM Reservation-based Cloud Service Broker and Its Performance Evaluation," in CloudComp, 2014.

[19] P. J. Brockwell and R. A. Davis, Introduction to Time Serires and Forecasting, Springer, 2002.

[20] W. Fang, Z. Lu, J. Wu, Z. Cao, "RPPS: A Novel Resource Prediction and Provisioning Scheme in Cloud Data Center," in SCC, 2012.

[21] "R", available: http://r-project.org/

Heejae Kim received the B.S. and M.S. degree in electrical engineering from KAIST, Daejeon, Korea, in 2011 and 2013, respectively. Currently, he is a Ph.D. candidate in electrical engineering at KAIST. His research interests include cloud-based network, cloud service broker, and resource management in cloud datacenters.

Yun-Gi Ha received the B.S. degree in electronic and electrical engineering from Sungkyunkwan University, Suwon, Korea, in 2013 and the M.S. degree from KAIST, Daejeon, Korea, in 2014. Currently, he is a Ph.D. candidate in electrical engineering at KAIST. His research interests include scientific workflow system management and efficient resource management in cloud computing environment.

Chan-Hyun Youn received the B.S. and M.S. degrees in electronics engineering from Kyungpook National University, Korea, in 1981 and 1985, respectively, and the Ph.D. degree in electrical and communications engineering from Tohoku University, Japan, in 1994. From 1986 to 1997, he worked for KT Telecom Network Research Laboratories, where he was a team leader of high-speed networking projects. Since 2009, he has been a professor in the department of electrical engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea. Where, he is a dean of office of planning and budgets and is also a director of Grid Middleware Research Center, KAIST. His research areas are cloud, computing workflow system, distributed system, communication middleware, and e-Health application services. Dr.Youn was the Editor-in-Chief in the Korea Information Processing Society, an editor of the Journal of Healthcare Engineering, U.K., and was the head of Korea branch (computer section) of the IEICE, Japan, during 2009–2010. He is a member of the IEEE.

Passive Vehicle Localization System based on DSRC Signals

Qing Yang, Montana State University, MT, USA
Email: qing.yang@cs.montana.edu

1. Introduction

In a vehicular environment, knowledge of the location of cars and other objects is powerful information that can help prevent accidents, reduce traffic, and lead to overall safer roads. Other applications include: emergency vehicle management, train crossing, tolling, and taxi management. Tracking applications are fundamental to the future of vehicular safety. Currently, the U.S. Department of Transportation is developing DSRC (Dedicated Short Range Communication) as the foundation for a national network among vehicles and roadside access points. For vehicular safety applications, accuracy in vehicles' locations is an essential requirement. As a possible solution, we present our RSSI-based vehicular tracking algorithm built on the Kalman Filter.

The RSSI is a measurement of the power of a radio signal. A main challenge with RSSI ranging is that the effect of reflecting and attenuating objects in the environment can radically distort the received RSSI, making it difficult to infer distance without a detailed model of the physical environment. In our study, we use the Kalman Filter to combat the error inherent within RSSI readings. The Kalman filter is a recursive algorithm that provides an efficient, computational method to estimate the state of a process in a way that minimizes the mean of the squared error. The filter is very powerful in several aspects: it supports estimations of past, present, and even future states, and it can do so even when the precise nature of the modeled system is unknown. The ultimate objective in our study is to use RSSI as the modality for estimating distance in combination with the Kalman Filter to achieve accuracy in multilateration that is viable for vehicular safety.

We performed tests using two wireless sensing devices, one roadside unit (RSU) and one onboard unit (OBU). We moved the OBU in timesteps to simulate a vehicle in motion and obtained the RSSI data received at each position by the OBU. We then used this data along with the Kalman Filter to calculate the position estimates and compared them to the actual path of the device to assess the accuracy of our calculations.

2. Vehicle localization based on Kalman Filter algorithm

Received Signal Strength Indicator (RSSI) has been widely used in target localization in wireless networks [1, 2, 3]. However, in exchange for ease of use, the RSSI is extremely susceptible to noise as reported in literature and supported by our empirical tests. This noisiness in the RSSI measurements is a result of the fading and shadowing of the RF signal, which is caused by the signal's surroundings [4]. More explicitly, fading and shadowing encompass the reflection, refraction, diffraction, and obstruction of the RF signal. Furthermore, multipath propagation explicates that the observed signal at the receiver is the summation of these reflected, refracted, and diffracted RF signals, distorted by the environment; these signals can add constructively or destructively, skewing the expected power of the received RF signal. Ultimately, all these effects make the RSSI measurement an erroneous indication of the distance between the transmitter and receiver.

To address this issue, we adopt the Kalman Filter algorithm [5] to achieve accurate vehicle localization based on DSRC signals [6]. Our underlying model of the mobile nodes trajectory is driven by the nodes velocity. Here, we make the assumption that the mobile vehicles velocity is accessible by our algorithm. This is not such an unfair assumption because it is fairly plausible for future vehicles to have embedded computer systems with velocity sensor readings available to them. In each of our applications of the Kalman Filter, we also utilized a Kalman smoother to further improve the position estimates.

Our localization algorithm is a two-pass filter that performs two executions of the Kalman Filter. The first Kalman Filter is used to rarefy the distance estimates given by each anchor node. Once the filtered distances are obtained, our multilateration algorithm calculates the optimal position at each time step, creating an estimation of the mobile nodes path. These new positions are then passed to our first application of the Kalman Filter to then further refine the path.

The idea behind our approach is to first improve the distance estimates before using them for multilateration. In our first application, we filtered on the mobile node's position. However, because the distance estimates are inherently noisy, the multilateration algorithm combines the errors of the distances and produces an inaccurate estimate of the mobile node's position. Therefore, to counteract this issue, we decided to implement a two-pass Kalman Filter that would refine the distance estimates before multilateration.

3. Experimental setup

We designed an experiment to gather RSSI readings received at the OBU from the RSUs, and we use this data in combination with the Kalman Filter to estimate the node's position. We used two devices made by Arada, the Locomate on board unit (OBU) and the Locomate road side unit (RSU), both equipped with a full DSRC WAVE software solution. The OBU functioned as a mobile node that would be placed in a vehicle and the RSU functioned as an anchor node. The RSSI data was always collected by the OBU.

The experiment was conducted outdoors in an open parking lot, assumed line of sight. We used a 30 by 30 ft. grid to help with localization precision. In a vehicular environment, any vehicle being localized would be in motion more often than not, so to simulate a vehicle in motion we moved the OBU along a specified path and gathered RSSI data in 3-ft. increments. Each increment represented a time step as if the device were moving. For each 3-ft. increment we gathered 500 samples of RSSI measurements from each anchor node. We collected this much data because the fading, attenuation and multipath propagation causes the RSSI to fluctuate, and taking the mean of 500 RSSI values gave us the most precise RSSI for each position.

Our experiment was limited in that we only had one RSU, whereas in a more realistic vehicular environment many RSU would be present. In order to emulate this, we placed the RSU in three different positions: (0,0), (30,0), and (0,30) and gathered RSSI data from each anchor node position every time the OBU was moved. With the RSSI data gathered from the anchor node positions, we used the path loss model to estimate the distance based on the RSSI alone. Using the aggregate anchor distance estimates, we could estimate the position of mobile node.

4. Experimental results and analysis

Figure 1 and 2 show the results of our application of the Kalman Filter algorithm in our first experiment. Figure 1 is a graphical display of a position coordinate system while Figure 2 presents the error of our Kalman Filter implementation. From both figures, it is evident that the implementation is compellingly accurate in the presence of noisy data on our initial experiment. The unfiltered estimated path from the multilateration of the distance estimates, i.e. the green path, does not encompass knowledge about the true mobile node's path; the green path is the lateration of the filtered distances, and the system model of the filtered distances does not directly relate to the mobile node's path.

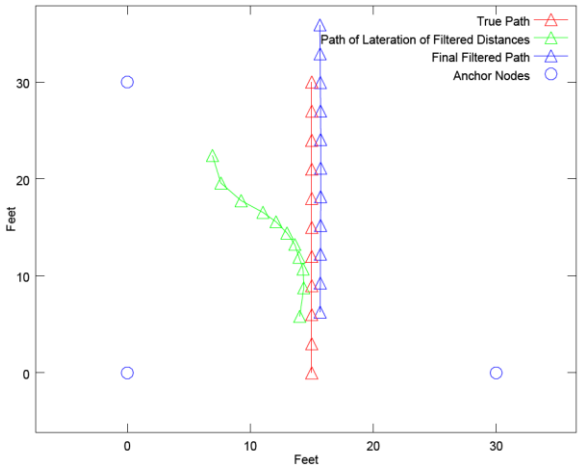


Figure 1 The local coordinate system used in the experiment

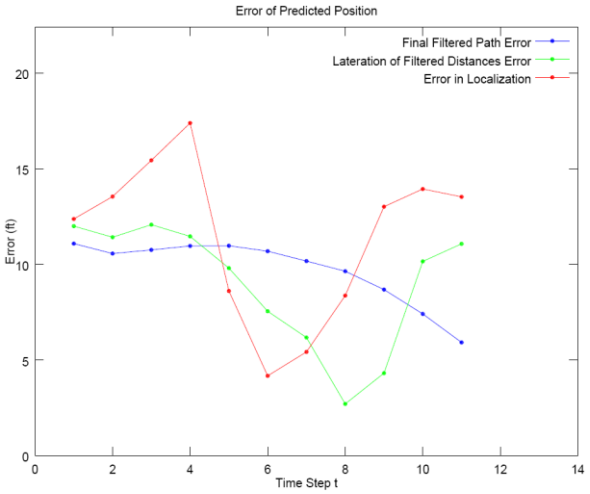


Figure 2 The errors of the steps of Kalman Filter implementation

Consequently, the shape of the resulting path is not reflective of the true path. In comparison, the final filtered path is a more accurate representation of the true path's shape. The final filtered path incorporates both the refined distances and the position model to improve the accuracy of the estimated path.

5. Conclusions

We have implemented a compelling tracking algorithm for vehicular or mobile tracking using the Kalman Filter. We utilized RSSI measurements as our medium for estimating the distance between a mobile node and an anchor node. With a set of three or more distance measurements, the position of the mobile node can be determined through various multilateration algorithms. In our experiment, we solved the linear least squares regression equation to choose the position that minimized the summation of the squared differences between the distance measurements and the estimated distances.

References

[1] C. Papamanthou, F. P. Preparata, and R. Tamassia, "Algorithms for location estimation based on RSSI sampling" in Algorithmic aspects of wireless sensor networks, S. P. Fekete, Verlag, Germany: Springer, 2008, pp. 72-86.

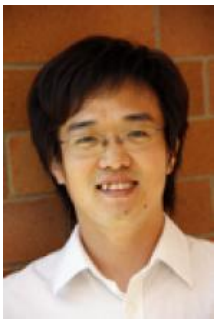
[2] X. Zhou, M. Li, and F. He, "Research on the optimization of wireless sensor network localization based on real-time estimation," in Proc. of the 2012 Int'l Conference on Information Technology and Software Engineering, 2012, pp. 23-31.

[3] G. Blumrosen, B. Hod, T. Anker, D. Dolev, B. Rubinsky, "Enhanced calibration technique for RSSI-based ranging in body area networks," in Ad hoc networks, vol. 11, (1), pp. 555-569, 2013.

[4] M. Rossi, "RSSI based tracking algorithms for wireless sensor networks: theoretical aspects and performance evaluation," M.S. Thesis, University of Padua, Padua, Italy, 2007.

[5] G. Welch, and G. Bishop, "An introduction to the kalman filter," 1995.

[6] J. Yin, G. Holland, T. Ebatt, F. Bai, and H. Krishnan, "DRSC channel fading analysis from empirical measurement," Proc. Ist IEEE Int'l Workshop Vehicle Comm &. Apps, Oct. 2006.



Qing Yang, is an RightNow Technologies Assistant Professor in the Department of Computer Science, Montana State University (MSU). He received B.S. and M.S. degrees in Computer Science from Nankai University and Harbin Institute of Technology, China, in 2003 and 2005, respectively. He received Ph.D degree in Computer Science from Auburn University in 2011. His research interests lie in the areas of wireless vehicular networks, target localization and tracking, and trust in online social networks. He has published more than 30 papers in his research areas. He serves as a TPC for several international conferences and on the editorial boards of several journals. His research has been funded by U.S. Federal Highway Administration (FHWA) and Montana Department of Transportation (DOT).

Mutual Information and Minimum Mean-Square Error in Multiuser Gaussian Channels

Samah A. M. Ghanem
Mobile Communications Department, Eurecom,Sophia Antipolis, France
Email: samah.ahanem@eurecom.fr

1. Introduction

Due to the lack of explicit closed form expressions of the mutual information for binary inputs, which were provided only for the BPSK and QPSK for the single input single output (SISO) case, [1], [2], [3], it is of particular importance to address connections between information theory and estimation theory for the multiuser case.

Connections between information theory and estimation theory dates back to the work of Duncan, in [4] who showed that for the continuous-time additive white Gaussian noise (AWGN) channel, the filtering minimum mean squared error (causal estimation) is twice the input-output mutual information for any underlying signal distribution. Recently, Guo, Shamai, and Verdu have illuminated intimate connections between information theory and estimation theory in a seminal paper, [1]. In particular, Guo et al. have shown that in the classical problem of information transmission through the conventional AWGN channel, the derivative of the mutual information with respect to the SNR is equal to the smoothing minimum mean squared error (noncausal estimation); a relationship that holds for scalar, vector, discrete-time and continuous-time channels regardless of the input statistics. There have been extensions of these results to the case of mismatched input distributions in the scalar Gaussian channel in [5] and [6].

However, the fundamental relation between the derivative of the mutual information and the MMSE, known as I-MMSE identity, and defined for point to point channels with any noise or input distributions in [1] is not anymore suitable for the multiuser case. Therefore, in this paper, we revisit the connections between the mutual information and the MMSE for the multiuser setup. We generalize the I-MMSE relation to the multiuser case. In particular, we prove that the derivative of the mutual information with respect to the signal to noise ratio (SNR) is equal to the minimum mean squared error (MMSE) plus a covariance induced due to the interference, quantified by a term with respect to the cross correlation of the users inputs' estimates, their channels, and their precoding matrices. Further, we capitalize on this unveiled multiuser I-MMSE relation to derive the components of the multiuser mutual information. In particular, we derive the derivative of the conditinal and non-conditional mutual information with respect to the SNR.

Further extensions of this result allows a generalization of the relations of linear vector Gaussian channels in [7] to multiuser channels. In particular, [8], [9] generalize the I-MMSE relation to the per-user gradient of the conditional, non-conditional and joint mutual information with respect to the MMSE, channels and precoders (power allocation) matrices of the user and the interferer.

Such new unveiled relation allows, the derivation of new closed form expressions of the mutual information for single user and multiuser channels driven by BPSK/QPSK inputs, and to provide asymptotic expansions of the mutual information and the MMSE for multiuser setups, [10].

Throughout the paper, the following notation is employed, boldface uppercase letters denote matrices, lowercase letters denote scalars. The superscript, $(.)^{-1}$, $(.)^{T}$, $(.)^{*}$, and $(.)^{\dagger}$ denote the inverse, transpose, conjugate, and conjugate transpose operations. The (∇) denotes the gradient of a scalar function with respect to a variable. The $\mathbb{E}[.]$ denotes the expectation operator. The $||.||$ and $Tr\{.\}$ denote the Euclidean norm, and the trace of a matrix, respectively.

The rest of the paper is organized as follows; section 2 introduces the system model. Section 3 introduces the new fundamental relation between the multiuser mutual information and the MMSE. Section 4 provides the conditional and non-conditional components of the I-MMSE identity.

2. System Model

Consider the deterministic complex-valued vector channel,

$$\mathbf{y} = \sqrt{snr}\ \mathbf{H_1}\mathbf{P_1}\mathbf{x_1} + \sqrt{snr}\ \mathbf{H_2}\mathbf{P_2}\mathbf{x_2} + \mathbf{n}, \qquad (1)$$

where the $n_r \times 1$ dimensional vector $\mathbf{y}$ and the $n_t \times 1$ dimensional vectors $\mathbf{x_1}$, $\mathbf{x_2}$ represent, respectively, the received vector and the independent zero-mean unit-variance transmitted information vectors from each user input to the multiuser channel[1]. The distributions of both inputs are not fixed, not necessarily Gaussian nor identical. The $n_r \times n_t$ complex-valued matrices $\mathbf{H_1}$, $\mathbf{H_2}$ correspond to the deterministic channel gains for both input channels (known to both encoder and decoder) and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the $n_r \times 1$ dimensional complex Gaussian noise with independent zero-mean unit-variance components. The $n_t \times n_t$ $\mathbf{P_1}$, $\mathbf{P_2}$ are precoding matrices that do not increase the transmitted power.

3. New Fundamental Relation between the Mutual Information and the MMSE

The first contribution is given in the following theorem, which provides a generalization of the I-MMSE identity to the multiuser case.

[1]We consider the two-user case for ease of exploitation. However, the relations apply to the k-user case.

Theorem 1: The relation between the derivative of the mutual information with respect to the SNR and the non-linear MMSE for a multiuser Gaussian channel satisfies:

$$\frac{dI(snr)}{dsnr} = mmse(snr) + \psi(snr) \tag{2}$$

Where,

$$mmse(snr) = Tr\left\{\mathbf{H_1P_1E_1(H_1P_1)}^\dagger\right\} + Tr\left\{\mathbf{H_2P_2E_2(H_2P_2)}^\dagger\right\}, \tag{3}$$

$$\psi(snr) = Tr\left\{\mathbf{H_1P_1}\mathbb{E}_{\mathbf{y}}[\mathbb{E}_{\mathbf{x_1|y}}[\mathbf{x_1|y}]\mathbb{E}_{\mathbf{x_2|y}}[\mathbf{x_2|y}]^\dagger](\mathbf{H_2P_2})^\dagger\right\} - Tr\left\{\mathbf{H_2P_2}\mathbb{E}_{\mathbf{y}}[\mathbb{E}_{\mathbf{x_2|y}}[\mathbf{x_2|y}]\mathbb{E}_{\mathbf{x_1|y}}[\mathbf{x_1|y}]^\dagger](\mathbf{H_1P_1})^\dagger\right\},$$

Proof: See Appendix A ■

The per-user MMSE is given respectively as follows:

$$\mathbf{E}_1 = \mathbb{E}_{\mathbf{y}}[(\mathbf{x}_1 - \widehat{\mathbf{x}}_1)(\mathbf{x}_1 - \widehat{\mathbf{x}}_1)^\dagger] \tag{4}$$

$$\mathbf{E_2} = \mathbb{E}_{\mathbf{y}}[(\mathbf{x}_2 - \widehat{\mathbf{x}}_2)(\mathbf{x}_2 - \widehat{\mathbf{x}}_2)^\dagger]. \tag{5}$$

The non-linear input estimates of each user input is given respectively as follows:

$$\widehat{\mathbf{x}}_1 = \mathbb{E}_{\mathbf{x_1|y}}[\mathbf{x_1|y}] = \sum_{\mathbf{x_1},\mathbf{x_2}} \frac{\mathbf{x}_1 p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) p_{x_1}(\mathbf{x_1}) p_{x_2}(\mathbf{x_2})}{p_y(\mathbf{y})} \tag{6}$$

$$\widehat{\mathbf{x}}_2 = \mathbb{E}_{\mathbf{x_2|y}}[\mathbf{x_2|y}] = \sum_{\mathbf{x_1},\mathbf{x_2}} \frac{\mathbf{x_2} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) p_{x_1}(\mathbf{x_1}) p_{x_2}(\mathbf{x_2})}{p_y(\mathbf{y})} \tag{7}$$

The conditional probability distribution of the Gaussian noise is defined as:

$$p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) = \frac{1}{\pi^{n_r}} e^{-\left\|\mathbf{y}-\sqrt{snr}\mathbf{H_1P_1x_1}-\sqrt{snr}\mathbf{H_2P_2x_2}\right\|^2} \tag{8}$$

The probability density function for the received vector $\mathbf{y}$ is defined as:

$$p_y(\mathbf{y}) = \sum_{\mathbf{x_1},\mathbf{x_2}} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) p_{x_1}(\mathbf{x_1}) p_{x_2}(\mathbf{x_2}). \tag{9}$$

Henceforth, the system MMSE with respect to the SNR is given by:

$$mmse(snr) = \mathbb{E}_{\mathbf{y}}\left[\left\|\mathbf{H_1P_1}(\mathbf{x_1} - \mathbb{E}_{\mathbf{x_1|y}}[\mathbf{x_1|y}])\right\|^2\right] + \mathbb{E}_{\mathbf{y}}\left[\left\|\mathbf{H_2P_2}(\mathbf{x_2} - \mathbb{E}_{\mathbf{x_2|y}}[\mathbf{x_2|y}])\right\|^2\right], \tag{10}$$

$$= Tr\left\{\mathbf{H_1P_1E_1(H_1P_1)}^\dagger\right\} + Tr\left\{\mathbf{H_2P_2E_2(H_2P_2)}^\dagger\right\} \tag{11}$$

Note that the term $mmse(snr)$ is due to the users MMSEs, particularly, $mmse(snr) = mmse_1(snr) + mmse_2(snr)$ and $\psi(snr)$ are covariance terms that appear due to the covariance of the interferers. Those terms are with respect to the channels, precoders, and non-linear estimates of the user inputs.

When the covariance terms vanish to zero, the mutual information with respect to the SNR will be equal to the MMSE with respect to the SNR, this applies to the relation for the single user and point to point communications. Therefore, the result of Theorem 1 is a generalization of such connection between the two canonical operational measures in information theory and estimation theory - the mutual information and the MMSE - and boils down to the result of Guo et. al, [1] under certain conditions which are: (i) when the cross correlation between the inputs estimates equals zero (ii) when interference can be neglected, and (iii) under the single user setup.

Such generalized fundamental relation between the change in the multiuser mutual information and the SNR is of particular relevance. Firstly, such result allows us to understand the behavior of per-user rates with respect to the interference due to the mutual interference and the interference of other users in terms of their power levels and channel strengths. In addition, the result allows us to be able to quantify the losses incurred due to the interference in terms of bits.

Therefore, when the term $\psi(snr)$ equals zero. The derivative of the mutual information with respect to the SNR equals the total $mmse(snr)$:

$$\frac{dI(snr)}{dsnr} = mmse(snr), \tag{12}$$

which matches the result by Guo et. al in [1].

4. The Conditional and Non-Conditional I-MMSE

In this section, we capitalize on the new fundamental relation to extend the derivative with respect to the SNR to the conditional and non-conditional mutual information. To make this more clear, we capitalize on the chain rule of the mutual information which states the following:

$$I(\mathbf{x_1},\mathbf{x_2};\mathbf{y}) = I(\mathbf{x_1};\mathbf{y}) + I(\mathbf{x_2};\mathbf{y}|\mathbf{x_1}) \tag{13}$$

Therefore, through this observation we can conclude the following theorem.

Theorem 2: The relation between the derivative of the conditional and the non-conditional mutual information and their corresponding minimum mean squared error satisfies, respectively:

$$\frac{dI(\mathbf{x_2};\mathbf{y}|\mathbf{x_1})}{dsnr} = mmse_2(snr) + \psi(snr) \tag{14}$$

$$\frac{dI(\mathbf{x_1};\mathbf{y})}{dsnr} = mmse_1(\gamma snr) \tag{15}$$

Proof: Taking the derivative of both sides of (13), and subtracting the derivative of $I(\mathbf{x_1};\mathbf{y})$ which is equal to $mmse_1(\gamma snr)$, γ is a scaling factor, due to the fact that x_1 is decoded first considering the other users' input x_2 as noise. Therefore, Theorem 2 has been proved. ■

Of particular relevance is the implication of the derived relations on understanding the achievable rates of interference channels. In particular, such relation allows for better understanding of the changes in the rates due the interferer which is either decoded first or considered as noise. Additionally, further details on the generalized relation that expresses the gradient with respect to arbitrary parameters for the joint, conditional, and non-conditional mutual information can be found in [8], [9].

5. Conclusions

We generalize the fundamental relation between the derivative of the mutual information and the MMSE to multiuser setups. We prove that the derivative of the mutual information with respect to the SNR is equal to the MMSE plus a covariance induced due to the interference, quantified by a term with respect to the cross correlation of the multiuser inputs' estimates, their channels, and their precoding matrices. We provide such relations for conditional and non-conditional components of the multiuser mutual information.

Appendix A: Proof of Theorem 1

The conditional probability density for the two-user multiple access Gaussian channel can be written as follows:

$$p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) = \frac{1}{\pi^{n_r}} e^{-\left\|y-\sqrt{snr}\mathbf{H_1P_1x_1}-\sqrt{snr}\mathbf{H_2P_2x_2}\right\|^2} \tag{16}$$

Thus, the corresponding mutual information is:

$$I(\mathbf{x_1},\mathbf{x_2};\mathbf{y}) = \mathbb{E}\left[log\left(\frac{p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{p_y(\mathbf{y})}\right)\right] \tag{17}$$

$$I(\mathbf{x_1},\mathbf{x_2};\mathbf{y}) = -n_r log(\pi e) - \mathbb{E}\left[log\left(p_y(\mathbf{y})\right)\right] \tag{18}$$

$$I(\mathbf{x_1},\mathbf{x_2};\mathbf{y}) = -n_r log(\pi e) - \int p_y(\mathbf{y}) log\left(p_y(\mathbf{y})\right) d\mathbf{y} \tag{19}$$

Then, the derivative of the mutual information with respect to the SNR is as follows:

$$\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} = -\frac{\partial}{\partial snr}\int p_y(\mathbf{y}) log\left(p_y(\mathbf{y})\right) d\mathbf{y} \tag{20}$$

$$= -\int\left(p_y(\mathbf{y})\frac{1}{p_y(\mathbf{y})} + log\left(p_y(\mathbf{y})\right)\right)\frac{\partial p_y(\mathbf{y})}{\partial snr} d\mathbf{y} \tag{21}$$

$$= -\int\left(1 + log\left(p_y(\mathbf{y})\right)\right)\frac{\partial p_y(\mathbf{y})}{\partial snr} d\mathbf{y} \tag{22}$$

Where the probability density function of the received vector $\mathbf{y}$ is given by:

$$p_y(\mathbf{y}) = \sum_{\mathbf{x_1},\mathbf{x_2}} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) p_{x_1,x_2}(\mathbf{x_1},\mathbf{x_2}) \tag{23}$$

$$= \mathbb{E}_{x_1,x_2}\left[p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right] \tag{24}$$

The derivative of the conditional output with respect to the SNR can be written as:

$$\frac{\partial p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{\partial snr} = -p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\times \frac{\partial}{\partial snr}\left(\mathbf{y}-\sqrt{snr}\mathbf{H_1P_1x_1}-\sqrt{snr}\mathbf{H_2P_2x_2}\right)^\dagger \times \left(\mathbf{y}-\sqrt{snr}\mathbf{H_1P_1x_1}-\sqrt{snr}\mathbf{H_2P_2x_2}\right) \tag{25}$$

$$= -\frac{1}{\sqrt{snr}}\left((\mathbf{H_1P_1x_1})^\dagger - (\mathbf{H_2P_2x_2})^\dagger\right)\times \left(\mathbf{y}-\sqrt{snr}\mathbf{H_1P_1x_1}-\sqrt{snr}\mathbf{H_2P_2x_2}\right)\times p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) \tag{26}$$

$$= -\frac{1}{\sqrt{snr}}\left((\mathbf{H_1P_1x_1})^\dagger - (\mathbf{H_2P_2x_2})^\dagger\right)\nabla_{\mathbf{y}} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) \tag{27}$$

Therefore, we have:

$$\mathbb{E}_{x_1,x_2}\left[\nabla_{snr} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right] = \mathbb{E}_{x_1,x_2}[-\frac{1}{\sqrt{snr}}\left((\mathbf{H_1P_1x_1})^\dagger - (\mathbf{H_2P_2x_2})^\dagger\right)\times \nabla_{\mathbf{y}} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})] \tag{28}$$

Substitute (28) into (22), we get:

$$\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} = \frac{1}{\sqrt{snr}}\int\left(1+log\left(p_y(\mathbf{y})\right)\right)\times \mathbb{E}_{x_1,x_2}[\left((\mathbf{H_1P_1x_1})^\dagger - (\mathbf{H_2P_2x_2})^\dagger\right)\times \nabla_{\mathbf{y}} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})]d\mathbf{y} \tag{29}$$

$$= \frac{1}{\sqrt{snr}}\mathbb{E}_{x_1,x_2}[(\int\left(1+log\left(p_y(\mathbf{y})\right)\right)\times \left((\mathbf{H_1P_1x_1})^\dagger - (\mathbf{H_2P_2x_2})^\dagger\right)\times \nabla_{\mathbf{y}} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})d\mathbf{y})] \tag{30}$$

Using integration by parts applied to the real and imaginary parts of $\mathbf{y}$ we have:

$$\int\left(1+log\left(p_y(\mathbf{y})\right)\right)\frac{\partial p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{\partial t} d\mathbf{t} = \int\left(1+log\left(p_y(\mathbf{y})\right)\right) p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})|_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\frac{1}{p_y(\mathbf{y})}\frac{\partial p_y(\mathbf{y})}{\partial t} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) d\mathbf{t} \tag{31}$$

The first term in (31) goes to zero as $\|\mathbf{y}\| \to \infty$. Therefore,

$$\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} = \frac{1}{\sqrt{snr}}\mathbb{E}_{x_1,x_2}[-\int\left(((\mathbf{H_1P_1x_1})^\dagger - (\mathbf{H_2P_2x_2})^\dagger\right)\times \frac{p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{p_y(\mathbf{y})}\times \nabla_{\mathbf{y}} p_y(\mathbf{y})d\mathbf{y})] \tag{32}$$

$$\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} = -\frac{1}{\sqrt{snr}}\int\nabla_{\mathbf{y}} p_y(\mathbf{y})\times \mathbb{E}_{x_1,x_2}[(\mathbf{H_1P_1x_1})^\dagger - (\mathbf{H_2P_2x_2})^\dagger)\times \frac{p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{p_y(\mathbf{y})}]d\mathbf{y} \tag{33}$$

$$\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} = -\frac{1}{\sqrt{snr}}\int\nabla_{\mathbf{y}} p_y(\mathbf{y})\times \mathbb{E}_{x_1,x_2}[(\mathbf{H_1P_1})^\dagger\mathbb{E}_{x_1|y}\left[\mathbf{x_1}|\mathbf{y}\right]^\dagger - (\mathbf{H_2P_2})^\dagger\mathbb{E}_{x_2|y}\left[\mathbf{x_2}|\mathbf{y}\right]^\dagger]d\mathbf{y} \tag{34}$$

However,

$$\nabla_{\mathbf{y}} p_y(\mathbf{y}) = \nabla_{\mathbf{y}}\mathbb{E}_{x_1,x_2}\left[p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right] = \mathbb{E}_{x_1,x_2}\left[\nabla_{\mathbf{y}} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right] = -\mathbb{E}_{x_1,x_2}[p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\times \left(\mathbf{y}-\sqrt{snr}\mathbf{H_1P_1x_1}-\sqrt{snr}\mathbf{H_2P_2x_2}\right) = -\mathbb{E}_{x_1,x_2}\left[p_y(\mathbf{y})\left(\mathbf{y}-\sqrt{snr}\mathbf{H_1P_1x_1}-\sqrt{snr}\mathbf{H_2P_2x_2}\right)|\mathbf{y}\right] = -p_y(\mathbf{y})\times(\mathbf{y} - \sqrt{snr}\mathbf{H_1P_1}\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}] - \sqrt{snr}\mathbf{H_2P_2}\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]) \tag{35}$$

Substitute (35) into (34) we get:

$$\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} = \frac{1}{\sqrt{snr}}\int p_y(\mathbf{y})(\mathbf{y}-\sqrt{snr}\mathbf{H_1P_1}\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}] - \sqrt{snr}\mathbf{H_2P_2}\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}])\times \mathbb{E}_{x_1,x_2}\left((\mathbf{H_1P_1})^\dagger\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]^\dagger - (\mathbf{H_2P_2})^\dagger\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]^\dagger\right)d\mathbf{y}$$

$$\begin{aligned}\frac{d\mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} =\ & \frac{1}{\sqrt{\mathbf{snr}}}\mathbb{E}_y[\mathbf{y}\mathbf{x}_1^\dagger](\mathbf{H_1P_1})^\dagger - \frac{1}{\sqrt{snr}}\mathbb{E}_y[\mathbf{y}\mathbf{x}_2^\dagger](\mathbf{H_2P_2})^\dagger \\ & - \mathbb{E}_y[\mathbf{H_1P_1}\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]^\dagger](\mathbf{H_1P_1})^\dagger \\ & + \mathbb{E}_y[\mathbf{H_1P_1}\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]^\dagger](\mathbf{H_2P_2})^\dagger \\ & + \mathbb{E}_y[\mathbf{H_2P_2}\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]^\dagger](\mathbf{H_2P_2})^\dagger \\ & - \mathbb{E}_y[\mathbf{H_2P_2}\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]^\dagger](\mathbf{H_1P_1})^\dagger \quad (36)\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} =\ & \mathbf{H_1P_1}\mathbb{E}_{x_1}[\mathbf{x_1}\mathbf{x_1}^\dagger](\mathbf{H_1P_1})^\dagger \\ & - \mathbf{H_1P_1}\mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]^\dagger(\mathbf{H_1P_1})^\dagger \\ & + \mathbf{H_1P_1}\mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}_{\mathbf{x_2}|\mathbf{y}}[\mathbf{x_2}|\mathbf{y}]^\dagger](\mathbf{H_2P_2})^\dagger \\ & - \mathbf{H_2P_2}\mathbb{E}_{x_2}[\mathbf{x_2}\mathbf{x_2}^\dagger](\mathbf{H_2P_2})^\dagger \\ & + \mathbf{H_2P_2}\mathbb{E}_y[\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]^\dagger(\mathbf{H_2P_2})^\dagger \\ & - \mathbf{H_2P_2}\mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]^\dagger](\mathbf{H_1P_1})^\dagger \quad (37)\end{aligned}$$

According to (4) and (5), (37) simplifies to:

$$\begin{aligned}\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} =\ & \mathbf{H_1P_1E_1}(\mathbf{H_1P_1})^\dagger - \mathbf{H_2P_2E_2}(\mathbf{H_2P_2})^\dagger \\ & + \mathbf{H_1P_1}\mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]^\dagger](\mathbf{H_2P_2})^\dagger \\ & - \mathbf{H_2P_2}\mathbb{E}_y[\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]^\dagger](\mathbf{H_1P_1})^\dagger \quad (38)\end{aligned}$$

Therefore, due to the fact that the expectation remains the same if $\mathbf{y}\sim\mathcal{CN}(\sqrt{snr}\mathbf{H_1P_1x_1}+\sqrt{snr}\mathbf{H_2P_2x_2},\mathbf{I})$ or $\mathbf{y}\sim\mathcal{CN}(-\sqrt{snr}\mathbf{H_1P_1x_1}-\sqrt{snr}\mathbf{H_2P_2x_2},\mathbf{I})$. Then, $\mathbb{E}[\mathbf{x_1}\mathbf{x_1}^\dagger]=-\mathbb{E}[\mathbf{x_1}\mathbf{x_1}^\dagger]$ and $\mathbb{E}[\mathbb{E}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}[\mathbf{x_1}|\mathbf{y}]^\dagger]=-\mathbb{E}[\mathbb{E}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}[\mathbf{x_1}|\mathbf{y}]^\dagger]$, and similarly for the other input. The derivative of the mutual information with respect to the SNR and the per users mmse and input estimates (or covariances) is as follows:

$$\begin{aligned}\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} =\ & mmse_1(snr)+mmse_2(snr) \\ & + Tr\left\{\mathbf{H_1P_1}\mathbb{E}_y[\widehat{\mathbf{x}}_1\widehat{\mathbf{x}}_2^\dagger](\mathbf{H_2P_2})^\dagger\right\} \\ & - Tr\left\{\mathbf{H_2P_2}\mathbb{E}_y[\widehat{\mathbf{x}}_2\widehat{\mathbf{x}}_1^\dagger](\mathbf{H_1P_1})^\dagger\right\} \quad (39)\end{aligned}$$

Therefore, we can write the derivative of the derivative of the mutual information with respect to the SNR as follows:

$$\frac{dI(snr)}{dsnr} = mmse(snr)+\psi(snr) \quad (40)$$

Therefore, Theorem 1 has been proved as a generalization of the I-MMSE identity to the multiuser case.

References

[1] D. Guo and S. Shamai and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels", *IEEE Transactions on Information Theory*, vol. 51, pp. 1261–1282, Apr 2005.

[2] S. A. M. Ghanem, "Mutual Information for Generalized Arbitrary Binary Input Constellations", *MAP-Tele Workshop*, Feb 2010.

[3] S. A. M. Ghanem, "Analysis, Modeling, Design, and Optimization of Future Communications Systems: From Theory to Practice", *Ph.D. Thesis*, Dec 2013.

[4] T. E. Duncan, "On the calculation of mutual information", *SIAM Journal on Applied Mathematics*, vol. 19, pp. 215–220, Jul 1970.

[5] S. Verdu, "Mismatched Estimation and Relative Entropy", *IEEE Transactions on Information Theory*, vol. 56, pp. 3712–3720, Aug 2010.

[6] T. Weissman, "The Relationship Between Causal and Noncausal Mismatched Estimation in Continuous-Time AWGN Channels", *IEEE Transactions on Information Theory*, vol. 56, pp. 4256-4273, Sept 2010.

[7] D. P. Palomar and S. Verdu, "Gradient of mutual information in linear vector Gaussian channels", *IEEE Transactions on Information Theory*, vol. 52, pp. 141–154, Jan 2006.

[8] S. A. M. Ghanem, "MAC Gaussian Channels With Arbitrary Inputs: Optimal Precoding and Power Allocation", *IEEE International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct 2012.

[9] S. A. M. Ghanem, "Multiple Access Gaussian Channels With Arbitrary Inputs: Optimal Precoding and Power Allocation", *http://arxiv.org/abs/1411.0446*, 2014.

[10] S. A. M. Ghanem, "Multiuser I-MMSE", *in preparation*

Samah A. M. Ghanem was born in Beirut, Lebanon, 1979. She received her BS in Electrical Engineering from An-Najah National University, Palestine, in 2001, the M.Sc. in Scientific Computing from Birzeit University, Palestine in 2009, and the Ph.D. in Telecommunications Engineering (summa cum laude) from the University of Aveiro (MAP-Tele), Portugal in 2013. She worked for the industry between 2001 to 2008, and at Ericsson LMI R&D, Ireland in 2008. She joined the Mobile Communications Dept. at Eurecom, Sophia Antipolis, France as a postdoctoral researcher in March, 2014. She was a visiting researcher at Aalborg University, Denmark in 2012 and at the University of Duisburg-Essen, Germany in 2008. Between 2009 to 2013, she was a researcher at the Institute of Telecommunications, Porto, Portugal. Dr. Ghanem is an IEEE member, IEEE ComSoc member and IEEE ITSoc member. She served as a TPC of several IEEE conferences such as Globecom'14, and as a reviewer for IEEE journals, such as IEEE T-DSC. Her research interests include information theory, estimation theory, statistical signal processing, and network coding.

A Framework Towards SDAI

Qi Sun, Chih-Lin I, Shuangfeng Han, Zhengang Pan
China Mobile Research Institute, Beijing 100053, P. R. China
Email: {sunqiyjy, icl, hanshuangfeng, panzhengang } @ chinamobile.com

1. Introduction

With the global commercialization of the 4G LTE standard, the wireless community is now looking forward to the next generation mobile network. Worldwide initiatives on the 5th generation (5G) wireless communication have been extensively carried out, starting with an investigation on user demands, scenarios, key performance indicators (KPIs) and enabling technologies. A global consensus is forming that 5G network will be able to sustainably support 1000-fold mobile data traffic growth, improve energy efficiency (EE) and cost efficiency by over 100 times, provide fiber link access data rate and "zero" latency user experience, be capable of connecting 100 billion devices and capable of delivering a consistent experience across a variety of scenarios including the cases of ultra-high traffic volume density, ultra-high connection density and ultra-high mobility [1,2]. 5G is also expected to provide intelligent optimization based on services and user awareness. Though 5G is still in the initiative stage, there emerges some promising themes in how to design 5G network, e.g., green and soft, proposed by China Mobile [3].

The existing wireless networks from 2G, 3G to 4G are composed of specifically designed hardware and dedicated firmware which are lack of flexibility. Facing the diverse requirements of 5G, the traditional network design paradigm need to undergo a fundamental change. The 5G network needs to go soft, in terms of the network wide reconfigurability of deployments, from the core network to the radio access network (RAN), and the super adaptation of the communication protocols. The soft core network and the soft RAN are being actively studied. Software defined network (SDN) [4] and network function virtualization (NFV) [5] are generally identified as the most promising technologies for 5G mobile core network architecture. By decoupling the software and hardware, control and data, uplink and downlink, they can enable a super flat architecture of network with flexibility and scalability under cost-efficient network deployment. C-RAN (Centralized, Clean, Cooperative, and Cloud-RAN) [6], which was first introduced by China Mobile, has been envisioned as a key enabling elements of a soft cell infrastructure. It implements a soft and virtualized base station (BS) pool with multiple baseband units (BBUs) for centralized processing. The virtualized and centralized baseband processing allows for soft and dynamic cell reconfigurations. Though the C-RAN architecture integrated with SDN and NFV technologies is possible to achieve the network scalability and flexibility, to fulfill the soft concept of 5G network, it needs to be assisted with a flexible communication protocol, especially the air interface. In 5G, the air interface needs to go soft as well to facilitate the network as a service and provide users with massive diverse services and consistent quality of experience.

To date, the paradigm of air interface design focused on the most stringent operating condition and adopted a one-fits-all approach. As a result, a global optimized/trade-off air interface design which is not necessary optimized for each individual application scenario was adopted by existing standards. When it comes to 5G, facing the requirements of 1000+ SE and 100+ EE improvement, massive user connections and diverse services demands, the 5G air interface should be capable of addressing the following issues:

- ✧ Support all spectrum access: a global consensus is forming that 1000MHz bandwidth of additional mobile spectrum is required for 5G. Thus, the future air interface needs to be able to access the low frequency band (<6GHz), the high frequency band (>6GHz) and also needs to be able to efficiently utilize the fragmented spectrum to provide full spectrum of services.
- ✧ Fulfill the demands of diverse scenarios and services: e.g., IoT applications with massive small asynchronous burst connections, tactile Internet, high speed train and the ultra-reliable communication with ultra-low latency, etc. The 5G air interface is expected to enable wireless signals "dressed for the occasion".
- ✧ Strive for multiple objectives or KPIs optimization: a variety metrics are put forward in the 5G network design: higher peak data rates, better energy efficiency and lower cost, improved coverage, better scalability with number of devices, low latency etc., which is quite different from the previous generations that mainly focused on the peak data rate. Hence, when designing the 5G air interface, how to properly handle the existence and tradeoff between multiple objectives should be kept in mind.

The above listed challenges call for a revolutionary paradigm shift in the 5G air interface design. In [7], a context-aware approach to wireless rate adaptation, named WhiteRate, is proposed to enable the flexibility of the air interface, allowing the fine grained transmission adjustment. But it mainly focuses on the modulation and coding scheme, and channel width

adaptation based on the packet error rate performance, resulting in limited scope of application. In this paper, we propose the concept of software defined air interface (SDAI) as the framework of the 5G air interface, which riches the adaptation and context-awareness dimensionality in [7]. SDAI provides a scalable and configurable mechanism to customize air interface design to support different services and applications under different transmitting and receiving conditions. The essential design principle of SDAI is making air interface to be service oriented, green and soft via adaptation of the fundamental building blocks, such as frame structure, duplex mode, multiple access scheme, waveform, modulation and coding, etc. Aiming to shed some lights on the 5G air interface design in terms of SDAI and the adaptation mechanism involved, this paper is organized as following: In section II, the concept of SDAI is introduced, including the features and the design framework. In section III, we present the promising technology enablers and introduce the smart engine: intelligent adaptation controller of SDAI. The paper is summarized in Section VI.

2. Software Defined Air Interface

Air interface, the most fundamental aspect of a wireless standard, to a large extent, is considered as the defining technical feature of each wireless communication generation and has been continually evolving in each cellular generation. Looking at the air interface design paradigm in 2G/3G/4G, a single air interface with specific building blocks and parameters was elaborately designed to efficiently provide the voice, short message and even the packet data services. Besides, to ensure the radio link QoS under the dynamic variations of the instantaneous channel conditions, link adaptation is introduced. However, the link adaptation has been relatively limited to the rather trivial search space of some particular modules, e.g., power control, modulation and coding due to the hardware limitations.

When it comes to 5G, as the challenges outlined earlier, the 5G air interface needs to be unprecedentedly scalable and flexible. And the link adaptation in 5G can be also integrated as a part of the air interface to facilitate the reconfigurability. Some brand-new feature emerges for link adaptation and also affects the design of 5G air interface. The first is that the link adaptation metric needs to evolve from the continuously pursued SE maximization to multiple metrics, including energy efficiency, user-centric quality of experience, network scalability, and easy upgradability etc. Second, the emerging technologies inherently enable a much richer adaptation space for 5G networks, such as the frame structure, multiple access, waveform schemes and also the parameters for these schemes. The advanced new multiple access schemes, such as SCMA [8], non-orthogonal multiple access (NOMA) [9] and new waveforms [10], e.g., filter bank multi-carrier (FBMC), generalized frequency division multiplexing (GFDM) , unified frequency multi-carrier (UFMC), Filtered-OFDM are brought up and being actively studied to meet the diverse demands of different scenarios.

The development of software defined radio, using the software to implement the signal processing instead of the traditional hardware, offers great opportunities for the soft air interface design. SDAI is proposed as a comprehensive platform for 5G air interface, which can be also interpreted as a ***link adaptation revolution***, which involves multiple fundamental building blocks of the air interface with all programmable parameters, such as the multiple access, waveforms, modulation and coding, spatial processing, etc., as shown in Figure 1. SDAI provides a scalable and configurable mechanism to customize air interface design to support different services and applications under different transmitting and receiving conditions. To enable SDAI, the potential candidate sets for each building block may be first predefined, and then the preferable parameter configurations for each building block (e.g., bandwidth, number of access users, codewords) and best suitable combination of the building blocks could be chosen based on varying service requirements and network/UE capabilities via the unified link adaptation mechanism.

The essential features of SDAI are ***agility and efficiency.*** SDAI should provide flexible composition of sets of radio technologies and case-specific configurable parameters adaptation tailored to use case and business model. All the functions should be programmable and configurable. In addition, to be efficient, the air interface needs to be energy efficient, computational efficient and also highly efficient for fast implementation. It is expected to use uniform architecture as much as possible and share the foundational functionalities supporting different use cases and access technologies, while also allowing advanced functionalities customized for specific demands. To support the agility and efficiency, the feasible configuration sets and adaptation mechanism for SDAI needs to be designed with comprehensive and thorough investigations.

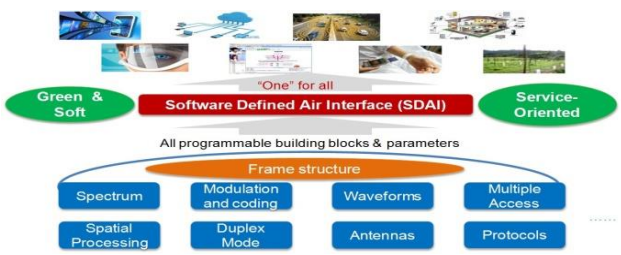


Figure 1 SDAI Framework

3. Promising Enablers & Smart Engine for SDAI

To shed some lights on the design of SDAI, in this section, we will first introduce some promising technology enablers and then propose the concept of "Smart Engine" for SDAI, offering intelligent context-aware link adaptation beyond the pure modulation and coding adaptation in the preliminary generations.

Hybrid Frame Structure

Due to the diversified traffic types, a hybrid frame structure is envisioned to handle a large set of demands in a unified framework. The structure of hybrid frame can support different length of transmit time intervals (TTIs), various delay tolerance, different uplink and downlink traffic ratios, flexible configurations for duplex mode supporting for XDD (a duplex scheme which is a super set of TDD, FDD and full duplex). Figure 2 shows an example of feasible structure of hybrid frame.

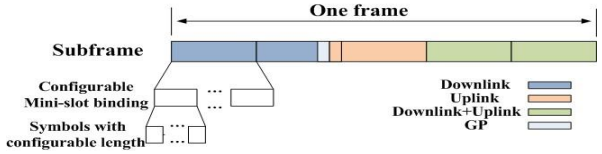


Figure 2 Example of hybrid frame structure

Flexible waveforms

Waveforms, as the vital components of the air interface, are expected to be more flexible in SDAI. Beyond the perfectly synchronized and orthogonal OFDM signals designed for high volume data transmission, the non-orthogonal asynchronous waveforms, like FBMC, UFMC, Filter-OFDM emerge as promising solutions for sporadic traffic in the IoT applications in 5G. The flexible compatible framework for theses waveforms can be based on the carrier/waveform aggregation as shown in Figure 3. Different waveforms located in different carriers can be aggregated in one air interface serving diverse 5G services. The waveform, sub-band bandwidth, sub-carrier spacing bandwidth, filter length and CP length in each wave can be flexibly chosen according to the dedicated scenarios and services.

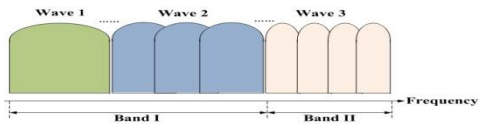


Figure 3 Flexible waveform based on wave aggregation

Adaptive Multiple Access

Non-orthogonal multiple access schemes have attracted considerable attention as a promising candidate for 5G systems since they can efficiently improve the spectral efficiency and accommodate the necessary scalability for massive IoT connectivity. However, the benefits of the typically identified candidates like SCMA, NOMA are usually achieved at the cost of higher signal processing complexity. Specifically, for SCMA, the decoding complexity dramatically increases with the overloading access users and modulation orders. The adaptation among different multiple access technologies or some parameters such as number of codeword of a SCMA codebook, spreading factor, max number of layers, number of nonzero elements of each codeword., based on different system requirements and network/UE capabilities such as coverage, connectivity, SE, EE, can be used to facilitate the SDAI adaptation. In [11], we studied and exemplified the adaptive multiple access scheme from EE-SE co-design perspective as a case study of the SDAI.

Configurable MIMO transmission

Different MIMO technologies perform differently in various channel conditions and antenna configurations. In 4G systems, MIMO mode switch was already introduced to achieve a consistent good performance. In the 5G SDAI, MIMO mode switch will be further enhanced with more new MIMO modes, e.g. the spatial NOMA or the so-called enhanced MU-MIMO [12], and be compatible with the hybrid digital and analog beamforming structure for massive MIMO antenna configurations

Smart Engine

Smart Engine is an essential component of SDAI, which works as a supervisor of link adaptation and a controller of the promising technology enablers and building blocks of SDAI. The decision making process of Smart Engine can be formulated as an optimization problem maximizing the system utility such as throughput, energy efficiency and latency under the constraints of configurable link technology enabler sets and the feasible link parameter sets. The relationship between the smart engine and the promising technology enablers for SDAI is illustrated in Figure 4**.**

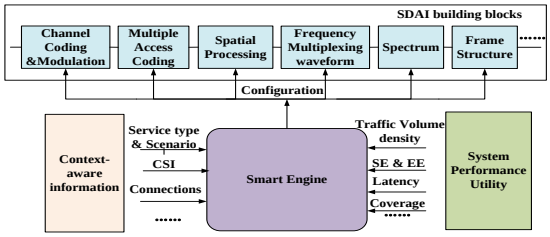


Figure 4 Illustration of smart engine

4. Conclusions

In this paper we have briefly introduced the challenges and proposed a potential paradigm of 5G air interface named software defined air interface. The key technology enablers in SDAI such as hybrid frame

structure, flexible waveforms, adaptive multiple access and configurable MIMO transmission has been pointed out. To facilitate SDAI, smart engine is introduced as an intelligent controller and decision maker for link adaptation and flexible configuration of build blocks and key technology enablers. For future works, it would be valuable to design the feasible configuration parameter sets of SDAI with considerations of the tradeoff between implementation complexity and system performance. In addition, it would be also important to design the utility function of smart engine in accordance with the diverse and perhaps contradicting metrics in 5G.

References

[1] Whiter paper, "5G vision and requirements", China IMT-2020 (5G) Promotion Group, May, 2014.

[2] 5G White Paper, Executive Version by NGMN Alliance, http://www.ngmn.org/uploads/media/141222_NGMN-Executive_Version_of_the_5G_White_Paper_v1_0_01.pdf, Dec. 2014.

[3] Chih-Lin I, C. Rowell, S. Han, et al., "Towards Green & Soft: A 5G Perspective," *IEEE Commun, Magazine*, vol. 52, no. 2, Feb. 2014, pp.66-73.

[4] D. Kreutz, F.M.V. Ramos, P.E. Verissimo, et al., "Software-Defined Networking: A Comprehensive Survey," *IEEE Proceedings.*, vol. 103, no. 1, Jan. 2015, pp. 14-76.

[5] M. Chiosi, D. Clarke, and P. Willis, "Network Functions Virtualization," Oct. 2012.

[6] Whiter Paper, "C-RAN: The Road towards Green RAN," V3.0, 2014, China Mobile.

[7] V. Pejovic, E. M. Belding, "WhiteRate: A Context-Aware Approach to Wireless Rate Adaptation," *IEEE Trans. Mobile Computing*, vol. 13, no. 4, Apr. 2014.

[8] H. Nikopour, H. Baligh, "Sparse Code Multiple Access," *IEEE PIMRC*, 2013, pp. 332-336.

[9] A. Benjebbovu, Y. Saito, Y. Kishiyama, et al., "Concept and practical considerations of non-orthogonal multiple access for future radio access," *IEEE ISPACS* 2013.

[10] G. Wunder, P. Jung, M. Kasparick, et al., "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications," *IEEE Commun, Magazine*, vol. 52, no. 2, Feb. 2014, pp.97-105.

[11] Q. Sun, C.-L. I, S. Han, et al., "Software Defined Air Interface: A Framework of 5G Air Interface," accepted by *IEEE WCNC workshop 5G Arch*, 2015.

[12] RP-140701, 3GPP TSG RAN Meeting #64. New study item proposal: Enhanced MU-MIMO and Network Assisted Interference Cancellation.

Qi Sun received the B.S.E. and Ph.D. degree in information and communication engineering from Beijing University of Posts and Telecommunications in 2009 and 2014, respectively. After graduation, She joined the Green Communication Research Center of the China Mobile Research Institute. Her research interests include MIMO, cooperative communications, green communications and 5G air interfaces.

Chih-Lin I received her Ph.D. in EE from Stanford University. She has led R&D and technology strategy programs in AT&T Bell Labs; the headquarters of AT&T, ITRI of Taiwan, and ASTRI of Hong Kong. She received the IEEE TransCom. Stephen Rice Best Paper Award and is a winner of the CCCP National 1000 Talent program. Currently, she is the Chief Scientist of wireless technologies in China Mobile. She established the Green Communications Research Center, and is spearheading major initiatives including 5G key technologies R&D; high energy efficiency system architecture, technologies and devices; green energy; C-RAN and soft base stations. Dr. I was an elected Board Member of IEEE ComSoc, Chair of the ComSoc Meetings and Conferences Board, and the Founding Chair of the IEEE WCNC Steering Committee. She is currently a member of the GreenTouch Executive Board, the ETSI NFV Network Operator Council, and the WWRF Steering Board, as well as an Adjunct Professor of BUPT. Her research interests focus on Green and Soft; including green communications, C-RAN, network convergence, soft air interface, massive MIMO, and active antenna arrays.

Shuangfeng Han received his M.S. and Ph.D. degrees in electrical engineering from Tsinghua University in 2002 and 2006 respectively. He joined Samsung Electronics as a senior engineer in 2006 working on MIMO, MultiBS MIMO etc. From 2012, he is a senior project manager in the Green Communication Research Center at the China Mobile Research Institute. His research interests are green 5G, massive MIMO, full duplex, NOMA and EE-SE co-design.

Zhengang Pan received his B.S.E. from Southeast University in Nanjing and his Ph.D. degree in E.E. from Hong Kong University. After graduation, he worked for NTT DoCoMo Beijing Labs and ASTRI in Hong Kong working on wireless communication, mobile digital TV, and wireline broadband for system design and terminal SoC chip implementation. He is currently a principal staff member of the Green Communication Research Center at the China Mobile and is now leading a team working on 5G technologies. His research interests include signal processing, space time coding, and cross-layer optimization.

INDUSTRIAL COLUMN: EMERGING TECHNIQUES & APPLICATIONS FOR 5G NETWORKS

Guest Editor: Kan Zheng, Beijing University of Posts & Telecommunications, China, zkan@bupt.edu.cn

The Fifth Generation (5G) mobile networks are expected to provide the significant increases in system capacity as well as spectrum efficiency, compared with the current Long Term Evolution (LTE) and LTE-advanced networks. Both standardization and technology developers are facing the challenge of diverse 5G technological requirements in the provision of 5G services and applications. This special issue of E-Letter focuses on the promising current progresses in 5G research area.

In the first article entitled "I-Net: Pave the Road towards Mobile Internet", *Jianquan Wang, Zhangchao Ma, Lei Sun and Bo Wang* (from China Unicom) propose a new internet-oriented, inter-connected, integrated and intelligent mobile network architecture, namely "I-Net", for 5G mobile networks. This new architecture has many merits as an inter-connected, integrated and intelligent network, which can achieve high efficiency, low cost and high value for Mobile Network Operators.

In the second article entitled "Coverage Analysis and Comparison between 3G and 4G Cellular Systems with mmWave", *Zheng Jiang, Bin Han, Peng Chen, Fengyi Yang, and Qi Bi* (from China Telecom) discuss and compare the coverage of mmWave communication in either 800MHz CDMA or 2.1 GHz LTE systems. They also suggest potential frequency bands for different scenarios.

The third article entitled "Multicarrier-Division Duplexing (MDD): A Duplexing Scheme Whose Time Has Come", by *Lie-Liang Yang* from (University of Southampton, UK) discusses the advantages and challenges of MDD by comparing it with the conventional Frequency-Division Duplexing (FDD) and Time-Division Duplexing (TDD) approaches. The work also considers the implementation of multicarrier modulation/demodulation and provides the approaches for channel estimation/prediction. According to the authors view it is now the time to put the MDD into practice.

The fourth article entitled "D2D Cooperation to Avoid Instantaneous Feedback in Non-reciprocal Massive MIMO Systems" by *Samah A. M. Ghanem and Laura Cottatellucci* (from Eurecom in France) propose a novel approach for fast fading Massive MIMO systems in non-reciprocal channels as FDD mode. The presented approach relies on Device-to-Device (D2D) cooperation to avoid instantaneous Channel State Information (CSI) feedback, unaffordable in practical systems. The authors show that the proposed approach outperforms state-of-art schemes under various user and antenna array distributions.

Finally, the fifth article entitled "Immersive Light Field Based 3D Telemedicine Applications in 5G", by Wei Xiang and Gengkun Wang (from the University of Southern Queensland, Australia) propose an immersive 3D telemedicine system based upon emerging light field technology for 5G applications. The system is able to provide a digitalized 3D model with a high level of realism and at the same time offer an immersive glasses-free 3D experience for doctors and patients situated far apart. It can be used in a wide range of telemedicine applications.

The articles of the Industrial Column provide different viewpoints for 5G, from network architecture to the potential applications. It is believed that 5G will come true in 2020 and totally change our life. I am very grateful to all the authors for making great contribution and the E-Letter Board for giving this opportunity to this Special Issue.

KAN ZHENG [SM'09] (zkan@bupt.edu.cn) is currently a Professor in Beijing University of Posts & Telecommunications (BUPT), China. He received the B.S., M.S. and Ph.D. degree from, BUPT in 1996, 2000 and 2005, respectively. He is author of more than 200 journal articles and conference papers in the field of communication signal processing, resource optimization in wireless networks, M2M, V2V networks and so on. He holds editorial board positions for several journals and also served in the Organizing/TPC Committees for more than 20 conferences such as IEEE PIMRC, IEEE VTC and many others.

I-Net: Pave the Road towards Mobile Internet

Jianquan Wang, Zhangchao Ma, Lei Sun and Bo Wang
China Unicom Group Company, Inc., China
Email: {wangjq, mazc, sunlei49, wangbo149}@chinaunicom.cn

1. Introduction

Mobile Internet is changing our daily life and will be the trend in 5G era. According to Cisco forecast, the mobile data traffic will keep surging globally, which will be 16 Exabyte in 2018, almost 11 folds of 2013 **Error! Reference source not found.**. Smartphone is becoming the main kind of mobile terminals. A lot of internet applications, are migrating from fixed network to mobile network, e.g., mobile commerce, mobile social networking, mobile video, and mobile gaming.

Mobile Internet is based on mobile network infra-structure; however, today's mobile network has not fully taken into account the internet requirements from the perspective of architecture design. In face of boosting data traffic and new internet service modes, traditional mobile network, featured vertical shaft, isolated islands and application-specific architecture, is under multilateral pressures from network capacity, user experience and CAPEX/OPEX. In order to be sustainable and flourish in the mobile internet era, this paper proposes the new internet-oriented, inter-connected, integrated and intelligent mobile network architecture, namely "I-Net", for 5G.

This paper includes three parts. In part one, we look at the challenges brought by mobile internet and analyze the main problems of today's mobile network. In the second part, I-NET and its solutions are introduced. Finally, we make the concluding remarks.

2. Mobile Internet Challenges and Network Status

Besides the opportunities brought by mobile internet, it also brings about new challenges for mobile network. Firstly, mobile internet requires dramatically larger capacity. IMT-2020 WG predicts [2] that in 2020 when 5G is planned to be deployed, the average user data rate will boost from around 20 Mbps to 1000 Mbps, and the connection density will increase from 140,000/km2 to 6,000,000/km2, which means the area capacity will increase 2000 folds compared to now.

Secondly, mobile internet brings in brand new traffic characteristics, where everyone need personalized, localized, mobile service and desire to share everything using mobile network, known as the SoLoMo (Social, Local and Mobile) mode.

In addition, mobile Internet requires substantially lower cost per bit for operators. According to the forecast [2], in 2020 the cost per bit has to be reduced to 200 times lower than today. This will impose enormous burden for mobile network operators (MNOs), which implies dramatically reducing network CAPEX and OPEX.

Facing these challenges, current GSM/UMTS/LTE mobile network architecture (Fig.1) defined by 3GPP [3] has several limitations, as summarized below.

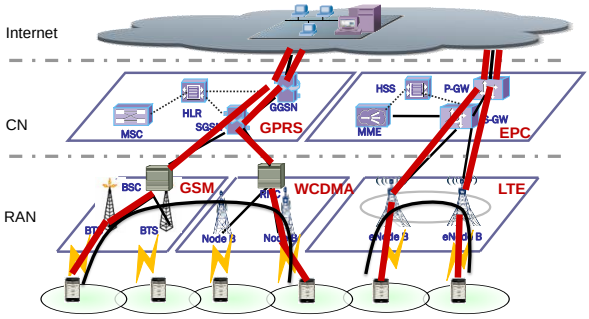


Figure 3 Conventional Network Architecture

2.1 Vertical Shaft Structure

All data flows have to go through the core network (CN) as shown in Fig. 1. However, in face of the enormous traffic volume and new service pattern of mobile Internet, this architecture will face severe problems. First, the pressure from radio side will be accumulated to upper layers, leading to congestion and endless expansion of core and transport network. Second, the centralized transmission route will cause unnecessary circuitous route and delay, when delivering localized interactive mobile internet service.

2.2 Isolated Island Modes

The whole mobile network exist islanding effect in many aspects. First, there are more than 10 kinds of independent standards, including 2G, 3G, 4G and WLAN, causing complex interoperability issues. Second, a lot of frequency bands co-exist in a fragmented manner. In China, 6-7 bands have been allocated, while each band runs independent system. In addition, the millions of base stations (BSs) deployed by operators is another form of islands. The BSs are "loosely connected" where interaction between them is limited to handover function, while effective radio signal cooperation and resource sharing are not possible, causing inconsistent user experience and unbalanced resource usage rate among cells.

2.3 Application-Specific Equipment

The whole mobile network, from BSs to the transport network to the core network, and to the service chains, all adopt application-specific telecom equipment, lack of flexibility and scalability, leading to complex and expansive upgrading and expansion. And when operators need to deploy new service, it usually takes very long period, and involves multiple network elements. In the era of mobile Internet, it is hard to respond to the rapidly changing needs of users.

3. New Mobile Internet Architecture: I-Net

In order to embrace the opportunities of mobile internet, the new mobile network architecture, namely "I-Net", i.e., internet-oriented network, is proposed to meet the new requirements, which featured three "I" concepts.

- **Inter-connected**

To break through the vertical shaft structure, I-Net will construct enhanced direct channel between mobile BSs, achieving effective inter-connection. Thus, the mobile network can adapt to internet service via flexible traffic steering.

- **Integrated**

To break the boundaries between standards, frequency bands, and the BSs, I-Net aims to integrate all available resources, and achieve the target of "One network", "One Carrier" and "One consistent experience".

- **Intelligent**

To reduce the cost and complexity of network construction and maintenance, I-Net will implement intelligent resource control and management, and flexible service deployment, via network virtualization.

Three features of I-Net and respective solutions are next introduced in detail.

3. 1 Inter-connected network adapts to traffic

To make the mobile network adapt to internet traffic, the mobile network has to be enhanced in two aspects as shown in Fig. 2. First, the data forwarding function is moved down to realize localized data transmission within or between BS(s), and thus Internet traffic can be absorbed within the mobile network via flexible traffic steering for interactive mobile services. Moreover, part of core network functions, e.g. mobility, charging and security need to be moved down to BS, to guarantee telecom level control and management for local service.

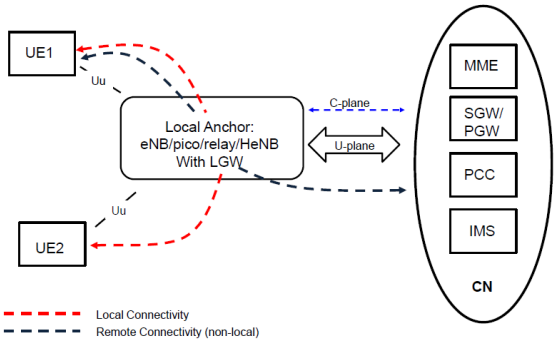


Figure 4 Inter-connected network architecture

This inter-connection ability can be implemented in different ways. The short-term feasible scheme is to deploy local gateway (L-GW) entity to local network, and this entity will be responsible for the local data forwarding operation. This scheme has least impact on existing mobile network and has mature standards, i.e., LIPA/SIPTO [4] to support. But it lacks enough flexibility and only areas where L-GW deployed can achieve local traffic offloading.

The long-term solution is based on SDN (Software defined network) and NFV (Network function virtualization). The basic principle is to first virtualize the BS and CN functions over physical network elements, and thus data forwarding between physical BS entities is transformed to virtual function entities, and the low layer data route can be automatically optimized via SDN functions. This scheme can truly realize flexible local traffic steering anywhere. But it still needs a lot of work, e.g., to define new standards and to develop new virtual devices.

With this inter-connected feature, the circuitous routing can be avoided and end-to-end delay can be reduced in the scenarios where local traffic is intensive, such as enterprise, government and campus; moreover, many proximity-based services can also be benefitted, such as mobile social networking. Another benefit is that it can make the service content deployed closer to the end user. For example, the CDN service or the local advertising content can be deployed beside BSs to ensure users nearby can get better experience.

3.2 Integrated Network Unleash Full Power

The co-existence of multi-standards, multi-bands, and millions of loosely connected BSs is the current situation. I-Net concerns about how to break the barriers between multi-systems, multi-bands and multi-BSs, and to integrate all kinds of resources, in order to maximize the value of the network. As shown in Fig. 3, the integrated network includes three aspects.

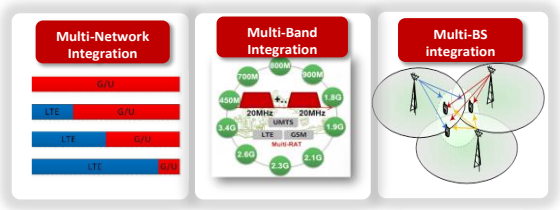


Figure 5 Integrated network aspects

- **Multi-Network Integration**

The first step is to integrate multiple systems. To current operator who usually owns multiple networks, the most important things are to provide consistent interoperation experience for users roaming among networks and to perform smooth spectrum refarming for network evolution. The solution of multi-system integration is based on united radio access controller (URC). The principle is to introduce a general control node over the 2G/3G/4G BSs and WLAN access points; then it can collect all information across the networks, including cell load, interference, service status and channel quality, etc. Thereafter, it is possible to carry out coordinated resource management, which can steer the UE to the network with best experience; it can also provide smooth spectrum refarming through dynamically allocate spectrums of 2G/3G to LTE on a cell by cell basis according to load variation.

- **Multi-Band Integration**

The second aspect is to integrate multiple frequency bands. When the networks reduce to one, this band integration can be easy, since the solution has already been supported by LTE-Advanced **Error! Reference source not found.**. Via carrier aggregation, it already supports up to 5 FDD bands' integration; soon it will also support TDD and FDD band aggregation. In future it may even support the integration of cellular and unlicensed bands where WLAN used to work on. Above all, the goal of I-Net multi-band Integration is to build one open carrier, which can integrate all spectrum resources available, maximizing the ability of the network.

- **Multi-BS integration**

The third aspect is to integrate BSs. The traditional loosely connected BSs have two problems: the signal interference between neighboring sites and the difficulty to achieve multi-BS load balance and resource sharing. The solution to BS integration is BBU pooling. When the baseband processing units (BBUs) of BSs are inter-connected, coordination required information can be exchanged to form a BBU pool. Thereafter, multiple BSs can cooperate to send and receive radio signals to reduce interference, namely the CoMP technique [5]; moreover, when high speed interface supported, multiple BSs can share processing resources with each other to achieve load balance.

3.3 Intelligent Network Adds Efficiency and Value

Intelligence is the most important target of I-Net. It is envisioned to greatly reduce the network cost, improve efficiency and enable more flexible service deployment and even can bring new values to network. Network virtualization is deemed to provide the foundation for Intelligent Network. Based on virtualization, opening network capabilities will be a powerful tool for operators to further explore the value of network.

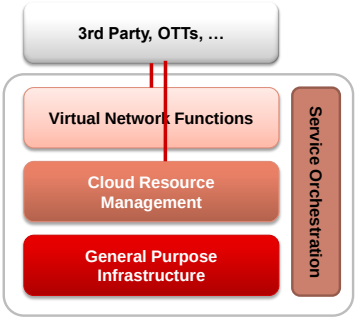


Figure 6 Intelligent Network Architecture

A basic model for intelligent network is provided in Fig. 4. In the physical layer, general purpose processor based hardware will replace traditional application-specific hardware. Above the physical layer is the cloud resource management layer, where all network resources, including communication, computation and storage, can be efficiently scheduled and shared. Above the cloud management platform is a virtual network function layer, where the traditional functions of the network, such as EPC, IMS, RAN, will be implemented with pure software, loaded in the cloud platform. This virtual network device is capable to share information and control abilities to 3rd Party via open interface, to add value of mobile network. Finally, the whole virtual network can realize quick and flexible service deployment. The flourishing of new services based on intelligent network will be the key for operators to win in the era of mobile internet. Then, the three main aspects of intelligent network will be explained in detail.

- **Network Virtualization**

Using GPP-based hardware platform plus software defined network function, can greatly improve the network flexibility and scalability, effectively reduce the network cost, and improve network efficiency. But the virtualization of the entire mobile network is not easy. From base station to transport to the core network to the service chains, often involves very complex and real-time computing. Thus, the virtualization of Mobile network needs to take a step by step path. Service chain, BSS, IMS core and EPC core are by far was the forerunners of virtualization. Virtual base station still

has a far way along the road of commercialization, since radio signal processing poses extremely strict requirements on delay, computation capacity and energy consumption. But it's still the future direction to build a pure virtual mobile network.

- **Open Network Capability**

By providing an open network interface API to 3rd party service developers, the network scan share some of the network information, on the premise to guarantee the private information security. This can enable faster service deployment, and create new values for network. In fact, the big data what network operators own is a real golden hill. For example, if the network can share the mobile users' positions around the shopping mall (hidden identity information) to third parties businesses, they can use the data to analyze the users' shopping habits, and further customize their products and marketing. Moreover, if the QoS and charging control functions is shared to third party, e.g., Amazon, then they can quickly provide accelerated web access speed and even free data traffic to their customers at certain shopping place and certain shopping promotion period, which can help improve OTT competitiveness.

- **Mobile CDN services**

Let take CDN as the example of new services brought by network intelligence. It is well known that CDN is one of the most important means to alleviate the pressure of the backbone network, and to improve the user experience. However, for mobile networks, traditional CDN services can only be deployed outside operator's gateway. Therefore, the traffic flow within the mobile network cannot enjoy the congestion alleviation and accelerating effect; in addition, the interface between mobile network and CDN server is private and defined by CDN service providers, resulting in inefficiency when coordinated with mobile network. The so-called mobile CDN scheme is designed to deploy internal cache within mobile network near the base station side, and then it can alleviate the pressure of the mobile transport and core network, and improve the end user experience. Moreover, it can enhance the differentiated QoS provision capability for MNOs, and thus provide one powerful means to cooperate and compete with OTTs in the era of mobile Internet.

4. Conclusions

The path of mobile network evolution toward mobile internet and the future so-called 5G era is long-term and full of difficulty. This paper intends to demonstrate the current challenging situation and the importance to break through traditional network constrains, including vertical shaft, isolated island and inflexible structure.

The proposed new mobile network, I-Net, which combines the advantages of inter-connected, integrated and intelligent network, is designed to achieve higher efficiency, lower cost and higher value for MNOs.

In order to flourish in the mobile internet era, we sincerely invite the telecom family, including academic and industry elites to endeavor on I-NET vision for a more sustainable and healthy eco-system in future.

References

[1] Cisco, "http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html", 2014.

[2] IMT-2020, "http://www.miit.gov.cn/n11293472/n11293832/n15214847/n15218234/16090851.html", 2014.

[3] 3GPP TS 23.002, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Network architecture", 2014

[4] 3GPP TR 23.829, "LIPA SIPTO Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO)", 2011.

[5] 3GPP TR 36.815, "Further advancements for E-UTRA; LTE-Advanced feasibility studies", V9.1.0, section 5.3.

[6] 3GPP TR 36.819, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Coordinated multi-point operation for LTE physical layer aspects", V11.0.0

Jianquan Wang is the manager of technologies within network construction department of China Unicom. He is dedicated to the deployment of HSPA+ and LTE network, responsible for mobile network technology strategies. He received the Ph.D degree from Beijing University of Posts and Telecommunications (BUPT) in 2003. His research interests include network technology and next generation mobile network.

Zhangchao Ma works for the department of network construction at China Unicom. His focus is on LTE and UMTS technologies and standards evolution. He received the B.S. and Ph.D degree from BUPT in 2006 and 2011.

Lei Sun received the B.S. and Ph.D degree from BUPT in 2006 and 2011. He works for the department of network construction at China Unicom. His research interests lie in the field of mobile communication technologies.

Bo Wang received the B.S. and Ph.D degree from BUPT in 2006 and 2011. He works for the department of network construction at China Unicom. He is focus on UMTS/LTE radio network technologies.

Coverage Analysis and Comparison between 3G and 4G Cellular Systems with mmWave

Zheng Jiang, Bin Han, Peng Chen, Fengyi Yang, and Qi Bi
Technology Innovation Center, China Telecom Corporation Limited, China
Email: jiangzh@ctbri.com.cn

1. Introduction

As the increasing popularity of smart phones and other mobile data devices such as Pads, wearable user devices and e-Books, user needs always online experience. The growth of mobile data traffic is predicted at 61 percent compound annual growth rate (CAGR) in 2013-2018[1], and the next generation (5G) is expected to meet this rapidly increasing traffic demand of the future. As the current fourth-generation (4G) systems including LTE and Mobile WiMAX have already use many advanced technologies to improve air interface capacity, there is limited room for further spectral efficiency improvement. Ultra-dense small cell deployment is considered as an approach to provide improvements on network capacity. However, it can only improve the network capacity linearly with the number of cells. Only depending on small cell is not able to meet the orders of magnitude capacity increase. Therefore, the allocation and use of new spectrum must be considered.

Since the operating mobile systems, e.g., 2G/3G/4G, have been deployed in 600MHz-3GHz frequency band, there is almost no room to allocate the new spectrum in sub-3GHz frequency band. On the other hand, a vast amount of spectrum in higher frequency, especially the 6–100GHz range, remains underutilized. Therefore, the new cellular systems design in 6–100GHz, referred as mmWave cellular communication systems, has recently gained significant interest [2-4].

The mmWave frequency bands have not favorable propagation characteristic as traditional cellular frequency bands below 6GHz. There are only short-distance and fixed wireless communication mmWave system on deployment. Compared with these systems, cellular systems need to provide larger coverage area and serve more number of users. Moreover, the operator controlled cellular network is expected to give the stable and consistent services to users in cell coverage area. Therefore, if mmWave frequency band is explored for cellular systems, the coverage of mmWave cellular systems has to been investigated. In this paper, the coverage of mmWave cellular systems is firstly analyzed and compared in 3G (CDMA) and 4G (LTE) systems. Then, the scenario and use case for mmWave cellular system are proposed.

2. Key factors for mmWave system coverage

There are three key factors to affect the coverage of a mmWave cellular system significantly, i.e., propagation characteristic, antenna gain and transmission power. In this section, we will analyze them, respectively.

2.1 Propagation characteristic

Generally, the propagation of mmWave signals is affected more seriously by oxygen and water vapor absorption than the sub-3GHz frequency band. The gaseous attenuation is given by [5]:

$$\gamma=\gamma_0+\gamma_w \quad \text{(dB/km)} \qquad (1)$$

where γ_0 and γ_w are the specific attenuations (dB/km) due to dry air (oxygen, pressure-induced nitrogen and non-resonant Debye attenuation) and water vapor, respectively. We can justify that the gaseous attenuation is less 2dB/km for most mmWave frequency band (6-100GHz) except around 60 GHz. The signal propagation in 57–64GHz band can experience about 15 dB/km attenuation as the oxygen molecule (O2) absorbs electromagnetic energy at around 60 GHz. Basically, excluding the oxygen and water absorption band (around 60GHz), we can ignore the effect of gaseous attenuation in 6-100GHz frequency band on the coverage of mmWave cellular systems.

In addition to the oxygen and water absorption, the penetration loss is another important factor to affect the coverage. The penetration loss effect of lower frequency and higher frequency on system coverage varies with different building material [6-9] are measured and presented in Table 1.

Table 1. Different material penetration losses

Material	**Attenuation (dB)**	
	<3GHz	**40GHz**
Clear glass(thickness 0.3/0.4cm)	6.4	2.5
Wood(thickness 0.7cm)	5.4	3.5
Concrete(thickness 10cm)	17.7	175
Foliage(depth 10m)	9	19

It can be seen that for wood and clear glass material, the penetration loss in 40GHz band is lower 2-3dB than that in sub-3GHz band; but for concrete and foliage, the penetration loss in higher frequency band is very seriously, which may severely reduce the coverage of mmWave cellular system.

Raindrop also bring attenuation to the mmWave signal propagation, the signal attenuation level varies with rain rates. Light rain at 2.5mm/h rate yields just about 1dB/km attenuation. So it is ignored in our coverage analysis.

The reflection and diffraction also cause signal power loss, the loss level depends on the NLOS environment. In link budget, Path Loss Exponent (PLE) values in the lower and higher frequency band are different to reflect the attenuation effect of reflection and diffraction on the system coverage.

- **Antenna gain**

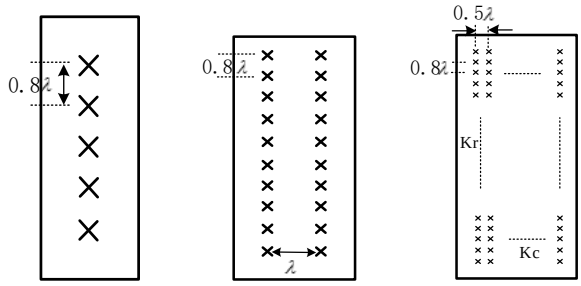
As the wave length of mmWave band is shorter and comparable with the size of the dust particles in the air, the signal propagation in mmWave band lacks the capability to bypass obstacles to spread uninterruptedly, thereby causing the signal power attenuation and being a disadvantage for system coverage. However, the short wave length also brings advantages for coverage. With the same antenna element gain, the size of the half-wavelength dipole in mmWave band is much smaller than that of traditional cellular frequency band. It means that more antenna elements can be packed into the same size antenna array of mmWave band than of sub-3 GHz band.

In real network environment, the antenna size of CDMA in 800 MHz is about 1500x260x100mm, containing one column of X polar antennas, and ten antenna elements with the antenna gain of one column elements of 15 dBi, as shown in Fig. 1 (a). For 2.1GHz LTE, the antenna size is about 1400x320x80mm, containing two column of X polar antennas, and forty antenna elements with 18dBi antenna gain of each column, as shown in Fig.1 (b). These commercial antenna products reveal that in the higher frequency, we can achieve the higher gain of antenna array. In this sense, if antenna size in 40 GHz mmWave band keep the same with that of LTE, the antenna array can contain 57 (K_c=57) columns of X polar antenna with 180 (K_r=180) antenna elements of each column, as shown in Fig.1 (c), each column antenna gain can increase to about 26dBi, the extra antenna gain can be used to overcome the path loss attenuation of mmWave signal propagation.

The similar antenna gain of mmWave system can be achieved in UE side. As the wave length of 40 GHz is only one-twentieth of that of 2 GHz, the more antenna can be packed in terminals. If considering 1/2 wave length antenna isolation, up to 10 dB extra antenna gain can be obtained in terminal side in mmWave band. Actually in current commercial UE, the antenna size can be 1/4-1/8 wave length depending on terminal design [10]. Therefore in mmWave band, extra antenna gain in terminal side may not be as significant as that in base station side.

- **Transmission power**

In current RFIC industrial level, the power efficiency of radio frequency power amplifier (RFPA) in mmWave band is about 10-20% while the power efficiency of RF PA in sub-3GHz band is about 40%. It means that with same transmission power, the power consumption of mmWave cellular system is larger than that of traditional cellular system. With current technology, the high cost and low power efficiency of mmWave RF devices is challenge.



(a) CDMA (b) LTE (c) mmWave

Figure 1 Antenna topology

3. Coverage comparison

In this section, link budget analysis is performed to compare the downlink (DL) coverage between mmWave cellular system and CDMA/LTE system.

The free space propagation model is given below [13]:

$$PL(d)=\overline{PL(d_0)}+10n\log(\frac{d}{d_0})+X_\sigma \qquad (2)$$

$$d=d_0 10^{\frac{PL(d)-\overline{PL(d_0)}-X_\sigma}{10n}} \qquad (3)$$

where X_σ (dB) is a zero-mean Gaussian distributed random variable with standard deviation σ (dB). With 90% coverage probability for urban area and σ =8 dB, X_σ =10.25 dB. The variable n is the path loss exponent which indicates the rate of the path loss increasing with distance. In NLOS environment, n=4 with f_c=800MHz and 2.6 GHz [12-13], n=4.5 with f_c =40GHz [4], d_0 is the reference distance at the far field of the antenna with value of 1 m in this paper, and d is the T-R separation distance.

$$PL(d_0)=20\log(\frac{4\pi d_0}{\lambda}) \qquad (4)$$

where λ is the carrier wavelength.

As shown in Table 2, the link budget for 800 MHz CDMA, 2.1 GHz LTE and the mmWave system in 40 GHz is performed and PL(d) in NLOS environment is calculated. By using equation (3) and (4), the coverage of CDMA, LTE and mmWave system can be obtained, i.e., about 719 m, 498 m and 52 m, respectively. Even with other approaches which can improve the coverage

of mmWave system, such as reducing system bandwidth, implementing beam-forming, the coverage gap between CDMA/LTE and mmWave systems still exists obviously. For instance, if the bandwidth of 40GHz mmWave system is reduced to 500 MHz, and 17dB (10log(K_c)) beam-forming gain can be achieved, the coverage of 40 GHz mmWave system can be improved to 144.6 *m*. Even so, under the same transmission power condition, if we want to improve the coverage of 40GHz mmWave system to the same size as 2. 1GHz LTE, there still needs the additional 24.2dB gain.

4. Scenario and use case discussion

Based on above coverage analysis, we can see that mmWave cellular systems, whose coverage is limited compared with traditional cellular system, is not suitable for contiguous coverage in outdoor or NLOS environment, especially outdoor-to-indoor (O2I) coverage or indoor-to-outdoor (I2O) coverage. In real network, inter-site distance (ISD) of CDMA in dense urban of typical developed city in China is about 450 *m*, ISD of LTE is about 350 *m*. If deploying mmWave system to provide the contiguous coverage, the number of mmWave base station deployed shall be triple that of LTE. On the contrary, the mmWave system is more suitable for providing coverage in LOS environment in case of PLE is 2.3 [12-13], and the cell coverage is increased to 2 *km*.

Therefore, the suitable scenarios for mmWave cellular system deployment are proposed as follows:

1) Virtual office
   In this scenario, office/team members distributed on different rooms or buildings need to communicate with 3D telepresence and multimedia devices, whose data-rate needed is around 1 Gbps [14].
2) Dense urban information society
   Users in dense urban areas require high communication data rates, they thus create local concentrations of very high data traffic demand. 500 Gbyte traffic per month per user is assumed. 95% area coverage with 100 Mbit/s is assumed to be needed [14].
3) Shopping mall
   With mobile network and manifold wireless sensor network deployment in a large shopping mall, high density of customers together with staff of shops and wireless sensors will create extended rich communications in indoor environment.
4) Stadium
   A lot of people watching an event such as sports events or live concerts in stadium are interested to high quality video contents vision and exchange.

Table 2. DL coverage comparison

Parameter	**CDMA**	**LTE**	**mm Wave**	
Carrier frequency (Hz)	800M	2.6G	40G	
System bandwidth	1.25M	20M	1G	
DL allocated Bandwidth (Hz)	1.25M	2.16M (12PRB)	1G	
Tx power (dBm)	43	39.8	46	a
Tx antenna gain (dBi)	15	18	26	b
Tx cable loss (dB)	2	0.5	0.5	c
EIRP (dBm)	56	57.3	71.5	d=a +b-c
Thermal PSD (dBm/Hz)	-174	-174	-174	
Thermal Noise (dBm)	-113.03	-110.66	-84.00	e
Noise figure(dB)	8	8	8	f
Date Rate (bps)	307.2K	1M	1G	
SINR (dB)	4	1.32	3.56	g
Receiver sensitivity (dBm)	-101.03	-101.34	-72.44	h=e +f+ g
RX antenna gain (dBi)	0	0	10	i
Cable loss (dB)	0	0	0	j
Minimal receive power (dBm)	-101.03	-101.34	-82.44	k=h- i+j
Body loss (dB)	2	2	2	l
Penetration loss (dB)	0	0	0	m
Shadow fading deviation (dB)	8	8	8	
Coverage probability (dB)	90%	90%	90%	
Shadowing fading margin (dB)	10.25	10.25	10.25	n
Interference link margin (dB)	0	0	0	o
PL(d)(dB)	144.78	146.38	141.69	PL= d-k- l-m- n-o

5) Open air festivals
   The scenario is an outdoor festival where an average of 100.000 users generates a traffic volume of 900 Gbps/km2 [14].

In these five scenarios, with low mobility users and high probability for LOS environment, the requirement of user data traffic is more than hundreds of megabytes, and need large chunks of underutilized spectrum to meet. Only mmWave band can support this requirement.

If the bandwidth requirement is about hundreds of megabytes, the frequency band may be found below or around 10 GHz. For example, 5.925-6.425 GHz or 9.9-10.6 GHz could be an option.

If the bandwidth requirement is more than one gigabytes, the frequency band has to be high over 20 GHz. For example, 40.5-42.5 GHz or 47.2-50.2 GHz could be an option.

As for spectrum access cost, a mixture of dedicated licensed access (mobile coverage), licensed shared access (LSA) and license-exempt approach is considered appropriate to keep OPEX/CAPEX for operators low.

5. Conclusions

In this paper, the coverage of mmWave cellular system is analyzed and compared with that of 800MHz CDMA and 2.1GHz LTE system deployed in real network. Based on this comparison, the suitable scenario and use case for future mmWave system deployment are presented. The potential frequency band to different scenario with the frequency requirements and frequency spectrum access scheme are also suggested.

References

[1] White Paper C11-481360, "Cisco Visual Networking Index: Forecast and Methodology, 2013-2018," June 2014.

[2] T. Kim et al., "Tens of Gbps Support with mmWave Beamforming Systems for Next Generation Communications," IEEE GLOBECOM'13, Dec. 2013, pp. 3790–95.

[3] W. Roh et al., "Millimeter-Wave Beamforming as an Enabling Technology for 5G Cellular Communications: Theoretical Feasibility and Prototype Results," IEEE Communications Magazine, Feb. 2014, pp. 106-113.

[4] T. Rappaport et al., "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!" IEEE Access, vol.1, May 2013, pp. 335–49.

[5] ITU-R P.676-8 Attenuation by atmospheric gases.

[6] C. R. Anderson and T. S. Rappaport, "In-Building Wide-band Partition Loss Measurements at 2.5 and 60 GHz," IEEE Trans. Wireless Commun., vol. 3, no. 3, May 2004.

[7] K. C. Allen et al., "Building Penetration Loss Measurements at 900 MHz, 11.4 GHz, and 28.8 GHz," NTIA rep. 94-306, May 1994.

[8] Z. Pi and F. Khan, "An Introduction to Millimeter-Wave Mobile Broadband Systems", IEEE Communications Magazine, June 2011, pp.101-107.

[9] A. Alejos, M. G. Sanchez, and I. Cuinas, "Measurement and Analysis of Propagation Mechanisms at 40 GHz: Viability of Site Shielding Forced By Obstacles," IEEE Trans. Vehic. Tech., vol. 57, no. 6, Nov. 2008.

[10] A. Diallo, et al., "Efficient two-port antenna system for GSM=DCS=UMTS multimode mobile Phones," ELECTRONICS LETTERS, 29th March 2007, Vol.43 No.7.

[11] C. H. Doann, "Millimeterwave CMOS Design," IEEE J. Solid-States Circuits, Jan. 2005.

[12] ITU-M.2135-Guidelines for evaluation of radio interface technologies for IMT-Advanced, ITU-R, 2008.

[13] T. S. Rappaport, "Wireless Communications: Principles and Practice (2nd Edition)", Prentice Hall, Jan., 10, 2002.

[14] ICT-317669-METIS/D5.1, "Intermediate description of the spectrum needs and usage principles," Aug. 2013.

Zheng Jiang is a senior engineer in the Technology Innovation Center, China Telecom Corporation Limited, China, where he leads the multi-cell multi-layer coordination communication research of National Science and Technology Major Project. He received M.S., and Ph.D. degrees in communication and information system from Beijing University of Posts and Telecommunications, Beijing, China.

Bin Han received the Ph.D. degree at Beijing University of Posts and Telecommunications (BUPT), China. He is currently working at China Telecom Corporation Limited Technology Innovation Center. His research interests include high frequency communications, cooperative communications.

Peng Chen received his Ph.D. degree in electronic engineering in 2006 from Beijing University of Posts and Telecommunications. He has been with China Telecom Innovation Center since 2010, working on LTE-A standardization in 3GPP RAN and cutting-edge technology for IMT-2020.

Fengyi Yang is the Vice-President of China Telecom Technology Innovation Center. He is the expert group member of national key project and 863 "5G project", Vice Chairman of CCSA TC5. His research interests include mobile communication technology and service application.

Qi Bi is the President of the Technology Innovation Center and the CTO of Beijing Research Institute of China Telecom. He received his M.S. from Shanghai Jiao Tong University and Ph.D. from Pennsylvania State University. Prior to joining China Telecom in 2010 he worked at Bell Labs for 20 years and became a Bell Labs Fellow in 2002. He is an IEEE fellow.

Multicarrier-Division Duplexing (MDD): A Duplexing Scheme Whose Time Has Come

Lie-Liang Yang
School of ECS, University of Southampton, SO17 1BJ, United Kingdom
Email: lly@ecs.soton.ac.uk

1. Introduction

In recent years, research and development in wireless communications have put a lot of emphasis on the efficient use of the available resources via employment of high-flexibility and intelligent wireless techniques, so that wireless systems can be spectral and energy efficient, and have the highest possible flexibility for implementation and operation. Currently, wireless communications are evolving towards the 5th generation (5G). Many new technologies, such as millimeter and visible light communications, full duplex transmission, massive MIMO, etc., have been considered as the technical options. When we are looking forward to the future, we have the vision that the future wireless communication systems are merging into a highly complex and comprehensive network, in which resources are well shared among different systems and wireless terminals/devices, many wireless systems are multicarrier systems, and wireless transmission distance between two terminals/devices is becoming short.

On the other side, a lesson we learned from MIMO theory [1] is that, in order to attain the highest possible capacity, especially, of the MIMO systems where transmitters have a large number of degrees of freedom, such as in massive MIMO systems [3], transmitters are required to exploit the channel state information (CSI) of outgoing channels, which is usually estimated with the aid of the reciprocity from their corresponding incoming channels. Furthermore, in order to carry out resource allocation and other types of scheduling in wireless networks, wireless transmitters often require the CSI of outgoing channels. All these request that future wireless transceivers employ the high-efficiency and high-feasibility mechanisms to provide reliable channel estimation/prediction. Explicitly, both the FDD and TDD are incapable of meeting all the requirements demanded by such future wireless systems.

Against the background, considering the development of wireless communications systems and the challenges from the future wireless systems, in this paper, we recommend the MDD scheme [3], which employs the advantages of both the FDD and TDD schemes, and also some unique merits. Specifically, we first analyze the advantages of the MDD as well as the challenges it may face by contrasting it to the conventional FDD and TDD schemes. Then, the principles of the MDD are discussed through its implementation in the principles of OFDM. Furthermore, we provide some channel estimation/ prediction approaches to show the high-flexibility of channel estimation/prediction in MDD-based wireless systems.

2. Multicarrier-devision duplexing

The principles, advantages and disadvantages of the FDD and TDD can be found in many references, such as, in [3]. The concept of MDD has originally also been stated in [3]. The principles of MDD can be readily understood with referring to Fig. 1, where 1/4 of sub-bands are allocated for supporting uplink transmission, while the rest 3/4 of sub-bands are allocated for supporting downlink transmission. Generally, both the uplink and downlink channels in MDD are operated within the same frequency band. A fraction of sub-bands (subcarriers) are allocated for supporting uplink (incoming) transmission, while the others for downlink (outgoing) transmission. Usually, the subcarrier signals are chosen to be orthogonal with each other. In MDD, according to the practical requirements, the number of sub-bands allocated to the uplink or downlink of a user can be fixed or dynamical. The number of sub-bands allocated to a user can also be different from that allocated to another user. Explicitly, MDD belongs to the family of FDD.

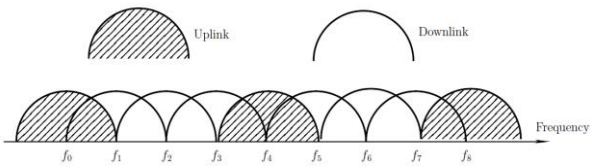


Figure 1 Illustration of multicarrier-division duplexing (MDD)

It can be shown that MDD employs all the advantages of TDD. First, MDD is flexible for supporting asymmetric and variable traffics for uplink and downlink. This can be readily achieved by dynamically allocating the corresponding number of subbands. Second, in MDD-based systems, the channel knowledge required for carrying out transmitter preprocessing can be readily obtained with the aid of F-domain channel estimation or prediction. MDD also inherits the merits of FDD. For example, under the FDD, there is no switch-over between transmission and receiving. Under the MDD, there is no or only very low chance of switch-over between transmission and receiving. Furthermore, MDD may employ some

merits that TDD and FDD are incapable of providing. MDD has the highest flexibility for design or online reconfiguration, since, in comparison with TDD and FDD, MDD employs a higher number of parameters, which can be adjusted according to the requirements in practice.

One typical challenge with MDD is the possibly added intercarrier interference resulted from frequency and time offsets. This may significantly degrade the achievable performance of MDD systems, when channel fading becomes time-selective or when there is a big frequency offset. However, nowadays, there are many signal processing algorithms available for mitigation of interference. Furthermore, the vision about future wireless communications systems is that most wireless services will be provided by short distance signal transmission. In these wireless systems, synchronization among wireless terminals becomes relatively accurate, which may substantially relieve the problem of inter-carrier interference in multicarrier systems.

3. Implementation of multicarrier-division duplexing

In this section, we propose an approach in the principles of OFDM for implementation of MDD-based systems. Assume a multicarrier system with M subcarriers. Let $\boldsymbol{F}$ be a ($M \times M$) discrete Fourier transform (DFT) matrix, which has the property of $\boldsymbol{F}\boldsymbol{F}^H = \boldsymbol{F}^{\mathrm{H}}\boldsymbol{F} = \boldsymbol{I}_M$. Let us assume that N out of the M subcarriers are allocated for supporting uplink transmission, while $\bar{N} = M - N$ subcarriers are allocated for downlink transmission. Let the reduced-DFT (RDFT) matrices for the uplink and downlink be defined as $\boldsymbol{F}_U$ and $\boldsymbol{F}_D$, respectively. The RDFT matrix $\boldsymbol{F}_U$ is constructed from $\boldsymbol{F}$ by keeping only the columns corresponding to the subcarriers allocated to uplink, but removing all the columns corresponding to the subcarriers for downlink. Similarly, the RDFT matrix $\boldsymbol{F}_D$ can be obtained. Explicitly, the RDFT matrices have the properties:

$$\boldsymbol{F}_U^T\boldsymbol{F}_U^* = \boldsymbol{I}_N, \boldsymbol{F}_D^T\boldsymbol{F}_D^* = \boldsymbol{I}_{\bar{N}}, \tag{1}$$

$$\boldsymbol{F}_U^T\boldsymbol{F}_D^* = \boldsymbol{0}, \boldsymbol{F}_D^T\boldsymbol{F}_U^* = \boldsymbol{0} \tag{2}$$

where $\boldsymbol{0}$ represents a matrix with entries of zeros.

Let $\boldsymbol{F}_U^{(k)}$ and $\boldsymbol{F}_D^{(k)}$ be two RDFT matrices for supporting uplink and downlink transmissions of user k. Depended on the uplink data rate of user k, $\boldsymbol{F}_U^{(k)}$ may contain some or all of the columns of $\boldsymbol{F}_U$. Similarly, depended on the downlink data rate of user k, $\boldsymbol{F}_D^{(k)}$ may contain a fraction or all of the columns of $\boldsymbol{F}_D$. In order to facilitate the channel estimation and prediction using the reciprocal principles, the columns chosen from $\boldsymbol{F}_U$ or $\boldsymbol{F}_D$ should be evenly distributed. It can be readily shown that $\boldsymbol{F}_U^{(k)}$ and $\boldsymbol{F}_D^{(k)}$ have the following properties:

$$(\boldsymbol{F}_U^{(k)})^T(\boldsymbol{F}_U^{(k)})^* = \boldsymbol{I}, \tag{3}$$

$$(\boldsymbol{F}_D^{(k)})^T(\boldsymbol{F}_D^{(k)})^* = \boldsymbol{I}, \tag{4}$$

$$(\boldsymbol{F}_U^{(k)})^T(\boldsymbol{F}_D^{(j)})^* = \boldsymbol{0}, (\boldsymbol{F}_U^{(k)})^T(\boldsymbol{F}_U^{(j)})^* = \boldsymbol{0}, \tag{5}$$

$$(\boldsymbol{F}_U^{(k)})^T\boldsymbol{F}_D^* = \boldsymbol{0}, (\boldsymbol{F}_D^{(k)})^T\boldsymbol{F}_U^* = \boldsymbol{0} \tag{6}$$

With the above assumptions/definitions, the multicarrier modulation/demodulation can now be implemented as follows. Let, after transmitter preprocessing, $\boldsymbol{X}_U^{(k)}$ and $\boldsymbol{X}_D^{(k)}$ be the signal vectors to be transmitted respectively on the uplink and downlink by user k. Then, the signals transmitted on the uplink and downlink are:

$$\begin{aligned} \boldsymbol{s}_U^{(k)} &= (\boldsymbol{F}_U^{(k)})^*\boldsymbol{X}_U^{(k)} \\ \boldsymbol{s}_D^{(k)} &= (\boldsymbol{F}_D^{(k)})^*\boldsymbol{X}_D^{(k)}, k = 1,2,...,K \end{aligned} \tag{7}$$

It can be understood that the signaling in (7) includes both the OFDMA signaling and the SC-FDMA signaling [3, 4]. Specifically, if $\boldsymbol{X}_U^{(k)}$ and $\boldsymbol{X}_D^{(k)}$ are obtained by the operations in the F-domain, $\boldsymbol{s}_U^{(k)}$ and $\boldsymbol{s}_D^{(k)}$ are OFDMA signals. By contrast, if $\boldsymbol{X}_U^{(k)}$ and $\boldsymbol{X}_D^{(k)}$ are obtained via another DFT stage associated with certain subcarrier mapping, $\boldsymbol{s}_U^{(k)}$ and $\boldsymbol{s}_D^{(k)}$ are SC-FDMA signals. For more details, please refer to [3].

Let the channel impulse response (CIR) with respect to the kth user be expressed as

$$h_k(\tau) = \sum_{l=0}^{L_k-1} h_l^{(k)}\delta(\tau - lT_\psi) \tag{8}$$

where L_k represents the number of T-domain resolvable paths, $h_l^{(k)}$ is the complex channel gain of the l th resolvable path, and $T_\psi = 1/W_s$ is the resolution, where W_s is the overall frequency bandwidth supporting both uplink and downlink. Then, when signals in the form of (7) are transmitted over the wireless channels characterized by (8), the received observation for detection of user k (either at BS for uplink or at user k for downlink) can be expressed as

$$\begin{aligned} \boldsymbol{r} &= \sum_l \breve{\boldsymbol{H}}_U^{(l)}\boldsymbol{s}_U^{(k)} + \breve{\boldsymbol{H}}_D^{(k)}\sum_{l'}\boldsymbol{s}_D^{(k)} + \boldsymbol{n}_U(\boldsymbol{n}_D^{(k)}) \\ &= \sum_l \breve{\boldsymbol{H}}_U^{(l)}\left(\boldsymbol{F}_U^{(l)}\right)^*\boldsymbol{X}_U^{(l)} + \breve{\boldsymbol{H}}_D^{(k)}\sum_{l'}\left(\boldsymbol{F}_D^{(l')}\right)^*\boldsymbol{X}_D^{(l')} + \boldsymbol{n}_U(\boldsymbol{n}_D^{(k)}) \end{aligned} \tag{9}$$

after removing the cyclic prefix (CP). In the above formulas, $\{l\}$ contains the indexes of uplink users, while $\{l'\}$ contains the indexes of downlink users, $\{\breve{\boldsymbol{H}}_U^{(l)}\}$ are ($M \times M$) circulant matrices of the uplink channels, while $\breve{\boldsymbol{H}}_D^{(k)}$ is the ($M \times M$) circulant matrices of the downlink channel to user k. Furthermore,

$\boldsymbol{n}_U(\boldsymbol{n}_D^{(k)})$ represents the background noise received at BS (or at user *k*). The circulant matrices have the properties:

$$\boldsymbol{F}\breve{\boldsymbol{H}}\boldsymbol{F}^H = \boldsymbol{H} \quad (10)$$

where the index for uplink or downlink as well as that for user *k* or *l* are ignored for simplicity. In (10), $\boldsymbol{H} = \mathrm{diag}\{H_0, H_1, \cdots, H_{M-1}\}$ associated with

$$H_m = \sum_{l=0}^{L-1} h_l \exp\left(-j\frac{2\pi ml}{M}\right), m = 0,1,...,M-1 \quad (11)$$

which is the channel fading gain of the m th subcarrier of a considered user.

Based on (9), the demodulation for the uplink and downlink of user *k* can be implemented with the aid of the reduced inverse DFT (RIDFT) as

$$\boldsymbol{y}_U^{(k)} = \left(\boldsymbol{F}_U^{(k)}\right)^T \boldsymbol{r}, \quad k = 1,2,...,K$$

$$\boldsymbol{y}_D^{(k)} = \left(\boldsymbol{F}_D^{(k)}\right)^T \boldsymbol{r}, \quad k = 1,2,...,K \quad (12)$$

Upon substituting (9) into the above equations and invoking the properties from (3) to (6) as well as (10), it can be readily shown that the demodulation outputs can be expressed as

$$\boldsymbol{y}_U^{(k)} = \sum_l \left(\boldsymbol{F}_U^{(k)}\right)^T \breve{\boldsymbol{H}}_U^{(l)} \left(\boldsymbol{F}_U^{(l)}\right)^* \boldsymbol{X}_U^{(l)} + \bar{\boldsymbol{n}}_U$$

$$\boldsymbol{y}_D^{(k)} = \sum_{l'} \left(\boldsymbol{F}_D^{(k)}\right)^T \breve{\boldsymbol{H}}_D^{(k)} \left(\boldsymbol{F}_D^{(l')}\right)^* \boldsymbol{X}_D^{(l')} + \bar{\boldsymbol{n}}_D^{(k)} \quad (13)$$

where $\bar{\boldsymbol{n}}_U$ and $\bar{\boldsymbol{n}}_D^{(k)}$ are the noise obtained after the RIDFT. Equations in (13) show that the uplink and downlink are successfully separated. Furthermore, when different users are allocated with different subcarriers, it can be shown that

$$\boldsymbol{y}_U^{(k)} = \boldsymbol{H}_U^{(k)} \boldsymbol{X}_U^{(k)} + \bar{\boldsymbol{n}}_U$$

$$\boldsymbol{y}_D^{(k)} = \boldsymbol{H}_D^{(k)} \boldsymbol{X}_D^{(k)} + \bar{\boldsymbol{n}}_D^{(k)} \quad (14)$$

Explicitly, all the interfering user signals are removed. Typically, the transmitter and receiver in MDD-based systems may have the structures as shown in Fig.2 and Fig.3, when OFDM principles are invoked for their implementation.

4. Channel estimation and prediction

For channel estimation/prediction, we assume that the maximum Doppler frequency shift is f_D. Then, from [3] the T-domain auto-correlation can be formulated as

$$E\left[H_m[i]\left(H_m[i+\lambda]\right)^*\right] = J_0\left(2\pi f_D \lambda T_s\right)\sum_{l=0}^{L-1}\Omega_l \quad (15)$$

where i is the time-index, T_s is the symbol duration, $\Omega_l = E[|h_l[i]|^2]$, and $J_0(x)$ is the zeroth order Bessel function of the first kind. Eq. (15) shows that the auto-correlation function is independent of index *m*. Hence, the T-domain correlation coefficient is

$$\rho_T(\lambda) = J_0\left(2\pi f_D \lambda T_s\right) \quad (16)$$

It can be shown that the F-domain auto-correlation function is given by

$$E\left[H_m[i]\left(H_n[i]\right)^*\right] = \sum_{l=0}^{L-1}\Omega_l \exp\left(-j\frac{2\pi(m-n)l}{M}\right) \quad (17)$$

and, hence, the F-domain correlation coefficient is

$$\rho_F(m,n) = \left(\sum_{l=0}^{L-1}\Omega_l\right)^{-1}\sum_{l=0}^{L-1}\Omega_l \exp\left(-j\frac{2\pi(m-n)l}{M}\right) \quad (18)$$

Note that, owing to the expectation operation, $\rho_F(m,n)$ is independent of the index *i*. However, it is dependent on the multipath delay profiles (MDP) of the T-domain CIRs.

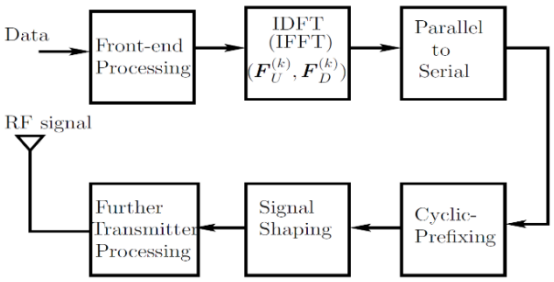


Figure 2 Transmitter block diagram in MDD-assisted wireless systems

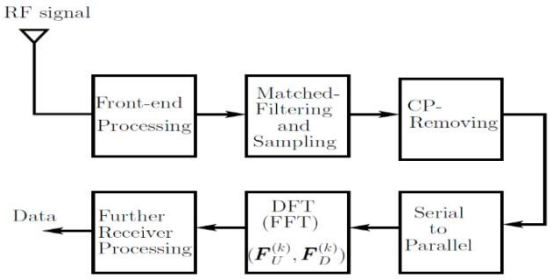


Figure 3 Receiver block diagram in MDD-assisted wireless systems

Having obtained the T- and F-domain channel correlation properties of the MDD systems, the downlink (uplink) channels may be estimated from the uplink (downlink) channels by employment various channel estimation/prediction algorithms [5]. As an example, below we briefly show the principles of channel estimation/prediction based on the Wiener filter theory [5].

Consider a cellular system, where downlink channels are estimated/predicted from the observations obtained from the uplink channels. Correspondingly, the channel estimation/prediction problem can be described in principle as shown in Fig.4, where the filled boxes are the observations obtained from uplink, and the blank boxes corresponding to downlink are estimated with the aid of the information provided by the filled boxes. The estimation/prediction can be carried out as follows:

Step 1 For each of $i = 0,1,...,G-1$ symbol durations, the channel gains of the $j = 0,1,...,\bar{N}$, which are represented by the blank boxes, are estimated in the F-domain.

Step 2 After the F-domain estimation, the estimation can be enhanced for each of the $i = 0,1,...,G-1$ and $j = 0,1,...,\bar{N}$ downlink channels in the T-domain.

Step 3 The above two steps may be repeated to further improve the channel estimation.

Step 4 Finally, the channel gains corresponding to the boxes filled with crossed lines are predicted in the T-domain.

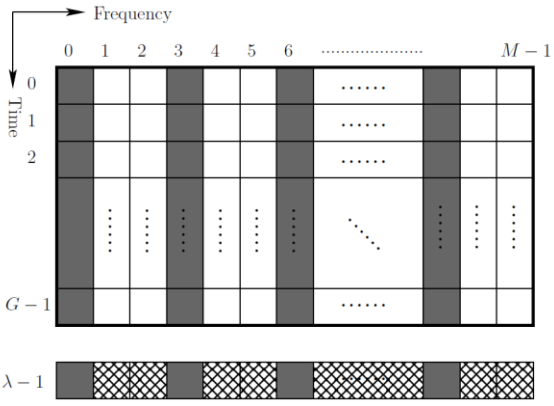


Figure 4 Illustration of the uplink (filled) and downlink (blank) channels in MDD-based systems using M subcarriers

Below we provide some further details. First, we assume that the uplink channel gains of a considered user at time *i* are given by

$$H_U[i] = \left[H_{U,0}[i], \cdots, H_{U,N-1}[i] \right]^T, i = 0,...,G-1 \quad (19)$$

where, without loss of any generality, we assume that the user uses all the *N* uplink subcarriers. According to (11), we can know that, $\boldsymbol{H}_U[i]$ can be represented as

$$\boldsymbol{H}_U[i] = \boldsymbol{Q}\boldsymbol{h}[i] \quad (20)$$

where $\boldsymbol{Q}$ is a ($N \times L$) matrix with its elements provided by some of the coefficients seen in (11). Furthermore, it can be shown that $\boldsymbol{Q}$ is independent of the time index *i*. In (20), $\boldsymbol{h}[i] = [h_0[i], h_1[i], ..., h_{L-1}[i]]^T$ to be estimated.

If $N \geq L$, from (20), we can solve to obtain

$$\boldsymbol{h}[i] = (\boldsymbol{Q}^H \boldsymbol{Q})^{-1} \boldsymbol{Q}^H \boldsymbol{H}_U[i], i = 0,...,G\text{-}1 \quad (21)$$

which provides full information for the T-domain CIR. Consequently, the fading gains of the $\bar{N}$ downlink subcarrier channels can be directly generated with the aid of (11).

In the case that the estimator can only access noisy observations of the uplink channels, from (14) the observation equation can be expressed as

$$\boldsymbol{y}_N[i] = \boldsymbol{H}_U[i] + \boldsymbol{n}[i] = \boldsymbol{Q}\boldsymbol{h}[i] + \boldsymbol{n}[i], i = 0,...,G-1 \quad (22)$$

Then, $\boldsymbol{h}[i]$ may be estimated with the aid of many available algorithms [3, 5], either directly in the F-domain or, as (20) - (21), by first estimating the T-domain CIR and, then, computing the F-domain channel gains. Specifically, if we follow (20) - (21) and when the Wiener filter [5] is employed, we have the estimation

$$\hat{\boldsymbol{h}}[i] = \boldsymbol{P}_h^H \boldsymbol{Q}^H (\boldsymbol{Q}\boldsymbol{P}_h\boldsymbol{Q}^H + \sum_n)^{-1} \boldsymbol{y}_U[i] \quad (23)$$

for $i = 0,...,G-1$, where $\boldsymbol{P}_h = E[\boldsymbol{h}[i]\boldsymbol{h}^H[i]] = \text{diag}\{\Omega_0, \Omega_1, ..., \Omega_{L-1}\}$ and $\sum_n$ is the covariance matrix of $\boldsymbol{n}[i]$. Having obtained $\hat{\boldsymbol{h}}[i]$, the downlink channel gains can be computed, again, by using (11).

After the estimation for all $i = 0,1,...,G-1$ and $j = 0,1,...,\bar{N}$, the fading gains of the *m*th, $m = 0,...,\bar{N}$, downlink channel can be represented as

$$\hat{\boldsymbol{H}}_{D,m} = [\hat{H}_{D,m}[0], \hat{H}_{D,m}[1], ..., \hat{H}_{D,m}[G\text{-}1]]^T \quad (24)$$

Based on these estimates, the estimate to the channel of a given time-index of a given downlink channel may be further enhanced in the T-domain. Furthermore, the future downlink channel can be predicted, e.g., using the Wiener filter principles [5]. For more details, please refer to [3].

5. Conclusions

MDD is capable of integrating the advantages of both the FDD and TDD, while simultaneously circumventing their shortcomings. It also employs some unique merits for practical operation. It can provide high feasibility for channel estimation/prediction, which can be flexibly implemented in the T- or F-domain, or in the joint T- and F-domain. Considering the practice that there are now many wireless systems, such as LTE/LTE-A cellular systems, WiFi, WIMAX, etc., which are based on multicarrier communications, and that there are a lot of short-distance wireless services, it seems that the time has come for us to put the MDD into practice.

References

[1] I. E. Telatar, "Capacity of multiantenna Gaussian channels," European Trans. on Telecom., vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[2] F. Rusek and et.al, "Scaling up MIMO: Opportunities and challenges with very large arrays," IEEE Sig. Proc. Mag., vol. 30, no. 1, pp. 40 – 60, Jan. 2013.

[3] L.-L. Yang, Multicarrier Communications.Chichester, United Kingdom: John Wiley, 2009.

[4] E. Dahlman, S. Parkvall, and J. Skold, 4G LTE/LTE-Advanced for Mobile Broadband, 2nd ed.Oxford, UK: Academic Press Inc, 2013.

[5] S. Haykin, Adaptive Filter Theory, 3rd ed.Upper Saddle River, New Jersey: Prentice Hall, 1996.

D2D Cooperation to Avoid Instantaneous Feedback in Non-reciprocal Massive MIMO Systems

Samah A. M. Ghanem and Laura Cottatellucci
Eurecom, Sophia Antipolis, France
Email: fsamah.g hanem; laura.cottatelluccig@eurecom.fr

1. Introduction

A system where the number antennas at the access point (AC) is much larger than the number of users is referred to as very large antenna array or massive MIMO system [1], [2]. Such a technology is widely believed to be a key enabler of the future 5G networks and fuels intense research activities.

When the channel reciprocity can be exploited, massive MIMO systems have impressive advantages in combatting interference without the costly signalling for cooperation and/or coordination of current cooperative multi-point (CoMP) architectures. In [2] it is shown that massive MIMO systems are not limited by data interference from adjacent cells, allow for a vanishing power per bit as the number of antennas grows, and simple linear detectors such as matched filters can be implemented at the receiver without substantial loss in performance.

However, as well known, an effective design of the downlink communications is strongly sensitive to CSI knowledge. Then, it is of paramount importance the CSI acquisition, which may cause a huge feedback overhead, particularly in fast fading channels with short coherence time. In [2], Marzetta limited the applicability of massive MIMO networks to time division duplex (TDD) mode since heavy load feedback can be avoided by resorting channel reciprocity and adopting open loop feedback.

When uplink-downlink channel reciprocity does not hold, as for example in FDD systems, a closed-loop CSI feedback is required. In a closed loop feedback scheme, each user needs to retransmit an amount of information proportional to the number of antennas at the AP. This task becomes unfeasible during the channel coherence time when the number of antennas grows large. Promising schemes for FDD mode have been proposed in [3], [4], and [5]. Most of these contributions leverage on the reduced rankness of the channel covariance matrices of user terminals in a massive MIMO system. In [6] and [7], the authors pointed out the low rankness of the channel covariance matrices for wide classes of antenna array configurations and exploited it for channel estimation. In [3] and [8], the authors exploit the reduced rankness of the correlation structure of UTs channels in the proposed Joint Spatial Division and Multiplexing (JSDM) scheme. The image of a user channel covariance matrix, whose dimension is substantially lower than the number of antennas at the AP, is referred to as *effective channel subspace* (ECS). The JSDM approach clusters users that share the same ECS and schedules simultaneous transmissions to clusters with orthogonal ECS. The signals for a certain cluster are projected onto the cluster ECS before transmission. Within a cluster, the signal is precoded based on the feedback for the ECS. Further feedback reductions are also possible by restricting the pre-beamformers to a subset of the ECS called *reduced-dimension effective channel subspace.* Recently, in [5], the authors capitalize on D2D communications and allow the users in a cluster to share their CSI, obtaining significant reduction in the CSI feedback. In [9], an iterative compensation algorithm, that reduces the complexity of a two-tier precoding for massive MIMOs has been proposed. Additionally, in [4], the authors introduce a sparse model representation of multi-cell massive MIMO systems and exploit compressive sensing (CS) techniques to reduce the training as well as the feedback overhead for the CSI acquisition at the AP.

Under realistic conditions, with arbitrary configurations of large antenna arrays and randomly located users, the ECSs of different users are in general different and the dimensions of a cluster subspace, union of the ECSs of the cluster users, might be still too large. The selection of a reduced dimension ECS is not obvious and an arbitrary subspace selection can severely impair the system performance [8].

In this paper, we address this issue and we provide a solution that leverages on D2D communications to create virtual MIMOs. As in the JSDM protocol, we cluster n users and project the signals to be transmitted onto the ECS via a beamformer. In contrast to the JSDM that requires instantaneous CSI feedback, we design the precoder as for a point-to-point MIMO system with only statistical CSI of the ECS at the transmitter. At each channel use, n information data streams for a target user in the cluster are precoded and subsequently beamformed for transmission. Each non-target user in the cluster amplifies and forwards the received signal such that the target user receives $n-1$ additional independent versions of the transmit signal to create a virtual MIMO system.

This article is organized as follows. Section 2 describes the system model. In Section 3 we define the proposed solution and derive the corresponding performance in terms of achievable total rate of a cluster. In Section 4 we assess the performance of the proposed system by numerical simulations and compare it with the benchmark JSDM protocol in [8].

The following notation has been used throughout the paper: Boldface uppercase and lowercase letters denote matrices and vectors, respectively. Scalars are in italic. $\mathbf{I}_n$ is the identity matrix of size $n \times n$. The Hermitian operator of a matrix $\mathbf{X}$ is denoted by $\mathbf{X}^H$. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. $\mathbb{E}\{\cdot\}$ is the expectation operator; $\mathrm{tr}(\cdot)$ denotes the traces of

the matrix argument. Finally, $\mathrm{diag}(\mathbf{v})$ is the square diagonal matrix having the element of vector $\mathbf{v}$ as diagonal elements.

2. System Model

We consider a single-cell massive MIMO system with the access point equipped with M antennas and serving single antenna users. The channel is fast fading and non-reciprocal, for example because the system operates in FDD mode. Let $\mathbf{R}_k$ denote the covariance matrix of user k channel and let us refer to its image as ECS of user k. As in [3], users with almost overlapping ECSs are clustered together. For a cluster $\mathcal{C}$, we introduce the cluster covariance matrix defined as $\mathbf{R}_c = \sum_{k\in\mathcal{C}} \mathbf{R}_k$ and we refer to its image as the cluster ECS. Simultaneous transmissions are scheduled to clusters with orthogonal (or quasi-orthogonal) cluster ECSs. Under strict orthogonality of the cluster ECSs, by projecting the signals meant for a single cluster onto its ECS by beamforming, the cluster signals do not cause interference to the other simultaneously scheduled clusters. Thus, in the following we can focus on a single cluster with n users.

Down-link Transmission

The downlink transmission to a single cluster $\mathcal{C}$ is modeled by:

$$\mathbf{y} = \mathbf{H}_c^H \mathbf{B}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{y}$ is the n-dimensional complex column vector of received signals at all the cluster users; $\mathbf{s}$ denotes the vector of i.i.d. Gaussian signals with zero-mean and unit-variance; and $\mathbf{n}$ represents the spatially and temporally white additive Gaussian noise (AWGN) with zero-mean and element-wise variance σ_n^2. Finally, $\mathbf{B}$ is the down-link beamformer such that $\mathrm{tr}\{\mathbf{B}\mathbf{B}^H\} = P_{max}$, if P_{max} is the total transmit power constraint. The down-link channel between the access point and the k-th user in the cluster $\mathcal{C}$ is denoted by the M-dimensional complex vector $\mathbf{h}_k$. Therefore,

$$\mathbf{H}_c = \left[\mathbf{h}_1, \mathbf{h}_2, \dots \mathbf{h}_n\right] \in \mathbb{C}^{M\times n} \tag{2}$$

and $\mathbf{R}_k = \mathbb{E}\{\mathbf{h}_k \mathbf{h}_k^H\}$. By leveraging on the low rankness of $\mathbf{R}_c$ and assuming its rank equals b, the cluster covariance matrix can be expressed as

$$\mathbf{R}_c = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{U}}^H \tag{3}$$

where $\widetilde{\mathbf{\Lambda}}$ is the $b\times b$ matrix of the nonzero eigenvalues of $\mathbf{R}_c$ in non increasing order and $\widetilde{\mathbf{U}}$ is the $M \times b$ matrix whose columns are the normalized eigenvectors of $\mathbf{R}_c$. Similarly, $\mathbf{R}_k = \widetilde{\mathbf{U}}_k \widetilde{\mathbf{\Lambda}}_k \widetilde{\mathbf{U}}_k^H$, with analogous meaning for b_k, $\widetilde{\mathbf{\Lambda}}_k$ and $\widetilde{\mathbf{U}}_k$. Since, by construction, the subspace spanned by the column vectors of $\widetilde{\mathbf{U}}_k$ lies into the subspace spanned by the column vectors of $\widetilde{\mathbf{U}}$, then $\widetilde{\mathbf{U}}_k = \widetilde{\mathbf{U}}\mathbf{T}_k$, where $\mathbf{T}_k$ is a $b \times b$ matrix whose i-th column elements are the coefficients of the i-th column of $\widetilde{\mathbf{U}}_k$ in the basis $\widetilde{\mathbf{U}}$. Additionally, the Gaussian vector $\mathbf{h}_k$ can be expressed as $\mathbf{h}_k = \widetilde{\mathbf{U}}_k \widetilde{\mathbf{\Lambda}}_k^{1/2} \widetilde{\mathbf{h}}_k$ being $\widetilde{\mathbf{h}}_k$ a b_k-dimensional vector of zero-mean, unit variance, independent Gaussian elements. Then,

$$\mathbf{H}_c = \widetilde{\mathbf{U}} \left[\mathbf{T}_1 \widetilde{\mathbf{\Lambda}}_1^{1/2} \widetilde{\mathbf{h}}_1, \cdots \mathbf{T}_n \widetilde{\mathbf{\Lambda}}_n^{1/2} \widetilde{\mathbf{h}}_n\right] = \widetilde{\mathbf{U}}\mathbf{A}_c, \tag{4}$$

where $\mathbf{A}_c$ is a matrix of Gaussian elements, in general, column-wise correlated. In the following we adopt the notation $\mathbf{a}_{ck} = \mathbf{T}_k \widetilde{\mathbf{\Lambda}}_k^{1/2} \widetilde{\mathbf{h}}_k$.

Finally, by projecting the signals to be transmitted onto the cluster ECS, which implies to structure the matrix $\mathbf{B}$ as $\mathbf{B} = \widetilde{\mathbf{U}}\widetilde{\mathbf{B}}$, being the first operator $\widetilde{\mathbf{U}}$ the projection beamformer, we obtain an equivalent system model in the ECS with reduced dimensions

$$\mathbf{y} = \mathbf{A}_c^H \widetilde{\mathbf{B}}^H \mathbf{s} + \mathbf{n}. \tag{5}$$

The received signal at user k is given by

$$y_k = \mathbf{a}_{ck}^H \widetilde{\mathbf{B}}^H \mathbf{s} + \mathbf{n} \tag{6}$$

and the corresponding averaged received power is

$$P_k = \mathrm{tr}\{\mathbf{B}^H \mathbf{R}_k \mathbf{B}\} + \sigma_n^2. \tag{7}$$

Intra-cluster D2D Communications

By D2D communications, the users in a cluster retransmit the received signals in orthogonal time intervals. User ℓ amplifies and forwards its received signal y_ℓ such that its transmitted signal is

$$x_\ell = \sqrt{\frac{P_r}{P_\ell}} y_\ell$$

where P_r is the average transmit power constraint as user ℓ acts as relay. To keep the notation easy, we assume that P_r is equal to all the users.

As likely from physical considerations and the analysis in [7], users within a cluster are closely located and far apart from users in other clusters. Then, we can assume that the simultaneous D2D transmissions in other clusters do not interfere with the intra-cluster transmissions. Thus, the signal received by user k from user ℓ is given by

$$\begin{aligned} r_k &= g_{k\ell} x_\ell + w_k \\ &= \sqrt{\frac{P_r}{P_\ell}} g_{k,\ell} \mathbf{a}_{ck}^H \widetilde{\mathbf{B}}^H \mathbf{s} + \sqrt{\frac{P_r}{P_\ell}} g_{k\ell} n_\ell + w_k \end{aligned} \tag{8}$$

where $g_{k,\ell}$ is the channel coefficient of the fast fading link from user ℓ to user k, realization of a Gaussian random process with variance $\gamma_{k,\ell}$. Finally, w_k is the additive Gaussian noise with unit variance at user k when it acts as receiver in the D2D communications.

At the end of the relaying phase, user k has n independent received versions of the original transmitted signal and can act as a virtual MIMO. The corresponding system model is given by

$$\mathbf{r}_k = \mathbf{\Gamma}_k \mathbf{A}_c^H \widetilde{\mathbf{B}}\mathbf{s} + \mathbf{z} \tag{9}$$

where $\mathbf{\Gamma}_k$ is an $n \times n$ diagonal matrix given by

$$\mathbf{\Gamma}_k = \mathrm{diag}\left[\sqrt{\frac{P_r}{P_1}} g_{k,1}, \cdots, \sqrt{\frac{P_r}{P_{k-1}}} g_{k,k-1}, 1, \sqrt{\frac{P_r}{P_{k+1}}} g_{k,k+1} \cdots \sqrt{\frac{P_r}{P_n}} g_{k,n}\right]$$

and $\mathbf{z}$ is the equivalent Gaussian noise with diagonal covariance matrix, at UT k, given by:

$$\mathbf{\Sigma}_k = \text{diag}\left[\frac{P_r}{P_1}\gamma_{k,1}\sigma_n^2+1,\cdots,\frac{P_r}{P_{k-1}}\gamma_{k,k-1}\sigma_n^2+1,\right.$$
$$\left.\sigma_n^2,\frac{P_r}{P_{k+1}}\gamma_{k,k+1}\sigma_n^2+1\cdots\frac{P_r}{P_n}\gamma_{k,n}\sigma_n^2+1\right].$$

3. Design and Analysis of the Proposed System

In this section we detail the proposed massive MIMO system which combats intra-cluster interference via virtual MIMO systems in the cluster. The access point transmits at each channel use information streams to the same final user in the cluster refered to as target user throughout this work. With a simple round robin algorithm, a transmission to each of the users in the cluster is performed each n time intervals. In contrast to the JSDM communication scheme, instantaneous CSI is not fed back to the access point to optimally design a precoder. Additionally, in this work, we assume that also the statistics that characterize the D2D communications are not fed back to the access point. Each user simply feeds back the rate at which its own data have to be transmitted. As detailed later, this rate can be computed locally by each user terminal. Then, the access point is substantially oblivious of the presence of varying virtual MIMO systems at the receiver side and the precoder design at the transmitter can be optimized for the fictitious point-to-point MIMO system in (1) or, equivalently, for the MIMO system in the cluster ECS in (5) by keeping in mind that the global precoder $\mathbf{B}$ consists of the cascade of the precoder $\widetilde{\mathbf{B}}$ in the cluster ECS followed by the beamformer $\widetilde{\mathbf{U}}$.

Although, in general, an optimal design of the system to maximize the rate at the fictitious point-to-point MIMO system would require the transmission of the information over b symbol streams and a careful design of the precoder $\widetilde{\mathbf{B}}$ along the lines of available works in literature for general correlated MIMO channels [10], in this contribution we opt for a suboptimal precoder to keep the exposition simple and concise. The proposed precoder transmits n streams with identical powers along the directions of the eigenvectors with the n highest eigenvalues of the matrix $\mathbf{R}_c$, equivalently

$$\widetilde{\mathbf{B}} = \frac{P_{max}}{n}\begin{bmatrix}\mathbf{I}_n\\ \mathbf{0}_{(b-n)\times n}\end{bmatrix} \tag{10}$$

where $\mathbf{0}_{(b-n)\times n}$ is a $(b-n)\times n$ matrix of zeros. Thus, (9) reduces to

$$\mathbf{r}_k = \frac{P_{max}}{n}\mathbf{\Gamma}_k\widetilde{\mathbf{A}}_c^H\mathbf{s}+\mathbf{z}$$

where $\widetilde{\mathbf{A}}_c$ is the submatrix obtained from $\mathbf{A}_c$ by extracting the first n rows. Then, the achievable rate for the target user k is

$$R_k = \frac{1}{n}\mathbb{E}\left[\log\det\left[\mathbf{I}_n+\frac{P_{max}}{n}\mathbf{\Gamma}_k\widetilde{\mathbf{A}}_c^H\widetilde{\mathbf{A}}_c\mathbf{\Gamma}_k^H\mathbf{\Sigma}_k^{-1}\right]\right]$$

where the factor n^{-1} is due to the fact that user k receives each n channel uses. Note that user k feeds back the value nR_k that should be used for encoding at the AP.

The total achievable rate in cluster $\mathcal{C}$ is given by

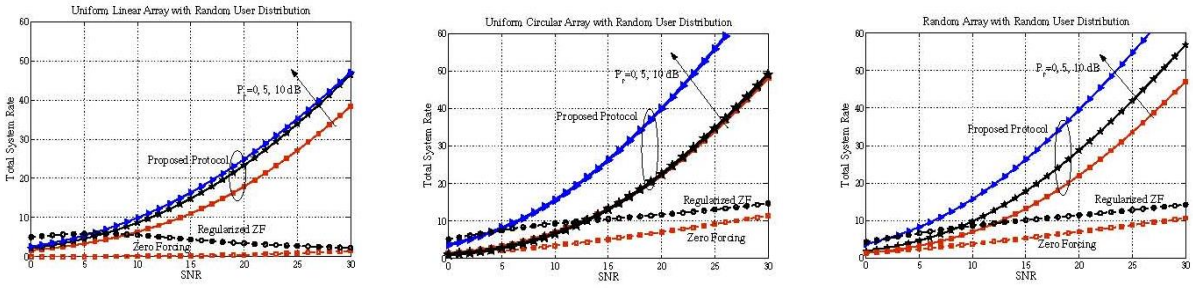
$$R_c = \sum_{k\in\mathcal{C}} R_k.$$

4. Simulation Results

In this section we assess the performance of the proposed communication system by simulations and we compare it with the benchmark JSDM system [3]. In all the simulations we adopt the following setting. We consider a system with $M = 64$ antennas and 16 users divided into 4 clusters with the same number of users, i.e. $n = 4$. At the access point, we adopt various configurations of the antenna array: Uniform Linear Array (ULA), Uniform Circular Array (UCA), and Random Array (RA). The downlink channel is fast fading with Gaussian distributed coefficients. In order to generate the covariance matrices of users' channels, we utilize a one-ring model as in [3] and the same configuration adopted in [3]: the ECSs have size $b \geq r$ with $r = 11$ where r is the number of dominant eigen-modes. In order to better understand the simulation results is relevant to recall that for the ULA the users are randomly distributed over a sector of 180 degrees while for the UCA and RA they are distributed over the full coverage area. Additionally, also the channels between pairs of users are fast fading, independent each other, and follow a complex Gaussian distribution with zero mean and unit variance.

Figures 1(a)-(c) provide a comparison between the proposed massive MIMO system with cooperation via D2D and the non-cooperative JSDM system. All the figures show the achievable total rate of the system as a function of the SNR in the downlink transmission between the access point and the user terminals. For the JSDM benchmark system, we show the performance (dashed lines) in the two cases when a zero forcing (ZF, square markers) and a regularized zero forcing (RZF, circles markers) precoding are implemented at the access point. No penalty is imposed to account for the cost of the feedback, which is essential to properly design ZF and RZF precoding. Both precoders are designed under the assumption of perfect CSI. It is worth recalling that the feedback requires a relevant portion of the useful bandwidth which can be dominant compared to the bandwidth allocated to the information when the dimensions of the ECSs are large and/or a high accuracy for the CSI is required. The performance of the proposed system is plotted in solid lines for various values of the average transmit powers at the user terminals when they act as relays, more specifically, for $P_r = 0$ dB (square markers), $P_r = 5$ dB (star markers), and $P_r = 10$ dB (triangle markers).

Since the clusters in the system with ULA antenna configuration suffer from a higher inter-cluster interference than the other systems, for this antenna configuration we obtain the poorest performance and, not surprisingly, the RZF has poorer performance at high SNR than at lower SNR. Intercluster interference is less significant for the proposed system since we adopt a reduced ECS of minimum size. For all the considered antenna configurations and the range of SNR of practical interest, the proposed system outperforms



(a) (b) (c)

Figure 1: Average Total Rate vs SNR in downlink for the proposed system (solid lines) with varying average relay powers (square marker $P_r = 0$ dB, star marker $P_r = 5$ dB, triangle marker $P_r = 10$ dB) and the benchmark JSDM system (dashed lines) with zero forcing (square marker) and regularized zero forcing (circle marker). Users randomly distributed in the coverage area and (a) ULA, (b) UCA, and (c) RA antenna configurations.

the benchmark system with ZF beamforming, even at very low levels of P_r, the averaged transmitted power at the relays. The comparison to the benchmark system with RZF beamforming is more critical: while at high SNR the proposed system has always better performance, at low SNR the same performance of the benchmark are attained at the expenses of higher transmit powers in the relaying phase. However, it is worth recalling that our plots do not account for the bandwidth required by the feedback. Additionally, this work does not explore the improvements of the proposed system that can be obtained by adopting reduced ECS of larger size and optimal precoding matrices. These further analyses are left for further studies.

References

[1] T. Marzetta, "How much training is required for multiuser MIMO?" in *Fortieth Asilomar Conference on Signals, Systems and Computers*, Oct 2006, pp. 359–363.

[2] ——, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, Nov 2010.

[3] J. Nam, J.-Y. Ahn, A. Adhikary, and G. Caire, "Joint Spatial Division and Multiplexing: Realizing massive MIMO gains with limited channel state information," in *Information Sciences and Systems (CISS), 2012 46th Annual Conference on*, Mar 2012, pp. 1–6.

[4] X. Rao and V. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3261–3271, Jun 2014.

[5] H. Yin, L. Cottatellucci, and D. Gesbert, "Enabling massive MIMO systems in the FDD mode thanks to D2D communications," *Asilomar Conference on Signals, Systems and Computers*, Nov 2014.

[6] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[7] H. Yin, D. Gesbert, and L. Cottatellucci, "Dealing with interference in distributed large-scale MIMO systems: A statistical approach," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 8, no. 5, pp. 942–953, Oct 2014.

[8] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint Spatial Division and Multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, Oct 2014.

[9] J. Chen and V. Lau, "Two-tier precoding for FDD multi-cell massive MIMO time-varying interference networks," *IEEE Journal on Selected Areas in Communications,*, vol. 32, no. 6, pp. 1230–1238, Jun 2014.

[10] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers ond noise: A (not so) large N analysis," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2545–2561, Oct 2003.

Samah A. M. Ghanem was born in Beirut, Lebanon, 1979. She received her BS in Electrical Engineering from An-Najah National University, Palestine, in 2001, the M.Sc. in Scientific Computing from Birzeit University, Palestine in 2009, and the Ph.D. in Telecommunications Engineering (summa cum laude) from the University of Aveiro (MAP-Tele), Portugal in 2013. She worked for the industry between 2001 to 2008, and at Ericsson LMI R&D, Ireland in 2008. She joined the Mobile Communications Dept. at Eurecom, Sophia Antipolis, France as a postdoctoral researcher in March, 2014. She was a visiting researcher at Aalborg University, Denmark in 2012 and at the University of Duisburg-Essen, Germany in 2008.

Laura Cottatellucci (S'01, M'07) is assistant professor at EURECOM since December 2006. Prior to joining EURECOM, she was research fellow at University of South Australia, Australia (Jan.-Nov. 2006) and in INRIA, Sophia Antipolis, France (Oct.-Dec. 2005). She obtained the PhD from the Technical University of Vienna, Austria (2006) and the Master degree from La Sapienza University, Rome, Italy (1995). After specialization in networking at the School for Advanced Studies Guglielmo Reiss Romoli (1996, Italy), she worked in Telecom Italia (1995-2000) as responsible of industrial projects. From April 2000 to September 2005 she worked as senior research in FTW Austria. She held visiting appointments at EURECOM, France, FAU Erlangen-Nürnberg, Germany, and NTNU, Trondheim, Norway.

Immersive Light Field Based 3D Telemedicine Applications in 5G

Wei Xiang and Gengkun Wang, University of Southern Queensland, Australia
Email:{wei.xiang, gengkun.wang}@usq.edu.au

1. Introduction

The demand for quality healthcare continues to grow, especially in regional areas, where the lower health status remains a concern. Telemedicine has the potential to enable better healthcare services for regional residents [1] [2]. Conventional videoconferencing telemedicine consultations have been widely explored in the past to support remote healthcare. However, today's conferencing systems are deficient in many ways, such as their ability to support effective collaboration among conference participants, engendering a feeling of remoteness amongst the participants that is contrary to the intended benefits of the system [3]. Aside from low resolution and unpredictable delays, two pronounced limitations have been identified in hindering the widespread adoption of 2D telehealth: (a) the difficulty in obtaining the desired 2D camera views, and (b) the lack of depth perception.

Immersive 3D videoconferencing system enables geographically distributed users interact in a simulated environment, which especially benefits telehealth services by providing an immersive 3D communication between patients and health professions [4]. An ideal 3D videoconferencing system is not hard to imagine: combining the best computer graphics, audio, computer simulation, and image, connecting with networking as good as direct memory access, providing software and hardware to track gaze, gesture, facial expression, and body position [5]. However, current 3D videoconferencing system suffers from issues as *row resolution, awkward, virtual reality (VR) glasses, and significant two-way communication latency*. Many of the barriers in 3DTI systems, such as capturing, rendering and display modules are market-based, but the other components are true technical research issues.

Light field photography is able to capture both spatial and angular information and enables new possibilities for digital imaging [6]. Light field displays (LFDs) can provide wide fields of view (the viewing zone where observers are able to see the clear view) and large depths of field (how far virtual objects can "pop out" of the screen) without specialized eyewear.

The advent of fifth generation (5G) mobile communication technologies provides unique opportunities to catalyze changes in the way healthcare is delivered. 5G wireless networks will support 1,000-fold gains in capacity, connections for at least 100 billion devices, and a 10 Gb/s individual user experience capable of extremely low latency and response times [7]. The 5G mobile network is able to support ultra-high definition video and virtual reality applications, which is perfect for immersive 3D telemedicine applications.

In this letter, a framework of a novel immersive 3D telemedicine for 5G networks is proposed. Using the state-of-the-art light field rendering and streaming technique, this system is able to provide a glasses-free real-time immersive 3D video communication experience under the high speed 5G network. The proposed framework of the immersive 3D telemedicine has the potential to revolutionize next-generation of 3D video communication technology, which can significantly improve the current videoconferencing experience through providing a high level of details, and low-latency communications. In the following, we will introduce light field technology, and later the architecture of this novel 3D telemedicine system and its modules. The experiment results will also be present.

2. Light field technology

Light field was first introduced in an 1846 lecture entitled "Thoughts on Ray Vibrations" that light should be interpreted as a field, much like the magnetic fields. A light ray in space can be parameterized by three coordinates and two angles, which is a five-dimensional function. It was not until recent advance in computational photography that enables us to capture light fields digitally [8].

Contrary to traditional cameras, light field cameras (also called plenoptic cameras) have a microlense array in front of the imaging sensor. Such arrays consist of many microscopic lenses, which split up what would have become a 2D-pixel into individual light rays just before reaching the sensor. The resulting raw image is a composition of as many tiny images as there are microlenses [9]. Next, sophistic software is used to find matching light rays across all these images. Once it has collected a list of (1) matching light rays; (2)

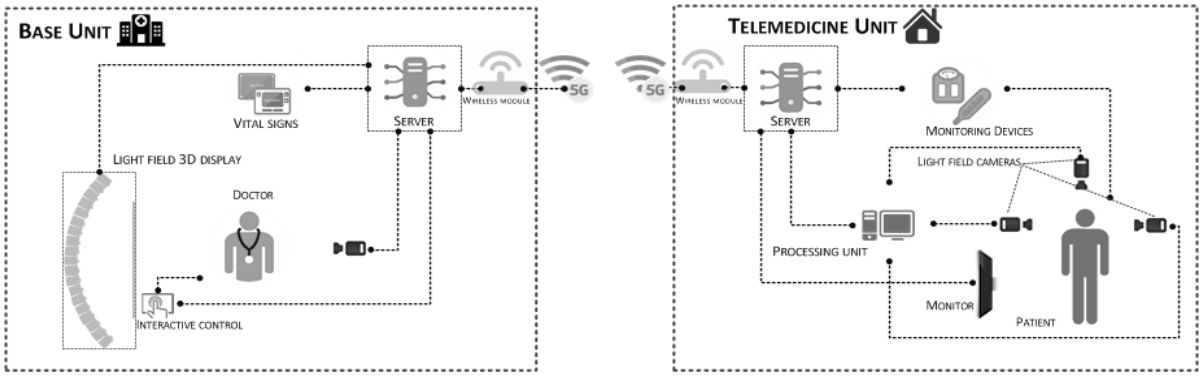


Fig. 1 System architecture

their position in the microlense array and (3) within the sub-image, the information can be used to reconstruct a sharp 3D model of the scene. The additional information inherent in a light field can be used to solve various computer vision challenges: 1) rendering images from a virtual viewpoint; 2) virtual refocusing and synthetic aperture; 3) stereoscopic display; 4) depth map construction for 3D modeling [10].

Light field display, on the other hand, reproduces light fields having both horizontal and vertical parallax. Instead of separate 2D views of a 3D scene, they reconstruct the 3D light field as a set of light rays. In most light field displays, this is achieved by using an array of projection modules emitting light rays and a custom made holographic screen. The light rays generated in the projection modules hit the holographic screen at different points and the holographic screen makes the optical transformation to compose these light rays into a continuous 3D view. Each point of the holographic screen emits light rays of different color to various directions. Light rays leaving the screen spread in multiple directions, as if they were emitted from points of 3D objects as fixed spatial locations. Light field displays are capable of providing 3D images with a continuous motion parallax on a wide viewing zone, without wearing glasses. Objects appear behind or even in front of the screen just like on holograms.

With the advantages in 3D model capture and immersive glasses-free 3D display, light field technology is ideally suitable for a wide range of telemedicine applications, e.g., tele-surgery, tele-wound and tele-psychiatry.

3. System Architecture

The architecture of the proposed system is illustrated in Fig. 1. The unit located at the patient's size is called "*telemedicine unit*", while the unit located at doctor's site is called "*base unit*". The *telemedicine unit* is responsible for collecting and transmitting biosignals and 3D model of the patient from a remote-located home to the doctor's location, while the base unit is responsible for receiving and displaying incoming data.

A. Telemedicine unit

The telemedicine unit mainly consists of four modules: 1) 3D data acquisition module, which is responsible to digitalized the patient in 3D format; 2) a processing unit, which implement ligh rendering and 3D mesh reconstruction; 3) a biogsignal monitoring unit, which collects the patient's biosignals, e.g., heart rate and blood pressure; 4) a server station that is responsible of the storage, compression and transmission of the data; and 5) a wireless module that is in charge of wireless transmission and reception for 5G network.

The 3D data acquisition module is designed to capture both RGB and depth images for 3D reconstruction. In our proposed system, three Lytro Illum light field cameras are used to capture the full 3D data of the patient. The light fields captured by the cameras are converted to 3D model and transferred to the server for transmission.

B. Base unit

The base unit mainly consists of a server for data collection, decoding and decompression, a display unit for vital signs, a display and interactive unit, and also a wireless module to connect to the 5G network. A HoloVizio 3D display unit for autostereoscopic light field display is used in this sytem. The HoloVizio 3D display is able to render the input 3D mesh model into light fields and realize an immersive autostereoscopic 3D experience. Moreover, via the interactive touch control, the doctor is able to interact with the display 3D model and determine his/her interested area.

Fig. 3 shows our experiment results of using light field images to conduct extract depth map and reconstruct mesh model and 3D model. It can be seen that our algorithm can achieve a 3D model with high level of detail. Our experiment also shows that the reconstruction speed is much faster than 3D mesh reconstruction using multi-view and depth images.

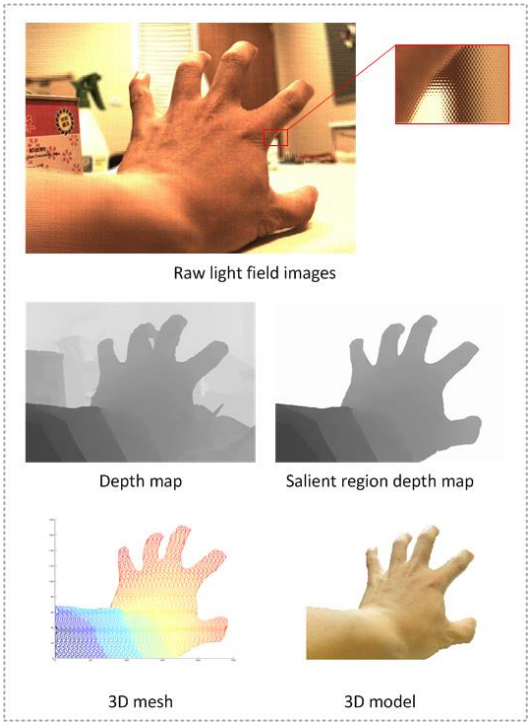


Fig. 3 Experiment results

3. Conclusions

An immersive 3D light field based telemedicine system for 5G networks is proposed in this letter. Utilizing the state-of-the-art light field technology under the future generation mobile communication network, it can provide a digitalized 3D model with high levels of detail and at the same time offer an immersive glasses-free 3D experience. Such a system has the potential to revolutionize next-generation of 3D video communication technology, and significantly affect the medical service in regional areas.

The proposed system is ideally suitable for a wide range of telemedicine applications. In future work, we will explore the various applications of this novel 3D telemedicine system, e.g., identifying the healing process of the wound by accurate measurement, subtle facial expression recognition in tele-psychiatry.

References

[1] M. B. Steventon, J. Billings, J. Dixon, H. Doll, S. Hirani, M. Cartwright, L. Rixon, M. Knapp, C. Henderson, A. Rogers, R. Fitzpatrick, J. Hendy and S. Newman, "Effect of telehealth on use of secondary care and mortality: findings from the whole system demonstrator cluster randomised trial," *BMJ: British Medical Journal*, p. 344, 2012.

[2] K. Ho, H. Karlinsky, S. Jarvis-Selinger and J. May, "Videoconferencing for telehealth: Unexpected challenges and unprecedented opportunities," *British Columbia Medical Journal*, vol. 46, pp. 285-289, Aug 2004.

[3] K. Ho, H. Karlinsky, S. Jarvis-Selinger and J. May, "Videoconferencing for telehealth: Unexpected challenges and unprecedented opportunities," *British Columbia Medical Journal*, vol. 46, pp. 285-289, Aug 2004.

[4] A. Stranieri, R. Collmann and A. Borda, “High definition 3d telemedicine: The next frontier?,” *Selected papers in Global Telehealth 2012*, pp. 133-141, 2012.

[5] S. Beck and e. al, "Immersive group-to-group telepresence," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 4, pp. 616-625, 2013.

[6] M. Lenoy, "Light fields and computational imaging," *IEEE Computer*, vol. 39, no. 8, pp. 46-55, 2006.

[7] Andrews, J.G.; Buzzi, S.; Wan Choi; Hanly, S.V.; Lozano, A.; Soong, A.C.K.; Zhang, J.C., "What Will 5G Be?," *IEEE Journal on Selected Areas in Communications*, vol.32, no.6, pp.1065-1082, June 2014

[8] C. Zhang and T. Chen, "A survey on image-based rendering - representation, sampling and compression," *Signal Processing: Image Communication*, vol. 19, no. 1, pp. 1-28, 2004.

[9] R. Ng, Digtial light field photography, *Standford University*, 2006.

[10] R. Ng, M. Levoy, M. Bredif, G. Duval and M. Horowitz, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, pp. 1257-1265, Apr 2005.

Wei Xiang (S’00–M’04–SM’10) received the B.Eng. and M.Eng. degrees, both in electronic engineering, from the University of Electronic Science and Technology of China, Chengdu, China, in 1997 and 2000, respectively, and the Ph.D. degree in telecommunications engineering from the University of South Australia, Adelaide, Australia, in 2004.

Since January 2004, he has been with the Faculty of Engineering and Surveying, University of Southern Queensland, Toowoomba, Australia, where he currently holds a faculty post of PPP. In 2008, he was a visiting scholar at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. During Oct. 2010 and Mar. 2011, he was a visiting scholar at the University of Mississippi, Oxford, MS, USA. He received the Best Paper Award at 2011 IEEE WCNC. He was named a Queensland International Fellow (2010-2011) by the Queensland Government of Australia, an Endeavour Research Fellow (2012-2013) by the Commonwealth Government of Australia, and a Smart Futures Fellow (2012-2015) by the Queensland Government of Australia. His research interests are in the broad area of communications and information theory, particularly coding and signal processing for multimedia communications systems.

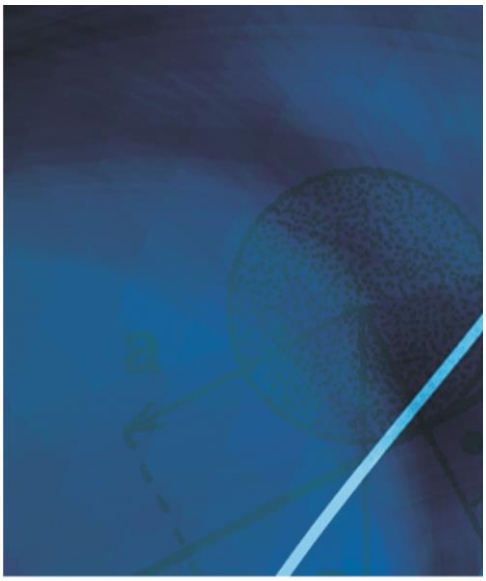
Gengkun Wang received the B.Eng. degree in mechanical engineering and M.Eng. degree in mechatronic engineering, from the University of Science and Technology, Beijing, China, in 2006 and 2009, and the Ph.D in telecommunications engineering in University of South Queensland, Toowoomba, Australia, respectively.

He is now working as a postdoctoral research fellow at the school of electrical and mechanical in the University of Southern Queensland. His research interests are Physical-layer network coding, Channel Coding, and light field technology. He has published a few high quality papers. He also received the Best Paper Award from the IEEE Wireless Communications and Networking Conference, Cancun, Mexico, in 2011.

IEEE MultiMedia

January–March 2015 Vol. 22, No. 1

Published by the IEEE Computer Society
in cooperation with the IEEE Communications Society
and IEEE Signal Processing Society



Interactive Sonification

Departments

IEEE Transactions on Circuits and Systems for Video Technology (CSVT) Newsletter

(November/December 2014 issue)

A. Editor-in-Chief Message: A recap of some of the major milestones for the IEEE Transactions on Circuits and Systems for Video Technology over the past year is summarized in the Editor-in-Chief Message. See http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7029761 and http://tcsvt.polito.it/editor/EIC_Message.html.

B. Introduction of New Associate Editors: An introduction to the new associate editors and editorial board members of the IEEE Transactions on Circuits and Systems for Video Technology is available on the CSVT Xplore website: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7029760.

C. Revised EDICS Categories: The CSVT editorial board has recently adopted revised Editors' Information Classification Scheme (EDICS) categories. The new EDICS categories reflect the reorganized and expanded scope of CSVT, allow us to more efficiently process paper submissions to CSVT, and inform the readers of the full scope of the transactions. The revised EDICS categories can be found on the CSVT website: http://tcsvt.polito.it/edics.html.

D. Call for Best Paper Nominations: We would like to solicit nominations for the annual "CSVT Best Paper Award". The deadline for submission of nominations is February 15, 2015. The Circuits and Systems for Video Technology Best Paper Award recognizes the best paper published in the "Transactions on Circuits and Systems for Video Technology" publication. The award is based on general quality, originality, contributions, subject matter and timeliness. Anyone who is an author of a paper published in the "Transactions on Circuits and Systems for Video Technology" during the three calendar years preceding the award is eligible for nomination. Nominations should be submitted online at http://ieee-cas.org/2015-ieee-circuits-and-systems-society-best-paper-award-nominations-form.

E. Call for Special Issue Proposals: We would like to solicit special issue proposals on new and emerging topics of great interest to the CSVT community. Special issue proposals should be within the scope of CSVT, which covers all aspects of visual information relating to video or that have the potential to impact future developments in the field of video technology and video systems. Special issue proposals should include (a) a cover letter outlining the topic of the proposed special issue, timeliness of the topic, related special issues, and interest to the CSVT community; (b) a draft call for papers summarizing the scope of the proposed special issue, list of topics, suggested timeline, and list of guest editors; and (c) a short biography for each of the guest editors. The proposal should include 3-5 guest editors who are senior-level researchers on the topic of the proposed special issue (e.g., full professor or equivalent research positions) and have a diverse background (i.e., geographically, ethnically, gender, expertise, industry/academia, etc.). Special issue proposals should be sent to the editor-in-chief, Dan Schonfeld tcsvt-eic@tcad.polito.it, and the deputy editor-in-chief, Shipeng Li shipeng.li@microsoft.com. Examples of approved special issues including a list of the guest editors and sample call for papers can be found on the CSVT webpage. See http://tcsvt.polito.it/special_issue.html.

F. Call for Papers - Special Issue Series on Visual Computing in the Cloud - Special Issue on Mobile Computing
Guest Editors:
Yonggang Wen, Nanyang Technological University, Singapore
Pascal Frossard, EPFL, Switzerland
Qibin Sun, Cisco Systems Inc., USA
Wenjun Zeng, University of Missouri, USA
Jacob Chakareski, University of Alabama, Tuscaloosa, AL, USA
Di Wu, Sun Yat-Sen University, China
Deadline: July 15, 2015 (Publication: June 2016)

G. Call for Papers - Special Issue on Augmented Video
Guest Editors:
Peter Eisert, Fraunhofer HHI, Germany
Henry Fuchs, University of North Carolina, USA
Yebin Liu, Tsinghua University, China
Kyoung Mu Lee, Seoul National University, Korea
Didier Stricker, DFKI, Germany
Graham Thomas, BBC R&D, UK
Deadline: September 1, 2015 (Publication: September 2016)

Call for Papers

IEEE Transactions on Multimedia

Special Issue on "Multimedia: The Biggest Big Data"

Multimedia is increasingly becoming the "biggest big data" as the most important and valuable source for insights and information. It covers from everyone's experiences to everything happening in the world. There will be lots of multimedia big data --- surveillance video, entertainment and social media, medical images, consumer images, voice and video, to name a few, only if their volumes grow to the extent that the traditional multimedia processing and analysis systems cannot handle effectively. As such, multimedia big data will emerge as the next "must have" competency in our society, and is spurring on tremendous amounts of research and development of related technologies and applications. As an active and inter-disciplinary research field, multimedia big data also presents a great opportunity for multimedia computing in the big data era. The challenges and opportunities highlighted in this field will foster some interesting future developments in the multimedia research and applications.

Scope

The goal of this special issue is to provide a premier forum for researchers working on the aforementioned multimedia big data aspects to present their recent research results. It also provides an important opportunity for multidisciplinary works connecting big data to multimedia computing. Topics of interest include, but are not limited to:

- New theory and models for multimedia big data computing
- Ultra-high efficiency compression, coding and transmission for multimedia big data
- Content analysis and mining for multimedia big data
- Semantic retrieval of multimedia big data
- Deep learning and cloud computing for multimedia big data
- Green computing for multimedia big data (e.g., high efficiency storage)
- Security and privacy in multimedia big data
- Interaction, access, visualization of multimedia big data
- Multimedia big data systems
- Novel and incentive applications of multimedia big data in various fields (e.g., search, healthcare, transportation, and retail)

Important dates

Submission deadline: **February 28, 2015**
First notification: April 28, 2015
Revision due: May 31, 2015
Final notification of acceptance: July 5, 2015
Camera-ready manuscript due: July 21, 2015
Tentative publication date: August 2015

Submission procedure

Papers should be formatted according to the IEEE Transactions on Multimedia guidelines for authors (see: http://www.signalprocessingsociety.org/tmm/tmm-author-info/). By submitting/resubmitting your manuscript to this transactions, you are acknowledging that you accept the rules established for publication of manuscripts, including agreement to pay all over-length page charges, color charges, and any other charges and fees associated with publication of the manuscript. Manuscripts (both 1-column and 2-column versions are required) should be submitted electronically through the online IEEE manuscript submission system at http://mc.manuscriptcentral.com/tmm-ieee. When selecting a manuscript type, the authors must click on BigMM Special Issue. All the submitted papers will go through the same review process as that for the regular TMM paper submissions. Referees will consider originality, significance, technical soundness, clarity of exposition, and relevance to the special issue topics above.

Guest Editors

Shu-Ching Chen, Florida International University, USA (chens@cs.fiu.edu)
Ramesh Jain, University of California, Irvine, USA (jain@ics.uci.edu)
Yonghong Tian, Peking University, China (yhtian@pku.edu.cn)
Haohong Wang, TCL Research America, USA (haohongwang@gmail.com)

Call for Papers

IEEE Transactions on Cloud Computing

Special Issue on "Mobile Clouds"

Mobile cloud computing represents one of the latest developments in cloud computing advancement. In particular, mobile cloud computing extends cloud computing services to the mobile domain by enabling mobile applications to access extern al computing and storage resources available in the cloud. Not only mobile applications are no longer limited by the computing and data storage limitations within mobile devices, nevertheless adequate offloading of computation intensive processes also has the potential to prolong the battery life.

Besides, there is also an incentive for mobile devices to host foreign processes. This represents a new type of mobile cloud computing services. Ad-hoc mobile cloud is one instance that mobile users sharing common interest in a particular task such as image processing of a local happening can seek collaborative effort to share processing and outcomes. Vehicular cloud computing is another instance of mobile cloud computing that exploits local sensing data and processing of vehicles to enhance Intelligent Transportation Systems.

This Special Issue will collect papers on new technologies to achieve realization of mobile cloud computing as well as new ideas in mobile cloud computing applications and services. The contributions to this Special Issue may present novel ideas, models, methodologies, system design, experiments and benchmarks for performance evaluation. This special issue also welcomes relevant research surveys. Topics of interest include, but are not limited to:

- Trends in Mobile cloud applications and services
- Architectures for mobile cloud applications and services
- Mobile cloud computing for rich media applications
- Service discovery and interest matching in mobile cloud
- Collaboration in mobile clouds
- Process offloading for mobile cloud computing
- Mobile device virtualization
- Mobile networks for cloud computing Mobile cloud monitoring and management
- Security and privacy in mobile clouds
- Performance evaluation of mobile cloud computing and networks
- Scalability of mobile cloud networks
- Software defined systems for mobile clouds
- Self-organising mobile clouds
- Mobile vehicular clouds
- Disaster recovery in mobile clouds
- Economic, social and environmental impact of mobile clouds
- Mobile cloud software architecture

Important Dates

Paper submission: **May 1, 2015**
First Round Decisions: August 15, 2015
Major Revisions Due: September 15, 2015
Final Decisions: November 15, 2015
Publication: 2016

Submission & Major Guidelines

This special issue invites original research papers that present novel ideas and encourages submission of "extended versions" of 2-3 Best Papers from the IEEE Mobile Cloud 2015 conference. Every submitted paper will receive at least three reviews and will be selected based on the originality, technical contribution, and scope. Submitted articles must not have been previously published or currently submitted for publication elsewhere. Papers should be submitted directly to the IEEE TCC at https://mc.manuscriptcentral.com/tcc, and must follow TCC formatting guidelines. For additional information, please contact Chuan Heng Foh (c.foh@surrey.ac.uk).

Editor-in-Chief

Rajkumar Buyya, the University of Melbourne, Australia

Guest Editors

- Chuan Heng Foh, University of Surrey, UK
- Satish Narayana Srirama, University of Tartu, Estonia
- Elhadj Benkhelifa, Staffordshire University, UK
- Burak Kantarci, Clarkson University, USA
- Periklis Chatzimisios, Alexander TEI of Thessaloniki, Greece
- Jinsong Wu, Alcatel Lucent, China

MMTC OFFICERS (Term 2014 — 2016)

CHAIR

Yonggang Wen
Nanyang Technological University
Singapore

STEERING COMMITTEE CHAIR

Luigi Atzori
University of Cagliari
Italy

VICE CHAIRS

Khaled El-Maleh (North America)
Qualcomm
USA

Liang Zhou (Asia)
Nanjing University of Posts &Telecommunications
China

Maria G. Martini (Europe)
Kingston University,
UK

Shiwen Mao (Letters & Member Communications)
Auburn University
USA

SECRETARY

Fen Hou
University of Macau, Macao
China

E-LETTER BOARD MEMBERS (Term 2014—2016)

Periklis Chatzimisios	Director	Alexander TEI of Thessaloniki	Greece
Guosen Yue	Co-Director	Broadcom	USA
Honggang Wang	Co-Director	UMass Dartmouth	USA
Tuncer Baykas	Editor	Medipol University	Turkey
Tasos Dagiuklas	Editor	Hellenic Open University	Greece
Chuan Heng Foh	Editor	University of Surrey	UK
Melike Erol-Kantarci	Editor	Clarkson University	USA
Adlen Ksentini	Editor	University of Rennes 1	France
Kejie Lu	Editor	University of Puerto Rico at Mayagüez	Puerto Rico
Muriel Medard	Editor	Massachusetts Institute of Technology	USA
Nathalie Mitton	Editor	Inria Lille-Nord Europe	France
Zhengang Pan	Editor	China Mobile	China
David Soldani	Editor	Huawei	Germany
Shaoen Wu	Editor	Ball State University	USA
Kan Zheng	Editor	Beijing University of Posts & Telecommunications	China