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

E-LETTER

Vol. 6, No. 10, October 2011

IEEE COMMUNICATIONS SOCIETY

CONTENTS

MESSAGE FROM MMTC CHAIR

Dear MMTC fellow members,

It was wonderful to meet many of our TC members in ICME 2011 (Barcelona, Spain) this July, and witness the recording-breaking event in ICME history. Please see below photos of our TC meeting at a classroom (conference room) at Ramon Llull University-La Salle. Next year, ICME 2012 will be held in Melbourne, Australia, Prof. Jianfei Cai, Nanyang Technological University, Singapore, has been elected to lead this event as the TPC Chair representing MMTC. Please do not hesitate to contact Prof. Cai directly for serving opportunities to support this important event.

Our next TC meeting will be held on 10:00-12:00, December 7, 2011 at GLOBECOM 2011 (Houston). We encourage all our TC members to attend this meeting.

On the other hand, we would like to congratulate Drs. Xiaoli Chu, Yung-Hisang Lu, Thomas Magedanz, and Jiangtao Wen, the Co-Chairs of MMTC's first annual workshop, MMCOM 2011, to be held in conjunction with the GLOBECOM 2011. Clearly this event is impossible to be successful without their tremendous efforts. The workshop program consists of Keynote talks and paper sessions on topics including wireless communications, video coding, green computing and next generation network infrastructures. We invite our members to participate in this workshop, more details of this event can be found at http://www.fuseco-workshop.org/mmcom2011/

As usual, it is time to nominate the following Awards (current TC chairs/vice chairs and Award board members and chairs are not eligible to be nominated):

- Distinguished Service Award: this award is set to recognize long-term TC member who had made significant contribution to this community;
- Outstanding Leadership Award: this award is set to recognize our IG Chairs, Co-Chairs, or Board Directors, who have made significant contribution to TC.

Please send your nomination to me at haohongwang@gmail.com by **October 15, 2011** with a justification letter in MS-Word format.

At the meantime, the Award Board will start to meet to select the Best Paper Awards winners from the recommended papers in the past R-Letters. If you think a paper is in Award quality, please help to nominate it to our R-Letter Board, as all papers reviewed in the R-Letter will automatically enter the pool of Best Paper Award nominees.

Again, we look forward to meeting you all at GLOBECOM 2011 in both TC meeting and MMCOM workshop. Thank you very much!

Haohong Wang
Chair of Multimedia Communication TC of IEEE ComSoc

SPECIAL ISSUE ON SECURE MEDIA STREAMING

Secure Media Streaming

Yufeng Shan, Cisco Systems, USA
yshan@cisco.com

Video is the “killer app” in the current Internet. With the explosive deployment of mobile devices and new IPTV services, security becomes more and more important in media streaming. The objective of the special issue of the E-Letter on secure media streaming is to identify and promote necessary enabling techniques in this area.

In the first paper, “Secure Media Streaming: a Joint Layer approach ”, the authors make the best use of the knowledge on the importance of source encoded media contents and the channel conditions to build up a joint layered coding scheme for unified reliable and secure media transmission over wireless channels. This scheme integrates the authentication into the media error protection components to ensure that every source-decodable media unit is authenticable. The proposed scheme protects both the source coded media stream and the authentication data from channel impairments.

In the second paper, “Content Dissemination And Protection. In Socially Consumed Video: A Network Coding Approach”, the authors summarize recent research in network coding (NC) for multi-network collaborative communication systems and multiscreen content, especially in a social viewing environment. They propose an approach for distributed content verification without the need to contact a centralized trusted authority. This relies on constructing homomorphic encryption. The encryption remains decryptable under coding, without the need to perform decoding, even at a single packet level for peer-to-peer system with untrusted nodes.

In the third paper, “Secure Adaptive Video Streaming in the Compressed-Domain: Benefits and Drawbacks”, the authors investigated a compressed-domain adaptive video streaming method which is to adapt H.264 videos for a target screen size, processing power, and/or bandwidth. A DCT domain watermarking approach as well as a macroblock and a slice-based encryption technique for real-time communication have been designed and implemented. This approach to secure adaptive streaming gave the flexibility of codec independence to achieve compressed-domain authentication, encryption, and spatio-temporal adaption of the encoded video.

In the forth paper, “Video Streaming using Multiple Description FEC and Network Coding”, the authors proposed methods to improve the user experience through combinations of scalable video coding, multiple-description FEC (MD-FEC), and practical network coding. They studied variations on multiple description (MD) coding together with forward error-control coding (FEC) and practical network coding (PNC) using real distortion-rate curves from state-of-art scalable video coders. Present work is connected with grouping the receivers for MD-FEC and extensions of MD-PNC to accommodate scalability in frame rate.

Yufeng Shan is an R&D engineer at Cisco Systems, USA. He performs research and development primarily in the areas of video and networking, including video coding, processing, streaming and network protocols. His other interests are in high performance and high reliability software design and development, distributed systems and networked embedded systems. He was a Visiting Professor at INRIA, Sophia Antipolis, France 2000, and a Research Scientist at Department of ECSE, University of California Berkeley 2001-2002. Dr. Shan is a Senior Member of IEEE.

Secure Media Streaming : A Joint Layered Approach

Xinglei Zhu and Chang Wen Chen, State University of New York at Buffalo, Buffalo, NY
{xzhu4, chencw}@buffalo.edu

1. Introduction

Secure media streaming becomes a critical issue in the rapid growing media streaming applications because of the public accessibility of the content delivery network in the media streaming system. In particular, in the popular wireless media applications, the media stream is unavoidably accessible to third party and thus is more vulnerable to malicious attacks. Therefore it is necessary to enable the end users to authenticate the integrity of the media source and media content, i.e. to verify that the received media is indeed sent by the claimed sender and the media content is not maliciously changed.

Traditional digital signature stands out to be a natural solution to the above sender identification and integrity verification issues. It verifies the data in a strict fashion such that even a flip of a single bit will fail the verification. This strict verification fashion prevents the direct application of data authentication scheme on media streaming because in media streaming, it should be the integrity of the media content rather than strict data integrity to be verified. Furthermore, channel loss is usually inevitable, particularly in wireless networks, and retransmission is not desired in real-time media streaming applications.

To utilize the desired provable security property of digital signature while being able to address the channel loss issue, typical stream authentication schemes [1]-[3] partite the media stream into small units, usually transmission packets, which are the basic authentication units. Robust authentication performance against channel error is achieved by keep appending authentication information of units to other units to form a directed acyclic graph (DAG) and finally the sink noe is signed using digital signature. If a unit is damaged, it will be considered as useless and will be removed from the DAG. Because there are multiple paths from a unit to the signature, it has high authentication probability even if some units are damaged or lost during transmission.

Designing efficient secure media streaming scheme is challenging due to several reasons. First, to achieve high authentication probability over lossy channel, multiple copies of authentication data of a unit are appended to other units. Therefore high authentication probability and low overhead become two inherent conflicting requirements under the traditional secure media streaming framework. Second, in media streaming applications, the quality of the received media stream is very important, while traditional schemes usually focuses only on the security merits without taking the media content characters into consideration. Although recent research works [4][5] try to address these challenges by joint consideration of media content and channel condition, there is still a pressing need of a paradigm shifting scheme for secure media streaming.

In this research, we make the best use of the knowledge on the importance of source encoded media contents and the channel conditions to build up a joint layered coding scheme for unified reliable and secure media transmission over wireless channels. This scheme integrates the authentication into the media error protection components to ensure that every source-decodable media unit is authenticable. The proposed scheme protects both the source coded media stream and the authentication data from channel impairments. The proposed scheme is fundamentally different from many existing schemes that consider the security separately from other components in the media streaming system.

2. Proposed Joint Layered Scheme

The general mobile media transmission context is briefly illustrated in Figure 1. The original media content is first encoded into codestream with source coding standard. In this paper to avoid confusion between media packet and network packet, we use "*slice*" to denote the encoded media unit and "*packet*" to denote the network transmission unit. The codestream is further packetized with optional error resilient techniques, for delivery over the lossy channel. The transmission channel is assumed to be neither reliable nor secure, i.e. packets transmitted over the channel may be lost (e.g. due to network congestion) or modified by malicious attacks before being received. At the receiver, the received packets are assembled and the reconstructed media content is consumed.

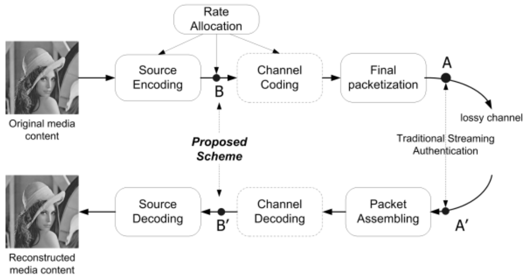


Fig. 1. Media delivery over lossy channel. a) A: traditional streaming authentication generation; A': traditional streaming authentication verification b) B: proposed authentication generation; B': proposed authentication verification

Let the encoded media stream have the format of a sequence of T slices $M_1, \ldots, M_T$ with utility $U_1,\ldots, U_T$ and size $S_1,\ldots, S_T$ respectively. Here the utility is defined as the amount by which the overall distortion is reduced if slice M_i is consumed. We assume that each slice is coding dependent only on previous slices. This can be achieved by proper sorting of the source slices.

At the server site the compressed media slices are protected by channel coding at L different levels using $(N, k_1),\ldots, (N, k_L)$ erasure codes respectively, where k_i refers to the number of packets required to recover the i^{th} slice. Next, a hash is generated for all the media slices at each level and appended to the media slices at the immediate higher level, until level 1 is reached. In other words, the media slices being protected at the same level are considered as a uniform group for the purpose of both error protection and authenticity protection. Compared with the transmission packet based scheme, the number of hash required is greatly reduced because the number of error protection levels is much less than the number of packets. Finally from the media slices at level 1 and the appended hash from level 2, a hash is generated and signed. The signature would also be protected at level 1, same as the most important media slices. After all the media slices are authenticated, the error control codes based channel coding is applied uniformly on the media slices and on the authentication information within each protection level, followed by packetization for transmission. This process is shown in Figure 2.

At the receiver site, if the number of the received packets is no less than k_i, all the slices and authentication information at i^{th} level and higher levels can be successfully channel decoded. Since the information to authenticate slices at i^{th} level is protected at higher level, these slices are verifiable as long as they can be correctly reconstructed in the proposed system. Hence, by the nature of the proposed scheme, all the decodable media slices are verifiable. Finally, the authenticated media slices are source decoded and assembled to produce media content.

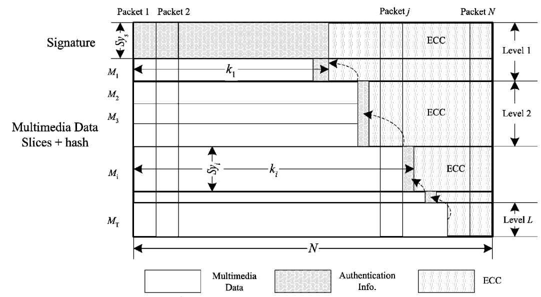


Fig. 2. Structure of transmission packets. The dashed arrow represent hash appending.

3. Analysis of the Proposed Scheme

Under the proposed scheme, the media quality degradation issues in the existing schemes can be simultaneously resolved. We elaborate the details of these benefits in the following:

1. *Coding dependency impact*: In this joint protection scheme, the authentication dependency has been made consistent with the coding dependency. Therefore, no additional quality degradation will be introduced by the mismatch of verifiable media slices and decodable media slices.

2. *Channel impairment impact*: The design of unequal error protection for the layered coding-based media authentication scheme will enable an optimal rate allocation (described in next Section) so as to minimize the channel impairment impact.

3. *Guaranteed authentication verification*: The smart placement of hash data under layered coding based scheme is able to achieve 100% verification probability. Hence, the potential additional distortion introduced by unverifiable media slices can be eliminated.

4. *Minimum authentication overhead*: In the proposed scheme, only one hash is generated for all media slices at any given protection level. Because the number of protection levels is generally much less than the number of the transmission packets or the number of media slices, the overhead introduced by authentication is minimized. Traditionally, two or more copies of each hash per packet/slice are needed in order to maintain high verification probability.

Since the proposed scheme does not change the coding dependency arbitrary truncation in transcoding does not affect the verification of the transcoded codestream. Hence the proposed scheme fits into the wireless transmission well in the means of compatibility with transcoding.

In summary, the proposed joint error and authentication protection scheme is able to simultaneously address several issues in media quality degradation in wireless networks. Given the rate-distortion knowledge of the source encoded media slices, optimal rate allocation scheme can be developed to achieve the optimal end-to-end quality under the proposed scheme.

4. Results

We have implemented an optimal rate allocation scheme and conducted the experiment on JPEG-2000 images [6] and H.264/AVC video [7]. In this section we report the result on JPEG2000 encoded images.

The network is modeled by an i.i.d distributed packet loss rate channel with the average packet loss rate ranges from 0.01 to 0.3. The number of transmission packets is set to 200. Reed-Solomon code is used for erasure coding. The hash function is SHA-1 with hash size 160 bits. A signature of length 1024 bits is generated by RSA for each picture.

We present the end-to-end R-D curve of the JMEAP scheme and compare the performance with the existing schemes in Figure 3. Optimal unequal error protection (OUEP) is the upper bound in which there is no authentication but only error protection. Joint Layered scheme is the one proposed in this paper. Joint source-channel-authentication (JSCA) [5] is the state-of-the-art joint designed scheme. SAIDA [8] is an error correction code based scheme and we implemented two different version, SAIDA with joint source-channel (JSC) and SAIDA with equal protection (EP) to address the channel errors. The proposed joint layered design provides the highest end-to-end quality. Due to the low authentication overhead, the proposed scheme suffers only 0.1 to 0.3dB quality decrease compared with the upper bound – OUEP scheme in which no authentication is applied. In comparison with the state-of-the-art JSCA scheme, there is a performance improvement in the range between 0.5dB and 4dB.

5. Conclusion

A novel joint layered scheme for unified reliable and secure media streaming over lossy networks is presented. This scheme simultaneously protects both compressed media content and the authentication data from wireless channel impairments. All source decodable media slices are verifiable and the authentication overhead is reduced to a very low level. Therefore, the media quality degradation incurred by both channel noise and authentication constraints can be minimized with the proposed scheme.

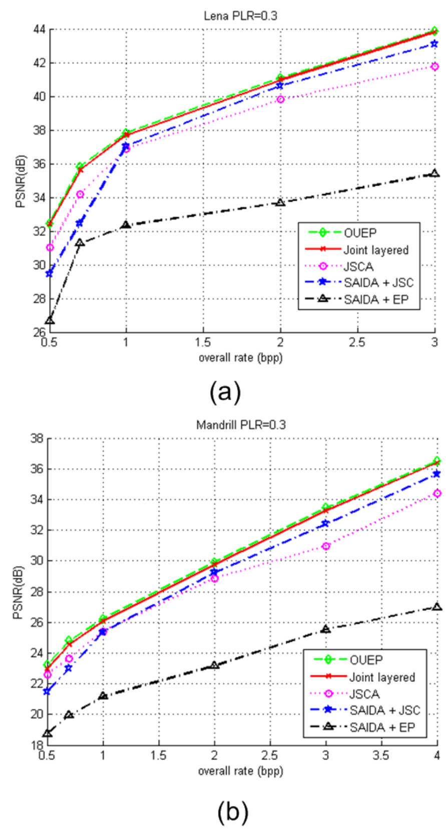


Fig. 3. End-to-end R-D curves on Lina and Mandrill, at packet loss rate PLR=0.3. (a) Lena (b) Mandrill.

References

[1] A. Perrig, R. Canetti, J. Tygar and D.Song, “Efficient Authentication and Signing of Multicast Streams over Lossy Channels”, IEEE Symposium on Security and Privacy, 2000.

[2] R. Gennaro, P.Rohatgi, “How to Sign Digital Streams”, Advances in Cryptology, pp.180-197, 1997.

[3] C. S. Lu and H. Y. M. Liao, “Structure Digital Signature for image authentication: An Incidental Distortion Resistant Scheme”, IEEE Trans. on Multimedia, vol. 5, no. 2, pp. 161-173, 2003.

[4] Z. Zhang, Q. Sun, J. Apostolopoulos, W.-C. Wong, “Rate-Distortion-Authentication Optimized Streaming with Generalized Butterfly Graph Authentication”,

Proc of ICIP 2008.

[5] Z. Li, Q. Sun, Y. Lian, C. W. Chen, "Joint Source-Channel-Authentication Resource Allocation and Unequal Authenticity Protection for Multimedia Over Wireless Networks", IEEE Trans. on Multimedia, vol. 9, no. 4, pp. 837-850, June 2007.

[6] X. Zhu, Z. Zhang and C. W. Chen, "A joint layered coding scheme for unified reliable and secure media transmission with implementation on JPEG 2000 images", IEEE Inter.Conf. on Multimedia and Expo (ICME), pp.710-717, 2009.

[7] X. Zhu and C. W. Chen, "A Joint Source-Channel Adaptive Scheme for Wireless H.264 Video Authetication", IEEE Inter.Conf. on Multimedia and Expo (ICME), pp.13-18, 2010.

[8] Park, J., Chong, E., Siegel, H., " Efficient Multicast Stream Authentication Using Erasure Codes, ACM Trans on Information and System Security" vol.6, no.2, pp.258-285, 2003.

Xinglei Zhu received B.E. degree from the University of Science and Technology of China in 2002 and M.S. degree from National University of Singapore in 2004. He is now pursuing the Ph.D. degree in the Department of Computer Science and Engineering, State University of New York at Buffalo. His current research interests include multimedia security, multimedia transmission and multimedia coding.

Chang Wen Chen (F'04) received the B.S. degree from the University of Science and Technology of China in 1983, the M.S.E.E. degree from the University of Southern California, Los Angeles, in 1986, and the Ph.D. degree from the University of Illinois at Urbana-Champaign in 1992.

He is a Professor of Computer Science and Engineering at the State University of New York at Buffalo. Previously, he has been Allen Henry Endow Chair Professor at the Florida Institute of Technology from 2003 to 2007, on the faculty at the University of Rochester from 1992 to 1996, on the faculty at the University of Missouri-Columbia from 1996 to 2003.

Prof. Chen served as the Editor-in-Chief for IEEE Trans. Circuits and Systems for Video Technology for two terms from January 2006 to December 2009. He has been an Editor for numerous IEEE Transactions and Journals, including Proceedings of IEEE, IEEE Journal of Selected Areas in Communications, IEEE Trans. Multimedia, and IEEE Multimedia Magazine. He has also served as Conference Chair for several major IEEE, ACM and SPIE conferences related to mobile wireless video communications and signal processing. He has received numerous research, service, and best paper awards, including the 2003 Sigma Xi Excellence in Graduate Research Mentoring Award and 2009 Alexander von Humboldt Research Award. He is also an SPIE Fellow.

Content Dissemination And Protection In Socially Consumed Video: A Network Coding Approach

Marie-José Montpetit and Muriel Médard, Massachusetts Institute of Technology, Cambridge MA, USA
João Barros, University of Porto, Portugal,
Frank Fitzek,University of Aalborg, Denmark
{mariejo, medard}@mit.edu, jbarros@fe.up.pt, ff@es.aau.dk

1. Key Challenges in Video Transmission

The behavior of video consumers has evolved tremendously in the past five years. Traditional television delivery mechanisms, based on content that is acquired and distributed to a single end device under the control of a single operator, is in upheaval and moving from classical broadcast to web based platforms. DVDs are already in the process of being replaced by streaming services. The ubiquitous TV set is becoming a much more personal multi-device ecosystem. Video consumption is now geo-localized and shifted in time and place. Furthermore, video streaming is now used in combination with Web 2.0 messaging and widgets. Television content is consumed more and more in a social way over multi-screens and heterogeneous environments.

Clearly, novel networking approaches are needed to provide the quality video experience demanded by today's users over any network in and out of the home. At the same time, the demands of content providers and consumers for secure and private mechanisms to procure and protect video content must be met in a convincing manner. Since the main content is often combined with ancillary features inserted anywhere in the network, traditional end-to-end architectures result in inefficient implementations.

On the other hand, the challenges of video-rich data distribution over wireless are growing with the uncompromising demand for mobile content. Missed opportunities for video transmission over wireless are mainly due to the multitude and limitations of devices and networks.

It is our contention that the solution to surviving this video glut is not to add more resources to the silos. The key is to use the available networks and rendering devices opportunistically and cooperatively. This includes (a) improving the reliability of the connected information nodes, (b) adding content protection where needed, and (c) minimizing access to core elements over expensive wireless access networks. In essence, the goal is to create a truly "video centric wireless network" with "streamware" that can be adapted to local transmission and existing requirements in terms of privacy and content protection.

The purpose of this overview paper is thus to summarize recent research in network coding (NC) for multi-network collaborative communication systems and multiscreen content. NC is paving new ground for novel services in particular as it applies to video streaming with high quality of experience and security guarantees in a social viewing environment.

2. Network Coding in a Nutshell

NC views digital traffic not merely as groups of bits but rather as algebraic entities. NC adds functionality to networking elements beyond the usual forwarding of data units from one interface to another. It assumes that added functionality inside the network can be provided with simple finite field arithmetic: packets can be multiplied by coefficients and combined additively. NC transforms the usual data stream into a set of linearly combined entities that can be recovered with simple equation solving algorithms.

Having been a very active topic of research in the last 10 years, NC is now moving into the real world, particularly in streaming and storage applications. This aligns it to emerging content-centric architectures for future Internet.

Since network codes do not need to be end to end, NC allows for the tailoring of coding strategies to the dynamics and topology of the network, as well as to the features of the receiver ecosystem. This is very beneficial when streaming video traffic over heterogeneous networks to a variety of end devices. NC allows content to be modified and/or stored in the network nodes without tight controls. Protection can be added only where needed, thus freeing vital resources and enabling peer-to-peer distribution with local features. This is particularly of interest for "social distribution" over small community networks. With network coded video there is no need to know exactly where a piece of video is located: to regenerate a file there is only a need to gather "useful" combinations of content until enough degrees of freedom have been accumulated to allow for perfect decoding. There is a small trade-off between using a complex implementation, which requires state information to be kept in network nodes, and the added acquisition and decoding delay of NC. Our research has shown that there exist coded approaches, particularly quasi-systematic coding, that can provide the required streams within the acceptable completion requirement of video delivery networks and indeed improve over approaches requiring state information.

3. New Strategies for Secure Dissemination of Video Streams

There are now major incentives to address delivery mechanisms beyond traditional client-server architectures. Because of the scarcity of wireless resources, we observe a growing requirement for content to be shared locally without wasting bottleneck resources for digital rights management (DRM), retransmission of lost segments and other non-revenue generating traffic. Peer to peer and distributed storage as well as quality of experience (QoE) with device augmentation have emerged as key enablers for layered content protection, which socializes viewing over the device ecosystem. However, without network coding the complexity of distributed approaches increase rapidly and lead to inefficient implementations and reduced scalability. NC creates the underlying network infrastructure that will deliver the next generation of streaming protocols.

It is important to stress that getting the information to its destination is not sufficient for the business success of any strategy: the intellectual property rights of content generated commercially or by private users must be protected, whereas user privacy cannot be compromised. Traditional encryption and DRM approaches need to be updated to meet the bandwidth constraints of wireless networks, the delay requirements of user applications and the need to separate business models from content protection. Commercially produced content shared among devices and "friends" that gets annotated and enhanced in this exchange is not suitable for traditional DRM. Key exchanges necessitating numerous cross-network round trips are inappropriately using precious wireless bandwidth and affecting QoE through excessive delay. Encryption and decryption of large packets stress the capabilities of end-devices by generating added delays, wasting bottleneck resources and compromising end-to-end delay budgets.

We propose an approach for distributed content verification without the need to contact a centralized trusted authority. This relies on constructing homomorphic encryption. The encryption remains decryptable under coding, without the need to perform decoding, even at a single packet level for peer-to-peer system with untrusted nodes. Our system uses standard schemes based on verifying if the received instance belongs to a subspace generated by the data itself. The power of this approach is that it allows us to consider distributed verification of data, whether it originates from a single source or is provided in a distributed fashion. The original content will be a subspace of the content with additional user-generated content. The original content can be seen as consisting of the total possible space with the portions allotted to users' content as having been set to having zero coefficients. Thus, even in the presence of user content, the original content may be verified. Note that original content may be reduced (in effect setting some portions of the contributions to be nil) but not modified. A modification of the original content will lead to rejection of the homomorphic signature for any packet. We may also deal with modifications through corrections. The goal is to take content that was modified and either restore it to its original state or discard it without the need to resort to techniques such as hash value computation and exchanges with trusted authorities.

4. Implementation

Our approach to content protection is currently being implemented on Android, iOS and Symbian based smart phones using a peer-to-peer NC dissemination system developed at the University of Aalborg. The use case is simple. Registered users have ready access to content that is self-verifying and distributed locally. For unregistered users the current implementation uses advertisements to carry the required keys and signature but other mechanisms could also be implemented. Additionally there can be a mode where some of the received packets at any node would need to be verified against the source (or its proxy) to add more protection. While this needs to go back to a centralized server the amount of exchanged information remains very small when compared to traditional approaches. Moreover, we have flexibility in which data is sent, and when it is sent, since any coded packet will adequately represent the whole space from which it was generated.

Figure 1 - NC video stream rendered on the iPod Touch.

5. Conclusions

We have proposed to use network coding and the inherent peering of wireless nodes to provide a distributed video dissemination network with better use of scarce resources. This results in distributed content verification without the need for a trusted authority. Our techniques build upon our earlier work on constructing secure storage over unsecured networks. The encryption of the NC coefficients is very lightweight hence the decryption is significantly faster than traditional methods and allows us to add further functionality. Thus, we can define an approach where commercial content is signed in a centralized way and the ancillary content is signed within the viewer group. Although outside peers can help to disseminate the video content, they will not be able to decode it without the right signature or see the viewer comments without the author's authorized signature.

This approach combines a number of elements: (a) proximity networking to take advantage of the availability of edge resources, (b) network coding to protect and secure the video content and (c) homomorphic signatures to allow the identification of multiple content sources in converged streaming applications.

Acknowledgements

The authors would like to thank Raluca Ada Popa of the MIT Computer Science and Artificial Intelligence Laboratory for her advice on the use of homomorphic signatures. NBC Universal, the Fundação para a Ciência e Tecnologia (MIT Portugal) and the European Commission are also acknowledged for their support to the research.

References

[1] MIT News, "The Power of Random", http://web.mit.edu/newsoffice/2010/network-coding-part1.html

[2] MIT News, "Rethinking Networking", http://web.mit.edu/newsoffice/2010/network-coding-part2.html

[3] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Networking*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[4] M. Kim, Lima, L., Zhao, F., Barros, J, Médard, M., Koetter, R., Kalker, T. and Han, K., "On Counteracting Byzantine Attacks in Network Coded Peer-to-Peer Networks", accepted to the *IEEE Journal on Selected Areas in Communications: Special Issue on Mission-Critical Infrastructure, 2010.*

[5] P. F. Oliveira, L. Lima, T. T. V. Vinhoza, M. Médard, J. Barros, "Trusted Storage over Untrusted Networks.", *Globecom 2010*.

[6] P. Vingelmann and F.H.P. Fitzek and M.V. Pedersen and J. Heide and H. Charaf. "Synchronized Multimedia Streaming on the iPhone Platform with Network Coding." *IEEE Consumer Communications and Networking Conference - Multimedia & Entertainment Networking and Services Track (CCNC), 2011*

[7] M.-J. Montpetit and Médard, M., "Video-centric Network Coding Strategies for 4G Wireless Networks: An Overview", *2010 IEEE Consumer Communications and Networking Conference*.

[8] A. ParandehGheibi, M. Médard, S. Shakkottai and A. Ozdaglar, "Avoiding Interruptions QoE Trade-offs in Block-coded Streaming Media Applications", *ISIT 2010.*

[9] J. Sundararajan, D. Shah, M. Medard, Michael Mitzenmacher, J. Barros, "Network coding meets TCP", *INFOCOM 2009, April 2009.*

[10] M.J. Montpetit and M. Médard, "Community Viewing meets Network Coding:
New Strategies for Distribution, Consumption and Protection of TV Content", *2nd W3C Web and TV Workshop*, February 2011.

Marie-José Montpetit is a research scientist in the Research Laboratory of Electronics at MIT focusing on network coding for video transmission. She was previously an invited scientist at the MIT Media Laboratory where she is still involved in a class on

Dr. Montpetit received a Ph.D. in EECS from the Ecole Polytechnique in Montreal, Canada. She is a member of the IEEE Standing Committee on DSP. She was the recipient of the Motorola Innovation Prize in 2007 for the development of a multi-screen and multi-network video mobility system. Her work on converged video applications, social and multi-screen IPTV has gotten her many invited papers and keynote presentations as well as a MIT Technology Review TR10 in 2010. Dr. Montpetit is a Senior Member of the IEEE.

Muriel Médard is a Professor in the Electrical Engineering and Computer Science at MIT. Professor Médard received B.S. degrees in EECS and in Mathematics in 1989, a B.S. degree in Humanities in 1990, a M.S. degree in EE 1991, and a Sc D. degree in EE in 1995, all from the Massachusetts Institute of Technology (MIT), Cambridge. Professor Médard's research interests are in the areas of network coding and reliable communications, particularly for optical and wireless networks. She was awarded numerous prizes for her publications and was named a 2007 Gilbreth Lecturer by the National Academy of Engineering. Professor Médard is a Fellow of IEEE and member of the Board of Governors of the IEEE Information Theory Society.

João Barros is an Associate Professor of Electrical and Computer Engineering at the University of Porto and the Director of the Instituto de Telecomunições in Porto, Portugal. Since 2008 he has had a visiting appointment with the Massachusetts Institute of Technology (MIT). In February 2009, Dr. Barros was appointed National Director of the Carnegie-Mellon University CMU-Portugal Program. In recent years, João Barros has been PI and Co-PI of numerous national, European and industry funded projects, publishing extensively in the fields of information theory, networking and security, with a special focus on network coding, physical-layer security, sensor networks, and intelligent transportation systems. Dr. Barros has received several awards, including the 2010 IEEE Communications Society Young Researcher Award for the Europe, Middle East and Africa region, the 2011 IEEE ComSoC and Information Theory Society Joint Paper Award, and a state-wide best teaching award by the Bavarian State Ministry of Sciences, Research and the Arts. He received his undergraduate education in Electrical and Computer Engineering from the Universidade do Porto (UP), Portugal and Universitaet Karlsruhe, Germany, and the Ph.D. degree in Electrical Engineering and Information Technology from the Technische Universitaet Muenchen (TUM), Germany.

Frank H. P. Fitzek is a Professor in the department of Electronic Systems, University of Aalborg, Denmark heading the Mobile Device group. He received his diploma (Dipl.-Ing.) degree in electrical engineering from the University of Technology - Rheinisch-Westfälische Technische Hochschule (RWTH) - Aachen, Germany, in 1997 and his Ph.D. (Dr.-Ing.) in Electrical Engineering from the Technical University Berlin, Germany in 2002 and became Adjunct Professor at the University of Ferrara, Italy in the same year. He co-founded the start-up company acticom GmbH in Berlin in 1999. He has visited various research institutes including Massachusetts Institute of Technology (MIT), VTT, and Arizona State University. In 2005 he won the YRP award for the work on MIMO MDC and received the Young Elite Researcher Award of Denmark. He was selected to receive the NOKIA Champion Award several times in a row from 2007 to 2011. In 2008 he was awarded the Nokia Achievement Award for his work on cooperative networks. In 2011 he received the SAPERE AUDE research grant from the Danish government. His current research interests are in the areas of wireless and mobile communication networks, mobile phone programming, network coding, cross layer as well as energy efficient protocol design and cooperative networking.

.

Secure Adaptive Video Streaming in the Compressed-Domain: Benefits and Drawbacks

Razib Iqbal and Shervin Shirmohammadi, Distributed and Collaborative Virtual Environment Research (DISCOVER) Laboratory, School of Electrical Engineering and Computer Science, University of Ottawa, Canada
Shervin Shirmohammadi, Multimedia Processing Laboratory, School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Iran
{riqbal, shervin}@discover.uottawa.ca, sshirmohammadi@ut.ac.ir

1. Compressed Domain Adaptation

Today's multimedia communication platforms involve the coexistence of a number of complementary as well as competing access, delivery, and consumption technologies. While there is a wealth of multimedia data on the Internet today, the delivery path of multimedia content to a receiver is not straightforward in the heterogeneous world of end devices. On one side, smart handhelds with built-in Wi-Fi and rich media capabilities continue to enter the consumer market on a daily basis, while on the other side variations of these devices and their network access technologies make it challenging to transmit and render video on them. Eventually, diverse network types, device variations, user preferences, and video codec types lead to manifold of possible adaptation requirements. In order to reduce complexity and to support the concept of Universal Multimedia Access (UMA), it is desired to have an effective adaptation system that is generic and format independent.

At the same time, while Peer-to-Peer (P2P) content sharing has gained momentum in the Internet world, limited availability of device-to-device video P2P solutions is a particular area that is still lagging. Until now, the video sharing/streaming concept has been limited to sharing only bandwidth, and little has been done to investigate the utilization of idle CPU resources of the participating peers, which can be utilized to perform adaptation operations in an intermediary peer to serve a less capable peer like a low-end mobile handheld device.

The combination of the above two challenges; i.e., adaptation, and P2P video streaming, motivated us to investigate compressed-domain adaptive video streaming [1][9][10][11][12][13]. Our goal of compressed-domain adaptive streaming is to adapt H.264 videos for a target screen size, processing power, and/or bandwidth, by reducing frame rate and/or spatial size while streaming the video, where both streaming and adaptation are performed by the peers themselves and without the need for centralized servers. For compressed-domain adaptive streaming, we relied on metadata description (gBSD) of the compressed bitstream as shown in Figure 1 [2]. In this scheme, an adapted version of the video is retrieved by removing segments

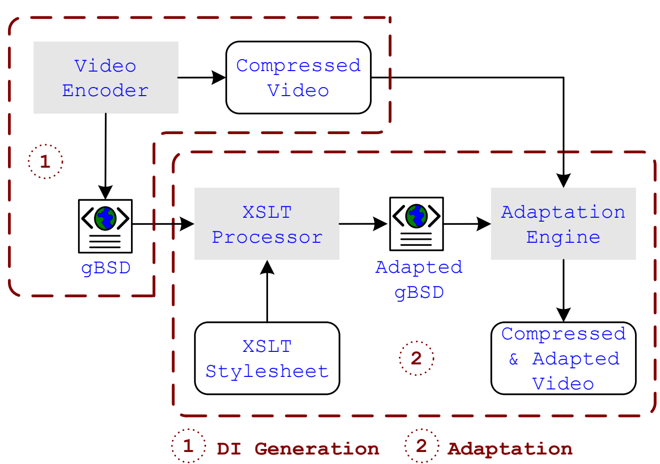


Figure 1. Metadata based adaptation. DI: Digital Item (video, in this case). See [2] for details.

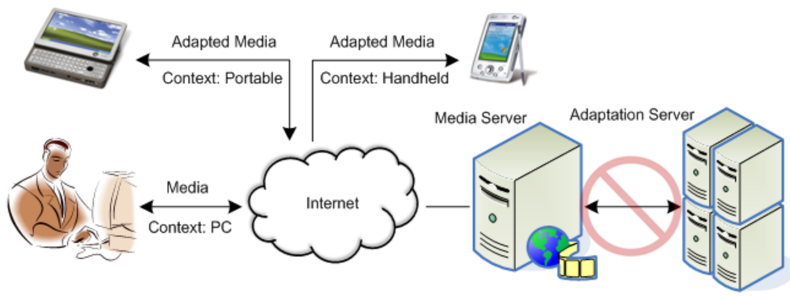


Figure 2. Online adaptation and streaming by the peers themselves.

belonging to the temporal domain or the spatial domain, depending on the desired video output. What makes compressed-domain adaptation different from other adaptation schemes is that the intermediary node which performs the adaptation, for example a peer in the P2P delivery network, can perform the adaptation without decoding and re-encoding of the video, and simply by directly manipulating the compressed bitstream as specified by the metadata. This has two advantages: 1- the intermediary node does not need to have codec knowledge and complexity, and 2- the process is faster than cascaded decoding and re-encoding of the video. This minimizes or even removes (in ideal cases) the need for dedicated adaptation servers, as shown in Figure 2. Cascaded decoding and encoding operations also requires a large amount of processing time or power on the intermediary nodes. An alternative solution would be to produce and store content in several formats taking into account a wide variety of possible user devices and preferences, making an appropriate selection at the delivery time. However, new consumer devices with different capabilities and types make it eventually impractical to pre-store all media for all possible contexts.

2. Secure Adaptive Video Streaming

The choice of metadata support also allows us to offer security as an embedded feature of the adaptation practice. A DCT domain watermarking approach [3][14] as well as a macroblock [4][15] and a slice-based encryption technique [5] for real-time communication have been designed and implemented which are resilient to compressed-domain adaptation operations. Our approach to secure adaptive streaming gave us the flexibility of codec independence to achieve compressed-domain authentication, encryption, and spatio-temporal adaption of the encoded video. Knowledge of the bitstream syntax and the hierarchical structure of metadata enable adaptation, authentication and encryption processes to be customized and easily deployed in an intermediary node such as a peer in a P2P network. Figures 3 and 4 show respectively the encryption and decryption operations in the compressed domain, where: C= compressed, A= adapted, and E= encrypted.

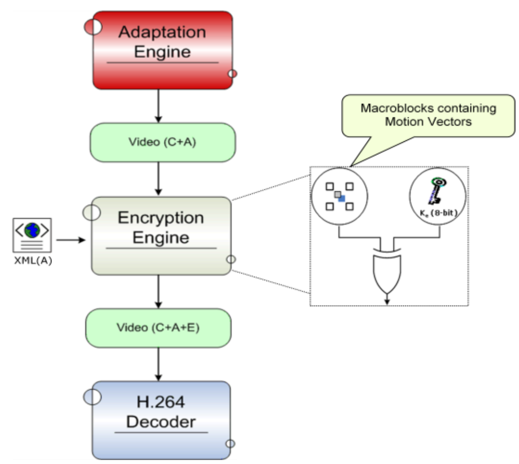


Figure 3. Metadata based encryption; see [4][14].

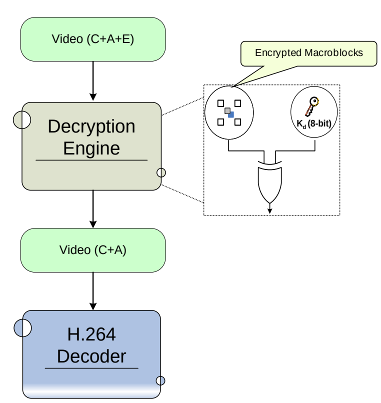


Figure 4. Metadata based decryption; see [4][14].

For authentication, on the receiver side the original content is not required; rather, a separate authenticator can verify the validity of the received video data. The authenticator can be independent of the decoder, in which case there will be no lag added while decoding the video. Now, in order to achieve multipoint-adaptation of the compressed bitstream, the bitstream description needs to be preserved. Certainly, metadata preservation, delivery and processing for decryption, authentication, and adaptation lead to an additional overhead. Our experimental results showed that the size of the metadata usually varies from 5 to 10 percent of the encoded video file size for all three compressed-domain operations. However, having a metadata description of the content allows us to operate on selective segments of the content as opposed to applying the encryption and authentication on the whole bitstream. Moreover, we need to generate the metadata only once during the lifetime of a video content (i.e. while encoding the video for the first time). It also enables the watermarking and encryption processes to be customized and easily deployable within the adaptation engine, as shown in figure 5. Experimental results showed that this approach is suitable for real time systems and commercial deployments.

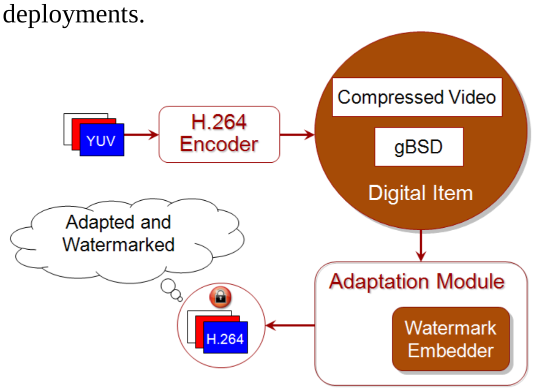


Figure 5. Authentication of live video in the compressed domain [14]

3. Drawbacks

As a tradeoff to the benefits discussed above, secure adaptive video streaming in the compressed domain leads to some drawbacks which are discussed in this section. As mentioned, compressed-domain adaptation approach does add an overhead of generating and transmitting a metadata portion; however, this is balanced by the fact that it reduces the cost of computation and processing at intermediary nodes. Also, depending on the application scenario, if video decryption or verification is insignificant - for example, no encryption or authentication has been made - then transmitting the corresponding metadata can be omitted.

Real-time generation of metadata is another issue for live video streams, where we need to process the stream in chunks in order to generate the metadata. As a consequence, live stream viewers experience a startup delay between the live capture of the video and it’s viewing, equal to the aggregate of the adaptation delay, network transmission delay, and time for receiver side rendering. As such, it might not be practical for conversational applications such as video conferencing. However, we have been able to apply this approach to presentational live videos [5][14][16], as well as a distributed surveillance camera network architecture supporting device heterogeneity [7].

Although scalable or layer coded video [8] could be an option instead of video adaptation, we may not achieve a fine grained adaptation performance in layer encoded video since the higher the number of layers, the less the coding efficiency.

Finally, for the streaming part, the practical impact of our approach might be limited if the participating peers in the P2P network are unwilling to share their idle CPU resources, and for commercial deployments, content distribution stakeholders might be reluctant to adopt the framework standardizing the metadata descriptions.

4. Conclusions

Secure adaptive video streaming in the compressed domain leads to the major benefit of encoding the video only once and adapting it using metadata-based manipulation of the compressed bitstream, which is easier, less computational intensive, and faster than cascaded decoding and re-encoding. Encryption and authentication can also be supported efficiently in the compressed domain, as we have shown. However, such compressed-domain adaptation and streaming requires an initial investment in the encoding process to generate and store metadata, which leads to a slight overhead, and more importantly might not be fast enough for conversational applications such as video conferencing, although it can be uses for presentational live video as our experiments proved.

References

[1] R. Iqbal and S. Shirmohammadi, “DAg-stream: Distributed Video Adaptation for Overlay Streaming to Heterogeneous Devices”, Peer-to-Peer Networking and Applications, Springer, Vol. 2, No. 3, 2009, pp. 202 - 216.

[2] R. Iqbal, S. Shirmohammadi, A. El Saddik, and J.

Zhao, "Compressed-Domain Video Processing for Adaptation, Encryption, and Authentication", IEEE Multimedia, Vol. 15, No. 2, April-June 2008, pp. 38-50.

[3] S. Shahabuddin, R. Iqbal, S. Shirmohammadi, and J. Zhao, "Compressed-Domain Temporal Adaptation-Resilient Watermarking For H.264 Video Authentication", Proc. IEEE Workshop on Multimedia Security and Content Protection, in Proc. IEEE International Conference on Multimedia & Expo, New York, USA, June 28 – July 3 2009, pp. 1752 – 1755.

[4] R. Iqbal, S. Shirmohammadi, and J. Zhao, "Hard Authentication of H.264 Video Applying MPEG-21 Generic Bitstream Syntax Description (GBSD)", Proc. IEEE International Conference on Multimedia and Expo, Beijing, China, July 2-5 2007, pp. 875 – 878.

[5] R. Iqbal, S. Shahabuddin, and S. Shirmohammadi, "Compressed-Domain Spatial Adaptation Resilient Perceptual Encryp

[6] tion of Live H.264 Video", Proc. International Conference on Information Sciences, Signal Processing and their Applications, Kuala Lumpur, Malaysia, May 10-13 2010, pp. 472 – 475.

[7] R. Iqbal, S. Ratti, and S. Shirmohammadi "A Distributed Camera Network Architecture Supporting Video Adaptation", Proc. ACM/IEEE International Conference on Distributed Smart Cameras, Como, Italy, August 30 - September 2 2009, (7 pages).

[8] K. Mokhtarian and M. Hefeeda, "Analysis of Peer-assisted Video-on-Demand Systems with Scalable Video Streams", in Proc. ACM Multimedia Systems, Scottsdale, Arizona, USA, February 22-23, 2010, pp. 133-144.

[9] R. Iqbal, S. Shahabuddin, and S. Shirmohammadi, "Compressed-Domain Spatio-Temporal Adaptation System for Video Delivery", Proc. ACM Multimedia, Beijing, China, October 19 – 24 2009, pp. 1007-1008.

[10] S. Shahabuddin, R. Iqbal, A. Nazari, and S. Shirmohammadi, "Compressed Domain Spatial Adaptation for H.264 Video", Proc. ACM Multimedia, Beijing, China, October 19 – 24 2009, pp. 797-800.

[11] R. Iqbal and S. Shirmohammadi, "A Light-Weight Federated Video Adaptation System for P2P Overlays", Proc. IEEE International Conference on Multimedia & Expo, New York, USA, June 28 – July 3 2009, pp. 1126 – 1129.

[12] R. Iqbal and S. Shirmohammadi, "Online Adaptation for Video Sharing Applications", Proc. ACM Multimedia, Vancouver, Canada, October 27 – 31 2008, pp. 853 – 856.

[13] R. Iqbal, B. Hariri, and S. Shirmohammadi, "Modeling and Evaluation of Overlay Generation Problem for Peer-assisted Video Adaptation and Streaming", Proc. ACM International Workshop on Network and Operating Systems Support for Digital Audio and Video, Braunschweig, Germany, May 28-30, 2008, pp. 87 – 92.

[14] R. Iqbal, S. Shirmohammadi, J. Zhao, "Compressed Domain Authentication of Live Video", Proc. IEEE Int. Conference on Signal Processing and Communication, Dubai, UAE, November 24-27 2007, pp. 1443-1446.

[15] R. Iqbal, S. Shirmohammadi, and A. El Saddik, "A Framework for MPEG-21 DIA Based Adaptation and Perceptual Encryption of H.264 Video", Proc. SPIE/ACM Multimedia Computing and Networking Conference, San Jose, California, Jan. 28. - Feb. 1 2007, pp. 650403-1 – 650403-12.

[16] R. Iqbal, S. Shirmohammadi, and C. Joslin, "MPEG-21 Based Temporal Adaptation of Live H.264 Video", Proc. IEEE International Symposium on Multimedia, San Diego, California, Dec. 11-13 2006, pp. 457 – 464.

Razib Iqbal received his Ph.D. degree in Computer Science from the University of Ottawa, Canada in 2011. His research interests are in the area of Multimedia communication and Distributed systems. He has published over 20 articles in leading journals, conferences, and workshops. In 2010, Dr. Iqbal received the Ottawa Center for Research and Innovation (OCRI) *Student Researcher of the Year Award* in recognition of his significant research accomplishment in Compressed-domain video processing and Peer-to-Peer adaptive video distribution to heterogeneous devices. He has served on the TPC of well-known conferences in his research areas, including ACM Multimedia and IEEE ICME. At present, Dr. Iqbal is affiliated with Amdocs as a Subject Matter Expert in Policy Products Verification for AAA and LTE Technologies.

Shervin Shirmohammadi (*M'04-SM'04*) received his Ph.D. degree in electrical engineering from the University of Ottawa, Canada, where he is currently an Associate Professor at the School of Electrical Engineering and Computer Science, and an Adjunct Professor with the School of Electrical and Computer Engineering, University of Tehran, Iran. He is the Associate Director of the Distributed and Collaborative Virtual Environment Research Laboratory (DISCOVER Lab), and Co-director of the Multimedia Communications Research Laboratory (MCRLab). His research interests are in multimedia systems and networking, specifically in adaptive P2P video streaming, mobile video, 3D video, video analysis, and gaming and virtual environments. The results of his research have lead to more than 190 publications, over a dozen technology transfers to the private sector, and a number of awards and prizes. He is Associate Editor of *ACM Transactions on Multimedia Computing, Communications, and Applications*, *IEEE Transactions on Instrumentation and Measurement*, and Springer's *Multimedia Tools and Applications*, and also chairs or serves on the program committee of a number of conferences in multimedia, virtual environments, and games. Dr. Shirmohammadi is an IEEE Distinguished Lecturer, a University of Ottawa Gold Medalist, a licensed Professional Engineer in Ontario, a Senior Member of the IEEE, and a Professional Member of the ACM.

Video Streaming using Multiple Description FEC and Network Coding

John W. Woods, Koushik Kar, Adarsh K. Ramasubramonian, and Hang Zhang,
Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180
{woods, koushik}@ecse.rpi.edu, {ramasa,zhang10}@rpi.edu

1. Introduction: Video streaming is an important subject because of the high demand for video content on the present day Internet. Due to limited bandwidth and loss constraints, good quality video stresses today's delivery networks. We are working on methods to improve the user experience through combinations of scalable video coding, multiple-description FEC (MD-FEC), and practical network coding.

2. MD-FEC: Multiple-description FEC coding of video can allow us to address many video delivery challenges. With MD-FEC coding, video is coded as multiple *descriptions* (packets), and the coded data is structured in such a way that the "more important" parts of the video are protected more against losses, by appropriate FEC provisioning.

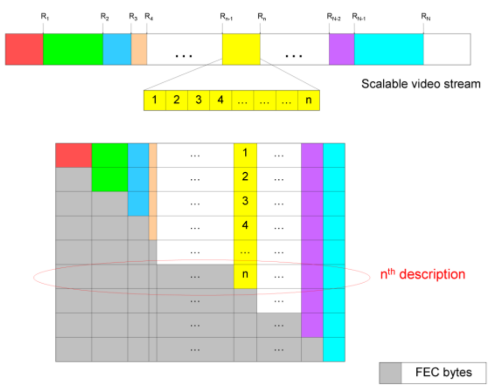


Figure 1. Example MD-FEC packet structure

As shown in Figure 1, MD-FEC video coding nicely adapts itself to changes in available link bandwidths and the packet loss rates. In general, MD-FEC provisions redundancy across different descriptions in a way that is rate-distortion optimal, i.e. for a given number of descriptions and total amount of allowed redundancy, the redundancy is split across the different descriptions such that the total rate distortion measure between the received video (after possible channel losses) and transmitted video is minimized. Puri and Ramchandran [1] have published a widely known method to optimize MD-FEC when the distortion-rate (D-R) function is convex decreasing. Pre-stored D-R functions are used and updated on a group of pictures (GOP) basis, typically every one-half second.

3. Network Coding: Routing of packets has been the primary method by which data have been transmitted across packet networks, such as the Internet. When multiple users are interested in the same data from the same source, multicast routing (or, routing with replication) is typically used. However, routing (with or without replication) is not sufficient to achieve the *homogeneous multicast capacity* of a network. By allowing intermediate nodes in a network to code incoming packets, Ahlswede et al. [2] have shown that the homogeneous multicast capacity equal to the min-cut max-flow capacity of the network can be achieved.

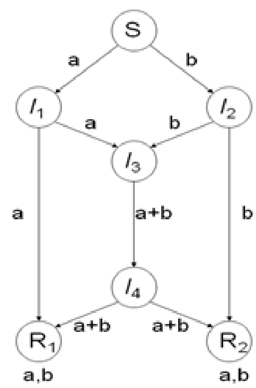


Figure 2. Classic butterfly network from [2]

Figure 2 shows the classic *butterfly network* from Ahlswede et al [2]. In order to get the two messages a and b to both receivers R_1 and R_2, multipath routing has a conflict in the center link I3-I4. But, if we allow node I3 to code a+b, then the conflict goes away in this example, and thruput to R_1 (and R_2) rises from 1.5 to 2 messages per channel use. (Note that arithmetic here is in $GF(2^8)$ to match the common 8-bit byte.)

4. Practical Network Coding: Chou et al. [3] presented a randomized network coding scheme *practical network coding* (PNC), which can be implemented in real networks. In PNC, the data stream is divided into groups, called *generations*, and each generation is grouped into a set of packets (say

N of them). At the source, the packets belonging to the same generation are linearly combined *using random coefficients*. In order to facilitate decoding at the receiver, these random coefficients are attached to the generated packets. At the intermediate nodes, packets are collected in a buffer and grouped according to the generation number. When a sufficient number of packets of a generation arrive, the node linearly combines these packets and sends them to the next node. The encoding vectors that are attached to the packet are also linearly combined. At the receiver, if *N* linearly independent packets of a generation arrive, the entire generation can be decoded. The use of random coding coefficients causes a small probability of non-decodability, but it is quite manageable. Also, the need to transmit the coefficients in a header causes a small overhead.

5. MD-PNC: As suggested in [3], MD-FEC can be combined with PNC. In this synthesis, the Reed Solomon codes of MD-FEC are replaced by the random codes of PNC. We call the combination MD-PNC and present initial results for video coding and network transmission in [4].

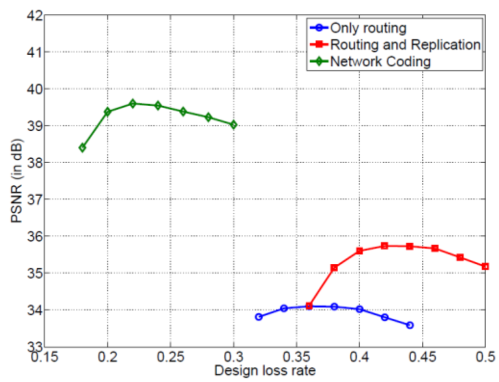


Figure 3. MD-PNC video transmission of *Foreman* (CIF,30fps) on butterfly net of [2].

Figure 3 shows simulation results (NS-2) on the butterfly network for the popular test clip *Foreman* (CIF, 30fps), where we compare unicast (routing), multicast (routing with replication), and MD-PNC. The figure shows a 3 dB advantage for the latter at the high 20% link packet-loss rate. In a more recent work [4], we have introduced *quality of service* (QoS) into MD-PNC and applied the concept to random hierarchical networks. This method can shape video quality based on the receiver's available bandwidth, thus achieving a *heterogeneous multicast*. While the work in [4] is restricted to the lossless case, we have recently extended this result to networks with loss [5].

6. Grouping Receivers: MD-FEC (-PNC) coding provides the flexibility, easy adaptivity and optimization that are desirable for delivering streaming video. However, the total distortion would in general depend on the number of separate streams that can be used to serve a given set of receivers. The total distortion will be minimized if each receiver is served using a different MD-FEC (-PNC) coded video stream, that is optimized for that individual receiver. However, even if the individual receiver bandwidth and/or loss characteristics are known, such an approach does not scale to a large number of receivers, as the video source or relay node has to produce a separate video stream for each receiver. In our recent (yet published) work, therefore, we consider multi-stream MD-FEC coded video delivery to a set of heterogeneous receivers, where practical considerations limit the number of streams that can be produced to serve the receivers of a given video. The question that we address is how a given set of *M* (typically large) receivers can be divided into *Q* (typically small) groups, where each group is served by a separate MD-FEC coded video stream. The overall objective in this grouping strategy is to minimize the total video distortion across all *M* receivers.

Specifically, we study the performance of a *multi-threshold* (sequential) grouping solution, where (Q+1) capacity thresholds $\Delta_0, \Delta_1, \ldots, \Delta_Q$, (with $\Delta_0 = 0$, $\Delta_Q = \infty$) are chosen, and receivers whose path capacities lie within Δ_i and $\Delta_{(i+1)}$ are included in group (i+1), i = 0, …, Q-1. We have found that for linear distortion functions, a multi-threshold solution (when the thresholds $\Delta_0, \Delta_1, \ldots, \Delta_Q$ are chosen "optimally") is also globally optimal. For non-linear distortion functions, while the optimal grouping structure may not be of multi-threshold type in general, numerical experiments with realistic D-R functions reveal that the difference between the optimal grouping solution and the best multi-threshold solution, is typically small. It can also be shown that the optimal thresholds of the multi-threshold solution can be computed in polynomial time using dynamic programming. We also observe decreasing performance returns with increasing values of Q, as we intuitively expect.

Conclusions

We study variations on multiple description (MD) coding together with forward error-control coding (FEC) and practical network coding (PNC) using real distortion-rate curves from state-of-art scalable video coders. Present work is connected with grouping the receivers for MD-FEC and extensions of MD-PNC to accommodate scalability in frame rate.

References

[1] R. Puri and K. Ramchandran. "Multiple description source coding through forward error correction," *Proc. of the 33rd Asilomar Conference on Signals, Systems and Computers*, pages 342–346, October 1999.

[2] R. Ahlswede, N. Cai, S.R. Li, and R.W. Yeung. "Network information flow," *IEEE Transactions on Information Theory*, 46(4):1204–1215, July 2000.

[3] P. Chou, Y. Wu, and K. Jain. "Practical network coding," *Proc. 51st Allerton Conf. Communication, Control and Computing*, October 2003.

[4] A. K. Ramasubramonian and J. W. Woods, "Video multicast using Network Coding," Proc. SPIE VCIP, Jan. 2009.

[5] A. K. Ramasubramonian and J. W. Woods, "Multiple Description Coding and Practical Network Coding for Video Multicast," IEEE Signal Process. Letters, vol. 17, no. 3, March 2010.

[5] A. K. Ramasubramonian and J. W. Woods, "MD-PNC with feedback for heterogeneous video multicast in lossy networks," *Proc. SPIE VIPC*, Jan. 2011, San Francisco.

John W. Woods is Professor in the Department of ECSE at Rensselaer. He is a Fellow of the IEEE. He has authored or co-authoed more than 100 papers in the areas of image and video signal processing and communications. He is the author of *Multidimensional Signal, Image, and Video Processing and Coding*, 2nd Ed., Elsevier and co-author with Henry Stark of *Probability, Statistics, and Random Processes for Engineers*, 4th Ed., Pearson.

Koushik Kar is an Associate Professor in the ECSE department at Rensselaer. He received his B.Tech. degree in Electrical Engineering in 1997 from the Indian Institute of Technology, Kanpur, and his M.S. and Ph.D. degrees in Electrical & Computer Engineering from the University of Maryland, College Park, in 1999 and 2002, respectively. Dr. Kar's primary research interests are in modeling, performance optimization and protocol development for communication networks. Dr. Kar received the CAREER Award from the NSF in 2005, and is an Associate Editor for IEEE/ACM *Transactions in Networking*.

Adarsh K. Ramasubramonian is a doctoral student in Electrical Engineering at Rensselaer. He received his B. Tech degree in Electrical Engineering from Indian Institute of Technology Madras, India in 2006 and his M.S in Electrical Engineering from Rensselaer in 2008. His research interests include signal processing, image/video coding and communications, and network coding.

Hang Zhang is a doctoral student in Computer & Systems Engineering at Rensselaer. His research interests are video networking, and optimization and game theory as applied to network design and performace analysis. He received his B.S. and M.S. degrees in Electrical Engineering from Tsinghua University, China in 2006 and 2009.

TECHNOLOGY ADVANCES

Mobile Cloud Computing

Guest Editor: Kyungtae Kim, NEC Laboratories America, Princeton NJ USA
kyungtae@nec-labs.com

Mobile devices are resource constrained; however mobile service demands greater resources and improved interactivity for better user experience. Resources in cloud computing platforms are a natural fit to remedy the lack of local resources in mobile devices. The objective of the special issue of the E-Letter on Emerging Technology for Mobile Cloud Computing is to identify and promote necessary enabling mobile cloud computing technology, particularly emerging technologies to fully utilize the benefits of cloud computing under dynamic mobile environments.

In the first paper, "Challenges and Enabling Technologies in Mobile Cloud Computing", current state of mobile cloud computing is reviewed and key enabling technology elements are identified. Specifically, it presents collaborative programming abstraction, dynamic workload profiling and scheduling, real-time elastic resource provisioning and privacy protection as key enabling technologies.

The second paper, "Mobile Cloud Computing", presents the current development of the mobicloud system at Arizona State University and its design features and capabilities to assist mobile applications. Several critical research issues are presented to highlight our on-going and future work.

In the third paper, "Mobile Cloud Computing: Opportunities and Challenges", examines key areas where cloud computing benefits mobile computing; (1) recognition (content-based), (2) energy efficiency (computation offload), and (3) disaster rescue and recovery.

The fourth paper, "Enabling Mobile Multimedia with the Cloud", addresses an overview of mobile cloud computing and presents mobile visual search technique migrating computation to cloud-based environment to alleviate resource issues for intensive search and matching operations. It has been shown that the remote rendering ideas works well when there is bandwidth availability and computing is the main constraint.

In the fifth paper, "Cloud Centric Mobile Application Development Using Domain Specific Languages" discusses the difficulty to develop cloud-mobile hybrid applications for heterogeneous mobile front-ends and cloud back-ends. This paper presents the advantage of MobiCloud utilizing proper abstractions by being provided Domain Specific Languages.

The sixth paper, "Energy-Efficient Mobile BitTorrent Solutions", discusses three examples to show the benefits of traffic shaping in peer-to-peer content sharing.

In the seventh paper, "Leveraging Service Clouds for Power and QoS Management for Mobile Devices" discusses the design of a QoS and power management framework to achieve power saving on mobile devices by leveraging the capabilities of service clouds. It has been shown that the proposed approach can achieve up to 70% power savings without sacrificing quality of services.

The eighth paper, "Towards a Secure Personal Cloud Computing for Mobile Thin-Clients" presents an overview of personal cloud computing and the security requirements of the personal cloud. The proposed security framework shows an example to realize the security requirements in the personal cloud computing.

In the ninth paper, "Using the Cloud to Build Content Adaptation Systems for Smart Phones" discusses how content adaptation method provides better user experience for smart phone users. It has been shown that moving content adaptation system to the cloud platform can achieve the best blend of good performance and cost optimization.

The last paper of the issue, "Augment Mobile Cloud Computing with Cognitive Radio", discusses possible opportunities of augmenting mobile cloud computing with cognitive radio and presents opportunistic cloud computing providing dynamic spectrum access, wireless adaptive learning and reconfiguration, and cognitive power management.

Kyungtae Kim received a M.S. degree in Department of Computer Science from the Columbia University in 2000, NY and his Ph.D. degree in the department of Electrical and Computer Engineering at Stony Brook University in 2006. He is working for NEC Laboratories America Inc during 9 years until now in the areas of multimedia interactive communications over the wireless network, cognitive radio, wireless mesh networks, and mobility management over the heterogeneous networks including 802.16/802.11 networks.

Challenges and Enabling Technologies in Mobile Cloud Computing

Kyung-Ah Chang and Il-Pyung Park, Samsung Advanced Institute of Technology, Samsung Electronics, Korea
{kachang, ilpyung.park}@samsung.com

1. Convergence of Mobile Device and the Cloud

Traditional cloud computing is an IT service model for enabling seamless, on-demand network access to computing resources in the form of IaaS (Infrastructure as a Service), PaaS (Platform as a Service) and SaaS (Software as a Service). Recently, with proliferation of mobile devices such as smart phones, tablets and Wi-Fi devices, cloud computing technologies are being slowly migrated to mobile environment. Also mobile applications, with most of data processing and storage occurring in the cloud, are gaining popularity. Although this simple way of utilizing the cloud is very effective, it is still a passive way of using cloud computing technology in mobile devices.

As mobile devices continue to evolve and cloud infrastructures gets increasingly more capable, we envision a proactive way mobile devices and the cloud collaborate to seamlessly provide more intelligent services. However, in order for mobile devices to fully utilize all the benefits of cloud computing, and possibly to establish a new mobile cloud computing paradigm, we need to address the following fundamental technical challenges.

Mobile devices are resource-poor compared to stationary devices. Even as mobile devices continue to evolve and improve, basic mobile related properties such as weight, power and size will always put a limitation on computational resources such as processor speed, memory size, and storage capacity. In mobile cloud computing, mobile device needs to be able to cooperate with the cloud to overcome the resource limitation.

Mobile network is characterized by lower bandwidths, higher error rates, and less reliable connections. Requirements on latency and delay are different per each application through mobile network. In mobile cloud computing, a mobile device should be able to deal with various demands from applications for latency and delay [8].

Mobile devices rely on finite energy source. Enhancement to battery technology has not been able to match the power requirements of the increasing resource demand. Energy efficiency has always been critical for mobile devices and the importance is continuously increasing. In mobile cloud computing, a mobile device should be able to have balanced energy consumption between computation and communication [6, 7].

2. Current State of Mobile Cloud Computing

In this section, we review some existing research efforts in mobile cloud computing [3, 4, 5]. There are three major approaches in current mobile cloud computing technology, which are augmented execution [1], dynamic partitioning of applications and proxy-based mobile cloud [9].

Augmented execution, used by CloneCloud [1], is a method in which a replica of the application and data is kept in the cloud, and the computation is offloaded to the clone in the cloud. It tries to overcome resource and battery limitations of the mobile device by using its clone for computationally intensive applications. In this approach, there is an overhead in the mobile device for determining which function to be sent to the cloud and later for integrating the result from the augmented execution. Another issue is that this requires a reliable connection to the clone in the cloud.

Dynamic partitioning of applications, used by MAUI [2], is a method where applications are partitioned and fine-grain offload of the code is sent to the cloud. Profiling information of the application, network connectivity measurements, bandwidth and latency estimations are used as input parameters to optimize and decide which piece and when it should be offloaded. This method requires developers to manually annotate their code to indicate portions to be offloaded for remote execution. Although theoretically attractive, this has practical issues in that it requires manual application partitioning and that it can be applied only to pre-defined configurations.

Proxy-based mobile cloud method, used by Cloudlet [9], is a system where mobile device uses a nearby local server as a proxy between itself and the distant cloud. When mobile devices do not want to offload to the cloud, they can find a nearby Cloudlet. In this way, mobile users may meet the demand for real-time interactive response by low-latency, one-hop, high-bandwidth wireless access to the Cloudlet. If no cloudlet is available nearby, the mobile device may refer to the default mode that will send requests to a distant cloud. However, it may not be practical to deploy this system in real-world as it requires well-defined infrastructure and maintenance operation for Cloudlet (e.g., Cloudlets should be installed everywhere).

3. Enabling Technologies for Mobile Cloud Computing

In addition to the existing studies, we present the following four enabling technologies to fully overcome fundamental challenges addressed in section 1. The two key questions are which computation in a partitioned program to offload, and how to structure the parallelism across mobile device, local server and the cloud. Key enabling technologies are collaborative programming abstraction, dynamic workload profiling and scheduling, real-time elastic resource provisioning and privacy protection.

Collaborative Programming Abstraction is a technology for a new programming model to hide the complexity of underlying cloud technologies. This model allows run-time system to execute mobile applications among various mobile devices and clouds in a seamless and collaborative way.

Dynamic Workload Profiling and Scheduling is a technology to dynamically identify workload of each task for the purpose of of-floading appropriate amount of computation between mobile device and the cloud based on the profiling information. Based on the information from the optimal cost model derived from workload profiling and scheduling, the cloud system dynamically shift computation between mobile devices and the cloud. The cost model generally consists of execution time, resource consumption, battery life, and network bandwidth.

Real-time Elastic Resource Provisioning is a technology to provide rapid resource provisioning by detecting and managing nearby available and faulty resources. Status of available resources is highly likely to change in dynamic mobile environments, and user transparency to those changes is critical for seamless service provisioning.

Privacy Protection is a technology to protect user data and computation in dynamically changing computing platforms. Due to the above three enabling technologies, a location of computation can be changed at any time during the execution. For example, user data and computation can be processed in mobile device or the cloud or both at the same time. Despite this fact, privacy protection needs to be satisfied throughout the lifecycle of a program regardless of its location.

4. Conclusions

In this paper, we briefly reviewed current state of mobile cloud computing and identified key enabling technology elements needed for proactive cloud computing of the future.

References

[1] B. G. Chun et al, “Augmented Smartphone Applications through CloneCloud Execution”, Workshop on Hot Topics in Operating Systems, USENIX, 2009.

[2] E. Cuervo et al, “MAUI: Making Smartphones Last Longer with Code Offload”, International conference on Mobile systems, applications, and services, ACM, pp. 49-62, 2010.

[3] G. Huerta-Canepa et al, “A Virtual Cloud Computing Provider for Mobile Devices”, Workshop on Mobile Cloud Computing & Services Social Networks and Beyond, ACM, pp. 1-5, 2010.

[4] I. Giurgiu et al, “Calling the Cloud: Enabling Mobile Phones as Interfaces to Cloud Applications”, Conference on Middleware, Springer, pp. 1-20, 2009.

[5] K. S. Rim et al, “Distributed execution for resource-constrained mobile consumer devices”, IEEE Transactions on Consumer Electronics, IEEE, vol.5, pp. 376-384, 2009.

[6] Kumar K. et al, “Cloud computing for mobile users: Can offloading computation save energy?”, IEEE Computer, IEEE, vol.43, pp. 51-56, 2010.

[7] Miettinen A. P. et al, “Energy efficiency of mobile clients in cloud computing”, Conference on Hot topics in cloud computing, USENIX, 2010.

[8] M. R. Ra et al, “Odessa: Enabling Interactive Perception Applications on Mobile Devices”, International Conference on Mobile Systems, Applications and Services, ACM, pp. 43-56, 2011.

[9] M. Satyanarayanan et al, “The Case for VM-Based

Cloudlets in Mobile Computing", IEEE Pervasive Computing, IEEE, vol. 8, no. 4, pp. 14-23, 2009.

Kyung-Ah Chang is a Research Staff Member in Samsung Advanced Institute of Technology of Samsung Electronics. She received her M.S. and Ph.D. in Computer Science from Korea University in 1999 and 2001 respectively. Her areas of expertise and interests include: device centric cloud computing, operating system-level virtualization, distributed computing system and security, and digital rights management on CE devices. She has more than 20 granted and filed EU and US patent applications, and more than 20 international publications.

Il-Pyung Park is Vice President and Director of Intelligent Computing Laboratory in Future IT Research Center of SAIT, Samsung Electronics. He leads advanced software research in various areas, but not limited to, operating systems, cloud computing, storage systems, data analytics, artificial intelligence, digital contents technology, future web technology and advanced user experience. Prior to joining Samsung in 2006, he was Department Head of Security and Platform Technologies at Panasonic Princeton Laboratory. He worked in areas of operating system security, embedded Linux systems, and secure download for software-defined radio. He was one of the key members responsible for establishing CE Linux Forum, and served as the founding chair of the Architecture Group.

From 1999 to 2001, he was Senior Director of Software Development at Timecruiser Computing Corporation, an Internet software company specializing in building Java-based enterprise systems. From 1993 to 1999, he was a faculty member at New York Institute of Technology (NYIT) in the computer science department. During this period, he was also a visiting researcher at Bellcore.
He received B.S. in Computer Science from Seoul National University, M.S. and Ph.D. in Computer Science from Columbia University.

Mobile Cloud Computing

Dijiang Huang, Arizona State University, Arizona, USA

dijiang@asu.edu

1. Introduction

Mobile cloud computing is an emerging cloud service model following the trend to extend the cloud to the edge of networks. It includes numerous mobile devices that are closely associated with their users. They will be directly involved in many cloud activities that extend the cloud boundaries into the entire cyber physical system. As predicted by Gartner, mobile phones will overtake PCs as the most common Web access devices worldwide by 2013 [1]. Thus, mobile devices will become more important and will be involved in almost all aspects of our daily life.

In this letter, we describe what mobile cloud is computing, including its scope, current developments, and research challenges. Our discussion is based on a mobile cloud computing platform that is currently developing at Arizona State University [2]. Then we present applications relying on mobile cloud computing. Finally, we conclude this article.

2. What is mobicloud computing?

Mobile computing research is to study how mobile device sense and learn the status of other devices and the context related to their mobility and networking in order to better access the Internet and/or support mobile applications in an ad hoc communication environment. Cloud computing research mainly focuses on how to manage computing, storage, and communication resources that are shared by multiple users in a virtualized and isolated environment (NIST provides a more completed definition of cloud computing in [3]). Mobile cloud computing cannot be simply illustrated as merging mobile computing and cloud computing technologies.

An illustrative example of mobile cloud computing is how a smart phone can best utilize the cloud resource to reduce its energy consumption. A computing task can be either executed on the mobile device or outsourced to the cloud. Where to compute depends on the overhead tradeoffs between computation and communication while considering the requirements of applications' Quality of Service (QoS) and users' Quality of Experience (QoE) [4]. On one hand, the dual computing model involving both the cloud and the mobile device should minimize the entire system cost, usually with more focus on reducing resource consumptions on mobile devices. On the other hand, a mobile cloud computing service model should improve mobile users' QoE to fulfill their satisfaction when using mobile cloud applications. To address the aforementioned issues, researchers need to have a comprehensive view of mobile cloud computing. If research is limited to within each individual research domain, it will not be sufficient enough to address complex problems arising from the new mobile cloud service model.

Transdisciplinarity of mobile cloud computing research: Mobile cloud computing originally is rooted from interdisciplinary research of mobile computing and cloud computing. Existing research tries to cross the disciplines' boundaries by applying cloud computing solutions into mobile applications or incorporating mobile features when constructing new cloud services. However, the immense information involved in mobile cloud applications and the high complexity of designing mobile cloud applications demanded a new transdisciplinary research to better understanding the natures and principles of mobile cloud computing. We call this new transdisciplinary research as *mobicloud computing*. To simplify the presentation in the following context, the term *mobicloud* is more frequently used.

One important feature of mobicloud applications is functional collaboration. For example, mobile social network based data mining requires collaborations among mobile users. To this end, mobicloud will serve as not only a nexus that interconnects information sources gathered from both cloud computing service domains and mobile computing domains, but also a knowledge center to help mobile users in their daily activities. With these capabilities, mobicloud can assist a mobile entity for discovering other interested entities to construct self-organized mobicloud subdomains that may only have a short lifespan.

IaaS delegation: Mobicloud infrastructure can be established in a bottom-up fashion based on an Infrastructure-as-a-service (IaaS) [3] framework. IaaS provides isolations at the Virtual Machine (VM) level. One or multiple VM(s) can be dedicated to a mobile client providing customized computing, communication, and storage support. Many operations that are not suitable to run on a mobile device, such as multimedia data processing or data mining, can be performed in a VM dedicated to it. We call this feature of mobicloud infrastructure as *IaaS delegation*.

NaaS intra-mobicloud service model: Current IaaS systems lack support to establish a flexible and secure virtual networking environment that can be initiated and managed by mobile users. For example, private data (e.g., pictures, video clips, etc.) can be shared among trusted users through their mobile devices. Based on *IaaS delegation*, the sender can transfer the data into his/her dedicated VM(s), where the data can be shared among all VMs dedicated to trusted data receivers with the data source's permission. Since mobile users can control and operate their dedicated VM(s) within the cloud system, it is desirable that the mobicloud protects the transmitted data. An obvious solution is to establish virtual private networks (VPNs) among VMs within the mobicloud system. Based on existing Open vSwitch solutions [5] at the hypervisor level and OpenFlow switches [6] at the physical network level, VLANs (Virtual LANs) can be deployed to build the intra-mobicloud VPNs.

The software controlled feature of vSwitch and OpenFlow switch enables mobicloud to establish a VLAN-based VPN management framework. Compared to traditional VPN solutions, mobicloud allows VLANs to be established in an ad hoc fashion. Additionally, mobicloud needs to address the need for a large numbers of VLANs with traffic engineering capability that provides bandwidth guarantee to support the application layer intra-mobicloud communication sessions. We call this on-demand network system provisioning as Network-as-a-Service (NaaS) for mobicloud. The described VLAN capability is important for mobicloud computing since inter-VM communication can dominate mobicloud applications. This is fundamentally different from existing remote computing and storage provisioning based cloud applications.

Mobicloud PaaS platform: The described IaaS delegation and NaaS mobicloud service model are restricted by its low-level realization that usually creates a high-level learning barrier for mobicloud application developers. To overcome this software development hurdle, Platform-as-a-Service (PaaS) [3] must be established to provide a standard Application Platform Interface (API). The main challenge in establishing mobicloud PaaS is the compatibility issue among many mobile operating systems in the current market, e.g., Andriod, iOS, Windows 7, Symbian, etc. A general application platform is required to integrate the IaaS delegation and NaaS service model to support different mobile application platforms. To this end, we propose an XMPP [7] plus OSGi [8] (i.e., Extensible Messaging and Presence Protocol plus Open Services Gateway initiative) solution. XMPP is a set of open technologies. It provides a lightweight middleware solution to coordinate the operations among mobicloud components. It also supports multimedia data transmissions and presence services that can be easily integrated as general interfaces for mobicloud PaaS. By interfacing with XMPP, OSGi provides multi-platform supported java programming capabilities. Standard OSGi bundles supporting mobicloud basic functions can be created as a software development kits for easier programming. Distributed OSGi allows mobicloud to run applications in a distributed fashion to balance performance and energy consumptions of mobile devices.

High QoE for mobile users: In mobicloud, QoE measures the subjective feeling of mobile users' experiences when using mobicloud based applications. QoS objectively measures the services such as communication delay, throughput, etc. QoS usually has a strong impact on QoE. However, a better QoS measurement does not mean improved QoE. For example, running powerful CPU on mobile devices may improve the performance of applications; however it may also shorten the device's use time due to high power consumption that will decrease users' QoE. If migrating computing tasks to the cloud is a solution to address the high energy consumption, mobicloud service model needs to consider the trade-offs caused by migration delay, communication overhead, and power consumption due to the migration. Thus, mobicloud needs to consider the relationship between QoS and QoE as a measurement to evaluate the performance of the cloud system. We call this measurement as *QoSE*.

To improve QoSE, mobicloud should adopt a geography-based service model (or geo-based model). This is because mobile users can access the cloud services from any location in the world. The geo-based model can push the cloud computing and storage service nodes to the mobile device as close as possible. In this way, service delay, which is a major

QoSE measure, can be greatly reduced. The geo-based model requires mobicloud server nodes (or clusters) to mirror VMs (or through VM migration) in different geographic locations. The states of the mirrored VMs must be synchronized.

Context awareness between a mobile device and its dedicated VM(s) is another important feature to improve QoSE. Nowadays, a smart mobile device usually serves as an information gateway for the mobile user involving many personalized activities such as checking emails, making an appointment, surfing the web, calling a number, locating some interested spots, etc. Context awareness can be at either the device level or the application level. When it is at the device-level, the device's states such as CPU utilization, available memory and storage, and battery levels need to be learned by its VM to decide whether computing operations should be migrated or not. At the application level, mobile devices can be interfaced to various wireless-capable devices such as sensors, RFIDs, and other smart devices. Mobicloud will serve as a knowledge base to instruct mobile devices on how to locate them and interface with (or control) them? Without the support of mobicloud services, these tasks are difficult to be executed by individual mobile devices. Each mobile device can be considered as a sensing service node for the cloud. The number of mobile devices involved determines the quality of mobicloud services that utilize sensed data. This situation requires us to investigate two critical mobicloud features: security and privacy.

User-centric security and privacy: Mobicloud must be user-centric, i.e., mobile users should have the privilege to manage their own data in the cloud. The trust model of existing cloud service is one-way (or cloud-centric), in which users must trust the cloud services that manage their data. This cloud-centric trust model prevents many users from using cloud services and storing their critical data in the cloud's storage. To address this problem, we propose to build a user-centric trust model for mobicloud. The user-centric model incorporates a tri-rooted VM management model including user root, maintenance root, and auditing root. The user root has the root privilege to manage all security and data storage related functions. The maintenance root performs regular maintenance tasks including networking, software upgrade, and regular cloud resource management for VMs. The auditing root does not have privileges of change, write, or execute in the VM. Instead, it maintains a log to record the VM activities that can be viewed by both the end users and cloud administrators for investigating issues such as misuse of the cloud system and ineligible access to user's data, etc.

2. Mobicloud applications

Mobicloud is a cloud service platform supporting many mobile application scenarios. Here, we just name a few: mobile health, mobile learning, mobile banking, intelligent transportation, smart grid/home, mobile advertising, urban sensing, disaster recovery, mobile entertaining/gaming, mobile social networks, and mobile enterprise solutions. These mobile applications scenarios share some common features that we have discussed previously, which are summarized as follows:

- *User-centric security and privacy protection*: A major incentive for mobile users using mobicloud applications is to protect users' data and allow them to decide what information could be exposed and what information should be kept as secrets.
- *Individual and collective sensing capability*: Mobicloud applications can utilize the sensing capabilities from each mobile device to learn the context of a given application situation.
- *Personalized functions and features*: A mobile device and its associated VM(s) serve as personal information assistants to help mobile users (equipped with several devices) to learn his/her activities and behaviors. These past learned activities and behaviors can help mobile users to correct their bad habits and assist their daily activities.
- *Strong reliability and fail-over protection*: Damage and loss of mobile devices are common due to their small and portable nature. Mobicloud can provide a suite of solutions to protect the mobile users' data and provide data recovery it due to failure, lost (or stolen), and upgrade. Moreover, it provides security detection and monitoring services to prevent malicious traffic from accessing to mobile devices. Furthermore, the remote control feature allows mobile users to eliminate important data even if the device is physically captured.
- *Rich software development platform*: The integrated IaaS delegation and PaaS features allow mobicloud application developers to develop various applications enjoying all features provided by mobicloud. Moreover, it is

easier to integrate various applications in the mobicloud system with little compatibility issue.

- *Function, data, and trust presence*: Dedicated VMs are present 24/7. Functions, data, and trust validations of mobile applications can be delegated to the VMs even if mobile devices and mobile users are not reacheable.
- *Caching capability*: Mobicloud can help mobile devices maintain the states of mobicloud applications. Partially delivered data, lost connections, and half operated functions can be resumed; as a result, this will reduce the overhead due to the uncertainty caused by the mobility of mobile users.

3. Conclusion

In this article, we describe mobile cloud computing, a new transdisciplinary research area based on traditional mobile computing and cloud computing. The description is based on the current development of the mobicloud system at Arizona State University. We lay out the design features of mobicloud and its capabilities to assist modern mobile applications. Several critical research issues are presented to highlight our on-going and future work. We hope this article will help interested researchers and system builders develop new mobicloud features and get a better understanding of the transdiciplinary nature of the research in mobile cloud computing.

Acknowledgement

The author would thank Professor Mario Gerla for providing very insightful comments about mobicloud features and capabilities. The author also thanks Sean A. Williams for his comments to improve the presentation of this article.

References

[1] Mark Walsh. Gartner: Mobile To Outpace Desktop Web By 2013. Online Media Daily, January 13, 2010.

[2] Secure Networking And Communication (SNAC) Research Group, MobiCloud Project, available at http://mobicloud.asu.edu, Arizona State University, 2010.

[3] Peter Mell and Timothy Grance. The NIST Definition of Cloud Computing (Draft). January, 2011.

[4] Ramesh Jain. Quality of experience, IEEE *Multimedia*, vol. 11, no. 1, pp. 96–95, 2004

[5] Ben Pfaff, Justin Pettit Teemu Koponen, Keith Amidon, Martin Casado, and Scott Shenker. Extending Networking into the Virtualization Layer, HotNets-VIII, Oct. 22-23, 2009.

[6] OpenFlow Switch Specification. Version 1.1.0 Implemented, February 28, 2011.

[7] Peter Saint-Andre. Extensible Messaging and Presence Protocol (XMPP): Core. IETF RFC 3920, Oct. 2004.

[8] OSGi Alliance. OSGi Service Platform Core Specification, Release 4, Version 4.3, April 2011

Dijiang Huang received his B.S. degree from Beijing University of Posts and Telecommunications, China 1995. He received his M.S., and Ph.D. degrees from the University of Missouri—Kansas City, in 2001 and 2004, respectively. He joined Arizona Sate University as Assistant Professor in 2005. He is currently an associate professor in the School of Computing Informatics and Decision System Engineering at the Arizona State University.

His current research interests are computer networking, security, and privacy. He is an associate editor of the Journal of Network and System Management (JNSM) and an editor of IEEE Communications Surveys and Tutorials. He served as an organizer for many International conferences and workshops. Dr. Huang's research is supported by NSF, ONR, ARO, Consortium of Embedded System (CES), and HP. He is a recipient of ONR Young Investigator Award 2010.

Mobile Cloud Computing: Opportunities and Challenges

Karthik Kumar and Yung-Hsiang Lu,Purdue University, United States
*{*kumar25,yunglu*}@purdue.edu*

Cloud computing is changing how companies plan their IT infrastructures, consumers use computers, and researchers design their projects. Cloud computing is one of the Top 11 Technologies of the Decade" [1]. Using cloud computing, companies can grow their IT budgets as computing demands grow. They do not have to set up their IT facilities right at the beginning; this can save significant amounts of money in planning, purchasing, installation, and maintenance. Consumers use cloud services such as on-line office applications or database programs. They do not have to worry about upgrading software or backing up data. Consumers also use social networks to share their photos and videos. For researchers that have large amounts of computation, they can access thousands of processors at low costs and pay only when needed.

In the past few years, two important trends have occurred, together with the growth of cloud computing. (1) Sensors are widely deployed for monitoring environment and for security. These sensors acquire large amounts of data but the sensors have limited computing capabilities. (2) Mobile systems, such as smartphones, have become the primary computing platforms for millions of people. These platforms can generate large amounts of multimedia data and most of the data are stored on-line. In the next few years, sharing pictures and videos with friends would be only a small part of *mobile cloud computing*. Sensors and mobile platforms represent an entirely new set of input devices. Consider the possibility when millions of cameras, microphones, GPS, and many other types of sensors are connected. The amounts of data they can produce would be staggering. The information and knowledge that could be extracted would dwarf what we have called "information explosion."

In the next few years, we will see pressing needs for personalized management of multimedia data. This would be a natural progression of the Internet. Before the Internet became popular, people already had large amounts of on-line documents stored in their desktop computers or company mainframes. In 1990s, as the Internet became popular, many documents were posted on-line and keyword-based (text) search became necessary. Search engines were an important driving force for the Internet in the late 1990s. The first ten years of the 21st century marked the rapid growth of personal multimedia data: images, videos, and audios. According to the forecast of Cisco Visual Networking Index (VNI) in 2009, the mobile data traffic is expected to grow to 3.6 exabytes per month by 2014, and almost 66 percent of those will be video [2].

What can the cloud offer mobile devices?

As mobile devices become increasingly prevalent, we examine three areas where cloud computing may benefit mobile computing: (1) recognition, (2) energy efficiency, and (3) disaster rescue and recovery.

Recognition

As the amounts of multimedia data grow, users need better ways to manage their data than relying on file names, dates, and directories. It would be inconvenient to ask users to describe every image and every video by a set of keywords and then use keyword-based search. Content-based search (e.g., using a picture to find similar pictures) is a promising solution even though the technology is not mature yet. Fortunately, the Internet has taught us that people would accept a solution when it is necessary even though the solution is still imperfect. Text-based search is an example. The information stored on the Internet is changing every second and there is no perfect solution for web search; in fact, it is not even possible to define what a “correct" solution means. Yet, millions of people find keyword search engines helpful and use them everyday. Another example is natural language translation. The technology is not perfect but already useful to many users. The providers of natural language translation solicit feedback from users to improve the algorithms. The technology for content-based search may follow a similar path: in the next few years, companies will offer technologies to help users manage personal multimedia data. Some websites already offer limited content-based searches. We think this technology will become available on some popular portal sites within five years. The technologies will improve quickly as more companies enter the competition and massive user feedbacks are obtained. We may see major breakthroughs in search algorithms that benefit from user feedback. Content-based search requires some degrees of recognition. Recognition technology will improve as the demands grow rapidly. The billions of cameras on mobile phones will produce so many photographs and videos that recognition technologies become necessary. We also need to think beyond mobile phones and include surveillance cameras at street corners, as well as many sensors deployed in smart buildings, bridges, vehicles, and so on. These many different types of sensors will produce so much data that must be processed by thousands of cloud servers.

Energy Efficiency

Battery lifetime is an important concern for mobile users. A 2005 study of users in 15 countries [6] found longer battery life to be more important than all other features. A survey in 2009 by ChangeWave Research [7] revealed short battery life to be the most disliked characteristic of Apple's iPhone, while a Nokia poll showed that battery life was the top concern of music phone users.

Cloud computing may help prolong the battery lifetimes of mobile phones through *computation offloading:* migrating heavy computation to grid powered servers. This migration may be energy efficient if the energy saved by avoiding computation on the mobile device exceeds the energy spent in sending the program and data to the servers. Since many cloud services already host users' data (for example Facebook and Picasa host users' pictures and videos), the amount of data that needs to be transmitted may be small enough for offloading to be energy efficient. Prior analysis [8] has shown that applications with a large computation-to-communication ratio benefit more from offloading.

Computation offloading does have its challenges: privacy and security, reliability, and handling real-time data. There is risk of privacy violations when data is sent to a cloud vendor. There have been some instances of privacy and security breaches by cloud vendors [8]. Some potential solutions to avoid this include homomorphic encryption [9]. Cloud servers may have downtimes. In April 2011, Amazon's crash indicates that cloud's reliability still needs improvement. In addition, many applications using mobile and sensor data--- such as computer vision--- involve handling massive amounts of real-time data. In such instances, additional factors like the network latency and bandwidth need to be considered.

Disaster Rescue and Recovery

Another possible usage of mobile cloud computing is disaster rescue and recovery. After a disaster, up-to-date information is crucial in planning and conducting rescue and recovery [5]. However, existing facilities might be severely damaged. Many of today's datacenters are built using trailers of servers. Those trailers, along with mobile phone towers, might be redeployed quickly in disaster areas so that communication and computation could be re-established. This could be possible because tasks could migrate among different machines, not restricted to the damaged machines. The millions of mobile phones and sensors can serve as data sources for cloud servers. Mobile phones can take pictures, and microphones can sample the surrounding sounds. Phones may broadcast emergency information to users. The sensors can also be mounted on rescue robots [4]. The vast amounts of information gathered by phones and sensors can be analyzed at cloud servers to assess the damage of the disaster, as suggested in [5]. The most up-to-date information can be used by cloud servers to develop a plan for rescue and recovery.

Mobile devices and cloud servers are complementary: Mobile devices need resources, and cloud servers need data. To reap the true potential of mobile cloud computing, several key technologies must be developed. First, better technologies are required for multimedia processing and recognition. Second, cloud computing must have the ability to handle and process information from several thousands of mobile devices, and to be able to handle timing sensitive tasks. This requires new designs and implementations of cloud servers. Third, cloud and mobile computing, as well as sensor networks, must be integrated so that information can flow smoothly as needed. Writing applications that can integrate those very different types of systems would be challenging. Finally, privacy and data protection could be the most significant barrier for users to accept these technologies. Even though it could be helpful for a phone to automatically take pictures after a disaster, no user would accept the possibility of such privacy invasion in daily lives. The outcomes of research and development in these four directions could shape how we use cloud computing, mobile platforms, and sensor networks, in the second decade of this century.

References

[1] P.E. Ross, J.J. Romero, W.D. Jones, A. Bleicher, J. Calamia, J. Middleton, R. Stevenson, S.K. Moore,

S. Upson, D. Schneider, W.D. Jones, E. Guizzo, T.S. Perry, and G. Zorpette. "Top 11 Technologies of the Decade". *IEEE Spectrum, 48(1):27-63*, January 2011.

[2] CiscoSystems. "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2009-2014", February 9 2010.

[3] C. Shahabi. "Crowdsourcing what is where: Community-contributed photos as volunteered geographic information". *Multimedia, IEEE, 17(4):36-45*, 2010.

[4] Robin R. Murphy. "Rescue Robotics for Homeland Security". *Communications of the ACM, 47:66-68*, 2004.

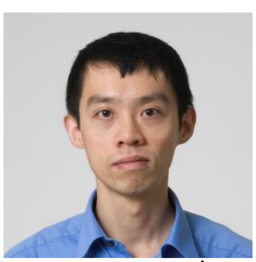
[5] Mahadev Satyanarayanan. "Mobile Computing: the Next Decade". *ACM Workshop on Mobile Cloud Computing and Services*, pages 1-6, 2010.

[6] CNN.com, "Battery Life Concerns Mobile Users," 23 Sept. 2005.

[7] J. Paczkowski, "Iphone Owners Would Like to Replace Battery," *All Things Digital*, 21 Aug. 2009.

[8] K Kumar and Y-H Lu, "Cloud Computing for Mobile Users: Can Offloading Save Energy?", *IEEE Computer*, April 2010.

[9] C Gentry, "Computing arbitrary functions of encrypted data", *Communications of the ACM*, 53:3:97-10 5, 2010



Yung-Hsiang Lu is an associate professor at the School of Electrical and Computer Engineering, Purdue University, USA. His research focuses on resource management in computer systems, especially mobile systems. He is a senior member of the IEEE and the ACM. In 2004, he received an NSF Career Award for investigating energy management using operating systems. He received the PhD degree from Stanford University in 2002. Contact him at yunglu@purdue.edu

Karthik Kumar obtained the PhD degree from Purdue University in August 2011, and the MS degree in August 2008. His research interests include energy efficiency in mobile and server systems. He received the Bilsland Dissertation Fellowship, and the Magoon Award for excellence in teaching from Purdue University. Contact him at kumar25@purdue.edu.

Enabling Mobile Multimedia with the Cloud

Hari Kalva, Multimedia Lab, Florida Atlantic University, Boca Raton, USA
Lai-Tee Cheok, Samsung Telecomm America, Dallas, USA
Hari.Kalva@fau.edu, l.cheok@samsung.com

1. Introduction

Mobile Cloud Computing refers to mobile services that use the computing resources and data on the cloud to provide a requested service. A request from a mobile client leads to a process in the cloud that operates on the data in the cloud and the result of this processing is returned to the mobile client. The main reasons for using the cloud to deliver mobile services are lack of sufficient computing and data resources on the mobile client This cloud computing model works especially well for mobile devices and enables services that are not otherwise possible because of limited computing and communication resources. Cloud computing can also improve the energy performance of mobile applications when computationally intensive tasks are off loaded to the cloud [1].

2. Mobile Cloud Computing

Multimedia applications require significantly more computing, storage, and bandwidth resources and the resource constraints on mobile devices are the main reason for the lack of resource intensive multimedia services on mobile devices

2.1 Computing Constraints

Computing resources increase the cost of the devices and lower-end devices are characterized by limited computing resources. Mobiles with computing constraints can use the cloud for computing needs and available bandwidth affects the types of services offered. Applications such as remote rendering and remote gaming fall into this category.

2.2 Bandwidth Constraints

Bandwidth availability varies with network service availability and costs increase with bandwidth usage. Cloud based services can be used to reduce the bandwidth usage (e.g,. video transcoding to lower bitrates) or a mobile device's computing resources can be used to pre-process the data and reduce the outgoing bandwidth needs (e.g, compact feature set instead of an image for a query).

2.3 Latencies

A key factor that affects the services offered is latency. Cloud based services can have higher latencies because of the distributed architecture and the higher latencies impact the services offered. For example, conversational services such as video conferencing are highly sensitive to latencies and hence are not suitable for cloud platforms. Services such as video on demand and broadcasting can hide latencies and can benefit from cloud computing [2].

2.4 Cloudlets

Problems such as high latencies can be addressed using a local cloud that uses the computing available in physical proximity [3]. A typical home environment has high bandwidth wireless connectivity and significant computing resources available in home devices such as PCs, game consoles, TVs, and entertainment units. These resources in the immediate proximity can be exploited to provide rich services to mobile users.

3. Mobile Cloud Computing Applications

Cloud based services are usually offered to enhance the mobile user experience and enable services that are otherwise not possible on mobile devices. In this section we present mobile cloud computing problems we have been working on and discuss the key problems and solutions.

3.1 Remote Gaming

Games are one of the most common applications on mobile devices. Complex games require powerful CPUs and dedicated graphics processing unit (GPU) to accelerate 3D graphics rendering. Resource constrained nature of mobile devices limits mobile devices from playing sophisticated games on such platforms. Gaming can be enabled on resource constrained mobile devices by offloading the computing tasks to the cloud.

We have developed a remote gaming platform where games are rendered on a server and the graphics updated by the game play are encoded as a video stream and streamed to a mobile device. The mobile device only needs computing resources to play the video and capture user interactions and sufficient bandwidth to enable game play without interruptions. Figure 1 shows the architecture of the proposed remote gaming platform. The user interactions are captured at the receiver/mobile device and transmitted to the server where the game play is advanced and graphics updated. High bandwidth and low latency is critical to uninterrupted game play. We studied the impact of latency and developed solutions to hide part of the latencies by controlling the gaming environment [4]. Lower bandwidth is not as a big a problem as high latencies. Bandwidth usage can be decreased by reducing the video bitrate and using content aware encoding solutions.

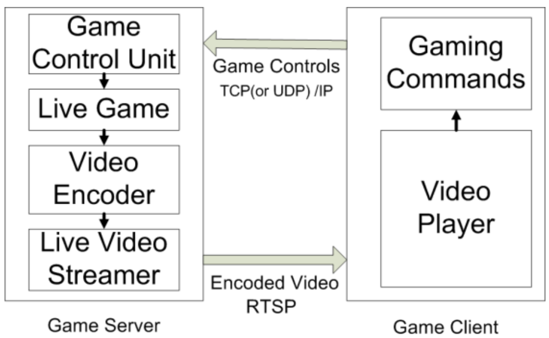


Figure 1. Remote gaming platform

3.2 Remote Rendering

Remote rendering, as in the case of remote gaming, uses cloud based rendering along with video streaming to enable compute intensive visualization applications. This instance of the cloud computing problem is compute constrained as scientific visualization has large data and compute requirements.

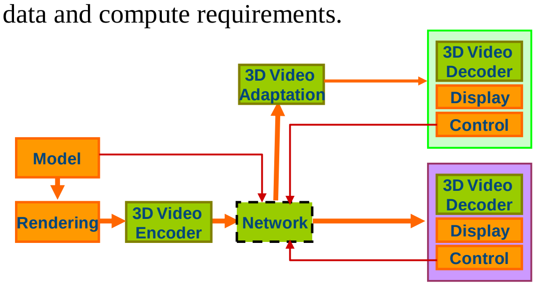


Figure 2. Remote rendering for data visualization

To enable this application on mobile devices, the rendering task is offloaded to a cloud and rendering is done on nodes that are close to the data storage. The rendered views are encoded as a video stream and streamed to the mobile device. A video stream is used to show continuous response to user interaction. Since the application is interactive data visualization, latency is not a significant problem as response times in visualization are much slower than the case of gaming. In our proposed solution, left and right views of rendered 3D volumes are encoded as a 3D video stream and streamed to the receiver [5]. Figure 2 shows the key components of the proposed solution. The bitrates of the coded video are kept at 500 Kbps and give very good quality due to low frame rate necessary to view the rendered models.

3.3 Distributed Transcoding in the Cloud

When available bandwidth is a constraint that needs to be addressed, cloud computing can play a key role in repurposing content to fit the available bandwidth for the mobile device. Distributed processing frameworks such as Hadoop can be used to quickly develop services on the cloud. We evaluated the performance of Hadoop based distributed systems for video transcoding. The goal is to transcode video to meet the bandwidth availability of a mobile device. This solution works well for on-demand services as a video can be segmented into chunks that are transcoded independently. A set of chunks can be assigned for processing on each node of the Hadoop cluster for a faster turnaround. Recent developments such as adaptive streaming over HTTP are based on the principle of segment videos into independent chunks that are adapted for mobile needs. We show that a Hadoop solution can increase the performance of video delivery services [2].

3.4 Cloud-based Mobile Visual Search

Visual search and augmented reality applications are becoming prevalent on mobile platforms [6]. These augmented reality services search and overlay information over images and video captured by the mobile phone camera. A typical scenario is a tourist using his/her mobile phone to point, click and identify names and history of buildings around him/her.

Applications of mobile visual search have received significant industry interest and this has resulted in a recent MPEG standards activity called the Compact Descriptor Visual Search (CDVS) [7]. CDVS aims to ensure interoperability of visual search applications and databases, among others. The Call-for-Proposals (CfP) has been issued during the 97^{th} MPEG meeting in Turin, Italy in July 2011 with planned evaluation of proposals to start in November, 2011.

Visual search on mobile platform imposes certain requirements on latency, processing, bandwidth and robust performance. Mobile devices have limited memory, processing power and battery life so the retrieval framework has to adapt to these stringent mobile system requirements. The computation on the mobile device has to be fast to achieve low power consumption. The descriptor size has to be small and compact to reduce transmission latency. The algorithms should be robust to allow objects to be recognized accurately under various conditions such as different viewing angles, lighting conditions, or when there are partial occlusions or blurred motion.

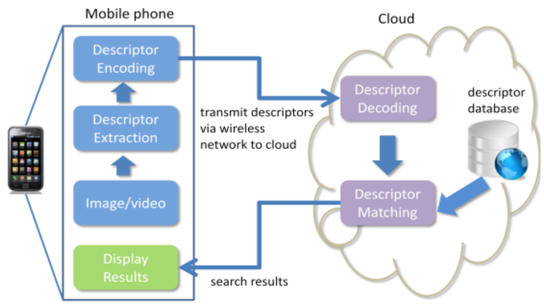


Figure 3. Cloud-based Mobile Visual Search

By deploying mobile visual search at the cloud, a few of these challenges can be alleviated. Figure 3 shows one of the possible client-server architecture supporting mobile visual search in the cloud. The process involves first extracting descriptors of images/video captured by the mobile phone camera. The descriptors are features used to uniquely identify the visual content. These query descriptors are then compressed/encoded and transmitted to the cloud where they are compared against descriptors of images/video stored in the cloud database during the descriptor matching process. The image/video of the closest matching descriptor in the database is then selected.

As shown in the diagram, the computation involved in matching query descriptors of images/video with the candidate descriptors in the reference database is offloaded to the cloud by leveraging the relatively higher power and processing resource at the cloud. However, feature extraction, especially of query image/video still has to be performed at the mobile device and the challenges of achieving a good tradeoff between compact descriptor size (to reduce bitrate and latency) and accuracy of object recognition and robust performance remains. Alternatively, when bandwidth is available, the captured image or video clip can be sent to the cloud for feature extraction and search.

4. Conclusion

We presented an overview of mobile cloud computing. Cloud computing is expected to become integral part of mobile services; especially for resource intensive multimedia services. Offloading computing to the cloud allows service providers to offer services that do not heavily rely on the receiver resources. Cloud computing can be used to address the computing and/or bandwidth constraints on mobile devices. Examples from our recent and ongoing research activities are presented. The remote rendering ideas presented enable multimedia services on resource constrained devices and have many applications. This solution works well when there is bandwidth availability and computing is the main constraint. Migrating mobile visual search to a cloud-based environment may alleviate resource issues for intensive search and matching operations, however many challenges still remain. It is anticipated that the emerging standardization efforts of CDVS will encourage development and standardization of the best possible solution and further increase both the research community's and the industry's interest in this area.

References

[1] K. Kumar and Yung-Hsiang Lu, "Cloud Computing for Mobile Users: Can Offloading Computation Save Energy?," Computer , vol.43, no.4, pp.51-56, April 2010.

[2] Adriana Garcia, Hari Kalva, and Borko Furht. "A study of transcoding on cloud environments for video content delivery," Proceedings of the 2010 ACM multimedia workshop on Mobile cloud media computing (MCMC '10). ACM, New York, pp. 13-18.

[3] Mahadev Satyanarayanan, Paramvir Bahl, Ramon Caceres, Nigel Davies, "The Case for VM-Based Cloudlets in Mobile Computing," IEEE Pervasive Computing, pp. 14-23, October-December, 2009.

[4] Waazim Reza, Hari Kalva and Richard Kaufman, "Remote gaming on resource-constrained devices", Proc. SPIE 7798, 77981Q (2010); doi:10.1117/12.863016

[5] Hari Kalva and Borko Furht, "3D Technologies and Visualization Tools," The 2nd Workshop on Frontiers of Multicore Computing, Sept. 21-23, 2010, University of Maryland, Baltimore County (UMBC), USA.

[6] Bernd Girod, Vijay Chandrasekhar, Radek Grzeszczuk, and Yuriy A. Reznik, "Mobile Visual Search: Architectures, Technologies, and the Emerging MPEG Standard," IEEE Multimedia, pp. 86-94, July-Sept. 2011.

[7] ISO/MPEG, "Compact Descriptors for Visual Search: Call for Proposals," MPEG output document N12201, July 2011.

Dr. Hari Kalva is an Associate Professor and the Director of the Multimedia Lab in the Dept. of Computer & Electrical Engineering and Computer Science at Florida Atlantic University. Dr. Kalva is an expert in the area of video compression and communication with over 17 years of experience in Multimedia research, development, and standardization.

Dr. Kalva's research interests include mobile multimedia services, exploiting human perception for video compression and bandwidth reduction, and content distribution. His publication record includes 2 books, 7 book chapters, 30 journal papers, 78 conference papers, 8 patents issued and 12 patents pending. One of his patents has been determined essential to the implementation of ATSC and Blu-ray standards. He is a recipient of the 2008 FAU Researcher of the Year Award and the 2009 ASEE Southeast New Faculty Research Award.

Dr. Kalva received a Ph.D. and an M.Phil. in Electrical Engineering from Columbia University in 2000 and 1999 respectively. He received an M.S. in Computer Engineering from Florida Atlantic University in 1994, and a B. Tech. in Electronics and Communications Engineering from N.B.K.R. Institute of Science and Technology, S.V. University, Tirupati, India in 1991.

Dr. Lai-Tee Cheok is a Staff Engineer/Manager at Samsung's Dallas Technology R&D Lab in Texas, US. She received her PhD degree in Electrical Engineering from Columbia University in the City of New York, US in 2006. She was with AT&T Labs Research, IBM T.J. Watson Research Center and Symbol Technologies Inc. (bought over by Motorola Inc.) during a few summer internships and was awarded the Jacob Millman Prize while at Columbia University. Prior to joining Samsung, she has worked in several companies including startups. Her areas of expertise and research interests are image/video processing and coding, semantic multimedia content analysis, computer vision, combinatorial optimization and multimedia standards. Dr Cheok has served on the conference committee of International conferences and has been an active participant in MPEG standards, having made several contributions to the standard.

Cloud Centric Mobile Application Development Using Domain Specific Languages

Ajith Ranabahu, Ashwin Manjunatha, Amit Sheth and Krishnaprasad Thirunarayan,
Kno.e.sis Center, Wright State University, Dayton OH USA
{ ajith,ashwin,amit,tkprasad}@knoesis.org

1. Introduction

Stellar growth in use of mobile platforms for business users, combined with rapid business growth in developing countries that have much higher utilization of mobile access have put cloud back-end hosted mobile applications in a sweet spot. By using a cloud mobile combination, computationally intensive services can be delivered right to the consumer anywhere, anytime. Two unmet challenges in developing such applications are managing the development of applications for heterogeneous mobile platforms with equivalent functionality, and maintaining a portable back-end to mitigate any catastrophic outage (such as the recent outage of Amazon EC2[1].)

This article discusses the mobile application aspects of the MobiCloud [1] approach for rapidly developing cloud-mobile hybrid applications for heterogeneous mobile front-ends and cloud back-ends. MobiCloud exploits the features of Domain Specific Languages (DSLs) to address the difficulty of programming for multiple mobile platforms. It also helps to overcome some of the limitations of mobile platforms by pairing with cloud back-ends.

2. Designing Hybrid Applications

Typical applications being developed as mobile device based front-ends for Cloud based applications have the following features.

1. The front-end and back-end applications are usually managed as separate projects. These projects depend on well-defined service interfaces to implement either the client or the service functions.
2. Front-end applications are tied strongly to the back-end application. Updates to the back-end applications would eventually have to be propagated to the front-end applications, vice versa. Such change propagation requires sweeping changes to the front-end application code, often to multiple front-end applications targeted towards different mobile devices.
3. A significant effort is needed to debug these applications, due to their use of remote procedure calls (RPC).

Regardless of the separation of the back-end and the front-end through a service layer, the application components still maintain their relevance to the well-known Model View Controller (MVC) design pattern [2, 3]. For example, the model data structure can be easily identified (equivalent in function yet different in implementation) in both the back-end and front-end.

Thus, in MobiCloud, the entire functionality, i.e. the functionality of the front-end mobile application as well as the back-end cloud application, is modeled as a single unit. The DSL based representation of this modeling can be transformed to the required software components.

This approach has the following benefits;
1. The same model can be used to generate functionally equivalent, mobile front-end applications as well as cloud back-end applications targeting different platforms.
2. Many developers are familiar with the MVC design pattern, thus this modeling has a gentler learning curve.
3. Modeling complexity is reduced by treating the entire functionality as a single unit.
4. Differences in service interfaces are non-existent since the service and the client are generated using the same specification. This relieves a significant burden in debugging.

In the next section, we present a brief overview of the MobiCloud language using an example.

3. The MobiCloud DSL

We present a simple, yet useful example application, made using MobiCloud. Listing 1 illustrates the DSL script to create a task manager application. This script can be generated using the online MobiCloud composer or just written using a basic text editor.

[1] http://blog.rightscale.com/2011/04/25/amazon-ec2-outage-summary-and-lessons-learned/

```
recipe(:todolist) do
metadata({:id => 'task-manager'})
    # models
   model(:task, {:name=>:string,
   :description => :string,
   :time => :date,
   :location => :string})
  #controllers
  controller(:taskhandler) do
    action :create, :task
    action :retrieve, :task
  end
  # views
  view :add_task, {
   :models =>[:task],
   :controller=> :taskhandler,
   :action => :create}
  view :show_tasks, {
   :models =>[:task],
   :controller=>:taskhandler,
   :action => retrieve}
end
```

Listing 1 : MobiCloud script to generate a task manager application.

Figure 1 illustrates the graphical representation of the same application, created using the MobiCloud composer. The composer also supports direct deployments to supported platforms.

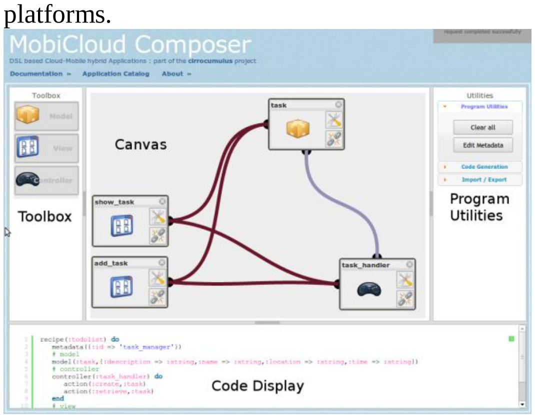


Figure 2 : The MobiCloud composer UI.

For example, an installable Android package (APK) can be downloaded as a result of the compilation when the Android platform is selected.

Figure 2 illustrates the relevance of the generated application to the MVC components. The *model* specification is used to generate two data structures, one for the mobile device and another for the service implementation. The *controller* specification mainly contributes to the server side implementation while the *view* specification contributes to the user interface, primarily on the mobile device.

We conclude the details of the language here for brevity. Manjunatha et al [1] has discussed in detail, the philosophy behind these constructs and their relationships.

4. Advantages to the Mobile Developer

There are many advantages to the mobile application developer, when using MobiCloud

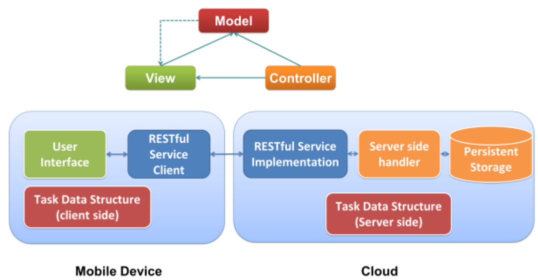


Figure 3 : The relevance of the generated artifacts to the MVC design.

I. *Ability to cover multiple platforms with one code base*: A number of mobile development platforms exist today, each with different development environments, Application Programming Interfaces (API), and programming languages. Fragmentation of APIs, even within a single platform forces mobile application developers to focus on only specific platforms and versions.

MobiCloud helps to alleviate these issues by providing the DSL and the generators that convert high level abstractions to platform specific code representations. Only a simpler high level DSL script needs to be saved and the platform specific code can be generated on demand.

II. *Reduced programming effort*: Developing Mobile applications take significant effort, yet most of it is focused on repetitive code segments. Furthermore some of these code segments depend on external service specifications that are hard to debug in case of errors.

MobiCloud can be used as a boilerplate code generator to cover most of the required but repetitive code segments. Developers can then pursue more creative aspects, such as improving the user interface. MobiCloud also takes care of generating the service clients, ensuring that there are no glitches in the communication aspects. This significantly lowers the programming effort.

III. *Platform specific UI and code optimizations*: There are many platform specific optimizations that can be done on each of the mobile platforms. Interpreter based universal solutions such as J2ME are not very popular primarily due to their inability to exploit these platform specific features. MobiCloud uses a generative approach and thus, has the ability to generate the optimal code for a given platform.
For example, in the case of Android, the UI components are created using Android widgets while for Blackberry; the corresponding Blackberry widget classes are extended. This enables the applications to present native widgets in their interfaces rather than adhering to universal look and feel.

5. Future Enhancements
There is a host of planned enhancement to MobiCloud and we highlight the key enhancements targeted towards the mobile front-ends.

1. *Sensor Integration*: Smartphones are equipped with myriad of sensors. These sensors play a vital role in developing intelligent applications. We plan to provide abstractions over commonly used sensors to improve the applicability of MobiCloud.
The following DSL snippet highlights how we plan to integrate the sensor data via special data types.

```
model(:bird,
{:name=>:string,
 :image => :camera_image,
 :time_seen=>:current_clock_time,
 :location_seen
=> :current_gps_coords})
```

Listing 2 : A model construct for a bird watchers application that integrates sensor data as data types.

In listing 2, data types like **:camera_image** and **:current_gps_coords** can generate platform specific code to obtain an image from the camera or populate the location from the built in GPS. The generators can also insert code that support graceful degradation when the sensors are absent or not functional. For example, when the GPS sensor is not active, the UI can simply present the user with two text boxes for the coordinates, indicating that the GPS sensor is inactive.

2. *Multimedia and third party integrations*:
Many mobile applications are integrated with third party services. Maps, videos and social network features are a few of the popular ‘must have’ integrations.
We have already introduced an extension mechanism to MobiCloud that enables third party extensions to the language constructs. We plan to add extensions to enable integrations with the popular services. The extension mechanism also gives the flexibility to advanced programmers to write custom integrations.

3. *Power and connectivity aware processing*:
Mobile devices have limited storage, processing power and connectivity. Applications can be made to make smarter use of the battery power and connectivity by dynamically changing the behavior of the application. This however, takes significant effort in programming.
MobiCloud can automatically generate the required code to efficiently manage the communication. For example, it can store content locally when the connectivity is via 3G and the data size is large, and complete the transfer when Wi-Fi is available, thereby saving power and bandwidth. We plan to add improvement to the code generators to automatically add data transfer optimization code.

6. Conclusion
Combining mobile devices and clouds is becoming important, but the difficulty of programming remains a serious problem. Our MobiCloud project has shown that by providing proper abstractions via a DSL, one can overcome the complexities in programming cloud-mobile hybrid applications. Given that mobile platforms are becoming more complex every day, we believe that proper abstractions will be the most important programming tool for mobile devices in the future.

References
[1] Manjunatha, A.; Ranabahu, A.; Sheth, A.; Thirunarayan, K.; , "Power of Clouds in Your Pocket: An Efficient Approach for Cloud Mobile Hybrid Application Development," Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on , vol., no., pp.496-503, Nov. 30 2010-Dec. 3 2010.

[2] T. Reenskaug. The Original MVC Reports. Xerox Palo Alto Research Laboratory, PARC, 1978.
[3] S. Burbeck. Applications Programming in Smalltalk-80: How to Use Model-View-Controller (MVC). Softsmarts,Inc., 1987.

Ajith Ranabah is a PhD candidate in computer science at the Kno.e.sis center in Wright State University. His primary research is focused on application and data portability in Cloud computing. Contact him at ajith@knoesis.org.

Amit Sheth is an IEEE Fellow, the LexisNexis Ohio Eminent Scholar and the director of the Ohio Center of Excellence in Knowledge-enabled Computing (Kno.e.sis) at Wright State University. Contact him at amit@knoesis.org

Ashwin Manjunatha is a recent graduate from the Ohio Center of Excellence in Knowledge-enabled Computing (Kno.e.sis) at Wright State University. His research interests include Cloud Mobile Hybrid Applications and Large Scale Data Analysis. He is presently working on a technology to bridge cloud computing and mobile computing. Contact him at ashwin@knoesis.org

Krishnaprasad Thirunarayan is a Professor in the Ohio Center of Excellence in Knowledge-enabled Computing (Kno.e.sis) at Wright State University. Contact him at tkprasad@knoesis.org

Energy-Efficient Mobile BitTorrent Solutions

Imre Kelényi, Budapest University of Technology and Economics, Hungary
Jukka K. Nurminen, Aalto University, Finland
imre.kelenyi@aut.bme.hu, jukka.k.nurminen@aalto.fi

1. BitTorrent on Mobile Devices

Most peer-to-peer applications, including BitTorrent, are originally designed for PC devices. When they are used on handheld devices, such as mobile phones, new problems arise because of limited resources and more constrained communication. A particular worry, which is also the target of our work, is how the applications influence the energy consumption of the device. Active BitTorrent usage can drain the battery of a standard smartphone in less than 5 hours [1]. Running BitTorrent on mobile phones is clearly interesting for users as evidenced by more than to 200.000 downloads of SymTorrent [2] and MobTorrent [3], our open-source BitTorrent clients for Symbian and Java ME phones. It shows that the resources of modern mobile phones are adequate to let them participate in P2P communities.

An obvious attempt to save energy is to make the mobile peers act in a selfish way and only download content. However, it turns out that this is not a good strategy because the tit-for-tat mechanism in BitTorrent penalizes the peers that do not upload content to others. The result is longer download time and bigger energy consumption. Besides, uploading content to others when the phone is actively downloading requires only little extra power. According to our measurements uploading during a download operation only consumes about 20% more energy [1].

Figure 1, based on the measurements data of [2], illustrates how the energy consumption per bit varies as a function of communication speed. The shape of the curves shows that the higher the bit rate, the more energy-efficient the communication is. This suggests that in order to save energy we should arrange the content download activity in a way that the mobile device experiences as high bitrate as possible. One way to achieve this is to alternate between active transfer periods with high bitrates and idle periods with no traffic. So instead of receiving data at a constant low speed we prefer to communicate in high-speed bursts separated by idle periods. The higher speed we are able to reach during the active periods the more energy-efficient the data transfer will be.

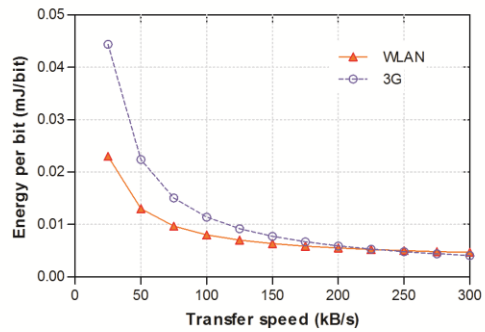


Figure 1: Power consumption per bit as a function of wireless communication speed

We present here three different solutions for the same problem: how we can download content to a mobile device via BitTorrent with consuming as less energy as possible.

4. BurstTorrent

BurstTorrent [3] is an extended version of the BitTorrent protocol that allows using scheduled bursty transfers to achieve energy efficient operation for mobile devices. The goal of the protocol is to achieve the lowest energy per bit ratio by minimizing the time that energy-limited peers are in active state, transferring data. Between these high-speed active phases, the peers are in low-power idle state.

In the context of BurstTorrent, peers are categorized as either energy-limited or regular. An energy-limited peer, which is downloading content from a BitTorrent swarm, negotiates time intervals with regular peers when the regular peers would promise to use all the necessary resources to send content to the downloading peer with the agreed speed. This way it can be ensured that when an energy-limited device is in active state, it receives data at full speed.

Regular peers maintain an upload schedule of time points when data is to be sent to limited peers. Similarly, energy limited peers follow a download schedule when requesting new pieces of data. The

[2] http://symtorrent.aut.bme.hu
[3] http://amorg.aut.bme.hu/projects/mobtorrent

scheduled transfers require regular peers to manage their upload speed and reserve a portion of their uplink bandwidth to serve limited peers at scheduled intervals. If a regular peer cannot transfer at the agreed speed the schedule can become corrupted. This results into extra energy consumption temporarily but otherwise does not harm the content download process.

We analyzed how BurstTorrent compares with standard BitTorrent via simulations. In a network where 50% of the peers are energy-limited their energy consumption was cut to half in comparison with standard BitTorrent. The downside of the energy saving was that their download time grew by 50%. The scheduling algorithm was designed in such a way that the performance regular peers did not degrade.

The BurstTorrent approach has a lot of potential in heterogeneous content sharing networks which consist of both energy-limited and regular peers. However, there are at least two difficulties to consider. First, although BurstTorrent is able to operate with current BitTorrent peers, they are not fully compatible. To support the energy-efficient operation both regular and energy-limited peers need to implement the new protocol. Thus, we cannot rely on existing BitTorrent users. Furthermore, we cannot necessarily assume that desktop clients will start using BurstTorrent since even if it does not affect their download times, they do not benefit from using the protocol either.

2. CloudTorrent

With CloudTorrent [4] the BitTorrent client is run on the cloud and mobile clients access it to download torrent content. CloudTorrent consists of two main parts: a phone application communicating with the cloud and a server hosting the remote BitTorrent client. All communication between the server and the client is carried out via HTTP connections. On the server side, we use Amazon EC2 to run uTorrent, which is a popular BitTorrent client having most of its functions available via an HTTP-based API. Downloading content using CloudTorrent is performed in two steps. First the server side uses the BitTorrent protocol to download the content to the CloudTorrent server. Once the torrent download is completed, the content is transferred to the phone via an HTTP connection.

The measured power consumption curves in Figure 2 show how CloudTorrent compares with SymTorrent, the native BitTorrent client for Symbian smartphones, when the same torrent is downloaded with both applications in 3G cellular network. In this case CloudTorrent achieved 65% energy saving and 60% time saving in comparison to SymTorrent. The power consumption of SymTorrent is almost constant with a 1.5W average during the whole 7 minute download. In the case of CloudTorrent, the transfer session starts with a low power phase with some high-power spikes (0 min – 0:45 min). During this period the server is downloading the torrent and a phone is idle waiting for the content to be ready. This is followed by the high power phase when content is sent from cloud server to client (0:45 min – 2:30 min).

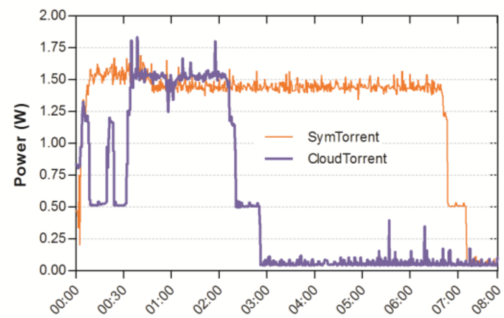


Figure 2: SymTorrent vs. CloudTorrent mobile client energy consumption

The savings arise from a several reasons. First, other peers favor the P2P client running on Amazon because it has a much higher uplink speed than the mobile phone. Via the tit-for-tat mechanism this results into a much shorter download time. When transferring the complete file from Amazon to mobile phone the solution is able to push the content to the phone in a very high speed again taking advantage of the high uplink bandwidth. Therefore the time when the phone is actually receiving the content, and spending a lot of power to keep the radio active, is short.

3. ProxyTorrent

ProxyTorrent [5] builds on the same idea as CloudTorrent: use an intermediate helper peer to cache the content and then push it to the phone in an energy-efficient way. However, there is a key difference: with ProxyTorrent, the proxy is hosted on a memory-limited home router. The router platform is attractive for a number of reasons. First, since most homes are equipped with them, the installed base of routers is big. Second, many of the router platforms allow modifying the operating system. Third, routers are typically powered up all the time and the energy consumption of the router is almost constant no matter how actively it is used. Running proxies on broadband routers would thus introduce no new costs.

When we use routers with small amount of resources, the limited memory allows us to store only part of the content. This raises interesting questions on what kind of memory management policy to adopt that ensures both a fast and fair torrent download and an energy-efficient transfer of content pieces from the proxy to the phone.

A key assumption of BitTorrent operation is that when a peer has completely downloaded a piece it announces the availability of the piece to its peers. The peers can then assume that the announced piece is available for downloading. This is, however, not the case if only part of the content fits the router memory. To be able to download the whole torrent, the router has to delete some pieces after they have been sent to the mobile device, and then reuse the memory to download additional pieces. The assumption that a piece a peer has downloaded is available for others is thus no longer valid.

In order to both serve other peers and ensure that all of the announcements the proxy peer makes are valid, our solution divides the available memory into two buffers. The download buffer holds transient data on the way to the mobile; the pieces are downloaded from peers, sent to the mobile device, and discarded. Then the same memory space is reused to download other pieces. The content of the download buffer is thus constantly changing as the download of the torrent progresses. The upload buffer, on the other hand, stores pieces that are served to the swarm. After a piece in the upload buffer has been downloaded and sent to the mobile device, it remains in the memory and is made available for other peers with the normal BitTorrent piece transfer mechanisms.

The 3G and WLAN based measurement results are depicted in Figure 3. In comparison with SymTorrent, using the router-based proxy consumes 40% less energy with 3G and 55% less with WLAN. As expected doing transfers at higher speeds significantly improves the energy efficiency. In addition to better bandwidth utilization, shorter download times and lower protocol overhead contribute to energy savings.

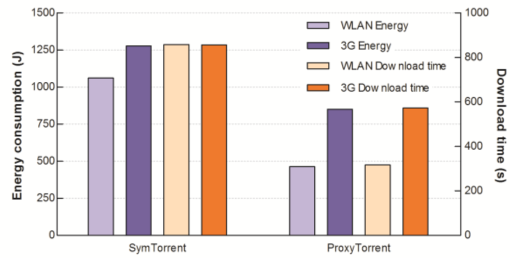


Figure 3. SymTorrent vs. ProxyTorrent mobile client energy consumption

5. Conclusions

The three solutions presented in this paper are all examples of how traffic shaping can be utilized to make peer-to-peer content sharing more energy-efficient. These concepts can be applied anywhere where wireless communication is a significant part of a system.

References

[1] J. K. Nurminen, J. Nöyränen. Energy-Consumption in Mobile Peer-to-Peer – Quantitative Results from File Sharing. In Proc. of 5th IEEE CCNC 2008, Las Vegas, US, 2008.

[2] J. K. Nurminen, Parallel Connections and their Effect on the Battery Consumption of a Mobile Phone. In Proc. of 7th IEEE CCNC, Las Vegas, USA, 2010.

[3] I. Kelényi, J. K. Nurminen: Bursty Content Sharing Mechanism for Energy-Limited Mobile Devices. 4th ACM Workshop on Performance Monitoring, Measurement and Evaluation of Heterogeneous Wireless and Wired Networks (PM2HW2N), 2009, Tenerife, Spain.

[4] I. Kelényi, J. K. Nurminen: CloudTorrent – Energy-Efficient BitTorrent Content Sharing for Mobile Devices via Cloud Services. 7th Annual IEEE Consumer Communications & Networking Conference (CCNC 2010), 2010, Las Vegas, USA.

[5] I. Kelényi, J. K. Nurminen, Ákos Ludányi: Using Home Routers as Proxies for Energy-Efficient BitTorrent Downloads to Mobile Phones. IEEE Communications Magazine Volume 49 Issue 6. pp. 142-147, 2011.

Imre Kelényi is currently an assistant lecturer in the Department of Automation and Applied Informatics, Budapest University of Technology and Economics (BME), Hungary. He received his Master's degrees in technical informatics from BME in 2007. He completed BME's PhD course in 2010. Currently he is preparing his PhD dissertation on energy efficient mobile peer-to-peer systems, which is also his primary research interest. He is the creator of the world's first ever mobile BitTorrent and Gnutella clients: SymTorrent and Symella. He has worked as a contract software engineer for various companies including Nokia, Nokia Siemens Networks and T-Mobile.

Jukka K. Nurminen is professor of data communication software at Aalto University in Finland. Before returning to academia he spent almost 25 years at Nokia Research Center working on a variety of topics related to mobile applications and services. His current interests are mobile computing, peer-to-peer applications, and the energy-efficiency of distributed solutions. Jukka received his M.Sc degree in 1986 and Ph.D. degree in 2003 from Helsinki University of Technology.

Leveraging Service Clouds for Power and QoS Management for Mobile Devices

I-Ling Yen, Yunqi Ye, Liangliang Xiao, Farokh Bastani, Department of Computer Science, University of Texas at Dallas
{ilyen, yxy078000, xll052000, bastani}@utdallas.edu

Introduction

Mobile devices are becoming the primary platforms for many users to roam around and to access service clouds. With the continuing advances in mobile devices and the supporting environment, the user tasks issued on mobile devices are increasing in complexity and sophistication. However, the limited battery life of mobile devices is a major concern, especially when recharging is not available. According to a survey from ChangeWave [1] on Apple's iPhone 3GS in 2009, its short battery life is considered as the most disliked feature by 41% of the respondents. A Nokia poll in 2009 [2] also shows that the battery life is the most important feature for a music phone, instead of sound quality. In critical applications, loss of power can adversely impact mission success and have severe consequences.

There have been a lot of works addressing the power management issues. For example, the dynamic power management (DPM) approaches [3] try to turn off the devices when they are idle. In the dynamic voltage and frequency scaling (DVFS) approaches [4] the processor can work under various performance-states (p-state) that provide tradeoffs between execution performance and energy consumption. These works focus on power management on local devices and can save 9-50% energy at the cost of reduced performance.

Cloud computing is a promising paradigm, enabling accesses to a pool of configurable computing resources such as networks, servers, storage, applications and services. In a service cloud, mobile users can access services provided in the cloud and can offload local tasks to the cloud for execution so that the local computation can be minimized, which has the potential of achieving very significant energy savings. In fact, for some computation-intensive services, it is possible to save energy as well as reduce the processing latency if they are executed in the cloud. However, task offloading involves communication costs for request uploading and response downloading. For some communication-intensive services, the power consumed for communication may exceed the potential savings if executed in the service cloud. Thus, to actually achieve power savings, it is necessary to make the execution platform selection decisions properly to balance the computation and communication cost tradeoffs.

This paper discusses the design of a QoS (quality of service) and power management (QPM) framework to achieve power saving on mobile devices by leveraging the capabilities of service clouds. Experimental results show that the approach is very effective, achieving up to 70% power savings without sacrificing quality of services.

QPM Framework

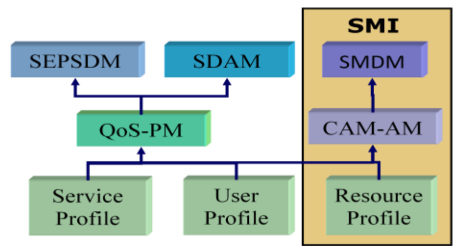


Figure 4. QPM framework

The architecture of the QPM framework is shown in Figure 1 [5]. The framework includes the service profile that provides the QoS and power behavior specifications of the services under different configurations and the user profile that specifies the users' historical service usage patterns. With the service cloud, many tasks activated by a mobile user may involve a composite service chain, including services on the mobile device and/or in the service cloud. One major component in QPM is the service execution platform selection decision module (SEPSDM) which selects the appropriate service execution platform, mobile device or service cloud, to satisfy the QoS constraints and achieve maximum energy savings. To support SEPSDM processes, it is necessary to predict the QoS behaviors with various input parameters and execution environments, which is done by the QoS prediction module (QoS-PM). To allow the selection of execution platforms, it is necessary for the service/application to be available on the mobile device as well as in the service cloud. When this is not the case, the service and associated data should be migrated to the mobile device, which is done by the service and data allocation module (SDAM). Another important component in the QPM framework is the service migration infrastructure (SMI) that is used to minimize the communication latency by migrating some frequently used services of a user to platforms in the service cloud that are closer to the user.

Accurate prediction of QoS behavior is the key to the success of the QPM framework. In QPM implementation, a function pool based prediction (FPP) model has been developed for **QoS-PM** to make accurate QoS behavior predictions. FPP is based on the parametric regression approach [6], but enhanced with a parametric function pool (FP) to facilitate the proper selection of the parametric function. FPP not only optimizes the parameters for the parametric functions but also optimizes the selections of the parametric functions based on the relations between the input and system characteristics data and QoS behavioral data stored in the service profile. It uses the conjugate-gradients method [7] to optimize the squared correlation coefficient for each QoS behavior and uses the selected relation function for QoS prediction. Experimental results show that FPP can predict QoS behaviors relatively accurately and yields higher prediction accuracy than the standard parametric regression prediction models.

The **SDAM** decisions are based on a migration cost model developed in the QPM framework. The migration cost and potential gains are evaluated in the cost model to make a service migration decision. For the backend data required in service execution, both the full and partial data migration policies are considered depending on the divisibility of the backend data, the total data size, and the QoS impacts due to the use of the partial data set. The data migration cost models for full and several partial migration strategies have also been developed in the QPM framework to effectively migrate data based on the data usage patterns in the user and service profiles.

In **SEPSDM**, an interface is provided to allow users to specify the end-to-end QoS constraints and objectives for the service chain execution. This QoS specification is used to guide the decision process for selecting the optimal service execution platforms. To support efficient on-the-fly decision making, a sub-chain selection approach is used to greatly reduce the decision space. Instead of making execution platform decisions for individual services, one entire sub-chain is selected for potential service cloud execution. To achieve the one sub-chain selection approach, a proxy based service execution architecture is used in the QPM framework (as shown in Figure 2). The proxy server mediates service execution in the cloud so that the mobile device only needs to process one sending and one receiving message, which further reduces the energy cost on the mobile device and has the potential to achieve improved QoS for the sub-chain execution.

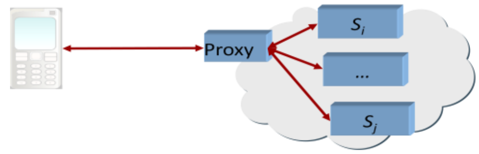


Figure 5. Proxy based service execution architecture.

Experimental Studies

Several experiments have been conducted to evaluate the effectiveness of the QPM techniques. The experimental system consists of a mobile device and a simulated service cloud. The mobile device is a Netbook with Intel® Atom(TM) CPU N270 @1.60GHz 1.60GHz, 1GB memory and Ubuntu 9.10 operating system. PowerTop [8] is used to obtain the battery power consumption data for the services running on the mobile device. The systems that simulate the service cloud are PCs with Intel® Core 2 E8400 @ 3.00GHz, 3.24 GB Memory and running Ubuntu Linux v9.10 operating system. Tomcat 6.0 is deployed on these platforms to host the services. The mobile device accesses the service cloud through a 54Mb wireless LAN. In the experiments, it is assumed that the services are already migrated to the platforms that are close to the user (with simulated communication cost that is equivalent to the communication cost between two platforms in Dallas and Houston).

A facial recognition service chain is used as a case study system. Consider a user attending a conference. She/he may recognize a face but cannot recall the person's name and other information. The user can compose a service chain as shown in Figure 3: first a picture of the person is captured by the mobile device, then the picture is resized to fit the size of the facial database, then a facial recognition service is activated to identify the person in the picture, next the detailed information about the person is retrieved, and finally the retrieved information is displayed on the device.

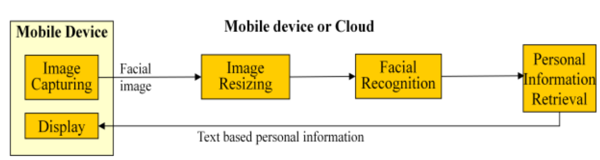


Figure 3. Service chain for facial recognition application

The execution platform selection decision for the three services in the case study is represented by "xxx", where x is "L" if the corresponding service is executed on the mobile device or "R" if the

corresponding service is executed in the service cloud. For example, LRL represents that image resizing and personal information retrieval services are executed on the mobile device while the facial recognition service is executed in the service cloud. Also, we use BASE to denote the base line configuration in which all services are executed on the mobile device and the facial image database is fully replicated on the mobile device (it is different from LLL in which SADM allocates partial database on the mobile device).
The experimental results for different input image sizes are shown in Table 1. The energy saving and latency improvements by QPM, compared to the baseline case (BASE) are summarized in the table. For this specific application, in the best case, energy consumption is reduced by 71.4% while the access latency is improved by 44.0%.

Input Size	Energy Saving	Latency improvement
50×50	66.7%	54.3%
90×100	42.9%	51.1%
200×200	71.4%	44.0%
300×300	50.0%	37.0%
400×400	37.5%	11.5%
500×500	33.3%	24.7%

Table 1. Energy saving and latency improvement by QPM

Conclusion

A QPM framework has been proposed in this paper to leverage service clouds to provide power saving on mobile devices while satisfying QoS requirements of the user tasks. Several case study service chains have been used to conduct experimental studies and one of the cases, the facial recognition service chain, is presented in Section 3. The results for all case studies confirm the effectiveness of the QPM framework approach, successfully making use of service clouds to achieve power saving and extend the battery life of mobile devices.

References

[1] http://digitaldaily.allthingsd.com/20090821/iphone-owners-would-like-to-replace-battery-att.

[2] http://conversations.nokia.com/2009/08/12/what-is-the-most-important-feature-in-a-music-phone/.

[3] Y. Lu and G. D. Micheli, “Comparing system-level power mangagement policies”, IEEE Design & Test of Computers, vol. 18, no. 2, Mar./Apr. 2001, pp. 10-19.

[4] X. Liu, P. Shenoy, M.D. Corner, “Chameleon: Application-level power management,” IEEE Trans. Mobile Computing, vol. 7, no. 8, Aug. 2008, pp. 995-1010.

[5] Y. Ye, L. Xia, I-L. Yen, F. Bastian, “Leveraging service clouds for power and QoS management for mobile devices,” Cloud 2011.

[6] J. Fox, Applied Regression Analysis, Linear Models and Related Methods, Sage Publications, Feb. 1997.

[7] Y. Notay, “Flexible conjugate gradients,” SIAM J. Sci. Comput., vol. 22, no. 4, 2000, pp. 1444-1460.

[8] http://www.lesswatts.org/projects/powertop/.

Towards a Secure Personal Cloud Computing for Mobile Thin-Clients

Yuan Tian, Biao Song and Eui-Nam Huh, KyungHee University, South Korea
{ytian, bsong, johnhuh}@khu.ac.kr

1. Introduction

The Personal Cloud [1-3] is a popular concept these days, it offers a range of task-oriented applications that users can retain full control of their data, and are free to tweak the available functionality to their likings [6]. Like everything, "Personal Cloud" has a rather fuzzy meaning, or at least one that is very changing, depending on the context. The analyst in Forrester Research, Frank Gillet, describes it is "an internal resource for organizing, preserving, sharing and orchestrating personal information and media[7]." He believes that digital devices and services will combine to create the Personal Cloud, which will shift individual computing "from being device-centric to information-centric". Similarly, Steve George[4], Canonical's vice president of business development, told InternetNews.com that the personal cloud, from the consumer's perspective, is to convert the operating system from a window to internet services.

However, some researchers do not think that personal cloud computing is much different from the cloud computing from businesses view. They believe the only differences are the types of features that are available to the users. Researchers in Business Cloud [5] concluded the notion of personal cloud as "one of the segments which refined and segmented from the cloud", which has a repository for personal information that enables users to organize and access. Essentially, the personal cloud describes a user-centric model of cloud computing where an individual's personal content and services are available anytime and anywhere, from whatever device they choose to access it.

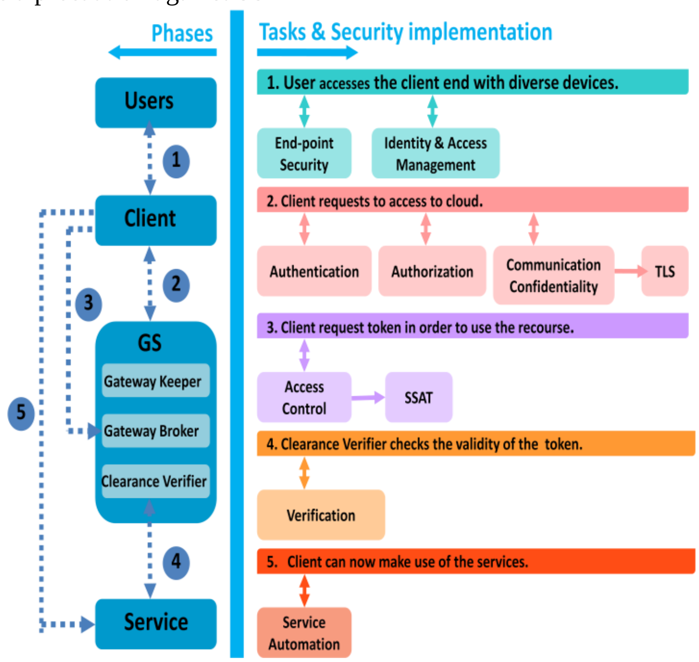
Today, most people have to juggle multiple devices to access all their services. What the personal cloud could provide is a single and portable access-point to multiple clouds. And in emerging economies, where people often share mobile devices, each individual would be able to log into their own clouds from the shared device. Meanwhile, with the increasingly development of cloud technology, the server side of thin client is applied to cloud, thus makes the full use of the features of the cloud, for example, virtualization, flexible, security and dynamic management.

This letter presents a brief review of personal cloud computing and a discussion about the security architecture of the personal cloud. Our ambitious objective is to provide a reference work to assist researchers in the definition of personal cloud and methodologies addressing the problems in the personal cloud, including the ecosystem of personal cloud and the requirements for the security framework.

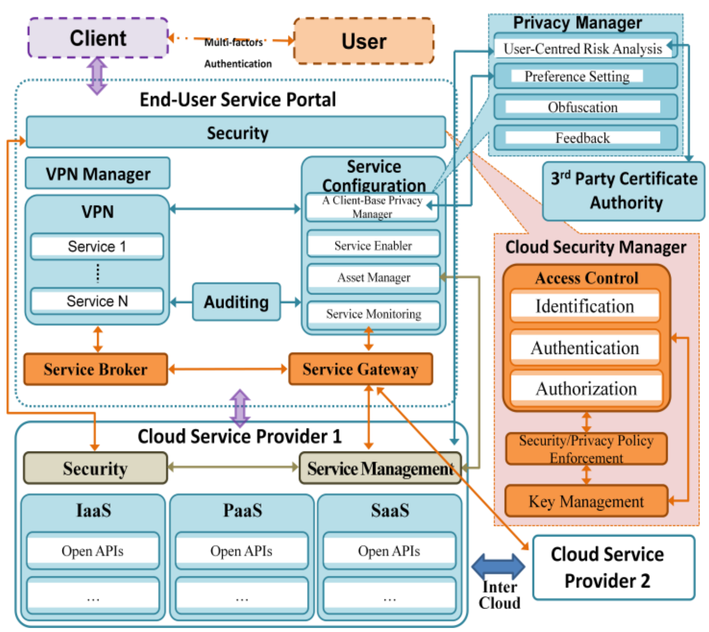
2. Personal Cloud Security Requirement Analysis

As the security for personal cloud is still in its infancy, there is an urgent need for an approach to cloud security that is holistic, adaptable, and reflects client requirements. Brock et al [8] proposed a security model which could be applied to Personal Cloud. Generally, there are two main security elements, a Gateway Server (GS) and a Single Sign-on Access Token (SSAT). GSs are hosted in clouds and manage the security of their host clouds. A SSAT is a time limited, non-forgeable and non-transferable entity that is granted to cloud clients. It is constructed and used according to the Information Flow Control model. This token identifies the client, services the client wishes to use and also verification tokens to prove the SSAT itself is valid. Only the intended client can use the token and once it expires, it cannot be reused. This addresses revoking rights. To ease management, the classification of services, and the resources behind them, is inherited from their providers.

We apply the related security technologies to each step in the Personal Cloud Model, which meets the requirements for Personal Cloud Security concern. The personal cloud security analysis is shown in Figure 1. A client program is installed on or downloaded to every endpoint (laptop, cell-phone, etc.) when user accesses the client end (phase 1). A server or gateway hosts the centralized security program, which verifies logins and sends updates and patches when needed. In the second phase, the user contacts the Gate Keeper (GK) service in the Gateway Server (GS) where the communication with the GK (or any other service) uses Transport Layer Security to protect against eves-dropping attacks (phase 2). The GS needs to securely identify their users through authentication and after that, a user must gain authorization for doing certain tasks. With the single sign-on access control token, a user logs in once and gains access to all systems without being prompted to log in again at each of them. The Clearance Verifier (CV) checks the validity of the token. If there is no verification in the SSAT, that service should contact the CV (phase 3). This step is a precaution against SSAT forging. If the CV reports back that the Gateway Server did not generate the SSAT, the request is blocked. If the SSAT is examined and proved valid, the CV attaches a verification token to the SSAT (phase 4). Client can now make use of the services.



Personal Cloud Security Requirement Analysis



Personal Cloud Security Framework

3. Personal Cloud Security Framework

In this section, we present the security framework for personal cloud based on the above analysis. The proposed framework is illustrated in Figure 2.

Users could access the client side (i.e.: web browser or host installed application) via diverse devices like

PDA, laptop, or mobile phone with multi-factors authentication provided by End-User Service Portal. The client side is the portal where users get their personal cloud. Multi-factors authentication based on certification issued by 3rd party CA. When clearance is granted, a Single Sign-on Access Token (SSAT) could be issued using certification of user. Thus, user could use services without limitation of service providers through the end-user service portal.

The security manager in the end-user portal provides significant protection for access control, security policy and key management against security threats. In the service configuration panel, the service enabler makes provision for personalized cloud service using user's profile. The user's profile is provided to the service management in cloud service provider for the integration and interoperation of service provisioning requests from user. The service monitoring system guarantees a high level of service performance and availability. The SPML can be used to share user's profile. The asset manager requests user's personalized resources with {user's profile}SPML to the service manager in the cloud service provider, and configure service via VPN connection. The service gateway manages network resources and VPN on the information lifecycle of service broker.

4. Conclusion

Personal cloud computing provides an efficient, scalable, and personalized way for individual information organizations. Meanwhile, the flexibility and openness of personal cloud have created a number of security issues. In this paper, we provided a personal cloud security framework based on the security requirement analysis. This work intends to serve as a technical reference for the development of security requirements methodologies aiming to the personal cloud.

5. Acknowledgment

This work was supported by the IT R&D program of MKE/KEIT [10035321, Terminal Independent Personal Cloud System]

References

[1] X Wu, W Wang, Ben Lin, K Miao, "Composable IO: A Novel Resource Sharing Platform in Personal Clouds", in CloudCom 2009.

[2] From Service Clouds to User-Centric Personal Clouds, Liliana ArdissonoAnna GoyGiovanna PetroneMarino Segnan, CLOUD '09, IEEE Computer Society.

[3] OpenGL Application Live Migration with GPU Acceleration in Personal Cloud, Yunbiao Lin, Wei Wang, Kyle Gui, HPDC '10,pp.280-283.

[4] Personal Cloud Computing : Ubuntu Maverick, Meerkat, http://itmanagement.earthweb.com/netsys/article.php/3907431/Personal-Cloud-Computing-Ubuntu-Maverick-Meerkat.htm

[5] The Personal Cloud, Storagezilla, http://storagezilla.typepad.com/storagezilla/2009/08/the-personal-cloud.html.

[6] Pioneering Personal Cloud - Finally the user owns the data, codemechanic, future engineering, 2010, http://www.codelathe.com/blog/index.php/2010/06/19/pioneering-personal-cloud-finally-the-user-owns-the-data/.

[7] Microsoft goes in hard as software war heats up with Google, simran, 2010, http://onlysoftwareblog.com/2010/09/microsoft-goes-in-hard-as-software-war-heats-up-with-google/.

[8] Michael Brock and Andrzej Goscinski, Toward a Framework for Cloud Security, Algorithms and Architectures for Parallel Processing, Volume 6082/2010, pp. 254-263.

Yuan Tian received her BS from Department of Computer Sciences, Hebei Normal University in 2008. Since 2008, she is working on her Master degree and currently he is a Ph.D candidate in Department of Computer Engineering at Kyung Hee University, Korea. Her research interests include privacy, security, grid and cloud computing.

Biao Song received his BS from Department of Computer Sciences, Hebei Normal University in 2008. He is currently working towards the master degree in the Department of Computer Engineering at Kyung Hee University, Korea. His research interests include task co-allocation, dynamic collaboration in cloud computing and privacy.

Eui-Nam Huh (Corresponding Author) has earned BS degree from Busan National University in Korea, Master's degree in Computer Science from University of Texas, USA in 1995 and Ph. D degree from the Ohio University, USA in 2002. He has also served for the WPDRTS/IPDPS community as program chair in 2003. He has been an editor of Journal of Korean Society for Internet Information and Korea Grid Standard group chair since 2002. He was also an Assistant Professor in Seoul Women's University, South Korea. Now he is with Kyung Hee University, South Korea as Professor in Dept. of Computer Engineering. His interesting research areas are: Cloud Computing, Ubiquitous Computing, High Performance Network, Sensor Network, Distributed Real Time System, Grid, Network Security.

Using the Cloud to Build Content Adaptation Systems for Smart Phones

Wei Hao, Northern Kentucky University, USA
Jicheng Fu, University of Central Oklahoma, USA
haow1@nku.edu, jfu@uco.edu

1. Content Adaptation Systems

With the arrival of iPhones and Android phones, smart phones are playing instrumental roles in our daily lives. Unfortunately, most of the web sites/pages are designed for desktops/laptops and not for smart phones. Researchers have proposed content adaptation approaches [1] [2] [3] [4] to present content in a suitable format for mobile devices.

Most content adaptation systems only adapt content format via column-wise, replacing, and page-splitting techniques to fit small screens of mobile devices. Those format adaptation approaches have some limitations. The adapted pages may not work well for smart phones. For example, if a web page contains flash code, the adapted page by the existing format adaptation approaches still contains the flash code, which is not supported by the iPhone. When the iPhone displays the adapted page, the flash code area will become a blank area. The URL link information in the flash area will be lost. Also, the page layout (content block and link layout) is unchanged. Mobile users need to give the same amount of clicking and scrolling as desktop users do to find their desired content. Compared with desktop computers, smart phones have not only smaller screens but also constrained input capabilities. Clicking and scrolling on smart phones are not as easy as on desktop computers. Also, more scrolling and clicking would lead to the consumption of more battery power.

A good content adaption system should perform intelligent adaptation for smart phones. Based on surveys and log files, [5] discovers access patterns of mobile users and designs an intelligent way to perform layout adaptation (block and link adaptation) for smart phones. The intelligent adaptation system comes at a price: computation overhead. Where should we run the content adaptation system? If the content adaptation runs on smart phones, it would cause the phones to consume more power, therefore decreasing battery life. If the content adaptation runs on the web server side, it can only serve one web site. Proxy servers are widely deployed around the Internet. They are the ideal platforms to perform content adaptation and information filtering for mobile users. If the content adaption runs on proxy servers, it can serve thousands of web sites. However, it is very hard for administrators to predict the workload on the proxy server. Over-provisioning of computing resources would be a big waste of money. Under-provisioning of computing resources would result in poor performance. How should we provision the proxy server with the content adaptation function to achieve the best blend of good performance and cost optimization?

2. Moving Content Adaptation Systems to the Cloud

Recently, Cloud computing [6] has drawn a lot of attention from industry and academia. The Cloud provides sharing of a pool of resources. It provides high scalability via dynamic provisioning of resources. It can relieve system administrators of worry about peak-load capacity. It can improve the utilization and efficiency of underutilized computing resources. For example, there are many computing resources in the university. Some of them are heavily used by students and faculty. Some of them are underutilized or even idle. Cloud computing would be a good solution to balance out computing resources and maximize utilization of computing resources.

The Cloud would be a good place for the proxy server with the content adaptation function. A private Cloud can be easily built on top of existing IT resources and open source software platform, such as Eucalyptus [7]. The private Cloud can scale up and scale down the exiting computing resources on demand to meet the fluctuating workload on the proxy server. In addition, the virtual machines in the Cloud can provide a secure platform for code and service migration [8] [9]. Code/service migration would achieve better performance for the content adaptation system.

3. How Can We Transparently Intercept Mobile Requests and Redirect them to the Cloud?

Since manually configuring phones to utilize proxy servers in the Cloud can become very tedious, WCCP (Web Cache Communication Protocol) [10] can be used to transparently redirect mobile requests. WCCP is a protocol designed by Cisco Systems to redirect traffic flows in real-time. It also supports load balancing, service-assurance, and fault tolerance.

WCCP can detect and transparently redirect mobile requests to one or multiple proxy servers in the Cloud. The redirected requests are forwarded via a GRE (Generic Routing Encapsulation) [11] tunnel to the proxy server. The GRE tunnel can handle multiple sessions keeping the data from interfering with any other web data going to the server.

5. Conclusions

With the current trend, we believe that more and more smart phones will be connected to the Internet. Content adaption systems can achieve better user experience for smart phone users. However, resource over-provisioning for content adaptation systems would cause a huge waste of money; while resource under-provisioning for content adaptation systems would suffer from poor performance. Due to the elastic nature of the Cloud, moving content adaptation systems to the Cloud will achieve the best blend of good performance and cost optimization.

References

[1] Y. Chen, W. Ma, H. Zhang “Detecting Web Page Structure for Adaptive Viewing on Small Form Factor Devices.” In Proc. World Wide Web 2003, Budapest, Hungary, 2003.

[2] Z. Hua, X. Xie, H. Lu, W. Ma, “Providing Adaptive Dynamic Web Content in Mobile Environment”, Tech Report, MSR-TR-2004-116, Microsoft Research

[3] Y.Chen, X. Xie, W.Ma, and H. Zhang, "Adapting Web Pages for Small-Screen Devices", IEEE Internet Computing, Vol. 9, No. 1, 2005, pp. 50-56.

[4] S. Yang, I. Chen, R. Chen, “Applying Content Adaptation Technique to Enhance Mobile Learning on Blackboard Learning System”, Seventh IEEE International Conference on Advanced Learning Technologies, Niigata, Japan, July, 2007.

[5] W. Hao, A. Lopez, and C. Walker, "A Novel Proxy-based Content Adaptation System for Mobile Devices", fifth IASTED International Conference on Human-Computer Interaction (HCI 2010), Maui, Hawaii, August, 2010.

[6] B. Hayes, “Cloud Computing”, Communications of the ACM, Volume 51, Issue 7, July 2008.

[7] D. Nurmi, R. Wolski, C. Grzegorczyk, G. Obertelli, S. Soman, L. Youseff, D. Zagorodnov, “The Eucalyptus Open-Source Cloud-Computing System”, 9th IEEE/ACM International Symposium on Cluster Computing and the Grid , Washington DC, 2009.

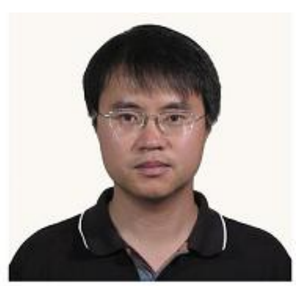
[8] W. Hao, J. Fu, P. Prapatanant, N. Dixon, "Achieving Better Cloud Access Experience with Server Virtualization", In 20th International Conference on Software Engineering and Data Engineering (SEDE 2011), Las Vegas, Nevada, June, 2011.

[9] W. Hao, I-L. Yen, and B. Thuraisingham, "Dynamic Service and Data Migration in the Clouds", 33rd Annual IEEE International Computer Software and Applications Conference, Seattle, Washington, July, 2009.

[10] Web Cache Communication Protocol v2, http://www.ciscosystems.com/en/US/docs/ios/12_0t/12_0t3/feature/guide/wccp.html

[11] RFC2784 - Generic Routing Encapsulation (GRE), http://www.faqs.org/rfcs/rfc2784.html

Wei Hao is an assitant professor in Computer Science Department at Northern Kentucky University. Prior to joining NKU in August 2008, he worked as a software engineer at Cisco Systems in San Jose, California. He also worked for Motorola and Alcatel. He earned his PhD in Computer Science from the University of Texas at Dallas in 2007. His research interests include Web Technologies (such as Web Caching, Web Services, and Web-Based Systems), Cloud Computing, Mobile Computing, and Operating Systems.



Jicheng Fu is an assistant professor in the Computer Science Department at the University of Central Oklahoma. He earned his PhD in Computer Science from the University of Texas at Dallas in 2009. His research interests include Automated Software Engineering, Artificial Intelligence Planning, and Bioinformatics.

Augment Mobile Cloud Computing with Cognitive Radio

Feng Ge, Cho-Yu J. Chiang, and Ritu Chadha, Telcordia Technologies, Inc., Pisataway, NJ
Ahmed M. Eltawil, Department of EECS, University of California, Irvine, CA
Heshan Lin and Wuchun Feng, Department of Computer Science, Virginia Tech, Blacksburg, VA
fge@telcordia.com, aeltawil@uci.edu, {hlin2, feng}@cs.vt.edu, {chiang, chadha}@research.telcordia.com

1. Introduction

Mobile cloud computing, which extends cloud computing to mobile devices via ubiquitous wireless access, is at the tipping point [1-3]. Its technology foundation has been laid down: first, mobile devices such as super-phones and smart-phones are pervasive in many people's daily life [1]; second, cloud computing infrastructure has been widely deployed [3]; third, the emerging broadband networks (LTE and WiMAX for 4G and IEEE 802.11n for Wi-Fi) provide high data rates with low latency, allowing easy access to the cloud from mobile devices. With the explosion of mobile Internet applications such as Facebook, mobile games, and Google Maps, mobile cloud computing is poised to profoundly shape the wireless landscape.

However, mobile cloud computing still faces key technical challenges in the above three areas: mobile devices, wireless networking, and cloud computing[4]. For mobile devices, the bottleneck is the limited energy that can be stored in a battery [4], a technology growing only 5% annually [5]. With service and data moved to the cloud, more applications—some demand more communications or computing resource—will run through the mobile devices. Supporting these applications will significantly shorten the battery life, thereby vexing users and slowing down the growth of mobile cloud computing. Further, the continuous increase of processor speed and wireless data rate—both imply more power consumption—exacerbates this problem. As users begin to own multiple mobile devices, synchronizing them through the cloud becomes a challenge. The solution by Apple's iCloud is to store contents almost exclusively over the cloud and allow access from different devices. By this solution, however, multi-media applications may drain out the mobile battery quickly.

[4]Here we only discuss challenges specific to mobile cloud, not cloud computing in general.

Wireless networking poses challenges to mobile cloud computing because users expect persistent connectivity, a high data rate, and low access latency. Even though the emerging 4G network (based on LTE or WiMAX) offers a high data rate (100 Mb/s for LTE downlink and 50 Mb/s for LTE uplink [6]) with better coverage (assisted by femtocells), the exponentially increasing wireless traffic, as shown by a Cisco report [7], may constrain the data rate assigned to a single user at peak times, similarly to the impact of the initial introduction of iPhone that resulted in an abrupt jump of data traffic and then congested Verizon's 3G network. Further, the available spectrum for 4G networks is limited; therefore, the total network bandwidth is constrained—as dictated by the Shannon–Hartley theorem. As more applications and users move to the cloud, a high data rate may not be guaranteed for each user at peak times.

Persistent connectivity is challenging because the wireless environment is dynamic and even hostile, characterized by multi-path, fading, and interference [8]. To mitigate their impact, LTE provides some level of coding-modulation adaptation; however, currently this capability is limited only to a set of pre-selected coding and modulation schemes. The adaptation must further take into account the constrained battery capacity because some coding-modulation consumes more power than others. Another added dimension of complexity is the heterogeneity of wireless networks, e.g., Wi-Fi and 4G networks. Users would expect consistent connectivity even when switching access among different networks and swapping devices.

Finally, decreasing mobile user access latency is a daunting task because mobile data must traverse many complex layers, each of which adds network latency and transmission delay. Scalable and dynamic network monitoring must be applied to optimize network and device costs against the users' perceived performance of cloud applications. Network intermittent performance and congestion must be mitigated through dynamic traffic re-routing and swapping, or handover, between cells based on traffic load patterns and user locations.

Cognitive radio (CR), capable of adaptive learning and reconfiguration [9], is a promising technology to address the above challenges in three aspects: (1) offering persistent connectivity by adapting to the environment [4]; (2) allowing flexible access to the cloud through dynamic spectrum access [9, 10]; and (3) achieving energy efficiency in mobile devices by effective power management [11]. This paper will detail our efforts in the above three areas.

Furthermore, one major challenge specific to the mobile cloud is choosing the right application computing model to mitigate mobile devices' energy consumption without introducing much latency. Several models have been proposed, including CloneCloud, fat client, thin client, and weblet as well as cloudlets and proxies [12]. Dynamic partition of application execution between mobile devices and the cloud is widely adopted; this approach reduces resource consumption even in heterogeneous and dynamic computing environments [12]. Associated with the approach, though, is the volatility of the available cloud computing resources. In this paper, we also discuss our opportunistic cloud designed to address this problem [13].

2. Cognitive Radio

One promising application of CR is dynamic spectrum access (DSA), which is endorsed by spectrum regulators worldwide (e.g., FCC in USA and OfCOM in UK) and many industry players such as Google, Microsoft, and Intel. By this method, a mobile device can dynamically access spectrum holes (e.g., TV whitespace) in an unlicensed manner without interfering with primary users. As major carriers are moving toward charge-by-volume plans while the data usage by mobile customers is increasing, the huge monthly bill may discourage adoption of mobile cloud applications. With DSA, carriers don't have to pay billions of dollars for exclusive spectrum licenses; therefore cheap wireless access to the cloud is likely. Further, with techniques of aggregating spectrum holes [6], high data rates can be supported. Finally, DSA allows new wireless service providers to access spectrum quickly by avoiding the prolonged spectrum license application process; thus, new services and applications can emerge quickly.

Cognitive radio is more than accessing spectrum dynamically. It is aware of and adapts to its environment as well as configures itself in an appropriate fashion, based on its knowledge of users' priorities, needs, operational procedures, and governing regulatory rules [9]. As such, consistent communications are possible. Towards that goal, we have been developing advanced DSA technologies by designing and demonstrating a decentralized, software defined radio (SDR)-based CR network prototype [10, 14]. Currently this network enables distributed cooperative spectrum sensing and management, dynamic spectrum access, and cross-layer network routing optimization. The prototype uses SDR technologies, signal detection and classification methods, distributed cooperative spectrum sensing systems, dynamic wireless protocols, and a multi-channel allocation algorithm. Systematic experiments were carried out to identify several performance determining factors for decentralized DSA networks [10].

Based on the developed technologies, we further built an experimental cognitive radio network for the U.S. Army Communications Electronics Research, Development, and Engineering Center (CERDEC) under the Tactical Information Technologies for Assured Networks (TITAN) program [15]. This network consists of 10+ SDR nodes [14] and provides advanced management capabilities for mobile ad hoc networks to enable adaptive, resilient, and reliable communications. Some demonstrated capabilities include (not exhaustively) (1) automatically re-distributing resources to address network capacity requirement changes in real time; (2) differentiating between congestion caused by legitimate users vs. malicious users and quarantining malicious nodes; and (3) automatically adjusting the sending rate of different applications based on available network capacity and traffic priority.

Finally, we developed a novel method for power management of mobile terminals [11] that results in significant savings by jointly factoring in knowledge about the application quality of service (QoS), the wireless channel, and the hardware status (reliability). Our approaches target embedded memory which is becoming a dominant portion of most transceivers in terms of area and power metrics. Buffering and intermediate memories store data samples off the wireless channel which introduces significant variability in the quality of the data. Thus, it is not efficient to store this data in a memory that always maintains a fixed hardware quality, independent of the quality of the data stored within. In our work, we utilize the knowledge of the channel to modulate the quality (hence power consumption) of embedded memories. When the received data experiences a “good” wireless channel (e.g., with high SNRs and less interference), the stringent constraints enforced on the underlying hardware is relaxed to reduce power consumption while maintaining the desired QoS. In this approach, each layer (e.g. hardware, network, and application) is aware of other layers’ status and adjusts its parameters with the goal of minimizing the power consumption while meeting the required QoS. Within this work, statistical memory failure models and error distributions have been developed and propagated through communication system models [16, 17]. Furthermore, Forward Error Correction (FEC) blocks that are resilient to both channel and hardware failures have been developed [18]. These techniques resulted in savings of up to 40% as compared to traditional methods.

3. Opportunistic Cloud Computing

Mobile devices are resource-poor; their weight, size, battery life, and ergonomics constrain the computational capacity such as processor speed, memory and disk size. Thus, elastic application models are usually used to augment the capabilities of mobile devices with cloud resources; mobile devices acquire and release resources on demand [12]. Given the large number of mobile users and the various granularities of mobile applications, cloud resources can be highly volatile while still supporting computationally intensive applications for none-mobile users. In this environment, existing programming models for cloud computing, e.g., MapReduce, perform poorly or even do not guarantee job completion.

As a preliminary step toward addressing this issue, we developed a technology called MOON–MapReduce On Opportunistic eNvironments [13], which allows us to build a private cloud using low-cost opportunistic PC resources (i.e., a desktop can only be used for cloud computing when it is not actively used by the owner—this mimics the volatile nature of mobile cloud). The key idea of the MOON design is a hybrid resource-provisioning model, where volatile nodes are supplemented by a small number of dedicated, reliable nodes. By carefully extending existing MapReduce designs to take advantages of the hybrid resources, MOON significantly improves the performance and service quality of MapReduce programs in volatile environments. For instance, in traditional MapReduce design, data is uniformly replicated on all compute nodes. On volatile resources, the cost of replication for high data availability can be prohibitively high. By placing data replicas on highly available dedicated nodes, MOON can achieve high data availability with much less data replicas. Also, on pure volatile resources, jobs with execution time longer than mean time between failures will have difficulties to finish. MOON can guarantee the completion of these jobs by placing them on dedicated nodes.

4. Conclusions

Augmenting mobile cloud computing with cognitive radio yields significant benefits. We have explored some of these opportunities and introduced our work in dynamic spectrum access, wireless adaptive learning and reconfiguration, and cognitive power management. We further developed MOON to resolve the volatility problem of available computing resource.

On the other hand, cognitive radio can offload to the cloud resource-intensive functions (e.g., computer vision, speech recognition, and machine translation) [4], thereby enhancing their performance while reducing mobile device's energy consumption. Thus, the integration of cognitive radio and mobile cloud computing—both a convergence of communications, computing, and network technologies—overcomes challenges faced by each [4].

References

[1] Chetan Sharma, “Mobile Cloud Computing,” 2011, Available: http://www.chetansharma.com/Mobile_Cloud_Computing_Final.pdf.

[2] iCloud is here. http://www.apple.com/icloud/

[3] M. Armbrust, et al., “Above the clouds: A berkeley view of cloud computing,” University of California, Berkeley, Tech. Rep. UCB/EECS-2009-28, 2009.

[4] F. Ge, et al., "Cognitive Radio Rides on the Cloud," Military Communications Conference, 2010.

[5] A. P. Miettinen, et al., “Energy efficiency of mobile clients in cloud computing,” in Proceedings of the 2nd USENIX conference on Hot topics in cloud computing, 2010.

[6] Linda Doyle, et al., "A Letter from the Editors," Cognitive Radio for LTE Advanced & Beyond: Special Issue of Physical Communication, 2011.

[7] Cisco, “Cisco visual networking index: Global mobile data traffic forecast update,” 2009.

[8] G. H. Forman, et al., "The challenges of mobile computing," Computer, vol.27, no.4, pp.38-47, 1994.

[9] F. Ge, et al., "Cognitive Radio: From Spectrum Sharing to Adaptive Learning and Reconfiguration," IEEE Aerospace Conference, 2008.

[10] F. Ge, et al., "Software radio-based decentralized dynamic spectrum access networks: A prototype design and experimental results," in IEEE Global Communications Conference, 2010.

[11] M. Khairy, A. K. Djahromi, A. Eltawil, and F. Kurdahi, "Autonomous Power Management Technique for Communication Systems," in Design Automation Conference, San Diego, CA, 2011.

[12] D. Kovachev, Y. Cao, and R. Klamma, "Mobile cloud computing: a comparison of application models," CoRR abs/1107. 4940, 2011.

[13] H. Lin, et al., "Reliable MapReduce Computing on Opportunistic Resources," Cluster Computing, 2011.

[14] F. Ge, et al., "GNU Radio-based digital communications: Computational analysis of a GMSK transceiver," IEEE Globecom, 2011.

[15] S. Rushen, "CERDEC PMs Watch as Demo Puts 'Field into a Box'," the Defense Systems Online magazine, September, 2010.

[16] A. Khajeh, et al., "A Unified Hardware and Channel Noise Model for Communication Systems," in Proceedings 2010 IEEE Global Communications Conference (GLOBECOM 2010), 2010, p. 5 pp.

[17] M. S. Khairy, et al., "FFT processing through faulty memories in OFDM based systems," in 2010 IEEE Globecom Workshops, 2010, pp. 1946-51.

[18] A. Hussien, et al., "Combined Channel and Hardware Noise Resilient Viterbi Decoder," in IEEE Asilomar Conference on Signals, Systems, and. Computers, Pacific Grove, CA, 2010.

Feng Ge is a Senior Research Scientist at Telcordia Technologies, Inc. His current research interests are in software-defined radio, cognitive radio, and adaptive wireless networks. Dr. Ge (with his team) won the grand prize in the 2007 inaugural Smart Radio Challenge.

Ahmed M. Eltawil is an Associate Professor at the University of California, Irvine. His current research interests are in low power digital circuit and signal processing architectures for wireless communication systems where he has published more than 70 technical papers, including 4 book chapters. He has received several distinguished awards, including NSF CAREER award in 2010 and the Best Paper Award in 2006 at ISQED. Since 2006, he has been actively involved in efforts towards integrating cognitive and software defined radio technology in first responder communication networks.

Heshan Lin is a Senior Research Associate in the Department of Computer Science at Virginia Tech. He received the M.S. degree in computer science from Temple University in 2004, and the Ph.D. degree in computer science from North Carolina State University in 2009. His current research interests include data-intensive parallel and distributed computing, bioinformatics, cloud computing, and GPU (graphics processing unit) computing. He received the Distinguished Paper Award from International Supercomputing Conference in 2008 and the HPC Storage Challenge Award from SC 2007.

Wu-chun Feng is an associate professor of computer science and electrical & computer engineering at Virginia Tech, where he directs the Systems, Networking, and Renaissance Grokking (SyNeRGy) Laboratory. His research interests span many areas of high-performance networking and computing from hardware to applications software. Dr. Feng is the recipient of three Best Paper Awards in human-computer interaction, high-performance networking, and bioinformatics, respectively, and three R&D 100 Awards in green supercomputing, high-speed networking, and bioinformatics, respectively. Dr. Feng's previous professional stints include IBM T.J. Watson Research Center, NASA Ames Research Center, Vosaic, University of Illinois at Urbana-Champaign, Purdue University, The Ohio State University, Orion Multisystems, and Los Alamos National Laboratory.

Cho-Yu J. Chiang is Chief Scientist and Director of Autonomous Wireless Communications group at Applied Research, Telcordia Technologies. His current research interests are in cognitive radio network, dynamic spectrum access, self adaptation and self organization, middleware for dynamic networks, testbeds for emerging wireless networks, policy-based network management, as well as reasoning and diagnosis of network faults. He has been leading multiple research programs in the above research areas.

Ritu Chadha is Executive Director at Telcordia, where she manages the Knowledge-Based Systems department in Applied Research at Telcordia Technologies. Dr. Chadha is an industry expert in the area of policy management and has recently authored a book on the subject. She received her Ph.D. in Computer Science from the University of North Carolina at Chapel Hill in 1991. Her research interests include policy-based management, network and service management for IP-based networks, ad hoc networking, and automated reasoning. Dr. Chadha is a Telcordia Fellow.

CALL FOR PAPERS

IEEE Multimedia Magazine: Special Issue on 3D Imaging Techniques and Multimedia Applications

Schedule:
Manuscript submission: Jan. 31, 2012
Acceptance/Revision notification: May 15, 2012
Revised manuscript due: July 15 2012
Final acceptance notification: Sept. 15, 2012
Final manuscript due: Oct. 15, 2012
Tentative publication: Jan. 2013

Guest Editors:
Ruzena Bajcsy, University of Berkeley (bajcsy@eecs.berkeley.edu)
Ruigang Yang, University of Kentucky (ryang@cs.uky.edu)
Pietro Zanuttigh, University of Padova (pietro.zanuttigh@dei.unipd.it)
Cha Zhang, Microsoft Research (chazhang@microsoft.com)

With the advances in sensing, transmission and visualization technology, 3D information has become more and more incorporated in real world applications, from architecture, entertainment, manufacturing, to security. By integrating the depth perception, a richer media interface can be presented. For example, in immersive telecommunication, spatialized audio and 3D parallax increases the effectiveness of communication; in medicine, 3D tracking of instruments enables more precise and safer operations; and new low-cost 3D cameras are starting a new chapter in interactive gaming and human-computer interaction.

One of the fundamental requests for these applications is the estimation of the scene depth information. The extraction of 3D information has been studied in the field of computer vision for over three decades, but it remains a very challenging problem, in particular under unconstrained environments, where the lighting can be adversary, the scene surface can be specular and deforming, the objects can occlude each other, etc. Multimedia researchers must take the imperfectness of the depth information into consideration when design their systems, making it a unique research opportunity.

This special issue aims to provide an overview of recent rapid advances in 3D acquisition systems, and many multimedia applications that can benefit from 3D integration and understanding. We solicit original contributions in the areas related to, but not limited to, the following:

- 3D Acquisition
  - Scene reconstruction from passive sensors: stereo vision, shape-from-x (silhouette, shading, focus, etc);
  - Active range sensors: laser scanning, structured light, time-of-flight, etc
  - 3D from audio sensors (e.g., microphone arrays)
  - Other 3D reconstruction techniques
  - Fusion of depth information acquired from different sensors and techniques
  - 3D sensor systems
  - Multi-view imaging and light fields
- 3D integration and understanding in multimedia applications
  - 3D audio/visual synthesis
  - Novel human computer interaction exploiting depth information
  - Augmented reality
  - Multimedia medical applications using volumetric information
  - Multimedia security applications using 3D information
  - User study of the effectiveness of 3D for multimedia applications

  - Other multimedia applications that integrates 3D understanding
- Transmission and visualization of 3D scenes
  - Compression and streaming of 3D representations
  - Spatialized audio rendering
  - Stereoscopic and autostereoscopic display technologies
  - 3D visualization for multimedia applications
  - 3D immersive telecommunication

Papers should be formatted according to the IEEE Multimedia Magazine guidelines for authors (see: http://www.computer.org/portal/web/peerreviewmagazines/acmultimedia). Manuscripts should be submitted electronically through the online IEEE manuscript submission system at https://mc.manuscriptcentral.com/mm-cs. When selecting a manuscript type, authors must click on Special Issue on 3D Imaging Techniques and Multimedia Applications.

Advances in Multimedia: Special Issue on Communication of Human Affect using Smart-Multimedia

Call for Papers

While CPU power, machine learning and artificial intelligence have gradually eliminated tedious manual procedures and, to a certain extent, have automated human analysis and decision making, human intentions and responses remain the central command in many applications. For example, Affective Computing attempts to understand human feelings, Quality of Experience focuses on understanding human satisfaction, and Visual Analytics aims at analyzing perceived content of human viewer. In social web and games, human is often embodied as a virtual character (avatar) either explicitly or implicitly, interacting with the virtual or augmented world. Interaction is a mean to reflect the emotion or thought of a human user. With the advancement of 3DTV technologies, e.g., stereoscopic display, multi-view video and high-definition immersive environment, strengthened by the many well developed signal processing and communication techniques, human affect, i.e., facial expression and body language, is able to be detected, understood and translated effectively.

Scope

The focus of this special issue is to explore this emerging area and provide a venue for exchanging novel algorithms and experimental findings. We invite submissions on, but are not inclusive to:

- Acquisition and modeling of human affect.
- Expressing participant's human affect in virtual and augmented multimedia environment.
- Coordination of human affect in a collaborative environment, including agent collaboration.
- Visualization and analysis of human affect using 3DTV technologies, e.g., stereoscopic display, multi-view video, free-viewpoint and glasses-free viewing.
- Enhancement of human affect using depth and multimedia information.
- Real-time transmission of human affect data over the Internet, which include cloud and mobile computing.
- Compression, quality assessment and optimization of human affect in the multimedia processing pipeline.
- Human affect manipulation interfaces, tools and applications.
- Analysis of human affect in healthcare, rehabilitation, games, and education.

Submission

Before submission authors should carefully read over the journal's Author Guidelines, which are located at: http://www.hindawi.com/journals/cdi/guidelines.html. Prospective authors should submit an electronic copy of their complete manuscript through the journal Manuscript Tracking System at http://mts.hindawi.com/ according to the following timetable:

Important Dates

- Submission: March 18, 2012
- First Round of Reviews: June 18, 2012
- Publication Date: September 18, 2012

Guest Editor

Irene Cheng, locheng@ualberta.ca
Mohamed Daoudi, mohamed.daoudi@telecom-lille1.eu
Pourang Irani, irani@cs.umanitoba.ca
Joonki Paik, paikj@cau.ac.kr