Evolutionary Network Formation for Dynamic Wireless Networks with Network Coding

Minhae Kwon, *Student Member, IEEE,* Hyunggon Park, *Senior Member, IEEE,*
and Mihaela van der Schaar, *Fellow, IEEE*

***Abstract*—In this paper, we aim to find a robust network formation strategy that can adaptively evolve the network topology against network dynamics in a distributed manner. We consider a network coding deployed wireless ad hoc network where source nodes are connected to terminal nodes through links constructed by intermediate nodes. We show that mixing operations in network coding can induce packet anonymity that allows the inter-connections in a network to be decoupled. Hence, the intermediate nodes can take action that is individually optimized for the transmission range based on the Markov decision process (MDP). This occurs in an intermediate node that can take optimal actions given its state, which are defined as the number of effective nodes. Then, the transmission ranges that correspond to the actions are determined. The optimal actions are followed by the optimal policy, which is obtained by solving the MDP. The proposed strategy is used to maximize long-term utility, which is achieved by considering both current network conditions and future network dynamics. We define the utility of an action to include network throughput gain and the cost of transmission power. We show that the resulting network of the proposed strategy eventually converges to stationary networks, which maintain the states of the nodes. Moreover, we propose how to initialize the network such that the convergence of the proposed algorithm can be expedited. Our simulation results confirm that the proposed strategy builds the network which adaptively changes the network topology in the presence of network dynamics. Therefore, the proposed strategy outperforms existing strategies in terms of system goodput and successful connectivity ratio.**

***Index Terms*—Network Formation, Network Topology Design, Markov Decision Process, Network Coding, Wireless Ad Hoc Networks, Mobile Network, Dynamic Network**

I. Introduction

The connected world which began with representative services such as connected cars, networked unmanned aerial vehicles (UAVs) and the Internet of things (IoT), results in network with inherent dynamics. The network entities of such services generally have high mobility, which causes frequent changes in member nodes associated with these networks and unstable channel conditions with high link failure rates. Hence, it is essential to form robust networks against such dynamics by adaptively reformulating inter-connections among network entities. This problem cannot be solved by conventional centralized solutions. Rather, it can be solved by decentralized and spontaneous network formation strategies that can proactively respond to the network dynamics by modifying the network topology based on the decision-making processes of each network entity.

However, it is not straightforward to design decentralized strategies that enable each network entity to make its *own* and *optimal* decisions, because the network entities are intimately coupled. Specifically, the network entities can be tightly inter-connected, so that the impact of small changes from a network entity may propagate over a large number of entities. Therefore, each network entity should consider the corresponding responses associated with its decisions to make optimal decisions. This may require significantly high computational complexity or may not be feasible in practice. Therefore, it is essential for the design of decentralized strategies to decouple the inter-connections among network entities.

In this paper, we show that the inter-connections among network entities can be decoupled by deploying network coding, which is referred to as *network decoupling*. Unlike the conventional store-and-forward approach, network coding [1] allows an intermediate entity to combine multiple packets that it has received and to forward the *mixed* packets. For a network coding enabled wireless ad-hoc network (which is widely considered as a network model of a connected world) a packet passes through many intermediate entities. Thus, it can be mixed with other packets multiple times. This leads to *packet anonymity*, where all packets in the network eventually include identical information as well as a terminal set. Packet anonymity allows an entity to consider the other entities as its *environment* so that complicated inter-connections among network entities can be decoupled, and only the connection directly associated with the entity is considered as a one-hop connection. This leads to network decoupling so that the interactions between network entities can be interpreted as a node-environment interaction at each entity.

Motivated by the node-environment interaction, we use the MDP to find a decentralized strategy, which is referred to as a policy, for network formation. We consider wireless entities to be autonomous decision-making agents and the state of an agent is defined as the number of effective nodes. Here, effective nodes are the entities that have *successfully* received packets from the agent. The probability density function of the states is modeled by the Poisson point process (PPP), which is widely used to characterize the behavior of mobile nodes. An agent can take an action by considering its current state, which determines the transmission range. The actions can be selected by the policy of the agent. The policy is optimal if it

M. Kwon and H. Park are with the Department of Electronic and Electrical Engineering, Ewha Womans University, Seoul, Republic of Korea (e-mail: minhae.kwon@ewhain.net, hyunggon.park@ewha.ac.kr).

M. van der Schaar is with the Department of Electrical Engineering, University of California, Los Angeles (UCLA), CA, 90095, USA (e-mail: mihaela@ee.ucla.edu) and the Department of Engineering Science, University of Oxford, UK (e-mail: mihaela.vanderschaar@oxford-man.ox.ac.uk).

enables the agent to maximize long-term utility.

As a node increases its transmission range, the number of hops required to reach the terminal node decreases, without loss of generality, leading to an improvement in network throughput. However, extending the transmission range increases power consumption and causes more inter-node interference. This is explicitly captured by the utility function, which is represented by both network throughput improvement and the additional corresponding transmission power. Therefore, the optimal policy enables each entity to successively determine the optimal changes in transmission range at each state, such that the entities can strike a balance between network throughput gain and power consumption. Finally, the consequences of the distributed decisions from each entity eventually determine the network topology.

Note that the proposed strategy allows the resulting network topology to evolutionarily adapt against network dynamics. This is because the state is defined by the effective nodes, which are directly dependent on link failure rates (i.e., channel condition) and node mobility (i.e., node distribution). For example, a larger transmission range may be required in a channel with higher link failure rates for the target number of effective nodes. Similarly, an agent can increase its transmission range to sustain connectivity in the case of sparse node density. The proposed strategy is also robust against frequent changes in member nodes of the considered network, which is widely observed in mobile networks. This is because the behavior of existing nodes is not affected by individual network members, instead it is only affected by *the number* of effective nodes included in its own transmission range.

Unlike conventional optimal solutions that focus on maximizing immediate utility, the proposed optimal policy determined by the MDP can provide a long-term strategy, which determines actions by explicitly considering future dynamics in the network. Specifically, the actions taken by the optimal policy can maximize the long-term utilities, which are expressed as the sum of discounted utilities over time. The discount factor can be determined by considering the consistency of network conditions. Therefore, the actions determined by the proposed policy can consider both current and future network dynamics.

The proposed system consists of two phases: initialization and adaptation. In the initialization phase, the optimal policy for each intermediate node is found and the state can be initialized. As will be shown in this paper, optimal actions can lead the network formation result to certain topologies, referred to as *stationary networks*. Hence, we design the initial network to be close to the stationary network. In the adaptation phase, each node adaptively and optimally changes its transmission range based on the optimal policy for the current state induced by network dynamics.

The main contributions of this paper are summarized as follows.

- We show that network coding can lead to packet anonymity where both the information and terminal of all packets in network asymptotically become identical,
- We show that the packet anonymity of network coding allows inter-connections among network nodes to be decoupled into node-environment interactions at each node, which is referred to as network decoupling,
- We formulate the problem of network topology formation in a MDP framework and provide a decentralized solution to the network formation strategy,
- The proposed strategy improves network robustness by adaptively rebuilding its topology in the presence of network dynamics which includes unstable channel conditions with high link failure rates, and high mobility of network nodes that causes frequent changes in member nodes associated with the considered network,
- The proposed strategy is a foresighted strategy that chooses the action maximizes a long-term utility by considering future network dynamics,
- The proposed strategy can determine the optimal transmission range that balances network throughput improvement and transmission power consumption,
- The resulting network of the proposed strategy converges to stationary networks, and
- We propose how to initialize a network such that the speed of convergence to the stationary network can be improved.

Note that the focus of this paper is neither on the code design for network coding which has been extensively studied in prior works [2]–[9], nor on perfect delivery which needs 100% reliability. Rather, our focus is on robust network formation based on network coding, which can proactively reform network topology against network dynamics in a decentralized manner.

The rest of the paper is organized as follows. In Section II, we briefly review related works. The wireless model for mobile users and detailed process of data delivery based on network coding are discussed in Section III. The MDP-based framework and a distributed network formation strategy are proposed in Section IV and Section V, respectively. Simulation results are presented in Section VI, and conclusions are drawn in Section VII.

For the reader's convenience, we summarize notations frequently used in this paper in Table I.

II. Related Works

Since network coding was first introduced in [1], it has shown excellent ability to improve throughput, robustness and complexity. The beginning of network coding was for throughput gain in a multicast scenario. In [1], it is shown that network coding can achieve maximum throughput via the max-flow min-cut theorem, and it is further proved that linear network coding can achieve the upper bound of capacity in [10]. Many works in network coding have been studied for random linear network coding (RLNC) [11] as it is a simple, randomized encoding approach that is decentralized [12], [13].

As well as throughput gain, it has been shown that network coding also enhances robustness against packet loss in lossy wireless networks [14]–[16]. Since packet loss is very prevalent in wireless network, typical erasure coding schemes have been widely employed by inserting a degree

TABLE I: Summary of Notations

Notation	Description	Notation	Description
v_i	a node with index i	$\mathbf{S}$	state space
$\mathbf{H}_t$	an index set of source nodes for a terminal node v_t	s	a state
$\mathbf{H}$	an index set of source nodes	$\mathbf{A}$	action space
$\mathbf{T}_h$	an index set of terminals for a source node v_h for $h \in \mathbf{H}$	a	an action
$\mathbf{T}$	total index set of all terminals in $\mathcal{G}_\tau$	$\bar{a}_\tau$	transmission range at time τ
$\mathbf{V}$	an index set of intermediate nodes	ρ	discount factor
λ	node density in network	ω	weight in (12)
$\Gamma_{i,\tau}(v_j, \mathbf{T}_h)$	node value function	$R(s, s')$	reward
$\delta_{i,\tau}(v_j)$	the Euclidean distance from node v_i to node v_j at time τ	π	policy
$\delta_{i,\tau}$	the radius of intermediate node v_i's transmission range at time τ	$V_\tau(s)$	state-value function
$\Delta_\tau(v_j, v_t)$	the number of hops between v_j and v_t at time τ	$\boldsymbol{\sigma}$	limiting distribution
$\boldsymbol{\Delta}_\tau(v_j, \mathbf{T}_h)$	a set of $\Delta_\tau(v_j, v_t)$ for all $t \in \mathbf{T}_h$	σ_s	limiting probability of state s
Φ	network coding function	$\mathbf{P}$	state transition matrix

of redundancy at the source node. Using network coding, redundancy can be inherently included at the intermediate nodes because wireless networks have broadcast links that reach multiple end nodes. In [16], a combination of RLNC and unequal error protection is proposed for eliminating feedback channel of video streaming applications. In [14], the problem of error-control in RLNC is considered, and it is shown that a minimum-distance decoder achieves correct decoding under certain conditions. The error correction capability for linear network coding is studied in [15] for various types of errors.

Another advantage of network coding is that there is a lower complexity requirement for network formation compared to a conventional store-and-forward approach. In a conventional store-and-forward approach, it is difficult to find the optimal routing path that can achieve the capacity upper bound. Even though an optimal routing solution exists in some cases, such as the Steiner tree in multicast routing, finding the solution is still very complex within a centralized setting [17]. Network coding, however, can transform complex network formation problems into low-complexity distributed problems. For example, a distributed solution that satisfies optimality condition to minimum delay and minimum energy consumption is proposed in [18]. Decentralized algorithms for network formation that can minimize cost per unit capacity are proposed in [19].

Even though network coding can reduce complexity in general, it is known that finding an optimal solution in network coding with multiple multicasts is an NP-hard problem [20]. Hence, suboptimal but practical solutions are often studied [21], [22]. A well-known practical solution to network formation is proposed based on linear optimization in [21]. In [22], a distributed network formation solution is developed for network coding deployed wireless networks that includes multi-source multicast flows. Using a game theoretical approach, each node in the network determines its transmission power and the use of network coding operations.

Network formation strategies for dynamic network conditions in conventional routing schemes have been widely studied in the context of a self-organizing network. Protocols for self-organization of wireless sensor networks where there exists a large number of static nodes with energy constraints are described in [23]. In [24], an emergency communication system based on UAV-assisted self-organizing network is considered. In this work, UAVs are used as a strong relay node to form a relay network in the air, and the nodes on the ground formed a self-organizing network automatically with the help of UAVs. In [25], a dynamic topology control that prolongs the lifetime of a wireless sensor network is proposed based on a non-cooperative game.

However, there have been few studies in network coding deployed network formation strategies that are robust in the presence of network dynamics.

III. Network Coding Deployed Wireless Ad Hoc Networks

A. Wireless Ad Hoc Network Model

We consider a wireless ad hoc network modeled by a directed graph $\mathcal{G}_\tau$, comprising a set $\mathcal{V}(\mathcal{G}_\tau)$ of nodes together with a set $\mathcal{E}(\mathcal{G}_\tau)$ of directed links at time τ. Let $v_i \in \mathcal{V}(\mathcal{G}_\tau)$ be the i-th element in $\mathcal{V}(\mathcal{G}_\tau)$, and there are three types of nodes in the network: source, intermediate and terminal. Let $\mathbf{H}$ be an index set of source nodes and its element is denoted as $h \in \mathbf{H}$. An index set of terminals for a source node v_h is denoted as $\mathbf{T}_h$, and the data that v_h generates at time τ are denoted as $x_{h,\tau}$. In this paper, we consider the multi-source multicast flows that have multiple source nodes, and each source node has an independent set of terminal nodes. Specifically, a source node v_h for $h \in \mathbf{H}$ aims to deliver its data $x_{h,\tau}$ to multiple terminal nodes $v_t, \forall t \in \mathbf{T}_h$ so that $\sum_{h \in \mathbf{H}} |\mathbf{T}_h|$ flows are simultaneously considered, where $|\cdot|$ denotes the size of a set. The total index set of all terminals in $\mathcal{G}_\tau$ is denoted as $\mathbf{T} = \bigcup_{h \in \mathbf{H}} \mathbf{T}_h$, and the number of source nodes and terminal nodes are denoted by N_H and N_T, respectively. The source nodes can only transmit data but cannot receive data, and the terminal nodes can only receive data but cannot transmit data, unlike the source nodes.

In cases where the source nodes are not able to directly transmit data to the terminal nodes, intermediate nodes can relay the data by receiving and transmitting the data. Let v_i for $i \in \mathbf{V}$ be an intermediate node where $\mathbf{V}$ denotes an index set of intermediate nodes, and N_V be the number of intermediate nodes. Then, the total number of nodes in the network can be represented by $|\mathcal{V}(\mathcal{G}_\tau)| = N_H + N_V + N_T$.

We consider the intermediate nodes as wireless mobile devices that can move around in a bounded region with energy constraints. We use a stochastic geometry model to capture the distribution of intermediate nodes to model the characteristics of mobility; the number of intermediate nodes in a bounded region follows an independent homogeneous PPP with node density λ, which is the expected number of Poisson points [26], [27]. Moreover, each intermediate node can adjust its transmission power, which determines the range of potential delivery of data from the node, which is referred to as *transmission range*. Let $\bar{\delta}_{i,\tau}$ and $\delta_{i,\tau}(v_j)$ be the radius of the transmission range of v_i and the Euclidean distance from node v_i to node v_j at time τ, respectively. Then, v_j is located in the transmission range of v_i if $\delta_{i,\tau}(v_j) \leq \bar{\delta}_{i,\tau}$ and v_j can receive the data from v_i. We assume that links between nodes may be disconnected with probability β, which is referred to as the link failure rate. Hence, the probability that v_j in the transmission range of v_i can successfully receive the data from v_i is given by $1-\beta$.

The deployment of intermediate nodes naturally leads to a multi-hop ad hoc network, and thus, the network throughput highly depends on the selection of paths constructed by the nodes. Therefore, node v_i needs to choose a node v_j, which can relay the data to terminal nodes $\mathbf{T}_j$ better than other neighbor nodes in terms of node values. The node value of v_j from the perspective of v_i is evaluated by the *node value function* $\Gamma_{i,\tau}(v_j, \mathbf{T}_h)$, expressed as

$$\Gamma_{i,\tau}(v_j, \mathbf{T}_h) = f(\delta_{i,\tau}(v_j), \mathbf{\Delta}_\tau(v_j, \mathbf{T}_h)) \tag{1}$$

where $\mathbf{\Delta}_\tau(v_j, \mathbf{T}_h) = \{\Delta_\tau(v_j, v_t) | \forall t \in \mathbf{T}_h\}$ and $\Delta_\tau(v_j, v_t)$ denotes the number of hops between v_j and v_t at time τ. The function $f(\delta_{i,\tau}(v_j), \mathbf{\Delta}_\tau(v_j, \mathbf{T}_h)) : (\mathbb{R}, \mathbb{W}^{|\mathbf{T}_h| \times 1}) \rightarrow \mathbb{R}$ is a decreasing function of $\delta_{i,\tau}(v_j)$ and $\mathbf{\Delta}_{jt,\tau}$, where $\mathbb{R}$ denotes the field of real numbers and $\mathbb{W} = \{0, \mathbb{Z}^+\}$ denotes the whole numbers which includes zero and the positive integers $\mathbb{Z}^+$. By defining the node value as a decreasing function of the distance and the number of hops from v_j to v_t, $\Gamma_{i,\tau}(v_j, \mathbf{T}_h)$ increases as v_j is located closer to v_i, and v_j is connected to $v_t, \forall t \in \mathbf{T}_h$ with a smaller number of hops. Therefore, v_i consumes lower transmission power with a smaller transmission range and reduces delay for data delivery for appropriate selection of v_j.

B. Network Coding Based Encoding Process

A source node v_h for $h \in \mathbf{H}$ generates a set of data $x_{h,\tau} = \left\{x_{h,\tau}^{(1)}, \ldots, x_{h,\tau}^{(L)}\right\}$ at time τ and it broadcasts $x_{h,\tau}$ with the transmission power $\bar{a}_h = g(\bar{\delta}_h)$, where the function $g : \mathbb{R} \rightarrow \mathbb{R}$ is determined based on a path loss model of wireless channels. We assume that the radius of transmission range $\bar{\delta}_h$ of v_h is stationary (i.e., time independent), so that the subscription τ is omitted. If an intermediate node v_i is located in the transmission range of v_h at time τ, v_i receives $x_{h,\tau}$, and puts $x_{h,\tau}$ into its buffer $\mathcal{L}_i$, i.e., $x_{h,\tau} \in \mathcal{L}_i$ wherein data are sorted by time stamp τ with the oldest time stamp at the head of the queue [28]. Note that the packet has a limited life span (e.g., time to live (TTL) in an internet packet) such that the packets with an expired time stamp can be discarded. For simplicity, we assume that the output capacity of a node is a single packet size such that a node transmits a single packet per unit time [29]–[31], and a node can receive multiple individual packets by applying multipacket reception techniques [32], [33].

The intermediate node v_i performs network coding operations by combining packets with the same time stamp in $\mathcal{L}_i$ and generates encoded data $y_{i,\tau+1}$ at time $\tau+1$ expressed as

$$y_{i,\tau+1} = \sum_{h=1}^{N_H} \bigoplus (C_{hi,\tau+1} \otimes x_{h,\tau}) \tag{2}$$

where $C_{hi,\tau+1}$ denotes the global coding coefficient of v_i for source data $x_{h,\tau}$. The network coding operations are performed in the Galois field (GF) and the operators $\oplus$ and $\otimes$ denote the addition and multiplication in GF, respectively. When the encoding process is performed in (2), the source data with the same time stamp are combined together, and a packet $p_{i,\tau+1}$ is constructed as

$$p_{i,\tau+1} = [\tau, C_{1i,\tau+1}, \ldots, C_{N_H i,\tau+1}, y_{i,\tau+1}]$$

which has the time stamp of the combined source data τ, the global coding coefficient $C_{hi,\tau+1}, \forall h \in \mathbf{H}$ as the header, and the encoded data $y_{i,\tau+1}$ as a payload. An index set of terminals for $p_{i,\tau+1}$ denoted by $\mathbf{T}_{p_{i,\tau+1}}$ can be expressed as

$$\mathbf{T}_{p_{i,\tau+1}} = \cup_{h \in \{h | C_{hi,\tau+1} \neq 0, h \in \mathbf{H}\}} \mathbf{T}_h. \tag{3}$$

This is because $p_{i,\tau+1}$ needs to be delivered to all terminals of combined source data, i.e., for all $v_t \in \mathbf{T}_h$ and $h \in \{h | C_{hi,\tau+1} \neq 0, \forall h \in \mathbf{H}\}$.

If the intermediate node v_i receives the encoded packets $y_{h,\tau+\alpha}$ at time $\tau+\alpha$, it recombines the received data and generates the encoded data $y_{i,\tau+\alpha+1}$ at time $\tau+\alpha+1$, i.e.,

$$\begin{aligned} & y_{i,\tau+\alpha+1} \\ & = \sum_{y_{j,\tau+\alpha} \in \mathcal{L}_i} \bigoplus (c_{ji} \otimes y_{j,\tau+\alpha}) \\ & = \sum_{y_{j,\tau+\alpha} \in \mathcal{L}_i} \bigoplus \left(c_{ji} \otimes \left(\sum_{h=1}^{N_H} \bigoplus (C_{hj,\tau+\alpha} \otimes x_{h,\tau}) \right) \right) \\ & = \sum_{h=1}^{N_H} \bigoplus \left(\sum_{y_{j,\tau+\alpha} \in \mathcal{L}_i} \bigoplus (c_{ji} \otimes C_{hj,\tau+\alpha}) \right) \otimes x_{h,\tau} \\ & = \sum_{h=1}^{N_H} \bigoplus (C_{hi,\tau+\alpha+1} \otimes x_{h,\tau}) \end{aligned} \tag{4}$$

where $\alpha > 0$, and c_{ji}[1] denotes the local coding coefficient for data from v_j to v_i. In this paper, network coding is implemented based on RLNC [34] so that c_{ji} is uniformly and randomly chosen from GF with a size of 2^M (GF(2^M)),

[1] The time stamp for c_{ji} is omitted because the local coding coefficient is used only for one time slot.

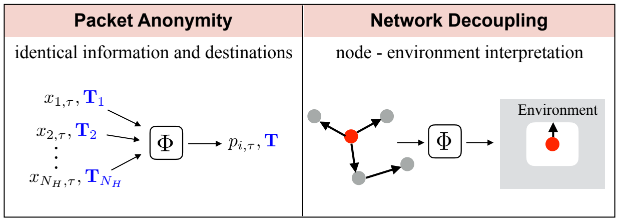

Packet Anonymity
identical information and destinations
$x_{1,\tau}, \mathbf{T}_1$
$x_{2,\tau}, \mathbf{T}_2$
$x_{N_H,\tau}, \mathbf{T}_{N_H}$
Φ
$p_{i,\tau}, \mathbf{T}$
Network Decoupling
node - environment interpretation
Φ
Environment

Fig. 1: Two characteristics of the network coding function Φ.

i.e., $c_{ij} \in$ GF(2^M). However, the proposed strategy is not limited to RLNC, and deterministic code designs [2]–[9] can be considered as well.

For a large-scale multi-hop wireless network, the process of recombining incoming packets in (4) can be performed significantly many times, which eventually allows the recombined packet to include all source data. Therefore, the terminal set of all packets in the network asymptotically converges to $\mathbf{T}$ by making all packets identical. This is defined as *packet anonymity* of network coding, which is expressed in Proposition 1.

Proposition 1 (Packet Anonymity)**.** *Network coding can asymptotically make both the information and terminal of each packet identical.*

We next consider the impact of network coding on the node value. Let Φ be the network coding function in (4). Then, the node value function $\Gamma_{i,\tau}(v_j, \mathbf{T}_h)$ in (1) that is transformed by the network coding function Φ can be expressed as

$$\begin{aligned}\Phi(\Gamma_{i,\tau}(v_j, \mathbf{T}_h)) &= \Phi\left(f(\delta_{i,\tau}(v_j), \boldsymbol{\Delta}_\tau(v_j, \mathbf{T}_h))\right) && (5)\\ &= f(\delta_{i,\tau}(v_j), \boldsymbol{\Delta}_\tau(v_j, \mathbf{T})) && (6)\\ &= \Gamma_{i,\tau}(v_j, \mathbf{T}). && (7)\end{aligned}$$

The equality between (5) and (6) is based on packet anonymity. $\mathbf{T}$ in (6) is constant, so that the node value is only a function of v_j, which concludes (7). Therefore, we conclude that if the network coding function Φ is employed in intermediate nodes, multi-hop connections to terminals (i.e., $\mathbf{T}_h$) need not be considered for node v_i. Rather, only the links directly associated with it should be considered as in a one-hop connection (i.e., v_j), leading to *network decoupling* described in Proposition 2.

Proposition 2 (Network Decoupling)**.** *Network coding can decouple one-hop connections from the overall network formation.*

Network decoupling can lead to the design of decentralized solutions by only considering node-environment interactions at each node, which are captured by the MDP framework introduced in Section IV. The characteristics of network coding function Φ are shown in Fig. 1.

C. Source Reconstruction at Terminal Nodes

We next discuss the decoding process. Let $\mathbf{H}_t = \{h | t \in \mathbf{T}_h, \forall h \in \mathbf{H}\}$ be an index set of source nodes for a terminal node v_t and $\tilde{p}_1, \cdots, \tilde{p}_K \in \mathcal{L}_t$ be the packets with the same time stamp of source data that v_i received. Then, we can construct a vector of network coded data $\tilde{\mathbf{y}} = [\tilde{y}_1, \cdots, \tilde{y}_K]^T$ and the global coding coefficient matrix $\tilde{\mathbf{C}}$ is expressed as

$$\tilde{\mathbf{C}} = \begin{bmatrix} \tilde{C}_{11} & \cdots & \tilde{C}_{N_H 1} \\ & \vdots & \\ \tilde{C}_{1K} & \cdots & \tilde{C}_{N_H K} \end{bmatrix} = [\tilde{\mathbf{c}}_1, \cdots, \tilde{\mathbf{c}}_{N_H}] \tag{8}$$

where $\tilde{\mathbf{c}}_h = [\tilde{C}_{h1}, \cdots, \tilde{C}_{hK}]^T$ for $1 \leq h \leq N_H$.

Node v_t is then able to perfectly reconstruct its source data, if $\tilde{\mathbf{C}}$ satisfies the following two conditions: 1) $\tilde{\mathbf{c}}_h \neq \mathbf{0}_K$ for all $h \in \mathbf{H}_t$, where $\mathbf{0}_K$ denotes all zero vector with length K, and 2) $\tilde{\mathbf{C}}'$ is full-rank, where $\tilde{\mathbf{C}}'$ is the matrix where all $\tilde{\mathbf{c}}_h = \mathbf{0}_K$ for $1 \leq h \leq N_H$ are removed from $\tilde{\mathbf{C}}$. Condition 1) ensures that the received packets include all data that should be reconstructed. This is a widely accepted condition under the wireless network settings because of the broadcasting nature of wireless communications [35], [36]. The condition 2) guarantees that the decoding process can uniquely reconstruct data $\hat{x}_h$ for all $h \in \mathbf{H}_t$. Because it is shown that RLNC makes the global coding coefficient matrix be full-rank with high probability [12], [13][2], the condition 2) can be satisfied. The decoding process can then be implemented based on well-known approaches such as Gaussian elimination in a GF [37].

While the conditions for perfect reconstruction can generally be satisfied with high probability, some special applications (e.g., delay-sensitive applications, error-prone networks with a high packet loss rate, etc.) may cause a perfect reconstruction to fail. That is, random mixing in the intersession network coding may lead to an increased decoding delay if only a subset of the coded sources of interest arrives at the terminal node. In this case, alternative decoding algorithms [38]–[41] can be deployed.

In the rest of this paper, we propose a distributed strategy for robust network formation.

IV. MDP-Based Network Formation

In this section, we propose an MDP-based framework for network formation, where intermediate nodes $v_i, \forall i \in \mathbf{V}$ of the network are considered as autonomous decision making agents to find the optimal strategy. An illustrative overview for the proposed framework is shown in Fig. 2.

For an agent v_i, an MDP is a tuple $\langle \mathbf{S}, \mathbf{A}, P(s'|s,a), U(s,a,s'), \rho \rangle$, where $\mathbf{S}$ is the state space, $\mathbf{A}$ is the action space, and $P(s'|s,a) : \mathbf{S} \times \mathbf{A} \times \mathbf{S} \rightarrow [0,1]$ is the state transition probability that action $a \in \mathbf{A}$ in state $s \in \mathbf{S}$ leads to the next state $s' \in \mathbf{S}_i$, which is a real number between 0 and 1. $U(s,a,s') : \mathbf{S} \times \mathbf{A} \times \mathbf{S} \rightarrow \mathbb{R}$ is a utility obtained after transition to state s' from state s with action a, and $\rho \in [0,1]$ is the discount factor. The details are explained as follows.

[2] It is shown in [12] that if RLNC is employed, the probability that the global coding coefficient matrix is full-rank is at least $(1 - |\mathcal{D}|/2^M)^{|\mathcal{E}(\mathcal{G})|}$. In general settings, a GF size of 2^M is significantly larger than the number of terminals $|\mathcal{D}|$ in the network. Hence, it is widely accepted that the global coding coefficient matrix is full-rank with high probability if RLNC is used.

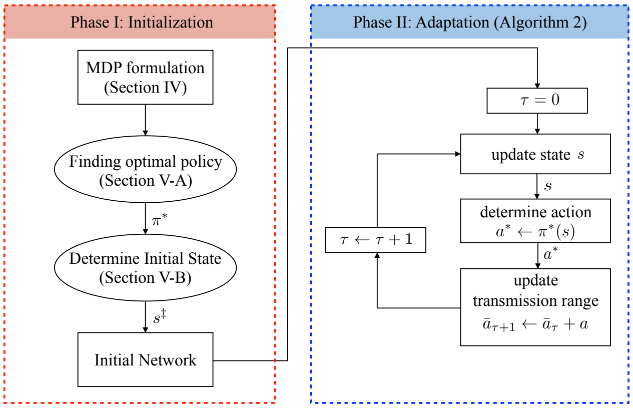

Phase I: Initialization
MDP formulation (Section IV)
Finding optimal policy (Section V-A)
π^*
Determine Initial State (Section V-B)
$s^\ddagger$
Initial Network
Phase II: Adaptation (Algorithm 2)
$\tau = 0$
update state s
s
determine action $a^* \leftarrow \pi^*(s)$
a^*
update transmission range $\bar{a}_{\tau+1} \leftarrow \bar{a}_\tau + a$
$\tau \leftarrow \tau + 1$

Fig. 2: An illustrative overview of the proposed system.

1) State Space $\mathbf{S}$*:* A state $s \in \mathbf{S}$ represents the expected number of effective nodes in the transmission range of the agent. Since a node can always be the effective node, at most $(N_V + N_T)$ nodes can be located in the transmission range, so that $1 \leq s \leq \lceil (N_V + N_T)/(1-\beta) \rceil$ in the channel with a β link failure rate.

Note that the definition of the state allows the network to be robust against network dynamics since a node can adaptively change its transmission ranges by considering node mobility and channel condition. If the network is static with node density λ, simply determining $\bar{\delta}_{i,\tau}$ can provide a solution for network topology, which directly determines the number of successfully received nodes, i.e., s, based on node density λ. If the network is dynamic, however, s at time τ cannot be directly determined by $\bar{\delta}_{i,\tau}$ since λ may not be a true value in the transmission range of the agent because of node mobility and the link failure of the channel. Hence, we design topology based on s rather than $\bar{\delta}_{i,\tau}$, which allows a node to adaptively change $\bar{\delta}_{i,\tau}$ against network dynamics, leading to a robust network.

2) Action Space $\mathbf{A}$*:* An action $a \in \mathbf{A}$ represents the increases in the transmission range as compared to a previous transmission range. Hence, the action at time τ becomes $a = \bar{a}_\tau - \bar{a}_{\tau-1}$, where $\bar{a}_\tau$ and $\bar{a}_{\tau-1}$ denote the transmission ranges at time τ and $\tau - 1$, respectively. If $a > 0$, the agent increases the transmission range (i.e., $\bar{a}_\tau > \bar{a}_{\tau-1}$). Similarly, if $a < 0$, the agent decreases the transmission range (i.e.,$\bar{a}_\tau < \bar{a}_{\tau-1}$). The agent may keep the same transmission range by taking action $a = 0$.

3) State Transition Probability $P(s'|s,a)$*:* A state transition probability represents the probability that a node in state s moves to state s' if action a is taken. Thus, $P(s'|s,a)$ means the probability that s' effective nodes will be included in the transmission range of the node in the next time stamp by taking action a from current s effective nodes. Since the number of intermediate nodes in a bounded region follows an independent homogeneous PPP with node density λ, the state transition probability can be described as in Theorem 3.

Theorem 3. *The state transition probability $P(s'|s,a)$ is given by*

$$P(s'|s,a) = \begin{cases} \frac{(\lambda a)^{\xi'-\xi} \cdot e^{-\lambda a}}{(\xi'-\xi)!} & a > 0 \\ 1 & a = 0 \\ \binom{\xi}{\xi'} (1 - \frac{|a|}{\bar{a}_\tau})^{\xi'} (\frac{|a|}{\bar{a}_\tau})^{\xi - \xi'} & a < 0. \end{cases} \tag{9}$$

where $\xi \triangleq \lceil \frac{s}{1-\beta} \rceil$ and $\xi' \triangleq \lceil \frac{s'}{1-\beta} \rceil$.

Proof. Based on the Kolmogorov definition of conditional probability, the state transition probability given in (9) can be expressed as

$$P(s'|s,a) \cdot P(s) = P(s|s',-a) \cdot P(s'). \tag{10}$$

Let ξ be the number of nodes at time τ included in $\bar{a}_\tau$, and thus, the corresponding probability is given by

$$\Pr\{\xi; \bar{a}_\tau\} = \frac{(\lambda \bar{a}_\tau)^\xi \cdot e^{-\lambda \bar{a}_\tau}}{\xi!}. \tag{11}$$

By considering the link failure rate β of the channel, the expected number of effective nodes becomes $s = (1-\beta) \cdot \xi$. Hence, $P(s) = \Pr\{\xi; \bar{a}_\tau\}$ with integer value $\xi \triangleq \lceil \frac{s}{1-\beta} \rceil$.

We assume that the state transition interval is short enough under mild regularity conditions [42]. Hence, $a = 0$ (which does not change the transmission range) does not lead to a state transition, i.e., $s' = s$. Therefore,

$$\begin{aligned} P(s'|s,a) \cdot P(s) &= 1 \cdot P(s) \\ &= 1 \cdot P(s') = P(s|s',-a) \cdot P(s'). \end{aligned}$$

If $a > 0$ (which enlarges the transmission range), more nodes can be included in the transmission region. Hence, $s' \geq s$. In this case, $P(s'|s,a) \cdot P(s')$ in (10) can be derived as

$$\begin{aligned} & P(s'|s,a) \cdot P(s) \\ &= P(s'|s,a) \cdot \Pr\{\xi; \bar{a}_\tau\} \\ &= \frac{(\lambda a)^{\xi'-\xi} \cdot e^{-\lambda a}}{(\xi'-\xi)!} \cdot \frac{(\lambda \bar{a}_\tau)^\xi \cdot e^{-\lambda \bar{a}_\tau}}{\xi!} \\ &= \frac{(\lambda(\bar{a}_{\tau+1} - \bar{a}_\tau))^{\xi'-\xi} \cdot e^{-\lambda(\bar{a}_{\tau+1} - \bar{a}_\tau)}}{(\xi'-\xi)!} \cdot \frac{(\lambda \bar{a}_\tau)^\xi \cdot e^{-\lambda \bar{a}_\tau}}{\xi!} \\ &= \frac{(\lambda(\bar{a}_{\tau+1} - \bar{a}_\tau))^{\xi'-\xi} \cdot e^{-\lambda(\bar{a}_{\tau+1})} (\lambda \bar{a}_\tau)^\xi}{(\xi'-\xi)!\xi!} \cdot \frac{(\lambda \bar{a}_{\tau+1})^{\xi'}}{(\lambda \bar{a}_{\tau+1})^{\xi'}} \cdot \frac{\xi_{i,\tau+1}!}{\xi_{i,\tau+1}!} \\ &= \frac{(\lambda \bar{a}_{\tau+1})^{\xi'} e^{-\lambda(\bar{a}_{\tau+1})}}{\xi'!} \cdot \frac{\xi'!}{(\xi'-\xi)!\xi!} \cdot (1 - \frac{\bar{a}_\tau}{\bar{a}_{\tau+1}})^{\xi'-\xi} \cdot (\frac{\bar{a}_\tau}{\bar{a}_{\tau+1}})^\xi \\ &= \Pr\{\xi'; \bar{a}_{\tau+1}\} \cdot \binom{\xi'}{\xi} \cdot (\frac{a_i}{\bar{a}_\tau})^{\xi'-\xi} \cdot (1 - \frac{a_{i,\tau}}{\bar{a}_{\tau+1}})^\xi \\ &= P(s|s',-a) \cdot P(s'). \end{aligned}$$

Similarly, $P(s'|s,a) \cdot P(s) = P(s|s',-a) \cdot P(s')$ in (10) is obtained for $a < 0$.

Therefore, we conclude that the state transition probability can be expressed as (9). □

Theorem 3 implies that the state transition probability is the probability that $(\xi' - \xi)$ nodes are included in the transmission range a for $a > 0$. If $a < 0$, the probability that ξ' nodes are in $\bar{a}_{\tau+1}$ given ξ nodes in $\bar{a}_\tau$ is the probability that $(\xi - \xi')$ nodes are included in $|a|$.

4) Utility Function $U(s, a, s')$: We define the utility function of node v_i as a quasi-linear function that consists of a reward and a cost, i.e.,

$$U(s, a, s') = u + \omega \cdot R(s, s') - (1 - \omega) \cdot a \qquad (12)$$

where $R(s, s')$ is the reward function that represents immediate throughput improvement given the state transition from s to s' at the cost of taking action a, which increases the transmission range. The weight $\omega (0 \leq \omega \leq 1)$ can be used to balance the reward and the cost. For example, if $\omega = 1$, the cost associated with taking action a can be ignored, but only the throughput improvement is considered. Since a utility is generally non-negative, a constant u is introduced in (12), and it can be set such that $U(s, a, s') \geq 0$. The reward function $R(s, s')$ is defined as

$$R(s, s') = \gamma(s') - \gamma(s)$$

where $\gamma(s)$ denotes network throughput when the node is in state s, which is a concave increasing function.

5) Discount factor ρ: The discount factor $\rho \in [0, 1]$ represents the degree of utility reduction over time, so that it determines the cumulative long-term utility. The discount factor can be determined based on the consistency of the network condition (e.g., [43], [44]). For example, if the network condition is static, a large value of ρ can be used by imposing a high weight on the predicted future utilities whereas a lower value of ρ needs to be used in more dynamically changing network conditions.

Next, we show that the proposed framework satisfies the Markov property.

Theorem 4. *The tuple $\langle \mathbf{S}, \mathbf{A}, P(s'|s, a), U(s, a, s'), \rho \rangle$ satisfies the Markov property.*

Proof. Let $\langle s_1, a_1 \rangle, \langle s_2, a_2 \rangle, \ldots, \langle s_\tau, a_\tau \rangle$ be the sequence of events, where $\langle s_\tau, a_\tau \rangle$ is an event which includes an action at time τ (i.e., a_τ) and a corresponding resulting state (i.e., s_τ)[3]. The initial transmission range and corresponding state are denoted by $\bar{a}_0$ and s_1, respectively. To show that the tuple $\langle \mathbf{S}, \mathbf{A}, P(s'|s, a), U(s, a, s'), \rho \rangle$ satisfies the Markov property, our aim is to prove

$$P\left(s_{\tau+1} \middle| \langle s_\tau, a_\tau \rangle, \ldots, \langle s_1, a_1 \rangle\right) = P(s_{\tau+1} | s_\tau, a_\tau).$$

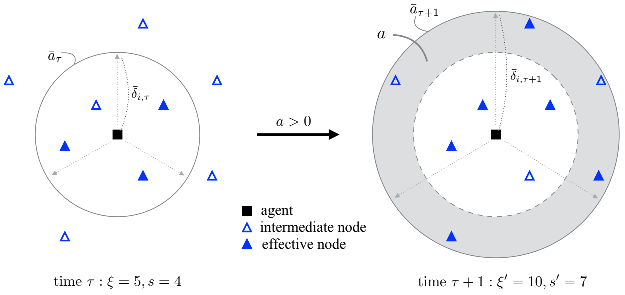


Fig. 3: An illustrative explanation of the proposed framework.

The state transition probability that action a_1 leads a node in state s_1 to a new state s_2 can be expressed as

$$P\left(s_2 | s_1, a_1\right) = \Pr\{s_1 + (s_2 - s_1); \bar{a}_0 + a_1 | s_1; \bar{a}_0\}$$

which implies that $(s_2 - s_1)$ nodes are additionally included in the transmission range expanded by a_1. Similarly, $P\left(s_3 \middle| \langle s_2, a_2 \rangle, \langle s_1, a_1 \rangle\right)$ can be expressed as

$$\begin{aligned} & P(s_3 | \langle s_2, a_2 \rangle, \langle s_1, a_1 \rangle) \\ & = \Pr\{s_1 + (s_2 - s_1) + (s_3 - s_2); \lambda(\bar{a}_0 + a_1 + a_2) \\ & \qquad |s_1 + (s_2 - s_1); \lambda(\bar{a}_0 + a_1)\} \\ & = \Pr\{s_1 + \sum_{t=1}^{2}(s_{t+1} - s_t); \lambda(\bar{a}_0 + \sum_{t=1}^{2} a_t) \\ & \qquad |s_1 + \sum_{t=1}^{1}(s_{t+1} - s_t); \lambda(\bar{a}_0 + \sum_{t=1}^{1} a_t)\}. \end{aligned}$$

By induction, $P\left(s_{\tau+1} \middle| \langle s_\tau, a_\tau \rangle, \ldots, \langle s_1, a_1 \rangle\right)$ can be expressed as

$$\begin{aligned} & P(s_{\tau+1} | \langle s_\tau, a_\tau \rangle, \ldots, \langle s_1, a_1 \rangle) \\ & = \Pr\{s_1 + \sum_{t=1}^{\tau}(s_{t+1} - s_t); \lambda(\bar{a}_0 + \sum_{t=1}^{\tau} a_t) \\ & \qquad |s_1 + \sum_{t=1}^{\tau-1}(s_{t+1} - s_t); \lambda(\bar{a}_0 + \sum_{t=1}^{\tau-1} a_t)\} \qquad (13) \\ & = \Pr\{s_{\tau+1} - s_\tau; \lambda \cdot a_\tau\} \qquad (14) \\ & = P(s_{\tau+1} | s_\tau, a_\tau). \end{aligned}$$

The equality between (13) and (14) can be derived by

$$s_{t+1} - s_t = \sum_{t=1}^{\tau}(s_{t+1} - s_t) - \sum_{t=1}^{\tau-1}(s_{t+1} - s_t)$$

and

$$\lambda \cdot a_\tau = \lambda(\bar{a}_0 + \sum_{t=1}^{\tau} a_t) - \lambda(\bar{a}_0 + \sum_{t=1}^{\tau-1} a_t)$$

which completes the proof. □

Theorem 4 shows that the proposed framework in this section can be modeled by the MDP. Fig. 3 shows an illustration of the proposed framework.

In the next section, we show how the strategy enables each node to make its own optimal decisions.

V. Distributed Network Formation Strategy

A. MDP-Based Optimal Strategy for Network Formation

The solution to an MDP is the optimal policy that maps the optimal actions performed in a particular state. Specifically,

[3] In this proof, we add time stamps τ on the notation of states and actions, e.g., s_τ and a_τ, to clearly specify the time that an action is taken.

Algorithm 1 Algorithm for ϵ-Optimal Policy

```
Require: state space $\mathbf{S}$, action space $\mathbf{A}$, utility function $U(s,a,s')$, weight $\omega$, discount factor $\rho$, state transition probability $P(s'|s,a)$, optimality level $\epsilon$
Initialize: $V_0(s) \leftarrow 0, \forall s \in \mathbf{S}$, $\tau \leftarrow 0$
while $V_\tau(s) - V_{\tau-1}(s) > \frac{1-\rho}{2\rho}\epsilon$ for any $s \in \mathbf{S}$ do
    for $\forall s \in \mathbf{S}$ do
        update $V_{\tau+1}(s) \leftarrow \max_{a\in\mathbf{A}} \left(\sum_{s'\in\mathbf{S}} P(s'|s,a) \times (U(s,a,s') + \rho V_\tau(s'))\right)$
    $\tau \leftarrow \tau + 1$
choose $\epsilon$-optimal policy $\pi^{\epsilon^*}(s) \leftarrow \arg\max_a V_\tau(s), \forall s \in \mathbf{S}$
return $\epsilon$-optimal policies $\pi^{\epsilon^*}$
```

the policy is a function $\pi : \mathbf{S} \rightarrow \mathbf{A}$ which returns an action for a state, i.e, $\pi(s) = a$. The policy π is optimal if it can maximize the *state-value function* $V_\tau(s)$.

The state-value function $V_\tau(s)$ represents a cumulative utility at time τ, starting from state s, expressed as

$$\begin{aligned} V_{\tau+1}(s) &= U(s,a,s') + \rho U(s',a,s'') + \rho^2 U(s'',a,s''') + \cdots \\ &= U(s,a,s') + \rho V_\tau(s') \end{aligned} \tag{15}$$

where state s sequentially moves into s', s'', and s'''. In (15), $V_\tau(s)$ includes the *immediate utility* $U(s,a,s')$ and the discounted state-value of successive states $\rho V_\tau(s')$. The expected value for the state-value function is thus expressed as

$$\mathbb{E}(V_{\tau+1}(s)) = \sum_{s'\in\mathbf{S}} P(s'|s,a)\left(U(s,a,s') + \rho V_\tau(s')\right). \tag{16}$$

The *optimal state-value function* $V^*(s)$ is the maximum state-value function over all policies, i.e,

$$V^*(s) = \max_\pi \mathbb{E}(V^\pi(s)) \tag{17}$$

where $V^\pi(s)$ is the state-value achieved by the actions determined by policy π at every state. Finally, the optimal policy π^* is the policy that leads to $V^*(s)$, and it is defined as

$$\pi^* = \arg\max_\pi \mathbb{E}(V^\pi(s)), \forall s \in \mathbf{S} \tag{18}$$

This is also known as the Bellman optimality equation [45]. Given the optimal policy π^*, optimal action a^* for each state s can be determined such that

$$\begin{aligned} a^* &= \pi^*(s) \\ &= \arg\max_{a\in\mathbf{A}} \sum_{s'\in\mathbf{S}} P(s'|s,a)\left(U(s,a,s') + \rho V^*(s')\right). \end{aligned}$$

In practice, a near-optimal policy is widely used as it requires lower computational complexity. π^*_ϵ is an ϵ-optimal policy if

$$||V^{\pi^*_\epsilon}(s) - V^*(s)||_\infty \leq \epsilon$$

which means that the error between $V^{\pi^*_\epsilon}(s)$, the state-value derived by π^*_ϵ, and $V^*(s)$ is bounded by the optimality level ϵ. The ϵ-optimal policy can be found using Algorithm 1.

Using stopping criteria, we show that Algorithm 1 converges to ϵ-optimal policies in Theorem 5.

Theorem 5. *The ϵ-optimal policy can be achieved by Algorithm 1 if the iteration stops with condition*

$$||V_\tau(s) - V_{\tau-1}(s)||_\infty \leq \frac{1-\rho}{2\rho}\epsilon$$

for all $s \in \mathbf{S}$.

Proof. See Appendix B. □

It can also be shown in Theorem 6 that Algorithm 1 converges to the optimal policy π^* by setting $\epsilon = 0$.

Theorem 6. *The optimal policy π^* can always be achieved by setting $\epsilon = 0$ in Algorithm 1, i.e.,*

$$\lim_{\tau\to\infty} V_\tau(s) = V^*(s) \tag{19}$$

for all $s \in \mathbf{S}$.

Proof. See Appendix C. □

The proof of Theorem 6 is shown in (56) in Appendix C, Based on this proof, we concluded that

$$\lim_{\tau\to\infty} ||V_\tau(s) - V^*(s)||_\infty \leq \lim_{\tau\to\infty} \rho^\tau ||V_0(s) - V^*(s)||_\infty$$

where $||||_\infty$ denotes the infinite norm. This shows that the convergence speed of Algorithm 1 significantly depends on the discount factor ρ. Hence, the convergence speed can be controlled by discount factor ρ.

Using the optimal policy, a node now can adaptively change its transmission range against network dynamics, which leads to a robust network. It is worth noting that the complexity to find the optimal policy at each node does not change, even if the total number of nodes in network increases. Hence, as the number of nodes increases, the total complexity to find the optimal policies of all nodes in the network increases linearly. This is because the proposed MDP framework of each node is not affected by individual network member nodes, instead it is only affected by node density λ in network. In the next section, we study how to determine the initial state for each node which determines the initial transmission range.

B. Stationary Network with Optimal Policy

Each node can periodically change its transmission range according to the optimal policy obtained from Algorithm 1. The resulting network can be in stationary, i.e., the number of network nodes is unchanged if each node takes action based on the optimal policy. In this section, we discuss how to initial conditions are determined such that the convergence speed for the optimal policy can be expedited in practice.

With the optimal policy π^*, the proposed MDP framework is reduced to the Markov chain with a state transition matrix $\mathbf{P}$ whose element at (s,s') is denoted by $\mathbf{P}(s,s')$, which is expressed as

$$\mathbf{P}(s,s') = P(s'|s,\pi^*(s)) \tag{20}$$

This is the state transition probability $P(s'|s,a)$ in (9) with the optimal action $a = \pi^*(s)$. The state transition probability $\mathbf{P}(s, s')$ provides the probability that a single state transition changes a node in s to s'. Then, the limiting matrix $\lim_{n\to\infty} \mathbf{P}^n$ and the limiting distribution $\boldsymbol{\sigma} = [\sigma_1, \ldots, \sigma_s \ldots, \sigma_{|\mathbf{S}|}]$ which can be obtained as

$$\sigma_s = \lim_{n\to\infty} \frac{\sum_{s''\in\mathbf{S}} \mathbf{P}^n(s'', s)}{\sum_{s'\in\mathbf{S}} \sum_{s''\in\mathbf{S}} \mathbf{P}^n(s'', s')} \tag{21}$$

where σ_s denotes the probability of being in state s after an infinite number of state transitions. Finally, the initial state $s^\dagger$ can be determined by choosing the state with the highest limiting distribution, i.e.,

$$s^\dagger = \arg_{s\in\mathbf{S}} \max \sigma_s \tag{22}$$

which allows the initial network to be formed close to the stationary network with the highest probability.

1) The optimal action includes no change of transmission range: If the optimal action at a state s^* is not to change its transmission range, i.e., $a^* = \pi^*(s^*) = 0$, the s^*th row of $\mathbf{P}$ can be expressed by the definition of $P(s'|s, a)$ in (9) as

$$\mathbf{P}(s^*, s') = \begin{cases} 1, & s' = s^* \\ 0, & s' \in \{\mathbf{S}\backslash s^*\} \end{cases} \tag{23}$$

where $\mathbf{S}\backslash s^*$ denotes the set of elements in $\mathbf{S}$ excluding s^*. This allows the state transition matrix $\mathbf{P}$ to be formulated in canonical form as

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \tag{24}$$

where $\mathbf{Q} \geq 0$ is a nonnegative matrix, $\mathbf{R} > 0$ is a strictly positive matrix, $\mathbf{0}$ denotes the matrix with zeros and $\mathbf{I}$ denotes the identity matrix. The size of matrix $\mathbf{I}$ becomes the number of states whose actions are zero.

Then, the limiting matrix of $\mathbf{P}$ in (24) becomes

$$\begin{aligned} \lim_{n\to\infty} \mathbf{P}^n &= \lim_{n\to\infty} \begin{pmatrix} \mathbf{Q}^n & \mathbf{Q}^{n-1}\mathbf{R} + \cdots + \mathbf{Q}\mathbf{R} + \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{0} & \mathbf{F}\mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \end{aligned} \tag{25}$$

where $\mathbf{F} = (\mathbf{I} - \mathbf{Q})^{-1}$ is the fundamental matrix of $\mathbf{Q}$. $\lim_{n\to\infty} \mathbf{Q}^n = \mathbf{0}$ as the element of $\mathbf{Q}$ is in $[0, 1)$. Then, an element σ_s in the limiting distribution $\boldsymbol{\sigma}$ can be obtained based on (21)

$$\sigma_s = \frac{\sum_{\forall i} (\mathbf{F}\mathbf{R})_{ij}}{\zeta}$$

where $(\mathbf{F}\mathbf{R})_{ij}$ denotes the (i, j)th element of the matrix $\mathbf{F}\mathbf{R}$ and $\zeta = \sum_{\forall j} \left(\sum_{\forall i} (\mathbf{F}\mathbf{R})_{ij} + 1 \right)$ is a constant. Therefore, the initial state $s^\dagger$ can be determined as

$$\begin{aligned} s^\dagger &= \arg_{s\in\mathbf{S}} \max \sigma_s \\ &= \arg_j \max \sum_{\forall i} (\mathbf{F}\mathbf{R})_{ij}. \end{aligned} \tag{26}$$

In the case where the set of optimal actions does not include no change in transmission range, the state transition matrix $\mathbf{P}$ cannot be formulated as shown in (24). This is discussed next.

2) Optimal action does not include no change of transmission range: Since $a^* = \pi^*(s^*) = 0$ is not available as an action, the optimal actions are either to enlarge or reduce the transmission range.

If the optimal action at a state s is to enlarge the transmission range (i.e., $a^* = \pi^*(s^*) > 0$), the sth row of $\mathbf{P}$ becomes

$$\mathbf{P}(s, s') = \begin{cases} 0, & s' < s \\ \frac{(\lambda a)^{\xi'-\xi} \cdot e^{-\lambda a}}{(\xi'-\xi)!}, & s' \geq s \end{cases}. \tag{27}$$

Similarly, if the optimal action at a state s is to reduce the transmission range (i.e., $a^* = \pi^*(s^*) < 0$), the sth row of $\mathbf{P}$ becomes

$$\mathbf{P}(s, s') = \begin{cases} \binom{\xi}{\xi'} (1 - \frac{|a|}{\bar{a}_\tau})^{\xi'} (\frac{|a|}{\bar{a}_\tau})^{\xi-\xi'}, & s' \leq s \\ 0, & s' > s \end{cases}. \tag{28}$$

The state transition matrix $\mathbf{P}$ can be correspondingly expressed as a canonical form of

$$\mathbf{P} = \begin{pmatrix} \mathbf{U} & \mathbf{Q}_1 \\ \mathbf{Q}_2 & \mathbf{L} \end{pmatrix} \tag{29}$$

where $\mathbf{U}$ and $\mathbf{L}$ denote the upper and lower triangular matrices and $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are strictly positive matrices. Since the optimal actions are determined by considering both rewards and costs, an optimal action can be determined to enlarge the current transmission range if a node is in a state with too few nodes. On the other hand, if a node is in a state with too many nodes, the optimal policy may determine the optimal action that reduces the transmission range such that the cost can be reduced. Hence, the state transition matrix $\mathbf{P}$ in (29) consists of $(\mathbf{U} \quad \mathbf{Q}_1)$ and $(\mathbf{Q}_2 \quad \mathbf{L})$.

Note that $\mathbf{P}^n$ for $n \geq 2$ is a strictly positive matrix. For example,

$$\begin{aligned} \mathbf{P}^2 &= \begin{pmatrix} \mathbf{U} & \mathbf{Q}_1 \\ \mathbf{Q}_2 & \mathbf{L} \end{pmatrix} \begin{pmatrix} \mathbf{U} & \mathbf{Q}_1 \\ \mathbf{Q}_2 & \mathbf{L} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{U}^2 + \mathbf{Q}_1\mathbf{Q}_2 & \mathbf{U}\mathbf{Q}_1 + \mathbf{Q}_2\mathbf{L} \\ \mathbf{Q}_2\mathbf{U} + \mathbf{L}\mathbf{Q}_2 & \mathbf{Q}_2\mathbf{Q}_1 + \mathbf{L}^2 \end{pmatrix} \end{aligned}$$

which becomes a strictly positive matrix. Hence, Perron-Frobenius theorem [46] guarantees that there is a unique largest eigenvalue and the largest eigenvalue is 1 since $\mathbf{P}$ is a stochastic matrix. Therefore, the unique limiting distribution $\boldsymbol{\sigma}$ can be found as a row eigenvector of $\mathbf{P}$ associated with eigenvalue 1, i.e., $\boldsymbol{\sigma}\mathbf{P} = \boldsymbol{\sigma}$, and the initial state becomes the state with the largest limiting distribution as shown in (22).

The initialization phase of the proposed system can be expedited by choosing the initial state of each node that leads to the initial network.

VI. Simulation Results

In this simulation, we consider a wireless ad hoc network with multi-source multicast flows where multiple intermediate nodes aim to relay source data to multiple terminal nodes using network coding. All intermediate nodes are policy-compliant, meaning that each node builds its own optimal policy and correspondingly changes its transmission range based on the number of nodes included in its transmission range. In this

Algorithm 2 Algorithm for adaptation phase with packet transition

```
Require: optimal policy $\pi^*$
Initialize: $s \leftarrow s^\dagger$ // build the equilibrium network
while network is active do
    // receive and combine packets
    store received packets in buffer
    if $\mathcal{L}_i \neq \emptyset$ then
    //if the buffer is not empty
        build a network coded packet based on (4)
        // update network topology
        check the current state $s$
        find the optimal action: $a^* \leftarrow \pi^*(s)$
        update the transmission range: $\bar{a}_{\tau+1} \leftarrow \bar{a}_\tau + a$
        broadcast the network coded packet
    $\tau \leftarrow \tau + 1$
```

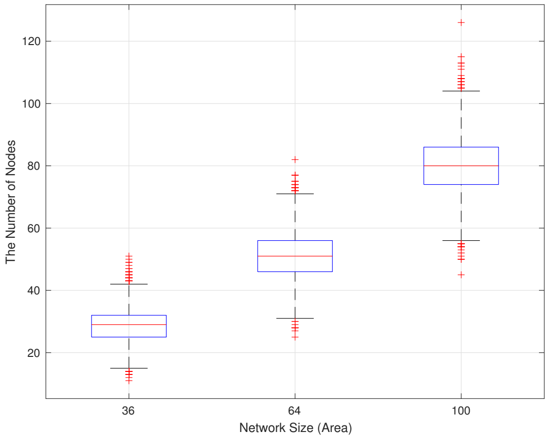


Fig. 4: The number of appeared nodes for given network size with a node density of $\lambda = 4/5$.

section, we present a network formation result based on the proposed strategy, and then we show a performance comparison with other existing network formation strategies in applications with Wi-Fi Direct.

A. Numerical Results for the Proposed Strategy

We consider a network with two source nodes, two terminal nodes and multiple intermediate nodes. The number of intermediate nodes follows the PPP with a node density of $4/5$. The network size denotes the size of the area in the network, and three different network sizes are considered. The results presented in this section are based on $1,000$ independent experiments with a randomly generated number of nodes in a network size.

Fig. 4 shows the number of nodes in given network areas that are determined by the PPP with a node density of $\lambda = 4/5$. The line in the middle of each box in Fig. 4 denotes the median of the experiments, which are 29, 51, and 80 for network sizes of 36, 64, and 100, respectively. The top and bottom of each box are the 25th and 75th percentiles, respectively. Hence, it is confirmed that intermediate nodes are well generated by the PPP based on the node density.

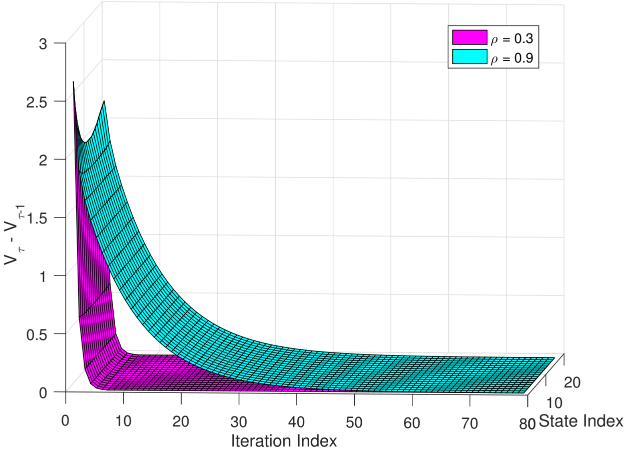


Fig. 5: Convergence process in Algorithm 1

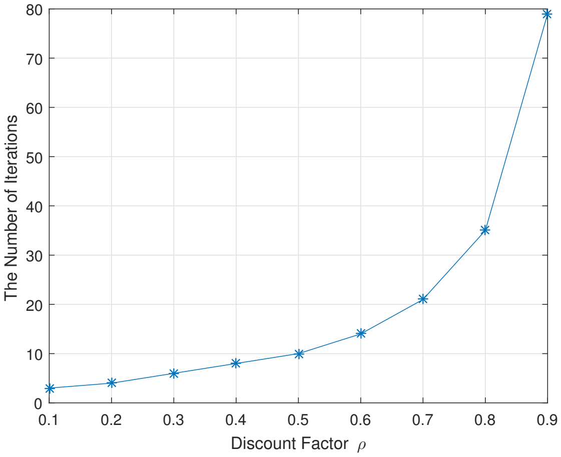


Fig. 6: Convergence speed of Algorithm 1 as a function of discount factor

Each agent builds ϵ-optimal policy based on Algorithm 1 with the parameter $\epsilon = 0.01$, twenty states and five actions, i.e., $|\mathbf{S}| = 20$ and $|\mathbf{A}| = 5$. Fig. 5 shows the values of $V_\tau(s) - V_{\tau-1}(s)$ for all $s \in \mathbf{S}$ over iterations and it is observed that $V_\tau(s) - V_{\tau-1}(s)$ approaches 0 as the number of iterations increases. Specifically, the iteration is terminated if $V_\tau(s) - V_{\tau-1}(s) \leq \frac{1-\rho}{2\rho}\epsilon$, and thus eventually $V_\tau(s) = V_{\tau-1}(s)$ with large enough iterations. Therefore, we conclude that the proposed algorithm converges. The convergence speed is dependent on the discount factor ρ as shown in Fig. 6. For larger ρ, which takes into account longer future utilities, it takes a longer time (i.e., more iterations) to find ϵ-policy. On the other hand, it takes less time to find the ϵ-policy for a small ρ.

We next study the resulting network in terms of two connectivity measures: the number of constructed links and algebraic connectivity [47]. The number of links constructed in the network reflects the extrinsic connectivity and can be quantified by counting the number of links in the network. In contrast, the algebraic connectivity is the measure of intrinsic connectivity, i.e., how well the overall network is constructed. Fig. 7 shows the impact of weight ω in utility function (12) on network connectivity. Since ω is the weight of reward in the utility function, it is expected that the resulting networks are formed such that the rewards (or the cost) are given more weight than the cost (or the rewards) if ω is high (or low). In the simulations, we assume that there is no link failure in the channels (i.e., $\beta = 0$) and the discount factor is $\rho = 0.5$. Fig. 7(a) shows that the number of links is proportional to both network size and ω. A node with high ω may increase the transmission range such that a larger number of links can be covered, leading to throughput gain over power consumption. This is also confirmed in Fig. 7(b), which shows high algebraic connectivity with high ω. However, Fig. 7(b) shows that the algebraic connectivity decreases as network size increases. This is because the proposed strategy does not consider to retain the same algebraic connectivity. Hence, if the same algebraic connectivity is required, a higher ω should be considered in a larger network.

The network connectivity as a function of the link failure rate β ($0 \leq \beta \leq 0.3$) is shown in Fig. 8. Fig. 8(a) shows that the proposed strategy enables nodes to make more links as β increases. This enables the networks to maintain approximately the same number of effective nodes. Moreover, it is confirmed from Fig. 8(b) that the algebraic connectivity increases as β increases. This is because the proposed approach increases the degree of connectivity of the network to overcome unstable channel conditions. Therefore, we conclude that the proposed strategy is successful at adaptively changing network topology by explicitly considering the link failure rates of the channels.

B. Performance Comparison in Wi-Fi Direct Application

In this section, we consider an illustrative application with Wi-Fi Direct where data are transmitted over dynamic wireless networks in a 60×60 [m^2] area. Mobile nodes are located at a density of 8×10^{-3}[nodes/m^2] and are connected by Wi-Fi Direct with IEEE 802.11ac standard MCS-9. The parameters used in the simulations are shown in Table II, and they are specified by the IEEE 802.11ac standard [48], [49]. The RLNC is used in the GF(2^8). The performance of the proposed strategy is evaluated based on the system goodput [50], which is defined as the sum of data rates successfully delivered to terminal nodes, expressed as

$$\sum_{h=1}^{N_H} \sum_{v_t \in \tilde{\mathbf{T}}_h} \frac{L}{\bar{\tau}(x_h, v_t)}$$

TABLE II: Simulation Parameters

Parameter	Value	Parameter	Value
η	1	α	2
Channel Bandwidth	80 MHz	TX-RX Antennas	3×3
Modulation Type	25-QAM	Coding Rate	5/6
Guard Interval	400ns	PHY Data Rate	1300 $Mbps$
MAC Efficiency	70%	Throughput	910 $Mbps$

where $\tilde{\mathbf{T}}_h \subseteq \mathbf{T}_h$ denotes a set of successfully delivered terminal nodes of x_h, L represents the size data set x_h, and $\bar{\tau}(x_h, v_t)$ denotes the travel time for data set x_h to arrive to a terminal node $v_t \in \tilde{\mathbf{T}}_h$. Moreover, the transmit power is measured by a path loss model, expressed as

$$P_{TX} = P_{RX} \cdot \left(\frac{4\pi}{\lambda} \cdot d \right)^{\alpha} = \eta \cdot d^{\alpha}$$

where P_{TX}, P_{RX}, λ, and d denote transmit power, receive power, wave length, and the distance between transmitter and receiver, respectively.

In this simulation, we simultaneously consider three types of network dynamics: changes in member nodes of considered network, link failure rates, and node locations. To produce realistic dynamic network settings, the location of network nodes is changed in every time stamp, and the network member is updated, and the link failure rates β are updated (i.e., randomly selected in $[0, 0.3]$) every 5 time stamps. The simulation parameters are set to $\omega = 0.53$, $u = 0.2$, $\epsilon = 0.01$, $|\mathbf{S}| = 18$ and $|\mathbf{A}| = 7$.

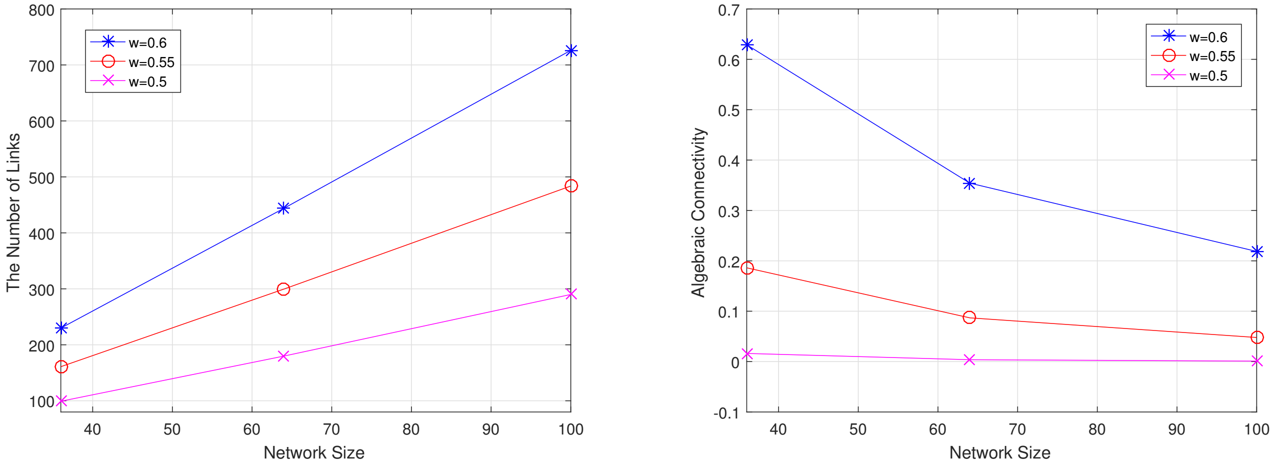
We compare the performance of the proposed strategy with the following three existing network formation strategies.

1) *Myopic*: A myopic strategy is a special case of the proposed strategy with the setting of $\rho = 0$ in (18). The myopic solution does not consider the future utilities. Rather, it focuses on maximizing the immediate utility only, i.e.,

$$\pi^{myop} = \arg\max_{a \in \mathbf{A}} \sum_{s' \in \mathbf{S}} P(s'|s, a) U(s, a, s'), \forall s \in \mathbf{S}.$$

2) *Traskov* [21]: A well-known centralized network formation strategy for network coding deployed networks. Traskov can provide a static network topology for a given node distribution by exploiting network coding opportunities. Hence, in the simulations, we consider the network where (network size $\times$ λ) nodes are uniformly distributed, and we find the network topology based on Traskov. Since Traskov determines individual links, the transmission range of a node is assigned to include all links determined from Traskov. To ensure a fair comparison, the assigned transmission range is not changed over time since the computational complexity for Traskov is much higher than that of other distributed strategies.

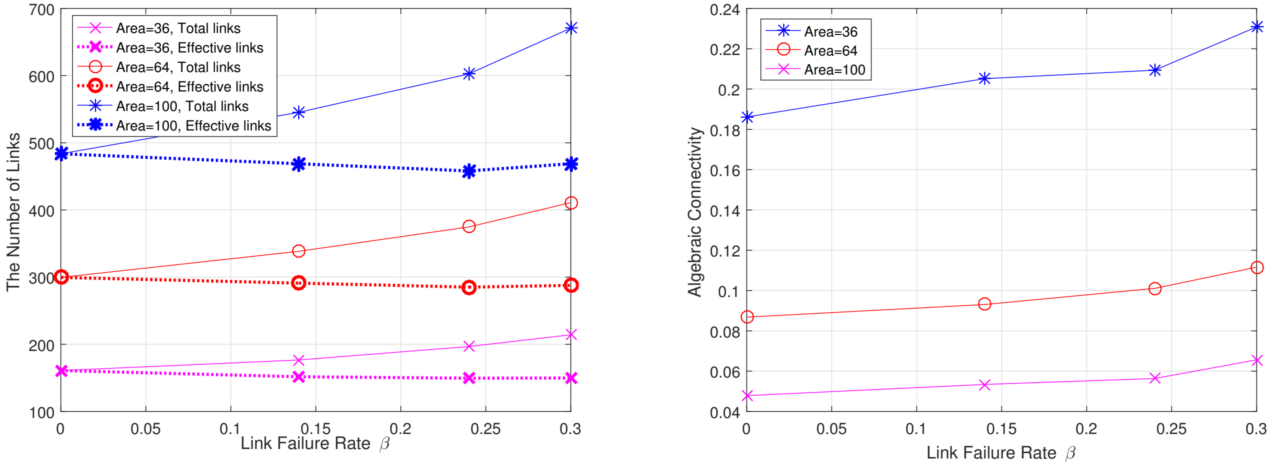
3) *TCLE* [25]: A state-of-the-art distributed strategy for network formation based on a non-cooperative game. In this strategy, a node chooses its transmission power by balancing the target algebraic connectivity against transmission energy dissipation. To ensure a fair comparison, the same set of actions are employed as the proposed strategy and the target algebraic connectivity



(a)

(b)

Fig. 7: The impact of ω on network connectivity ($\beta = 0, \rho = 0.5$)



(a)

(b)

Fig. 8: Network connectivity over link failure rate ($0 \leq \beta \leq 0.3, \omega = 0.55, \rho = 0.5$)

of TCLE is set as 0.1, which is the average algebraic connectivity of the proposed strategy with $\omega = 0.53$. Note that the network topology determined by TCLE does not adaptively change

To combat network dynamics, we allow TCLE to re-calculate its solution every 5 time stamps. Note that in terms of computational complexity, these TCLE settings require higher complexity than the proposed and myopic strategies, where nodes simply lookup the optimal policies during all the simulations, which are obtained in the beginning of the simulations.

The average numerical results from $1,000$ time stamps are summarized in Table III, and illustrative results in the time stamp range of $[380, 440]$ are shown in Fig. 9 and Fig. 10 for the radius of transmission range of a node and the number of total links in the overall network, respectively.

As shown in Table III, the proposed strategy provides the highest system goodput as well as high successful connectivity ratio. The proposed strategy always outperforms the myopic strategy. This is because the policy of the myopic strategy focuses only on the immediate utility, while the proposed strategy considers future utilities. For example, time stamps in $[430, 440]$ of Fig. 10 show that the proposed strategy more proactively responds to network dynamics than the myopic strategy by changing larger number of network links. Moreover, it is confirmed that the myopic strategy tends to result in smaller transmission ranges (shown in Fig. 9), which leads to a lower total number of links in the network (shown in Fig. 10).

While the second highest system goodput is achieved by the TCLE, it shows the second lowest successful connectivity ratio in Table III. This implies that the TCLE can make successful connections between a source and a terminal based

TABLE III: Numerical Results of Wi-Fi Direct Application

Strategy	System Goodput $[Mbps]$	Successful Connectivity Ratio $[\%]$	Power Consumption $[dBm]$
Proposed	324.60	75.63	89.07
Myopic	276.51	70.28	80.17
Traskov	226.59	38.19	103.14
TCLE	317.36	45.68	76.22

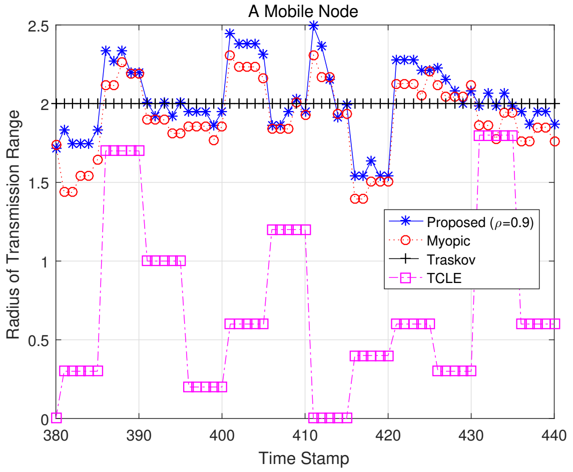


Fig. 9: Radius of transmission range of a node in presence of network dynamics.

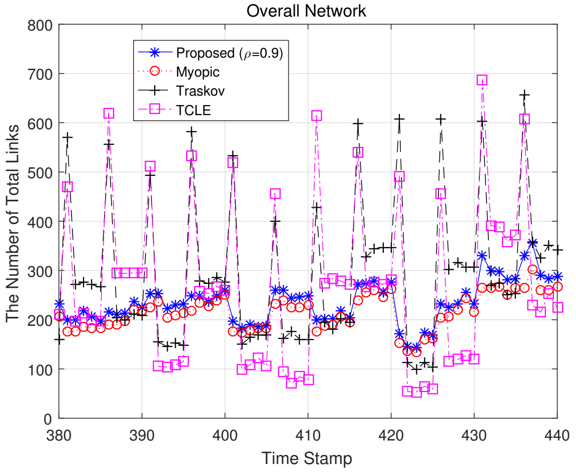


Fig. 10: The number of links in the resulting network in the presence of network dynamics.

on significantly short paths. However, the TCLE is not an appropriate solution for applications that count on successful delivery over throughput. Rather, it is the most energy-efficient strategy (Table III) as highlighted in [25], and it can determine smaller transmission ranges (in Fig. 9).

Traskov shows the lowest performance in terms of system goodput and successful connectivity ratio while it requires the highest power consumption. As shown in Fig. 9, Traskov does not change the transmission range once it is determined in the beginning of the simulation such that the result of network formation fails to overcome network dynamics.

VII. Conclusions

In this paper, we focus on a distributed network formation strategy that can build a robust network against network dynamics. We show that network coding induces packet anonymity and network decoupling such that the MDP framework can be employed at each intermediate node. The intermediate nodes determine an optimal policy based on MDP, and the policy allows the nodes to determine optimal transmission ranges that maximize the long-term cumulative utilities. The optimal transmission ranges are determined by explicitly considering current network conditions and future network dynamics. We further show that the resulting network of the proposed strategy converges to the stationary networks, and we propose how to determine an initial network that can rapidly converge to the stationary network. Simulation results confirm that the resulting network of the proposed strategy can adaptively change by responding to network dynamics such as unstable channel condition with high link failure rate, node mobility, and corresponding changes in member nodes associated with the considered network.

Appendix A
Bellman Operation and Its Properties

Let $\mathcal{T}^*$ be the Bellman optimality operator [45] for $V_\tau(s)$, which maps a foresighted state-value function to a foresighted state-value function (i.e., $\mathcal{T}^* : \mathbb{R}^{|\mathbf{S}|} \to \mathbb{R}^{|\mathbf{S}|}$), defined as,

$$(\mathcal{T}^* V_\tau)(s) = \max_{a \in \mathbf{A}} \left(\sum_{s' \in \mathbf{S}} P(s'|s,a) \left(U(s,a,s') + \rho V_\tau(s') \right) \right). \tag{30}$$

This updates the state-value function with the action that provides the maximum expected long-term state value.

The Bellman optimality operator $\mathcal{T}^*$ has monotonicity, additivity and ρ-contraction properties as shown below.

Property 7. *(Monotonicity of Bellman Optimality Operator)* $V_\tau(s) \leq V'_\tau(s), \forall s \Rightarrow (\mathcal{T}^* V_\tau)(s) \leq (\mathcal{T}^* V'_\tau)(s)$

Proof. For the optimal action $a^* = \pi^*(s) = \arg\max_a \sum_{s \in \mathbf{S}} P(s'|s,a) \left(U(s,a,s') + \rho V^*(s') \right)$ and when $V_\tau(s) \leq V'_\tau(s), \forall s$, $(\mathcal{T}^* V_\tau)(s) - (\mathcal{T}^* V'_\tau)(s)$ becomes:

$$\begin{aligned} &(\mathcal{T}^* V_\tau)(s) - (\mathcal{T}^* V'_\tau)(s) \\ &= \left(\sum_{s' \in \mathbf{S}} P(s'|s,a^*) \left(U(s,a^*,s') + \rho V_\tau(s') \right) \right) \\ &- \left(\sum_{s' \in \mathbf{S}} P(s'|s,a^*) \left(U(s,a^*,s') + \rho V'_\tau(s') \right) \right) \end{aligned}$$

$$
\begin{aligned}
&= \rho \sum_{s'\in\mathbf{S}} P(s'|s,a^*)\left(V_\tau(s') - V'_\tau(s')\right) && (31)\\
&\le 0 && (32)
\end{aligned}
$$

The inequality between (31) and (32) is satisfied because $0 < \rho < 1$, $P(s'|s,a^*) \geq 0$, and $V_\tau(s') - V'_\tau(s') \leq 0$. Therefore, if $V_\tau(s) \leq V'_\tau(s)$, then $(\mathcal{T}^*V_\tau)(s) \leq (\mathcal{T}^*V'_\tau)(s)$. □

Property 8. *(Additivity of Bellman Optimality Operator)* $(\mathcal{T}^*V_\tau + d)(s) = (\mathcal{T}^*V_\tau)(s) + \rho d, \forall s \in \mathbf{S}_i$

Proof.

$$
\begin{aligned}
&(\mathcal{T}^*V_\tau + d)(s)\\
&\quad= \max_{a\in\mathbf{A}} \left(\sum_{s'\in\mathbf{S}} P(s'|s,a)\left(U(s,a,s') + \rho(V_\tau(s') + d)\right) \right)\\
&\quad= \max_{a\in\mathbf{A}} \left(\sum_{s'\in\mathbf{S}} P(s'|s,a)\left(U(s,a,s') + \rho V_\tau(s')\right) \right. && (33)\\
&\qquad \left. + \rho d \sum_{s'\in\mathbf{S}} P(s'|s,a) \right) && (34)\\
&\quad= (\mathcal{T}^*V_\tau)(s) + \rho d
\end{aligned}
$$

where $\sum_{s'\in\mathbf{S}} P(s'|s,a) = 1$ in (34). Therefore, $(\mathcal{T}^*V_\tau + d)(s) = (\mathcal{T}^*V_\tau)(s) + \rho d, \forall s \in \mathbf{S}$. □

Property 9. *(ρ-Contraction Property of Bellman Optimality Operator)* $||\mathcal{T}^*V_\tau(s) - \mathcal{T}^*V'_\tau(s)||_\infty \leq \rho||V_\tau(s) - V'_\tau(s)||_\infty, \forall s \in \mathbf{S}$

Proof. We define d as,

$$d = ||V_\tau(s) - V'_\tau(s)||_\infty, \tag{35}$$

where $||\cdot||_\infty$ denotes the infinite norm, defined as

$$||V_\tau(s)||_\infty = \sup\left\{|V_\tau(s)| : s \in \mathbf{S}\right\}. \tag{36}$$

Then the following equations can be obtained from (35).

$$
\begin{aligned}
&V_\tau(s) - d \leq V'_\tau(s) \leq V_\tau(s) + d && (37)\\
&\mathcal{T}^*(V_\tau(s) - d) \leq (\mathcal{T}^*V'_\tau)(s) \leq \mathcal{T}^*\left(V_\tau(s) + d\right) && (38)\\
&\mathcal{T}^*V_\tau(s) - \rho d \leq (\mathcal{T}^*V'_\tau)(s) \leq \mathcal{T}^*V_\tau(s) + \rho d && (39)\\
&||\mathcal{T}^*V_\tau(s) - \mathcal{T}^*V'_\tau(s)||_\infty \leq \rho d && (40)
\end{aligned}
$$

The Bellman optimality operator is used between (37) and (38), and Property 8 is used between (38) and (39). By substituting d in (40) for (35), we conclude the following equation.

$$||\mathcal{T}^*V_\tau(s) - \mathcal{T}^*V'_\tau(s)||_\infty \leq \rho||V_\tau(s) - V'_\tau(s)||_\infty$$

□

Appendix B
Proof of Theorem 5

In this proof, we show that for all $s \in \mathbf{S}$,

$$||V_\tau(s) - V_{\tau-1}(s)||_\infty \leq \frac{1-\rho}{2\rho}\epsilon \Rightarrow ||V^{\epsilon^*}(s) - V^*(s)||_\infty < \epsilon. \tag{41}$$

By using the definition of infinite norm, $||V^{\epsilon^*}(s) - V^*(s)||_\infty$ can be written as follow.

$$
\begin{aligned}
&||V^{\epsilon^*}(s) - V^*(s)||_\infty && (42)\\
&\quad\leq ||V^{\epsilon^*}(s) - V_\tau(s)||_\infty + ||V_\tau(s) - V^*(s)||_\infty && (43)
\end{aligned}
$$

We now bound each part of the summation in (43) individually:

$$
\begin{aligned}
&||V^{\epsilon^*}(s) - V_\tau(s)||_\infty && (44)\\
&= ||\mathcal{T}^{\epsilon^*}V^{\epsilon^*}(s) - V_\tau(s)||_\infty && (45)\\
&\leq ||\mathcal{T}^{\epsilon^*}V^{\epsilon^*}(s) - \mathcal{T}^*V_\tau(s)||_\infty + ||\mathcal{T}^*V_\tau(s) - V_\tau(s)||_\infty && (46)\\
&= ||\mathcal{T}^{\epsilon^*}V^{\epsilon^*}(s) - \mathcal{T}^{\epsilon^*}V_\tau(s)||_\infty + ||\mathcal{T}^*V_\tau(s) - \mathcal{T}^*V_{\tau-1}(s)||_\infty && (47)\\
&\leq \rho||V^{\epsilon^*}(s) - V_\tau(s)||_\infty + \rho||V_\tau(s) - V_{\tau-1}(s)||_\infty && (48)\\
&\leq \frac{\rho}{1-\rho}||V_\tau(s) - V_{\tau-1}(s)||_\infty
\end{aligned}
$$

where $\mathcal{T}^{\epsilon^*}$ denotes the Bellman ϵ-optimality operation and it satisfies $V^{\epsilon^*}(s) = \mathcal{T}^{\epsilon^*}V^{\epsilon^*}(s)$ because $V^{\epsilon^*}(s)$ is the fixed point of $\mathcal{T}^{\epsilon^*}$, which is used in (45). The inequality between (45) and (46) is obtained by using the definition of infinite norm. Since π^{ϵ^*} is maximized over the actions using $V_\tau(s)$ in (47), this implies that $\mathcal{T}^{\epsilon^*}V_\tau(s) = \mathcal{T}^*V_\tau(s)$. The inequality between (47) and (48) is based on Property 9.

Similarly, the second part of the summation in (43) becomes

$$
\begin{aligned}
&||V_\tau(s) - V^*(s)||_\infty && (49)\\
&\quad\leq ||V_\tau(s) - \mathcal{T}^*V_\tau(s)||_\infty + ||\mathcal{T}^*V_\tau(s) - V^*(s)||_\infty\\
&\quad\leq \rho||V_{\tau-1}(s) - V_\tau(s)||_\infty + ||V_\tau(s) - V^*(s)||_\infty\\
&\quad\leq \frac{\rho}{1-\rho}||V_\tau(s) - V_{\tau-1}(s)||_\infty. && (50)
\end{aligned}
$$

By substituting (48) and (50) in (43) and using the condition $||V_\tau - V_{\tau-1}||_\infty \leq \frac{1-\rho}{2\rho}\epsilon$ in (41), we conclude the following.

$$
\begin{aligned}
||V^{\epsilon^*}(s) - V^*(s)||_\infty &\leq \frac{2\rho}{1-\rho}||V_\tau(s) - V_{\tau-1}(s)||_\infty\\
&< \frac{2\rho}{1-\rho}\frac{1-\rho}{2\rho}\epsilon\\
&= \epsilon
\end{aligned}
$$

Therefore, for all $s \in \mathbf{S}$, (41) is satisfied.

Appendix C
Proof of Theorem 6

In this proof, we show that Algorithm 1 converges to the optimal policies π^* by showing that the infinite interactions of the Bellman optimality operation converge to the optimal state-value function $V^*(s)$, such as

$$\lim_{\tau\to\infty} V_\tau(s) = V^*(s), \forall s \in \mathbf{S}. \tag{51}$$

This is identical to the following equation based on the definition of the infinite norm in (36).

$$\lim_{\tau\to\infty} ||V_\tau(s) - V^*(s)||_\infty = 0, \forall s \in \mathbf{S} \tag{52}$$

Hence, in this proof, we prove (52) as below.

$$\lim_{\tau\to\infty} ||V_\tau(s) - V^*(s)||_\infty \quad (53)$$

$$= \lim_{\tau\to\infty} ||\mathcal{T}^* V_{\tau-1}(s) - \mathcal{T}^* V^*(s)||_\infty \quad (54)$$

$$\le \lim_{\tau\to\infty} \rho ||V_{\tau-2}(s) - V^*(s)||_\infty \quad (55)$$

$$\le \cdots \le \lim_{\tau\to\infty} \rho^\tau ||V_0(s) - V^*(s)||_\infty \quad (56)$$

$$= 0$$

Based on the definition of $V^*(s)$ in (17), $V^*(s) = \mathcal{T}^* V^*(s)$ is used in (54), and the inequality between (54) and (55) is based on Property 9. Since $0 < \rho < 1$, $\lim_{\rho\to\infty} \rho^\tau = 0$ in (56). Therefore, $\lim_{\tau\to\infty} ||V_\tau(s) - V^*(s)||_\infty = 0, \forall s \in \mathbf{S}$ so that $\lim_{\tau\to\infty} V_\tau(s) = V^*(s), \forall s_{i,\tau} \in \mathbf{S}$.

References

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[2] C.-C. Wang and N. B. Shroff, "On wireless network scheduling with intersession network coding," in *IEEE Annual Conference on Information Sciences and Systems*, 2008, pp. 30–35.

[3] Y. Kim and G. D. Veciana, "Is rate adaptation beneficial for inter-session network coding?" *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 5, pp. 635–646, Jun. 2009.

[4] A. Khreishah, C. C. Wang, and N. B. Shroff, "Rate control with pairwise intersession network coding," *IEEE/ACM Transactions on Networking*, vol. 18, no. 3, pp. 816–829, Jun. 2010.

[5] E. Bourtsoulatze, N. Thomos, and P. Frossard, "Decoding delay minimization in inter-session network coding," *IEEE Transactions on Communications*, vol. 62, no. 6, pp. 1944–1957, Jun. 2014.

[6] ——, "Distributed rate allocation in inter-session network coding," *IEEE Transactions on Multimedia*, vol. 16, no. 6, pp. 1752–1765, Oct. 2014.

[7] Hulya and Athina, "Distributed rate control for video streaming over wireless networks with intersession network coding," in *2009 17th International Packet Video Workshop*, May 2009, pp. 1–10.

[8] A. Douik, S. Sorour, T. Y. Al-Naffouri, and M. S. Alouini, "Decoding delay controlled completion time reduction in instantly decodable network coding," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2016.

[9] "Virtual overhearing: An effective way to increase network coding opportunities in wireless ad-hoc networks," *Computer Networks*, vol. 105, pp. 111 – 123, 2016.

[10] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[11] T. Ho, R. Koetter, M. Médard, D. Karger, and M. Effros, "The benefits of coding over routing in a randomized setting," in *IEEE International Symposium on Information Theory*, Cambridge, MA, USA, Jun/Jul. 2003.

[12] T. Ho, M. Médard, J. Shi, M. Effros, and D. R. Karger, "On randomized network coding," in *Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Oct. 2003.

[13] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Annual Allerton Conference on Communication, Control, and Computing*, Monticell, IL, USA, Oct. 2003.

[14] R. Koetter and F. R. Kschischang, "Coding for errors and erasures in random network coding," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3579–3591, Aug 2008.

[15] Z. Zhang, "Linear network error correction codes in packet networks," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 209–218, Jan 2008.

[16] M. Esmaeilzadeh, P. Sadeghi, and N. Aboutorab, "Random linear network coding for wireless layered video broadcast: General design methods for adaptive feedback-free transmission," *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 790–805, Feb 2017.

[17] K. Jain, M. Mahdian, and M. R. Salavatipour, "Packing steiner trees," in *ACM-SIAM symposium on Discrete algorithms*, 2003, pp. 266–274.

[18] Y. Cui, Y. Xue, and K. Nahrstedt, "Optimal distributed multicast routing using network coding: Theory and applications," *ACM SIGMETRICS Performance Evaluation Review*, vol. 32, no. 2, pp. 47–49, Sep. 2004.

[19] D. S. Lun, N. Ratnakar, R. Koetter, M. M/'edard, E. Ahmed, and H. Lee, "Achieving minimum-cost multicast: a decentralized approach based on network coding," in *IEEE Annual Joint Conference of the Computer and Communications Societies*, vol. 3, March 2005, pp. 1607–1617.

[20] X. Yan, J. Yang, and Z. Zhang, "An outer bound for multisource multisink network coding with minimum cost consideration," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2373–2385, Jun. 2006.

[21] D. Traskov, N. Ratnakar, D. S. Lun, R. Koetter, and M. Medard, "Network coding for multiple unicasts: An approach based on linear optimization," in *IEEE International Symposium on Information Theory*, July 2006, pp. 1758–1762.

[22] M. Kwon and H. Park, "Distributed network formation strategy for network coding based wireless networks," *IEEE Signal Processing Letters*, vol. 24, no. 4, pp. 432–436, April 2017.

[23] K. Sohrabi, J. Gao, V. Ailawadhi, and G. J. Pottie, "Protocols for self-organization of a wireless sensor network," *IEEE Personal Communications*, vol. 7, no. 5, pp. 16–27, 2000.

[24] T. Gao, F. Lang, and N. Guo, "An emergency communication system based on uav-assisted self-organizing network," in *International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, July 2016, pp. 90–95.

[25] M. Xu, Q. Yang, and K. S. Kwak, "Distributed topology control with lifetime extension based on non-cooperative game for wireless sensor networks," *IEEE Sensors Journal*, vol. 16, no. 9, pp. 3332–3342, 2016.

[26] J. G. Andrews, R. K. Ganti, M. Haenggi, N. Jindal, and S. Weber, "A primer on spatial modeling and analysis in wireless networks," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 156–163, November 2010.

[27] K. Huang and V. K. N. Lau, "Enabling wireless power transfer in cellular networks: Architecture, modeling and deployment," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 902–912, Feb. 2014.

[28] P. A. Chou and Y. Wu, "Network coding for the internet and wireless networks," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 77–85, Sep. 2007.

[29] S. Katti, H. Rahul, H. Wenjun, D. Katabi, M. Médard, and J. Crowcroft, "Xors in the air: practical wireless network coding," *IEEE/ACM Transactions on Networking*, vol. 16, no. 3, pp. 497 –510, Jun. 2008.

[30] T. Nad and A. Krishnamurthy, "Problems with network coding in overlay networks," *Techinical Report, Yale University*, 2004.

[31] H. Topakkaya, "Network coding for wireless and wired networks: Design, performance and achievable rates," Ph.D. dissertation, Iowa State University, 2011.

[32] J. Cloud, L. M. Zeger, and M. Medard, "Mac centered cooperation - synergistic design of network coding, multi-packet reception, and improved fairness to increase network throughput," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 2, pp. 341–349, Feb. 2012.

[33] S. M. Mirrezaei, M. Dosaranian-Moghadam, and M. Yazdanpanahei, "Effect of network coding and multi-packet reception on point-to-multi-point broadcast networks," *Wireless Personal Communications*, vol. 79, no. 3, pp. 1859–1891, 2014.

[34] M. Médard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413–4430, Oct 2006.

[35] S. Katti, D. Katabi, W. Hu, H. Rahul, and M. M/'edard, "The importance of being opportunistic: Practical network coding for wireless environments," in *Allerton Annual Conference on Communication*, 2005.

[36] J. Liu, D. Goeckel, and D. Towsley, "Bounds on the gain of network coding and broadcasting in wireless networks," in *IEEE International Conference on Computer Communications*, May 2007, pp. 724–732.

[37] K.-J. Bathe and E. L. Wilson, "Numerical methods in finite element analysis," *Prentice-Hall Englewood Cliffs, NJ*, 1976.

[38] M. Kwon, H. Park, and P. Frossard, "Compressed network coding: Overcome all-or-nothing problem in finite fields," in *IEEE Wireless Communications and Networking Conference*, Apr. 2014, pp. 2851–2856.

[39] Z. Yan, H. Xie, and B. W. Suter, "Rank deficient decoding of linear network coding," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 5080–5084.

[40] M. Kwon, H. Park, N. Thomos, and P. Frossard, "Approximate decoding for network coded inter-dependent data," *Signal Processing*, vol. 120, pp. 222–235, Mar. 2016.

[41] M. Kwon and H. Park, "The impact of network coding cluster sizes on the approximate decoding performance," *KSII Transactions on Internet and Information Systems*, vol. 10, no. 3, pp. 1144–1158, Mar. 2016.

[42] T. W. Anderson, *The statistical analysis of time series*. John Wiley & Sons, 2011, vol. 19.

[43] H. Park and M. Van der Schaar, “On the impact of bounded rationality in peer-to-peer networks,” *IEEE Signal Processing Letters*, vol. 16, no. 8, pp. 675–678, 2009.

[44] ——, “A framework for foresighted resource reciprocation in p2p networks,” *IEEE Transactions on Multimedia*, vol. 11, no. 1, pp. 101–116, 2009.

[45] R. E. Bellman and S. E. Dreyfus, *Applied dynamic programming*. Princeton university press, 2015.

[46] S. U. Pillai, T. Suel, and S. Cha, “The perron-frobenius theorem: some of its applications,” *IEEE Signal Processing Magazine*, vol. 22, no. 2, pp. 62–75, 2005.

[47] J. L. Gross and J. Yellen, *Handbook of graph theory*. CRC press, 2004.

[48] “Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications,” *IEEE Std. 802.11-2013*, 2013.

[49] “802.11ac: The fifth generation of wi-fi - technical white paper,” *CISCO*, Mar. 2014.

[50] G. Miao, J. Zander, K. W. Sung, and S. B. Slimane, *Fundamentals of Mobile Data Networks*. Cambridge University Press, 2016.