Joint Dynamic Rate Control and Transmission Scheduling for Scalable Video Multirate Multicast Over Wireless Networks

Chenglin Li, *Member, IEEE*, Hongkai Xiong, *Senior Member, IEEE*, Junni Zou, *Member, IEEE*, and Dapeng Oliver Wu, *Fellow, IEEE*

***Abstract*—In this paper, we consider the time-varying characteristics of practical wireless networks, and propose a joint dynamic rate allocation and transmission scheduling optimization scheme for scalable video multirate multicast based on opportunistic routing (OR) and network coding. With OR, the decision of optimal routes for scalable video coding layered streaming is integrated into the joint optimization formulation. The network throughput is also increased by taking advantage of the broadcast nature of the wireless shared medium and by network coding operations in intermediate nodes. To maximize the overall video reception quality among all destinations, the proposed scheme can jointly optimize the video reception rate, the associated routes to different destinations, and the time fraction scheduling of transmitter sets that are concurrently transmitting in the shared wireless medium. By using dual decomposition and primal-dual update approach, we develop a cross-layer algorithm in a fully distributed manner. Simulation results demonstrate significant network multicast throughput improvement and adaptation to dynamic network changes relative to existing optimization schemes.**

***Index Terms*—Dynamic rate allocation, multirate multicast, network coding, opportunistic routing, scalable video coding.**

I. Introduction

With the extensive growth of global mobile data traffic and the widespread use of smart devices, wireless video streaming has experienced extensive growth in the last decades and been utilized for a wide range of applications, such as mobile video services. According to Cisco visual networking index [1], mobile video traffic took more than 50 percent of the total traffic by the end of 2014, and will increase to nearly three-fourths of the world's mobile data traffic by 2019. However, transmitting video streams via wireless networks with guaranteed quality of service (QoS) metrics is still a challenging problem since the wireless channels are time-varying and error-prone with respect to wired transmissions. Multirate multicast for scalable video streams, as an important method for video content distribution over wireline networks, can benefit the overall network utility by adapting to different user requirements and heterogeneous network conditions [2]. A scalable video coding (SVC) stream comprising a base layer and one or multiple enhancement layers with a flexible multi-dimension layer structure, provides various operating points in spatial resolution, temporal frame rate, and video reconstruction quality. Multirate multicast allows different SVC layers to be delivered in different IP multicast groups and subscribed by heterogeneous receivers with different computation/communication resources and capabilities. As will be discussed in detail in Section I-A, due to the unique features of wireless multi-hop networks, such as the lossy behavior, the contention and interference among links, and the time-varying characteristics, designing an efficient and reliable SVC multirate multicast scheme for wireless networks is not a straightforward extension from the schemes designed for wireline networks.

A. Related Work

Traditional optimization based rate control schemes for multirate multicast in general networks have been proposed in literature [3]–[6], and often formulated as a network utility maximization (NUM) problem with distributed implementation that maximizes the total receiver utility for all multirate multicast sessions. As multirate multicast applications in scalable video streaming, several resource rate and network flow control schemes [7], [8], [9] adopted video reception quality as destination's utility and attempted to maximize the overall video reception quality for all destinations and all SVC video layers, while satisfying SVC layer dependency and network capacity constraints. In [10], random linear coding was designed as an efficient solution for scalable video multicast over one hop transmission. Thomos *et al.* in [11] jointly investigated network

Manuscript received January 13, 2016; revised April 22, 2017; accepted July 31, 2017. Date of publication August 29, 2017; date of current version January 17, 2018. This work was supported in part by the NSFC under Grants 61501293, 61529101, 61425011, 61622112, and 61472234, in part by the Program of Shanghai Academic Research Leader under Grant 17XD1401900, in part by the China Postdoctoral Science Foundation under Grants 2016T90372 and 2015M570365, in part by the China Scholarship Council, and in part by the NSF under Grant ECCS-1509212. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Christian Timmerer. *(Corresponding author: Dapeng Oliver Wu.)*

C. Li is with the Signal Processing Laboratory (LTS4), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland (e-mail: chenglin.li@epfl.ch).

H. Xiong is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: xionghongkai@sjtu.edu.cn).

J. Zou is with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zou-jn@cs.sjtu.edu.cn).

D. O. Wu is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: dpwu@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2017.2745709

coding and scalable video coding for efficient multicast over multi-hop networks, and designed a distributed receiver-driven streaming solution. Similarly, a resource allocation scheme was proposed in [12] to support scalable video multicast for WiMAX relay networks.

These schemes, although proved to have optimal or near optimal performance, suffer from two major limitations. First, they are based on the assumption that network characteristics (e.g., topology, link capacity, etc.) are static and do not change over time. For example, Yan *et al.* in [5] supposed that the capacity of each link is fixed and given in the optimization formulation, while the packet reception rate in [9] was assumed to be a given and constant value for each wireless link. Second, one or multiple routes between each source-destination pair have to be predefined before the optimization scheme operates. For example, in [7] and [9], the candidate path set from the source node to all the destination nodes were predefined and denoted by a fixed link-route matrix. In practical wireless networks, however, these two assumptions would lead to infeasible or poor performance since the wireless channel is time-varying by nature, and the associated wireless channel state changes over time frequently. For example, the wireless node's movement, or the rate decrease (or even link failure) of an incoming wireless link to the destination node may cause the original optimal solution to be infeasible and previously obtained maximum network utility unreachable. In addition, it would also be impractical and costly to determine the routes whenever the network change occurs, since the shortest paths may not be the optimal due to flow congestion in practical networks and the total number of possible paths between a source-destination pair would increase exponentially with the scale of the network. Therefore, the scalable video multirate multicast problems over time-varying wireless networks are still technical challenges that require further investigation to capture the temporal dynamics in practice. To this end, the distributed robust algorithm in [9] tried to enhance the robustness of NUM formulation by reserving partial bandwidth for several backup paths in case of possible link failures of primary paths, and by introducing an uncertainty set of the wireless medium capacity to represent the uncertain and time-varying property of the wireless channel. However, the limitation of this work is that either the primal paths or the backup paths between each source-destination pair have to be predefined and will not change before the optimization scheme operates, which might not work well in the dynamic networks with the frequent change of routes. Furthermore, in order to obtain a tradeoff between optimization performance and robustness, it has to sacrifice network throughput to protect from infeasible solution caused by dynamic network changes.

In order to efficiently deal with time-varying wireless channels and imperative needs for dynamic routing, opportunistic routing (OR) approaches have been investigated in [13]–[16], to exploit the broadcast nature and the spatial diversity of the wireless shared medium. Compared to traditional routing which specifies only one forwarding candidate for a transmitter, involving multiple forwarding candidates by OR could combine multiple weak links into one strong link, and increase the probability of at least one next-hop node having successfully received the packets. Such increase of packet forwarding success probability within one wireless transmission decreases the probability for retransmission, which in turn enhances the network throughput [13], [14]. Zeng *et al.* in [15], [16] proved that with opportunistic routing, the end-to-end throughput can be significantly improved in multirate and multi-hop wireless networks compared to the traditional routing. Another advantage of OR is that the routes for a packet between any source-destination pair are not required to be specified in advance. On the contrary, the next-hop forwarder is determined for a packet at the transmitter node on the fly based on OR strategy, and thus the route is also dynamic. However, to avoid duplicated transmission of one packet, a specific design on the MAC protocol is required for these schemes to ensure the coordination of nodes in the forwarder set and to specify which one of them should forward the packet that has been received by multiple nodes. It is also difficult to extend these existing opportunistic routing approaches from the unicast case to the multicast case, since the availability of multiple destinations in a multicast tree may make the MAC protocol design much more complicated.

To eliminate the shortcomings incurred by tying MAC with opportunistic routing, network coding is applied to significantly simplify the design of opportunistic routing protocols and result in substantial throughput gains compared with non-coding based protocols, and several network coding-based opportunistic routing schemes were proposed in [17]–[22]. With network coding, both the source node and the relay nodes will randomly mix the packets that have been received before forwarding them. Such random linear coding operation at each node ensures that, different forwarder nodes that may have received the same packet could still send linearly independent coded packet with high probability. In [18] and [19], the coding-aware opportunistic routing protocol that combines hop-by-hop opportunistic forwarding and localized inter-flow network coding was proposed for improving the throughput performance of wireless mesh networks. In [20] and [21], the network coding-based opportunistic routing approach was investigated for the unicast case, where significant throughput gain is demonstrated over traditional opportunistic routing schemes. As an extension to multiple flows, [21] suggested an optimization framework based on network utility maximization, by defining the link rate constraints per broadcast region instead of links in isolation. Khreishah *et al.* in [22] addressed the multicast scenario by jointly considering link loss rate, correlation among links and reachability of nodes. However, they neglect the characteristic of wireless link contention and interference, which makes the wireless transmission scheduling a critical problem that would affect the performance of the wireless streaming. In addition, these works were mainly designed for the general data flows without any priority. However, in the scalable video streaming applications, the transmission failure of lower layer video packets would be propagated to the higher layers and greatly impact the decoding of the higher layers of the scalable video. Therefore, the decoding process of the higher layers is highly dependent on the lower layers, and different priority has to be assigned for each scalable video layer.

B. Proposed Research and Paper Organization

To address the challenges of scalable video multirate multicast over time-varying wireless networks, in this paper, we are motivated by the network coding based opportunistic routing, and incorporate the multi-period formulation of dynamic network utility maximization to formulate a joint dynamic rate allocation, transmission scheduling, and opportunistic routing optimization for scalable video multirate multicast based on network coding. With the dynamic channel information of wireless links either known or predicted, the proposed scheme can jointly optimize the video reception rate, the associated routes to destinations, and the time fraction scheduling of transmitter sets that are concurrently transmitting in the shared wireless medium. For practical implementation, we develop a fully distributed algorithm to solve the joint rate allocation, link routing and transmission scheduling problem. Extensive simulation results show that significant network multicast throughput improvement and adaptation to dynamic network changes can be achieved compared to existing optimization schemes. The main contribution of this paper is as follows: 1) Different from existing NUM-based scalable video multirate multicast schemes, this work takes node mobility and time-varying channel state into account, where a packet's actual route between the source and destination pair is determined on the fly based on network coding and opportunistic routing, so as to deal with the time-varying unreliable transmissions and to increase the successful transmission probability. 2) The multi-period formulation of dynamic network utility maximization is incorporated with the transmission conflict graph to formulate a joint dynamic rate allocation, transmission scheduling, and opportunistic routing optimization problem **P1**. 3) Different from a simple extension to directly applying network coding-based opportunistic routing to the scalable video streaming, the SVC layer dependency constraint is jointly considered with the network coding-based routing to ensure that each destination node subscribes to the video layers in an incremental order. 4) In order for a destination node to meet the SVC decoding requirement and successfully recover a video layer, there is another requirement on the total amount of information belonging to that layer delivered through the opportunistic routes to the destination over the entire transmission time interval. Such QoS requirement over the time interval can also be viewed as an SVC video delivery contract that couples the reception rate of each destination node across time and is promised by the proposed scheme. 5) We develop a stochastic dynamic version of the problem **P1** in **P2**, by using the finite state Markov channel (FSMC) model with consideration of node mobility. In **P2**, the channel state information of the future time slots is unknown and only predicted based on previous and current channel state information. The resulting performance after solving **P1** can be treated as the theoretical upper bound for the stochastic dynamic problem **P2** with imperfect channel state information.

The rest of this paper is organized as follows. Section II gives system models. In Section III, we formulate a joint dynamic rate allocation and transmission scheduling problem based on both OR and network coding, and further develop a stochastic dynamic problem by using the FSMC model for prediction. Section IV develops a distributed cross-layer algorithm with implementation details. Simulation results are shown and analyzed in Section V. Section VI concludes this paper.

II. System Models

A. Wireless Network Under Time-Varying Channels

Consider the scalable video streaming over a wireless network with all wireless nodes randomly placed on a plane. We model such wireless network as a directed graph $G = (V, E)$, where V is the set of wireless nodes and E is the set of wireless links. Alternatively, the node set V can be viewed as the union of three subsets, i.e., $V = \{s\} \cup N \cup D$, where s represents the single video source node, N denotes the set of relay nodes, and D is the set of destination nodes, respectively. By utilizing scalable video coding, the video stream is encoded at the source node s, and then multicasted to all the destination nodes via multi-hop wireless nodes with network coding and opportunistic routing.

In a practical wireless network, the channel state between two nodes might frequently change (e.g., link failures or link fluctuations) due to node mobility and wireless channel fading [23]. In accordance with the time-varying characteristics of wireless links, we assume that within the entire period of scalable video streaming, time is slotted with slots normalized to integral units $t \in \mathcal{T} = \{1, 2, \ldots T\}$, and channels hold their states within the duration of a time slot [24]. Such time-varying channel state information (CSI) can be captured either through direct measurement (if the duration of a time slot is sufficiently long compared with the required measurement time), or through the combination of measurement and channel state prediction. Since the time domain is discretized and t can only take some discrete values, throughout the paper, for any variable x and a given discrete-time slot $t \in \mathcal{T}$, we denote $x(t)$ as the discrete-time variable at time slot t.

At time slot t, each node i ($i \in \{s\} \cup N$) can broadcast data packets to its neighboring nodes with capacity C_i. Denote $p_{(i,j)}(t)$ as the packet reception ratio (PRR) from transmitting node i to receiving node j through link $l = (i, j)$, then a directed link (i, j) exists if and only if $p_{(i,j)}(t)$ is larger than a positive PRR threshold p_{td} [15], [16]. Alternatively, let $dist_{(i,j)}$ denote the Euclidean distance between nodes i and j, then there is a usable directed link (i, j) when $dist_{(i,j)} < D_t$, where the effective transmission range D_t is defined as the maximum allowable sender-receiver distance at which the PRR equals to the threshold p_{td}.

B. Rate Allocation for Scalable Video Multirate Multicast

According to [7], the aggregate throughput for a multicast over heterogeneous destinations may be improved by providing multirate to the destinations. Layered data multicast, as a commonly used multirate approach, could benefit such aggregate throughput by adding data layers to destination nodes with higher min-cut so as to further utilize the available capacity. In this work, the multirate multicast for scalable video streaming is considered as a layered multicast with the transmission of each

video layer corresponding to a sperate multicast session. Specifically, assume that an SVC source video stream is encoded as a set of M layers $\{L_1, L_2, \ldots, L_M\}$ with predefined encoding rate for each layer, and then multicast to the destination node set D via the wireless network. The transmission of each video layer m corresponds to a multicast session m through the network.

To study the multicast transmission's impact on the video quality of each individual user, the encoding/decoding dependency among different SVC layers should be taken into account, which indicates that, within the multiple multicast sessions of different video layers, different priority has to be assigned to each multicast session according to its importance. Specifically, since the failure of the transmission of lower layer video packets would be propagated to the higher layers and greatly impact the decoding of the higher layers of the scalable video, the decoding process of the higher layers is highly dependent on the lower layers. For each individual destination node, in order to achieve the best video quality, it is supposed to join as many SVC multicast sessions as possible (or, to be allocated with a reception rate that supports receiving as many SVC video layers as possible) constrained by its capacity in an incremental order, since layer $m+1$ is only decodable with the existence and successful decoding of all the previous layers 1 to m. Therefore, the SVC layer dependency constraint as proposed in [9] is taken into account in this paper to guarantee that a higher priority is assigned to the multicast session of more important layers.

Also, we introduce a tolerable rate region $[r^m_{\min}, r^m_{\max}]$ and suppose that each layer m is distributed over a multicast session at a variable transmission rate chosen from that rate region. Technically, the upper bound $r^m_{\max}$ and the lower bound $r^m_{\min}$ are specified as a confidence interval of the transmission rate in layer m. Therefore, for each layer, the achievable transmission rate is extended from a single encoding rate point to a tolerable rate region, which would lead to adaptation to network fluctuation and a broader feasible region for the corresponding SVC multirate multicast problem. Let $R^{md}(t)$ represent the received rate by destination node d for video layer m at time slot t, then, according to [9], the SVC encoding/decoding constraint is proposed as follows:

$$\left[\sum_{t=1}^{T} R^{md}(t)\cdot\Delta T\right]\cdot\left[\sum_{t=1}^{T} R^{md}(t)\cdot\Delta T - r^m_{\min}\cdot T\cdot\Delta T\right]\cdot\left[\sum_{t=1}^{T} R^{md}(t)\cdot\Delta T - r^m_{\max}\cdot T\cdot\Delta T\right]\le 0,\ \forall m,d, \quad (1)$$

where ΔT denotes the duration of a time slot. Since $R^{md}(t)$ is nonnegative, constraint (1) indicates that a video layer is either not subscribed with zero rate, or received by a destination node with a sufficient amount of information for that video layer. Together with constraint (1), the aforementioned SVC layer dependency constraint is guaranteed and given by [9]:

$$\frac{\sum_{t=1}^{T} R^{(m+1)d}(t)\cdot\Delta T}{r^{(m+1)}_{\max}\cdot T\cdot\Delta T}\le\frac{\sum_{t=1}^{T} R^{md}(t)\cdot\Delta T}{r^m_{\min}\cdot T\cdot\Delta T},\ \forall m,d. \quad (2)$$

From the perspective of application-layer QoS, to accurately measure the satisfaction perceived by a destination node, the following rate-distortion (RD) model in [25] is adopted as the utility for video applications:

$$D_e(R_e)=\frac{\theta}{R_e-R_0}+D_0, \quad (3)$$

where D_e is the distortion of the encoded video sequence and R_e is the encoding rate. The remaining variables, θ, R_0 and D_0, are the parameters of the RD model, which depend on the actual video content and are estimated from empirical rate distortion curves using regression techniques.

At time slot t, we associate with $R^{md}(t)$ a strictly increasing, differentiable and concave utility function $U(R^{md}(t))$. Within the context of SVC [9], this utility function is defined as the QoS improvement in the form of distortion decrement when destination node d successfully receives and decodes layer m:

$$U(R^{md}(t))=-\left[D_e\left(\sum_{\tau=1}^{m} R^{\tau d}(t)\right)-D_e\left(\sum_{\tau=1}^{m-1} R^{\tau d}(t)\right)\right]. \quad (4)$$

C. Network Coding-Based Opportunistic Routing

In order to efficiently cope with time-varying channels and unreliable transmissions, opportunistic routing [14] has been proposed to achieve high throughput for wireless networks. In opportunistic routing, one wireless packet transmission originated from a transmitter can be overheard by multiple neighbors within its effective transmission range. Since there is no specific next-hop node, any node that successfully receives the packet can potentially forward it. In this paper, we adopt the geographic opportunistic routing protocols [13], [16], [47] and select the set of nodes that overhear the transmission and are closer to the destination as the forwarder set to participate in the local opportunistic forwarding. The reason why we choose the geographical distance as the metric to determine the relay forwarder set and to assign corresponding relay priorities is as follows. First, the geographical distance is one of the most commonly used metrics in opportunistic routing protocols, since the geographic OR can help overcome the lack of infrastructure in wireless networks and adapt to its frequently varying topologies [47]. In addition, the performance evaluation in [13] shows that the geographic OR achieves a very close performance (in terms of the average number of hops to reach a destination) to that of the idealized scheme, where the best relay is always selected. It is also proved in [16] that by giving higher relay priorities to the candidate nodes closer to the destination, the opportunistic routing procedure can achieve the maximum expected advancement rate (EAR). The physical meaning of EAR is the expected bit advancement per second towards the destination when the packet is forwarded by further taking into account the link state information. Both the smaller average number of hops and the higher EAR could result in an overall lower end-to-end delay experienced by the packets, which is thus suitable for video applications. However, a difficulty is also introduced by opportunistic routing. That is, more than one node in the forwarder set may receive a packet and unnecessarily forward the same packet. To avoid such duplication, a specific design on the MAC protocol is required to guarantee the coordination

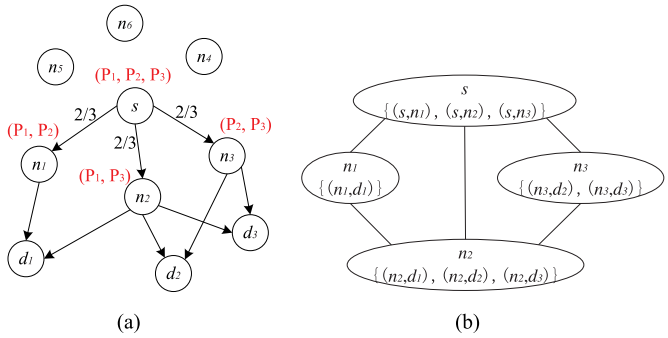


Fig. 1. (a) Basic module of opportunistic routing, where at time slot t node s is forwarding a packet to destination nodes $d_1 \sim d_3$ with forwarder candidate set $F_s(t) = \{n_1, n_2, n_3\}$, and (b) the transmitter conflict graph.

of nodes in the forwarder set and to specify which one of them should forward the packet that has been received by multiple nodes.

To make the idea clear, illustrated in Fig. 1(a) is the basic module of opportunistic routing. Suppose that at time slot t, node s is multicasting three packets, denoted by P_1, P_2 and P_3, to destination nodes $d_1 \sim d_3$.All the next-hop neighbors $n_1 \sim n_6$ within the effective transmission range of node s can overhear these packets. Based on the geographic opportunistic routing, however, only nodes $\{n_1, n_2, n_3\}$ are chosen as the forwarder set $F_s(t)$, since they are geographically closer to the destination nodes than node s. In Fig. 1(a), we further set the packet reception ratio of the link connecting s to each node within $F_s(t)$ to $2/3$. Since wireless receptions at different receiving nodes are highly independent [26], [27], we assume that after three transmissions each node receives the packets as indicated in Fig. 1(a), i.e., n_1 receives P_1 and P_2, n_2 receives P_1 and P_3, and n_3 receives P_2 and P_3.

Next, the packets received by nodes $\{n_1, n_2, n_3\}$ will be further forwarded and multicast to the three destination nodes. We first take destination node d_2 as an example and illustrate how the traditional opportunistic unicast routing scheme works for the packet forwarding from the relay nodes to this destination node. Considering the reachability to d_2 of relay nodes in $F_s(t)$, both nodes n_2 and n_3 need to forward the packets they received to d_2. Without coordination, both nodes want to forward packet P_3, which may cause the chance of duplicated reception of P_3 at d_2. Therefore, traditional opportunistic routing assigns ordered relay priorities to nodes n_2 and n_3 based on their distances to destination d_2. If the same packet is received by more than one node, then only one of them will continue forwarding the received packet according to their relay priorities. In this case, the relay priority of n_2 is higher than that of n_3 since n_2 is closer to d_2. Therefore, if node n_2 has correctly received the packet, it is arranged to forward the packet towards d_2 while the remaining node n_3 avoids duplicate forwarding. Otherwise, node n_3 is assigned to forward the packet if it has received the packet successfully. In this way, node n_2 would forward both received packets, P_1 and P_3, to d_2, while n_3 only transmits P_2. The same forwarder set selection and prioritization criteria can also be applied to the other two destination nodes so as to extend to the multicast case. However, it should be noted that the shortcoming of opportunistic routing is that, a specific MAC protocol is required to determine and leverage the selection of forwarder nodes and the coordination among them [17], [22]. Another disadvantage is the difficulty of the extension from the unicast case to the multicast case, since the availability of multiple destinations in a multicast tree may make such MAC protocol design much more complicated.

In contrast, by using network coding that allows relay nodes to perform algebraic operations on the received packets, these two disadvantages of opportunistic routing can be eliminated [17], [22]. For example, instead of forwarding the original source packets, n_1 may send coded packet $P_1 + P_2$ to d_1, and n_3 may send coded packet $P_2 + P_3$ to d_2 and d_3, while n_2 is still transmitting P_1 and P_3. Then, P_1, P_2 and P_3 can be successfully recovered by all the three destinations with network coding applied in the multicast scenario. The advantage of integrating opportunistic routing with network coding is that the destinations no longer need to receive specifically all the original source packets, but can receive any K (suppose the number of original packets is K) linearly independent coded packets and recover the original packets. Since we do not insist on receiving specific packets, we are no longer bothered by the problem of designing a MAC protocol.

In this paper, network coding-based multicast opportunistic routing is utilized to deal with time-varying unreliable transmissions as well as to achieve high multicast throughput. In order to reach the optimal throughput of the shared wireless network, an intuitive way to multicast multiple SVC layers through that network is inter-session network coding where coding operations are performed across layers. However, integrating data packets from multiple layers makes it difficult for destination nodes that only receive partial layers to recover the original data packets. Therefore, in this work, we adopt intra-session network coding [28] and only allow data packets from the same layer to be combined and coded. Specifically, the source node s keeps sending multiple flows of coded packets corresponding to different layers of the encoded SVC stream. Upon receiving a coded packet for a specific layer, each relay node would check whether this coded packet is linearly independent from the packets that have been previously received. If so, the relay node keeps this coded packet as an innovative packet, generates a random linear combination of the coded packets it has heard from the same layer, and broadcasts it. Otherwise, this packet is considered as non-innovative and thus dropped.

Defining $g^{m\,d}_{(i,j)}(t)$ as the virtual information flow rate for destination node d within multicast session m at time slot t, we have the following information flow conservation condition:

$$\sum_{j:(i,j)\in E} g^{m\,d}_{(i,j)}(t) - \sum_{j:(j,i)\in E} g^{m\,d}_{(j,i)}(t) = \sigma_i^{m\,d}(t),\ \forall m, d, i, t, \tag{5}$$

$$\sigma_i^{m\,d}(t) = \begin{cases} R^{m\,d}(t), & \text{for } i = s; \\ -R^{m\,d}(t), & \text{for } i = d; \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Meanwhile, by being coded together with intra-session network coding, flows to different destination nodes of a specific SVC

video layer are able to share the network capacity. Such flow sharing property requires the actual physical flow on each link to be the maximum of the individual destinations' virtual information flows, which leads to the following network coding constraint as in [4], [6]:

$$g^{m\,d}_{(i,j)}(t) \leq f^{m}_{(i,j)}(t),\ \forall m, d, (i,j), t, \tag{7}$$

where $f^{m}_{(i,j)}(t)$ denote the actual physical flow rate of layer m forwarded from node i to node j.

In addition, with network coding-based opportunistic routing, if a packet is overheard by multiple next-hop nodes which are able to reach a specific destination node d, only one of them can actually forward this packet through increasing its corresponding physical rate. According to [21], [22], the following feasible physical flow rate constraint is proposed for each broadcast region:

$$\sum_{j\in J}\sum_{m=1}^{M} f^{m}_{(i,j)}(t) \leq C_{(i,J)}(t),\ \forall J \subseteq F_i(t), i, t, \tag{8}$$

where $F_i(t)$ denotes the set of candidate next-hop nodes of i which comprises all the nodes that are geographically closer to the destinations than node i, J is a subset of $F_i(t)$, hyperedge (i, J) represents the broadcast region of links from node i to each node within J, and $C_{(i,J)}(t)$ is the effective transmission rate of packets that can be actually delivered from node i to any of the nodes in J. To better illustrate the idea of this effective transmission rate of a broadcast region (i, J), we consider again the example in Fig. 1(a). For example, destination d_1 is reachable from both nodes n_1 and n_2. We accordingly assume that $J = (n_1, n_2)$ and that source node s has sent K randomly coded innovative packets, then both n_1 and n_2 would receive $(2/3)\cdot K$ linearly independent packets. Due to the independence of the links associated with these two nodes, $2/3 \times 2/3 = 4/9$ of the K coded innovative packets will be received by both nodes, while $2/3 \times (1 - 2/3) = 2/9$ of the K coded innovative packets will be received by node n_1 (or n_2) but not received by node n_2 (or n_1). Therefore, the total innovative packets that should be further forwarded by either node n_1 or n_2 cannot exceed $(4/9 + 2/9 + 2/9)\cdot K = (8/9)\cdot K$, which is the effective transmission capacity of the broadcast region $(s, J = \{n_1, n_2\})$.

D. Wireless Transmission Scheduling

1) Transmitter Conflict and Concurrent Transmitter Sets: In wireless networks, the throughput of a wireless link is interrelated and affected by adjacent wireless links because of wireless link contention and interference in the shared transmission medium. The physical model and the protocol model [29], [30] are two interference models widely used to define the conditions for a successful transmission. Although the physical model is considered as a more realistic and accurate model for the real system, the corresponding signal-to-inference ratio (SINR) calculation is a non-convex function with respect to the transmission power, which might lead to high computational complexity in solving a capacity related optimization problem [31]. To circumvent the complexity issue, the protocol interference model has been widely adopted in the wireless networking research community. For example, several recent studies have confirmed that insights and solutions based on the protocol model are adequate in many application scenarios under rather general conditions [32], [33]. Specifically, the recent study in [32] presents how to properly set the interference range in the protocol model such that the protocol model can offer comparable results as those under the physical model through narrowing the solution gap between these two models. Therefore, in this paper, we take the protocol interference model into account. A typical protocol model [29] takes the location information of wireless nodes into consideration and proposes the following two conditions for a successful wireless transmission between the sender-receiver pair: 1) The intended recipient locates in the effective transmission range of the sender; and 2) Any other node in the carrier sensing range of the receiver is not transmitting simultaneously.

Under the protocol model, two or more wireless links will interfere with each other and thus have a conflict when they are not able to make successful transmissions at the same time. Considering the broadcast nature in opportunistic routing, we extend from the link conflict and further define the transmitter conflict as two or more transmitters that cannot be transmitting simultaneously due to the conflict among links to their associated forwarding candidates. To better understand the conflict relationship between transmitters and their associated forwarding candidates for opportunistic routing [15], [16], we construct an illustrative transmitter conflict graph in Fig. 1(b), with the original network topology shown in Fig. 1(a). Suppose that in the original network topology, source node s is multicasting SVC packets to three destination nodes $d_1 \sim d_3$, via three relay nodes $n_1 \sim n_3$. With opportunistic routing, nodes s, n_1, n_2 and n_3 can be selected as transmitters or forwarders, however, they cannot be forwarding packets simultaneously because of the transmitter conflict. In the transmitter conflict graph, each vertex denotes a transmitter in the original network topology associated with a set of links to its forwarding candidates. The conflict of two transmitters exists if they cannot be transmitting at the same time due to the conflict among their associated links, and is accordingly represented as an edge between two vertices in the transmitter conflict graph.

To characterize the impact of wireless interference and the opportunistic nature of opportunistic routing, in the following, we introduce the concepts of concurrent transmitter set (CTS) as defined in [15], [16]. Under the protocol model, a CTS is defined as a set of transmitters, when all nodes within that set are transmitting packets at the same time, all the links associated with them to their forwarding candidates can make successful transmissions. The basic idea of CTS is to avoid transmitter conflict by requiring all the opportunistic receivers to be interference-free at the same time. In order to make full use of the wireless shared medium and achieve the capacity bound of the network, a maximum CTS is defined as a CTS, if adding any one more node into it will lead to a non-CTS. For illustration, based on the transmitter conflict graph shown in 1(b), all the possible maximum CTS of the original network graph in Fig. 1(a) can be obtained as $\{s\}$, $\{n_2\}$ and $\{n_1, n_3\}$.

2) Effective Transmission Rate and Maximum CTS Scheduling: In a specific maximum CTS, all the transmitters can deliver packets via their associated links to the corresponding forwarder candidates. Let $\{\Gamma^\alpha(t)|\alpha = 1, 2, \ldots, A\}$ denote the set of all the maximum CTSs in the wireless network at time slot t, where α represents the index of a maximum CTS and A is the number of the maximum CTSs contained in the set. Denote an indicator function $\psi_i^\alpha(t)$ to represent the relationship between node i and a particular maximum CTS $\Gamma^\alpha(t)$. Specifically, $\psi_i^\alpha(t) = 1$ if node i is in maximum CTS $T^\alpha(t)$ within time slot t, and $\psi_i^\alpha(t) = 0$ otherwise. It should be noted that once the set $\{\Gamma^\alpha(t)|\alpha = 1, 2, \ldots, A\}$ is determined based on the protocol interference model, the indicator function $\psi_i^\alpha(t)$ is also known based on the set information without any thresholding. Take the network topology in Fig. 1(a) for example, we have $\Gamma^1(t) = \{s\}$, $\Gamma^2(t) = \{n_2\}$, and $\Gamma^3(t) = \{n_1, n_3\}$ in accordance with the transmitter conflict graph in Fig. 1(b). Based on the set information of $\Gamma^1(t)$ - $\Gamma^3(t)$, we can find that the values of $\psi_s^1(t)$, $\psi_{n_2}^2(t)$, $\psi_{n_1}^3(t)$ and $\psi_{n_3}^3(t)$ are 1, and all the other values are 0.

For any $J \subseteq F_i(t)$, a packet sent by i is considered as successfully received by J if it is received by at least one of the nodes in J. Therefore, the effective transmission rate of hyperedge (i, J) in maximum CTS $\Gamma^\alpha(t)$ is derived as:

$$C_{(i,J)}^\alpha(t) = \psi_i^\alpha(t) \cdot C_i \left[1 - \prod_{j \in J}(1 - p_{(i,j)}(t))\right], \tag{9}$$

where C_i is the broadcast capacity of node i and $p_{(i,j)}(t)$ denotes the packet reception ratio (PRR) of the wireless link (i, j) at time slot t. For any destination node $d \in D$, since all the received packets can be used for decoding and thus considered as useful, the effective transmission rate of link (i, d) is:

$$C_{(i,d)}^\alpha(t) = \psi_i^\alpha(t) \cdot C_i \cdot p_{(i,d)}(t). \tag{10}$$

Let $\{\Gamma^1(t), \Gamma^2(t), \ldots, \Gamma^A(t)\}$ denote the set of all the maximum CTSs in the wireless network at time slot t. Due to the transmitter conflict between any two maximum CTSs, at any time within that time slot, no more than one CTS can be scheduled to transmit while all the transmitters in that CTS can forward packets simultaneously. Set $\lambda^\alpha(t)$ as the fraction of time scheduled to a specific maximum CTS $\Gamma^\alpha(t)$, then the effective transmission rate of hyperedge (i, J) is derived as:

$$C_{(i,J)}(t) = \sum_{\alpha=1}^{A} \lambda^\alpha(t) \cdot C_{(i,J)}^\alpha(t), \ \forall i, J, t. \tag{11}$$

We can also have an opportunistic routing based scheduling problem that aims to achieve the maximum network utility by optimally scheduling the transmission of all the maximum CTSs. Accordingly, the constraints for the maximum CTS scheduling are given as follows:

$$\sum_{\alpha=1}^{A} \lambda^\alpha(t) \leq 1, \ \forall t, \tag{12}$$

where constraint (12) ensures that no more than one maximum CTS is scheduled to transmit at any time within time slot t.

III. A Joint Dynamic Rate Allocation and Transmission Scheduling Problem With Network Coding-Based Opportunistic Routing

A. Problem Formulation

In wireless networks, the mobility of wireless nodes would often occur. Thus the channel state information of a wireless link varies over time. In this work, we assume that the SVC multicast problem is required to be adapted to link (i, j)'s time-varying packet reception ratio $p_{(i,j)}(t)$, which is selected to characterize the dynamic network changes in practical wireless networks. Previous work on scalable video multicast is mainly based on static network assumption and predetermined routes between source-destinations, and thus fails to capture the temporal dynamics in practical networks. Different from these static optimization formulation, we assume in this work that all the links' PRRs are dynamic with regard to time slots and either known or predicted for all time slots $\{1, 2, \ldots, T\}$ within the SVC multicast period. And the decision of optimal routes for each destination node subscribing to different SVC layers is integrated by opportunistic multicast routing. In such scalable video multirate multicast applications, in order for a destination node to meet the SVC decoding requirement and successfully recover layer m, there is a requirement on the total amount of information belonging to layer m delivered through the opportunistic routes to the destination over the entire transmission time interval from time slot 1 to T. Such requirement can also be viewed as an SVC video delivery contract that couples the reception rate of each destination node across time [34] and is given in (1). Accordingly, we adopt the multi-period formulation of dynamic network utility maximization with delivery contract to formulate a joint dynamic rate allocation and transmission scheduling optimization problem based on opportunistic multicast routing, where the received rates by destination nodes are coupled across time by SVC encoding/decoding constraints. The proposed opportunistic multicast routing formulation for scalable video streaming is as follows:

$$\mathbf{P1}: \quad \max_{(\mathbf{R},\mathbf{g},\mathbf{f},\boldsymbol{\lambda}) \succeq \mathbf{0}} \sum_{t=1}^{T} \sum_{d \in D} \sum_{m=1}^{M} U\left(R^{m\,d}(t)\right)$$

$$\mathbf{s.t.} \quad \text{Constraints in (1), (2), (5), (7), (8), (11), and (12),} \tag{13}$$

where the objective is to achieve the maximum total utility of scalable video layers received by all recipients within the duration of SVC multicast streaming. The SVC encoding/decoding constraint in (1) is also the SVC video delivery contract [34] that couples the reception rate of each destination node across time. On one hand, since the video rate variable $\boldsymbol{R}$ is coupled across time in (1), problem **P1** cannot be split into separate sub-problems with each sub-problem corresponding to a time slot, and thus the choice of $\boldsymbol{R}$ over all time slots need to be coordinated. On the other hand, compared to the sub-optimal performance achieved by the separate optimization within each time slot, such coordination among $\boldsymbol{R}$ over all time slots can achieve the global optimum and benefit the overall multicast throughput by allocating a larger video reception rate than the

total bitrate of the SVC stream during time slots with good channel conditions and reserving such extra rate for time slots with bad conditions. To achieve the best overall video reception quality, the proposed scheme jointly optimizes the video reception rate $\mathbf{R}$, the routes to destination nodes with associated virtual information flow $\boldsymbol{g}$ and physical flow $\boldsymbol{f}$, and the time fraction of maximum CTSs $\boldsymbol{\lambda}$.

B. Stochastic Dynamic Problem

In the above formulation **P1**, we assume that the channel state information (specifically, the dynamic packet reception ratio for each link) is known a priori for all time slots, and the joint rate allocation, opportunistic routing and transmission scheduling is determined based on it. In practice, such channel state information is not fully known ahead of time, and is revealed only at each time interval. In the following, we will describe the extension to **P1** and develop accordingly the stochastic dynamic version of the problem formulation **P1**. In this case, the resulting performance after solving **P1** can be treated as the theoretical upper bound for the stochastic dynamic problem with imperfect channel state information.

Based on the model predictive control [34], [35], we impose a causality constraint that the links' PRRs at time slot t must be a function of the PRRs up to time t. Suppose that at time slot $t=\tau$, the links' PRR vector $\boldsymbol{p}(1),\ldots,\boldsymbol{p}(\tau)$ are known. Then, given the information available at time slot τ, the expected value of the link's PRR $\hat{\boldsymbol{p}}(t|\tau)$ can be defined as:

$$\hat{\boldsymbol{p}}(t|\tau)=\boldsymbol{E}[\boldsymbol{p}(t)|\boldsymbol{p}(1),\ldots,\boldsymbol{p}(\tau)],\ \forall t\in\{\tau+1,\ldots,T\}. \tag{14}$$

Then, by replacing the link PRR with its expected value in (9), we have the expected transmission rate of hyperedge (i,J) in maximum CTS $\Gamma^{\alpha}(t)$:

$$\hat{C}^{\alpha}_{(i,J)}(t|\tau)=\psi^{\alpha}_{i}(t)\cdot C_i\left[1-\prod_{j\in J}(1-\hat{p}_{(i,j)}(t|\tau))\right],\ \forall t. \tag{15}$$

Then the expected effective transmission rate of (i,J) is:

$$\hat{C}_{(i,J)}(t|\tau)=\sum_{\alpha=1}^{A}\lambda^{\alpha}(t)\cdot\hat{C}^{\alpha}_{(i,J)}(t|\tau),\ \forall i,J,t. \tag{16}$$

Therefore, the stochastic dynamic problem can be formulated as:

$$\mathbf{P2}:\quad \max_{(\mathbf{R},\mathbf{g},\mathbf{f},\boldsymbol{\lambda})|_{t=\tau}^{T}\succeq\mathbf{0}}\sum_{t=\tau}^{T}\sum_{d\in D}\sum_{m=1}^{M}U\left(R^{m\,d}(t)\right)$$

$$\mathbf{s.t.}\quad \text{Constraints in (1) and (2),}$$

$$\text{Constraints in (5), (7), and (12)},\forall t\in\{\tau,\ldots,T\},$$

$$\sum_{j\in J}\sum_{m=1}^{M}f^{m}_{(i,j)}(\tau)\le C_{(i,J)}(\tau),\ \forall i\in V,J\subseteq F_i(\tau),$$

$$\sum_{j\in J}\sum_{m=1}^{M}f^{m}_{(i,j)}(t)\le\hat{C}_{(i,J)}(t|\tau),$$

$$\forall i\in V,J\subseteq F_i(t),t\in\{\tau+1,\ldots,T\}. \tag{17}$$

In the stochastic dynamic problem **P2**, at a given time slot $t=\tau$, all the variables for the time slots up to $\tau-1$ are already known or solved, and thus the optimization variable becomes the vector $(\mathbf{R},\mathbf{g},\mathbf{f},\boldsymbol{\lambda})|_{t=\tau}^{T}$ for the current time slot τ and future time slots $\tau+1,\ldots,T$. Here we use the exact value of the current effective transmission rate of hyperedge (i,J), $C_{(i,J)}(\tau)$, since its value is known. However, for the future effective transmission rate, we use instead the conditional mean $\hat{C}_{(i,J)}(t|\tau)$ in (16).

It can be seen from the formulation of **P2** that its performance depends on the prediction of the links' PRR vector given in (14). According to [36]–[38], a good approximation in modeling the time-varying error process of a wireless link (i,j) can be provided by the ergodic finite state Markov channel (FSMC) model. It not only provides a good approximation to the statistical properties of the real channels, but is mathematically tractable. In practical applications, it has also been widely accepted in the literature as an effective approach to characterizing the correlation structure of the fading process. By using the FSMC model, suppose that the wireless channel of link (i,j) has a finite set of H states corresponding to H different packet reception probabilities, denoted by vector $\mathbf{P}_{(i,j)}=[P^{1}_{(i,j)},\ldots,P^{h}_{(i,j)},\ldots,P^{H}_{(i,j)}]^{T}$, and a transition probability matrix $\mathbf{Tr}_{(i,j)}\in\mathbb{R}_{+}^{H\times H}$ which has the following structure:

$$\mathbf{Tr}_{(i,j)}=\begin{pmatrix}\omega_{1,1}&\omega_{1,2}&0&0&0&\ldots&0&0\\ \omega_{2,1}&\omega_{2,2}&\omega_{2,3}&0&0&\ldots&0&0\\ 0&\omega_{3,2}&\omega_{3,3}&\omega_{3,4}&0&\ldots&0&0\\ \vdots&\vdots&\vdots&\vdots&\vdots&\ddots&\vdots&\vdots\\ 0&0&0&0&0&\ldots&\omega_{H,H-1}&\omega_{H,H}\end{pmatrix}, \tag{18}$$

where $\omega_{x,y}$ is the transition probability from state index x to y. Suppose that at time slot $t=\tau$, we have observed the state of link (i,j) with associated PRR, as $p_{(i,j)}(\tau)=P^{h}_{(i,j)}$. Based on the transition probability that is independent of time slot, the expected PRR of that link at time slot $\tau+1$ given the information available at time slot τ can be predicted as:

$$\begin{aligned}\hat{p}_{(i,j)}(t=\tau+1|\tau)&=[\mathbf{Tr}_{(i,j)}]_{(h)}\cdot\mathbf{P}_{(i,j)}\\&=\omega_{h,h}P^{h}_{(i,j)}+\omega_{h,h-1}P^{(h-1)}_{(i,j)}+\omega_{h,h+1}P^{(h+1)}_{(i,j)},\end{aligned} \tag{19}$$

where $[\mathbf{Tr}_{(i,j)}]_{(h)}$ denotes the h-th row of $\mathbf{Tr}_{(i,j)}$. By recursively using (19), all the expected value of future PRRs can be predicted by the h-th row of the $t-\tau$-th power of $\mathbf{Tr}_{(i,j)}$:

$$\hat{p}_{(i,j)}(t|\tau)=[\mathbf{Tr}^{(t-\tau)}_{(i,j)}]_{(h)}\cdot\mathbf{P}_{(i,j)},\quad t=\tau+1,\tau+2,\ldots,T. \tag{20}$$

Observing (19) and (20), the future wireless links' PRR prediction is determined by the packet reception probability vector $\boldsymbol{P}_{(i,j)}$ and the transition probability matrix $\boldsymbol{Tr}_{(i,j)}$. For Rayleigh fading channel [37], denote $\boldsymbol{\Omega} = [\Omega^1, \ldots, \Omega^h, \ldots, \Omega^{H+1},]^T$ as the received signal-to-noise (SNR) thresholds for all the H states in a increasing order with $\Omega^1 = 0$ and $\Omega^{H+1} = \infty$. The channel is in state h if the received SNR located in the rage $[\Omega^h, \Omega^{h+1})$. Since $\boldsymbol{P}_{(i,j)}$ is fixed for a given $\boldsymbol{\Omega}$, in the following, we will further analyze the impact of the transition probability on the PRR prediction by taking into account the path loss and node movement. According to [37], the transition probability is given by:

$$\omega_{h,h+1} = \frac{N(\Omega^{h+1}) \cdot T_p}{\pi^h},\ \omega_{h,h-1} = \frac{N(\Omega^h) \cdot T_p}{\pi^h}, \tag{21}$$

where T_p is the time duration of each packet; π^h represents the steady-state probability; and $N(\Omega^h)$ denotes the level cross-rate of state h (either in the positive direction only, or in the negative direction only) and is given by [37]:

$$N(\Omega^h) = \sqrt{\frac{2\pi\Omega^h}{\gamma_0}} \cdot \psi_m \cdot \exp\left(-\frac{\Omega^h}{\gamma_0}\right). \tag{22}$$

Here, γ_0 is the average received SNR. Considering the node movement, the fading characteristics of the signal envelop are determined by the Doppler frequency due to the motion of a wireless node. Accordingly, ψ_m in (22) denotes the maximum Doppler frequency caused by the node movement at a certain speed v, and can be defined as the node movement speed v divided by the carrier wavelength [36]. Thus, the duration of a time slot, i.e., the time duration over which the channel's response (or state) remains invariant, can be approximated as [39]:

$$\Delta T \approx \frac{1}{\psi_m}. \tag{23}$$

In addition, it can be observed from (21) and (22) that the transition probability is also determined by the average SNR γ_0, which (in dBm) can be derived by subtracting the receiver noise power (dB) and the path loss (dB) due to radio propagation from the transmitter output power (dBm). Here, we adopt the following simplified model for path loss as a function of the distance between transmitter and receiver [40]:

$$PL\ (\text{dB}) = 10\sigma \log_{10}\left[\frac{dist}{dist_0}\right] - K\ (\text{dB}). \tag{24}$$

This path loss model is commonly used for wireless system design, where σ represents the path-loss exponent, $dist_0$ denotes a reference distance for the antenna far field, and K is a unitless constant dependent on the antenna characteristics and the average channel attenuation. Due to the scattering phenomena in the antenna near field, (24) is valid for transmission distance $dist > dist_0$, where $dist_0$ is typically assumed to be 1–10 meters indoors and 10–100 meters outdoors.

IV. Distributed Cross-Layer Algorithm

A. Optimization Decomposition and Distributed Solution

For distributive implementation, decomposition methods (primal or dual decomposition) are commonly used [41] to decompose a large optimization problem into a set of small subproblems, which can be solved by distributed and often iterative algorithm converging to the global optimum. Due to space limit, we will only introduce the distributed solution to the stochastic dynamic optimization problem **P2**, and the distributed solution to **P1** can be developed in a similar way. With dual decomposition, problem **P2** can be decoupled by relaxing all the coupling constraints with Lagrange multiplier vector $\boldsymbol{\vartheta} = (\mu, \eta, \theta, \beta, \zeta, \nu)$. Specifically, the Lagrangian of **P2** is obtained in (25) shown at the top of the next page. Accordingly, the Lagrange dual function is given by:

$$g(\boldsymbol{\vartheta}) = \sup_{\mathbf{R},\mathbf{g},\mathbf{f},\boldsymbol{\lambda}\succeq\mathbf{0}} L(\mathbf{R}, \mathbf{g}, \mathbf{f}, \boldsymbol{\lambda}, \boldsymbol{\vartheta}). \tag{26}$$

And the Lagrange dual problem of **P1** is expressed as:

$$\min_{\eta,\theta,\beta,\zeta,\nu\succeq\mathbf{0}} g(\boldsymbol{\vartheta}). \tag{27}$$

Since the objective function is concave and all the constraints are convex with respect to the primal variables, **P2** is a convex optimization problem. Optimization theory [41] ensures that the convex constrained optimization problem **P2** is equivalent to its Lagrange dual problem in (27), and thus can be decomposed into a master dual problem in (27) with several cross-layer subproblems of the transport, network, and link layers. These subproblems correspond to the maximization of the Lagrangian in (25) with regard to the primal variables $\mathbf{R}$, $\mathbf{g}$, $\mathbf{f}$, and $\boldsymbol{\lambda}$, respectively. Since the Lagrangian in (25) is partially differentiable with respect to the primal dual variables, both the master dual problem and the cross-layer sub-problems can be solved by the gradient algorithm [42]. Based on the primal-dual algorithm that simultaneously updates primal and dual variables, the distributive cross-layer iteration algorithm is developed in (28)–(33), where subscript k denotes the iteration index, $\boldsymbol{\delta}$ is positive step-size, and $[\cdot]^+$ represents the projection onto $\mathbb{R}_+$. Specifically, at each iteration, each primal/dual variable is updated by a new value in its gradient ascent/descent direction which is closer to the local maximum/minimum.

1) Source Rate Control, $\forall t \in \{\tau, \ldots, T\}$:

$$\begin{aligned} R^{md}(t)|_{k+1} &= \left[R^{md}(t) + \delta^{(R)} \cdot \frac{\partial L(\mathbf{R}, \mathbf{g}, \mathbf{f}, \boldsymbol{\lambda}, \boldsymbol{\vartheta})}{\partial R^{md}(t)}\right]^+_k \\ &= \Bigg[R^{md}(t) + \delta^{(R)}\Bigg(U'(R^{md}(t)) + \mu_s^{md}(t) - \mu_d^{md}(t) \\ &\quad - \frac{\nu^{(m-1)d}}{r^m_{\max}} + \frac{\nu^{md}}{r^m_{\min}} - \zeta^{md}\Bigg\{3\left[\sum_{t=1}^{T} R^{md}(t)\right]^2 \\ &\quad - 2T(r^m_{\min} + r^m_{\max})\left[\sum_{t=1}^{T} R^{md}(t)\right] + T^2 r^m_{\min} r^m_{\max}\Bigg\}\Bigg)\Bigg]^+_k. \end{aligned} \tag{28}$$

$$L(\mathbf{R},\mathbf{g},\mathbf{f},\boldsymbol{\lambda},\boldsymbol{\vartheta}) = L(\mathbf{R},\mathbf{g},\mathbf{f},\boldsymbol{\lambda},\mu,\eta,\theta,\beta,\zeta,\nu) = \sum_{t=\tau}^{T}\sum_{d\in D}\sum_{m=1}^{M} U(R^{md}(t)) - \sum_{t=\tau}^{T}\sum_{d\in D}\sum_{m=1}^{M}\sum_{i\in V}\mu_i^{md}(t)\Bigg[\sum_{j:(i,j)\in E} g_{(i,j)}^{md}(t) - \sum_{j:(j,i)\in E} g_{(j,i)}^{md}(t) - \sigma_i^{md}(t)\Bigg] - \sum_{t=\tau}^{T}\sum_{d\in D}\sum_{m=1}^{M}\sum_{(i,j)\in E}\eta_{(i,j)}^{md}(t)\Big[g_{(i,j)}^{md}(t) - f_{(i,j)}^{m}(t)\Big] - \sum_{t=\tau}^{T}\theta(t)\Bigg[\sum_{\alpha=1}^{A}\lambda^{\alpha}(t) - 1\Bigg] - \sum_{i\in V}\sum_{J\subseteq F_i(\tau)}\beta_{(i,J)}(\tau)\cdot\Bigg[\sum_{j\in J}\sum_{m=1}^{M} f_{(i,j)}^{m}(\tau) - \sum_{\alpha=1}^{A}\lambda^{\alpha}(\tau)\cdot C_{(i,J)}^{\alpha}(\tau)\Bigg] - \sum_{t=\tau+1}^{T}\sum_{i\in V}\sum_{J\subseteq F_i(t)}\beta_{(i,J)}(t)\Bigg[\sum_{j\in J}\sum_{m=1}^{M} f_{(i,j)}^{m}(t) - \sum_{\alpha=1}^{A}\lambda^{\alpha}(t)\cdot\hat{C}_{(i,J)}^{\alpha}(t|\tau)\Bigg] - \sum_{d\in D}\sum_{m=1}^{M}\zeta^{md}\Bigg[\sum_{t=1}^{T}R^{md}(t)\Bigg]\Bigg[\sum_{t=1}^{T}R^{md}(t) - r_{\min}^{m}T\Bigg]\Bigg[\sum_{t=1}^{T}R^{md}(t) - r_{\max}^{m}T\Bigg] - \sum_{d\in D}\sum_{m=1}^{M-1}\nu^{md}\cdot\left[\frac{\sum_{t=1}^{T}R^{(m+1)d}(t)}{r_{\max}^{(m+1)}} - \frac{\sum_{t=1}^{T}R^{md}(t)}{r_{\min}^{m}}\right]. \tag{25}$$

2) Link Flow Control and Routing, $\forall t\in\{\tau,\ldots,T\}$:

$$g_{(i,j)}^{md}(t)|_{k+1} = \left[g_{(i,j)}^{md}(t) + \delta^{(g)}\cdot\frac{\partial L(\mathbf{R},\mathbf{g},\mathbf{f},\boldsymbol{\lambda},\boldsymbol{\vartheta})}{\partial g_{(i,j)}^{md}(t)}\right]_k^+ = \left[g_{(i,j)}^{md}(t) + \delta^{(g)}\Big(\mu_i^{md}(t) - \mu_j^{md}(t) - \eta_{(i,j)}^{md}(t)\Big)\right]_k^+. \tag{29}$$

$$f_{(i,j)}^{m}(t)|_{k+1} = \left[f_{(i,j)}^{m}(t) + \delta^{(f)}\cdot\frac{\partial L(\mathbf{R},\mathbf{g},\mathbf{f},\boldsymbol{\lambda},\boldsymbol{\vartheta})}{\partial f_{(i,j)}^{m}(t)}\right]_k^+ = \left[f_{(i,j)}^{m}(t) + \delta^{(f)}\Bigg(\sum_{d\in D}\eta_{(i,j)}^{md}(t) - \sum_{\substack{J\subseteq F_i(t)\\ \text{and } j\in J}}\beta_{(i,J)}(t)\Bigg)\right]_k^+. \tag{30}$$

3) Maximum CTS Scheduling, $\forall t\in\{\tau,\ldots,T\}$:

$$\lambda^{\alpha}(\tau)|_{k+1} = \left[\lambda^{\alpha}(\tau) + \delta^{(\lambda)}\cdot\frac{\partial L(\mathbf{R},\mathbf{g},\mathbf{f},\boldsymbol{\lambda},\boldsymbol{\vartheta})}{\partial\lambda^{\alpha}(\tau)}\right]_k^+ = \left[\lambda^{\alpha}(\tau) + \delta^{(\lambda)}\Bigg(\sum_{i\in V}\sum_{J\subseteq F_i(\tau)}\beta_{(i,J)}(\tau)C_{(i,J)}^{\alpha}(\tau) - \theta(\tau)\Bigg)\right]_k^+, \tag{31}$$

and for any $t\in\{\tau+1,\ldots,T\}$, we have

$$\lambda^{\alpha}(t)|_{k+1} = \left[\lambda^{\alpha}(t) + \delta^{(\lambda)}\cdot\frac{\partial L(\mathbf{R},\mathbf{g},\mathbf{f},\boldsymbol{\lambda},\boldsymbol{\vartheta})}{\partial\lambda^{\alpha}(t)}\right]_k^+ = \left[\lambda^{\alpha}(t) + \delta^{(\lambda)}\Bigg(\sum_{i\in V}\sum_{J\subseteq F_i(t)}\beta_{(i,J)}(t)\hat{C}_{(i,J)}^{\alpha}(t|\tau) - \theta(t)\Bigg)\right]_k^+. \tag{32}$$

4) Dual Variable Update, $\forall t\in\{\tau,\ldots,T\}$:

$$\boldsymbol{\vartheta}|_{k+1} = \left[\boldsymbol{\vartheta} - \delta^{(\vartheta)}\cdot\frac{\partial L(\mathbf{R},\mathbf{g},\mathbf{f},\boldsymbol{\lambda},\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right]_k^+. \tag{33}$$

It should be noted that $[\cdot]^+$ will not be applied to the update iteration of μ since it is associated with the equality constraint and thus might have negative value.

B. Algorithm Design and Implementation Details

To implement the proposed distributed cross-layer algorithm, each node $i\in V$ in the wireless network is treated as an entity capable of processing, storing, and communicating information. In practice, the associated wireless link (i,j) and hyperedge (i,J) are both delegated to their sender node i, and all computations related to such links and hyperedges will be executed on node i. A distributed implementation of the proposed iterative algorithm is summarized in Algorithm 1.

Once the appropriate rate allocation for destination nodes and the time fraction scheduling among different maximum CTSs are determined through Algorithm 1, the source node s will keep sending multiple flows of coded packets that are belonging to different SVC layers in accordance with the optimal source rate. Each transmitter node will be allocated the corresponding time fraction for sending or relaying the coded packets to all of its next-hop nodes through opportunistic routing. Upon receiving a coded packet for a specific layer, each relay node would check whether this coded packet is linearly independent from the packets that have been previously received. If so, the relay node keeps this coded packet as an innovative packet, generates a random linear combination of the coded packets it has heard from the same layer, and broadcasts it. Otherwise, this packet is considered as non-innovative and thus dropped.

In terms of amount of message passing, the overhead of the proposed distributed algorithm consists of two parts: the

Algorithm 1: Distributed Algorithm for the Joint Rate Allocation, Link Routing and Transmission Scheduling Problem

Initialization

Set an initial primal/dual point $(\mathbf{R}, \mathbf{g}, \mathbf{f}, \boldsymbol{\lambda}, \boldsymbol{\vartheta})|_{k,\ k=0}$ to some nonnegative value for $t \in \{\tau, \ldots, T\}$.

repeat

for the source node s:

1) Receive $\mu_d^{md}(t)|_k$ from the destination node d;
2) Receive $\mu_j^{md}(t)|_k$ from the next-hop node j;
3) Fetch $R^{md}(t)|_k$, $\mu_s^{md}(t)|_k$, $\nu^{md}|_k$, $\zeta^{md}|_k$, $\eta^{md}_{(s,j)}(t)|_k$, $\beta_{(s,J)}(t)|_k$, $\theta(t)|_k$ and $C^{\alpha}_{(s,J)}(\tau)$ stored in the local processor;
4) Calculate the expected value of the effective transmission rate $\hat{C}^{\alpha}_{(s,J)}(t|\tau)$ according to (9)–(11) and (20);
5) Calculate $R^{md}(t)|_{k+1}$, $g^{md}_{(s,j)}(t)|_{k+1}$, $f^{m}_{(s,j)}(t)|_{k+1}$, $\lambda^{\alpha}(t)|_{k+1}$ and $\boldsymbol{\vartheta}|_{k+1}$ according to (28)–(33), respectively;
6) Send $\theta(t)|_{k+1}$ to the nodes in $\{N \cup D\}$.

for each node $i \in N \cup D$:

1) Receive $\theta(t)|_k$ from the source node s;
2) Receive $\mu_j^{md}(t)|_k$ from the next-hop node j;
3) Fetch $\mu_i^{md}(t)|_k$, $\eta^{md}_{(i,j)}(t)|_k$, $\beta_{(i,J)}(t)|_k$ and $C^{\alpha}_{(i,J)}(\tau)$ stored in the local processor;
4) Calculate the expected value of the effective transmission rate $\hat{C}^{\alpha}_{(i,J)}(t|\tau)$ according to (9)–(11) and (20);
5) Calculate $g^{md}_{(i,j)}(t)|_{k+1}$, $f^{m}_{(i,j)}(t)|_{k+1}$, $\lambda^{\alpha}(t)|_{k+1}$ and $\boldsymbol{\vartheta}|_{k+1}$ according to (29)–(33), respectively;
6) Send $\mu_i^{md}(t)|_{k+1}$ to the sender nodes in $\{j \in V | (j,i) \in E\}$;
7) Send $\mu_i^{md}(t)|_{k+1}$ to the source node s, if node $i = d, d \in D$.

until All the primal and dual variables converge to the optimums, or the predefined maximum number of iterations is achieved.

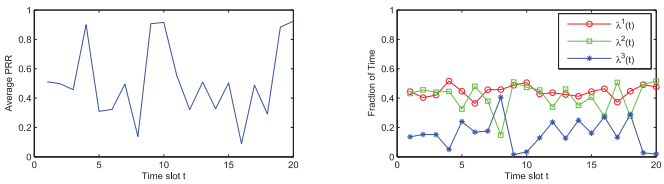
network coding overhead and the communication overhead. For network coding operations, we adopt a packet format similar to [17], [22] with each packet carrying 1500 bytes of data. Therefore, the side information required by network coding is the global encoding vector in the header of each packet, which is related to the number of source packets that need to be transmitted. Therefore, the cost of the practical network coding scheme is the overhead of transmitting extra symbols for the encoding vector in each packet. Considering a large size packet with the number of both header and payload symbols far more than the number of header symbols, such overhead is about 1–2%, and can be neglected [22].

On the other hand, as shown in Algorithm 1, the communication overhead introduced by the proposed algorithm is the transmission of updated $\theta(t)$ value from the source node s to the downstream nodes $\{N \cup D\}$, as well as the the transmission of updated $\mu_i^{md}(t)$ value to the previous-hop sender nodes of node i and to the source node s if node i is a destination node. Consider the implementation issues [43] and take the network topology in Fig. 1(a) for example. If we further assume that each updated variable is in float type which takes up 4 bytes, and let $M = 3$ and $T = 5$, then the communication overhead for source node s to transmit $\theta(t)$ is $5 \times 4 = 20$ bytes, while the communication overhead for each node i to transmit $\mu_i^{md}(t)$ in the upstream direction requires $3 \times 3 \times 5 \times 4 = 180$ bytes. Thus, the total communication overhead sums up to 200 bytes. For the same configuration of IP packet with a packet size of 1500 bytes, the communication overhead introduced by the proposed algorithm is approximately $200/1500 = 13\%$. Furthermore, it can be noted that these dual variable ($\theta(t)$ and $\mu_i^{md}(t)$) in practical implementation do not need to be communicated as separate packets. Instead, variable $\theta(t)$ can be delivered through a field in the video data packets, while variable $\mu_i^{md}(t)$ can be conveyed through a field in the acknowledgement (ACK) packets. At each time when the algorithm starts, these dual variables are required to be transmitted, as the communication overhead of the first several packets, for the updating iterations in Algorithm 1. However, once the convergence is reached, the optimization outputs are determined and will hold within the whole time slot. From then on, the communication overhead for carrying these dual variables is no longer needed.

V. Experimental Results

In this section, we evaluate the performance of the proposed distributed cross-layer algorithm (PDCA) under different network settings, and illustrate the effectiveness of PDCA over four baseline schemes: 1) JAOR, joint source and flow rate optimization algorithm [9] based on OR in [15]; 2) JATR, joint source and flow rate optimization algorithm [9] based on traditional routing and link conflict graph in [29]; 3) LATR, layered multicast algorithm [7] based on traditional routing and link conflict graph in [29], and 4) DRDA, distributed receiver-driven streaming algorithm [11] that exploits randomized network coding and prioritized distributed video delivery for efficient multicast. It should be noted that OR is not utilized in the original formulation of JATR [9]. Here, we integrate OR into JAOR to achieve better performance than JATR as proposed in [9]. Alternatively, the key difference among the above five schemes is that JAOR/JATR(or LATR) are based on the static network assumption with/without OR and DRDA is suitable for dynamic network without utilization of OR, respectively, while PDCA can deal with the practical dynamic networks with OR.

In the following, we use Joint Scalable Video Model 7_ 10 reference codec of H.264/AVC extension standard [44] at the source node, to encode three well-known test-sequences (*Bus*, *Football*, and *Foreman*) at frame rate of 30 frames per second, CIF (352×288) resolution, and a GOP-length of 32 frames. These sequences are encoded with 256 Kbps at the base layer, and 384 Kbps, 512 Kbps and 1024 Kbps at the three enhancement layers by fine granularity scalable coding. As shown in [45], the objective QoE of video streaming may be affected by factors such as the average and temporal variability of video quality, time spent in re-buffering, and the initial startup delay.



(a) (b)

Fig. 2. (a) A random PRR sample, and (b) optimal time fraction scheduling of the maximum CTSs.

In this paper, we mainly focus on the dynamic network utility maximization problem formulation for the SVC multicasting to achieve a maximum aggregate network utility in terms of video quality. Therefore, the average video quality measure in the form of PSNR is selected as the quality metric to evaluate the overall users' satisfaction level in the performance evaluation.

A. Performance Evaluation on Illustrative Topology

In this section, each evaluation point in our analysis is the average performance computed over 20 simulation runs with the same simulation configuration. First, we simulate PDCA and compare its performance to the other four baseline schemes on the illustrative network topology as shown in Fig. 1(a). Here, s, n_i, and d_i represent the source node, relay node, and destination node, respectively. Suppose that the SVC streaming period is divided into twenty time slots. The transmission rate for each transmitter node is set to 7.5 Mbps, and the packet reception rate (PRR) of each wireless link is varied over different time slots. In the following of this section, we will evaluate the performance achieved by all the four algorithms both with random link PRR samples and with the dynamic link PRRs predicted by the FSMC model.

1) Random PRR Sample: Assume that the dynamic PRR of each wireless link in Fig. 1(a) is randomly generated over time. Fig. 2(a) shows the average PRR value of all links at each time slot, and thus indicates the time-varying network condition. Accordingly, the optimal time fraction of the three maximum CTSs scheduled by PDCA is shown in Fig. 2(b), where $\Gamma^1(t) = \{s\}$, $\Gamma^2(t) = \{n_2\}$, and $\Gamma^3(t) = \{n_1, n_3\}$. It can be seen that, since s is the source node which always generates video packets for SVC streaming, no matter whether the channel condition is good or not, the optimal time fraction allocated for $\Gamma^1(t)$ at each time slot remains a relatively large value (e.g., around 0.4) with little temporal variation. On the other hand, when the network condition is good with higher PRRs (e.g., at $t = 4$, $t = 10$, or $t = 20$), PDCA would schedule a larger transmission time fraction for transmitter nodes with more next-hop nodes (e.g., node n_2 in $\Gamma^2(t)$). In contrast, when the network condition is not good, wireless links become unreliable with lower PRRs (e.g., at $t = 8$, or $t = 16$). At this time, PDCA could increase the time fraction and thus the utilization of other transmitter nodes (e.g., nodes n_1 and n_3 in $\Gamma^3(t)$) to combine multiple weak links into one strong link.

To further study the impact of different transmission schedules, as illustrated in Fig. 3, we compare the rate allocation results achieved by the optimal transmission time scheduling of PDCA to two fixed time scheduling settings, i.e., $\lambda^1(t) = 1/2$, $\lambda^2(t) = \lambda^3(t) = 1/4$ and $\lambda^1(t) = \lambda^2(t) = \lambda^3(t) = 1/3$. As can be noted, for some destination nodes (e.g., d_2 and d_3), the allocated rate by the optimal transmission time scheduling might be similar to the schedule of letting $\lambda^1(t) = 1/2$, $\lambda^2(t) = \lambda^3(t) = 1/4$. However, much more rate will be allocated for destination node d_1 by the optimal transmission time scheduling, and thus the overall allocated rate for all the three destination nodes is larger than the other two fixed transmission time schedules.

The convergence behavior of the proposed algorithm is illustrated in Fig. 4, where we take time slot $t = 3$ for example and show the iterations of the total allocated rates for all the three destination nodes at $t = 3$. It can be seen that with different setups of the initial rate values, all the allocated rates can simultaneously and quickly converge to the corresponding optimal values in a few tens of iterations (e.g., 90 iterations for d_1 and d_3, and 125 iterations for d_2). If the initial values of the allocated rates are set closer to the optimal solution, fewer number of iterations is needed for convergence. In addition, it is also the case that the allocated rates converge to modest accuracy within less number of iterations than needed for convergence. For example, the allocated rates for both d_1 and d_3 approach within 10% of their optimal values after 50 iterations, while the allocated rate for d_2 reaches within 10% of its optimal value after 90 iterations, which is sufficient for the practical applications. This observation indicates that in some practical cases where the convergence to high accuracy is relatively slow, a sub-optimal performance with modest accuracy can be achieved with much less computational iterations and thus much lower computational complexity. In terms of the algorithm running time, the average running time spent in one iteration in the simulation is 0.03 ms. Therefore, the total algorithm running time to approach the steady state (e.g., after 100 iterations) is 3 ms, which is very short compared with the duration of the time slot ($\Delta T = 100$ ms) with a ratio of 0.03.

Fig. 5 shows the destination node reception rates allocated by different algorithms. It can be seen that by PDCA and JAOR, more video reception rate is allocated for each destination node at each time slot than JATR, LATR and DRDA. Such network multicast throughput gain is achieved due to the key difference between OR and traditional routing. That is, with OR the wireless shared medium time utility is enhanced by allowing multiple neighbor nodes to forward a packet opportunistically, which in turn makes throughput take place concurrently at multiple outgoing links from the same transmitter node. With traditional routing used in JATR, LATR and DRDA, in contrast, since one transmitter node cannot simultaneously send a packet to multiple relay nodes due to wireless link interference, the node transmission capacity is constrained by one of its outgoing links with the highest transmission rate. In addition, compared to JAOR, the network multicast throughput (i.e., the total video reception rate allocated for all the three destination nodes) is further improved by PDCA. This is because PDCA considers the network dynamics by introducing the SVC video delivery contract and optimizes the total network utility in the form of video reception qualities by balancing and adapting network

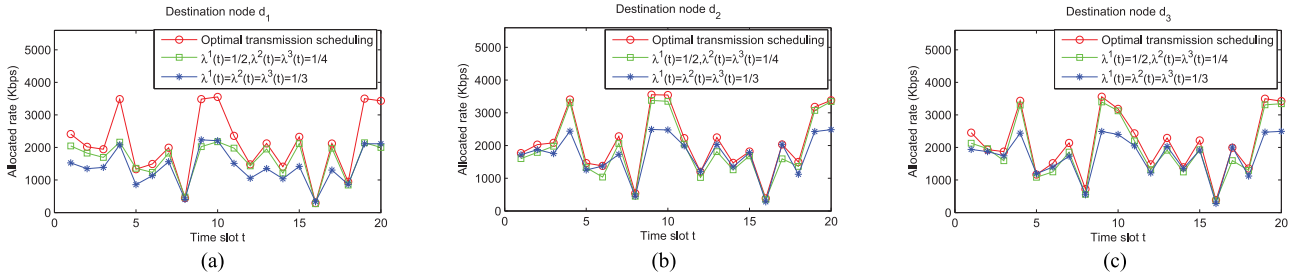


Fig. 3. Impact of different time fraction schedules on allocated rate for destination nodes (a) d_1, (b) d_2, and (c) d_3.

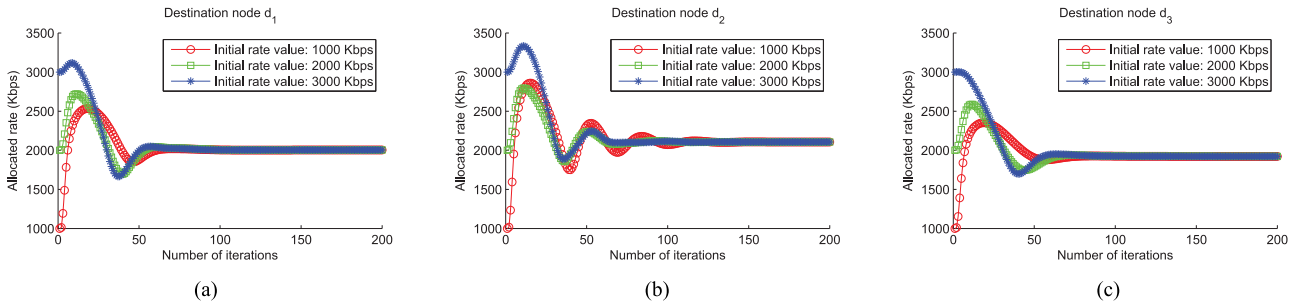


Fig. 4. Convergence behavior of allocated rate at time slot $t = 3$ for destination nodes (a) d_1, (b) d_2, and (c) d_3.

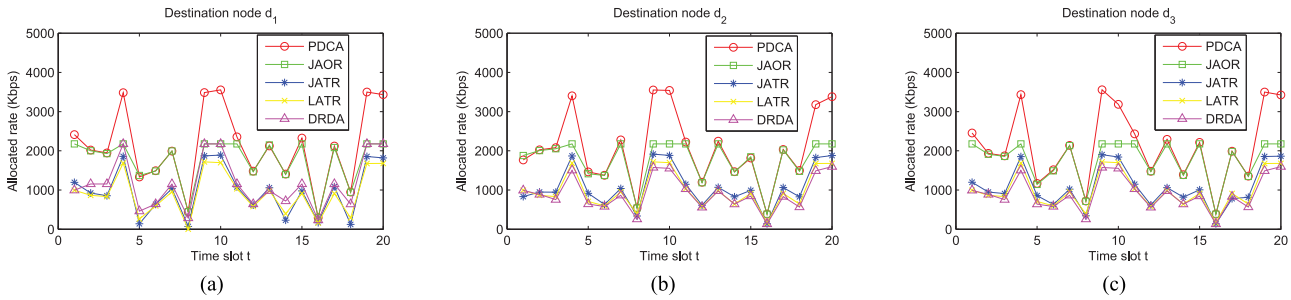


Fig. 5. Comparison of allocated rate achieved by different algorithms for destination nodes (a) d_1, (b) d_2, and (c) d_3.

flow over all time slots within the SVC streaming period, while JAOR only individually optimizes within each time slot. Specifically, when the network condition is not good (e.g., at $t = 8$, or $t = 16$), both algorithms assign the same amount of reception rate for all destination nodes. However, during the time slots with higher PRRs (e.g., at $t = 4$, $t = 9$, $t = 10$, or $t = 20$), PDCA would allocate more data rate than needed for that time slot and reserve such extra rate for time slots with bad conditions to increase the overall utility, while JAOR fails to do so because all the four video layers are already fully received at $t = 4$, $t = 9$, $t = 10$, $t = 20$ and no more rate could be allocated. This extra rate allocation of PDCA is achieved by solving the proposed optimization problem **P1** (when the future channel information is known in a priori) or the stochastic dynamic problem **P2** (when the future channel information is predicted based on the FSMC model) to obtain the optimal video rate allocation variable $\boldsymbol{R}$ for all the time slots. For example, the total allocated rate for destination d_1 by JAOR at $t = 4$ is 2176 Kbps, which is equal to the total rate of all the four video layers. In comparison, PDCA implements Algorithm 1 to solve the optimization problem **P1** for $t = 1, 2, \ldots, 20$ and to obtain the video rate allocation variable $\boldsymbol{R}$ for destination d_1 at $t = 4$, with the total rate being $\sum_{m=1}^{4} R^{m d_1}(t = 4) = 3485$ Kbps. The extra rate of $3485 - 2176 = 1309$ Kbps at $t = 4$ will then be used to send the video packets for the latter time slots $t = 5$ and $t = 6$. Likewise, an extra rate is allocated by PDCA for $t = 9$, while the only difference is that this extra will be used to send the video packets for the previous time slot $t = 8$.

In Fig. 6, we vary the value of node transmission rate and present the relationship between the average throughput over time and the node transmission capacity. Generally, for each algorithm, the average throughput of each destination node would increase with the increment of the node transmission capacity. Furthermore, within all the five algorithms, PDCA could achieve the highest average throughput for all destinations. However, when the node transmission capacity is relatively low (e.g., 5 Mbps in Fig. 6), due to the rate bound constrained by each source-destination pair's end-to-end throughput, there is no extra rate region for PDCA to balance reception rate over different time slots even at time slots with good network condition. At this time, both PDCA and JAOR would achieve similar performance. When the node transmission capacity gradually increases to

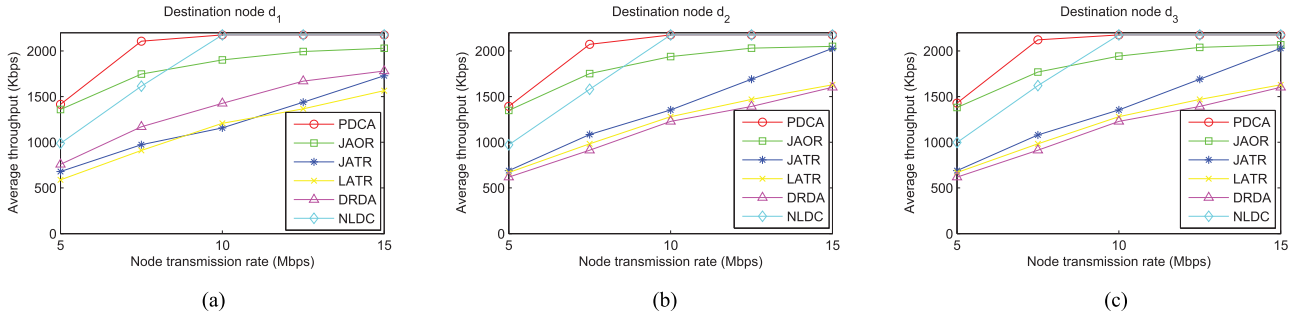


Fig. 6. Comparison of average throughput vs. node transmission rate achieved by different algorithms for destination nodes (a) d_1, (b) d_2, and (c) d_3.

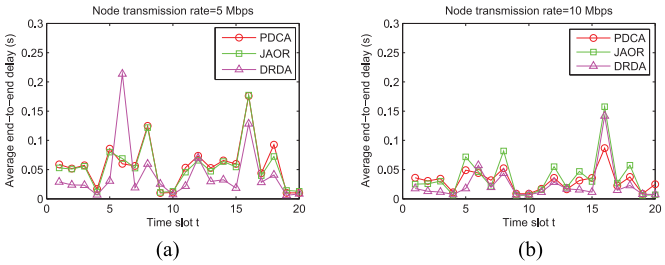


Fig. 7. Comparison of average end-to-end delay achieved by different algorithms.

allow such extra rate region, the average throughput achieved by PDCA would be higher than JAOR. It should also be noted that with the increment of the node transmission capacity, the average throughput of each destination achieved by PDCA would eventually reach its upper bound that is constrained by the sum rate of all the four SVC layers. To study the impact of the SVC layer dependency constraint in (2), we further show in Fig. 6 the result achieved by PDCA but with no layer dependency constraint (NLDC). It can be seen that the average throughput achieved by NLDC is lower than PDCA when the node transmission rate is relatively low (e.g., 5 Mbps and 7.5 Mbps). The reason is that, since all the four layers are treated equally by NLDC without any priority, the reception rate allocated for higher layers might not be decodable due to the lack of the lower layers, which is useless and thus cannot contribute to the average throughput of the destination nodes.

As stated before, all these algorithms aim at a maximum aggregate network utility in terms of video quality, which is proportional to the throughput. They allocate the appropriate video reception rate $\boldsymbol{R}$ for each destination to achieve the maximum throughput, while respecting the network transmission capacity (i.e., avoiding congestion at the relay nodes). Therefore, the end-to-end delay performance of different algorithms are expected to be similar. To justify this, we illustrate in Fig. 7 the average end-to-end delay of packets within each time slot, when the node transmission rate is 5 Mbps and 10 Mbps, respectively. We use the UDP protocol for transmission with 1500-byte payload in each packet and set the buffer size of each relay node to 300 packets. In Fig. 7(a), the average end-to-end delay over all the $T = 20$ time slots for PDCA, JAOR and DRDA is 43.0, 41.5 and 29.9 ms, respectively. Both PDCA and JAOR have similar average end-to-end delay, while DRDA achieves a lower average end-to-end delay because a smaller number of packets (corresponding to a lower throughput) is allocated for transmission under the same network condition. When the node transmission rate increases to 10 Mbps in Fig. 7(b), the network transmission capacity is increased, which causes the average end-to-end delay over all the $T = 20$ time slots for PDCA, JAOR and DRDA to reduce to 24.3, 29.9 and 16.0 ms, respectively. Since we adopt in this paper the network coding based OR, the source node will keep sending coded packets of a flow in accordance with the optimal allocated rate $\boldsymbol{R}$ to each destination until the original video packets of that flow can be fully recovered. In this sense, the packet delivery ratio is 1. In addition, we assume that the buffer of each destination node is large enough to store all the packets received during the $T = 20$ time slots. After $T \cdot \Delta T = 2$ s of the initial startup delay, users then start watching the video. In this way, no buffer stall will occur during the video playback.

2) Finite State Markov Channel Model: To characterize and represent the time-varying behavior of the wireless Rayleigh fading channels, the FSMC model in [37] is adopted, with SNR thresholds and symbol error probabilities for different channel states listed in [37, Table I]. From (15) and (16), it can be observed that the accuracy of the expected effective capacity expression depends on the accuracy of the expected link's PRR $\hat{p}_{(i,j)}(t|\tau)$, which is determined by the accuracy of two parameters, the transition probability matrix $\mathbf{Tr}_{(i,j)}$ and the packet reception probability vector $\mathbf{P}_{(i,j)}$ as shown in (20). The FSMC model accuracy of these two parameters are justified in [37], which therefore validates the accuracy of the expected effective capacity expression in (16). For a completeness of this paper, we include this Table here with slight adaptation to our scenario. In Table I, γ_0 represents the average SNR, Ω^h denotes the received SNR threshold, P_s^h is the symbol error rate, and the channel is in state h if the received SNR locates in the range $[\Omega^h, \Omega^{h+1})$. In addition, the values of the other critical model parameters used for simulation are listed in Table II.

Fig. 8 shows the average throughput vs. node transmission rate curves achieved by PDCA under FSMC model and the average received SNR $\gamma_0 = 10$ dB, with different setups of the initial channel state. According to Table I, the channel state index of the FSMC is ordered with decreasing symbol error probabilities. A smaller channel state index therefore indicates a worse channel condition with lower PRR for each wireless link.

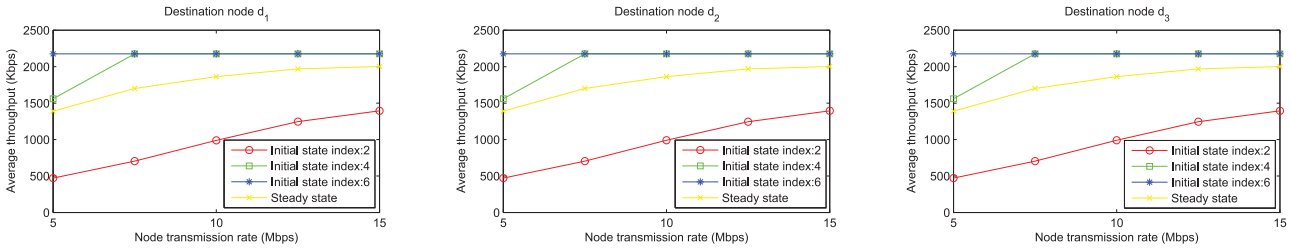


Fig. 8. Average throughput vs. node transmission rate achieved by the proposed algorithm under FSMC model and average SNR = 10 dB for destination nodes (a) d_1, (b) d_2, and (c) d_3.

TABLE I
FSMC MODEL PARAMETERS IN [37]

h	Ω^h/γ_0 (dB)	P_s^h (γ_0 = 5 dB)	P_s^h (γ_0 = 10 dB)
1	$-\infty$	$7.529 \cdot 10^{-1}$	$5.570 \cdot 10^{-1}$
2	–12.0474	$4.153 \cdot 10^{-1}$	$1.313 \cdot 10^{-1}$
3	–6.0158	$1.555 \cdot 10^{-1}$	$9.900 \cdot 10^{-3}$
4	–2.4754	$4.008 \cdot 10^{-2}$	$2.429 \cdot 10^{-4}$
5	0.0499	$7.079 \cdot 10^{-3}$	$1.813 \cdot 10^{-6}$
6	2.0232	$8.840 \cdot 10^{-4}$	$3.782 \cdot 10^{-9}$
7	3.6514	$6.614 \cdot 10^{-5}$	$2.003 \cdot 10^{-12}$
8	5.0454	$3.285 \cdot 10^{-6}$	–
9	6.2726	$9.867 \cdot 10^{-8}$	–
10	7.3777	$1.681 \cdot 10^{-9}$	–
11	8.3934	$1.214 \cdot 10^{-11}$	–
–	∞	–	–

TABLE II
CONFIGURATION OF MODEL PARAMETERS IN THE NETWORK

Parameter	Description	Value
D_t	Effective transmission range	50 m
v	Node movement speed	5 km/h
ψ_m	Maximum Doppler frequency	8.7963 Hz
ΔT	Time duration of a time slot	100 ms
T_p	Packet time duration	0.384 ms
σ	Path-loss exponent	2–4
γ_0	Average received SNR	5 dB/10 dB

It can be seen from Fig. 8 that if the initial channel state index is 6, the average throughput of each destination node would not change with the node transmission capacity, since at this time the channel condition of the wireless network is good enough to ensure that the maximum achievable throughput bound is already reached with the node transmission rate being 5 Mbps. When the initial channel state index decreases from 6 to 2, the respective initial channel condition becomes worse, and thus the average throughput at the same node transmission capacity will accordingly decrease. Besides the three curves studying the effect of the transient behavior of the wireless fading channel, the steady state performance is also shown in Fig. 8 with the corresponding curve located in the middle of these three curves. To obtain the steady state distribution of the fading channel, we first solve the steady state equation of the FSMC. Then, for each initial state index, we run the simulation and get the resulting average throughput vs. node transmission rate performance. Finally, the steady state performance is obtained by weighted averaging of the performance measure among different initial states using the steady state distribution of the FSMC.

B. Simulations for Larger Scale Network

To evaluate the effect for a more general network, we further generate a wireless network with 45 nodes randomly distributed in a 100 m × 100 m square region. Taking the node mobility into account, each node is assumed to be moving at a walking speed of 5 kilometers per hour. Here, we consider the similar mobile user's movement setup to that in [46], because it corresponds to a slow-fading channel where the first-order FSMC model is accurate and the time duration of a time slot[1] (ΔT = 100 ms in the simulation) is larger than the packet time duration (T_p = 0.384 ms in the simulation) and the video frame time interval (33 ms in the simulation). In our future work, we will study the high-speed user movement scenario (such as mobile users travelling in the vehicles or trains), which results in a fast-fading channel and thus needs to be modeled by a more complicated high-order FSMC. We fix the node nearest to the lower left corner as the source node, and select five nodes as destination nodes. The effective transmission range D_t is assumed to be 50 m, and the average number of hops between each source-destination pair is 6. In Fig. 9, we set the node transmission rate for each transmitter node to 12.5 Mbps and investigate the comparison of average video reception qualities in PSNR for *Bus*, *Football*, and *Foreman* video sequences achieved by different algorithms, while the dynamic PRR samples of wireless links within each time slot is the same as in Fig. 2(a). The video packets are transmitted according to the video rate allocation determined by each algorithm, and the average PSNR of that algorithm for the transmission of a specific video is measured over all received video packets within all the $T = 20$ time slots. Similarly, it can be observed that PDCA outperforms

[1]The time duration of a time slot can be viewed as the coherence time of the channel, which is the time duration over which the channel's response (or state) remains invariant and given in (23).

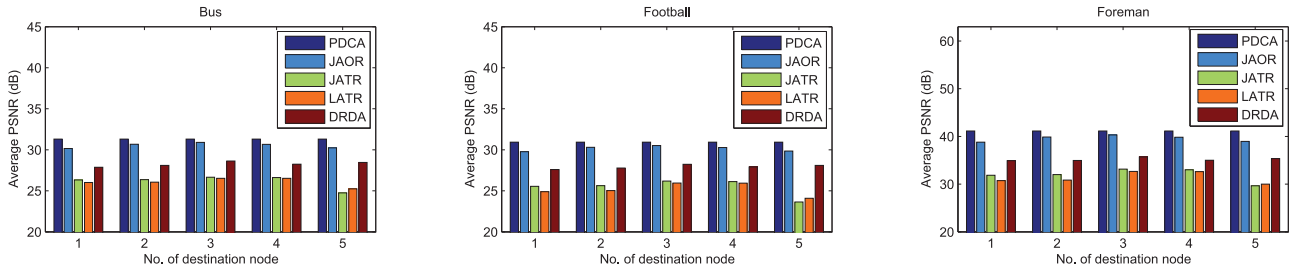


Fig. 9. Comparison of received average video quality in PSNR for (a) *Bus*, (b) *Football*, and (c) *Foreman* sequences, with the random PRR sample in Fig. 2(a).

TABLE III
COMPARISON OF RECEIVED AVERAGE VIDEO QUALITY IN PSNR UNDER FSMC MODEL WITH DIFFERENT AVERAGE SNR AND NODE TRANSMISSION RATE

Sequence		Bus				Football				Foreman			
Ave. SNR (dB)		5		10		5		10		5		10	
Trans. Rate (Mbps)		10	15	10	15	10	15	10	15	10	15	10	15
d_1	*PDCA*	***26.67***	***27.22***	***29.68***	***30.47***	***26.22***	***27.03***	***29.28***	***30.09***	***33.22***	***34.76***	***37.83***	***39.45***
	JAOR	26.34	26.69	28.95	29.90	25.59	26.27	28.53	29.51	31.93	33.31	36.33	38.28
	JATR	23.30	24.65	26.03	26.24	22.88	23.56	24.91	25.40	29.39	29.65	30.74	31.56
	LATR	22.59	23.38	25.23	26.21	22.55	22.92	24.08	25.35	29.25	29.40	29.99	31.44
	DRDA	24.86	26.20	26.91	28.18	23.73	25.33	26.68	27.85	29.69	31.41	34.14	34.99
d_2	*PDCA*	***27.04***	***27.80***	***30.14***	***30.72***	***26.87***	***27.52***	***29.75***	***30.33***	***34.64***	***34.90***	***38.77***	***39.94***
	JAOR	26.64	26.98	29.60	30.39	26.17	26.79	29.20	30.01	33.11	34.42	37.67	39.28
	JATR	23.53	24.95	26.08	26.27	22.98	23.82	25.09	25.46	29.43	29.75	30.94	31.67
	LATR	22.61	23.40	25.26	26.22	22.56	22.93	24.11	25.36	29.25	29.41	30.02	31.47
	DRDA	25.40	26.24	26.99	28.27	24.24	25.39	26.80	27.93	30.14	31.54	34.44	35.01
d_3	*PDCA*	***28.11***	***28.72***	***30.89***	***31.16***	***27.79***	***28.31***	***30.51***	***30.79***	***34.97***	***35.87***	***40.29***	***40.85***
	JAOR	27.19	27.74	30.51	30.85	27.00	27.47	30.12	30.47	34.75	34.88	39.51	40.21
	JATR	24.57	25.92	26.37	26.72	23.50	24.73	25.66	26.32	29.63	30.58	32.07	33.41
	LATR	23.74	24.69	26.19	26.59	23.08	23.59	25.30	26.07	29.47	29.65	31.36	32.91
	DRDA	26.07	26.41	27.59	28.70	25.07	25.72	27.35	28.29	30.90	32.20	34.85	35.83
d_4	*PDCA*	***27.57***	***28.24***	***30.59***	***30.95***	***27.33***	***27.90***	***30.20***	***30.57***	***34.84***	***35.00***	***39.67***	***40.42***
	JAOR	26.90	27.34	30.12	30.64	26.65	27.13	29.73	30.26	34.09	34.79	38.73	39.78
	JATR	24.48	25.74	26.32	26.70	23.43	24.57	25.55	26.27	29.61	30.43	31.85	33.31
	LATR	23.71	24.67	26.18	26.58	23.07	23.57	25.29	26.06	29.47	29.65	31.33	32.88
	DRDA	26.05	26.39	27.55	28.66	25.01	25.69	27.31	28.26	30.84	32.13	34.84	35.74
d_5	*PDCA*	***26.98***	***27.65***	***30.10***	***30.71***	***26.79***	***27.40***	***29.71***	***30.33***	***34.40***	***34.86***	***38.68***	***39.93***
	JAOR	26.63	26.95	29.60	30.29	26.14	26.75	29.20	29.90	33.05	34.29	37.66	39.08
	JATR	22.75	23.56	24.25	26.01	22.63	23.00	23.32	24.83	29.28	29.44	29.57	30.68
	LATR	22.58	23.27	24.17	25.90	22.55	22.87	23.28	24.72	29.25	29.38	29.55	30.57
	DRDA	25.75	26.31	27.28	28.43	24.58	25.53	27.08	28.07	30.44	31.80	34.77	35.28

the other four algorithms with higher overall video reception quality in peak signal-to-noise ratio (PSNR) for all destinations. Also, by exploiting the broadcast nature of the wireless shared medium, PDCA could allocate similar video reception quality for different destinations to promise the fairness among them.

In order to further study the behavior of the five different algorithms under the FSMC model, Table III shows the performance comparison on the average received video quality in PSNR versus the average received SNR of the wireless Rayleigh fading channel and the node transmission rate, for *Bus*, *Football*, and *Foreman* video sequences. Here, the average PSNR measures the steady state performance of the FSMC, which is computed as follows. For each initial state of the FSMC, we transmit the video packets and obtain the average PSNR of that initial state over all received video packets within all the time slots. Then, the average PSNR of the steady state performance is obtained by the weighted averaging of the average PSNR of each initial state by using the steady state distribution of the FSMC. It can be seen from Table III that for the same average received SNR, all the five algorithms will achieve higher average video reception PSNR as the node transmission capacity increases, which allows more transmission bit rate to be supported by the wireless channel of each link. On the other hand, when the node transmission rate is fixed, the average video reception PSNR for all the five algorithms would become higher with the increment of the average received SNR, since a larger average SNR corresponds to a better channel condition and thus a higher PRR for each wireless link. Generally, for given average received SNR and node transmission rate, the average video reception PSNR achieved by the proposed algorithm is higher than those of the other four schemes.

VI. Conclusion

Considering the time-varying characteristics of practical wireless networks, we proposed a joint optimization scheme for scalable video multirate multicast based on OR and network coding. The proposed scheme aimed at maximizing the overall video reception quality among all destinations over the multicast period of a scalable video stream. Different from traditional work on SVC streaming, the decision of optimal routes for SVC layered streaming has been integrated into the joint optimization formulation with OR. With the dynamic wireless link states either known or predicted, the proposed scheme jointly optimized the reception rate, routes to each destination, and time fraction of maximum CTSs. We developed a distributed dynamic cross-layer algorithm by using dual decomposition and primal-dual update approach. Simulation results have validated significant network multicast throughput improvement and adaptation to dynamic network changes relative to existing optimization schemes.

References

[1] "Global mobile data traffic forecast update, 2014–2019," Cisco Visual Networking Index (VNI) Forecast, 2015. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H. 264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[3] K. Kar, S. Sarkar, and L. Tassiulas, "Optimization based rate control for multirate multicast sessions," in *Proc. IEEE INFOCOM*, Apr. 2001, pp. 123–132.

[4] D. S. Lun *et al.*, "Minimum-cost multicast over coded packet networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2608–2623, Jun. 2006.

[5] J. Yan, K. Katrinis, M. May, and B. Plattner, "Media- and TCP-friendly congestion control for scalable video streams," *IEEE Trans. Multimedia*, vol. 8, no. 2, pp. 196–206, Apr. 2006.

[6] L. Chen, T. Ho, S. H. Low, M. Chiang, and J. C. Doyle, "Optimization based rate control for multicast with network coding," in *Proc. IEEE INFOCOM*, 2007, pp. 1163–1171.

[7] J. Zhao, F. Yang, Q. Zhang, Z. Zhang, and F. Zhang, "LION: Layered overlay multicast with network coding," *IEEE Trans. Multimedia*, vol. 8, no. 5, pp. 1021–1032, Oct. 2006.

[8] S. Hua, Y. Guo, Y. Liu, H. Liu, and S. S. Panwar, "Scalable video multicast in hybrid 3G/ad-hoc networks," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 402–413, Apr. 2011.

[9] C. Li, H. Xiong, J. Zou, and C. W. Chen, "Distributed robust optimization for scalable video multirate multicast over wireless networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 6, pp. 943–957, Jun. 2012.

[10] D. Vukobratović, V. Stanković, D. Sejdinović, L. Stanković, and Z. Xiong, "Scalable video multicast using expanding window fountain codes," *IEEE Trans. Multimedia*, vol. 11, no. 6, pp. 1094–1104, Oct. 2009.

[11] N. Thomos, J. Chakareski, and P. Frossard, "Prioritized distributed video delivery with randomized network coding," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 776–787, Aug. 2011.

[12] J.-P. Sheu, C.-C. Kao, S.-R. Yang, and L.-F. Chang, "A resource allocation scheme for scalable video multicast in WiMAX relay networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 1, pp. 90–104, Jan. 2013.

[13] M. Zorzi and R. R. Rao, "Geographic random forwarding (GeRaF) for ad hoc and sensor networks: Multihop performance," *IEEE Trans. Mobile Comput.*, vol. 2, no. 4, pp. 337–348, Oct.–Dec. 2003.

[14] S. Biswas and R. Morris, "ExOR: Opportunistic multi-hop routing for wireless networks," in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun.*, vol. 35, no. 4, 2005, pp. 133–144.

[15] K. Zeng, W. Lou, and H. Zhai, "On end-to-end throughput of opportunistic routing in multirate and multihop wireless networks," in *Proc. IEEE INFOCOM*, Apr. 2008, pp. 816–824.

[16] K. Zeng, W. Lou, and H. Zhai, "Capacity of opportunistic routing in multirate and multi-hop wireless networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5118–5128, Dec. 2008.

[17] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, "Trading structure for randomness in wireless opportunistic routing," in *Proc. Conf. Appl., Technol., Archit, Protocols Comput. Commun.*, 2007, pp. 169–180.

[18] Y. Lin, B. Li, and B. Liang, "CodeOR: Opportunistic routing in wireless mesh networks with segmented network coding," in *Proc. IEEE Int. Conf. Netw. Protocols*, 2008, pp. 13–22.

[19] Y. Yan, B. Zhang, J. Zheng, and J. Ma, "Core: A coding-aware opportunistic routing mechanism for wireless mesh networks [accepted from open call]," *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 96–103, Jun. 2010.

[20] D. Koutsonikolas, C.-C. Wang, and Y. C. Hu, "Efficient network-coding-based opportunistic routing through cumulative coded acknowledgments," *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1368–1381, Oct. 2011.

[21] B. Radunović, C. Gkantsidis, P. Key, and P. Rodriguez, "Toward practical opportunistic routing with intra-session network coding for mesh networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 2, pp. 420–433, Apr. 2010.

[22] A. Khreishah, I. M. Khalil, and J. Wu, "Distributed network coding-based opportunistic routing for multicast," in *Proc. 13th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2012, pp. 115–124.

[23] I. F. Akyildiz and X. Wang, "A survey on wireless mesh networks," *IEEE Commun. Mag.*, vol. 43, no. 9, pp. S23–S30, Sep. 2005.

[24] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 89–103, Jan. 2005.

[25] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1012–1032, Jun. 2000.

[26] A. Miu, H. Balakrishnan, and C. E. Koksal, "Improving loss resilience with multi-radio diversity in wireless networks," in *Proc. 11th Annu. Int. Conf. Mobile Comput. Netw.*, 2005, pp. 16–30.

[27] C. Reis, R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan, "Measurement-based models of delivery and interference in static wireless networks," in *Proc. Conf. Appl., Technol., Archit, Protocols Comput. Commun.*, 2006, pp. 51–62.

[28] R. W. Yeung, "Multilevel diversity coding with distortion," *IEEE Trans. Inf. Theory*, vol. 41, no. 2, pp. 412–422, Mar. 1995.

[29] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[30] K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," in *Proc. 9th Annu. Int. Conf. Mobile Comput. Netw*, 2003, pp. 66–80.

[31] Y. Cheng, H. Li, P.-J. Wan, and X. Wang, "Wireless mesh network capacity achievable over the CSMA/CA MAC," *IEEE Trans. Veh. Technol.*, vol. 61, no. 7, pp. 3151–3165, Sep. 2012.

[32] Y. Shi, Y. T. Hou, J. Liu, and S. Kompella, "How to correctly use the protocol interference model for multi-hop wireless networks," in *Proc. 10th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2009, pp. 239–248.

[33] L. Xiong, L. Libman, and G. Mao, "Uncoordinated cooperative communications in highly dynamic wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 2, pp. 280–288, Feb. 2012.

[34] N. Trichakis, A. Zymnis, and S. Boyd, "Dynamic network utility maximization with delivery contracts," in *Proc. IFAC World Congr.*, 2008, pp. 2907–2912.

[35] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[36] H. S. Wang and N. Moayeri, "Finite-state Markov channel-a useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, Feb. 1995.

[37] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.

[38] J. Chen, V. K. Lau, and Y. Cheng, "Distributive network utility maximization over time-varying fading channels," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2395–2404, May 2011.

[39] B. Sklar, "Rayleigh fading channels in mobile digital communication systems Part I: Characterization," *IEEE Commun. Mag.*, vol. 35, no. 9, pp. 136–146, Sep. 1997.

[40] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[41] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[42] S. Shakkottai, S. G. Shakkottai, and R. Srikant, *Network Optimization and Control*. Breda, the Netherlands: Now Publishers, Inc., 2008.

[43] K. Kar, S. Sarkar, and L. Tassiulas, "A scalable low-overhead rate control algorithm for multirate multicast sessions," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 8, pp. 1541–1557, Oct. 2002.
[44] *JSVM 1.0 Software*. ISO/IEC, Hong Kong, Document JCT1/SC29/WG11/N6900, Jan. 2005.
[45] V. Joseph and G. de Veciana, "NOVA: QoE-driven optimization of DASH-based video delivery in networks," in *Proc. IEEE INFOCOM*, 2014, pp. 82–90.
[46] C. Chen, R. W. Heath, A. C. Bovik, and G. de Veciana, "A Markov decision model for adaptive scheduling of stored scalable videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 1081–1095, Jun. 2013.
[47] N. Chakchouk, "A survey on opportunistic routing in wireless communication networks," *IEEE Commun. Surveys Tut.*, vol. 17, no. 4, pp. 2214–2241, Oct.–Dec. 2015.

Chenglin Li (S'12–M'15) received the B.S., M.S., and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2007, 2009 and 2015, respectively. From 2011 to 2013, he was with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA, as a visiting Ph.D. student. Since 2015, he has been a Postdoctoral Research Fellow in the Signal Processing Laboratory (LTS4) at École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. His main research interests include network oriented image/video processing and communication, network-based optimization for video sources, and adaptive multimedia communication systems. Dr. Li received the Microsoft Research Asia (MSRA) Fellow in 2011, and the IEEE VCIP Best 10% Paper award in 2016.

Hongkai Xiong (M'01–SM'10) received the Ph.D. degree from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2003. Since 2003, he has been with the Department of Electronic Engineering, SJTU. He is currently a Distinguished Professor in the School of Electronic Information and Electrical Engineering, SJTU. During 2007–2008, he was a research scholar in the Department of Electrical and Computer Engineering, Carnegie Mellon University. From 2011 to 2012, he was a Scientist in the Department of Biomedical Informatics, University of California, San Diego. He has published more than 170 refereed journal and conference papers. His research interests include signal processing, image and video coding, multimedia communication, along with computer vision and learning. Prof. Xiong was granted the National Science Fund for a Distinguished Young Scholar, in 2014. In 2016, he was granted the Yangtze River Scholar Distinguished Professor award from the Ministry of Education of China, and the Youth Science and Technology Innovation Leader award from the Ministry of Science and Technology of China. He received Shanghai Academic Research Leader and Shanghai Youth Science and Technology Talent awards as well. He received the Shanghai Shu Guang Scholar award in 2013; the First Prize of the Shanghai Technological Innovation Award for network-oriented video processing and dissemination: Theory and technology in 2011; the SMC-A Excellent Young Faculty Awards of Shanghai Jiao Tong University in 2010 and 2013; the New Century Excellent Talents in University award from the Ministry of Education of China in 2009; the Best Student Paper Award at the 2014 IEEE Visual Communication and Image Processing (IEEE VCIP'14); the Best Paper Award at the 2013 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (IEEE BMSB'13); and the Top 10% Paper Award at the 2011 IEEE International Workshop on Multimedia Signal Processing (IEEE MMSP'11). His research projects are funded by NSF of China, QUALCOMM, MICROSOFT, and INTEL. Since 2012, he has been a member of Innovative Research Groups of the National Natural Science. He was the Area Chair in the technical program committee of ICIP and ISCAS, and a TPC member of prestigious conferences such as ACM Multimedia, ICME, and ICASSP.

Junni Zou (M'07) received the M.S. and Ph.D. degrees in communication and information system from Shanghai University, Shanghai, China, in 2004 and 2006, respectively. Since then, she has been with the School of Communication and Information Engineering, Shanghai University, where she is a Full Professor. From 2011 to 2012, she was with the Department of Electrical and Computer Engineering, University of California, San Diego, as a Visiting Professor. Since 2017, she has been a Full Professor in the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. Her research interests include distributed resource allocation, multimedia networking and communications, and network information theory. She has authored more than 70 refereed journal/conference papers on these topics. Prof. Zou received the National Science Fund for Outstanding Young Scholar award in 2016 and Shanghai Young Rising-Star Scientist award in 2011. She acts as a Member of the Technical Committee on Signal Processing of Shanghai Institute of Electronics.

Dapeng Oliver Wu (S'98–M'04–SM'06–F'13) received the B.E. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 1990, the M.E. degree in electrical engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 1997, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2003. He is currently a Professor in the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA. His research interests include networking, communications, signal processing, computer vision, machine learning, smart grid, and information and network security. Dr. Wu received the University of Florida Term Professorship Award in 2017, the University of Florida Research Foundation Professorship award in 2009, the AFOSR Young Investigator Program (YIP) award in 2009, the ONR Young Investigator Program (YIP) award in 2008, the NSF CAREER award in 2007, the IEEE Circuits and Systems for Video Technology (CSVT) Transactions Best Paper award for Year 2001, and the Best Paper Awards in the IEEE GLOBECOM 2011 and the International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine) 2006. He is currently an Editor-in-Chief of the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, and an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS, and the IEEE SIGNAL PROCESSING MAGAZINE. He was the founding Editor-in-Chief of *Journal of Advances in Multimedia* between 2006 and 2008, and an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is also a Guest-Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, Special Issue on Cross-layer Optimized Wireless Multimedia Communications. He was the Technical Program Committee (TPC) Chair for the IEEE INFOCOM 2012, and the TPC Chair of the IEEE International Conference on Communications (ICC 2008), Signal Processing for Communications Symposium, and as a member of executive committee and/or technical program committee of more than 80 conferences. He was elected as a Distinguished Lecturer by the IEEE Vehicular Technology Society in 2016. He was the Chair for the Award Committee, and the Chair of Mobile and Wireless Multimedia Interest Group (MobIG), the Technical Committee on Multimedia Communications, and the IEEE Communications Society. He was an elected member of the Multimedia Signal Processing Technical Committee, the IEEE Signal Processing Society, from Jan. 1, 2009 to Dec. 31, 2012.