Received: 10 May 2018 | Revised: 4 August 2018 | Accepted: 8 October 2018

DOI: 10.4218/etrij.2018-0254

ORIGINAL ARTICLE

WILEY ETRI Journal

Genetic algorithm-based content distribution strategy for F-RAN architectures

Xujie Li[1] | Ziya Wang[1] | Ying Sun[1] | Siyuan Zhou[1] | Yanli Xu[2] | Guoping Tan[1]

[1]College of Computer and Information, Hohai University, Nanjing, China

[2]Department of Information Engineering, Shanghai Maritime University, Shanghai, China

Correspondence
Xujie Li, College of Computer and Information, Hohai University, Jiangning District, Nanjing, China.
Email: lixujie@hhu.edu.cn

Funding information
This research was supported by the Fundamental Research Funds for the Central Universities (No.2017B14214), Six Talent Peaks project in Jiangsu (DZXX-008) and National Natural Science Foundation of China (No. 61701168, 61601283, 61301110).

Fog radio access network (F-RAN) architectures provide markedly improved performance compared to conventional approaches. In this paper, an efficient genetic algorithm-based content distribution scheme is proposed that improves the throughput and reduces the transmission delay of a F-RAN. First, an F-RAN system model is presented that includes a certain number of randomly distributed fog access points (F-APs) that cache popular content from cloud and other sources. Second, the problem of efficient content distribution in F-RANs is described. Third, the details of the proposed optimal genetic algorithm-based content distribution scheme are presented. Finally, simulation results are presented that show the performance of the proposed algorithm rapidly approaches the optimal throughput. When compared with the performance of existing random and exhaustive algorithms, that of the proposed method is demonstrably superior.

KEYWORDS
content distribution, fog network, genetic algorithm, next-generation access network, throughput analysis

1 | INTRODUCTION

In recent years, cloud radio access network (C-RAN) architectures have been developed to satisfy ever increasing demands for wireless data. However, the performance of C-RANs is typically limited by the throughput-constrained fronthauls used to connect the distributed remote radio heads (RRHs) to the pool of centralized baseband units (BBUs) [1]. Fog computing-based fog radio access network (F-RAN) architectures have been proposed to overcome this limitation [2].

In F-RANs, services can be provided by a centralized unit such as the BBU pool in a C-RAN or hosted at smart terminal devices that are closer to the user equipment (UE) [3]. Fog computing relies on a set of low-power fog access points (F-APs) that are located close to the UEs to offload the services originally located at cloud data centers [4].

Another important network performance metric is the latency. In wireless networks, the end-to-end latency is largely made up of the delay through the air interface, core network, and public data network (PDN). Among these, the most significant is caused by the physical distance between the core network and PDN. When compared to cloud computing networks, F-RANs shorten the transmission distance and reduce the network latency by colocating core network functions and localized mobile data content [5,6]. The scalability and flexibility of F-RANs can be increased by coordinating the technologies used in the air interface, network architecture, and core network [7]. With these advantages, F-RAN architectures have been selected for use in the forthcoming 5G network system [8–12].

An emerging problem in F-RANs lies in determining how best to distribute content across the network. To answer this question, many scholars have researched the distribution and caching of content across networks. In [13], a content caching and distribution framework for smart grid enabled orthogonal frequency division multiplexing (OFDM) networks is presented that facilitates the wireless transmission of multimedia content to mobile users. In [14], a dynamic content distribution (DCD) method is proposed that considers the dynamic supply and demand for content in locations frequented by tourists. The authors in [15] propose a content centric based network architecture that includes a novel caching scheme for storing replicas of mobile content. In [16], a distributed cluster formation algorithm is described that reduces inter-F-AP interference by assisting F-APs in scheduling appropriate UE so as to optimize the system throughput. In [17], a set of dynamic content caching update rules are detailed that stipulate how much and which part of each piece of content should be cached. The authors in [18] propose a dynamic control algorithm to optimally locate content so as to minimize the overall operational cost overtime under service response time constraints. The novel internet access network model based on fog computing that was proposed in [19] dynamically moves cloud or web content to nodes located at the edge of the access network, and then performs proactive caching. The authors in [20] proposed three schemes to solve the energy-efficient dynamic content distribution problem while considering user requests, network resource capacity, and overall energy use. In [21], the authors derived the coverage rates and traversal rates of F-AP and device-to-device (D2D) users by considering different node locations, cache sizes, and user access patterns. In [22], a content distribution protocol is proposed that unifies piece and peer selection via swarming effects and reduces the average download time by up to 20%. In [23], the authors propose a new Bayesian coalition game (BCG) as-a-service for content distribution among objects with support from the cloud. In [24], the authors propose an F-RAN architecture that includes three candidate transmission modes: D2D, local distributed coordination, and global C-RAN. The authors in [25] present both centralized and distributed transmission aware cache placement strategies to minimize the average download delay experienced by users while considering storage throughput constraints. In [26], the authors introduce a new model to compute the total energy consumed by content distribution networks (CDNs), and the results show that increasing the number of surrogate servers decreases the transmission delay. In [27], the authors propose a cooperative cache placement framework for multipoint joint and single-cell transmissions, and present numerical results that confirm that the scheme achieves faster convergence and greatly reduces the content transmission time for mobile users (MUs). In [28], the authors investigate the joint optimization of content object caching and scheduling for in-radio access network (RAN) caches and prove the feasibility of in-RAN cooperative caching. In addition, they consider different time scales when constructing the joint content object caching and scheduling problem. In [29], the authors present an architecture for edge caching, discuss the challenges in implementing existing caching policies in the proposed framework, and provide experimental results demonstrating that with the proposed framework, the requirements of edge caching services can be guaranteed. However, while extensive research has been conducted on the above listed topics, little research has focused on developing practical content distribution schemes for optimally matching users and F-APs.

Many intelligent optimization algorithms have been developed, including the particle swarm, ant colony, and simulated annealing techniques. However, these suffer from a number of limitations.

- The particle swarm optimization method has difficulty searching for local optima, which may result in premature convergence, and the search process may become trapped in a local minima when there is insufficient population diversity in the search space.
- If the parameters of the ant colony algorithm, such as the pheromone volatilization coefficient, are set improperly, the quality of the obtained solution will suffer. In addition, this algorithm is complex and requires a significant amount of time to compute a solution.
- The limitation of the simulated annealing algorithm is that it remains unaware of the conditions of the entire search space, and therefore has difficulty identifying the most promising search area. Thus, the efficiency and convergence speed of this algorithm are low.

An alternative approach to optimization is based on the use of genetic algorithms, which were first proposed by Holland and his students at Michigan University in the late 1960s and early 1970s [30]. Genetic algorithms have the following advantages:

- The probabilistic mechanisms of crossover and mutation are used in every iteration process, which greatly enhances the search ability via a more random and extensive search space.
- The genetic algorithm has good scalability and can be easily combined with other algorithms.

For these reasons, the method proposed in this paper for optimizing content delivery over F-RANs is based on a genetic algorithm.

The main contributions of this work are as follows:

- A genetic algorithm-based content distribution strategy is proposed for use in F-RAN architectures.
- Simulation results are provided that demonstrate the performance, system throughput, and convergence speed of the proposed method.

The remainder of this paper is organized as follows. The F-RAN system model is described in Section 2 along with the method of allocating content to each F-AP and demand content to each UE. Section 3 describes the problematic content distribution scenario in detail, and Section 4 describes the proposed genetic algorithm-based content distribution scheme. Simulation results that demonstrate the feasibility and superiority of proposed content distribution scheme are provided in Section 5 and the paper is concluded in Section 6.

2 | SYSTEM MODEL

2.1 | Network model

Let us consider a single macrocellular network, as shown in Figure 1. The numbers of F-APs and UEs are N and M, respectively, where $N > M$. Then, we let $\mathbf{S} = (N, M)$. The F-APs are numbered 1, 2, …, N and the UEs are numbered 1, 2, …, M. The set of F-APs in the simulated scenario were uniformly distributed in a cell with radius R and were configured to cache the content requested by the UEs in advance from the cloud data centers through the backhaul and fronthaul links. A set of UEs was also uniformly distributed in this area and configured to request content over the F-APs. Each F-AP was assumed to have a fixed transmit power P_t. The wireless links in this scenario were configured to use the orthogonal frequency division multiple access scheme.

2.2 | Caching model

The set of content provided by the network to the UEs is $C = \{C_1, C_2, C_3, \ldots, C_K\}$. The content that was cached by the F-APs was selected based on the respective popularities of each piece of content. The content caching probability follows a Zipf distribution, which can be computed as follows:

$$P_{C_i}^F(\sigma_1, K) = \frac{\frac{1}{i^{\sigma_1}}}{\sum_{n=1}^{K} \frac{1}{n^{\sigma_1}}}, \tag{1}$$

where $\sum_{j=1}^{K} P_{C_i}^F(\sigma_1, K) = 1$, i represents the index of the content, n is an integer variable used in the summation, and $\sigma_1(\sigma_1 > 0)$ is the parameter describing the Zipf distribution, which determines the relative popularity of the cached content. Larger values of σ_1 indicate that the probability caching the most popular content is larger. The total number of pieces of content is K. From (1), it can be seen that the smaller i is, the larger $P_{C_i}^F$ is, which indicates that the content with smaller index values are more likely to be cached by the F-APs. In other words, the probability of caching is greater for smaller index values. For C_i and C_j, if $i < j$, then $P_{C_i}^F(\sigma_1, K) > P_{C_j}^F(\sigma_1, K)$.

Similarly, the content demanded by the UEs was determined using the Zipf distribution. The content demand probability can be calculated as follows:

$$P_{C_i}^U(\sigma_2, K) = \frac{\frac{1}{i^{\sigma_2}}}{\sum_{n=1}^{K} \frac{1}{n^{\sigma_2}}}, \tag{2}$$

where the parameters in this equation are as defined above and σ_2 is a parameter describing the Zipf distribution. The smaller σ_2 is the more popular is the content requested by the UEs.

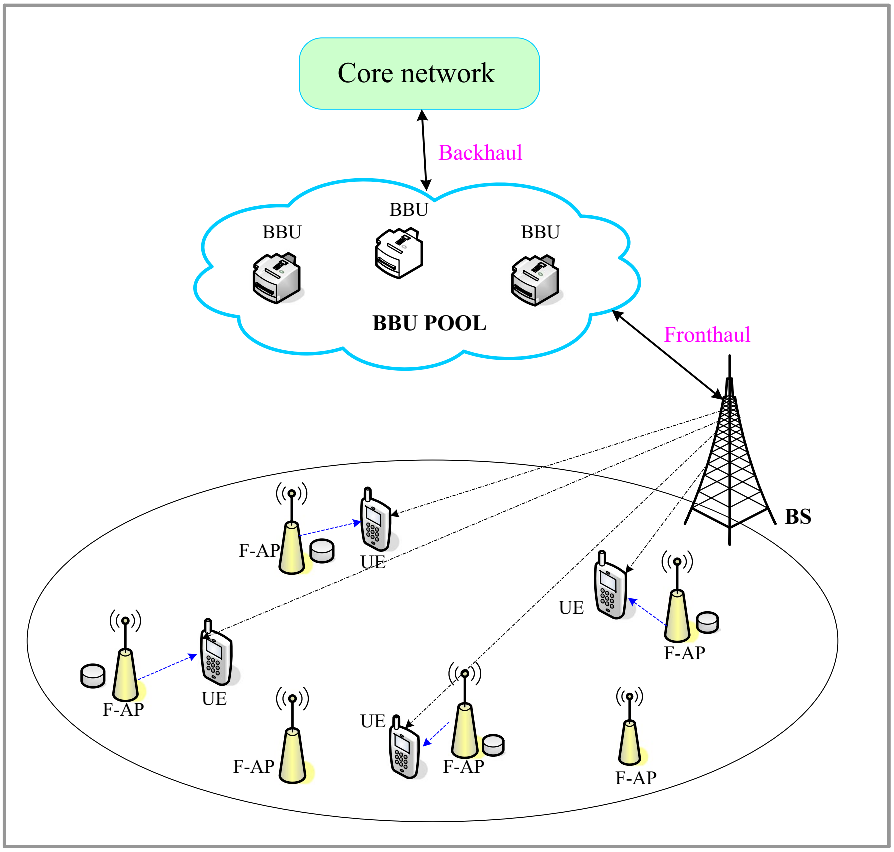


FIGURE 1 F-RAN system model

3 | PROBLEM FORMULATION

The content distribution problem of interest in this scenario is as follows. When UE j is paired with F-AP i, then UE j obtains any required content from F-AP i. The signal-to-interference plus noise ratio (SINR) value of UE j is given by:

$$\beta_j = \frac{P_t |h_{i,j}|^2 / r_{i,j}^{\alpha}}{I_{\text{inter-cell}} + I_{\text{intra-cell}} + n_0 B}, \tag{3}$$

where P_t is the transmit power of the F-AP, $r_{i,j}$ is the distance from F-AP i to UE j, α is the path loss factor, $I_{\text{inter-cell}}$ is the inter-cell interference, and $I_{\text{intra-cell}}$ is the intra-cell interference. In this study, it was assumed that the noise was additive white Gaussian noise (AWGN) with a noise power n_0B, where n_0 represents the power spectral density of the noise and B represents the subchannel bandwidth. As in Rayleigh fading channels, $|h_{i,j}|^2$ follows an exponential distribution with unit mean.

In previous works, many researchers have proposed methods to mitigate inter-cell interference, such as power control [31], scheduling [32], and coordinated multipoint [33]. As mentioned before, we only consider a single macrocellular network where the inter-cell interference is ignored in this paper. In the system model evaluated in this study, there was assumed to be N content providers (F-APs) and M users (UEs) in a cell. When a user requested certain content, an optimal content provider was selected to deliver it. As the F-AP only stores a limited amount of content, the content requested by the user may not be available from the nearest F-AP. Note that if the same frequency is adopted for every pair in the network, then simultaneous communication becomes difficult due to the introduced co-channel interference. While a resource reuse scheme, such as non-orthogonal multiple access, can be used to improve the spectral efficiency, it will result in high levels of interference and a higher implementation complexity [34]. Thus, the intra-cell interference in this model can be ignored. Then, the SINR value of UE j can be rewritten as:

$$\beta_j = \frac{P_t |h_{i,j}|^2 / r_{i,j}^{\alpha}}{n_0 B}. \tag{4}$$

The throughput for UE j can be computed as follows:

$$C_j = B \log_2 \left(1 + \beta_j\right) \tag{5}$$

and the system throughput can be written as:

$$\begin{aligned} C &= \sum_{j=1}^{M} C_j \\ &= \sum_{j=1}^{M} B \log_2 \left(1+\beta_j\right) \\ &= \sum_{i=1}^{M} B \log_2 \left(1 + \frac{P_t |h_{i,j}|^2 / r_{i,j}^{\alpha}}{n_0 B}\right). \end{aligned} \tag{6}$$

Then, the optimization goal can be obtained as:

$$\begin{aligned} &\max \sum_{j=1}^{M} B \log_2 \left(1 + \frac{P_t |h_{i,j}|^2 / r_{i,j}^{\alpha}}{n_0 B}\right) \\ &\quad \text{s.t.} \quad \beta_j \geq \Gamma, \, i \in (1, M), \end{aligned} \tag{7}$$

where Γ is the SINR threshold and M is the number of UEs.

In a certain cell, it is assumed that there are three UEs and five F-APs that are uniformly distributed as shown in Figure 2. The user is interested in content $C_1, C_2, C_3, \ldots, C_{10}$. For example, it is assumed that users 1, 2, and 3 wish to receive content C_1, C_2, and C_3, respectively. The content cached at each F-AP are shown in Figure 2. It can be seen from Equation that if the content requested by a particular user is cached at the nearest neighbor node of the user, then this user will receive the optimized throughput value. Assume that the distance between UE i and F-AP j is d_{ij}. Then, with reference to Figure 2, the distance relationship is as follows:

$$d_{11} < d_{13} < d_{14}, \tag{8}$$

$$d_{22} < d_{23} < d_{21} < d_{25}, \tag{9}$$

$$d_{35} = \min \left\{d_{3j}\right\} \; (j = 1, 2, \ldots, 5). \tag{10}$$

As shown in Figure 2, the content requested by user 3 is cached at all F-APs, and F-AP 5 is closest to user 3. Therefore, if user 3 is paired with F-AP 5, then the optimized throughput can be achieved. However, the content requested by user 1, namely C_1, was not cached at F-AP 1, which was the nearest. In addition, the content requested by user 2, namely C_2, was not cached at F-AP 2, which is closest to the user 2. Thus,

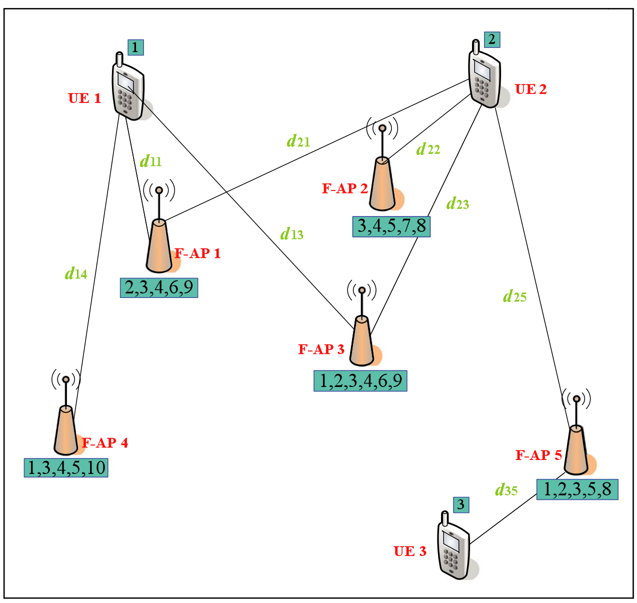


FIGURE 2 Scene to be solved by genetic algorithm

users 1 and 2 must locate other F-APs from which to source their requested content. In such a scenario, it is important to determine how best to distribute the content from the F-APs to all UEs. Ideally, a rational distribution strategy can improve resource utilization, reduce the transmission delay, and achieve optimal system throughput.

This optimization problem is solved in the following section using the genetic algorithm.

4 | GENETIC ALGORITHM-BASED OPTIMAL CONTENT DISTRIBUTION SCHEME

The proposed optimal content distribution scheme is described in this section. In this approach, the system throughput is optimized by pairing the F-APs and UEs based on the genetic algorithm.

The advantages of genetic algorithms were introduced in Section 1. The main principles of genetic algorithms are as follows.

- Produce an initial population.
- Construct a fitness function according to the objective function of the problem.
- Continue choosing and multiplying according to the fitness value.
- After several generations, the best fitness value is the optimal solution.

A flowchart of the proposed algorithm is shown in Figure 3 and contains the following six steps.

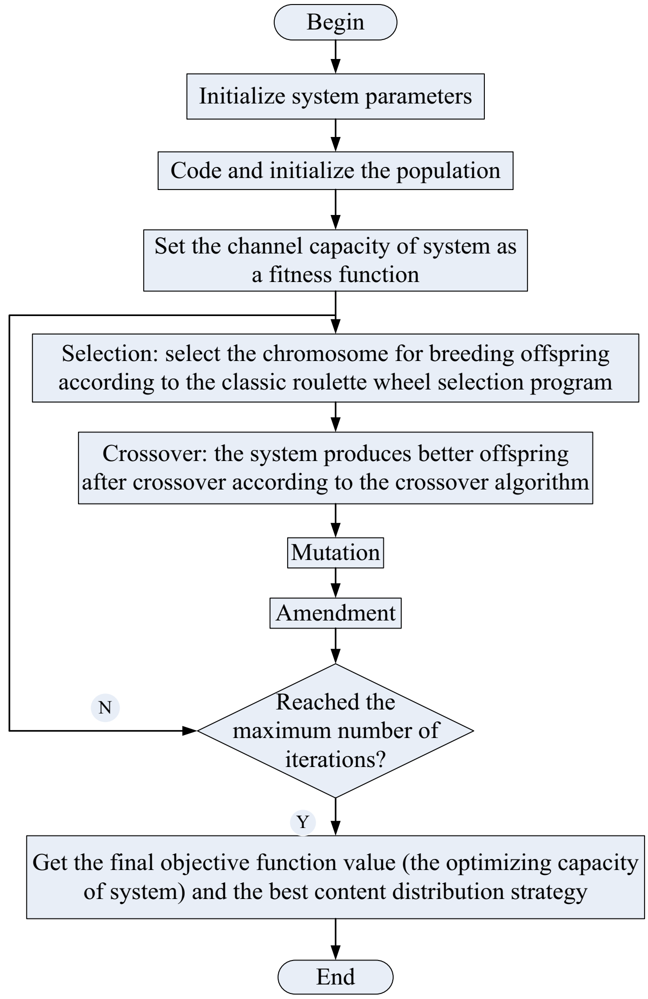


FIGURE 3 Flowchart of the proposed algorithm

4.1 | Coding

Each chromosome is encoded as an M-dimensional vector $\mathbf{U}_x = \left(u_1, \ldots, u_m, \ldots, u_M\right)$, $u_m \in \{1, 2, \ldots, N\}$ and the elements in the chromosome can be repeated. Here, x is the index of the individual in a population. For $\mathbf{S} = (7, 4)$, there is a chromosome encoded as $\mathbf{U}_x = (3, 1, 6, 2)$, which means that UE 1 acquires its content from F-AP 3, and UEs 2, 3, and 4 acquire their content from F-APs 1, 6, and 2, respectively, as illustrated in Figure 4.

4.2 | Population Initialization

Define the number of initialized populations as NP. The elements in chromosome $\mathbf{U}_x$ are discrete random variables from 1 to N. Note that it is necessary to verify the initial population.

- If none of the content requested by the UEs is present in the F-AP corresponding to one or more genotypes of an individual, the individual should be reinitialized until the condition is satisfied.
- In terms of UEs, if multiple UEs are paired with the same F-AP, the content requested by these UEs should be the same. Otherwise, the individual should be reinitialized.

4.3 | Select the fitness function

Each individual in the genetic algorithm is assigned a fitness value, which is a deterministic index describing the individual survival opportunity in the group. As mentioned in Section 3, the system throughput was selected as the fitness function in the proposed approach. This fitness function is as per (6) and the optimization goal is as per (7).

4.4 | Breeding process

In the genetic algorithm, the population evolves toward the optimal solution via a breeding process that comprises four steps: selection, crossover, mutation, and amendment.

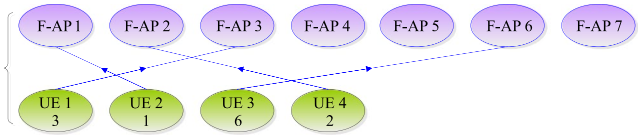


FIGURE 4 Example of the proposed content distribution scheme

4.4.1 | Selection

In the genetic algorithm, the most commonly used selection strategy is the proportional selection strategy, that is, the probability that each individual is selected for genetic operation is the ratio of the fitness of the individual to the sum of all individual fitness values in the population. Here, for chromosome $\mathbf{U}_x$, the probability of being selected for breeding offspring is:

$$P_x = \frac{C\left(U_x\right)}{\sum_{m=1}^{NP} C\left(U_m\right)}. \tag{11}$$

According to the classic roulette wheel selection program [35]:

$$PP_0 = 0, \tag{12}$$

$$PP_x = \sum_{j=1}^{x} P_j. \tag{13}$$

The total number of turns of the roulette wheel is NP. For the kth turn, $\xi_k \in U(0, 1)$ is generated randomly. The individual $\mathbf{U}_x$ will be selected to reproduce the offspring when the relationship $PP_{x-1} \le \xi_k \le PP_x$ has been established.

4.4.2 | Crossover

The purpose of crossover is to produce better offspring. In this process, two chromosomes are simultaneously manipulated so that the properties of each can be combined to produce new offspring. The crossover method in the proposed algorithm is a single cutoff point crossover. The chromosome crossover point is selected randomly. The genes of the parent chromosomes located at the right side of the crossover point are exchanged to produce offspring.

For example, for $\mathbf{S} = (9, 5)$, the crossover point is assumed to be three. If the two chromosomes $\mathbf{U}_1 = (3, 5, 2, 1, 6)$ and $\mathbf{U}_2 = (4, 1, 5, 2, 3)$ are selected as parents to crossover, the process proceeds as shown in Figure 5. After crossover, the resulting chromosomes are $\mathbf{U}_1 = (3, 5, 2, 2, 3)$ and $\mathbf{U}_2 = (4, 1, 5, 1, 6)$.

However, not all selected parents should cross. The value of the crossover probability is p_c. Many scholars recommend selecting a crossover probability between 0.6 and 0.9 [36]. This is because a higher crossover probability results in a larger solution space, thereby reducing the likelihood of finding a nonoptimal solution. However, if the cross probability is too high, it will increase the computational burden required to search the oversized solution space. It is therefore important to ensure that p_c is selected appropriately.

To improve the performance of the proposed algorithm, an adaptive crossover probability mechanism is introduced [36,37]. This addition efficiently enhances the searching ability and increases the convergence speed. The adaptive crossover probability for individual i can be written as:

$$p_c(i) = \begin{cases} p_{c_0}\left(1 + \alpha \frac{f_i - f_{\min}}{f_{\text{average}} - f_{\min}}\right), & f_i \le f_{\text{average}} \\ 1, & f_i > f_{\text{average}} \end{cases}, \tag{14}$$

where p_{c0} is the initial crossover probability with a reference range from 0.6 to 0.9, α is the coefficient factor, $f_{\min}$ and f_{average} are the minimum and average fitness values of the current population, respectively, and f_i is the fitness value of the individual i. As detailed in previous work [36], the values of p_{c0} and α were set to 0.6 and 0.5, respectively. As per (14), the greater the fitness value of the individual, the greater the

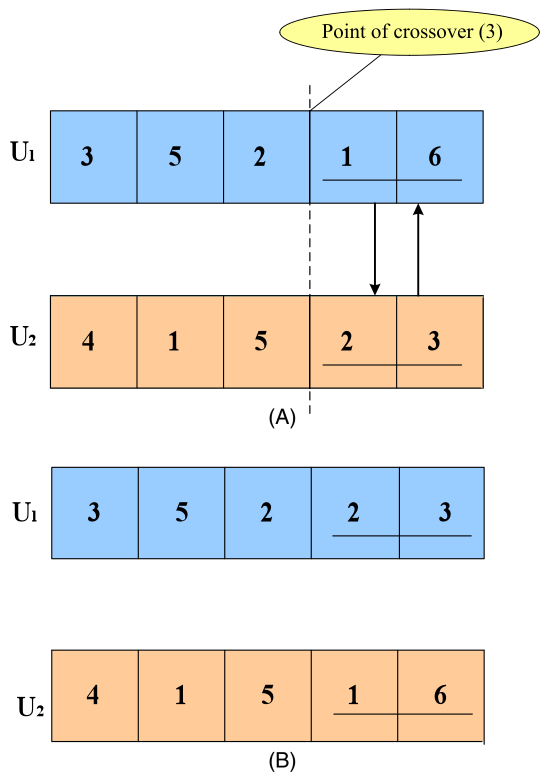


FIGURE 5 Example of the crossover process: (A) chromosomes selected to crossover and (B) offspring after crossover

probability of being selected for crossover. If the fitness value of an individual is higher than the average fitness value in this generation, then the crossover probability of the individual is one.

4.4.3 | Mutation

The purpose of mutation in this context is to maintain the diversity of the population. Here, mutation refers to the changing of random gene bits $U_{i,j}$ in the population according to the mutation probability p_{m}, where $U_{i,j} \in \{1, 2, \ldots, N\}$, $i = 1, 2, \ldots, NP$, $j = 1, 2, \ldots, M$. In genetic algorithms, mutations can produce genes that did not appear in the initial population, which adds new content to the population.

The mutation probability governs the probability of introducing genetic variation into an individual. In previous work [35], typical values of the mutation probability are in the range of 0.005–0.05. However, as these mutation probability values are extremely small, an adaptive method of setting the mutation probability was not included in the proposed algorithm. Instead, the mutation probability was set to a constant value of $p_{\mathrm{m}} = 0.05$.

4.4.4 | Amendment

To guarantee the quality of service (QoS) to the UE, the SINR value must be greater than the SNIR threshold. In addition, proofreading is required for each new generation after breeding. Here, the revision method is similar to that used for population initialization. For an individual:

- If the UE has not requested any content from the F-AP that corresponds to one or more genotypes in the new individual, then the individual is replaced with the corresponding one in the population of the last generation.
- If the same element exists, that is, different individuals are paired with the same F-AP, then these individuals must be interested in the same content. Otherwise, the individual should be replaced by the corresponding one in the population of the last generation.

In addition, if the content requested by the UE does not exist in any F-AP, the UE will request that the required content be delivered from the cloud data centers.

4.5 | Optimal preservation strategy

For each generation of the population, the best individual in the population is retained as part of the optimal set, which is used in the selection of the next generation. In this way, the optimal individuals in each generation are not lost.

4.6 | Stopping criteria

We denote the number of iterations as NG. In general, the population will converge after NG iterations, resulting in an optimal chromosome, which can be used to calculate the optimal result. In this paper, the population stops breeding when the required number of iterations is reached as this represents the optimized throughput of the system.

5 | SIMULATION RESULTS AND ANALYSIS

A MATLAB simulation was conducted to evaluate the performance of the proposed content distribution scheme. In the simulation, it was assumed that there were nine F-APs and four UEs uniformly distributed in a cell with radius R. The locations of the UEs and F-APs in the cell

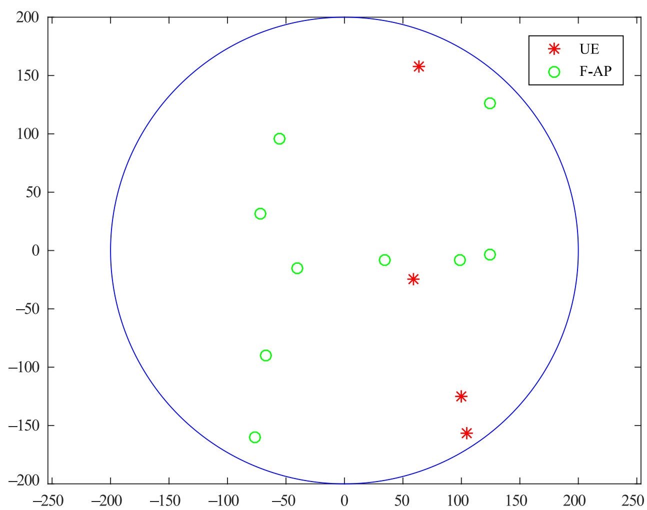


FIGURE 6 Distribution of the UEs and F-APs

TABLE 1 Simulation parameters

Parameters	Value
Population number (NP)	11
Number of iterations (NG)	50
Number of pieces of popular content (K)	100
Cell radius (R)	200 m
SINR threshold (β)	4.6 dB
Noise power (N_0)	−174 dBm
Bandwidth of the subchannel (B)	0.2 MHz
Transmit power of the F-APs (P_t)	200 mW
Path loss exponent (α)	4
Zipf index 1 (σ_1)	1.4
Zipf index 2 (σ_2)	1.5

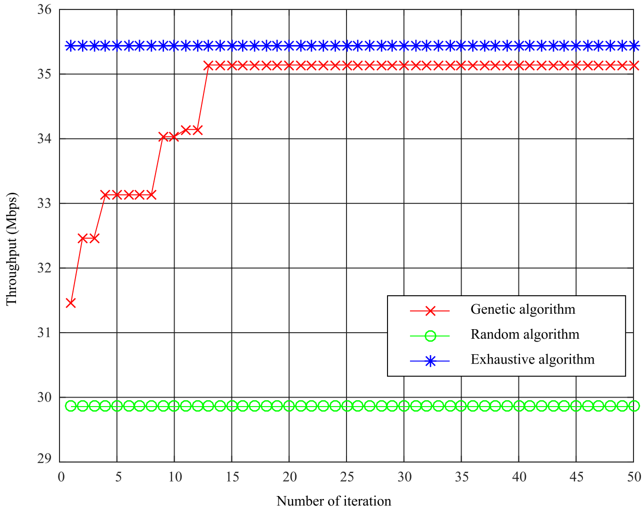


FIGURE 7 Throughput of the F-RAN system

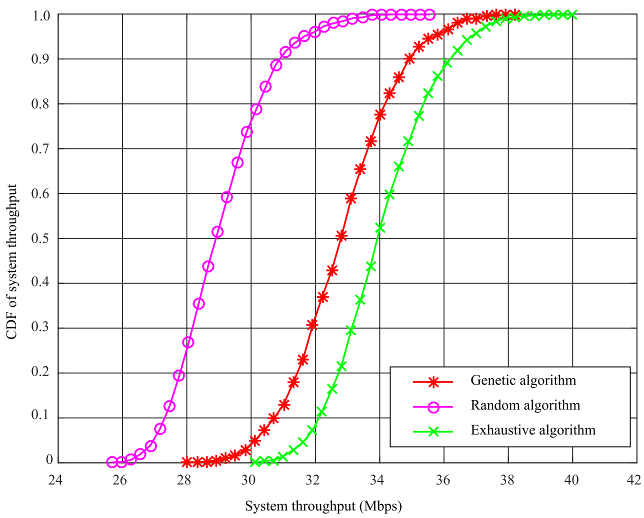


FIGURE 8 CDF curve of system throughput

are illustrated in Figure 6. The simulation parameters are shown in Table 1.

A comparison of the system throughput obtained via the genetic, exhaustive, and random algorithms is shown in Figure 7. In the figure, it can be seen that the exhaustive algorithm achieved the best results as it evaluates an exhaustive list of all possible situations. However, in the simulations in this study, the exhaustive algorithm was observed to have a high complexity as it required N^M computations ($9^4 = 6{,}561$), which takes a long time to complete. In contrast, the throughput based on the genetic algorithm quickly converged to a value close to the optimal value obtained via the exhaustive algorithm and was far superior to that obtained by the random algorithm. The population reached the optimal value after only approximately 13 iterations.

The cumulative distribution function (CDF) curves of the system throughput for the different algorithms are shown in Figure 8, where it can be seen that the proposed genetic algorithm obtain a 4 Mbps system throughput gain compared with the random algorithm. At the same time, while the system throughput of the proposed genetic algorithm was very close to that of the exhaustive algorithm, the complexity of the proposed genetic algorithm was significantly lower. These results demonstrate the excellent performance of the proposed content distribution strategy.

6 | CONCLUSION

In this paper, an optimal content distribution scheme based on a genetic algorithm was proposed. In the target scenario, N F-APs and M UEs were uniformly distributed around the cell. When the F-APs and UEs were paired using the genetic algorithm, the optimal system throughput was obtained. The final simulation results show that the performance of the proposed method was far superior to that of the random algorithm and close to that of the exhaustive algorithm. It is anticipated that these results will prove useful in the design and optimization of F-RAN architectures.

ACKNOWLEDGMENTS

This work was supported by the Fundamental Research Funds for the Central Universities (No. 2017B14214), the Six Talent Peaks project in Jiangsu (DZXX-008), and the National Natural Science Foundation of China (Nos. 61701168, 61601283, and 61301110).

ORCID

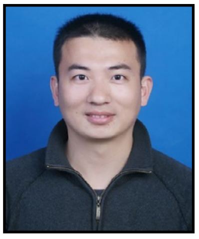
Xujie Li https://orcid.org/0000-0001-5486-5702

REFERENCES

1. C. Zhang et al., *Social-aware content downloading for fog radio access networks supported device-to-device communications*, in Int. Conf. Ubiquitous Wireless Broadband (ICUWB), Nanjing, China, 2016, pp. 1–4.
2. M. Peng et al., *Fog-computing-based radio access networks: issues and challenges*, IEEE Netw. **30** (2016), no. 4, 46–53.
3. M. Peng, K. Zhang, *Recent advances in fog radio access networks: performance analysis and radio resource allocation*, IEEE Access **4** (2016), 5003–5009.
4. H. Zhang et al., *Computing resource allocation in three-tier IoT fog networks: a joint optimization approach combining Stackelberg game and matching*, IEEE Internet Things J. **4** (2017), no. 5, 1204–1215.
5. Y. Shih et al., *Enabling low-latency applications in fog-radio access network*, IEEE Netw. **31** (2017), no. 1, 52–58.
6. M. Chiang, T. Zhang, *Fog and IoT: an overview of research opportunities*, IEEE Internet Things J. **3** (2016), no. 6, 854–864.

7. D. Zeng et al., *Joint optimization of task scheduling and image placement in fog computing supported software-defined embedded system*, IEEE Trans. Comput. **65** (2016), no. 12, 3702–3712.
8. S. Jia et al., *Hierarchical content caching in fog radio access networks: ergodic rate and transmit latency*, China Commun. **13** (2016), no. 12, 1–14.
9. S. Avik, T. Ravi, S. and Osvaldo, *Fog-aided wireless networks for content delivery: fundamental latency tradeoffs*, IEEE Trans. Inf. Theory **63** (2017), no. 10, 6650–6678.
10. X. Wang, S. Leng, and K. Yang, *Social-aware edge caching in fog radio access networks*, IEEE Access **5** (2017), 8492–8501.
11. H. Zhang et al., *Fog radio access networks: mobility management, interference mitigation, and resource optimization*, IEEE Wireless Commun. **24** (2017), no. 6, 120–127.
12. D. Ahmed and S. Sameh, *Data dissemination using instantly decodable binary codes in fog-radio access networks*, IEEE Trans. Commun. **66** (2018), no. 99, 2052–2064.
13. X. Huang and A. Nirwan, *Content caching and distribution in smart grid enabled wireless networks*, IEEE Internet Things J. **4** (2017), no. 2, 513–520.
14. K. Shahriar et al., *Dynamic content distribution for decentralized sharing in tourist spots using demand and supply*, in Int. Wireless Commun. and Mobile Comput. Conf. (IWCMC), Valencia, Spain, 2017, pp. 2121–2126.
15. Z. Su and Q. Xu, *Content distribution over content centric mobile social networks in 5G*, IEEE Commun. Mag. **53** (2015), no. 6, 66–72.
16. Y. Sun, T. Dang, and J. Zhou, *User scheduling and cluster formation in fog computing based radio access networks*, in IEEE Int. Conf. Ubiquitous Wireless Broadband (ICUWB), Nanjing, China, 2016, pp. 1–4.
17. K. Karla et al., *A methodology for the design of self-optimizing, decentralized content-caching strategies*, IEEE/ACM Trans. Netw. **24** (2016), no. 5, 2634–2647.
18. X. Qiu et al., *Cost-minimizing dynamic migration of content distribution services into hybrid clouds*, IEEE Trans. Parallel Distrib. Syst. **26** (2015), no. 12, 3330–3345.
19. N. Lotti et al., *Improving quality of experience in future wireless access networks through fog computing*, IEEE Internet Comput. **21** (2017), no. 2, 26–33.
20. L. Jaime et al., *Energy efficient dynamic content distribution*, IEEE J. Sel. Areas Commun. **33** (2015), no. 12, 2826–2836.
21. S. Yan, M. Peng, and W. Wang, *User access mode selection in fog computing based radio access networks*, in Int. Conf. Commun. (ICC), Kuala Lumpur, Malaysia, 2016, pp. 1–6.
22. E. Flavio and C. Walter, *Integrating piece and peer selection in content distribution networks*, in Global Commun. Conf. (GLOBECOM), San Diego, CA, USA, 2015, pp. 1–6.
23. K. Neeraj, J. P. C. R. Joel, and C. Naveen, *Bayesian coalition game as-a-service for content-caching distribution in internet of vehicles*, IEEE Internet Things J. **1** (2014), no. 6, 544–555.
24. H. Xiang et al., *Joint mode selection and resource allocation for downlink fog radio access networks supported D2D*, in Heterogeneous Netw. for Quality, Reliability, Security and Robustness (QSHINE), Taipei, Taiwan, 2015, pp. 177–182.
25. J. Liu et al., *Cache placement in Fog-RANs: from centralized to distributed algorithms Fog-RAN*, IEEE Trans. Wireless Commun. **16** (2017), no. 11, 7039–7051.
26. B. Andrea, M. and Reza, M. Michela, *Energy consumption for data distribution in content delivery networks*, in IEEE Int. Conf. Commun. (ICC), Kuala Lumpur, Malaysia, 2016, pp. 1–6.
27. H. Li et al., *Cooperative RAN caching based on local altruistic game for single and joint transmissions*, IEEE Commun. Lett. **21** (2017), 853–856.
28. Z. Wang, H. Li, Z. and Xu, *Real-world traffic analysis and joint caching and scheduling for in-RAN caching networks*, Sci. China Inf. Sci. **60** (2017), no. 6, 1–15.
29. L. Lei et al., *Collaborative edge caching through service function chaining: architecture and challenges*, IEEE Wireless Commun. **25** (2018), no. 3, 94–102.
30. H. John, *Holland, genetic algorithms and the optimal allocation of trials*, SIAM J. Comput. **2** (1972), no. 2, 88–105.
31. S. Luca et al., *Energy-efficient power control: a look at 5G wireless technologies*, IEEE Trans. Signal Process. **64** (2016), no. 7, 1668–1683.
32. D. Bai, J. Lee, and I. Kang, *Advanced interference management for 5G cellular networks*, IEEE Commun. Mag. **52** (2014), no. 5, 52–60.
33. K. Manolakis et al., *The role of small cells, coordinated multipoint, and massive MIMO in 5G*, IEEE Commun. Mag. **52** (2014), no. 5, 44–51.
34. J. Montalban et al., *Multimedia multicast services in 5G networks: subgrouping and non-orthogonal multiple access techniques*, IEEE Commun. Mag. **56** (2018), no. 3, 91–95.
35. A. Lipowski and D. Lipowska, *Roulette-wheel selection via stochastic acceptance*, Phys. A **391** (2012), 2193–2196.
36. M. Srinivas and L. Patnaik, *Adaptive probabilities of crossover and mutation in genetic algorithms*, IEEE Trans. Syst. Man Cybern. Syst. **24** (1994), no. 4, 656–667.
37. L. Zhang, N. Wang, and X. He, *RNA genetic algorithm with adaptive crossover probability for estimating parameters of heavy oil thermal cracking model*, - in Chinese Control Conf., Xi'an, China, 2013, 1866–1870.

AUTHOR BIOGRAPHIES



Xujie Li received his PhD degree in communications engineering from the National Mobile Communications Research Lab at Southeast University in 2012. He is currently working as an Associate Professor in the College of Computer and Information Engineering at Hohai University. From April 2016 to April 2017, he was a visiting associate professor at the University of Warwick, UK. Dr. LI has served as a reviewer for many journals and conferences, such as IEEE JSAC, TCOM, TWC, TSP, IET Com, Globecom, etc. His research interests include device-to-device (D2D) communications, energy-efficient wireless communications systems design, and small cells.

Ziya Wang received her BS degree from Henan University in 2016 and is currently pursuing her MS degree at the College of Computer and Information Engineering at Hohai University, majoring in signal and information processing. Her current research interests include D2D communications, optimization algorithms, resource allocation, and 5G networks.

Ying Sun received her BS degree in communication engineering from Hohai University in 2012 and is currently pursuing her MS degree at the College of Computer and Information Engineering at Hohai University. Her current research interests include resource allocation, D2D communications, and system performance analysis.

Siyuan Zhou received his BS degree in electronic engineering from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2008, an MS degree in communication and information systems from Southeast University, Nanjing, China, in 2011, and a PhD degree in electronics and telecommunications engineering from Politecnico di Torino, Torino, Italy. At present, he is an assistant professor with the College of Computer and Information at Hohai University. His current research interests include wireless communications theory and cooperative communications.

Yanli Xu receive her BS degree from the Department of Electrical and Electronics Engineering of Shandong Normal University in 2006. She received her PhD degree from the National Mobile Communication Key Laboratory of Southeast University in China in 2012. She is now working as an Assistant Professor at Shanghai Maritime University. Her research interests are in mobile and ad hoc networks, with a current focus on D2D communications.

Guoping Tan is a full Professor with the College of Computer and Information Engineering at Hohai University, Nanjing, China. He received his PhD degree in Computer Science from Saarland University, Germany, in 2009. He has been an IEEE Senior member since 2017. He was a senior engineer at ZTE Corporation China from 2000 to 2003. From September 2003 to August 2005, he was a PhD Candidate with the National Mobile Communication Research Lab at Southeast University, Nanjing, China. From September 2005 to April 2009, he was a researcher at the Telecommunications Laboratory, Saarland University, Germany under a DAAD scholarship. His current research interests include wireless multi-hop networks, wireless multimedia communications, and signal processing for mobile communications.