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

MMTC Communications – Review

IEEE COMMUNICATIONS SOCIETY

Vol. 8, No. 3, June 2017

TABLE OF CONTENTS

Message from the Review Board Directors

Welcome to the June 2017 issue of the IEEE ComSoc MMTC Communications – Review.

This issue comprises five reviews that cover multiple facets of multimedia communication research including confidential cloud-based image processing, correlated image transmission within cloud, QoE enhancement in 5G networks, and secure multimedia transmission in wireless sensor networks. These reviews are briefly introduced below.

The first paper, published in IEEE Transactions on Multimedia and edited by Debashis Sen. Its key idea is to represent unlabeled video by a fixed-length tag vector. Each dimension of the tag vector uniquely corresponds to a specific tag, with the value reflecting the relevance of the tag with respect to the video content.

The second paper is published in IEEE Conference on Computer Vision and Pattern Recognition and edited by Jun Zhou. A two-stage network structure was proposed, which includes an encoder-decoder stage for alpha matte estimation, and a residual network for outcome refinement.

The third paper, published in IEEE Conference on MultiMedia Signal Processing (MMSP) and edited by Gwendal Simon. It focuses in their article on a series of popular technologies (especially TCP, DASH, and ISOBMFF) and they design solutions that are specific to these technologies.

The fourth paper is published in IEEE Transactions on Multimedia and edited by Roger Zimmermann. The paper investigates the optimization of video transcoding task scheduling in recently popular crowdsourced live streaming systems.

The fifth paper is published in IEEE Transactions on Communication and edited by Cong Shen. A novel analytical framework is proposed for multi-BSS IEEE 802.11 networks with universal frequency reuse in this paper.

All the authors, nominators, reviewers, editors, and others who contribute to the release of this issue deserve appreciation with thanks.

IEEE ComSoc MMTC Communications – Review Directors

Pradeep K. Atrey
State University of New York at Albany, USA
Email: patrey@albany.edu

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

Wei Wang
San Diego State University, USA
Email: wwang@mail.sdsu.edu

Jun Wu
Tongji University, China
Email: wujun@tongji.edu.cn

Towards Unsupervised Representation Learning for Video Event Detection

A short review for "TagBook: A Semantic Video Represenation Without Supervision for Event Detection"

Edited by Debashis Sen

M. Mazloom, X. Li and C.G.M. Snoek, "TagBook: A Semantic Video Representation Without Supervision for Event Detection," IEEE Transactions on Multimedia, vol. 18, no. 7, Jul. 2016.

Video event detection strives to detect events such as *dog show*, *felling a tree*, and *birthday party* in arbitrary video content. Compared with video categorization that focuses on recognizing individual concepts, event detection is more challenging as it requires an effective modeling of co-occurrence of multiple concepts and their interactions along the temporal dimension. The key to event detection is to have a discriminative video representation. Traditional video representations for event detection rely on low-level features extracted from a video's audiovisual content [1]. Not surprisingly, learned representations derived from deep convolutional neural network layers are becoming popular for video event detection as well [2]. As these feature vectors are of high dimensionality, they perform well when provided with sufficient carefully labeled examples for each event. However, as events become more specific, it becomes increasingly difficult to find adequate relevant examples for supervised deep learning. Moreover, in scenarios where only a textual description of an event is available, *i.e.* without any video training example, even the low-level representations by themselves are less useful.

In this paper, the authors consider event detection where video examples of the event are scarce (few-example) or even completely absent (zero-example). To conquer these two challenging settings, previous works build semantic representations on top of pre-trained concept detectors such as 'dog', 'tree', and 'groom'. Consequently, the previous need for sufficient event examples was seemingly replaced by the complicate issue of specifying appropriate concepts and acquiring concept examples. In contrast, the authors of this paper propose a simple yet effect solution of learning exclusively from freely available social-tagged videos, without the need for training any intermediate concept detectors.

The authors' primary contribution is *TagBook*, a new semantic video representation applicable for both few-example and zero-example event detection. The key idea of TagBook is to represent unlabeled video by a fixed-length tag vector. Each dimension of the tag vector uniquely corresponds to a specific tag, with the value reflecting the relevance of the tag with respect to the video content. To compute the relevance, the authors develop an intuitive algorithm that propagates tags to a given video from many social-tagged videos.

The tag propagation process is inspired by image neighbor voting [3], and the authors take two steps further to make it more suited for the video domain. The first improvement is to perform video-level tag propagation instead of frame-level tag propagation. As the amount of videos is much smaller than the amount of video frames, this design allows one to learn a TagBook representation from very large-scale video sets within reasonable time. The second improvement is to conduct tag refinement on the social-tagged training set before using it as a training resource for tag propagation. This resolves to some extent the inaccuracy and the incompleteness of social tags associated with the source videos.

TagBook works effectively for both zero- and few-example event detection. For zero-example, an event of interest is expressed by a textual description. By converting the description into a tag vector, the cross-media relevance between the event and a specific video can be computed in light of the cosine similarity between their tag vectors. As for the few-example setting, the authors build the event's TagBook representation by weighted combination of the TagBooks of the event's learning examples. In particular, noting the combination's high resemblance to the decision function of a linear Support Vector Machine (SVM), the authors nicely optimize the weights by a linear SVM solver.

Tested on more than 1,000-hours unconstrained Internet videos from the 2013 and 2014 editions of the NIST TRECVID Multimedia Event Detection Task, and the Columbia Consumer Video dataset, the proposed TagBook video representation is shown to be remarkably effective, even outperforming very recent state-of-the-art alternatives, e.g., VideoStory [4] and Concept Prototypes [5], that are built on supervised representations. As for the construction of TagBook, results show that tag propagation by improved neighbor voting is superior to training SVM models per tag. This is quite attractive, as neighbor voting is more easy to implement and more flexible to handle novel tags than the model-based alternative.

TagBook is a significant effort towards unsupervised representation learning for video event detection. It can learn from social-tagged videos alone. The tag propagation algorithm is easy to implement. Presumably, its performance can be improved further with better low-visual video features used for tag propagation. More importantly, as each dimension of a TagBook vector corresponds to a specific tag, the representation is quite interpretable. As such, for event detection given only a few or even zero examples, TagBook, in near future, should emerge among the first choices to use.

Acknowledgement

The editor would like to thank the authors for providing a preliminary draft of this review.

References:

[1] Y.-G. Jiang, S. Bhattacharya, S.-F. Chang, and M. Shah, "High-level event recognition in unconstrained videos," IJMIR, vol. 2, no. 2, pp. 73–101, 2013.

[2] Z. Xu, Y. Yang, and A. Hauptmann, "A discriminative CNN video representation for event detection," in Proc. of CVPR, 2015.

[3] X. Li, C. Snoek, and M. Worring, "Learning social tag relevance by neighbor voting," TMM, vol. 11, no. 7, pp. 1310–1322, 2009.

[4] A. Habibian, T. Mensink, and C. Snoek, "VideoStory: A new multimedia embedding for few-example recognition and translation of events," in Proc. of ACM MM, 2014.

[5] M. Mazloom, E. Gavves, and C. Snoek, "Conceptlets: Selective semantics for classifying video events," TMM, vol. 16, no. 8, pp. 2214–2228, 2014.

Debashis Sen is an Assistant Professor in Department of Electronics and Electrical Communication Engineering of Indian Institute of Technology - Kharagpur. He received his Ph.D. in Image Processing from Jadavpur University, India. He was a postdoctoral researcher at the Multimedia Analysis and Synthesis Laboratory, National University of Singapore and at the Center for Soft Computing Research, Indian Statistical Institute. He currently heads the Vision, Image and Perception research group in his department. His current research interests are in Vision, Image and Video Analysis, Uncertainty Handling, Eye Movement Analysis, Artificial Models of Cognition, Computational Visual Perception, Applied Visual Neuroscience, Applied Multi-modal Deep Learning, and Multimedia Signal Processing. He has authored/co-authored many research articles in reputed IEEE, IET, Springer and Elsevier journals and conferences. Dr. Sen is on the editorial board of IET Image Processing and Springer's Circuits, Systems and Signal Processing. He has received a young scientist award from The Institution of Engineers (India) and a couple of best paper awards from IET.

Another Success of Deep Learning on Image Matting

A short review for "Deep Image Matting"

Edited by Jun Zhou

Ning Xu, Brian Price, Scott Sohen, and Thomas Huang. "Deep Image Matting", IEEE Conference on Computer Vision and Pattern Recognition, 2017.

Image matting is an important technique in image processing and computer graphics. It is widely adopted for image editing and scene synthesis by multimedia and movie industries. Given an image and a trimap, the aim of image matting is to extract an alpha matte so that the image can be represented as a linear combination of foreground and background colors [1].

Quite a few methods have been reported to solve the image matting problem. They can be roughly divided into two categories. The first category performs local affinity analysis of the statistical correlation among neigboring pixels to propagate alpha values from known image regions to unknown pixels [2]. Another category is based on sampling of known regions to find candidate colors that can produce the composite colors [3]. A common property of these approaches is that they are based on color features and require trimap as a guidance, which makes them sensitive to overlapping foreground and background color distribution and trimap placement. Furthermore, most approaches were developed on

In order to better utilize the image structure and semantic information, deep learning has been introduced into image matting. Shen *et al.* [4] proposed an end-to-end solution based on convolutional neural network (CNN) to create the trimap of a person in a portrait, and then used the method in [2] to estimate the matte results. Cho *et al.* [5] also used CNN for matting, but on natural images. They took the output of [2], KNN matting [6], and normalized color image as inputs, and learned an end-to-end deep CNN to reconstruct the alpha mattes. It is interesting to see that CNN was used for different purposes other than directly creating an alpha matte in these two approaches. Nevertheless, thanks to the powerful deep learning models, these two methods still outperformed previous methods.

This latest CVPR oral paper from Xu *et al.* adopted a different strategy. A two-stage network structure was proposed, which includes an encoder-decoder stage for alpha matte estimation, and a residual network for outcome refinement. This allows the alpha matte be directly learned rather than relying on previous affinity analysis or sampling approaches.

In the first stage, the input to the deep learning model is a four channel image with the color image concatenated with the trimap. The encoder network uses 14 convolutional layers and 5 max-pooling layers to downsample feature maps. Then the decoder network with 6 convolutional layers, 5 unpooling layers, and an alpha prediction layer is used to estimate the alpha matte. In the network structure, two losses are calculated for alpha prediction, i.e., the difference between the estimated and the ground truth alpha values, and for image composition, i.e., the difference between the ground truth colors and the predicted colors via alpha matting. In order to increase the number of training samples, the training data are augmented via randomly cropping images from unknown regions and different sizes, flipping images, and randomly dilating the trimaps.

The first stage of alpha prediction tends to give smoothed results. The authors therefore proposed to use another network to refine the alpha mattes to achieve shaper edges. The input to this network is image and predicted alpha values. A simpler network structure is adopted with 4 convolutional layers including a ReLU layer.

Since the training deep learning methods require large amount of data, authors produced a new synthesized matting dataset of 49,300 images with 493 unique foreground objects for training. The testing set includes 1000 images with 50 unique objects. This dataset is significantly larger than the widely used matting benchmark

introduced by Rhemann *et al*. [7].

This method was compared with seven state-of-the-art methods on aforementioned datasets, and a real image dataset. The results show significant improvement over alternative approaches, with more than halved SAD, MSE, Gradient, and Connectivity results. In particular, this method is robust to trimap changes and shows stable error rates compared with rapid rising errors from other approaches when the trimap is changed.

This paper is yet another example on improvement that deep learning can bring to computer vision tasks. Clearly, a key factor to the success of the proposed method is from building a very large dataset so as to sufficiently train the deep learning model. Another idea that can be learned from this paper is that given a complex tasks with several steps, different deep learning models can be adopted for each step. Optimized performance in single step will be accumulated to generate much bigger benefit to the overall task.

References:

[1] Y. Chuang, B. Curless, D. Salesin, and R. Szeliski. "A Bayesian approach to digital matting," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 264–271, 2001.

[2] A. Levin, D. Lischinski, and T. Weiss. "A closed-form solution to natural image matting," IEEE Trans. Pattern Analysis and Machine Intelligence, 30(2):228-242, 2008.

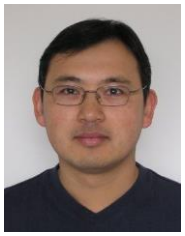
[3] J. Wang and M. F. Cohen. "Optimized color sampling for robust matting," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. 2007.

[4] X. Shen, X. Tao, H. Gao, C. Zhou, and J. Jia. "Deep automatic portrait matting." In Proceedings of the European Conference on Computer Vision, 2016.

[5] D. Cho, Y.-W. Tai, I. Kweon, D. Cho, Y.-W. Tai, and I. Kweon. "Natural image matting using deep convolutional neural networks," In Proceedings of the European Conference on Computer Vision, 2016.

[6] Q. Chen, D. Li, and C. Tang. "KNN matting," IEEE transactions on pattern analysis and machine intelligence, 35(9):2175–2188, 2013.

[7] C. Rhemann, C. Rother, J. Wang, M. Gelautz, P. Kohli, and P. Rott. "A perceptually motivated online benchmark for image matting," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1826-1833, 2009.



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 is now a senior lecturer in the School of Information and Communication Technology in Griffith University. 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 spectral imaging pattern recognition, computer vision, and their applications to and environmental informatics and remote sensing.

Fast Bootstrap of Video Delivery: Technologies, Science, and Research

A short review for "Fast DASH Bootstrap"

Edited by Gwendal Simon

Nassima Bouzakaria, Cyril Concolato, and Jean Le Feuvre. " Fast DASH Bootstrap " in IEEE Conference on MultiMedia Signal Processing (MMSP), 2015

The performance of multimedia systems is essentially related to the so-called Quality of Experience (QoE) and to the various criteria that make the content consumers enjoy the service. To identify the main challenges of multimedia systems, researchers have thus to conduct large-scale measurement campaigns and to analyze a vast amount of data. In the recent years, several studies have shown that, regarding video delivery in Video-on-Demand services (whether it is for long movies such as Netflix or in crowdsourced platforms such as YouTube), consumers expect the video to start as fast as possible. It means that the delay between the time at which the user clicks to see the video, and the time at which the fist frames are actually displayed must be as short as possible. In this area, every effort that results in a millisecond gain is a progress, which improves the QoE, and by consequences the success of the content provider.

For the past decade, the content providers have massively used Content Delivery Network (CDN) to meet this demand for fast video bootstrap. By reducing the distance between the client and the server, the content providers enable fast reaction of the service. Thousands of papers have been published on the optimization of CDNs. The authors of "*Fast DASH Bootstrap*" were among the first to specifically address the challenge of reducing the delay without regards of the server placement, but rather by analyzing the flow of messages during the "handshake" (messages that are required to introduce both entities and to properly set the system parameters).

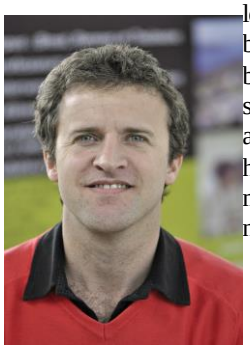
Multimedia Systems are at the frontier of multiple scientific disciplines, such as networking, operating systems, and algorithms. To address the big challenges in the area, it is sometimes necessary to explore a bit of several untelated areas, in the hope of making bridges between different advances that have insufficiently been combined. In the article "*Fast DASH Bootstrap*", the authors have typically to deal with the network layer (especially the behavior of TCP), with the application layer (especially the process in adaptive streaming systems), and with video technologies (especially the video containers).

Multimedia Systems are also applied science, where the researchers consider the technologies that are the most widely implemented, including their potential weaknesses, and no matter there exists another potentially better but not used option. Researchers have to deal here with various constraints, some of them being the context, the market, and the technological trends. With that respect, the authors of "Fast DASH Bootstrap" focus in their article on a series of popular technologies (especially TCP, DASH, and ISOBMFF) and they design solutions that are specific to these technologies. It is a limitation of such work, since the impact of this article is valid only until these technologies become obsolete (but given their popularity, the risk of fast obsolescence is very low in this case). But it is also a strength since significant gains can be obtained by exploiting the features of these technologies.

Finally, researchers in Multimedia Systems can contribute to the global knowledge by the means of fundamental results (academic theoretical results related to signal processing), and also sometimes by contributions to standards, which are eventually implemented worldwide by various actors. In this latter case, the academic papers aim at demonstrating the efficiency of the proposal by going into the details of the scientific process that has led the authors to make this proposal in the standard. In the case of the article "Fast DASH Bootstrap", the authors are definitely in this process. Their main contribution is specifically a proposal to modify DASH standard. They demonstrate that this slight change enables significant gains regarding the bootstrap, without any modification in the delivery infrastructure architecture, so at a minimum investment overhead for content providers.

In summary, the article “Fast DASH Bootstrap” is an original contribution, which some may judge less noble than various other papers because 1/ it is not generic, 2/ it is not backed by any long theoretical development, and 3/ it involves several scientific disciplines. However, I would like to emphasize the quality of this paper because, in a very pragmatic approach, the authors succeed in obtaining significant gains for a problem that is considered as a challenge (as revealed by the massive investment in CDN), simply by combining in a smart way multiple well-known technologies. The outcome of this paper is especially strong since it results in an adopted standard add-on and in implementation in currently active services.

While many developers considered that it was no longer possible to reduce bootstrap unless investing billion dollars into sophisticated and complex fog architecture, the authors show here that gains as high as 100 ms can be obtained without new investment.



Gwendal Simon is Associate Professor at Telecom Bretagne. He received his Master Degree in Computer Science in 2000 and his PhD degree in Computer Science in December 2004 from University of Rennes 1 (France). From 2001 to 2006 he was a researcher at Orange Labs, where he worked on peer-to-peer networks and social media innovations. Since 2006, he has been Associate Professor at Telecom Bretagne, a graduate engineering school within the Institut Mines-Telecom. He has been a visiting researcher at University of Waterloo from September 2011 to September 2012. His research interests include large scale networks, distributed systems, optimization problems and video delivery systems.

User-Generated Live Streaming with Transcoding in the Cloud

A review for "Coping With Heterogeneous Video Contributors and Viewers in Crowdsourced Live Streaming: A Cloud-Based Approach"
Edited by Roger Zimmermann

Qiyun He, Jiangchuan Liu, Chonggang Wang, Bo Li, "Coping With Heterogeneous Video Contributors and Viewers in Crowdsourced Live Streaming: A Cloud-Based Approach", IEEE Transactions on Multimedia, vol. 18, no.5, pp. 916–928, May 2016.

The authors investigate and consider the optimization of video transcoding task scheduling in recently popular crowdsourced live streaming systems. Because of technological advances in both smartphones and wireless networks a number of crowdsourced live streaming applications have become popular with users. These systems allow individuals to live broadcast events to an audience world-wide. Examples of such live streaming systems are Twitch.tv, YouTube Live and Facebook Live. Such systems need to manage very heterogeneous upload stream formats, widely varying viewership numbers and users who are geographically distributed.

To provide a solution that can scale, the authors consider the situation where a live stream is transmitted into a server environment from which it is then broadcast to end-users. Specifically the authors assume that the dissemination of streams is accomplished via the dynamic adaptive streaming over HTTP (DASH) [1] approach. With DASH, a source stream is transcoded into a number of different representations, from low to high quality, requiring low to high transmission bandwidth. Each representation is also temporally divided into so-called chunks (usually of 2 to 10 seconds length) and a DASH client can adaptively switch between the different representations at chunk-boundaries, depending on the bandwidth that is available at the time and to ensure a good and continuous viewing experience. One of the key elements in DASH, especially for live transmissions, is the heavy computation load that real-time transcoding requires.

Since the user numbers and their geographic location vary widely, the authors propose to use a flexible cloud environment for transcoding a scalable number of video streams. In such an application setting a key issue is the optimized number and location of the transcoding tasks. The authors judiciously assume that transcoding should occur close to the viewers of a stream in order to ensure a good end-user experience. Public cloud instances, such as Amazon Web Services (AWS) server and network resources can be rented in different regions of the world. The optimization goal is to match one or more cloud servers to one broadcaster (i.e., user live stream) to maximize the *reward* and to minimize the overall *cost*. The authors first present a generic framework with an optimal scheduler for allocating cloud instances with no regional constraints. They then extend the solution to accommodate regional constraints, and discuss a series of practical enhancements.

For their study the authors have collected a significant trace dataset captured through the Twitch.tv API. They first provide some insights based on this dataset in order to further motivate the presented approach. From the statistics that are provided the reader learns that heterogeneity is one of the key aspects. For example, 177 different source video resolutions ranging from 116p to 1600p were detected. Furthermore, the viewership is geographically widely distributed and the last-mile Internet subscriber bandwidth also varies considerably from country to country. Additionally, on the expense-side server instances and network bandwidth costs of AWS also vary across regions. This heterogeneity in the conditions motivates the design of an optimizing scheduler which balances the tradeoff between opposing objectives, namely achieving high viewer quality of experience (QoE) while maintaining low overall system cost. The reward is the total viewer satisfaction, i.e., the overall QoE, which consists of the average satisfaction level of a channel and the popularity level of a broadcaster. On the other hand,

the cost stems from renting servers, outbound data traffic charges and cross-boundary network traffic.

The authors investigate the optimal resource scheduling of assigning cloud server instances in different regions to live channels for video transcoding in multiple steps. They first consider the case where sufficient cloud service resources are available in every region. Under such conditions it is shown that a greedy approach results in an optimal solution, which the authors call the Greedy Rental Scheduler (GRS). However, in reality service supply may not be unlimited in a region and hence it is more realistic to consider what the authors term Scheduling with Limited Cloud Service Supply (SLCS). With the additional constraints a greedy approach does not work and the authors introduce a dynamic programming framework. As is shown later in the experimental results, SLCS runs sufficiently fast for a small to medium system scale. For larger environments, the authors propose a third algorithm called SLCS* which reduces the complexity of SLCS by using a heuristic to rank cloud instances ahead of time, and then selects the generally most preferred instance first when assigning the rental schedule.

The authors evaluate the three proposed scheduling algorithms GRS, SLCS, and SLCS* first in a simulation environment and then in a Planetlab/Amazon EC2 setup. As a fourth method, Top-N, the current strategy of Twitch.tv is compared. Top-N provides full video representations to a small number (top *N*) most popular broadcasters, but only the original video quality for everybody else. For the dataset, the authors use their previously mentioned Twitch.tv traces, obtained from February to June 2015. The costs are modeled after Amazon EC2 c3 server instances. The three different schedulers are evaluated based on a number of metrics, for example the comprehensive cost, just the rental cost and the total viewer satisfaction. The experimental evaluation reveals that overall SLCS performs best, but conversely it also requires the largest scheduler execution time. SLCS* is on the order of 2.5 times faster than SLCS and provides very competitive performance. All three methods, GRS, SLCS, and SLCS*, outperform Top-N, thus clearly demonstrating that optimizing the scheduling has its benefits. The authors further explore a number of parameter settings, for example different weights in the reward function. They also performed experiments in a Planetlab/Amazon EC2 environment where additional issues such as realistic streaming delays and network throughput are measured.

Overall the proposed framework is interesting and demonstrates that the scheduling of transcoding tasks can be optimized with the SLCS and SLCS* algorithms in recently popular crowdsourced live streaming systems.

References:

[1] MPEG-DASH Specification. ISO/IEC 23009-1:2012 Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats, April 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 Smart Systems Institute (SSI). His research interests are in both spatio-temporal and multimedia information management, for example distributed 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, seven patents and more than two hundred conference publications, journal articles and book chapters in the areas of multimedia and databases. He has received the best paper award at the ACM IWGS 2016 workshop and the IEEE ISM 2012 conference. He has been involved in the organization of conferences in various positions, for example program co-chair of ACM Multimedia 2013. He 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://www.comp.nus.edu.sg/~rogerz.

Universal Frequency Reuse for High-Density Wi-Fi Networks

A short review for "Throughput Optimization of Multi-BSS IEEE 802.11 Networks with Universal Frequency Reuse"

Edited by Cong Shen

Yayu Gao, Lin Dai and Xiaojun Hei, "Throughput Optimization of Multi-BSS IEEE 802.11 Networks with Universal Frequency Reuse," IEEE Transactions on Communications, early access, 2017.

The low cost and easy deployment of Wi-Fi devices, as well as the soaring demand of high data rate performance, have led to the worldwide emergence of high-density IEEE 802.11 networks [1]. Yet it faces significant challenges due to the increasing interference from neighboring co-channel basic service sets (BSSs). To address this issue, the majority of previous studies have been focused on allocating orthogonal sub-channels among neighboring BSSs to minimize co-channel interference [2-5]. With the growing bandwidth of each sub-channel in the latest IEEE 802.11ac standard [6], nevertheless, the number of available orthogonal sub-channels is expected to be greatly reduced. It implies that the current frequency-division approach may not be in line with the evolution of 802.11 standards. Instead, for Wi-Fi networks longing for high data rates, they may have to share the frequency band.

Although the universal frequency reuse approach has been widely adopted in cellular networks, for Wi-Fi networks, the performance with universal frequency reuse remains little understood. In contrast to cellular networks where centralized access is adopted in each cell, IEEE 802.11 networks are based on *random access*. Due to the lack of proper models, performance optimization of random-access networks has been known as notoriously difficult. With multiple BSSs sharing the frequency band, the co-channel interference needs to be carefully characterized, which further complicates the modeling and performance analysis.

In this paper, a novel analytical framework is proposed for multi-BSS IEEE 802.11 networks with universal frequency reuse. Fundamental performance metrics, such as the steady-state points and the network throughput, are obtained as functions of system parameters including the number of nodes and the initial backoff window size of each node. The network throughput is further maximized by optimally choosing the initial backoff window sizes of all the nodes, and shown to be closely dependent on the percentage of nodes that can be heard by multiple access points (APs).

At the core of this analytical framework lies in an elegant tradeoff between accuracy and modeling complexity. To capture the essence of co-channel interference, a multi-group model is proposed in this paper, where nodes in each BSS are divided into different groups according to the set of APs by which they can be heard, and each group is modeled to have a distinct probability of sensing an idle channel as it suffers from different levels of co-channel interference. To reduce the modeling complexity, it is assumed that each node in the same BSS has an identical steady-state probability of successful transmissions, which is a good approximation when the initial backoff window sizes of nodes are sufficiently large.

The approximation leads to a simple, yet powerful, analytical model that sheds important light on the optimal design of multi-BSS IEEE 802.11 networks. In particular, it is found that a close-to-optimal network throughput can be achieved by individual BSS optimization that requires no inter-BSS coordination. This important result implies that in practical high-density Wi-Fi networks, each AP can calculate the optimal initial backoff window size only based on the local information of the node population in its own BSS. Moreover, the analysis shows that the maximum network throughput of a multi-BSS network is equal to that of a single-BSS network if each node can be

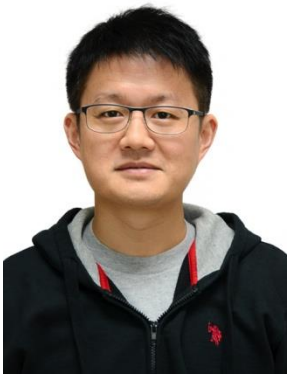
heard by all the APs. It indicates that for a high-density Wi-Fi network, to ensure that the optimal network throughput performance can be improved by increasing the number of APs, they need to have distinct coverage areas.

Most importantly, this paper provides a new perspective for the future design of high-density Wi-Fi networks. The comparison to the conventional orthogonal frequency division reveals that despite the degradation of maximum network throughput due to co-channel interference among BSSs, significant gains in the network data rate can be achieved by universal frequency reuse. It suggests that similar to cellular networks, universal frequency reuse could be a preferable option for multi-BSS IEEE 802.11 networks thanks to better spectral efficiency and more flexibility in network planning.

In summary, the analytical framework proposed in this paper provides a simple and effective way to evaluate and optimize the network performance of multi-BSS IEEE 802.11 networks with universal frequency reuse. It not only contributes to the fundamental understanding of modeling and performance optimization of random-access networks, but also sheds plenty of insights to practical system design of high-density WiFi networks.

References:

[1] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach* (5th Edition), Addison Wesley, 2009.
[2] A. Mishra, V. Brik, S. Banerjee, A. Srinivasan, and W. Arbaugh, "A client-driven approach for channel management in wireless LANs," in *Proc. IEEE INFOCOM*, pp. 1–12, 2006.
[3] M. Yu, A. Malvankar, and W. Su, "A Distributed Radio Channel Allocation Scheme for WLANs with Multiple Data Rates," *IEEE Trans. Commun.*, vol. 56, no. 3, pp. 454–465, Mar. 2008.
[4] X. Yue, C. Wong, and S. Chan, "CACAO: Distributed Client-Assisted Channel Assignment Optimization for Uncoordinated WLANs," *IEEE Trans. Parallel and Distributed Systems*, vol. 22, no. 9, pp. 1433–1440, Sept. 2011.
[5] S. Chakraborty, B. Saha, and S. Bandyopadhyay, "Dynamic Channel Allocation in IEEE 802.11 Networks," *Int. J. Comp. App.*, vol. 149, no. 1, Sept. 2016.
[6] *IEEE Std. 802.11ac-2013, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications–Amendment 4: Enhancements for Very High Throughput for Operation in Bands below 6 GHz*, Dec. 2013.



Cong Shen received his B.S. and M.S. degrees, in 2002 and 2004 respectively, from the Department of Electronic Engineering, Tsinghua University, China. He obtained the Ph.D. degree from the Electrical Engineering Department, UCLA, in 2009. From 2009 to 2014, He worked for Qualcomm Research in San Diego, CA. In 2015, he returned to academia and joined University of Science and Technology of China (USTC) as the 100 Talents Program Professor in the School of Information Science and Technology. His general research interests are in the area of communication theory, wireless networks, and information theory. He is a Senior Member of IEEE and serves as an editor for the IEEE Transactions on Wireless Communications

Paper Nomination Policy

Following the direction of MMTC, the Communications – Review 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 Review Board Directors: Pradeep K. Atrey (patrey@albany.edu), Qing Yang (qing.yang@montana.edu), Wei Wang (wwang@mail.sdsu.edu), and Jun Wu (wujun@tongji.edu.cn). 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 Review quality, a board editor will be assigned to complete the review (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.

Best Paper Award

Accepted papers in the Communications – Review 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://mmc.committees.comsoc.org/.

MMTC Communications – Review Editorial Board

Multimedia Communications Technical Committee Officers

Chair: Shiwen Mao, Auburn University, USA
Steering Committee Chair: Zhu Li, University of Missouri, USA
Vice Chair – America: Sanjeev Mehrotra, Microsoft, USA
Vice Chair – Asia: Fen Hou, University of Macau, China
Vice Chair – Europe: Christian Timmerer, Alpen-Adria-Universität Klagenfurt, Austria
Letters & Member Communications: Honggang Wang, UMass Dartmouth, USA
Secretary: Wanqing Li, University of Wollongong, Australia
Standard Liaison: Liang Zhou, Nanjing University of Posts and Telecommunications, China

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.