Hindawi
Journal of Electrical and Computer Engineering
Volume 2017, Article ID 4758471, 7 pages
https://doi.org/10.1155/2017/4758471

Research Article

Secure-Network-Coding-Based File Sharing via Device-to-Device Communication

Lei Wang[1] and Qing Wang[2]

[1] *School of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210023, China*
[2] *School of Tongda, Nanjing University of Posts and Telecommunications, Yangzhou 225127, China*

Correspondence should be addressed to Lei Wang; leiwang@njupt.edu.cn

Received 18 March 2017; Revised 2 May 2017; Accepted 16 May 2017; Published 8 June 2017

Academic Editor: Arun K. Sangaiah

In order to increase the efficiency and security of file sharing in the next-generation networks, this paper proposes a large scale file sharing scheme based on secure network coding via device-to-device (D2D) communication. In our scheme, when a user needs to share data with others in the same area, the source node and all the intermediate nodes need to perform secure network coding operation before forwarding the received data. This process continues until all the mobile devices in the networks successfully recover the original file. The experimental results show that secure network coding is very feasible and suitable for such file sharing. Moreover, the sharing efficiency and security outperform traditional replication-based sharing scheme.

1. Introduction

Sharing large scale files such as high-resolution videos with many friends through mobile devices at the same time is becoming a popular application. Smart phones are always used to upload and download the shared files through WiFi, 3G, or LTE, but these ways will naturally incur high expense and security threat when large scale data needs to be shared. Actually, it is unnecessary to share data via commercial networks in some scenarios. If the devices are located in the same area, the users could share the files through direct link between devices so that the traffic fee could be saved. The mobile devices can be strategically switched to soft AP (Access Point) mode so that the other users could connect to it and receive the files. However, there are two constraints in this method. First, from the technical perspective, the number of users connecting a soft AP is often limited from four to eight. Second, some users cannot connect to the soft AP within one hop. Therefore, after the users close to the source receive the files, they are supposed to share the data with their neighbors by switching to a new soft AP. Through the sharing of multiple hops, all the users in the network could obtain the files.

When sharing files with many users, sharing efficiency and security should be focused. When a large scale data needs to be shared, it would be better to split the original file into multiple slices before sharing because the direct link between devices may be disconnected during the transmission. After splitting the files into multiple slices, as long as the other users receive all the slices, the original file could be recovered. However, this method has a drawback which could be optimized. When a node requires a specific block of the original file, its neighbors may not have it either. Therefore, they have to wait until the slice is received. In order to overcome this drawback, network coding [1] could be introduced in such applications. Network coding has been considered as a promising technology in big data transmission. Network coding has been proposed for more than ten years, and it has attracted much attention of researchers. Li et al. [2] proposed linear network coding, and then Ho et al. [3] and Jaggi et al. [4] proposed RLNC (Random Linear Network Coding) and DLNC (Deterministic Linear Network Coding), respectively. Network coding has been studied in many areas, such as information security [5], distributed storage [6], video communication [7], and content sharing [8].

The main feature of network coding is that it requires reencoding operation at the intermediate devices of the network. Benefiting from the reencoding operations, the network performance could be increased such as bandwidth and energy efficiency. Moreover, the data is highly mixed at the source node and intermediate nodes, which means that the data transmitted in the channel is no longer the original data. Therefore, the security is significantly increased. However, it is very difficult to change the traditional network architecture, which impedes the development of network coding because traditional intermediate devices such as routers and switches cannot perform additional computational operation. Currently, the development of mobile devices and 4G/5G networks makes the computational operation at the mobile devices feasible. Therefore, network-coding-based applications are becoming more and more popular in mobile devices [9, 10].

D2D communication is a key supporting technology for the fifth-generation communication network. In D2D communication, the mobile devices could communicate with others directly via physical links without the relay of the base station, and it is feasible to perform network coding operation at the devices. Therefore, 5G network is a perfect place to apply network coding. The aim of this paper is to model and analyze the sharing efficiency of large scale data in D2D communication when network coding is introduced.

The remainder of this paper is organized as follows. In Section 2, the authors introduce some closely related studies. In Section 3, the authors model the secure network-coding-based file sharing scheme. In Section 4, the authors evaluate the proposed scheme. Finally, the conclusion is made in Section 5.

2. Related Works

There are some existing papers closely related to this study. M. Yang and Y. Yang [11] proposed a network-coding-based file sharing scheme for peer-to-peer networks. They encode the original files and then deploy the encoded subfiles on a web server. All the clients not only download the encoded subfiles but also forward the encoded subfiles for each other. Their scheme achieves 15%–20% higher throughput than previous schemes, and it achieves good reliability and robustness to link failure. Their scheme shows that network coding is promising in the file sharing application on the Internet. Our research is for future wireless networks. Moreover, the network model in their study is abstracted as a combination network. Based on the network model, they proposed a deterministic algorithm to encode the files, while the network model in our system is based on RLNC.

Lin et al. [12] presented a stochastic analytical framework to study the performance of epidemic routing using network coding in opportunistic networks. They showed that network coding is superior when bandwidth and node buffers are limited. The application scenario they described is similar to ours. This paper made some modification based on the traditional epidemic model. Moreover, our scheme is designed for the mobile devices. In order to establish the network, the devices in our scheme have to switch between ordinary mode and AP mode. Therefore, even if some devices are very close to each other, they may be unable to communicate.

There are also some studies [13, 14] on the ad hoc networks in which the nodes are mobile devices. In these studies, the mobile devices can connect to each other through working in ad hoc mode. BATMAN [14] is a representative protocol in such application. However, a precondition for this protocol is that those devices in the network have to be rooted because there are very rare commercial released operation systems which could work in ad hoc networking mode. Therefore, this paper studies the file sharing scheme for mobile devices without the support of ad hoc mode.

The contribution of this paper can be summarized as follows. First, the authors analyze and model the secure network-coding-based file sharing scheme for the network with a number of mobile devices. In the scheme, the mobile devices are not required to be rooted before sharing files, which is more realistic. Second, the authors evaluate the scheme and show that file sharing among mobile devices is an ideal place to apply network coding.

3. Proposed Scheme

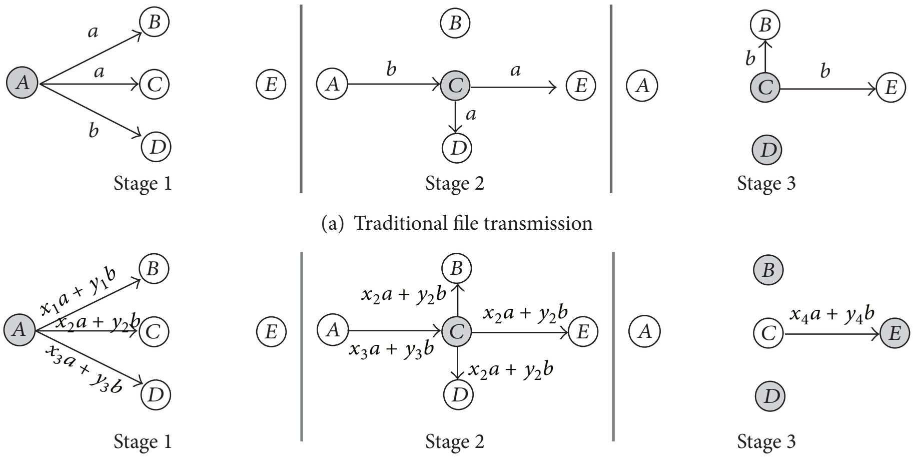
In order to accelerate the sharing rate, this paper proposes a principle data sharing scheme which is based on network coding. The source device needs to encode the original data with network coding. When an intermediate device receives some or all of the data slices, it could reencode the received data with RLNC and then spread the data to its neighbors via D2D communication. Because the data is highly mixed during the reencoding operation, each device could receive and decode the data as long as sufficient slices are received with high probability.

It is feasible to implement direct communication between devices for mobile devices via IEEE 802.11n. When a device X receives part of the encoded slice, it could configure itself as a soft AP and then allow other devices to connect for data transmission. When a device Y joins the network of X, device X reencodes the slices it received and then forwards the slices after reencoding to device Y. Through strategically switching between AP mode and ordinary mode, all the devices in the network could receive and decode the original data.

3.1. Network Model

3.1. Network Model. Instead of the traditional store-and-forward working mode of network devices, network coding technology uses the storage-coding-forwarding working mode at intermediate devices. Through the operation at the intermediate network devices, it can effectively improve the file transmission rate.

Network coding could work in unicast or multicast networks [15]. In most cases, network coding works in multicast networks. However, after a mobile device configures itself to AP mode, the device could not forward multicast message. Some authors consider that multiple devices could overhear the data transmitting to someone at the same time via unencrypted wireless channel [16], but that is another story. Therefore, this paper assumes that the device sends data to its neighbors via unicast connections.



(a) Traditional file transmission

(b) Linear-network-coding-based data sharing

Figure 1: Data sharing in different schemes.

Figure 1 shows the advantages of network coding in sharing files.

The example in Figure 1 shows the principle why network coding helps accelerate the rate of data sharing. Nodes A–E in Figure 1 are a subset of a network. When an intermediate device needs to forward data to its neighbors, it switches to a soft AP. In the second stage of Figure 1(a), after node C switches to AP mode, node B can no longer receive data from C because they have the same data a. Node B can only receive data b at stage 3. After using network coding, this performance could be significantly improved. Node B could decode data a and b in stage 2. From an overall perspective, nodes B, D, and E could become soft APs and spread data to their own neighbors in stage 2 of Figure 1(b), while only node D could become the AP in stage 2 of Figure 1(a). In the third stage of Figure 1(b), nodes B, D, and E could work as soft APs and spread data to its own neighbors, while only node D could become a soft AP in Figure 1(a). Therefore, the sharing rate in the network-coding-based scheme is faster than that in traditional way.

3.2. Network Coding Strategy. Network coding scheme could be divided into linear network coding and nonlinear network coding. RLNC is a practical scheme, and RLNC is suitable for the network with dynamical topology. Deterministic algorithm has higher computation efficiency compared with randomized algorithm, but it is heavily dependent on the network topology. In our scheme, the mobile devices may change their modes from ordinary mode to AP mode, which will change the network logical topology. Therefore, we select RLNC in our scheme. First, the device who starts the sharing process needs to equally split the original file p into k slices $p_1, p_2, \dots, p_k$. In each transmission session, this device randomly selects k elements $a_{i1}, a_{i2}, \dots, a_{ik}$ from the finite field GF(256) and then obtains the encoded slice X_I with

$$X_i = a_{i1} \times p_1 + a_{i2} \times p_2 + \cdots + a_{ik} \times p_k. \tag{1}$$

The reencoding operation at the intermediate nodes could increase the performance of network transmission, including the throughput and security. For each intermediate device, when it needs to transmit a slice to its neighbor, it has to follow the same strategy. It randomly selects m ($m \leq k$) elements to be coefficients from the field GF(256) and then linearly reencodes the m slices it received with the m coefficients. After the reencoding operation, the linear dependency of the data is reduced. Therefore, the receiver could obtain a linearly independent slice with high probability.

As long as a device successfully accumulates k linearly independent slices, it could recover the original files with Gauss-Jordan elimination method.

3.3. Analysis Model. Through the analysis for Figures 1(a) and 1(b), we observe that both the schemes transmit data in a complex network environment. The second scheme is more complex because the data are linearly mixed at the source device and intermediate devices. In order to clarify the difference of the two schemes, we consider this kind of problem as complex-network-based epidemic model and then model the two schemes.

3.3.1. Classical Propagation Model. There are many disease propagation models proposed by previous researchers. In the researches about complex network, the most widely used models are SIS (Susceptible-Infected-Susceptible) model and SIR (Susceptible-Infected-Removed) model. This paper assumes that each device is a node in the network and then makes analysis for both the two models.

When SIS model is used, the nodes could be divided into two categories. One is the mobile devices that have become soft APs, and the other is the devices which are working in AP mode but switched to ordinary mode soon afterwards. However, after using network coding, there exists the third category, namely, the devices that received part of encoded slice but have not switched to AP mode. The devices of this kind cannot be expressed in SIS mode.

Compared with SIS model, there is one more category in SIR model, namely, removal individual. Removal individual is equivalent to the devices which become AP nodes, and then its neighbors received all the data. Finally, these devices permanently close the AP mode. In other words, the devices leave the network permanently.

In accordance with the above analysis, both the two schemes lack the expression for the devices that receive part of encoded data but have not become soft AP. Therefore, traditional SIS and SIR models cannot be directly used in our network environment. We have to make some improvement based on the SIR mode for our scheme.

3.3.2. Analysis Model for the Proposed Scheme. In our model, the concept of hidden nodes is introduced to indicate the devices which could switch to AP mode even if only part of encoded slices is received. Moreover, for any device in the network, it is not allowed to stay in suspended mode, which means that the devices neither switch to AP mode nor receive data from others at that state. Therefore, the switch time of AP mode is very important during the transmission. The transform is described in (2) in which m_j refers to the occupied cache of node j, n_j refers to the cache size of node j, and α_j refers to the proportion of received data

$$\alpha_j = \frac{m_j}{n_j} \quad (\alpha_j \leq 0.5) . \tag{2}$$

Theoretically speaking, any intermediate node could switch to AP mode at any time in the ad hoc network. In order to guarantee the efficiency, when an intermediate node is receiving data, it cannot switch to AP mode. Only when the condition $\alpha_j \leq 0.5$ is satisfied can the device be allowed to start sharing. So the number of mobile devices working in AP mode in the network shows a kind of dynamic distribution. A node in the network will experience the following states:

(1) The data is transmitted from the source node to its neighbor.

(2) The neighbors receive part of encoded data.

(3) Some nodes receive part of data and switch to AP mode.

(4) The nodes decode and recover the original data.

Then this paper makes the following analysis.

(a) All the N mobile devices are divided into three categories, in which N is a dynamic value, and each device is randomly distributed.

(i) For the devices that have recovered all the data and switched to AP, we called them infected group.

(ii) For the devices that have received part of encoded data but been switched to AP mode, we called them hidden group.

(iii) For the devices that have not received any data, we called them healthy group.

(b) Due to the random distribution of mobile devices, the number of adjacent nodes of each device is different. We assume that all the mobile devices are subject to uniform distribution, and each device has λ neighbors. Moreover, this paper assumes that the number of devices working in AP mode at time t is $I(t)$, and the number of devices working in ordinary mode is $S(t)$.

(c) We assume that all the hidden devices could become infected group with a probability P. P is a variable related to the generation depth K, total resource number L, and time t.

Traditional file transmission mode is very different from the mode of RLNC in generation depth. When we set k to be 1, the scheme based on linear network coding will be degenerated to traditional scheme. Therefore, we make the analysis in two kinds of conditions.

(1) When k equals 1, the network-coding-based scheme is equivalent to traditional file sharing scheme. The probability that the hidden AP devices could recover the original file will be influenced by the total resource number L and the transmission time.

This paper assumes that the received data at hidden devices j cannot exceed local cache capacity n_j. When the generation number K is great, the hidden node has to receive all the data so that it could recover the original data. Therefore, the probability that the hidden AP node could recover the original file decreases as K increases. The relation can be expressed by the following equation:

$$P_1 \propto \frac{1}{L}. \tag{3}$$

As time passed, hidden nodes receive more and more slices it requires, and then the probability of successfully decoding will accordingly increase:

$$P_1 \propto t. \tag{4}$$

Through the analysis above, we observe that the transform probability P_1 and time t in traditional scheme have the following relation:

$$P_1 = \left(\frac{C_1}{L} \times t\right) \times \alpha_i. \tag{5}$$

(2) When K does not equal 1, all the data transmitted on the network is encoded with RLNC, and the intermediate devices have to send the linear combinations to its neighbors. Then the transfer probability P_2 will be influenced by the generation K and the transmission time t.

When K becomes greater, hidden nodes have to receive more slices to decode and recover the original file. Therefore, the probability of recovering the original file in a specific time will be reduced.

As time goes on, the probability that the slices required by a node exist in its neighbor will increase.

$$P_2 = \frac{C_2}{k} \times \alpha_j \times t \times L. \tag{6}$$

We assume each soft AP could make $\lambda S(t)$ ordinary devices become soft AP and then set up differential equations

$$N\frac{dI(t)}{d(t)} = \lambda S \times I \times N, \qquad Y = N \times I \times P, \tag{7}$$

$$\begin{aligned} &S(t) + I(t) = 1 \\ &P_2 = \frac{C_2}{k} \times \alpha_j \times L \quad (K \neq 1) \\ &P_1 = \left(\frac{C_1}{L} \times t\right) \times \alpha_j \quad (K = 1) \\ &Y_0 = 1 \\ &\alpha_j \geq \frac{1}{2}. \end{aligned} \tag{8}$$

The meaning of the parameters in (8) is listed in the Abbreviations.

4. Evaluation Result

According to the differential equations and the constraints in (7) and (8), the relation between the time t and the probability $I(t)$ could be expressed in

$$I(t) = \frac{1}{1 + (1/i_0 - 1)e^{-\lambda t}}. \tag{9}$$

No matter what the transmission mode we used is, only the data is different, and the transmission frameworks are the same. We then make the simulation based on this model. When the number of adjacent nodes $\lambda = 2$, the initial ratio $I_0 = 1$:

$$\frac{d_i}{d_t} = \lambda i(1 - i), \qquad I_0 = 0.1. \tag{10}$$

MATLAB is adopted to calculate the function in (10) and (11), and then we obtain the relation between the ratio i (the number of soft APs/the number of all devices), time t, and d_i/d_t, which are displayed in Figure 2 and Figure 3.

According to Figures 2 and 3, when sharing data in a network with N nodes, the number of nodes in the infected group reaches half of the whole nodes, and the sharing rate would reach the highest level which makes the number of successful devices increase at the highest rate.

When $K = 1$, the network-coding-based scheme degenerated to traditional replicate-based scheme. The relation between time t and the number of successful devices that could recover all the original data is calculated for both network-coding-based scheme $K > 1$ and traditional replicate-based scheme $K = 1$, respectively, which is shown in

$$Y(t) = \begin{cases} \dfrac{N}{1 + (1/i_0 - 1)e^{-\lambda t}} \times \left[\left(\dfrac{C_1}{L} \times t\right) \times \alpha_j\right] & (K = 1) \\ \dfrac{N}{1 + (1/i_0 - 1)e^{-\lambda t}} \times \left[\dfrac{C_2}{K} \times \alpha_j \times t \times L\right] & (K \neq 1). \end{cases} \tag{11}$$

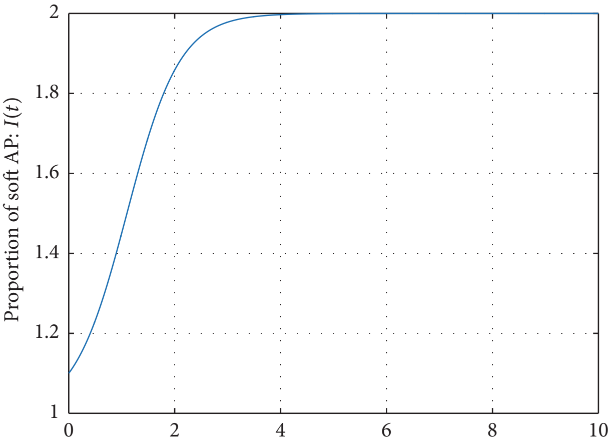


FIGURE 2: Relation between $I(t)$ and t.

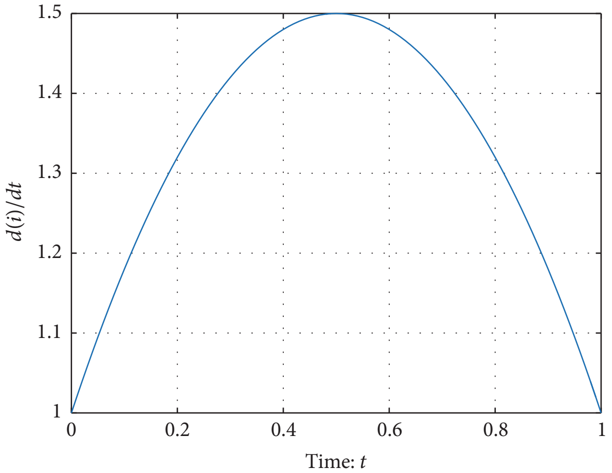


FIGURE 3: Relation between d_i/d_t and t.

Figure 4 is calculated with MATLAB. In the calculation, N is set to be 10, and file size L is set to be 4 M.

As shown in Figure 4, the network-coding-based scheme outperforms traditional replicate-based sharing scheme.

When network coding is used, the parameter K has influence on the data sharing efficiency.

It is clearly evident from Figure 5 that the sharing rate increases as K increases. However, a drawback is that the computational overhead would increase as K increases.

5. Conclusion

In order to realize the large scale date sharing in future networks, this paper studies a scheme based on secure network coding via D2D communication. Part of the mobile devices in the system may be switched to soft AP mode, and linear network coding operation will be performed on the AP before forwarding the file slices. Through the evaluation of analysis model, the authors observe that the time required for file sharing among multiple devices is less than that in traditional networks. In the future, the authors will implement the scheme in mobile devices such as smartphone networks.

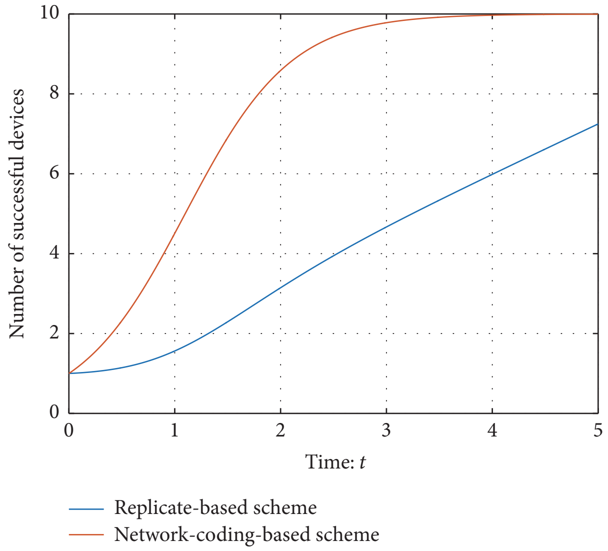


FIGURE 4: Performance of network-coding-based scheme and replicate-based scheme.

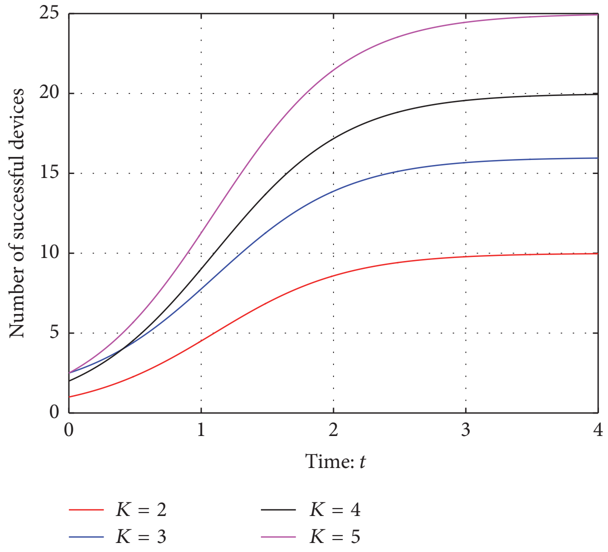


FIGURE 5: Influence of the generation size K.

Abbreviations

N:	The number of nodes in the network
S:	The proportion of devices that have no data but work in soft AP mode
I:	The proportion of mobile devices working in soft AP mode
Y:	The number of devices that can recover data
λ:	The average number of neighbors for each device
ε:	The probability of recovering data
K:	The generation of transmission
L:	The file size
α:	Proportion of cache occupation
P:	Probability of successful recovery
C_1, C_2:	Constant.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Project for Science and Technology of Jiangsu Province (BK20150846), the Scientific Research Foundation of NUPTSF (214167), the Huawei Innovation Research Program (HIRP20140116), the Postdoctoral Fund of Jiangsu (1501041C), and the Postdoctoral Fund of China (2016M590484).

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, 2000.

[2] S. Y. Li, R. W. Yeung, and N. Cai, “Linear network coding,” *IEEE Transactions on Information Theory*, vol. 49, pp. 371–381, 2003.

[3] T. Ho, M. Médard, R. Koetter, 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, 2006.

[4] S. Jaggi, P. Sanders, P. A. Chou et al., “Polynomial time algorithms for multicast network code construction,” *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1973–1982, 2005.

[5] M. Effros, S. El Rouayheb, and M. Langberg, “An equivalence between network coding and index coding,” *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2478–2487, 2015.

[6] K. V. Rashmi, N. B. Shah, K. Ramchandran, and P. V. Kumar, “Regenerating codes for errors and erasures in distributed storage,” in *Proceedings of the 2012 IEEE International Symposium on Information Theory (ISIT ’12)*, pp. 1202–1206, July 2012.

[7] B. Saleh and D. Qiu, “Performance analysis of network–coding–based P2P live streaming systems,” *IEEE/ACM Transactions on Networking*, vol. 24, no. 4, pp. 2140–2153, 2016.

[8] D. Li, H. Zhao, F. Tian, H. Bo, Y. Xu, and G. Zhang, “Multipath network coding and multicasting for content sharing in wireless P2P networks: a potential game approach,” *Computer Communications*, vol. 96, pp. 17–28, 2016.

[9] D. Ferreira, R. A. Costa, and J. Barros, “Real–Time network coding for live streaming in hyper–dense,” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 4, pp. 773–781, 2014.

[10] L. F. Xie, P. H. J. Chong, I. W.-H. Ho, and H. C. B. Chan, “Virtual overhearing: an effective way to increase network coding opportunities in wireless ad-hoc networks,” *Computer Networks*, vol. 105, pp. 111–123, 2016.

[11] M. Yang and Y. Yang, “Peer–to–peer file sharing based on network coding,” in *Proceedings of the 28th International Conference on Distributed Computing Systems (ICDCS ’08)*, pp. 168–175, IEEE, 2008.

[12] Y. Lin, B. Li, and B. Liang, “Stochastic analysis of network coding in epidemic routing,” *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 5, pp. 794–808, 2008.

[13] J. Thomas, J. Robble, and N. Modly, “Off grid communications with android meshing the mobile world,” in *Proceedings of the IEEE Conference on Technologies for Homeland Security (HST ’12)*, pp. 401–405, July 2012.

[14] R. Sanchez-Iborra, M.-D. Cano, and J. Garcia-Haro, “Performance evaluation of BATMAN routing protocol for VoIP services: a QoE perspective,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 4947–4958, 2014.

[15] M. Médard, F. H. P. Fitzek, M.-J. Montpetit, and C. Rosenberg, “Network coding mythbusting: why it is not about butterflies anymore,” *IEEE Communications Magazine*, vol. 52, no. 7, pp. 177–183, 2014.

[16] L. Keller, A. Le, B. Cici, H. Seferoglu, C. Fragouli, and A. Markopoulou, “MicroCast: cooperative video streaming on smartphones,” in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (MobiSys ’12)*, pp. 57–70, June 2012.

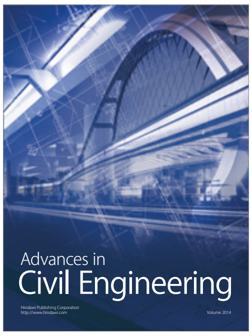
Journal of
Engineering

The Scientific World Journal

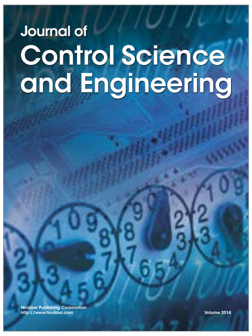
International Journal of
Rotating Machinery

Journal of
Sensors

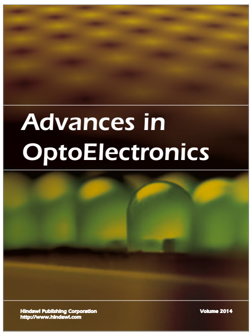
International Journal of
Distributed Sensor Networks


Advances in
Civil Engineering

Journal of
Robotics

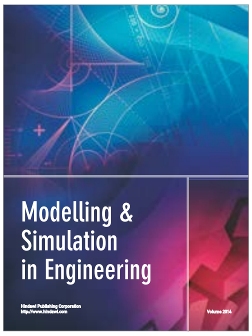

Journal of
Control Science and Engineering

Journal of
Electrical and Computer Engineering

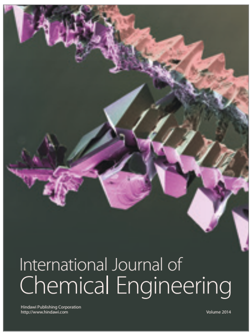

Advances in
OptoElectronics

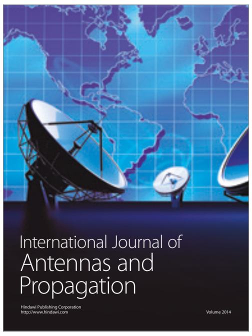
VLSI Design

International Journal of
Navigation and Observation

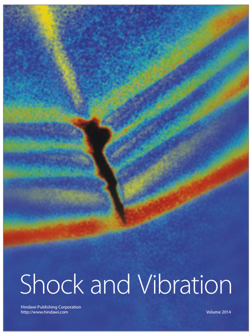

Modelling &
Simulation
in Engineering

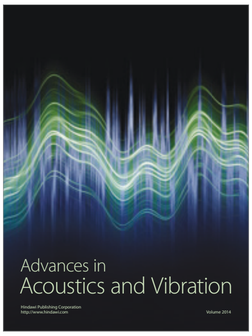
International Journal of
Aerospace Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and Propagation

Active and Passive
Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration