Linear Network Coding Based Fast Data Synchronization for Wireless Ad Hoc Networks with Controlled Topology

Die Hu¹², Xuejun Zhu², Min Gong², Shaoshi Yang¹³,*

¹ School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
² China Academy of Launch Vehicle Technology, Beijing 100076, China
³ Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing 100876, China
* The corresponding author, email: shaoshi.yang@bupt.edu.cn

I. INTRODUCTION

Wireless ad hoc networks, such as the networked sensors, robots and unmanned aerial vehicles (UAVs), constitute a distributed, flexible and cooperative information-sharing system [1–3]. Fast data synchronization among network nodes is important for wireless ad hoc networks, since it is expected to provide essential information reliably for high-layer real-time decision and control algorithms. Unfortunately, in general this is a great challenge, because in many cases the network topologies and the wireless channels between nodes are highly dynamic and random [4, 5].

Owing to the openness of wireless channels, all-to-all broadcast has the potential to serve as a highly efficient approach for achieving fast data synchronization in wireless ad hoc networks. However, the transmission efficiency of all-to-all broadcast is still constrained by the conventional store-and-forward protocol, which does not take advantage of more sophisticated data processing and thus limits the distributed system’s efficiency of sensing and reacting to the environment. The network coding (NC) technique [6, 7] offers an attractive approach for data synchronization in distributed systems, since it is capable of reducing the total time cost of data synchronization by exploiting both the broadcast property of wireless channels and the XOR operation in NC, thus increasing the effective data transmission rate [8, 9].

Regarding the related work, in [10] the authors proposed a proactive NC scheme for all-to-all broadcast while using the random access schemes of IEEE
802.11 and considering several different network topologies. In [11] an all-to-all broadcast protocol was designed for wireless ad hoc networks that use directional antennas, but NC was not considered. In [12] the reliability of a random neighbor NC based all-to-all broadcast approach was analyzed. In [13] an adaptive transmission protocol suite integrated with the random linear NC was proposed, where the modulation and channel coding parameters are adaptively selected for each packet to improve the packet loss performance.

However, all the above contributions are designed for general ad hoc networks, while ignoring the unique properties of specific systems. There indeed exist some particular scenarios, where the original distributed data synchronization problem can become less challenging or enjoy more benefits. For instance, compared with the data integrity and security, the latency requirement is less stringent for the data synchronization in distributed databases [14, 15]. Additionally, more benefits can be gleaned in the scenario where the network topology is under the operator’s control. Topology control is important and practical for wireless ad hoc networks, as it is beneficial for reducing energy consumption (thus extending the network lifetime) and radio interference (thus increasing the network communication capacity) [16–18]. By using topology control, a more stable and convenient graph representation of the network can be obtained. Wireless ad hoc networks with controlled topology have found applications in many areas, such as the flight formation in military operations, the truck platooning in Internet of Vehicles (IoV) as highlighted by 5G, and the low earth orbit (LEO) satellite constellations etc.

In this paper we propose a pair of linear NC and all-to-all broadcast based fast data synchronization algorithms for wireless ad hoc networks with controlled topology. We study the performance of the proposed algorithms with a large number of randomly generated network topology samples. It is shown that the average gain in terms of the time-slot usage reduction upon applying the proposed algorithms can be over five times compared with the time-division multiple-access (TDMA) based all-to-all broadcast algorithm that does not use NC. Furthermore, this substantial gain is achievable in a wide range of network topologies, in particular for the topologies that have a low or medium degree of network connectivity.

II. SYSTEM MODEL

As shown in Figure 1, we consider an ad-hoc network that consists of \( N \) nodes connected in any arbitrary topology. We assume that for each node that carries out linear NC, the corresponding decoding is performed by its immediate neighbouring nodes. For each hop, a single action of transmitting-and-receiving occupies one time slot. The data synchronization throughout the network is achieved as follows. The node \( n \) broadcasts its own data block \( p_n \), whose original or processed copy is then disseminated to the other \( N - 1 \) nodes by relaying through their respective neighbouring nodes, where \( n = 1, 2, \ldots, N \).

The number of neighbouring nodes of node \( n \) and the number of data blocks stored at node \( n \) are denoted as \( m_n \) and \( d_n \), respectively. It is required that all the nodes in the network obtain the set of data blocks \( A = \{p_1, p_2, \ldots, p_N\} \) as fast as possible, which means that \( d_n \) must equal \( N \) after the data synchronization is achieved throughout the network, i.e., \( \max \{d_n\} = N, \forall n = 1, 2, \ldots, N \).

In a given time slot the node \( n \) shares its data blocks with its neighbouring nodes \( n_k \), where \( 1 \leq k \leq m_n, 1 \leq n_k \leq N \) and \( n_k \neq n \). To this end, we define the following sets:

- \( A_n = \{p_1, p_2, \ldots, p_{d_n}\} \): The data blocks stored at node \( n \).
- \( \tilde{A}_{n_k} = \{p_j | p_j \in A_n, p_j \notin A_{n_k}, j = 1, 2, \ldots, N\} \): The data blocks that node \( n_k \) has not obtained.

Then \( \tilde{A}_{n_k} \cap A_n \) represents the data blocks that node \( n_k \) can obtain from node \( n \).

III. THE PROPOSED DATA SYNCHRONIZATION ALGORITHMS

Below we describe the proposed data synchronization algorithms in detail, whose flowchart is shown in Figure 2.

**Step 0:** This is the initialization stage, where each node carries out data acquisition by monitoring its own state or the state of the environment. Note that it is possible for a single node to have multiple data blocks.

**Step 1:** Each node broadcasts the data blocks, which are stored on it and encapsulated into packets, to its neighbours in a TDMA manner.
by examining if chronized in all nodes. This is achieved at each node block stored at node \( n \) the format of data blocks, and each node updates its data block storage status.

Otherwise, the packet is directly stored on the node in the initial stage. If the packet is not received correctly, each node updates its data block storage status.

![Figure 1. The system model of the wireless ad hoc network considered, where the shaded rectangle \( p_n \) denotes the data block stored at node \( n \) in the initial stage.](image)

**Step 2:** If a packet is received correctly and network-coded, it is fed into the network-decoding module deployed on the node that receives the packet, and the data blocks extracted from the decoder are stored on the node. If the packet is not received correctly, each node updates its data block storage status. Otherwise, the packet is directly stored on the node.

**Step 3:** Check whether all the data blocks are synchronized in all nodes. This is achieved at each node by examining if \( d_n = N \). If yes, the algorithms terminate. Otherwise, select an appropriate transmitting node that contributes on average the maximum innovative data to each neighbouring node according to (optional, used in Alg. 2, by examining the feedback from the subsequent data block selection (DBS), as elaborated in Sec. 3.3), and an appropriate set of data blocks that are innovative and decodable to the maximum number of the neighbouring nodes of the selected node according to (mandatory, used in both Alg. 1 and Alg. 2, by exploiting the output of the decoder in Step 2, as elaborated in Sec. 3.2). These selected data blocks are stored in their host node to participate in the subsequent network-encoding operation.

**Step 4:** Carry out network-encoding operation with respect to the data blocks selected, and jump to Step 1, where the above selected node broadcasts the network-coded data blocks to their respective neighbours. This process is repeated until all the data blocks on all the nodes are synchronized.

Note that the DBS results are fed back to serve as an input for the node selection (NS) module, according to the outputs of the decoding module introduced in Sec. 3.1.

### 3.1 Decoding

For each TX-RX pair, the receiver is anticipated to decode the packets sent by the transmitter and then keep the data blocks that are absent from this receiver previously. In other words, each receiver is only interested in the data blocks that are innovative to itself. To clarify our design philosophy, the following theorem is given.

**Theorem 1.** For packets generated by linear NC, the receiver is capable of decoding the packet if at most one component data block is unknown.

**Proof.** Let us use the mathematical induction method to prove the theorem. Firstly, we have the following proposition: the component data block \( x \) can be extracted from packet \( P = x \oplus a_1 \oplus a_2 \oplus \ldots \oplus a_l \oplus a_{l+1} \) provided that all the component data blocks \( a_l, l = 1, 2, \ldots, L \) are known, where \( L \in \mathbb{Z}_+ \) is a positive integer.

1) When \( L = 1 \), the proposition is obviously true.

2) Assume that the above proposition is true when \( L = r \), with \( r \in \mathbb{Z}_+ \) being an arbitrarily given positive integer. Then \( x \) can be solved from \( P = x \oplus a_1 \oplus a_2 \oplus \ldots \oplus a_l \oplus a_r \) if and only if \( a_l \) is known, \( \forall l \in \{1, 2, \ldots, r\} \).

3) When \( L = r + 1 \), the packet \( P \) is written as

\[
P = x \oplus a_1 \oplus a_2 \oplus \ldots \oplus a_l \oplus a_r \oplus a_{r+1}.
\]

Then if \( \forall l \in \{1, 2, \ldots, r+1\} \), where \( a_l \) is unknown, \( x \) is obviously unsolvable. Let us assume that at least one \( a_l \) is known. For convenience, assume that \( a_{r+1} \) is known. Then Eq. (1) is rewritten as

\[
P = P' \oplus a_{r+1},
\]

where \( P' = x \oplus a_1 \oplus a_2 \oplus \ldots \oplus a_r \) and it is decodable. Furthermore, to decode the data block \( x \), \( \forall a_l, l = 1, 2, \ldots, r \) must be known because of the above statement 2).

Thus the proof has been established.

Therefore, upon receiving the packet \( P \), the node \( n_k \)
is able to obtain innovative data block from node $n$ if
$$|\tilde{A}_n \cap \tilde{A}_{n_k}| = 1. \tag{3}$$

### 3.2 The DBS Operation and Encoding

To maximize the gain of all-to-all broadcast, the network encoding operation is expected to select the particular data blocks that enable as many neighbouring receivers as possible to decode the packet transmitted and each of these receivers must obtain innovative data block from the decoding. In other words, the network encoding should operate on the data blocks that are the solutions to the following optimization problem

$$\max_{\tilde{A}_n \subseteq A_n} \beta_n, \tag{4}$$

where $\tilde{A}_n$ denotes the set of data blocks selected on the node $n$ and $\beta_n$ is defined as the particular number of the neighbouring nodes that can decode and obtain innovative data from node $n$:

$$\beta_n = \sum_j \varepsilon_{nj}. \tag{5}$$

More specifically, $\varepsilon_{nj} = \begin{cases} 1 & |\tilde{A}_n \cap \tilde{A}_{nj}| = 1 \\ 0 & \text{otherwise} \end{cases}$, and $j$ represents the index of receiving nodes. (4) is a combinatorial optimization problem, which can be solved by any established search method $f(\cdot)$ over all the subsets of $A_n$.

To be more efficient, the node $n$ can select data blocks to transmit from the set $\mathcal{B} = (\tilde{A}_{n_1} \cap A_n) \cup (\tilde{A}_{n_2} \cap A_n) \cup \cdots \cup (\tilde{A}_{n_m} \cap A_n)$ with the specific approach $f(\cdot)$ introduced above$^1$, and encodes these component data blocks into packet $P$ with the aid of the bitwise XOR operation, i.e., $P = p_{d_1} \oplus p_{d_2} \oplus \cdots \oplus p_{d_j}$, where $\oplus$ stands for bitwise XOR, $1 \leq d_j \leq N$, $1 \leq j \leq |\tilde{A}_n|$ and the selected data blocks $p_{d_j} \in \tilde{A}_n = f(\mathcal{B})$.

### 3.3 The NS Operation

For a given network topology, it is important to determine which node should broadcast its data to the neighbours at a particular time instant. In other words, it is necessary to carry out NS, in addition to DBS.

Thus, based on $\beta_n$ we further consider the ratio of the neighbouring nodes that can be helped among all the neighbouring nodes of the selected node per broadcast or time slot. Then, the node is selected according to

$$\max_i \beta_i / m_i, \tag{6}$$

where $i = 1, 2, \cdots, N$. By solving (6), the node that enables the largest proportion of neighbour nodes to be capable of obtaining innovative data blocks is selected, thus the transmission efficiency is improved.

### IV. Simulation Results and Discussions

In this section, the performance of the proposed all-to-all broadcast based distributed data synchronization algorithms are demonstrated with Monte Carlo simulations. We consider a wireless ad hoc network, where the number of nodes is from 5 to 11. For each net-
Comparisons are made among three data synchronization schemes:

1. Uncoded+DBS (baseline, abbreviated as U-DBS): The data synchronization is performed without using NC. In a single transmission cycle, each node broadcasts its individual data in turn following a given order, and each node uses a single time slot. DBS is also invoked by each node in this scheme, using the criterion similar to (4), where $\tilde{A}_n$ changes from a multi-element set to a single-element set. In other words, each node broadcasts the data block that is innovative to the largest number of receivers. This process is repeated in the next cycle, until the data synchronization across the network is achieved using the conventional store-and-forward protocol.

2. Coded+DBS (Alg. 1, abbreviated as C-DBS): The data synchronization is achieved by employing the DBS based linear NC, as presented in Sec. 3.2, while NS is not invoked. The transmission protocol can now be termed as compute-and-forward. The remaining operations are the same as the scheme of U-DBS.

3. Coded+DBS+NS (Alg. 2, abbreviated as C-DBS-NS): In this scheme, not only the DBS based linear NC, but also the NS as introduced in Sec. 3.3, is invoked. Thus, in each time slot, only the node selected broadcasts its data. This process is repeated until the data synchronization across the network is achieved. As such, the concept of transmission cycle as used in the schemes of U-DBS and C-DBS becomes invalid, and the transmission protocol is also the compute-and-forward.

Firstly, by using the concept of average degree as defined in graph theory for characterizing the nodes’ and network’s connectivity, in Figure 3 we evaluate the average number of time slots used by the three schemes considered, under different values of network degree, network size and packet error rate $P_e$ of a single hop. The average number of time slots $S_d$ under a particular average degree $d$ is defined as

$$S_d = \frac{\sum_i s_{id}}{N_d},$$

where $s_{id}$ denotes the number of time slots used in the data synchronization process for the $i$th network sample under the average degree $d$, and $N_d$ denotes the number of randomly generated network samples, which essentially also represent different network topologies.

We can observe that the proposed C-DBS-NS scheme (the red curves) and C-DBS scheme (the blue curves) both enjoy a significant time slot usage reduction, compared to the U-DBS scheme (the green curves) in the weakly or moderately connected networks. Additionally, the C-DBS-NS scheme performs best among the three schemes, especially under the moderately connected network topology, but the ad-

**Figure 3.** The average number of time slots used by different schemes versus the average degree of a wireless ad hoc network.
vantage of the C-DBS-NS scheme over the C-DBS scheme is limited. Furthermore, the time slot usage increases when we have a larger $P_e$ and a larger network size.

Secondly, the average relative processing gain (RPG) of the proposed two schemes to the baseline U-DBS scheme under different values of network degree, network size and packet error rate $P_e$ of a single hop is depicted in Figure 4. Here the average RPG is defined as

$$G_d = \frac{\sum_{i=1}^{N_d} \tilde{s}_{id}/\tilde{s}_{td}}{N_d},$$

(8)

where $\tilde{s}_{td}$ and $\tilde{s}_{id}$ respectively denote the number of time slots used in the data synchronization process of the U-DBS scheme and of either the proposed C-DBS or C-DBS-NS schemes for the $i$th network sample under the average degree $d$. We observe that for either large or small average degree, the average RPG of the C-DBS scheme is similar to that of the C-DBS-NS scheme, which is a common phenomenon under different values of network size. For medium average degree, the average RPG of the C-DBS-NS scheme is significantly higher than that of the C-DBS scheme. In other words, for poorly connected networks, the C-DBS scheme has a similar performance compared to the most complicated C-DBS-NS scheme; for well connected networks, the data synchronization performance (i.e., synchronization speed) of all the three schemes, including the simplest U-DBS scheme, tends to be identical; while for moderately connected networks, the C-DBS-NS scheme is shown to have the highest synchronization speed.

Finally, in Figure 5 we compare the computational complexity of the three schemes in terms of the average number of FLOPS\(^3\) under different numbers of nodes. We see that as the number of nodes rises, the average number of FLOPS of the C-DBS-NS scheme grows much faster than that of the U-DBS scheme, while the C-DBS scheme has a moderately higher computational complexity than the U-DBS scheme. This implies that when choosing appropriate data synchronization algorithm for large-scale wireless ad hoc networks, both the time slot usage and the computational complexity of the algorithm should be considered. It is advised to use the divide-and-conquer approach to reduce the network size to a moderate value.
V. CONCLUSION

In this paper, we have proposed a pair of linear NC and all-to-all broadcast based fast data synchronization algorithms for wireless ad hoc networks that have controlled topology. For better exploiting the benefits of NC, the first algorithm C-DBS exploits data block selection, while the second algorithm C-DBS-NS exploits both the transmitting node selection and data block selection. Thus, compared with the conventional uncoded approach that uses store-and-forward protocol, a more efficient compute-and-forward protocol is used in our algorithms. We show that C-DBS-NS performs best among the three schemes in the moderately connected networks, while in the weakly connected networks, C-DBS achieves similarly good performance at a lower computational complexity compared to C-DBS-NS. For well connected networks, however, the three schemes have almost identical performance, hence there is no need to perform NC and the simplest U-DBS approach suffices.

NOTES

1 In other words, $\tilde{A}_n \subseteq A_n$ in (4) can be replaced with $\tilde{A}_n \subseteq B$, resulting in a smaller search space.

2 As the average degree increases, the network connection state is improved.

3 Due to the randomness of the topology and the network size, it is difficult to give analytic results of the computational complexity of the three schemes.

ACKNOWLEDGEMENT

This work is financially supported by Beijing Municipal Natural Science Foundation (No. L202012), the Open Research Project of the State Key Laboratory of Media Convergence and Communication, Communication University of China (No. SKLMMC2020KF008), and the Fundamental Research Funds for the Central Universities (No. 2020RC05).

References

[1] MENG X, WANG W, LEONG B. SkyStitch: A cooperative multi-UAV-based real-time video surveillance system with stitching[C]//Proc. 23rd ACM International Conference on Multimedia (MM). New York, USA: ACM, 2015: 261-270.
[2] BALASINGAM B, PATTIPATI K, LEVCHUCK G, et al. Robust collaborative learning by multi-agents[C]//Proc. IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA). Verona, USA: IEEE, 2015: 1-5.
[3] FEI Z, LI B, YANG S, et al. A survey of multi-objective optimization in wireless sensor networks: Metrics, algorithms, and open problems[J]. IEEE Communications Surveys & Tutorials, 2017, 19(1):550-586.
[4] XIE X, HUANG B, YANG S, et al. Adaptive multi-channel MAC protocol for dense VANET with directional antennas[C]//Proc. 6th IEEE Consumer Communications and Networking Conference. Las Vegas, USA: IEEE, 2009: 1-5.
[5] LV T, GAO H, LI X, et al. Space-time hierarchical-graph based cooperative localization in wireless sensor networks[J]. IEEE Transactions on Signal Processing, 2016, 64(2):322-334.
[6] AHLSWEDE R, CAI N, LI S Y R, et al. Network information flow[J]. IEEE Transactions on Information Theory, 2000, 46(4):1204-1216.
[7] BASSOLI R, MARQUES H, RODRIGUEZ J, et al. Network coding theory: A survey[J]. IEEE Communications Surveys & Tutorials, 2013, 15(4):1950-1978.
[8] KATTI S, RAHUL H, HU W, et al. XORs in the air: Practical wireless network coding[J]. IEEE/ACM Transactions on Networking, 2008, 16(3):497-510.
[9] CAO R, LV T, GAO H, et al. Achieving full diversity in multi-antenna two-way relay networks via symbol-based physical-layer network coding[J]. IEEE Transactions on Wireless Communications, 2013, 12(7):3445-3457.
[10] ASTERJADHI A, FASOLO E, ROSSI M, et al. Toward network coding-based protocols for data broadcasting in wireless ad hoc networks[J]. IEEE Transactions on Wireless Communications, 2010, 9(2):662-673.
[11] DUAN P, PENG L, XU R, et al. An all-to-all broadcasting protocol using directional antennas in multi-hop wireless networks[C]//Proc. International Conference on Wireless Commu-
[12] MA L, LIN Z, ZHANG Z, et al. Reliability of all-to-all broadcast with network coding[C]// Proc. IEEE Global Communications Conference (GLOBECOM). Atlanta, USA: IEEE, 2013: 1991-1996.

[13] BORKOTOKY S S, PURSLEY M B. Adaptive transmission for network-coded all-to-all broadcast in multi-hop ad hoc wireless networks[J]. IEEE Communications Letters, 2019, 23(12): 2412-2416.

[14] CHOI M, CHO E, PARK D, et al. A synchronization algorithm of mobile database for ubiquitous computing[C]//Proc. 5th International Joint Conference on INC, IMS, and IDC. Seoul, Korea: IEEE, 2009: 416-419.

[15] SON S H. Replicated data management in distributed database systems[J]. SIGMOD Record, 1988, 17(4):62-69.

[16] PAOLO S. Topology control in wireless ad hoc and sensor networks[J]. ACM Computing Surveys, 2005, 37(2):164-194.

[17] CHEN B, JAMIESON K, BALAKRISHNAN H, et al. Span: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks[J]. Wireless Networks, 2002, 8(5):481-494.

[18] RAJMOHAN R. Topology control and routing in ad hoc networks: A survey[J]. SIGACT News, 2002, 33(2):60-73.

Biographies

Die Hu received the B.Eng degree in communication engineering from Wuhan University, China, in Jul. 2018, and the M.Eng degree in system design of flight vehicle from China Academy of Launch Vehicle Technology in Jun. 2021. She was also a visiting student at Beijing University of Posts and Telecommunications from Jan. 2020 to Apr. 2021. She is currently a research engineer at China Academy of Launch Vehicle Technology. Her research interests include distributed information system, flying ad hoc networks and routing protocols.

Xuejun Zhu received the B.Eng degree in automatic control from National University of Defense Technology, China, in 1984, and the M.Eng degree in aircraft navigation and control from China Academy of Launch Vehicle Technology in 1987. She is currently a chief designer and professorial research fellow with China Academy of Launch Vehicle Technology. In 2019, she was elected to the Academician of Chinese Academy of Sciences. Her research interests include networking technologies, flight vehicle design and electrical system design.

Min Gong received the B.Eng degree in information engineering from Beijing Institute of Technology, China, in 2005, and the PhD degree in information and communication engineering from Tsinghua University, China, in 2010. He is currently a chief designer with China Academy of Launch Vehicle Technology. His research interests include mobile ad hoc networks, distributed information processing, flight vehicle design and game theory.
Shaoshi Yang received the B.Eng degree in information engineering from Beijing University of Posts and Telecommunications (BUPT), China, in 2006, and the PhD degree in electronics and electrical engineering from University of Southampton, U.K., in 2013. From 2008 to 2009, he was a researcher of WiMAX standardization with Intel Labs China. From 2013 to 2016, he was a Research Fellow with the School of Electronics and Computer Science, University of Southampton. From 2016 to 2018, he was a Principal Engineer with Huawei Technologies Co. Ltd., where he made significant contributions to the company’s products and solutions associated with 5G base stations, wideband IoT, and cloud gaming/VR. He is currently a Full Professor with BUPT. His research expertise includes 5G wireless networks, massive MIMO, iterative detection and decoding, mobile ad hoc networks, distributed artificial intelligence, and cloud gaming/VR. He is a member of the Isaac Newton Institute for Mathematical Sciences, Cambridge University, and a Senior Member of IEEE. He received the Dean’s Award for Early Career Research Excellence from University of Southampton in 2015, the Huawei President Award of Wireless Innovations in 2018, the IEEE Technical Committee on Green Communications & Computing (TCGCC) Best Journal Paper Award in 2019, and the IEEE Communications Society Best Survey Paper Award in 2020. He is an Editor for IEEE Systems Journal, IEEE Wireless Communications Letters, and Signal Processing (Elsevier). He was also an invited international reviewer of the Austrian Science Fund (FWF).