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

R-LETTER

Vol. 6, No. 3, June 2015

IEEE COMMUNICATIONS SOCIETY

TABLE OF CONTENTS

Message from the Review Board Directors

Welcome to the June 2015 issue of the Review Letter (R-Letter) of the IEEE Communications Society Multimedia Communications Technical Committee (MMTC). This issue comprises nine papers and is brought to you by review board members who independently nominated research papers published within IEEE MMTC sponsored publications and conferences.

We hope that this issue **stimulates your research in the area of multimedia communication** featuring topics:

- imaging technology (3 papers);
- multimedia networking (2 papers); and
- multimedia streaming and delivery (4 papers).

An overview of all reviews are provided in the following:

The **first paper**, published in the *IEEE Transactions on Multimedia* and *edited by Jun Zhou*, explores semantic attributes for large scale image search.

The **second paper**, published in the *IEEE Transactions on Multimedia* and *edited by Pavel Korshunov*, describes an approach for semantic-driven color imaging taking into account context information.

The **third paper** is *edited by Bruno Macchiavello* and has been published within the *IEEE Transactions on Image Processing*. It provides an annotated dataset and model for salient object detection in complex scenes.

The **forth paper**, published in the *IEEE Transactions on Multimedia* and *edited by Frank Hartung*, provides means for video QoE control in distributed OpenFlow networks.

The **fifth paper**, published in the *IEEE Journal on Selected Areas in Communications* and *edited by Lifeng Sun*, is about delay-aware WiFi offloading.

The **sixth paper**, published in *INFOCOM 2012* and edited by *Michael Zink*, unreels Netflix and helps to understand and improve multi-CDN movie delivery.

The **seventh paper**, published in *ACM Multimedia 2014* and edited by *Carsten Griwodz*, proposes an approach enabling the correction of depth compression for planar scenes.

The **eighth paper**, published in *IEEE Transactions on Multimedia* and edited by *Carl James Debono*, deals with the transmission of video chat applications over wireless networks.

Finally, the **sixth paper** is *edited by Roger Zimmermann* and published in *IEEE Transactions on Multimedia*. It alleviates the effects of early user departures in progressive steaming scenarios.

We would like to thank all the authors, nominators, reviewers, editors, and others who contribute to the release of this issue.

IEEE ComSoc MMTC R-Letter

Director: Christian Timmerer
Alpen-Adria-Universität Klagenfurt, Austria
Email: christian.timmerer@itec.aau.at

Co-Director: Weiyi Zhang
AT&T Research, USA
Email: wzhang@ieee.org

Co-Director: Yan Zhang
Simula Research Laboratory, Norway
Email: yanzhang@simula.no

IEEE COMSOC MMTC R-Letter

Exploring Semantic Attributes for Large Scale Image Search

A short review for "Fine-grained Image Search"
(Edited by Jun Zhou)

Lingxi Xie, Jingdong Wang, Bo Zhang, and Qi Tian. "Fine-grained image search", IEEE Transactions on Multimedia, Vol. 17, No. 5, pages 636-647, 2015.

With enormous amount of images available online, contents-based image retrieval has become a very challenging task. Several paths have been explored. These include recent progresses on deep learning methods which try to build high-level abstractions of data, normally in a bottom-up manner, by using hierarchical model architectures [1]. Due to the complexity in concept organization and mapping, ontology based knowledge representation has also been studied so as to construct top-down hierarchy of object classes [2,3]. To this end, large benchmark datasets, such as ImageNet [1], has used ontology in WordNet [4] to arrange image categories.

Researchers have proposed different ways to explore such semantic ontology. An example is assigning vague labels to objects so as to enable coarse to fine category optimization [5]. To cope with the naturalness of labels embedded in language models, entry-level concept is proposed to allow common naming convention of object classes be used in the modelling. It tries to maximize the difference between the naturalness and distance between two labels in the hypernym structure so the optimal concept mapping can be achieved from one label to the other [6].

While entry-level concept in [6] tries to use vague labels to represent several similar categories, the method proposed in this paper goes into an opposite direction. It explores fine-grained image database hierarchy in which image categories are defined on very specific concept. The authors defined three levels to model the relevance between query image and candidates. The first level is semantically matched, which means two images belong to the same fine-grained categories, such as the same species of dog. The second level is semantically similar, which refers to images belonging to the same general concept, for example, birds. This is similar to the entry-level concept mentioned above. The last level is irrelevant which indicates two images are not at the same basic level, for example a bird against a building.

Given an image databased composed of images from several fine-grained classes, near-duplicate instance groups, and complete irrelevant images, the overall framework of this approach contains two stages. The first stage is offline indexing, which aims at learning a fine-grained judger and several fine-grained classifiers. Given fine-grained classifiers, classification scores for images can be used to generate their normalized vector representations, i.e., the attribute vectors. These vectors are then used to construct semantically co-indexed inverted file as a preparation process to build high-level semantic property for the online query stage.

In the second stage, online querying, the learned fine-grained judger is used to predict whether a query image belongs to one of the fine-grained classes. If the answer is yes, its semantic attribute is calculated using the corresponding fine-grained classifier, and then compared with those targets in the database. During this step, hashing approach is used to speed up the searching process. On the contrary, if the answer from fine-grained judge is no, visual words of query is used to retrieve the near-duplicate search results based on the co-indexed inverted file.

In this paper only three fine-grained object categories and three levels of relevance are included. In practice, this framework shall be extended by incorporating more object categories and use more hierarchical semantic levels in the ontology. Nevertheless, authors showed that their method is effective in fine-grained recognition and searching problem by using a database of more than one million images.

In summary, this paper introduced a new image retrieval setting and system framework in fine-grained image search. This may enlighten new research on image database organization, semantic attributes usage, relevance evaluation, and novel applications.

References:

[1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge". arXiv:1409.0575, 2014.

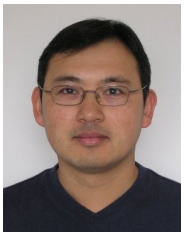
[2] J. Deng, A. Berg, and L. Fei-Fei, "Hierarchical semantic indexing for large scale image retrieval". Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 785-792, 2011.

[3] A. Popescu, C. Millet, and P. Moellic, "Ontology driven content based image retrieval". Proceedings of the 6th ACM International Conference on Image and Video Retrieval, pages 387394, 2007.

[4] G. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. "WordNet: An online lexical database". International Journal of Lexicograph. 3, 4, pp. 235–244, 1990.

[5] J. Deng, J. Krause, A. Berg, and L. Fei-Fei. "Hedging your bets: Optimizing accuracy-specificity trade-offs in large scale visual recognition". Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3450-3457, 2012.

[6] V. Ordonez, J. Deng, Y. Choi, A.. Berg, and T. Berg. "From Large Scale Image Categorization to Entry-Level Categories", Proceedings of the IEEE International Conference on Computer Vision, 2013.



Jun Zhou received the B.S. degree in computer science and the B.E. degree in international business from Nanjing University of Science and Technology, China, in 1996 and 1998, respectively. He received the M.S. degree in computer science from Concordia University, Canada, in 2002, and the Ph.D. degree in computing science from University of Alberta, Canada, in 2006.

He joined the School of Information and Communication Technology in Griffith University as a lecturer in June 2012. Prior to this appointment, he had been a research fellow in the Australian National University, and a researcher at NICTA. His research interests are in statistical pattern recognition, interactive computer vision, and their applications to hyperspectral imaging and environmental informatics.

Semantic-driven Color Imaging

A review for "Semantic-Improved Color Imaging Applications: It Is All About Context"
(Edited by Pavel Korshunov)

Albrecht Lindner and Sabine Süsstrunk. "Semantic-Improved Color Imaging Applications: It Is All About Context", in IEEE Transactions on Multimedia, volume 17, issue 5, pages 700-710, May 2015.

Classical computer vision approach infers semantic information by analyzing images or video. Examples of such applications include automatic image and video classification, object recognition, labeling of objects, and image annotation. Such approach can be viewed as a mapping from the image domain to the semantic domain.

This paper considers the inverse problem, which is a mapping from the semantic domain to the image domain. The focus of this paper is on translating the existing semantic contextual information into appropriate color imaging actions.

This reversed view on semantic imaging assumes that images have associated semantic context, which is a safe assumption in the current era of social networks and media sharing. Many online images are surrounded with various related contextual information, including titles, tags, geo-position, and comments.

Assuming a large dataset of images with contextual descriptors, the authors present an automated statistical framework that is able to relate keywords from the semantic domain with characteristics in the image domain. It uses the Mann-Whitney-Wilcoxon rank-sum test [1] to assess whether a characteristic from images annotated with a given keyword differ significantly from images without that keyword. The test is robust to outliers and efficiently scales to very large databases.

The effectiveness of the proposed statistical framework is demonstrated using three different color imaging applications: 1) semantic image enhancement: automatically re-render an image and adapt it to its semantic context, 2) color naming: determine the color triplet for a given color name, and 3) color palette extraction: extract the most suitable palette of five harmonic colors given a semantic expression.

Semantic image enhancement is implemented for both color and depth-of-field adjustments. Colors and focus are adapted in order to strengthen an associated semantic context. Example keywords illustrated are *autumn, strawberry, macro,* or *flower*. An important aspect of semantic image enhancement is that the enhanced image depends not only on the image content, but also on the associated keyword. Consequently, one image can have different enhanced versions for different keywords, e.g., *gold* keyword would lead to the emphasized golden color of the setting sun in the mountains, while *winter* keyword results in brightened image with a more salient snow on the mountains. Essentially, it means that beauty requires context.

The authors also demonstrate that the proposed statistical framework can be used for color naming. By taking 950 color names from XKCD color survey[1] (they also translated color names in 10 European and Asian languages) and about a million images harvested from Flickr and Google image search, the authors used the framework to estimate the color values that correspond to the names, essentially building a multi-language color thesaurus: www.colorthesaurus.com.

The goal for automated creation of color palettes is to choose their five colors in the way that best describes a given arbitrary keyword in terms of color and at the same time to respect harmonic principles common to the artistic community. For that purpose, the authors applied the framework to about 100'000 commonly used words and associated images found in Google image search. Then, using the computed histograms, which correspond to every commonly used word, the palettes are derived based on the five harmonic templates [2]. Created palettes can be explored at the following webpage: www.koloro.org.

The authors evaluated the performance of the proposed framework by comparing it with five popular statistical tests using color naming application as an example: Kolmogorov-Smirnov [3], Student's t-test [4], Earth mover's distance [5], Hodges-Lehmann estimator [6], and Chi-square test [7]. The results showed Mann-Whitney-Wilcoxon test to be optimal in terms of computational complexity and accuracy.
In summary, the paper takes an interesting look at the lesser common problem of finding visual representation for a given meaning (typically, researchers try to

[1] http://blog.xkcd.com/2010/05/03/color-survey-re-sults/

derive the meaning from a media content) and proposes a general enough statistical framework that can help solving this problem. Three interesting applications of this framework are presented: image enhancement, color naming, and palette construction, with latter two available online.

References:

[1] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," Ann. Math. Statist. vol. 18, no. 1, pp. 50–60, 1947.

[2] P. O'Donovan, A. Agarwala, and A. Hertzmann, "Color compatibility from large datasets," in Proc. ACM SIGGRAPH, no. 63, 2011.

[3] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions," Ann. Math. Statist., vol. 19, no. 2, pp. 279–281, 1948.

[4] Student, "The probable error of a mean," Biometrika, vol. 6, no. 1, pp. 1–25, 1908.

[5] S. Shirdhonkar and D. W. Jacobs, "Approximate earth mover's distance in linear time," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun 2008, pp. 1-8.

[6] J. L. Hodges and E. L. Lehmann, "Estimates of location based on rank tests," Ann. Math. Statist., vol. 34, no. 2, pp. 598–611, 1963.

[7] R. L. Plackett, "Karl Pearson and the chi-squared test," Int. Statist. Rev., vol. 51, no. 1, pp. 59-72, 1983.

Pavel Korshunov is a postdoctoral researcher in Multimedia Signal Processing Group at EPFL. He received his Ph.D. in Computer Science from National University of Singapore (NUS). He is a recipient of ACM TOMM Nicolas D. Georganas Best Paper Award in 2011, two top 10% best paper awards in MMSP 2014, and top 10% best paper award in ICIP 2014. He has over 50 research publications and is a co-editor of the new JPEG XT standard for HDR images. His research interests include computer vision and video analysis, video streaming, video and image quality assessment, crowdsourcing, high dynamic range imaging, ultra-high definition imaging, focus of attention, visual privacy protection mechanisms and their evaluation.

IEEE COMSOC MMTC R-Letter

Salient Object Detection in Complex Scenes: An Annotated Dataset and Model

A short review for: "What is a Salient Object? A Dataset and a Baseline Model for Salient Object Detection" (Edited by Bruno Macchiavello)

Borji, A., "What is a Salient Object? A Dataset and a Baseline Model for Salient Object Detection," IEEE Transactions on Image Processing, vol.24, no.2, pp.742-756, Feb. 2015.

One very important characteristic of the human visual system is the ability to accurately prioritize noticeable objects in a scene. This is known as salient object detection and segmentation. Recently, this fundamental problem has attracted a great deal of interest in the computer vision community. There are various applications for salient object detection, including object detection and recognition [1], image and video compression [2], video summarization [3], image cropping [4], photo collage [5], image segmentation [6], human-robot interaction [7], and so on.

A large number of saliency detection methods have been proposed in the past years [8]. However, most of the previously proposed models have focused they effort on segmentation of a salient object on scenes with a single object. It is unclear how those models perform on more complex cluttered scenes with several objects. There are several reason why salient object detection in more complex scenes has not been studied to the same extent. Two major reasons are: *(i)* the lack of a suitable benchmark dataset for scaling up models and model development and *(ii)* the lack of a widely-agreed objective definition of the most salient object.

In this work the author provides a less biased benchmark dataset containing scenes with multiple objects (by modifying a previously proposed dataset). A model based on superpixels for salient object segmentation is also proposed. This model aims to be a baseline for model evaluation and comparison.

In a previous work, of the same authors [9], 70 students were asked to draw a polygon around the object that stood out the most over 120 images. The degree to which annotations of participants agree with each other was measured. It was verified that images with several foreground objects showed a low value of agreement while images with highest annotations agreement had often one visually distinct salient object. Therefore, the difficulty of an automatic salient object detection method increases in relation to the complexity of the scene. In that work the authors also compare the relationship between the annotated images and freeviewing fixations. Results appear to indicate that the most salient object in a scene is the one that attracts the majority of fixations. This conclusion has been presented previously by similar works [10].

In this work in order to provide a more challenging dataset, the author decided to annotate scenes of the dataset by Judd et. al. [11]. The reason for choosing this dataset is because it is currently the most popular dataset for benchmarking fixations prediction models. The author discarded images without well-defined objects (e.g. mosaic, tiles, etc.) or images with very cluttered backgrounds. Two observers were asked to manually outline objects, following some basic rules. Five other observers were asked to choose the best segmentations between the two annotations for each image. This annotated dataset is referred as Judd-A. Then, a quantitative analysis of the relationship between fixations and annotations was performed. It was observed that in about 55% of the images the most salient object attracts more that 50% of fixations. The author also compare the Judd-A dataset with other three datasets, and verified that Judd-A was less center-biased than the others, in terms of number of fixations. The complexity of the scenes between all four datasets was also analyze. The author used a graph-based superpixel segmentation algorithm in order to segment a image into contiguous regions larger than 60 pixels each. The basic idea is that the more superpixels an image contains, the more complex and cluttered it is. It was verified that the most salient object in Judd-A dataset on average contains more superpixels than salient objects in the other datasets, even with smaller objects.

Besides this more complex dataset, the author propose a straightforward model to serve two

purposes: *(i)* to assess the degree to which the propose Judd-A dataset can be explained by a simple model and *(ii)* to gauge progress and performance of the state-of-the-art models. By comparing performance of the best models relative to this baseline model, over existing datasets and the Judd-A, it can be judge how powerful and scalable these models are. The propose model involves two stages. In the first, two fixation prediction models are used to find spatial outliers in scenes that attract human eye movements and visual attention, the models used were previously reported elsewhere. These salient regions are then fed, as a saliency map, to the segmentation component in the next step. In the second step, the image is segmented using the same graph-based superpixel segmentation algorithm. The saliency map is normalized and then thresholded. Then all unique image superpixels that spatially overlap with the truncated saliency map are included in the final segmentation. The holes inside the selected regions are considered as part of the salient object.

This model with 8 other state-of-the-art models were test in three different datasets, including Judd-A. The results showed a drop of performance of all models in the Judd-A dataset due to the fact of more complex scenes. The propose model outperformed others in the Judd-A dataset and was among the best in all cases. The number of fixations was used as ground truth for most salient regions.

In conclusion the proposed modified dataset can allow a more elaborate analysis of the interplay between saliency detection and fixation prediction. The more complex scenes are more challenging and can help develop better salient object detections algorithms. However, as mentioned in this work, human observers have low agreement on which objects are more salient in scenes with several foreground objects. Therefore, an objective definition of the most salient object in complex scenes is still and open issue.

References:

[1] D. Walther and C. Koch. Modeling attention to salient proto-objects. Neural Networks, 19(9):1395–1407, 2006.

[2] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," IEEE Trans. Image Process., vol. 19, no. 1, pp. 185–198, Jan. 2010.

[3] P. Bodesheim, "Spectral clustering of ROIs for object discovery," in Proc. 33rd DAGM-Symp., LNCS 6835. 2011, pp. 450–455.

[4] L. Marchesotti, C. Cifarelli, and G. Csurka. A framework for visual saliency detection with applications to image thumbnailing. In ICCV, pages 2232–2239, 2009

[5] S. Goferman, A. Tal, and L. Zelnik-Manor, "Puzzle-like collage," Comput. Graph. Forum, vol. 29, no. 2, pp. 459–468, 2010.

[6] M. Donoser, M. Urschler, M. Hirzer, and H. Bischof, "Saliency driven total variation segmentation," in Proc. IEEE 12th Int. Conf. Comput. Vis. (ICCV), Sep./Oct. 2009, pp. 817–824.

[7] D. Meger et al., "Curious George: An attentive semantic robot," Robot. Auto. Syst., vol. 56, no. 6, pp. 503–511, 2008.

[8] T. Liu, J. Sun, N.-N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2007, pp. 1–8.

[9] A. Borji, D. N. Sihite, and L. Itti, "What stands out in a scene? A study of human explicit saliency judgment," Vis. Res., vol. 91, pp. 62–77, Oct. 2013.

[10] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille, "The secrets of salient object segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 280–287.

[11] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in Proc. IEEE 12th Int. Conf. Comput. Vis., Sep./Oct. 2009, pp. 2106–2113.

Bruno Macchiavello is an assistant professor at the Department of Computer Science of the University of Brasilia (UnB), Brazil. He received his B. Eng. degree in the Pontifical Catholic University of Peru in 2001, and the M. Sc. and D.Sc. degrees in electrical engineering from the University of Brasilia in 2004 and 2009, respectively. Prior to his current position he helped develop a database system for the Ministry of Transport and Communications in Peru. He also was co-organizer of a special session on Streaming of 3D content in the 19th International Packet Video Workshop (PV2012). His main research interests include video and image coding, image segmentation, distributed video and source coding, multi-view and 3D video processing.

Video QoS Control in Distributed OpenFlow Networks

A short review for "Distributed QoS Architectures for Multimedia Streaming Over Software Defined Networks" (Edited by Frank Hartung)

Hilmi E. Egilmez, and A. Murat Tekalp, "Distributed QoS Architectures for Multimedia Streaming Over Software Defined Networks", IEEE Transactions on Multimedia, Vol. 16, No. 6, pp. 1597-1609, October 2014

The traditional Internet, as we know it, works. One of the reasons is its highly distributed and local control plane. However, this concept can be optimized. A major development in that respect is the advancement of the Software Defined Networking (SDN) concept and its most prominent representative, the OpenFlow framework [1]. In SDN and OpenFlow, the control and forwarding layers are decoupled. Compared to traditional routers, routing control is somewhat more concentrated in *controllers*, while data forwarding remains in the routers, now also called *forwarders*. Controllers and forwarders communicate via the OpenFlow protocol. This concept allows to dynamically change routing on a per-flow basis. Some vendors have started to produce OpenFlow-enabled switches and routers.

In their prior publications [2][3][4][5], the authors have already applied OpenFlow to dynamic QoS routing for scalable video streaming. However, they have considered a network controlled by a single OpenFlow controller. In real applications, this assumption is not realistic. Rather, one would envision connected multi-domain and multi-operator SDNs. In the present publication, the authors thus extend their work to that scenario.

In detail, the paper describes the following new contributions in the context of OpenFlow: (i) information gathering, in the form of topology aggregation and link summarization methods to efficiently acquire network topology and state information, (ii) utilization of this information, in the form of a general optimization framework for flow-based end-to-end QoS provision over multi-domain networks, and (iii) two distributed control plane designs utilizing controller-to-controller messaging for scalable and secure inter-domain QoS routing. The proposed extensions are then analyzed with respect to video streaming quality, cost, and overhead.

The authors suggest an architecture where multiple controllers communicate via the controller-controller-interface and share inter-domain routing information. Each controller fully manages its domain, including the functions of topology management, resource management, route calculations, flow management, queue management, call admission, and traffic policing. The authors propose that controllers have a complete view of their domain, but provide only a condensed view to the outside, i.e., other controllers. This condensed view only describes the externally visible topology and QoS parameters of a network, for example its border nodes and their (virtual) links, rather than the actual mesh of links within the domain. Given that information, controllers can do inter-domain routing. In each domain, the respective controller can do intra-domain routing.

Using the mentioned controller-controller-interface, controllers regularly pull QoS information from neighboring controllers, and push information in the case of sudden changes.

The authors propose two control plane designs: one in which all controllers are fully distributed equal peers, and one in which controllers are hierarchically distributed, i.e., super-controllers are responsible for inter-domain routing, and each controller for intra-domain routing in its own domain.

Another major contribution of the paper is a proposal for distributed QoS routing. First, as an ingredient, the authors analyze dynamic (centralized) QoS routing. For each route, a cost function, based on packet loss and delay variation, is established. The route which minimizes the cost function, subject to a maximum total delay, is identified using a Lagrangian relaxation based aggregated cost (LARAC) algorithm. For multiple flows, the LARAC algorithm is applied successively, starting from the flows with highest priority, and with cost parameters updated after each flow.

In order to extend that to distributed QoS routing, a distributed optimization framework is proposed

where QoS parameters (i.e., inter-domain routing information) are mapped on aggregated networks. The distributed QoS routing problem is separated into two problems: inter-domain routing, i.e., routing over the virtual aggregated networks, followed by intra-domain routing. The authors also propose solutions to these problems, in the form of three algorithms. Two of the algorithms apply to inter-domain routing in the super controllers and intra-domain routing in the regular controllers of the hierarchically distributed control plane design, respectively, while the third algorithm applies to inter-domain and intra-domain routing in all controllers of the fully distributed control plane design. All algorithm descriptions include the routing decision logic and the message passing routines for the controller-controller-interface.

The authors complement their paper by applying and analyzing their control framework for the case of scalable H.264 SVC based video streaming over unreliable UDP transport. The base layer is streamed using QoS routing with the proposed mechanisms, while enhancement layers are streamed using best-effort routing. The video is streamed across a simulated 180-node network in 6 domains and with heavy other data traffic congesting the network. The results show that the proposed QoS routing mechanisms work significantly better than best-effort routing. The fully distributed control plane design outperforms the hierarchically distributed control plane design, especially for controllers with spatial distances > 5000 km. The authors conclude that a fully distributed control plane is more suitable for service providers that serve multimedia across continents using SDNs, while the hierarchical control plane can be a better option for in-nation service providers, especially when the number of domains is large.

The results of the paper (and the previous work by the authors) gives valuable insights into the large-scale QoS control design of OpenFlow networks, especially for video transport. Since video is the main source of traffic in todays Internet, this is an important contribution to the advancement of OpenFlow and related SDN concepts. The paper is also well written.

Related work has been presented in [6]. However, a major difference is that [6] considers properties of the transmitted video streams, and aims at optimizing user's QoE, rather than "just" QoS, as in the present publication. A possible future extension of the publication discussed here could thus be an extension to QoE, rather than QoS, optimization.

References:

[1] N. McKeown, et al., "OpenFlow: Enabling innovation in campus networks," SIGCOMM Comput. Commun. Rev., vol. 38, no. 2, pp. 69–74, Mar. 2008.

[2] H. E. Egilmez, et al., "Scalable video streaming over OpenFlow networks: An optimization framework for QoS routing," in Proc. 18th IEEE Int. Conf. Image Process., Sep. 2011, pp. 2241–2244.

[3] H. E. Egilmez, et al., "OpenQoS: An OpenFlow controller design for multimedia delivery with end-to-end quality of service over software-defined networks," in 2012 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC), Dec. 2012, pp. 1–8.

[4] H. Egilmez, et al., "An optimization framework for QoS-enabled adaptive video streaming over OpenFlow networks," IEEE Trans. Multimedia, vol. 15, no. 3, pp. 710–715, Apr. 2013

[5] H. E. Egilmez, et al., "A distributed QoS routing architecture for scalable video streaming over multi-domain OpenFlow networks," in Proc. 19th IEEE Int. Conf. Image Process., Sep./Oct. 2012, pp. 2237–2240.

[6] P. Georgopoulos et al., "Towards Network-wide QoE Fairness Using OpenFlow-assisted Adaptive Video Streaming", Proceedings of the 2013 ACM SIGCOMM workshop on Future human-centric multimedia networking. ACM, 2013.

Frank Hartung is a full professor of multimedia technology at FH Aachen University of Applied Sciences, Aachen, Germany. He received a MSc in electrical engineering from RWTH Aachen University, Germany, and a PhD in Telecommunications from University of Erlangen, Germany. He has been working with Ericsson Research, as a research team leader in Multimedia Technologies, from 1999 to 2011. His research interests include media security, networked multimedia, immersive multimedia communication, streaming, and mobile video. He has authored or co-authored more than 50 publications in this domain, and is the co-inventor of 16 granted patents. Dr. Hartung is a member of IEEE, VDE and ITG, and was in 2003-2004 serving as chairman of the German IEEE Signal Processing Chapter.

Delay-Aware Wi-Fi Offloading

A short review for "DAWN: Delay-Aware Wi-Fi Offloading and Network Selection"
(Edited by Lifeng Sun)

Man Hon Cheung and Jianwei Huang, "DAWN: Delay-Aware Wi-Fi Offloading and Network Selection," IEEE Journal on Selected Areas in Communications (Special Issue on Recent Advances in Heterogeneous Cellular Networks), 2015.

Mobile video applications, such as YouTube, constitute a significant proportion of the global mobile data traffic. According to Cisco's forecast, mobile data traffic will grow to 24.3 exabytes per month by 2019, an increase by nearly 10-fold between 2014 and 2019 globally, which 71.6% of the traffic will be due to video [1]. The huge amount of mobile data traffic puts a lot of pressure on the operators' cellular networks. On the other hand, traditional network expansion methods, such as acquiring more spectrum and upgrading to more advanced communication technologies (e.g., LTE-A), are costly and time-consuming. As a result, it is likely that the mobile traffic demand will exceed the network capacity in the short to medium term. An efficient way to increase the network capacity in a cost-effective and timely manner is to use complementary technologies, such as Wi-Fi or small cells, to offload the traffic originally targeted towards the cellular network. In fact, Cisco estimated that 54% of the data traffic from the mobile devices will be offloaded by 2019 [1].

For most of the mobile video applications, they are delay-tolerant in nature, where a user can tolerate delays ranging from several minutes to several hours without having a significant satisfaction loss. For example, the survey in [2] reported that more than half of the respondents are willing to wait for 10 minutes to stream YouTube videos and 3-5 hours to download a file when a monetary incentive is given.

Taking into account this characteristic of delay-tolerant applications, the authors considered the user-initiated Wi-Fi offloading problem, where a user aims to minimize its total data usage payment under usage-based pricing, while taking into account the deadline of its application. Previous works on user-initiated Wi-Fi offloading policy mainly focused on reducing the cellular data usage without paying too much attention to the quality of service (QoS) of the user's application. As an example, under the on-the-spot offloading (OTSO) scheme that most smartphones are using by default, a user offloads its data traffic to a Wi-Fi network whenever possible. However, the authors suggested that it is not always desirable to use the OTSO scheme, especially when the Wi-Fi network is highly loaded and the deadline is tight. More specifically, when the deadline is short, it would be more preferable to use a cellular network with a higher data rate to complete the file transfer with a cost than to use the free but congested Wi-Fi. On the other hand, when the deadline is long, it would be more preferable to wait for a Wi-Fi hotspot to reduce the cellular usage price than to impatiently use the cellular network immediately as in the OTSO scheme. However, in general, it is challenging to achieve a good balance between the total payment and the QoS when taking various factors such as network conditions and delay deadlines into consideration.

To capture this design tradeoff analytically, the authors considered a general user offloading scenario, and formulated the delay-aware Wi-Fi offloading problem as a finite-horizon sequential decision problem. They proposed a general Delay- Aware Wi-Fi Offloading and Network Selection (DAWN) algorithm, which tradeoffs between the total payment and the QoS. To reduce the complexity of the proposed algorithm, the authors derived sufficient conditions under which the optimal policy exhibits threshold structures, which leads to the design of the monotone DAWN algorithm with a much lower computational complexity and is easier to implement. To the best of the authors' knowledge, this is the first paper that studies offloading algorithm design analytically, which tradeoffs a user's payment and QoS.

To summarize, the main contributions of their paper are:

- Optimal user-initiated offloading algorithm: The authors proposed a general DAWN algorithm for delay-tolerant applications that achieves a good tradeoff between total data usage payment and the user's QoS.

- Low-complexity approximation offloading algorithm: The authors proposed a monotone approximation DAWN algorithm with a much lower computational complexity by exploring the

threshold structure of the optimal policy.

- Insights on offloading decisions: Contrary to the belief that Wi-Fi offloading is always preferable at any time and location, with a deadline consideration, the authors showed that it may not be a good idea for a user to offload its data traffic under a tight deadline constraint and a congested Wi-Fi network.

Acknowledgement:

The R-Letter Editorial Board thanks the authors of the paper for providing a summary of its contributions.

References:

[1] Cisco Systems, "Cisco visual networking index: Global mobile data traffic forecast update, 2014-2019," White Paper, Feb. 2015.

[2] Juniper Research, "Mobile data offload & onload: Wi-Fi, small cell & carrier-grade strategies 2013-2017," Report, Apr. 2013.

[3] S. Sen, C. Joe-Wong, S. Ha, J. Bawa, and M. Chiang, "When the price is right: Enabling time-dependent pricing of broadband data," in Proc.of ACM SIGCHI, Paris, France, Apr. 2013.

[4] A. J. Nicholson and B. D. Noble, "BreadCrumbs: Forecasting mobile connectivity," in Proc. of ACM MobiCom, San Francisco, CA, Sept. 2008.

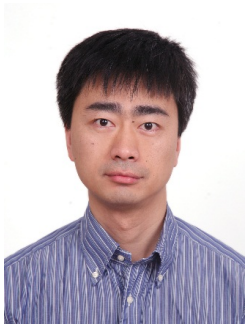
[5] S. Gambs, M. Killijian, and M. N. del Prado Cortez, "Next place prediction using mobility Markov chains," in Proc. of ACM MPM, Bern, Switzerland, Apr. 2012.

[6] T. M. T. Do and D. Gatica-Perez. Contextual conditional models for smartphone-based human mobility prediction. In UbiComp 2012, pages 163{172. ACM, 2012.

[7] J. Manweiler, N. Santhapuri, R. R. Choudhury, and S. Nelakuditi. Predicting length of stay at wi hotspots. In INFOCOM, 2013 Proceedings IEEE, pages 3102{3110. IEEE, 2013.

[8] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti. Wherenext: a location predictor on trajectory pattern mining. In ACM SIGKDD, pages 637{646. ACM, 2009.

[9] K. Lee, I. Rhee, J. Lee, S. Chong, and Y. Yi, "Mobile data offloading: How much can WiFi deliver?" in Proc. of ACM CoNEXT, Philadelphia, PA, Nov. 2010.



Lifeng Sun received the B.S. and Ph.D. degrees in System Engineering from National University, of Defense Technology China, in 1995 and 2000, respectively. He joined the Department of Computer Science and Technology (CST), Tsinghua University (THU), Beijing, China, in 2001. Currently, he is a Professor in CST of THU. His research interests include networked multimedia, video streaming, 3D/multi-view video coding, multimedia content analysis, multimedia cloud computing and social media. He has authored or coauthored about 80 high quality papers, one chapter of a book. He got the Annual Best Paper Award of IEEE TCSVT (2010), Best Paper Award of ACM Multimedia (2012) and Best Student Paper Award of Multimedia Modeling (2015).

Unreeling Netflix

A review for "Unreeling Netflix: Understanding and Improving Multi-CDN Movie Delivery"
(Edited by Michael Zink)

Vijay Kumar Adhikari, Yang Guo, Fang Hao, Matteo Varvello, Volker Hilt, Moritz Steiner and Zhi-Li Zhang, "Unreeling netflix: Understanding and improving multi-CDN movie delivery," INFOCOM, 2012 Proceedings IEEE , vol., no., pp.1620,1628, 25-30 March 2012

In this paper, the authors perform a measurement study of the online, over-the-top (OTT), video streaming service Netflix. At the time when this study was performed Netflix was the leading on-demand Internet video streaming provider in the US and Canada. In the US alone Netflix data accounted for 29.7% of the peak downstream traffic (single largest source of Internet traffic). Since then Netflix has expanded its service into many other countries (e.g., UK, Austria, Germany, etc.) and its share of peak downstream traffic in the US has increased to XX%. Thus, one can safely say that Netflix is currently the largest on-demand Internet video provider worldwide.

The goal of the work of this paper is to uncover Netflix' architecture and service strategy. Without giving away further interesting details on architecture and service strategy the authors uncover in their study, the most surprising outcome of this work is the realization that Netflix provides this large-scale service with virtually no own infrastructure. The results provided in this paper can be seen as a "blue print" for an architecture for a scalable, infrastructure-less content provider.

The initial analysis of the Netflix streaming platform reveals that the architecture consists of four main components, a Netflix owned data center, Amazon web services (AWS), CDNs (in this case Limelight, Level-3, and Akamai), and the Silverlight player. In addition, Netflix uses DASH for available bitrate streaming. This information about Netflix' architecture is obtained through traffic monitoring of a complete streaming session (starting from a new user registration) and subsequent DNS and WHOIS analysis of contacted servers and their respective IP addresses. After identifying the major components of the architecture the authors dissect the interaction between a client and the different servers involved in the streaming process. This process starts with the download of the Silverlight player for each streaming session. Next, the Silverlight player downloads a client-specific manifest file. This file contains a list of CDNs that can serve the requested content, the location of the trickplay data, video/audio chunk URLs for different quality levels, and timing parameters (e.g., timing and polling interval). To support simple trickplay, thumbnails for periodic snapshots are downloaded. The manifest file contains multiple audio and video quality levels with respective URLs for individual CDNs. For the actual streaming process 4-second chunks are downloaded. Downloads of these chunks are more frequent in the beginning of the streaming session to quickly fill the player buffer. After that, downloads show a more periodic pattern with 4-second intervals which corresponds to the video segment length. During the streaming phase, the player sends back periodic reports to a statistics server in AWS.

The authors also conduct a large-scale analysis to understand how geographic locations, client capabilities, and content type impact streaming parameters. To achieve this goal a set of manifest files from a combination of 25 movies, six accounts, four computers (Windows/Mac), and four different locations are analyzed. For obtaining additional manifest files from more geographically diverse locations the authors make use of Squid proxy servers which were installed at different PlanetLab nodes. The analysis of this large set of manifest files reveals the following information:

- The CDN ranking is only based on user account and stays the same independent of movie, computer type, time, and location.
- For identical parameters different users may see different CDN rankings.
- The CDN ranking stays unchanged for several days.

Based on these, somewhat surprising finding, the authors further investigate Netflix' CDN selection strategy and produce the most surprising result of this paper. For this investigation a movie is played from a single client and the bandwidth is throttled for the top-ranked CDN (with the aid of dummynet) in 100Kbps increments down to 100Kbps. Surprisingly, the client sticks with this CDN until streaming the video at even the lowest quality is not feasible, and

only then switches to the next, lower ranked CDN. The same procedure then repeats if bandwidth throttling is applied for the second ranked CDN.

In addition to the single-client measurements described above, the authors also analyze the overall performance of the CDNs that are used for the actual streaming of video content. To perform this analysis the bandwidth between end user locations and CDN servers is measured through the "replay" of get requests from these locations. The end user locations are comprised by a set of residential sites and PlanetLab nodes. The main highlights of the analysis of the data obtained from these measurements are that i) all CDNs perform relatively equal; ii) the last mile is still the bottleneck on the path between client and server; iii) at a few residential sites one CDN provider performs much better than the others. An additional analysis of the average bandwidth over a 24-hour period reveals that it varies significantly. These results indicate that a better CDN selection strategy could be chosen than the one currently applied by Netflix.

The final contribution made by this paper is the study of alternative video delivery strategies. Here, the authors try to answer the question how the shortcomings that have been identified through their measurement study can be improved and how much improvement these changes can theoretically bring? For this analysis the theoretical upper bound average bandwidth is determined. The results show that always assigning the top CDN to a user is only 6% worse than the theoretical optimal case and the optimal case is between 17% and 33% better than the average case. To identify the best performing CDN, the authors propose an approach where the player conducts instantaneous bandwidth measurements at the beginning of a streaming session. Finally, streaming in parallel from multiple CDNs can increase the overall available bandwidth by up to 70%.

Michael Zink is currently Assistant Professor in the Electrical and Computer Engineering Department at the University of Massachusetts in Amherst. Previously, he was a Research Assistant Professor in the Computer Science Department at the University of Massachusetts in Amherst. He received his PhD in 2003 from the Multimedia Communications Laboratory at Darmstadt University of Technology. He works in the fields of sense-and-response sensor networks, distribution of high-bandwidth, high-volume data, and the design and analysis of long-distance wireless networks and Systems Engineering. Further research interests are in wide-area multimedia distribution for wired and wireless environments and network protocols. He is one of the developers of the KOMSSYS streaming platform. He received his Diploma (M.Sc.) from Darmstadt University of Technology in 1997. From 1997 to 1998 he was employed as a guest researcher at the National Institutes of Standards and Technology (NIST), where he developed an MPLS testbed. In 2003 he received his Ph.D. degree (Dr.-Ing) from Darmstadt University of Technology; his thesis was on Scalable Internet Video-on-Demand Systems. Dr. Zink is a senior member of the IEEE. He is the associate editor for the ACM/Springer Journal on Multimedia Systems. He has served on the technical program committees of several professional conferences, including IEEE Infocom, ACM Multimedia, and ACM Multimedia Systems.

IEEE COMSOC MMTC R-Letter

Correction of Depth Compression for Planar Scenes

A short review for "Anahita: A System for 3D Video Streaming with Depth Customization"
(Edited by Carsten Griwodz)

Calagari, K. Templin,K., Elgamal, T., Diab, K., Didyk, P., Matusik, W., Hefeeda, M., "Anahita: A System for 3D Video Streaming with Depth Customization,", Proc. of ACM Multimedia 2014, pp. 337-346.

3D video has become an artistic medium frequently applied in feature films that aim at the movie market, and the standardization of formats by the DVB consortium has also led to a flood of 3D TV sets that are deployed in homes. Audiences can also enjoy 3D video on mobile phones or with the help of a variety of glasses.

The paper by Calagari et al. introduces the readers to a particular challenge that has appeared in this new 3D environment. The authors observe that the variety of displays, as well as the variety of viewing conditions when users enjoy stereoscopic video, leads to highly inconsistent user experiences. The authors distinguish two main reasons: first, the inability of device manufacturers and content creators to know the viewing conditions of auto-stereoscopic devices in audiences' homes, and second, the diversity of users themselves.

In this discussion, it may be understood implicitly, but is not spelled out, that the problem of unpredictable 3D-impressions is worse for TV and mobile phone displays that provide auto-stereoscopic 3D than for systems that rely on 3D glasses (such as such shutter glasses). Many auto-stereoscopic displays are tuned to provide an ideal 3D impression at a sweet spot that is located at a location that the device manufacturer expects to be the most common viewing angle and distance for their product, viewers outside this sweet spot may experience stronger or weaker depth impressions, or even skew if their viewing angle is too steep. It is stated explicitly in the paper that 3D content, once prepared, is generally only adapted to the format of the wide range of presentation devices that are in use, but that the depth encoding remain unchanged for all of these devices.

Calagari et al. point also out that there is a wide variance among the population in the way in which stereoscopic depth is perceived [1], [2]. An interesting piece of information that was provided by the authors during the presentation of the paper at ACM MM 2015 was that actually stereoscopically coded videos tend to be encoded with a compressed depth effect, due to the fact that users experience nausea if depth is perceived as unnaturally strong, while a weak depth effect can be ignored.

Based on these insights, the authors go about to develop an approach for automatically adjusting the depth information of 3D videos before it is presented on a particular device. In several user studies, Calagari et al. adapt this depth information not only to the specific device; they have also found that optimal depth adjustment differs by content. Interestingly, they have found relevant differences in optimal depth adjustment in spite of a fairly narrow range of contents, where all were outdoor arena sports.

The authors' system for adjusting depth information to specific devices, content, and (potentially) user perceptions, comprises a new way of encoding 3D data and the actual adjustment step. The delivery system is inspired by DASH, but delivers not only a variety of video qualities. The authors prepare a considerably larger number of videos to address the variety of DVB-compliant 3D video delivery systems, including multiview coding (MVC), side-by-side, frame-sequential formats, and more [3][4][5]. To account for this diversity and delivery appropriate formats, they create a tree of formats, and qualities within each of these formats, to deliver to each device in an appropriate DASH-like manner.

The contributions that seem of more long-term relevance to me, however, is the depth expansion and compression method proposed by the authors, and the user study that they performed to support their results. In their paper, depth expansion refers to the modification of existing 3D material to give an impression of greater depth than before the manipulation, depth compression refers to the opposite. Any depth compression method that accounts for an entire scene structure, and tries to re-create a correct view from a closer viewpoint, is hindered by an effect of perspective correction close to the viewing axis: the angular difference between left and right view increases, and edge pixels that were hidden at a smaller angle become visible. Depending on the distortion that is performed to compress depth, this may affect one of both

views. Calagari et al., take an entirely different approach, which is bound to sacrifice pixel ratio accuracy but avoids the problem of exposing hidden 3D elements.

The method works exclusively for mostly planar scenes, and is specifically designed for long shots in case of football, soccer, tennis, and similarly shaped scenes. It works adding a moderate amount of stretch and slant to each view of a frame. The authors acknowledge that this is a limited technique, and present a method for detecting long shots in 3D content that can potentially consist of such video material mixed with views from different angles that cannot be handled by the simplified technique.

Assuming that a scene is mostly planar, and the original scene's 3D encoding parameters are known, it is possible to apply stretch in parallel to the original horizon, where higher distances are stretched linearly more (depth compression) or linearly less (depth expansion) with increasing original distance value. With smaller values, the focal length will appear reduced and extended, respectively, but at higher values, pixel ratios for objects sticking out of the plane (such as players) will affect the quality of experience. Slant is added in opposite directions to both views to ensure that both eyes maintain a consistent impression of vanishing point for all lines parallel to the main plane in both views.

The design is obviously limited in its applicability, as well as the strength of the effect that can be achieved, but it provides a very efficient and fast means of adapting 3D depth impression in such planar scenarios.

The authors have evaluated the performance of their approach in a series of user studies. Although the data analysis should have been a bit deeper, stating also the relevance of the experimental findings rather than mostly average values and standard deviation, the studies are performed and described in a convincing manner and well enough for a future replication of the setup.

The study is divided into two parts: the first experiment was performed with a limited number of students to identify appropriate range limits for devices ranging from mobile phone to TV setup. The second experiment tested the entire system over the network, extended the number of users to 15 and the set of devices to five. The results document a consistent improvement of the depth-corrected views, but are also important documentation of fluctuations in depth perception.

I recommend that paper for its elegant presentation of the depth perception problem on today's variety of devices, and the innovative, efficient approach to depth correction in a particular set of cases. The authors are open about the limitations of their technique, and conducted quite thorough user studies in a field where objective quality is still a very hot research field [6].

References:

[1] B. E. Coutant and G. Westheimer, "Population distribution of stereoscopic ability.," Ophthalmic Physiol. Opt., vol. 13, no. 1, pp. 3–7, 1993.

[2] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel, "A perceptual model for disparity," ACM Transactions on Graphics, vol. 30, no. 4. p. 1, 2011.

[3] K. Diab, T. Elgamal, K. Calagari, and M. Hefeeda, "Storage optimization for 3D streaming systems," in MMSys 2014, 2014, pp. 59–69.

[4] EBU and DVB, "DVB Plano-stereoscopic 3DTV Part 3: HDTV Service Compatible Plano-stereoscopic 3DTV," 2012.

[5] EBU and DVB, "DVB Plano-stereoscopic 3DTV Part 2: Frame Compatible Plano-stereoscopic 3DTV," 2012.

[6] F. Battisti, E. Bosc, M. Carli, P. Le Callet, and S. Perugia, "Objective image quality assessment of 3D synthesized views," Signal Process. Image Commun., vol. 30, pp. 78–88, 2015.

Carsten Griwodz is chief research scientist in the Media Department at the Norwegian research company Simula Research Laboratory AS, Norway, and professor at the University of Oslo. His research interest is the performance of multimedia systems. He is concerned with streaming media, which includes all kinds of media that are transported over the Internet with a temporal demands, including stored and live video as well as games and immersive systems. To achieve this, he wants to advance operating system and protocol support, parallel processing and the understanding of the human experience. He was area chair and demo chair of ACM MM 2014, and general chair of ACM MMSys and NOSSDAV (2013), co-chair of ACM/IEEE NetGames (2011), NOSSDAV (2008), SPIE/ACM MMCN (2007) and SPIE MMCN (2006), TPC chair ACM MMSys (2012), and systems track chair ACM MM (2008). He is currently editor-in-chief of the ACM SIGMM Records. More information can be found at http://mpg.ndlab.net

IEEE COMSOC MMTC R-Letter

Transmission of Video Chat over Wireless Systems

A short review for "Rate and Power Allocation for Joint Coding and Transmission in Wireless Video Chat Applications" (Edited by Carl James Debono)

S.-P. Chuah, T.-P. Tan, and Z. Chen, "Rate and Power Allocation for Joint Coding and Transmission in Wireless Video Chat Applications," IEEE Transactions on Multimedia, vol. 17, no. 5, pp. 687-699, May 2015.

Video chat calls over Internet are a popular application especially in wired networks. However, when it comes to make such calls over wireless systems the quality may suffer due to the lack of resources available. Transmission of video chat over wireless demands an uplink and a downlink channel both carrying video and voice. This continuous stream of data in both directions consumes a lot of power from the limited supply available at the mobile device. Moreover, the base station has limited resources which must be shared between all the users that it is serving.

Advances in mobile device technology, advanced video coding standards, and increased bandwidths in wireless communication systems allow good quality transmission of video content. One possible application is video chat where users send video together with voice. The complex computations needed during video coding and the high bit rates required for high quality video consume a lot of power leading to quick exhaustion of battery power [1]. Video chat demands bi-direction data traffic between the communicating parties which must be relayed through base stations. The rate and power allocated to the user's video content and transmission not only impact the device's power consumption and network usage but also the receiver's Quality of Experience (QoE). From the network provider's perspective, energy efficiency and interference control are important parameters and thus ideally the power budget per user is minimized. In traditional multi-user video systems, the network resources are allocated by a network manager, such that the overall network utilization is maximized [2, 3]. This can result in large costs for the user, who typically seeks more advantageous pay rates.

The authors of the original paper look into this problem as a Stackelberg game [4] with the base station being the leader. Video chat is a high bit rate service and dynamic pricing based on use and resources allocated favor good use of the network. The base station dictates the price for relaying the content based on the relay transmitting power. The price given and the available power budget are used by the users to determine the optimal video rate and transmission power. The rate-distortion optimization during the video encoding is done such that the QoE of the user on the other side of the link is maximized for these constraints.

Most of the work in literature has used game theory approaches in wireless systems to determine fair spectrum sharing, such as [5], cognitive relay solutions, such as [6], and power allocation in [7]. These studies look at the overall network performance, mainly throughput, and ignore the characteristics of the video. Game theory was also applied to scheduling, rate adaptation and management of the buffers in multi-user systems in [8] and a pricing scheme was applied to allocate bit rate in [9]. These all apply to one-way traffic, which is more typical of communication networks, and do not consider a tight two-way video connection such as a video chat application.

The authors of the original paper develop a mechanism for an energy-efficient and fair transmission of video chat over wireless networks. The authors model the interaction of base stations and users using video chat using dynamic pricing with power allocation constraints as a Stackelberg game [4]. In their paper, they present a flow-based analysis for the rate-distortion and power usage. Fairness is also taken into consideration and the pricing trades this with energy efficiency. This results in a joint rate and power allocation problem that can be used during the coding of the video chat session to maximize the quality of the video. This type of video normally consists of low-motion characteristics and contains the head and the shoulders of the subject. Thus there is high correlation in space which can be exploited by the encoder. The authors propose a complexity-scalable video coding scheme that uses the correlation between neighboring macroblocks and a limited motion estimation search area. Selection is done between the possible coding parameters depending on the trade-off between complexity, rate and distortion. The authors present a model to describe these characteristics which can be adapted online during the video chat session.

With current smartphone technology video chat is becoming an important communication service as

vision helps improve communication information by conveying facial expressions and emotions. Further energy reductions, such as those related to the mobile display and camera, are needed to avoid fast draining of the battery. The wireless channel is a harsh environment and the conditions can change drastically, thus the optimization problem needs to be extended to cater for the random channel conditions and adapt also to these variables. Further signal processing is also needed to cater for the different lighting environments and noise of the captured cameras that can impinge on the coding performance and the quality of experience.

References:

[1] J. Baliga, R.W.A. Ayre, K. Hinton, and R.S. Tucker, "Energy consumption in wired and wireless access networks," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 70–77, June 2011.

[2] S. Cicalo and V. Tralli, "Distortion-fair cross-layer resource allocation for scalable video transmission in ofdma wireless networks," *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 848–863, April 2014.

[3] S.-P. Chuah, Z. Chen, and Y.-P. Tan, "Energy-efficient resource allocation and scheduling for multicast of scalable video over wireless networks," *IEEE Transactions on Multimedia*, vol. 14, no. 4, pp. 1324–1336, August 2012.

[4] Z. Han, D. Niyato, W. Saad, T. Başar, and A. Hjørungnes, *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*. Cambridge, U.K.: Cambridge University Press, 2011.

[5] X. Kang, R. Zhang, and M. Motani, "Price-based resource allocation for spectrum-sharing femtocell networks: A Stackelberg game approach," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 538–549, April 2012.

[6] Y. Xiao, G. Bi, and D. Niyato, "Game theoretic analysis for spectrum sharing with multi-hop relaying," *IEEE Transactions on Wireless Communications*, vol. 10, no. 5, pp. 1527–1537, May 2011.

[7] S. Ren and M. van der Schaar, "Pricing and distributed power control in wireless relay networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2913–2926, June 2011.

[8] J. W. Huang, H. Mansour, and V. Krishnamurthy, "A dynamical games approach to transmission-rate adaptation in Multimedia WLAN," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3635–3646, July 2010.

[9] M. Tiwari, T. Groves, and P. Cosman, "Bit-rate allocation for multiple video streams using a pricing-based mechanism," *IEEE Transactions on Image Processing*, vol. 20, no. 11, pp. 3219–3230, November 2011.

Carl James Debono (S'97, M'01, SM'07) received his B.Eng. (Hons.) degree in Electrical Engineering from the University of Malta, Malta, in 1997 and the Ph.D. degree in Electronics and Computer Engineering from the University of Pavia, Italy, in 2000. Between 1997 and 2001 he was employed as a Research Engineer in the area of Integrated Circuit Design with the Department of Microelectronics at the University of Malta. In 2000 he was also engaged as a Research Associate with Texas A&M University, Texas, USA. In 2001 he was appointed Lecturer with the Department of Communications and Computer Engineering at the University of Malta and is now an Associate Professor. He is currently the Deputy Dean of the Faculty of ICT at the University of Malta. Prof. Debono is a senior member of the IEEE and served as chair of the IEEE Malta Section between 2007 and 2010. He was the IEEE Region 8 Vice-Chair of Technical Activities between 2013 and 2014. He has served on various technical program committees of international conferences and as a reviewer in journals and conferences. His research interests are in wireless systems design and applications, multi-view video coding, resilient multimedia transmission, and modeling of communication systems.

Alleviating the Effects of Early User Departures with Progressive Streaming

A review for "Smart Streaming for Online Video Services"
(Edited by Roger Zimmermann)

Liang Chen, Yipeng Zhou, Dah Ming Chiu, "Smart Streaming for Online Video Services", IEEE Transactions on Multimedia, vol. 17, no. 4, pp. 485–497, April 2015.

In this manuscript the authors consider the natural phenomenon of users often terminating the viewing of a video stream early, rather than watching until the end of the content. This effect, often referred to as *early departure*, has been documented in a number of studies that have observed real-world systems [1]. There are a variety of reasons why users may elect to stop watching a stream before its end, but for a service provider the main consideration is that it may lead to resource, i.e., bandwidth wastage. This is especially the case in the context of progressive HTTP streaming that is currently in use by a number of large-scale video providers. The authors have collaborated with Tencent Video, one of the largest video streaming providers in China, and therefore have access to considerable trace data.

With progressive HTTP streaming, a video file is downloaded by the client at a high bandwidth after the initial play request has been initiated. Often the download bitrate is markedly higher than the playback bitrate of the video, depending on the available connection. This results in media data being buffered at the client side. The positive effect of this strategy is that stalls and playback freezes are reduced. Earlier studies have shown that such re-buffering events are generally much disliked by users and considerably reduce their Quality of Experience (QoE). However, one of the drawbacks of downloading a lot of data quickly is that some of that data may be wasted if a user decides to terminate watching a video early. Data that has already been pre-fetched into the client buffer then becomes useless and must be discarded. However, since the wasted data was transmitted from the server through the service provider's network, it resulted in costs. The authors investigate in this study how the bandwidth wastage can be reduced with a *smart streaming* strategy. It is noteworthy here to point out that progressive streaming is simpler than the more recently developed technique of Dynamic Adaptive Streaming over HTTP, now standardized as MPEG-DASH [2]. The authors acknowledge that DASH may alleviate some of the problems with progressive downloading, but they also remark that their smart streaming strategies can be integrated with DASH deployments.

In order to arrive at their smart streaming strategy, the authors first analyze a large set of trace data from their video streaming collaborator. Based on a set of over 550 million video streaming sessions, the authors extract and document that early departures follow a relatively stable pattern where approximately half of the departures fall into what is described as the *browsing* phase. This phase, for long form content (e.g., movies and TV shows), consists of roughly about the first 15% of a video's length. The other 50% of the early departures fall into the remaining *viewing* phase. The authors first describe four basic strategies that can be applied to a whole video, namely Simple Rate Control (SC), Best Effort Streaming (BE), Equal Buffer Streaming (EB), and Equal Waste-Rate (EW). The details of SC and BE are as follows:

- Simple Rate Control (SC): SC tries to maintain the playback rate and not go beyond it. SC does not incur any waste even if the user departs early.
- Best Effort Streaming (BE): BE corresponds to the strategy of progressive download, which is commonly implemented in HTTP-based streaming. The user end keeps requesting for video chunks. The server tries to respond with best effort, resulting in an equal rate when all other things are equal.

From their trace data analysis the authors postulate that a basic strategy should not be applied to the full video length, rather SC should be used in the browsing phase to meet users' QoE and yet minimize wastage, while for the viewing phase BE should be applied to help pre-fetch content while the termination rate is relatively low. The authors call this heuristic strategy Behavioral-Based Smart Streaming, or simply BB.

To understand the benefits of the different strategies the authors first execute a number of simulations with both a synthetic user arrival pattern (Poisson) and then a trace-driven pattern. The results indicate that a

number of metrics that are concerned with various aspects of video freezes can be significantly reduced with hybrid BB as compared to using only SC or only BE for the full length of a video. The number of users experiencing freezes as well as the freeze duration and rate are lowered. It is noted that BB does result in some wasted bandwidth – not as much as BE though. SC by design wastes almost no data. To further validate their results the authors also implemented a testbed that runs a significant number of client emulators and the server system. The testbed measurements largely confirm the simulation results, though the margins of improvement are somewhat reduced, which is to be expected, since a system incurs various implementation overheads.

Video content distributors are moving towards system deployments that leverage DASH, which contains mechanisms to optimize bandwidth usage. The authors briefly discuss that their smart streaming approach could be used to exploit the bandwidth available between different DASH quality levels, since the video rate is a quantized value. Overall the proposed method is interesting, especially also because it is based on insights from a large set of real-world session data. The issue of early departures can be further explored to optimize large-scale video streaming systems.

References:

[1] L. Chen, Y. Zhou, and D. M. Chiu. "Video browsing–A study of user behavior in online VoD services," in *Proceedings of the 22nd IEEE International Conference on Computer Communications and Networks*, Jul.–Aug. 2013, pp. 1–7.

[2] "Information Technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats." ISO/IEC 23009-1, 2012.

Roger Zimmermann is an associate professor with the Department of Computer Science at the School of Computing with the National University of Singapore (NUS) where he is also an investigator with the Interactive & Digital Media Institute (IDMI). His research interests are in both spatio-temporal and multimedia information management, for example distributed and peer-to-peer systems, spatio-temporal multimedia, streaming media architectures, georeferenced video management, mobile location-based services and geographic information systems (GIS). He has co-authored a book, six patents and more than hundred-ninety conference publications, journal articles and book chapters in the areas of multimedia and databases. He has received the best paper award at IEEE ISM 2012 and was part of the team who won second place at the ACM SIGSPATIAL GIS Cup 2013. He has been involved in the organization of conferences in various positions, for example program co-chair of ACM Multimedia 2013. He co-directs the Centre of Social Media Innovations for Communities (COSMIC) at NUS and is an investigator with the NUS Research Institute (NUSRI) in Suzhou, China. Roger Zimmermann is an Associate Editor of the ACM Transactions on Multimedia journal (TOMM, formerly TOMCCAP) and the Multimedia Tools and Applications (MTAP) journal. He is a Senior Member of the IEEE and a member of ACM. For more details, see http://eiger.ddns.comp.nus.edu.sg.

Paper Nomination Policy

Following the direction of MMTC, the R-Letter platform aims at providing research exchange, which includes examining systems, applications, services and techniques where multiple media are used to deliver results. Multimedia includes, but is not restricted to, voice, video, image, music, data and executable code. The scope covers not only the underlying networking systems, but also visual, gesture, signal and other aspects of communication.

Any HIGH QUALITY paper published in Communications Society journals/magazine, MMTC sponsored conferences, IEEE proceedings, or other distinguished journals/conferences within the last two years is eligible for nomination.

Nomination Procedure

Paper nominations have to be emailed to R-Letter Editorial Board Directors:

Christian Timmerer (christian.timmerer@aau.at), Weiyi Zhang (wzhang@ieee.org), and Yan Zhang (yanzhang@simula.no).

The nomination should include the complete reference of the paper, author information, a brief supporting statement (maximum one page) highlighting the contribution, the nominator information, and an electronic copy of the paper, when possible.

Review Process

Members of the IEEE MMTC Review Board will review each nominated paper. In order to avoid potential conflict of interest, guest editors external to the Board will review nominated papers co-authored by a Review Board member. The reviewers' names will be kept confidential. If two reviewers agree that the paper is of R-letter quality, a board editor will be assigned to complete the review letter (partially based on the nomination supporting document) for publication. The review result will be final (no multiple nomination of the same paper). Nominators external to the board will be acknowledged in the review letter.

R-Letter Best Paper Award

Accepted papers in the R-Letter are eligible for the Best Paper Award competition if they meet the election criteria (set by the MMTC Award Board).

For more details, please refer to http://committees.comsoc.org/mmc/rletters.asp

MMTC R-Letter Editorial Board

Multimedia Communications Technical Committee Officers

Chair: Yonggang Wen, Singapore
Steering Committee Chair: Luigi Atzori, Italy
Vice Chair – North America: Khaled El-Maleh, USA
Vice Chair – Asia: Liang Zhou, China
Vice Chair – Europe: Maria G. Martini, UK
Vice Chair – Letters: Shiwen Mao, USA
Secretary: Fen Hou, China
Standard Liaison: Zhu Li, USA

MMTC examines systems, applications, services and techniques in which two or more media are used in the same session. These media include, but are not restricted to, voice, video, image, music, data, and executable code. The scope of the committee includes conversational, presentational, and transactional applications and the underlying networking systems to support them.