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

MMTC Communications - Frontiers

Vol. 14, No. 3, May 2019

CONTENTS

SPECIAL ISSUE ON *Application of Age of Information, Caching and Mobile Edge Computing in Wireless Networks*

Guest Editor: Rui Wang
Tongji Univeristy, China
ruiwang@tongji.edu.cn

This special issue of Frontiers focuses on applying several recently developed techniques, including age of information, Caching and mobile edge computing, in wireless networks. These three research directions have received great attentions from both academia and industries. Various research groups all around the world are currently working on these topics. We invited three papers from three distinguished research groups. The main contributions are summarized as follows.

The first paper of the issue focuses on the problem of minimizing *age of information* AoI. It provides approach on how to intelligently schedule the IoT devices to sample and update their status information, in order to minimize the AoI.

In the second paper, issues related wireless network coding and caching are discussed. Authors try to design the linear wireless network coding in the wireless caching by exploiting the characteristics of wireless channel and interference. They propose the linear wireless network coding operated wireless caching, referred to as linear network coded (NC) wireless caching, consisting of linear network coding assisted cache placement phase and signal-space alignment (SSA) enabled content delivery phase.

The third paper is about computation offloading in mobile edge computing. They formulate the problem of sum offloading rate maximization by joint offloading-user scheduling, offloaded-data size control, and communication-and-computation time division and also propose an optimal algorithm with low complexity based on a decomposition approach and Dinkelbach method.

Rui Wang (ruiwang@tongji.edu.cn) received his Ph.D. degree in 2013 from Shanghai Jiao Tong University, China. From Aug. 2012 to Feb. 2013, he was a visiting Ph.D. student at the Department of Electrical Engineering of University of California, Riverside. From Oct. 2013 to Oct. 2014, he was with the Institute of Network Coding, the Chinese University of Hong Kong as a post- doctoral research associate. From Oct. 2014 to Dec. 2016, he was with the College of Electronics and Information Engineering, Tongji University as an assistant professor, where he is currently an associate professor. Dr. Wang is currently an associate editor of the journal of IEEE Access and editor of IEEE Wireless Communications Letters.

Minimizing Age of Information in the Internet of Things

Bo Zhou and Walid Saad
Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061, USA.
{ecebo,walids}@vt.edu

1. Introduction

To enable many time-sensitive Internet of Things (IoT) applications [1], such as environment monitoring, drone navigation, and autonomous driving, it is imperative to deploy a reliable wireless infrastructure that can deliver low-latency communications. Such low-latency communications is needed to ensure a timely delivery of the IoT data pertaining to the status of the physical processes that are being monitored or operated by the IoT devices. To evaluate the timeliness of the IoT status information, the concept of *age of information* (AoI) has been recently introduced as a key performance metric. The AoI allows one to quantify the elapsed time from the generation of the last received status update at a remote information destination [2]. The AoI naturally characterizes the IoT information freshness from the perspective of the remote destination, and can jointly account for the latency introduced by sampling the physical process and transmitting the generated status updates. Thus, this notion is fundamentally different from traditional performance metrics, such as delay and throughput.

Recently, the problem of minimizing AoI has been addressed for variety of communication system settings, such as, for example, wireless broadcast systems (e.g., see [3] and [4]), queueing systems as done in [5] and [6], as well as energy harvesting systems (e.g., see [7] and [8]). In general, in these existing works [2]-[8], there are two general models of the generation process of the status update: the first one in which status updates randomly arrive at the source (e.g., see [2]-[5]) and the second one in which status updates can be generated at will by the source (e.g., see [6]-[8]). Various optimal and suboptimal scheduling/ updating algorithms have been proposed to minimize the AoI in [2]-[8], through different mathematical tools, including queueing theory, dynamic programming, multi-armed bandit, and Lyapunov optimization.

However, two important practical issues in minimizing AoI have been largely overlooked in the existing literature, e.g., [2]-[8], from the perspectives of the generation and transmission of status updates, respectively. On one hand, to implement sophisticated artificial intelligence tasks [9], IoT devices will have to consume a significant amount of energy. On the other hand, for low-power IoT devices with limited transmission capability, the devices may only be able to transmit a few bits in one transmission slot. Thus, a single status update should be split into multiple transmission packets and more than one time slots will be needed to send the complete status update to the destination.

In presence of *the energy cost* pertaining to the sampling process, and the multi-time slot transmissions with *non-uniform sizes of the status updates*, how to intelligently schedule the IoT devices to sample and update their status information, in order to minimize the AoI is still an open problem. In this regard, based on our works in [10]-[13], this e-letter provides new approaches to address the aforementioned two issues for minimizing the AoI in real-time IoT monitoring systems. Particularly, our contributions include: i) A joint design of status sampling and updating processes that minimizes the AoI while meeting stringent device energy constraints, by taking into account the energy cost for generating and updating status updates [10], [11]; and ii) a joint design of device scheduling and status sampling that minimizes the average AoI, by taking into account the non-uniform sizes of status update packets [12], [13].

2. Joint Status Sampling and Updating under Energy Cost Constraints

We first focus on the issue of the energy cost for generating and updating status packets. In Fig. 1, we illustrate the status sampling and updating processes for a single IoT device in a real-time monitoring system. The IoT device can collect the real-time status of an underlying physical process with a sampling cost and can send status packets to a remote destination through a wireless channel, with an update cost that depends on the channel condition.

Due to the energy budget of the device and the energy cost for generating and sending status packets at each time slot, the device must decide whether to sample the physical process and whether to send the generate status packet to the destination so as to maintain the freshness of the status information. We adopt the AoI to measure the freshness of the status information. In particular, we introduce two concepts of AoI at the device and AoI at the destination, so as to quantify the age of the status update at the device and the most recently received update at the destination, respectively.

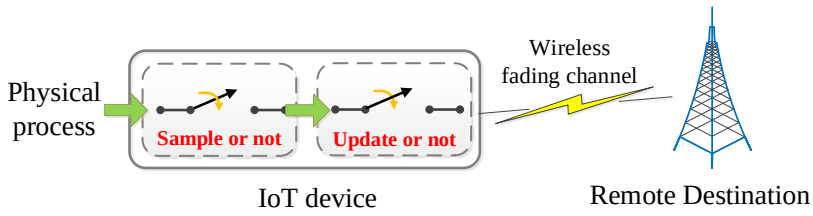


Fig. 1 Illustration of a real-time monitoring system with a single IoT device.

We are interested to find an optimal stationary sampling and updating policy that minimizes the time-average AoI at the destination, under the time-average energy constraint at the device. This stochastic problem is formulated as an infinite horizon average cost constrained Markov decision process (CMDP). Our solution approach to solve the CMDP is outlined as follows based on [10]:

- Using a Lagrangian formulation, we convert the CMDP into an unconstrained MDP parameterized with a Lagrange multiplier, and show the optimal policy for the CMDP can be expressed as a randomized mixture for two deterministic policies of the unconstrained MDP.
- For the unconstrained MDP, we characterize the structural properties of the optimal policy. Specifically, the optimal policy possesses a threshold-based structure with respect to the AoI at the device and the AoI at the destination. Such a threshold-based structure, as illustrated in Fig. 2., indicates the inherent tradeoff between the average AoI and the energy costs.
- By utilizing these structural properties and the Robbins-Monro algorithm, we propose a structure-aware low-complexity algorithm to obtain the optimal policy of the CMDP.

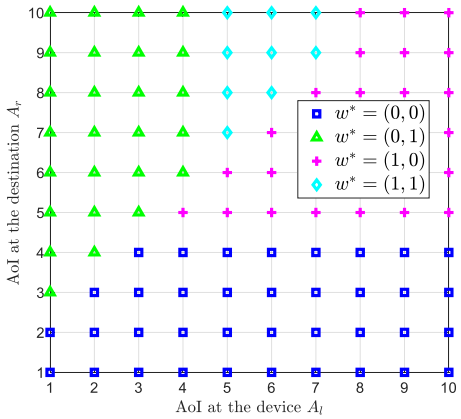


Fig. 2 Structure of the optimal policy for the unconstrained MDP under a given Lagrange multiplier and channel state [10]. $\boldsymbol{w}^*$= (sampling action, updating action) is the optimal action of the device.

Then, in [11], we have studied a more general scenario, in which multiple IoT devices sample the associated physical processed and send the status packet to a common destination through a shared wireless channel. Due to the exponential growth of the system state space with the number of devices, we focus on the design of a low-complexity suboptimal solution. Through a CMDP formulation, we develop a low-complexity semi-distributed learning algorithm with convergence guarantee to obtain a suboptimal sampling and updating policy so as to minimize the average AoI at the destination. The effectiveness of the proposed suboptimal is evaluated via extensive simulations in [11].

3. Joint Device Scheduling and Status Sampling with Non-uniform Status Packet Sizes

Next, we address the issue of the non-uniform status packet sizes. As illustrated in Fig. 3, consider a real-time IoT monitoring system with multiple IoT devices, which is similar to the one in Section 2. The major difference is that, we consider that for each device, a single status update may be composed of *multiple transmission packets*, and different devices may have different status packet sizes.

To avoid the collision among the transmissions from multiple devices, in each slot, the network has to decide which devices to be scheduled for updating their status. Note that, since multiple transmission slots are required to deliver a single status update, the current in-transmission status update could be obsolete and less useful for the destination. Thus, the network also needs to determine whether a scheduled device should continue its current in-transmission

update or sample and send a new one. Due to this distinct feature, for each device, in addition to the two concepts of AoI at the device and AoI at the destination, we need to introduce a particular system state to record the number of packets that are left to be sent to complete the transmission of the status update.

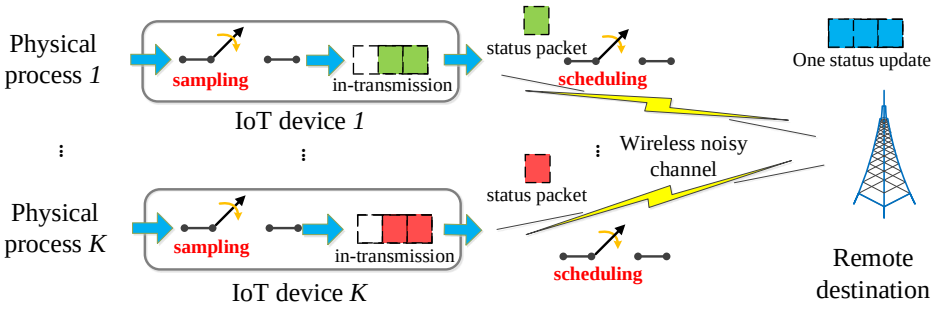


Fig. 3 Illustration of a real-time IoT monitoring system with non-uniform status packet sizes. Each status update is composed of more than one transmission packets.

We aim to jointly control the IoT device scheduling and status sampling processes to minimize the time-average AoI at the destination under non-uniform status update packet sizes. We formulate this problem as an infinite horizon average cost MDP. Our solution approach for the formulated MDP is outlines as follows based on [12].

- We characterize the structural properties of the optimal scheduling and sampling policy. Specifically, as shown in Fig. 4, the optimal policy is threshold-based with the AoI at each device. This means that, the device is more willing to sample and send a new status update, if the AoI at this device is larger. Such a threshold-based structure can be exploited to develop low-complexity optimal algorithms.
- To overcome the curse of dimensionality, we then develop a low-complexity suboptimal policy, by applying a linear decomposition method for the value function. The proposed policy offers significantly reduced complexity over the optimal algorithms and enjoys a similar structure to the optimal policy. Then, we develop a structure-aware algorithm to obtain this policy. The effectiveness of this policy is further demonstrated via extensive simulations in [13].

Using similar approaches, we extend the above framework to the IoT system in which the status updates randomly arrive at each IoT device. Similar structural properties of the optimal policy are characterized [13].

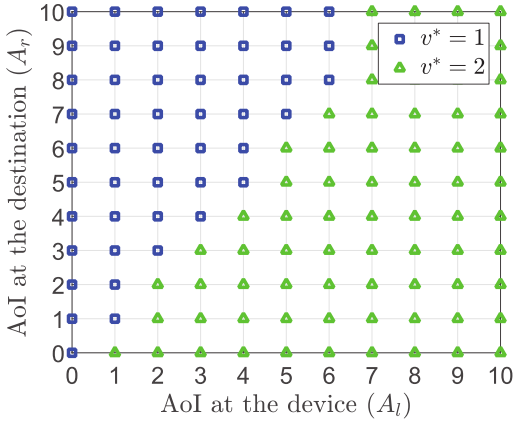


Fig. 4 Structure of the optimal policy in the single IoT device case [12]. $\boldsymbol{v}^*$ is the optimal sampling action.

4. Summary

In this e-letter, we have studied two optimization problems of minimizing the average AoI in IoT systems, by taking into account *the energy cost* pertaining to the sampling and updating processes, and the multi-time slot transmissions with *non-uniform sizes of the status updates*, respectively. To gain design insights for practical IoT systems, we have characterized that the structural properties of the optimal policies. To reduce the computational complexity, we have also proposed structure-aware low-complexity solutions. Simulation results have demonstrated the effectiveness of the proposed solutions in minimizing the average AoI. Future works will address extensions such as theoretically analyzing the performance of the proposed suboptimal policies and designing policies to for minimizing the AoI in

IoT monitoring systems with correlated underlying physical processes.

References

[1] W. Saad, M. Bennis, and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems", *arXiv preprint arXiv:1902.10265*, 2019.

[2] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, Orlando, FL, USA, March 2012, pp. 2731–2735.

[3] Y.-P. Hsu, "Age of information: Whittle index for scheduling stochastic arrivals," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Colorado, USA, June 2018.

[4] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," *IEEE/ACM Trans. Netw.*, vol. 26, no. 3, pp. 2637–2650, Dec 2018.

[5] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1807–1827, Mar 2019.

[6] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492–7508, Nov 2017.

[7] B. T. Bacinoglu and E. Uysal-Biyikoglu, "Scheduling status updates to minimize age of information with an energy harvesting sensor," *arXiv preprint arXiv:1701.08354*, 2017.

[8] X. Wu, J. Yang, and J. Wu, "Optimal status update for age of information minimization with an energy harvesting source," *IEEE Trans. Green Commun. and Netw.*, vol. 2, no. 1, pp. 193–204, March 2018.

[9] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial," *IEEE Commun. Surveys Tuts.*, to appear, 2019.

[10] B. Zhou and W. Saad, "Optimal sampling and updating for minimizing age of information in the Internet of Things," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, UAE, Dec. 2018.

[11] B. Zhou and W. Saad, "Joint status sampling and updating for minimizing age of information in the Internet of Things," *arXiv preprint arXiv:1807.04356*, 2018.

[12] B. Zhou and W. Saad, "Minimizing age of information in the Internet of Things with non-uniform status packet sizes," in *Proc. of IEEE International Conference on Communications (ICC)*, Shanghai, China, May 2019.

[13] B. Zhou and W. Saad, "Minimum Age of Information in the Internet of Things with Non-uniform Status Packet Sizes," *arXiv preprint arXiv:1901.07069*, 2019.

Bo Zhou (S'16, M'19) is currently a postdoctoral associate at the Bradley department of Electrical and Computer Engineering at Virginia Tech. He received his B.E. degree in electronic engineering from South China University of Technology, China in 2011 and his PhD from Shanghai Jiao Tong University, China in 2017. His research interests include age of information, wireless caching, stochastic network optimization, the Internet of things, and machine learning. He received the best paper award at IEEE GLOBECOM in 2018.

Walid Saad (S'07, M'10, SM'15, F'19) received his Ph.D degree from the University of Oslo in 2010. He is currently a Professor at the Department of Electrical and Computer Engineering at Virginia Tech, where he leads the Network Science, Wireless, and Security laboratory. His research interests include wireless networks, machine learning, game theory, security, unmanned aerial vehicles, cyber-physical systems, and network science. Dr. Saad is a Fellow of the IEEE and an IEEE Distinguished Lecturer. He is also the recipient of the NSF CAREER award in 2013, the AFOSR summer faculty fellowship in 2014, and the Young Investigator Award from the Office of Naval Research (ONR) in 2015. He was the author/co-author of seven conference best paper awards at WiOpt in 2009, ICIMP in 2010, IEEE WCNC in 2012, IEEE PIMRC in 2015, IEEE SmartGridComm in 2015, EuCNC in 2017, and IEEE GLOBECOM in 2018. He is the recipient of the 2015 Fred W. Ellersick Prize from the IEEE Communications Society, of the 2017 IEEE ComSoc Best Young Professional in Academia award, and of the 2018 IEEE ComSoc Radio Communications Committee Early Achievement Award. From 2015-2017, Dr. Saad was named the Stephen O. Lane Junior Faculty Fellow at Virginia Tech and, in 2017, he was named College of Engineering Faculty Fellow. He received the Dean's award for Research Excellence from Virginia Tech in 2019. He currently serves as an editor for the IEEE Transactions on Wireless Communications, IEEE Transactions on Mobile Computing, IEEE Transactions on Cognitive Communications and Networking, and IEEE Transactions on Information Forensics and Security. He is an Editor-at-Large for the IEEE Transactions on Communications.

Linear Network Coded Wireless Caching in Cloud Radio Access Network[1]

Long Shi, Kui Cai
Science and Math Cluster, Singapore University of Technology and Design, Singapore
slong1007@gmail.com; cai_kui@sutd.edu.sg

1. Introduction

In modern wireless networks such as cloud radio access network (C-RAN), fronthaul links are threatened by an alarming "digestive disease", due to the huge congestion caused by explosive growth of wireless traffic. How to alleviate the fronthaul congestion while meeting the peak traffic demands is a matter of great urgency. Recent research unveils that multimedia delivery is a driving factor of the wireless traffic, of which duplicate downloads of a few popular contents (e.g., music or videos) occupy a significant portion [1]. This finding drives us to reduce the redundant delivery through fronthaul to alleviate the traffic congestion. To deal with this challenge, caching revives and come into play in wireless networks. Following the spirit of web caching, one way is to employ memories distributed across the networks. Recently, wireless caching has been applied to a wide ranges of wireless networks [2-4]. Standing apart from web caching, wireless caching, operated in the physical layer, attains significant gains integrated with advanced physical-layer coding technologies.

Related Work

The seminal work in [5] proposed coded caching to investigate fundamental limits of cache-aided broadcasting networks. The bit-wise XOR network coding was employed in the delivery phase. Superior to uncoded caching, coded caching can explore the global caching gain from the coded multicasting transmission, in addition to the local caching gain [5]. Coded caching also manifests its benefits in wireless RANs [6,7]. The major challenges in the cache-aided wireless networks lie in the cache placement at transmitters and receivers and the interference management induced by different user requests in the content delivery. The ultimate goal of those works is to maximize DoF for the wireless coded caching networks, which in turn reduces the traffic burden over fronthaul.

Contributions

In this work, we address two issues that are not considered in the existing works on wireless coded caching. First, the bit-wise XOR network coding is not the optimal solution to accommodate the fading and interference, even in the wireless networks without cache [8,9]. The goal of this paper is to design the linear wireless network coding in the wireless caching by exploiting the characteristics of wireless channel and interference. This in turn brings the following issue. Second, the related works have not explored extra coding gain brought by the nature of wireless network coding, as the interference mitigation in the delivery phase mainly relies on the shared cache placement among different transmitters rather than the structure of wireless network coding. Targeting at the coding gain, interference management in the delivery phase catering to the wireless network coding operated caching remains open. To cope with these problems, we propose the linear wireless network coding operated wireless caching, referred to as linear network coded (NC) wireless caching, consisting of linear network coding assisted cache placement phase and signal-space alignment (SSA) enabled content delivery phase.

2. Cache-Aided C-RAN

To illustrate the proposed caching strategy, we consider a cache-aided C-RAN in Fig. 1 that consists of a central unit (baseband unit), K BSs (remote radio units), and M users. Each BS has N_T antennas, and each user is equipped with N_R antennas. The central unit owns M message vectors, where each vector contains multiple messages and represents a general file (e.g., text, music, video, etc.). Let $\boldsymbol{w}_{1\sim M} = [\boldsymbol{w}_1^T, \boldsymbol{w}_2^T, \dots, \boldsymbol{w}_M^T]^T$. All BSs are connected to a central unit through the error-free fronthaul links in the centralized manner. The local caches of finite size are equipped at BSs and users respectively. As depicted in Fig. 1(a), the wireless caching has two sequential phases.

Cache placement phase: All BSs and users have the access to the entire content in the central unit and prefetch some popular messages in their local caches, according to pre-assigned caching functions. Define $\boldsymbol{u}_k = \phi_k(\boldsymbol{w}_{1\sim M})$ and $\boldsymbol{v}_m = \psi_m(\boldsymbol{w}_{1\sim M})$ as the caching message vectors associated with the caching functions ϕ_k and ψ_m at BS k and user m, $k = 1,2,\dots,K$ and $m = 1,2,\dots,M$, respectively.

Content delivery phase: Each user m requests a message vector $\boldsymbol{w}_{\gamma_m}$ from the central unit, $\gamma_m \in \{1,2,\dots,M\}$. Let $\boldsymbol{\gamma} = [\gamma_1\ \gamma_2\ \dots\ \gamma_M]$ denote a request vector from all users, where γ_m corresponds to user m's request. In this phase, all BSs are informed of these requests and proceed by transmitting a function of caching messages over wireless channels.

[1]A short review for L. Shi, K. Cai, T. Yang, T. Wang, and J. Li, "Linear network coded wireless caching", submitted to IEEE Trans. Wireless Commun., under the 3rd round review.

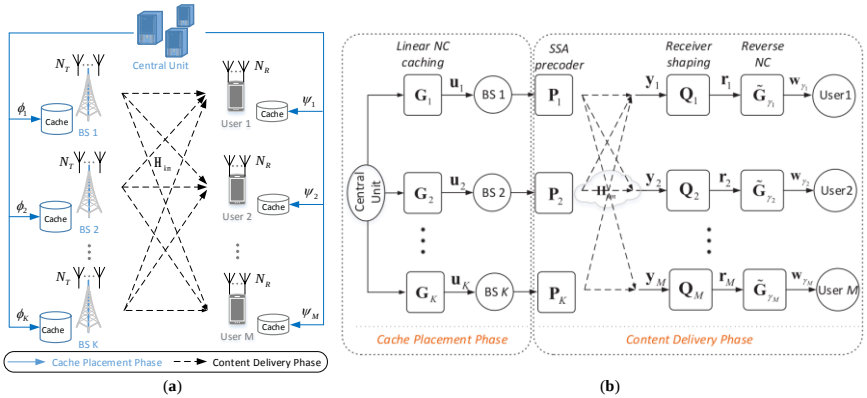


Figure 1 System models of (a) a cache-aided C-RAN and (b) the proposed linear NC wireless caching.

3. Linear Network Coding Assisted Cache Placement Phase

Let $\boldsymbol{w}_m = [w_{m,1}\ w_{m,2}\ \dots\ w_{m,L_m^{\mathrm{o}}}]^T$ denote the mth message vector in the central unit, $m = 1,2,\dots,M$. Consider that each element of $\boldsymbol{w}_m$ is drawn i.i.d. from a finite field $\mathbb{F}_q$, where the order q corresponds to the modulation cardinality. Let $L^{\mathrm{o}} = L_1^{\mathrm{o}} + \cdots + L_M^{\mathrm{o}}$ denote the total number of messages. In the cache placement phase, we design the linear network coding assisted caching functions ϕ_k at BS k to store a length-L_k^{b} caching message vector as

$$\boldsymbol{u}_k = \phi_k(\boldsymbol{w}_{1\sim M}) = \boldsymbol{G}_k \otimes \boldsymbol{w}_{1\sim M},\quad k = 1,2,\dots,K, \tag{1}$$

where $\boldsymbol{G}_k$ is the NC caching matrix of BS k and $\otimes$ denotes the multiplication operation in $\mathbb{F}_q$, i.e., $a \otimes b = ab \pmod q$. Let $\boldsymbol{g}_{k,l_k}$ denote the l_kth row vector of $\boldsymbol{G}_k$ and $g_{k,l_k}[j]$ denote the (l_k, j)th element in $\boldsymbol{G}_k$, respectively. We refer to $\boldsymbol{u}_k$ as the NC caching message vector stored at BS k. Let $\boldsymbol{u}_k = [u_{k,1}\ u_{k,2}\ \dots\ u_{k,L_k^{\mathrm{b}}}]$ with u_{k,l_k} being the l_kth NC caching message of $\boldsymbol{u}_k$, given by

$$u_{k,l_k} = \boldsymbol{g}_{k,l_k} \otimes \boldsymbol{w}_{1\sim M} = \bigoplus_{j=1}^{L^{\mathrm{o}}} (g_{k,l_k}[j] \otimes w_j), \tag{2}$$

where $\oplus$ denotes the addition operation in $\mathbb{F}_q$, i.e., $a \oplus b = a + b \pmod q$. From (1) and (2), BS k prefetchs L_k^{b} linear combinations of messages from the central unit rather than the messages themselves. The NC caching message vectors in (1) and (2) can be collectively expressed as $\boldsymbol{u}_{1\sim K} = [\boldsymbol{u}_1^T\ \boldsymbol{u}_2^T\ \dots \boldsymbol{u}_K^T] = \boldsymbol{G}_{1\sim K} \otimes \boldsymbol{w}_{1\sim M}$, where the joint NC caching matrix $\boldsymbol{G}_{1\sim K} = [\boldsymbol{G}_1^T\ \boldsymbol{G}_2^T\ \dots \boldsymbol{G}_K^T]^T$ has the full rank over $\mathbb{F}_q$. As shown in Fig. 1(b), the key design of placement phase lies in the joint NC caching matrix.

4. Signal-Space Alignment Enabled Content Delivery Phase

Each multi-antenna BS precodes their caching messages by a precoding matrix and broadcasts its precoded signals to all users simultaneously. The N_R-dimensional received signal vector at user m is given by

$$\boldsymbol{y}_m = \Sigma_{k=1}^{K} \boldsymbol{H}_{k,m}\, \boldsymbol{P}_k\, \boldsymbol{u}_k + \boldsymbol{z}_m, \tag{4}$$

where $\boldsymbol{P}_k$ denotes the precoding matrix at BS k with size of $N_T \times L_k^{\mathrm{b}}$. Let $\boldsymbol{p}_{k,l_k}$ denote the l_kth column of $\boldsymbol{P}_k$. Define the set that collects the vectors corresponding to the signal spaces of the NC caching messages received at user m as $\mathcal{V}_m = \{\mathcal{V}_{1,m}, \mathcal{V}_{2,m}, \dots, \mathcal{V}_{K,m}\}$, where the subset $\mathcal{V}_{1,m} = \{\boldsymbol{H}_{k,m}\, \boldsymbol{p}_{k,1}\ \boldsymbol{H}_{k,m}\, \boldsymbol{p}_{k,2}\ \dots\ \boldsymbol{H}_{k,m}\, \boldsymbol{p}_{k,L_k}\}$ with the vector $\boldsymbol{H}_{k,m}\, \boldsymbol{p}_{k,l_k}$ corresponding to signal space of the NC caching message u_{k,l_k}. Now it is clear that the size of signal spaces exceeds the spatial dimension of the received signal at each user. Under the dimension constraint at each user, the precoders should be jointly designed to deliberately align signal spaces of some desired NC caching messages at each user.

However, all BSs broadcast different NC caching message vectors to all users simultaneously, and the interference at each user comes from the transmission of requested messages by all the other users. Since each NC caching message is a linear combination of multiple user requests, it is not possible for each user to extract the requested messages from the caching message by interference alignment. Following the SSA designs in [10,11], we propose a new SSA pattern for the linear NC wireless caching to align the desired NC caching messages. The SSA for the wireless network coding advances interference alignment by exploring the structure of the NC operation. As shown in Fig. 1(b), the key designs of placement phase include the precoder, the receiver shaping, and reverse NC operation, respectively.

Towards this end, we first design a binary "bin" matrix to indicate the NC caching messages that should be aligned at each user, and then prove the existence of precoding matrices that realize the SSA in each bin. To determine the joint NC caching matrix, we select a proper number of rows from the bin matrix. Thus, the bin design is the core of the linear NC wireless caching, which bridges the cache placement and content delivery via the linear network coding.

After that, each user deploys the receiver shaping and reverse NC operation to decode its requested messages.

5. Numerical Results

To assess the sum DoF, we adopt the achievable sum rate analysis studied in [10]. We note that the sum DoF corresponds to the scaling factor of the sum rate as the SNR goes high. Fig. 2 plots the achievable sum rates of the proposed caching schemes with $K = M = 2, N_T = N_R = 3$ and $K = M = 3, N_T = 6, N_R = 3$ under the worst-case caching scenario [7], respectively. Fig. 2(a) shows that the proposed scheme can achieve the sum DoF of 5 with different user requests $\{\gamma_1, \gamma_2\} = \{1,2\}$ and $\{2,1\}$. Fig. 2(b) shows that the proposed scheme can achieve the sum DoF of 7 under different user requests $\{\gamma_1, \gamma_2, \gamma_3\} = \{1,2,3\}, \{2,3,1\}$, and $\{3,1,2\}$.

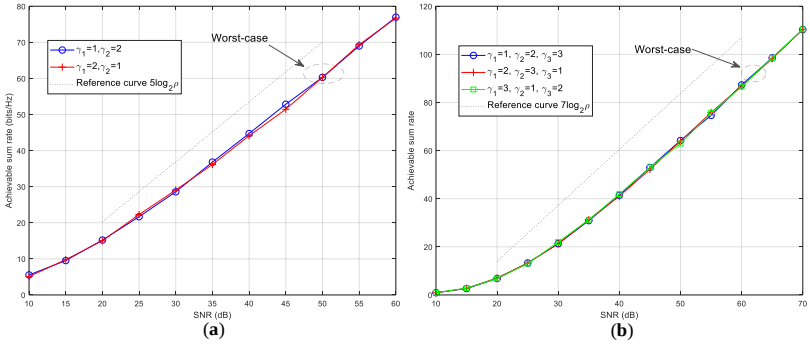


Figure 2 Achievable sum rates of the proposed wireless caching schemes with (a) $\boldsymbol{K = M = 2, N_T = N_R = 3}$ and (b) $\boldsymbol{K = M = 3, N_T = 6, N_R = 3}$.

5. Conclusion and Future Directions

We have proposed the linear wireless network coding operated caching network. In the cache placement phase, each BS stores the NC caching messages as a form of linear network coding. In the content delivery phase, we designed the SSA enabled precoding matrix to align the desired NC caching messages. With the receiver shaping and reverse NC operation, the proposed scheme can deal with distinct user requests using an invariant cache placement. Several interesting directions follow this work. First, this paper shows that the linear NC wireless caching is applicable in C-RAN. It is of interest to generalize the spirit of linear NC wireless caching into other wireless networks. Second, consider that each user randomly and independently pre-downloads the content without the central coordination in the placement phase. How to design the linear wireless NC aided decentralized caching deserves further investigation.

References

[14] G. Paschos, E. Bastug, I. Land, G. Caire, and M. Debbah, "Wireless caching: technical misconceptions and business barriers," IEEE Commun. Mag., vol. 54, no. 8, pp. 16-22, Aug. 2016.
[15] J. Li, H. Chen, Y. Chen, Z. Lin, B. Vucetic, and L. Hanzo, "Pricing and resource allocation via game theory for a small-cell video caching system," IEEE J. Sel. Areas Commun., vol. 34, no. 8, pp. 2115--2129, Aug. 2016.
[16] M. Ji, G. Caire, and A. F. Molisch, "Wireless device-to-device caching networks: basic principles and system performance," IEEE J. Sel. Areas Commun., vol. 34, no. 1, pp. 176-189, Jan. 2016.
[17] M. Tao, E. Chen, and W. Yu, "Content-centric sparse multicast beamforming for cache-enabled cloud RAN," IEEE Trans. Wireless Commun., vol. 15, no. 9, pp. 6118-6131, Sept. 2016.
[18] M. A. Maddah-Ali and U. Niesen, "Fundamental Limits of Caching," IEEE Trans. Inf. Theory, vol. 60, no. 5, pp. 2856-2867, May 2014.
[19] J. Hachem, U. Niesen, and S. Diggavi, "Degrees of freedom of cache-aided wireless interference networks," IEEE Trans. Inf. Theory, vol. 64, no. 7, pp. 5359-5380, Apr. 2018.
[20] Y. Cao, M. Tao, F. Xu, and K. Liu, "Fundamental storage-latency tradeoff in cache-aided MIMO interference networks," IEEE Trans. Wireless Commun., vol. 16, no. 8, pp. 5061-5076, Aug. 2017.
[21] L. Shi, S. C. Liew, and L. Lu, "On the subtleties of q-PAM linear physical-layer network coding," IEEE Trans. Inf. Theory, vol. 63, no. 8, pp. 2520-2544, May 2016.
[22] L. Shi and S. C. Liew, "Complex linear physical-layer network coding," IEEE Trans. Inf. Theory, vol. 62, no. 5, pp. 4949-4981, Aug. 2017.
[23] T. Yang, "Distributed MIMO broadcasting: reverse compute-and-forward and signal space alignment," IEEE Trans. Wireless Commun., vol. 16, no. 1, pp. 581-593, Jan. 2017.
[24] N. Lee, J.-B. Lim, and J. Chun, "Degrees of freedom of the MIMO Y channel: signal space alignment for network coding," IEEE Trans. Inf. Theory, vol. 56, no. 7, pp. 3332-3342, Jul. 2010.

Multiuser Computation Offloading in MEC Systems with Virtualization

Yuan Liu
School of Electronic and Information Engineering, South China University of Technology
eeyliu@scut.edu.cn

1. Introduction

Mobile-edge computing (MEC) is an emerging technology for enhancing the computational capabilities of the mobile devices and reducing their energy consumption via offloading complex computation tasks to the nearby servers [1], [2]. An essential technology for implementing multiuser MEC at servers is virtualization, referring to sharing of a physical machine (server) by multiple computing processes via creating virtual machines (VMs). Specifically, each VM is a virtual computer configured with a certain amount of the server's hardware resource (such as CPU, memory and I/O bus). Through virtualization, applications are deployed and executed within VMs that are well-isolated with others. To this end, the MEC server can isolate co-located applications and provide multi-service support. Nevertheless, it has been shown in the literature [3]–[5] that sharing the same physical platform can incur the so-called I/O interference among VMs, resulting in a certain degree of computation-speed reduction for each VM. As far as we know, prior research of this issue focuses on the interference modeling [4]–[6] and computation resource provisioning [7]. No previous works related to the computation offloading coupled with joint radio-and-computational resource allocation (RCRA) have been done before.

Omitting I/O interference in multiuser MEC based on virtualization leads to the unrealistic assumption that the total computation resource at a server remains fixed regardless of the number of VMs. In reality, the resource reduces as the number grows due to I/O interference. Thus, the number of VMs per server is usually constrained in practice, so as to maintain the efficiency in resource utilization. Despite its importance, I/O interference has received little attention in the literature. It motivates the current work on accounting for the factor in resource allocation for MEC systems.

Thus we study the RCRA problem in multiuser MEC systems in the presence of I/O interference. The I/O interference is modelled using a practical model developed based on measurement data [6]. Considering I/O interference introduces a *dilemma*: scheduling more offloading users increases the multiplexing gain in parallel computing but degrades the speeds of individual VMs due to their interference. Our key contributions are summarized as follows: 1) We formulate the problem of sum offloading rate maximization by joint offloading-user scheduling, offloaded-data size control, and communication-and-computation time division. 2) We propose an optimal algorithm with low complexity based on a decomposition approach and Dinkelbach method.

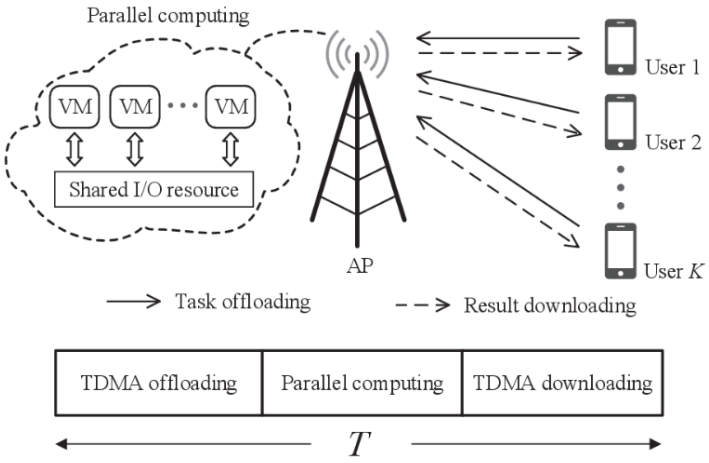


Figure 1 The considered multiuser MEC system.

2. System Model and Problem Formulation

Consider an MEC system shown in Fig. 1, consisting of one access point (AP) integrated with an MEC server, and K users within the set $\mathcal{K} = \{1, \dots K\}$. Partial offloading is assumed so that each user can partition its computation task into two independent parts for local computing and offloading to the server. All the users have to complete their tasks within a fixed duration T (in second) so as to meet a real-time requirement. The system operation is divided into *three*

sequential phases: 1) TDMA-based task offloading by users, 2) parallel computing at the server, and 3) TDMA-based computation-result downloading from the server. Corresponding models and assumptions are described as follows.

1) *Offloading and downloading phases*: We assume the uplink and downlink transmission rates are fixed within duration T and thus the offloading delay (t_i^u) and result-downloading delay (t_i^d) at each scheduled user can be formulated a linear function of user's offloaded data (l_i).
2) *Parallel-computing phase with virtualization:* After receiving all the offloaded tasks, the server executes them in parallel by creating multiple VMs. We consider the important factor of I/O interference in parallel computing [8] and adopt a model developed in the literature based on measurement data [1], [6]. Specifically, let $S \subseteq \mathcal{K}$ denote the set of scheduled offloading users, t_e the time allocated to the parallel-computing phase, and r_i the expected computing rate (bits/sec) of a VM given task i when running in isolation. Following [1], [6], a performance degradation factor $d > 0$ is defined to specify the percentage reduction in the computing rate of a VM when multiplexed with another VM. Suppose that one VM is created and assigned to a task. Then, for a given t_e, the numbers of offloadable bits are constrained by

$$0 \le l_i \le t_e r_i (1+d)^{1-|S|}, \quad \forall i \in S \tag{1}$$

The constraints in (1) show that the maximum number of offloadable bits per user decreases with the number of offloaded tasks (or offloading users) due to the I/O interference in parallel computing. Moreover, relaxing the duration for parallel computing t_e can accommodate more offloaded bits $\{l_i\}$, however, at the cost of less time for the offloading and downloading phases. This introduces a tradeoff between the three phases under the total-latency constraint $\sum_{i\in S} t_i^u + t_e + \sum_{i\in S} t_i^d \le T$.

Our problem can be described as maximizing the weighted sum of the users' offloading rates under the latency constraint, by joint offloading-user scheduling, offloaded-bits control, and three-phase time allocation. Here, the sum offloading rate is defined as the sum offloadable bits over the time duration T.

3. Optimal Algorithm

The formulated RCRA problem is a mixed-integer programming problem comprising both a combinatorial variable (S) and continuous variables ($\{l_i\}$, t_e) and non-convex constraints (1), which is NP-hard in general. By analyzing the problem properties, we propose a solution approach of decomposing the problem into *master and slave sub problems*. The master problem is to find the optimal number of offloading users. Given the number, each slave subproblem is to optimize offloading-user set, offloaded-data sizes, and time division (offloading, computing, downloading). By adopting Dinkelbach method, an efficient iteratively algorithm is designed to solve the slave problem that is a combinatorial-optimization problem. With the algorithm, the master problem can be then solved by a simple search over a finite integer set of possible numbers of offloading users. The details of the proposed algorithm can be found in [9].

4. Simulation Results

In this section, we provide simulation results to evaluate the proposed algorithm in comparison of three benchmark algorithm: *greedy method*, in which the offloading-user set S is obtained by selecting users in the descending order of the transmission rate until latency constraint is valid, *Linear Programming Relaxation (LR)*, in which S is obtained by solving K slave problems using linear programming relaxation, *All-Offloading*, in which all the users are scheduled to offload.

In Fig. 2, we compare the sum offloading rate performance of different algorithms when the number of users K varies from 4 to 12, where d is set as 0.1. We can see that the sum offloading rate is increasing with K for the optimal, LR and greedy algorithms, while for the scheme that all users offload, it grows slowly when $K = 10$ and begins to decrease afterwards. This is because the former three algorithms have more flexible user-scheduling schemes to balance the degradation impact caused by I/O interference and thus have superior system performance. In contrast, the last algorithm with no control on the number of offloading users, will suffer more severe performance degradation

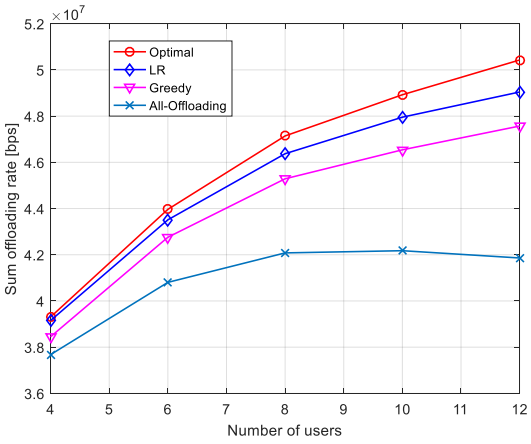


Figure 2 Sum offloading rate versus K.

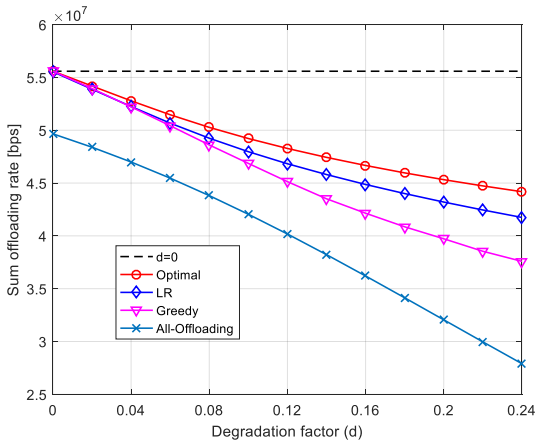


Figure 3 Sum offloading rate versus d.

as K increases. Besides, it is observed that the optimal algorithm outperforms the benchmark algorithms especially when K is large. For instance, when $K = 12$, the optimal algorithm obtains about 3%, 6%, and 20% performance improvements over the three benchmarks, respectively.

In Fig. 3, we illustrate the relationship between the degradation factor d and the sum offloading rate performance, where $K = 10$. As expected, the sum offloading rate is decreasing with d in all considered algorithms while the descending rate of the optimal algorithm is the slowest. This indicates that our proposed algorithm has the best performance resistance against the I/O interference effect. One observes that, the performance of LR and greedy algorithms is close-to-optimal when d is small. This coincides with the result of special case that when the degradation factor d is zero, the optimal solution can be obtained by the greedy approaches. On the other hand, the line of $d = 0$ can been seen as the sum offloading rate of the conventional case without considering the I/O interference issue. Its performance gap with the optimal algorithm can be interpreted as the overestimation of the system performance built on the optimistic assumption of no I/O interference.

5. Conclusion

We studied joint radio-and-computation resource allocation in a multiuser MEC system, where the computation interference issue has been considered. We formulated the problem of sum offloading rate maximization by joint offloading scheduling, offloaded-data sizes, and communication-and-computation time division. We proposed an optimal solution with low complexity to solve the non-convex problem. Simulation results demonstrated that our proposed algorithm achieves superior performance gain compared with the benchmark schemes.

References

[25] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, Fourthquarter 2017.

[26] European Telecommunications Standards Institute (ETSI), "Mobile-edge Computing-Introductory technical white paper," Sep. 2014.

[27] X. Pu, L. Liu, Y. Mei, S. Sivathanu, Y. Koh, and C. Pu, "Understanding performance interference of I/O workload in virtualized cloud environments," in *IEEE Cloud*, 2010, pp. 51–58.

[28] S. Ibrahim, B. He, and H. Jin, "Towards pay-as-you-consume cloud computing," in *IEEE SCC*, 2011, pp. 370–377.

[29] S.-g. Kim, H. Eom, and H. Y. Yeom, "Virtual machine consolidation based on interference modeling," *J. Supercomput.*, vol. 66, no. 3, pp. 1489–1506, Dec. 2013.

[30] D. Bruneo, "A stochastic model to investigate data center performance and QoS in IaaS cloud computing systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 3, pp. 560–569, Mar 2014.

[31] W. B. Slama, Z. Brahmi, and M. M. gammoudi, "Interference-aware virtual machine placement in cloud computing system approach based on fuzzy formal concepts analysis," in *IEEE WETICE*, 2018, pp. 48–53.

[32] X. Pu, L. Liu, Y. Mei, S. Sivathanu, Y. Koh, C. Pu, and Y. Cao, "Who is your neighbor: Net I/O performance interference in virtualized clouds," *IEEE Trans. Serv. Comput*, vol. 6, no. 3, pp. 314–329, Jul. 2013.

[33] Z. Liang, Y. Liu, T.-M. Lok, and K. Huang, "Multiuser computation offloading and downloading for edge computing with

virtualization," [Online]. Available: https://arxiv.org/pdf/1811.07517.pdf.

Yuan Liu received the B.S. degree from Hunan University of Science and Technology, Xiangtan, China, in 2006; the M.S. degree from Guangdong University of Technology, Guangzhou, China, in 2009; and the Ph.D. degree from Shanghai Jiao Tong University, China, in 2013, all in electronic engineering. Since Fall 2013, he has been with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou, where he is currently an associate professor. Dr. Liu serves as an editor for the IEEE Communications Letters and the IEEE Access. His research interests include 5G communications and beyond, mobile edge computation offloading, and machine learning in wireless networks.

SPECIAL ISSUE ON Artificial Intelligence and Machine Learning for Network Resource Management and Data Analytics

Guest Editor: Longwei Wang
Auburn Univeristy, AL USA
allenwang163@gmail.com

This special issue of Frontiers focuses on the recent progresses of application of artificial intelligence and machine learning in network and data analytics. Various research groups all around the world are currently working on artificial intelligence enabled network management and data analytics, which have recently also attracted the interest of the industry.

The first paper of the issue applies deep reinforcement learning (DRL) to tackle the congestion control problem. It proposes TCP-Drinc framework that offers effective solutions to several long-existing problems in congestion control: delayed environment, partial observable information, and measurement variations.

In the second paper, issues related physical layer wireless communications are discussed. Authors introduce two DL-aided key techniques, i.e., automatic modulation classification (AMC) and fast beamforming in physical wireless communications. AMC represents a supervised classification task with better performance than traditional algorithms, and fast beamforming is a typical unsupervised regression task, which is more effective and less complex than traditional algorithms at a cost of slight performance loss.

The third paper is about the information theory inspired multimodal data fusion. The authors give a short review of information theory inspired multimodal data fusion methods in literature. Three different methods are covered: Mutual Information Based Multimodal Data, Fusion Nature Encoded Multimodal Data Fusion Information, Resonance Based Multimodal Data Fusion.

LONGWEI WANG is currently with Auburn University. His current research interest includes statistical signal processing and machine learning, with applications in intelligent sensing, network optimization and data analytics.

TCP-Drinc: Smart Congestion Control Based on Deep Reinforcement Learning[2]

Kefan Xiao, Shiwen Mao, and Jitendra K. Tugnait
Dept. Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201 U.S.A.
E- KZX0002@tigermail.auburn.edu, smao@ieee.org, tugnajk@eng.auburn.edu

1. Introduction and Motivation

The unprecedented growth of network traffic, in particular, mobile network traffic, has greatly stressed today's Internet. Although the capacities of wired and wireless links have been continuously increased, the gap between user demand and what the Internet can offer is actually getting wider. Furthermore, many emerging applications not only require high throughput and reliability, but also low delay. Although the brute-force approach of deploying wired and wireless links with a higher capacity helps to mitigate the problem, a more viable approach is to revisit the higher layer protocol design, to make more efficient use of the increased physical layer link capacity.

Congestion control is the most important networking function of the transport layer, which ensures reliable delivery of application data. However, the design of a congestion control protocol is highly challenging. First, the transport network is an extremely complex and large-scale network of queues. The TCP end host itself consists of various interconnected queues in the kernel. When the TCP flow gets into the Internet, it traverses various queues at routers/switches along the end-to-end path, each shared by cross traffic (e.g., other TCP flows and UDP traffic) and served with some scheduling discipline. Significant efforts are still needed to gain good understanding of such a complex network to develop the queueing network theory that can guide the design of a congestion control protocol. Second, if following the end-to-end principle, agents at end hosts have to probe the network state and make independent decisions without coordination. The detected network state is usually error-prone and delayed, and the effect of an action is also delayed and depends on the actions of other competing hosts. Third, if to involve routers, the algorithm must be extremely simple (e.g., stateless) to ensure scalability, since the router may handle a huge amount of flows. Finally, as more wireless devices are connected, the lossy and capacity-varying wireless links also pose great challenges to congestion control design.

Many effective congestion control protocols have been developed in the past three decades since the pioneering work [1]. However, many existing schemes are based on some fundamental assumptions. For example, early generation of TCP variants assume that all losses are due to buffer overflow, and use loss as indicator of congestion. Since such assumption does not hold true in wireless networks, many heuristics have been proposed for TCP over wireless to distinguish the losses due to congestion from that incurred by link errors. Moreover, many existing schemes assume a single bottleneck link in the end-to-end path, and the wireless last hop (if there is one) is always the bottleneck. Given the high capacity wireless links and the complex network topology/traffic conditions we have today [2], such assumptions are less likely to be true. The bottleneck could be at either the wired or wireless segment, it could move around, and there could be more than one bottlenecks. Finally, when there is a wireless last hop, some existing work [3] assumes no competition among the flows at the base station (BS), which, as shown in [4], may not be true due to coupled wireless transmission scheduling at the BS.

2. TCP-Drinc Design and Contributions

In [15], we aim to develop a smart congestion control algorithm that does not rely on the above assumptions. Motivated by the recent success of applying machine learning to wireless networking problems [5], and based on our experience of applying deep learning (DR) and deep reinforcement learning (DRL) to 5G mmWave networks [6], edge computing and caching [7]-[9], and RF sensing and indoor localization [10]-[12], we propose to develop a model-free, smart congestion control algorithm based on DRL. The original methods that treat the network as a white box have been shown to have many limitations. To this end, machine learning, in particular, DRL, has a high potential in dealing with the complex network and traffic conditions by learning from past experience and extracting useful features. A DRL based approach also relieves the burden on training data, and has the unique advantage of being adaptive to varying network conditions.

[2] A short review for [15]. This work was supported in part by the NSF under Grant CNS-1702957, and by the Wireless Engineering Research and Education Center (WEREC), Auburn University, Auburn, AL, USA.

In particular, we present TCP-Drinc in [15], acronym for Deep reinforcement learning based congestion control. TCP-Drinc is a DRL based agent that is executed at the sender side. The TCP-Drinc architecture is presented in Fig. 1. The agent estimates features such as congestion window difference, round trip time (RTT), the minimum RTT over RTT ratio, the difference between RTT and the minimum RTT, and the inter-arrival time of ACKs, and stores historical data in an experience buffer. Then the agent uses a deep convolutional neural network (DCNN) concatenated with a long short term memory (LSTM) network to learn from historical data and select the next action to adjust the congestion window size (see Fig. 2).

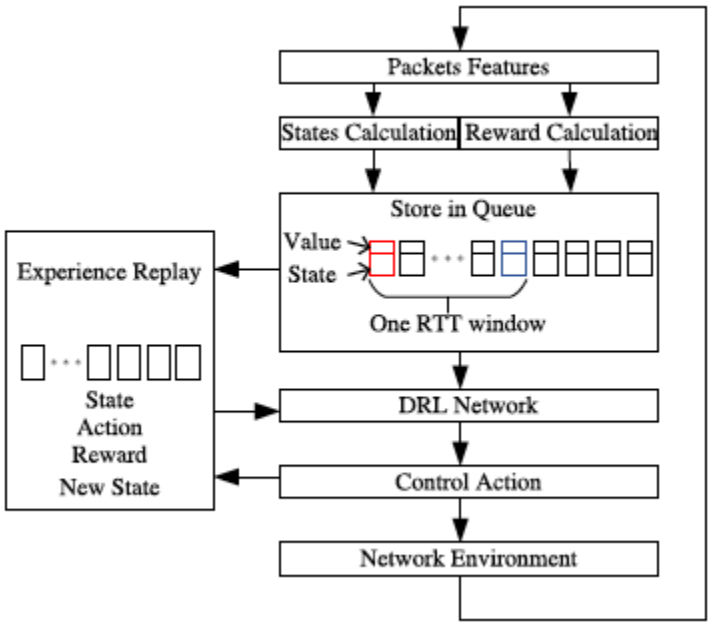


Figure 1 The proposed TCP-Drinc system architecture.

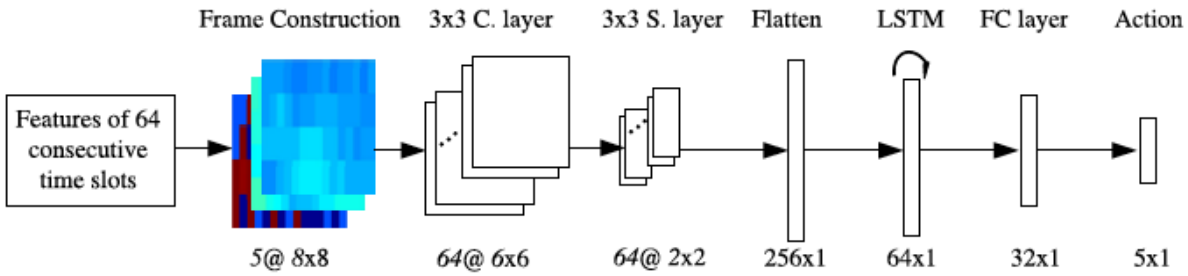


Figure 2 Design of the proposed DCNN (in the figure, "C." represents the convolutional layer, "S." represents the down sampling (pooling) layer, "FC" means fully connected).

The contributions of the work [15] are summarized as follows.

1) To the best of our knowledge, [15] is the first work that applies DRL to tackle the congestion control problem. Specifically, we propose a DRL based framework on (i) how to build an experience buffer to deal with the delayed environment, where an action will take effect after a delay and feedbacks are also delayed, (ii) how to handle the multi-agent competition problem, and (iii) how to design and compute the key components including states, action, and reward. We believe this framework could help to boost the future research on smart congestion control protocols.

2) The proposed TCP-Drinc framework also offers effective solutions to several long-existing problems in congestion control: delayed environment, partial observable information, and measurement variations. We apply DCNN as a filter to extract stable features from the rich but noisy measurements, instead of using exponential window moving average (EWMA) as a coarse filter as in previous works. Moreover, LSTM is utilized to handle the autocorrelation within the time-series introduced by delay and partial information that an agent senses.

3) We develop a realistic implementation of TCP-Drinc on the ns-3 [13] and TensorFlow [14] platforms. The DRL agent is developed with TensorFlow and the training and inference interfaces are built in ns-3 using TensorFlow C++. We conduct an extensive simulation study with TCP-Drinc and compare with five representative benchmark schemes, including both loss based and latency based TCP variants. TCP-Drinc achieves superior performance in throughput and RTT in all the simulations, and exhibits high adaptiveness and robustness under dynamic network environments.

3. Experimental Results

We examine the performance of TCP-Drinc and the baseline schemes under dynamic network settings. In particular, the simulation is executed 100 times, each lasting for 500s. The number of users is 5. The bottleneck capacity is varied at a frequency of 10 Hz; each capacity is randomly drawn from a uniform distribution in [5, 15] Mbps. The propagation delay is also varied at a 10 Hz frequency and each value is randomly drawn from a uniform distribution in [0:06, 0:16]s. In Fig. 10, we plot the combined RTT (*x*-axis) and throughput (*y*-axis) results in the form of of 95% confidence intervals. That is, we are 95% sure that the throughput and RTT combination of each scheme are located within the corresponding oval area. We find that TCP-Drinc achieves a comparable throughput performance with the loss based protocols, e.g., TCP-Cubic and TCP-NewReno. Furthermore, TCP-Drinc achieves a much lower RTT performance than the loss based protocols, e.g., at least 46% lower than TCP-NewReno and 65% lower than TCP-Cubic. Furthermore, TCP-Drinc achieves an over 100% throughput gain than TCP-Vegas at the cost of an only 15% higher RTT.

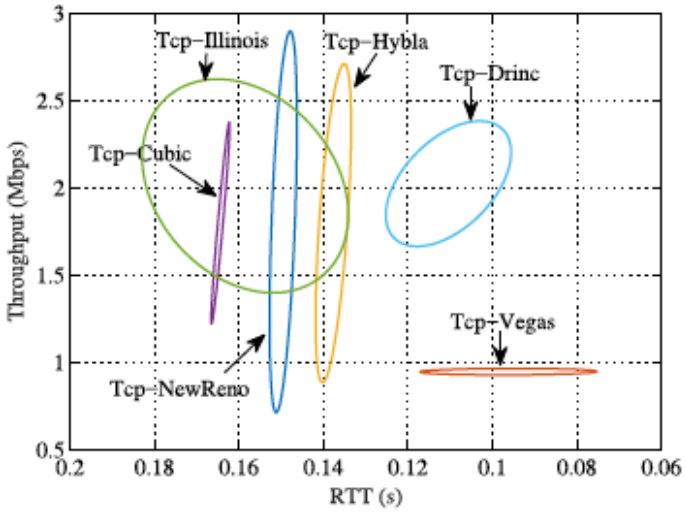


Figure 3 Throughput and RTT of the TCP variants under randomly varied network parameters. Each oval area represents the 95% confidence interval.

To study the fairness performance of the algorithms, we evaluate the Jain's index they achieve in the simulation. The average fairness index and the corresponding 95% confidence intervals are presented in Table 1. TCP-Vegas and TCP-Illinois achieve the best fairness performance among all the algorithms. TCP-Drinc can still achieve a considerably high fairness index (only 1.9% lower than the best). Note that the best fairness performance of TCP Vegas is achieved at the cost of a much poorer throughput performance. It is also worth noting that the 95% confidence interval of TCP-Drinc is the smallest among all the schemes, which is indicative of its robustness under varying network conditions.

Table I Jain's Fairness Index Achieved by the Congestion Control Schemes

	TCP-Cubic	TCP-Hybla	TCP-Illinois	TCP-NewReno	TCP-Vegas	TCP-Drinc
Average fairness index	0.7873	0.8025	0.8125	0.7562	0.8214	0.8058
95% Confidence Interval	[0.7005, 0.8741]	[0.7008, 0.9042]	[0.7114, 0.9136]	[0.6284, 0.8839]	[0.7115, 0.9313]	[0.7233, 0.8882]
Confidence Interval Span	0.1736	0.2034	0.2022	0.2555	0.2198	0.1649

4. Conclusions

In [15], we developed a framework for model-free, smart congestion control based on DRL. The proposed scheme

does not require accurate models for network, scheduling, and network traffic flows; it also does not require training data, and is robust to varying network conditions. The detailed design of the proposed TCP-Drinc scheme was presented and the trade-offs were discussed. Extensive simulations with ns-3 were conducted to validate its superior performance over five benchmark algorithms.

References

[34] V. Jacobson, "Congestion avoidance and control," *ACM SIGCOMM Comput. Commun. Rev*., vol. 18, no. 4, pp. 314-329, Aug. 1988.

[35] Y. Zhao, B. Zhang, C. Li, and C. Chen, "ON/OFF traffic shaping in the Internet: Motivation, challenges, and solutions," *IEEE Netw*., vol. 31, no. 2, pp. 48-57, Mar./Apr. 2017.

[36] K. Winstein, A. Sivaraman, and H. Balakrishnan, "Stochastic forecasts achieve high throughput and low delay over cellular networks," in *Proc. USENIX NSDI*, Lombard, IL, USA, Apr. 2013, pp. 459-471.

[37] Y. Zaki, T. Potsch, J. Chen, L. Subramanian, and C. Gorg, "Adaptive congestion control for unpredictable cellular networks," *ACM SIGCOMM Comput. Commun. Rev*., vol. 45, no. 4, pp. 509-522, Oct. 2015.

[38] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao. (Sept. 2018). "Application of machine learning in wireless networks: Key techniques and open issues." [Online]. Available: https://arxiv.org/abs/1809.08707

[39] M. Feng and S. Mao, "Dealing with limited backhaul capacity in millimeter wave systems: A deep reinforcement learning approach," *IEEE Commun.*, vol.57, no.3, pp.50-55, Mar. 2019.

[40] X. Chen, H. Zhang, C. Wu, S. Mao, Y. Ji, and M. Bennis, "Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning," *IEEE Internet Things J*., to be published. [Online]. Available: https://ieeexplore.ieee.org/document/8493155

[41] Y. Sun, M. Peng, and S. Mao, "Deep reinforcement learning based mode selection and resource management for green fog radio access networks," *IEEE Internet Things J*., vol.6, no.2, pp.1960-1971, Apr. 2019.

[42] Z. Chang, L. Lei, Z. Zhou, S. Mao, and T. Ristaniemi, "Learn to cache: Machine learning for network edge caching in the big data era," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 28-35, June 2018.

[43] X. Wang, X. Wang, and S. Mao, "RF sensing in the Internet of Things: A general deep learning framework," *IEEE Commun*., vol. 56, no. 9, pp. 62-69, Sept. 2018.

[44] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763-776, Jan. 2017.

[45] W. Wang, X. Wang, and S. Mao, "Deep convolutional neural networks for indoor localization with CSI images," *IEEE Trans. Netw. Sci. Eng*., to be published. [Online]. Available: https://ieeexplore.ieee.org/document/8468057

[46] G. F. Riley and T. R. Henderson, ``The ns-3 network simulator," in *Modeling and Tools for Network Simulation*, K.Wehrle, M. Gunes, and J. Gross, Eds. Berlin, Germany: Springer, 2010, pp. 15-34.

[47] M. Abadi et al., ``Tensorflow: A system for large-scale machine learning," in *Proc. USENIX OSDI*, Savannah, GA, USA, Nov. 2016, pp. 265-283.

[48] K. Xiao, S. Mao, and J.K. Tugnait, "TCP-Drinc: Smart congestion control based on deep reinforcement learning," *IEEE Access Journal*, Special Section on Artificial Intelligence and Cognitive Computing for Communications and Networks, vol.7, no.1, pp.11892-11904, Jan. 2019. DOI: 10.1109/ACCESS.2019.2892046.

Briefly Introduction of Deep Learning for Physical Layer Wireless Communications[3]

Guan Gui[1], Yu Wang[1], and Jinlong Sun [1]
[1]*Nanjing University of Posts and Telecommunications, Nanjing, China*
guiguan@njupt.edu.cn

Abstract
Current communication systems cannot meet future demands such as ultra-high speed, low latency, high reliability, and massive access. Hence, extensive researches are focusing on next generation wireless communications. Deep learning (DL) is recently considered a powerful and effective tool in mining deep-level structures from massive data, and it can be applied to optimize the overall system using large amounts of available historical data. In this article, we first introduce future challenges of wireless communications, and review some proposed DL-aided techniques in physical layer. Then, we present two DL-aided techniques, and compare them with traditional algorithms. Finally, future challenges and opportunities are pointed out. We believe that DL-aided physical layer wireless techniques will play important roles in future wireless communications.

1. Introduction

Current wireless communication systems are challenged by explosive growth in incremental data, high-speed streams, and low-latency communication requirements. Existing wireless communication techniques are weak when facing the popularization of smart terminals, the rapid development of Internet of Things (IoT), the breakthrough of artificial intelligence (AI), and the boost of big data. In addition, existing communication theories have inherent limitation in utilizing complex structural information and processing massive data. Therefore, new communication theories and techniques need to be established to meet the requirements of future wireless communication systems [1].

In recent years, deep learning (DL) is considered as one of the most powerful tools in numerous fields, because DL is expert in automatically extracting structural information from huge amounts of data. As a result, DL has been applied in physical layer wireless communications [2]-[4] and IoT [5]-[9]. We are also engaging in researches in the field of DL-aided physical layer wireless communications. In [10], a long short-term memory (LSTM) network was applied into typical non-orthogonal multiple access (NOMA) system to enhance spectral efficiency. In [11], we introduced DL into a massive multi-input multi-output (MIMO) system for super-resolution channel estimation and direction of arrival (DOA) estimation. In [12], [13], we designed a novel model-driven deep learning architecture termed as message passing (MP)-Net for resource allocation, and it achieved a great success. In [14], we proposed an effective automatic modulation classification (AMC) method based on a combined convolutional neural network (CNN). In [15], we proposed a fast beamforming technology based on unsupervised learning for downlink MIMO systems.

Based on the Review of previous works in physical layer wireless communications, we can find that DL is a powerful and potential weapon in the areas of performance optimization, channel estimation, and multiple access for the following reasons.

- DL can achieve better system-level performance by implementing an end-to-end optimization rather than a block-by-block optimization. It should be noted that existing communication systems are designed block by block, and the best performance of each block may not mean the global optimization.
- There will be various communication links considering rapidly changing channel conditions in future ultra-dense and large-scale scenarios, and DL have potentials in solving the problems where wireless channels might be modeled inaccurately.
- DL-aided signal processing algorithms can provide fast computing speed and better accuracy. In addition to taking advantage of finding structural information to improve the overall performance, DL can also apply parallel computing to handle massive data, and thus has the ability to adapt rapidly changing scenarios.

The rest of this paper is organized as follows. Section 2 introduces two DL-aided key techniques: AMC and fast beamforming, which respectively represent DL achievable tasks of classification and regression. In Section 3, we introduce future challenges and opportunities of wireless communications in physical layer. In Section 4, we conclude this paper.

2. Two DL-aided Key Techniques

In this section, we will introduce two DL-aided key techniques, i.e., automatic modulation classification (AMC) and fast beamforming in physical wireless communications. AMC represents a supervised classification task with better performance than traditional algorithms, and fast beamforming is a typical unsupervised regression task, which is more effective and less complex than traditional algorithms at a cost of slight performance loss.

CNN-based AMC

AMC is an essential technique for uncooperative communications to distinguish modulation mode of signal without any prior information. A typical AMC-based communication system is shown in Fig. 1. There are various civil and military applications adopting AMC. For instance, in the aspect of modern military applications, AMC is a key step to recover the intercepted signals in electronic warfare. AMC also contributes an assistance in analyzing interference signal and sensing spectrum for civilian scenarios.

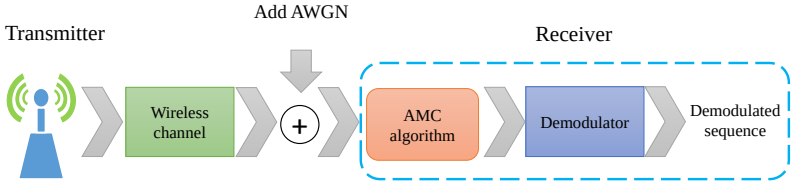


Figure 1 AMC-based system model.

AMC is generally considered as a pattern recognition problem, which typically includes three steps: pre-processing, feature extraction, and classification. In traditional AMC algorithms, the core technique is to design manmade features. And there are many effective and efficient features, such as instantaneous features, high order cumulants (HOC), wavelet transformation-based features, and so on. The classification step is based on these features and machine leaning algorithms.

To the best of our knowledge, CNN can directly replace the complex process of designing manmade features and conventional classifiers. Instead, it can straightforward extract more effective features and execute a classification process, relying on massive data. In this article, we transform baseband signals into in-phase and quadrature signals, respectively. Then, the in-phase part and quadrature part of signals are arranged in two-dimension matrices, which are denoted as IQ samples. The IQ samples form the training and testing datasets of the utilized CNN. In addition, each sample need to be labeled according to its modulation mode. Thus, the proposed CNN-based AMC scheme is a supervised learning algorithm.

In this article, we compare the performances of the CNN-based AMC and two HOC-based AMC, where support vector machine (MLP) and multilayer perceptron (MLP) are used as classifiers, respectively. These three AMCs aim to recognize three modulation modes, namely FSK, PSK, and QAM for AWGN channels. Simulation results are shown in Fig. 2. From Fig. 2, it can be observed that the accuracy of the CNN-based AMC is far beyond those of the other two AMCs. Moreover, the results also demonstrate that the CNN-based AMC is more powerful than the evaluated traditional methods in feature extraction.

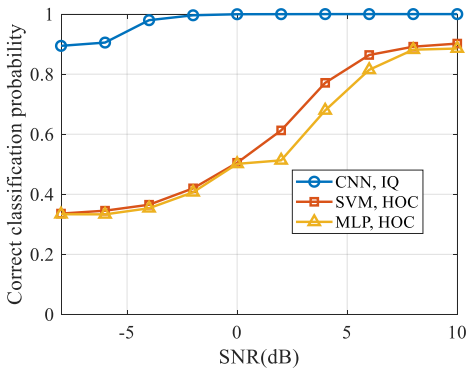
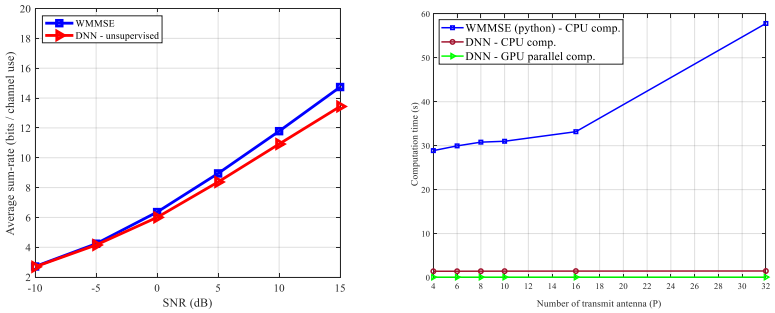


Figure 2 The performance of CNN-based AMC and two traditional AMC.

Unsupervised Learning-aided Fast Beamforming

In downlink transmission scenarios, power control and beamforming design at the transmitter side is very essential when using antenna arrays. Here, we consider multiple input multiple output broadcast channels (MIMO-BCs) and aim to maximize weight sum-rate (WSR) under certain power constraints. The conventional weighted minimum mean-square error (WMMSE) algorithm can obtain suboptimal solutions with high computational complexity. To reduce computational complexity and time consumption, we apply an unsupervised learning process to obtain the beamforming solution. We have trained a deep neural network (DNN) offline, and the obtained network can provide real-time service with just simple linear operations. The training process is based on an end-to-end method without any labeled samples.

As can be seen in Fig. 3 (a), the performance of the proposed DNN-based method is close to the WMMSE algorithm when calculating average sum-rates. Although the DNN-based algorithm with fixed structure presents slight performance loss in comparison with the WMMSE, the complexity of the DNN-based scheme decrease exponentially, especially when the number of antennas increases (see Fig. 3 (b)). In addition, DNN can be accelerated by GPUs, and WMMSE has no work under GPUs' parallel computing. Hence, the computing time of DNN can be further reduced.



(a) (b)

Figure 3 The performance and computation times of WMMSE and unsupervised DNN. (a) is the performance of WMMSE and our proposed unsupervised-DNN, and (b) is the computing times.

3. Future Challenges and Opportunities

Data Simulation and Actual Data Collection

In the recent years, DL has achieved unprecedented success in computer vision (CV), nature language processing (NLP), speech recognition (SR), and *et al*. One of the most important reasons is that training and testing frameworks in computer science are implemented on massive and effective dataset, such as ImageNet for large-scale image classification and Cornell Natural Language Visual Reasoning (NLVR) for NLP.

However, there are fewer generic and available dataset for DL-aided wireless communications. To facilitate researches, we can develop model-driven DL, such as orthogonal approximate message passing (OAMP)-based DL [16] and message passing algorithm (MPA)-based DL [13]. For those communication problems which can be modeled, the model-driven DLs can reduce the dependence on data. On the other hand, for modeless problems and problems cannot be accurately modeled, massive, reliable, and available dataset should be created to facilitate the use of data-driven DL. Relying on large amounts of data, the data-driven DL methods are important supplements to deal with the modeless problems. Hence, it is desirable to create more actual data collected from real and complex scenarios, and to develop software for creating simulation datasets for various wireless communication systems.

DL Models Selection

In DL-aided wireless communications, the core problem is to choose models and determine the parameters of the models. Models selection and parameters determination rely on experience from adequate experiments, which may occupy most time of a research period. In addition, experience-based parameters determination generally brings in overfitting problem, which means that neural networks have excessive redundancy hyper-parameters. Deep reinforcement learning (DRL) may act as an assistance to automatically choose models and adjust parameters.

DL Models Compression and Acceleration

DL is famous with its outstanding performance with high computing complexity, and DL-based algorithms generally require GPUs to be accelerated in practical applications. In addition, large memory resources are also needed for the

deployment of DL-based algorithms. However, a large amount of wireless communication devices, such as IoT devices, are usually not equipped with GPUs and have limited memory units. So if we intend to apply DL in wireless communications, there remains a crucial step to research on how to reduce computing complexity and compress the model sizes. The acceleration of DL-based structures is a key issue for future commercial DL-aided wireless communications.

4. Conclusion

In this article, inspired by state-of-the-art DL-based methods and their outstanding performance in various tasks in physical layer communications, we have reviewed and summarized the development of DL-aided physical layer wireless communications. We have presented a detailed description of CNN-based AMC and unsupervised learning-based fast beamforming, which can demonstrate the effectiveness of DL. There have been many essential breakthroughs in this filed, however, we firmly believe that DL-aided physical layer techniques are key directions of the future wireless communications.

References

[1] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, 2017.

[2] B. Mao *et al.*, "A Novel Non-Supervised Deep Learning Based Network Traffic Control Method for Software Defined Wireless Networks," *IEEE Wirel. Commun.*, vol. 25, no. 4, pp. 74–81, 2018.

[3] X. Gao, S. Jin, C. K. Wen, and G. Y. Li, "ComNet: Combination of Deep Learning and Expert Knowledge in OFDM Receivers," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2627–2630, 2018.

[4] C. K. Wen, W. T. Shih, and S. Jin, "Deep Learning for Massive MIMO CSI Feedback," *IEEE Wirel. Commun. Lett.*, vol. 7, no. 5, pp. 748–751, 2018.

[5] X. Sun, G. Gui, Y. Li, R. P. Liu, and Y. An, "ResInNet: A Novel Deep Neural Network with Feature Reuse for Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 679–691, 2018.

[6] H. Huang, Y. Song, and J. Yang, "Deep-Learning-based Millimeter-Wave Massive MIMO for Hybrid Precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, 2019.

[7] M. Liu, J. Yang, G. Gui, and S. Member, "DSF-NOMA : UAV-Assisted Emergency Communication Technology in a Heterogeneous Internet of Things," *IEEE Internet Things Journal, to be Published, doi 10.1109/JIOT.2019.2903165*.

[8] F. Tang, Z. M. Fadlullah, B. Mao, and N. Kato, "An Intelligent Traffic Load Prediction Based Adaptive Channel Assignment Algorithm in SDN-IoT: A Deep Learning Approach," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 5141–5154, 2018.

[9] F. Tang, B. Mao, Z. Fadlullah, and N. Kato, "On a Novel Deep-Learning-Based Intelligent Partially Overlapping Channel Assignment in SDN-IoT," *IEEE Commun. Mag.*, vol. 56, no. September, pp. 80–86, 2018.

[10] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep Learning for an Effective Nonorthogonal Multiple Access Scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, 2018.

[11] H. Huang, J. Yang, Y. Song, H. Huang, and G. Gui, "Deep Learning for Super-Resolution Channel Estimation and DOA Estimation based Massive MIMO System," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, 2018.

[12] M. Liu, T. Song, and G. Gui, "Deep Cognitive Perspective: Resource Allocation for NOMA based Heterogeneous IoT with Imperfect SIC," *IEEE Internet Things Journal, to be Publish, doi 10.1109/JIOT.2018.2876152*.

[13] M. Liu, J. Yang, T. Song, J. Hu, and G. Gui, "Deep Learning-Inspired Message Passing Algorithm for Efficient Resource Allocation in Cognitive Radio Networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 641–653, 2018.

[14] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-Driven Deep Learning for Automatic Modulation Recognition in Cognitive Radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, 2019.

[15] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, and X. Zhu, "Unsupervised Learning Based Fast Beamforming Design for Downlink MIMO," *IEEE Access*, vol. 7, pp. 7599–7605, 2019.

[16] Jing Zhang, Hengtao He, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li, "Deep Learning Based on Orthogonal Approximate Message Passing for CP-Free OFDM," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 1-10.

Information Theory Inspired Multi-modal Data Fusion

Longwei Wang[1], Yupeng Li[2]
[1]Auburn University, USA ,[2]Tianjin Normal University, China
allenwang163@gmail.com

1. Introduction

Information about a target of interest or a phenomenon can be obtained from different sensing modalities, such as visual, acoustic, text or from other types of measurement techniques and oriental views. This different information of a single phenomenon can be seen as multi-modal data. How to manage and process these multi-modal data has been a challenge in academia for decades.

There are many applications for multi-modal systems. One typical application is the interactive systems [1]. The multimodal interactive systems make the computers interact with users by various different modalities, such as gesture, voice or eye contact. The computer can deliver information by speech, sound, graphics or text. The multi-modal interactive system can be easily accessible to disabled people, such as visually impaired people. It is also more convenient and flexible for users' interaction with computers by switching to the users' preferred interaction modalities.

Another application for multi-modal system is the medical diagnosis [2]. The radiological appearances or the patterns of most diseases are highly complex and heterogeneous. Doctors can obtain the patients' information by different modalities, such as positron emission tomography (PET), magnetic resonance imaging (MRI), or computed tomography (CT). Developing efficient algorithms of multimodal image fusion that integrate the information from different modalities can make it possible to predict or find the symptoms which could be hidden when we consider the information of different modalities separately.

Fusion of data from heterogeneous sensor modalities [3] has been proved to improve monitoring and surveillance performance in many scenarios. The main reason is that multimodal sensors can capture more information than the single modality sensor. For example, in human speech communications, the voice with the help of body movement visualization can efficiently improve the understanding of speech. The visual information provides more information to make the speech easily being interpretable. Efficient Multimodal data fusion can provide as much information as possible to analyze and interpret the uncertain phenomenon.

Information theory is proposed as a great method to quantify uncertainty and manipulate the probability of uncertain phenomenon. Information-theoretic metrics, such as entropy and mutual information, have been applied to solve various problems in the areas of image processing and signal processing.

One major problem in multimodal data fusion is that the data forms of various modalities are heterogeneous, which makes the data fusion difficult to be performed. For example, the acoustic signal is usually one dimensional sequence, while the visual signals are two dimensional images. It is essential to extract the information from these heterogeneous data [3]. Recently, representation learning [5] is proposed as an efficient way to extract the information from data. Learning unified representations of the data makes it easier to build classifiers or predictors [6]. Learning an unified representations of multimodal data is shown to be critical for the downstream data fusion.

In this paper, we try to give a short review of information theory inspired multimodal data fusion methods in literature. Three different methods are covered: Mutual Information Based Multimodal Data, Fusion Nature Encoded Multimodal Data Fusion Information, Resonance Based Multimodal Data Fusion.

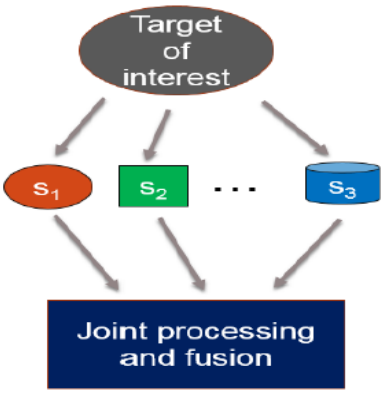


Figure 1 Multi-modal data fusion process [3].

2. Mutual Information Based Multimodal Data Fusion

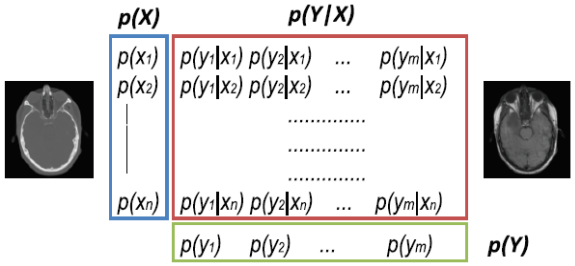


Figure 2 The communication channel between the two input data sets [2].

The channel in communication research is exploited to capture the relationship between multimodal data sets [2]. The mutual information is characterized for the multimodal data sets. For each intensity value in each data set, the probability information is computed, which is called the modality data specific information. The mutual information is decomposed in three different ways, based on which various information measures are derived. This mutual information based method for multimodal data fusion can efficiently express the relationship among the multimodal data, thus providing an useful way for data fusion. However, when the dimension of the data is very high, the computation of the probability information would be prohibitive.

3. Nature Encoded Multimodal Data Fusion

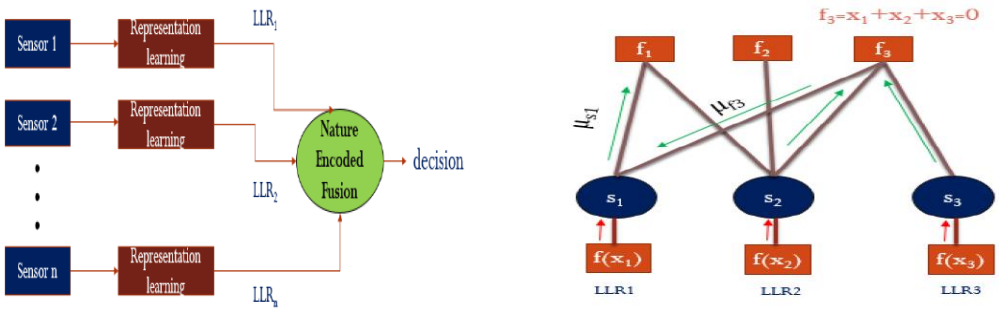


Figure 3 Nature encoded fusion and belief propagation [3].

In this work, a two-stage fusion framework is proposed to improve the target detection performance based on multimodal sensor data. First, each modality data is trained by an individual classifier and transformed into the same representation form. Then, the learned representation is used for the following probabilistic fusion. The inherent inter-

sensor relationship is exploited to encode the original sensor data on a graph. Iterative belief propagation is used to fuse the local sensing belief [3].

4. Information Resonance Based Multimodal Data Fusion

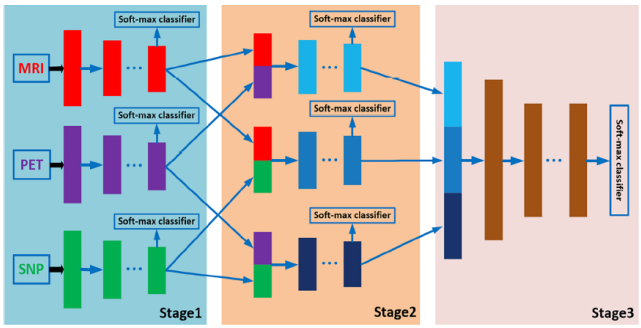


Figure 4 Representation combination for multimodal data fusion [4].

In this work, the authors consider a multiple stages fusion method based on deep learning [4]. First, large training data set is used to learn the individual representation of each modality, so the modality heterogeneity can be addressed by transforming into an unified representation. In this way, the representation of different modalities can be combined in the following stage. In the second stage, the joint representation (information resonance) of each pair of modality data can be learned based on the output of the first stage. Last, the prediction label can be learned by fusing the joint representations (information resonance) of the previous stage. The information resonance can boost the data fusion performance to some extent. But in some cases, there would be overfitting in the training of the fusion parameters.

5. Concluding Remarks and Future Directions

In this paper, we have reviewed several information-theory inspired multimodal data fusion frameworks. The mutual information based fusion method takes advantage of the communication channel concept in information theory, which forms an information channel between the data sets of different modalities. The nature encoded fusion method exploit representation learning and inherent relationship among the different sensors to fuse the multimodal information. The information resonance based multimodal data fusion combines the representation of different modality data in the training of the fusion process, which efficiently boosts the fusion performance.

For future work, the generalization of information theory inspired approach to the fusion of more generalized data sets is to be investigated. Traditional information theory mainly focus on applying probabilistic analysis to the uncertain data. We will study how to fuse the structural information by considering the spatial coherence and inherent behavior information.

References

[1] Lalanne, D., Nigay, L., Robinson, P., Vanderdonckt, J. and Ladry, J.F., 2009, November. Fusion engines for multimodal input: a survey. In *Proceedings of the 2009 international conference on Multimodal interfaces* (pp. 153-160). ACM.

[2] Bramon Feixas, R., Boada, I., Bardera Reig, A., Rodriguez, J., Feixas Feixas, M., Puig Alcántara, J. and Sbert, M., 2012. Multimodal data fusion based on mutual information. *© IEEE Transactions on Visualization and Computer Graphics, 2012, vol. 18, núm. 9, p. 1574-1587.*

[3] Wang, L. and Liang, Q., 2019. Representation Learning and Nature Encoded Fusion for Heterogeneous Sensor Networks. *IEEE Access*, *7*, pp.39227-39235.

[4] Zhou, T., Thung, K.H., Zhu, X. and Shen, D., 2019. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Human brain mapping*, *40*(3), pp.1001-1016.

[5] Bengio, Y., Courville, A. and Vincent, P., 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), pp.1798-1828.

[6] Le, N. and Odobez, J.M., 2016, October. Learning multimodal temporal representation for dubbing detection in broadcast media. In *Proceedings of the 24th ACM international conference on Multimedia* (pp. 202-206). ACM.

MMTC OFFICERS (Term 2018 — 2020)

CHAIR

Honggang Wang
UMass Dartmouth
USA

STEERING COMMITTEE CHAIR

Sanjeev Mehrotra
Microsoft
USA

VICE CHAIRS

Pradeep K Atrey (North America)
Univ. at Albany, State Univ. of New York
USA

Wanqing Li (Asia)
University of Wollongong
Australia

Lingfen Sun (Europe)
University of Plymouth
UK

Jun Wu (Letters&Member Communications)
Tongji University
China

SECRETARY

Shaoen Wu
Ball State University
USA

STANDARDS LIAISON

Guosen Yue
Huawei
USA

MMTC Communication-Frontier BOARD MEMBERS (Term 2016—2018)

Dalei Wu	Director	University of Tennessee at Chattanooga	USA
Danda Rawat	Co-Director	Howard University	USA
Melike Erol-Kantarci	Co-Director	University of Ottawa	Canada
Kan Zheng	Co-Director	Beijing University of Posts & Telecommunications	China
Rui Wang	Co-Director	Tongji University	China
Lei Chen	Editor	Georgia Southern University	USA
Tasos Dagiuklas	Editor	London South Bank University	UK
ShuaiShuai Guo	Editor	King Abdullah University of Science and Technology	Saudi Arabia
Kejie Lu	Editor	University of Puerto Rico at Mayagüez	Puerto Rico
Nathalie Mitton	Editor	Inria Lille-Nord Europe	France
Zheng Chang	Editor	University of Jyväskylä	Finland
Dapeng Wu	Editor	Chongqing University of Posts & Telecommunications	China
Luca Foschini	Editor	University of Bologna	Italy
Mohamed Faten Zhani	Editor	l'École de Technologie Supérieure (ÉTS)	Canada
Armir Bujari	Editor	University of Padua	Italy
Kuan Zhang	Editor	University of Nebraska-Lincoln	USA