A Data-Driven Robustness Algorithm for the Internet of Things in Smart Cities

Tie Qiu, Jie Liu, Weisheng Si, Min Han, Huansheng Ning, and Mohammed Atiquzzaman

The authors propose an approach to improve the robustness of network topology based on a multi-population genetic algorithm (MPGA). First, the geographic information and neighbor list of nodes are extracted from a big data server. Then a novel MPGA with a crossover operator and a mutation operator is proposed to optimize the robustness of topology.

Abstract

The Internet of Things has been applied in many fields, especially in smart cities. The failure of nodes brings a significant challenge to the robustness of topologies. The IoT of smart cities is increasingly producing a vast amount different types of data, which includes the node's geographic information, neighbor list, sensing data, and so on. Thus, how to improve the robustness of topology against malicious attacks based on big data of smart cities becomes a critical issue. To tackle this problem, this article proposes an approach to improve the robustness of network topology based on a multi-population genetic algorithm (MPGA). First, the geographic information and neighbor list of nodes are extracted from a big data server. Then a novel MPGA with a crossover operator and a mutation operator is proposed to optimize the robustness of topology. Our algorithm keeps the initial degree of each node unchanged such that the optimized topology will not increase the energy cost of adding edges. The extensive experiment results show that our algorithm can significantly improve the robustness of topologies against malicious attacks.

Introduction

The Internet of Things (IoT) [1] is an integration of multiple disciplines, including fifth generation (5G) ultra-dense cellular networks [2], heterogeneous ad hoc networks [3], hybrid mobile networks [4], wireless sensor networks [5], and so on. The IoT has a broad range of applications in smart cities. Typically, it deploys a large number of networking nodes within a certain area, and these nodes communicate with each other to collect data and provide reference for smart cities as shown in Fig. 1, such as industry, agriculture, security, transportation, smart home, and healthcare. Meanwhile, these are producing a vast amount and different types of data. Therefore, how to improve the robustness of IoT against node failure based on big data servers of smart cities has become an essential issue in recent years [6].

The scale-free model is one of the classic models in complex network theory. It is mainly used for modeling homogeneous network topologies, in which the node degree follows power-law distribution [7]. Scale-free topology has better performance in withstanding random attacks than small world topology, but it is fragile under malicious attacks [8]. Therefore, researchers and developers focus on how to construct a scale-free topology with high robustness against malicious attacks.

In recent years, some researchers have employed evolutionary algorithms to enhance the robustness of topologies [9]. In this article, we are interested in achieving this by genetic algorithms (GAs) particularly, which is one class of evolutionary algorithms. The population of candidate solutions is used to evolve toward the optimal solution in GA. For the conventional GA, there is a typical limitation called premature convergence. However, one type of GA, the multi-population genetic algorithm (MPGA), can effectively overcome this limitation by using multiple populations to co-evolve.

In this article, the main contributions are as follows:

- Based on the topology characteristics extracted from the big data servers of smart cities, we propose an MPGA with novel crossover and mutation operators to enhance the robustness of topologies.
- Our algorithm keeps the nodes' degrees unchanged during exchanging edges of topology. It will not increase the energy cost of adding edges.

Initial Topology Construction

Due to the limitations of energy and communication range, the scale-free IoT topology in smart cities has the following two constraints:

1. The communication range of its nodes cannot be arbitrarily long.
2. Its node degrees cannot be arbitrarily large.

Because of these two constraints, the traditional method of constructing scale-free topologies, such as the Barabási-Albert (BA) model [10], cannot be directly applied. In order to simulate the scale-free topology information extracted from the big data server in smart cities [11], we employ the following method, which adapts the BA model, to construct the initial scale-free topologies for IoT in smart cities. We add edges between nodes asynchronously during the process of constructing IoT topology after judging whether they are in the communication range. This means that a pair of nodes cannot generate a new edge at the same time. The local world of the newly joined node is composed of all the nodes within its communication range. If a node has been connected with the newly joined node or has reached the maximum

Digital Object Identifier: 10.1109/MCOM.2017.1700247

Tie Qiu, Jie Liu, and Min Han are with Dalian University of Technology; Weisheng Si is with Western Sydney University; Huansheng Ning is with the University of Science and Technology Beijing; Mohammed Atiquzzaman is with the University of Oklahoma.

 0163-6804/17/$25.00 © 2017 IEEE

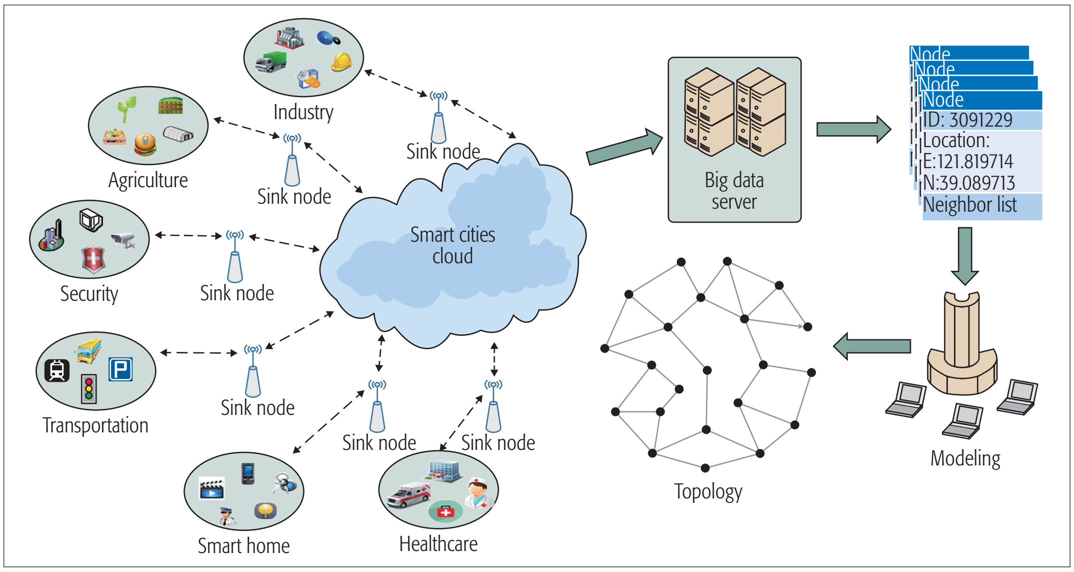

Figure 1. The data-driven robustness topology of IoT in smart cities.

degree [12], it will be removed from the local area of the newly joined node. Furthermore, the newly joined node chooses its neighbors to establish connection by a roulette method according to the node's degree. The newly joined node prefers to connect with higher degree nodes in its local area.

Preliminary

After the initial topologies for IoT nodes in smart cities are constructed, we try to optimize the robustness of topology based on MPGA. First, we convert an adjacency matrix of topology to a binary-coded chromosome. To further illustrate this operator, a topology with four nodes is converted to a chromosome, as shown in Fig. 2. The topology consists of node *i*, node *j*, node *k*, and node *l*, and the adjacency matrix is a binary matrix. It is feasible to convert the adjacency matrix into a chromosome directly. However, the storage space is wasted and the operating complexity in GA is increased. The adjacency matrix is a symmetric matrix, and its upper triangular matrix is able to represent the connection relation between nodes in networks. We convert the upper triangular matrix to a chromosome as shown in Fig. 2. It can shorten the length of the chromosome and improve the efficiency of the proposed algorithm.

•The probability of selecting an individual depends on its fitness value. When the fitness value of an individual is much higher than other individuals' in the current population, this individual will be selected several times in the operation of constructing the next generation. Finally, the population will be controlled by this individual and become uncompetitive, which causes stagnant evolution in population.

•The frequency of the crossover operator and mutation operator is influenced by crossover probability P_c and mutation probability P_m. The values of P_c and P_m directly affect the balance between global search and local search in GA. The result of evolution is quite sensitive to the values of P_c and P_m.

•The size of population has a great influence on the optimization performance. If the size of the population is small, the diversity of population will be decreased and the competition among individuals will be weakened. The population will become a single group soon, and the effect of the crossover operator will gradually disappear. The update of population only relies on the mutation operator. If the size of the population is large, the calculation cost will be increased and the efficiency will be affected.

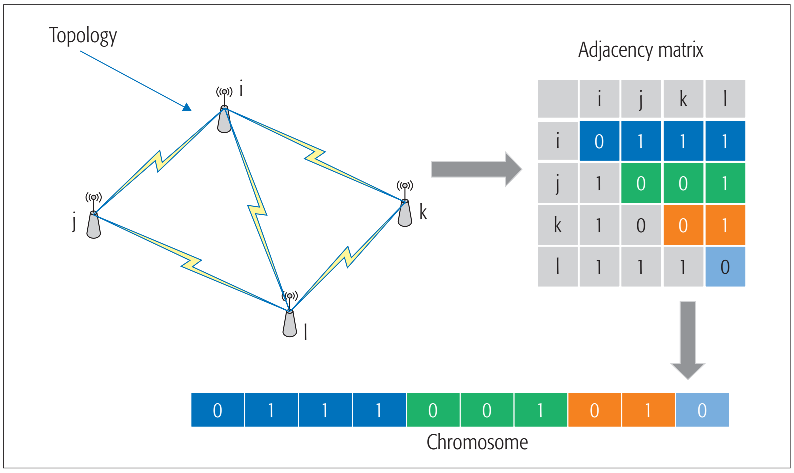

Figure 2. The adjacency matrix is converted to a chromosome.

In order to solve these problems that exist in a conventional GA, the following improvements are applied in an MPGA.

The framework of a GA, which only uses a single population to search for the optimal solution, is broken by using several populations at the same time. For different populations, different crossover probability P_c and mutation probability P_m are selected. Although the ranges of crossover probability and mutation probability are suggested as $P_c(0.7–0.9)$ and $P_m(0.001–0.05)$, respectively, their ranges are wide enough to allow plenty of values to be selected. The optimization results will greatly differ with different P_c and P_m. Given the global search and local search at the same time, we conduct experiments with various P_c and P_m

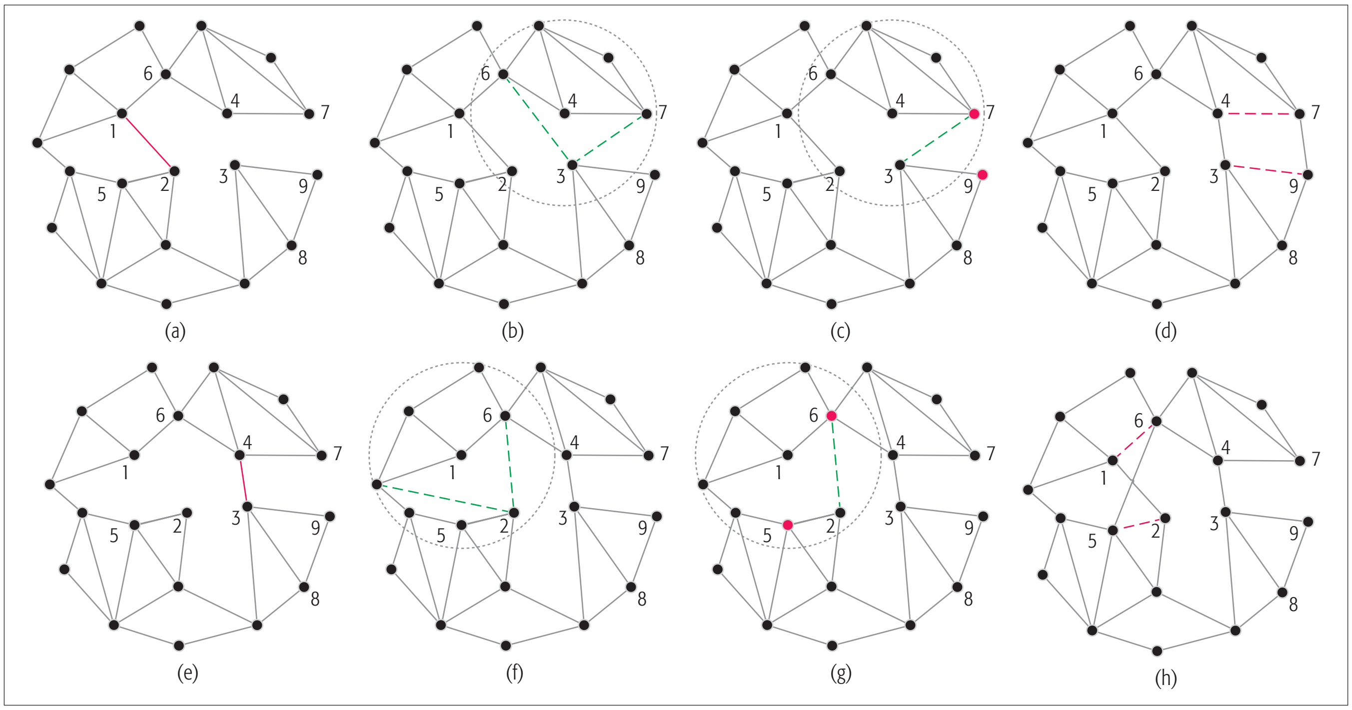

Figure 3. The process of the crossover operator: a) father; b) search in neighbors; c) swap edges; d) son; e) mother; f) search in neighbors; g) swap edges; h) daughter.

to pick the best values to prevent the populations from falling into the local optimum.

Each population is independent and communicates with other populations through the immigration operator. The immigration operator periodically moves the optimal individual that appeared on each population during evolution to other populations (every certain number of generations), which achieves the gene exchange among populations.

The best individual that appears in each generation is selected to compose the immigration population. The operations of crossover and mutation are not executed in immigration population, which ensures that the best individual of each population will not be destroyed. The immigration population is the foundation of the immigration operator, which increases the diversity of genetics and guarantees the fitness function to search for the optimal solution in a wide range.

To evaluate the network robustness, Schneider *et al.* [13] proposed a new metric, R. It considers the maximal connected subgraphs after removing the highest degree node repeatedly to measure the robustness of network topology, which means that the important nodes in IoT of smart cities are attacked. The value of R lies in the range (0, 0.5). We employ metric R to measure the robustness of the IoT topologies as the fitness function of MPGA. Furthermore, Herrmann *et al.* [14] have found that an onion-like structure is more stable and robust against malicious attacks. Thus, we make the evolution of individual topology toward the onion-like structure in the mutation operator to improve the robustness of topologies against malicious attacks. Basically, the connections among nodes in an onion-like structure exhibit the following characteristics:

- Nodes with similar degrees connect to each other.
- Node degrees gradually decrease from inner nodes to outer nodes.
- The majorities of the nodes have small degrees and are located in the outer layers of the onion-like structure.

Crossover Operator in MPGA

The crossover operator has no fixed evolution direction in a GA. Parent topologies generate new children topologies by the crossover operator, which obtains a larger solution space. Thus, the fitness function will search for the best solution in a larger space. Generally, the crossover operator retains a part of the father and mother genes, and eventually generates new children topologies. The crossover operator in this article keeps the initial degree of each node unchanged, which means that we cannot change the degree distribution of parent topologies. The degree distribution of new children topologies is the same as in the parent topologies.

Taking into account the limitation of communication range in smart cities, the crossover operator is designed as follows.

First, the parents are chosen by crossover probability P_c. We assume that the son topology inherits its father topology, and the daughter topology inherits its mother topology. Second, we get the sets of the father's exclusive edges and the mother's exclusive edges through the set of the father's edges and the set of the mother's edges. Here "exclusive" means that an edge only exists in one parent's set but not the other. Finally, the son topology disconnects the existing edges to build every mother's exclusive edges, during which the initial degree of each node is kept unchanged. The construction process of the daughter topology is similar to the above operation. Figure 3 illustrates the process of the crossover operator.

Figures 3a and 3e represent the connection between nodes in the father topology and mother topology. It can be seen that the father has an exclusive edge e_{12} between node 1 and node 2 in Fig. 3a, and the mother has an exclusive edge e_{34} between node 3 and node 4 in Fig. 3e. According to the criteria in the crossover operator, we build the mother's exclusive edge e_{34} in the son topology (Fig. 3d), and the father's exclusive edge e_{12} in the daughter topology (Fig. 3h).

Here is the detailed description about how father topology (Fig. 3a) generates its son topology (Fig. 3d). In order to generate a new edge e_{34} in Fig. 3a, we select the candidate nodes that have no edge with node 3 from the neighbors of node 4. Then we calculate the distance of the candidate nodes to node 3. Finally, we sort the distances to generate a candidate list in ascending order. As shown in Fig. 3b, node 7, which is a neighbor of node 4 and has no edge with node 3, is the nearest node to node 3. Node 3 searches each of its neighbor nodes in Fig. 3c until finds a node that is in the communication range of node 7 and has no edge with node 7. As shown in Fig. 3c, node 3 chooses its neighbor node 9, and we disconnect the edges e_{47} and e_{39} in Fig. 3d. After that we generate the edges e_{34} and e_{79}. Finally, we successfully generate a new edge, e_{34}, in the son topology (Fig. 3d). The degrees of node 3 and node 4 both equal 3 before the crossover operator, and after the operator they remain unchanged. Therefore, it is consistent with the criterion that keeps the initial degree of each node unchanged. The process that the mother generates its daughter topology is similar to the above operations, as shown in Figs. 3e–3g. Finally, we can see the father's exclusive edge e_{12} in the daughter topology in Fig. 3h.

Besides, when the father topology generates its son topology, if node 3 cannot find an eligible node to match node 7, which is the candidate neighbor of node 4, node 4 will sequentially choose another candidate node in the candidate list. And node 3 will search all of its neighbors for each candidate node until it finds an eligible node to match the candidate node. If node 3 still cannot find an eligible node after traversing the candidate list of node 4, we give up generating this edge.

Mutation Operator in MPGA

The mutation operator is an important way to generate new individuals in GA. We choose the individual by the mutation probability P_m. The goal of the mutation operator is to increase the robustness of the selected individual through exchange edges, during which the initial degree of each node is unchanged. Metric R is used to measure the robustness of topology. We search for the optimal solution within the local area by the mutation operator.

Nodes with similar degrees connect to each other in an onion-like structure. If a node with a large degree fails, another node with a large degree will replace its function. Therefore, we can minimize the adverse effects of failure nodes as much as possible, and the network topology will remain robust. In order to make the evolution of individual topology like the onion-like structure, we generate a new edge between two nodes that have similar degrees, during which the initial degree of each node is unchanged. We propose a criterion to sort degree and exchange edges as follows.

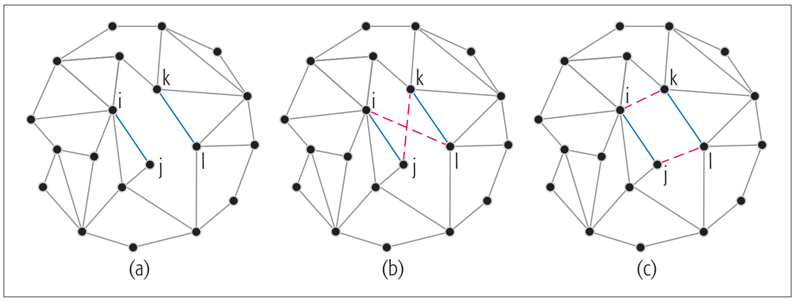

Figure 4. The candidate for the topology connection: a) initial topology; b) candidate 1; c) candidate 2.

We select two edges in the individual topology, and judge the four end nodes of these two edges as to whether they are in the communication range of each other to guarantee that we can generate a new edge in these four nodes. As shown in Fig. 4a, we select e_{ij} and e_{kl}. First, we sort the degrees of node i, node j, node k, and node l in descending order, and name them d_1, d_2, d_3, and d_4. Second, we add the absolute value of the difference between d_1 and d_2 to the absolute value of the difference between d_3 and d_4 as s_1. Then we add the absolute value of the difference between d_i and d_j to the absolute value of the difference between d_k and d_l as s_2. Finally, we let s_1 divide s_2 to get p and compare p with the exchange threshold φ. If p is less than φ, we exchange edges according to d_1, d_2, d_3, d_4. There are two candidate strategies in Figs. 4b and 4c.

Based on the criteria mentioned above, the nodes that have similar degrees will connect with each other, thus enabling the evolution of individual topology toward the onion-like structure. Besides, the exchange threshold φ is defined in [0, 1), and it cannot be 1 because the two edges will not be exchanged in that case. We control the efficiency of the mutation operator by changing the value of φ. The appropriate exchange threshold φ can effectively avoid inefficient exchange edges operation.

Simulation Results

In order to extract valid information of nodes in smart cities, we simulate deployment of IoT using Matlab. The nodes are deployed randomly in a circular area with diameter equal to 500 m. Considering that each node must have sufficient neighbors during the process of building initial topology, the communication range is set to 200 m. Then the node's geographic information and neighbor list are extracted for our algorithm. The parameters of our algorithm are obtained by many experiments. Finally, we set the optimal number of a population to 10, and the optimal number of individuals in a population to 20. Because more iterations means more time cost, we set iterations to 200 by considering various factors.

The threshold of exchange edges *pChange* slides from 0.1 to 1, and the sliding interval is 0.1. All the results of the experiment are taken from the average value of the $k(k \geqslant 10)$ times independent simulation experiments. The vertical axis

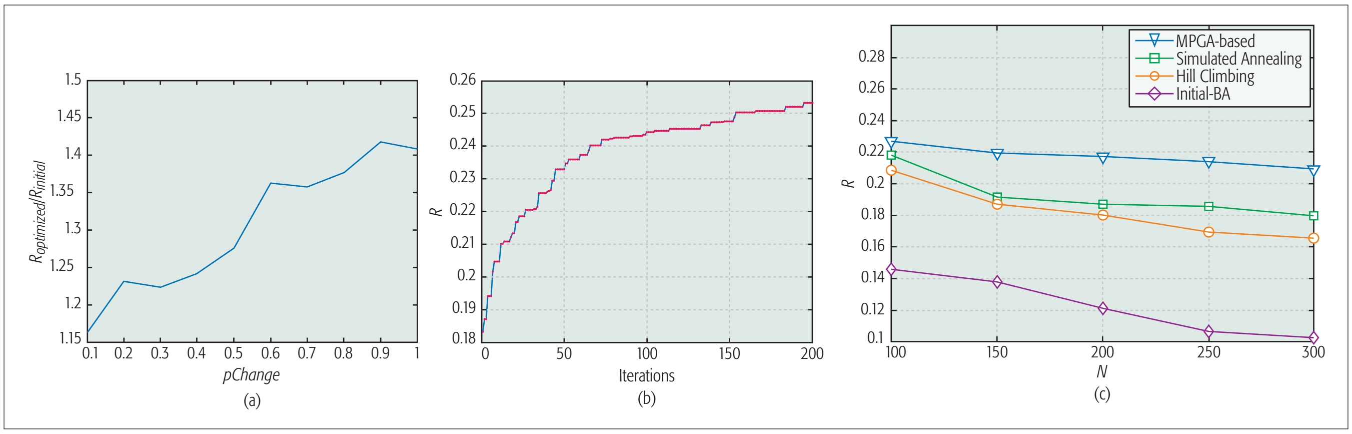

Figure 5. Performance of our algorithm: a) the threshold of exchange edges; b) the robustness evaluation of our algorithm; c) the compaction with other algorithms.

represents the ratio of the *r* value after optimizing to the initial *r* value. As shown in Fig. 5a, the value of the vertical axis changes with the different values of *PChange*. The curve shows a growing trend with the increase of *PChange*. As the curve grows, some local peaks can be seen, but the maximal value is gotten at 0.9. Thus, the value of *PChange* is set to 0.9 in the following simulation experiments.

Figure 5b illustrates that the metric *r* increases with the number of iterations of MPGA. At the beginning, the metric *r* of the initial topology is low; the value of *r* is increased obviously from the 1st to the 70th generation. After the 70th generation, the optimization result increases slowly due to the value of *r* having increased to a high level.

Based on the established initial IoT topology, we compare our algorithm to two existing algorithms, namely the Hill Climbing algorithm [14] and the Simulated Annealing algorithm [15]. Both of these algorithms keep the initial degree of every node unchanged. Figure 5c shows that the optimization results of the Hill Climbing algorithm, the Simulated Annealing algorithm, and our MPGA-based algorithm in different sizes of IoT topology. The size of topologies are set to 100, 150, 200, 250, and 300 nodes. The results are the average of k ($k > 10$) independent experiments, and each IoT topology remains connected after optimization. As shown in Fig. 5b, the value of *r* presents a downward trend with the increase of network sizes, and our algorithm has better performance than the other two algorithms.

Conclusion

Based on the big data of smart cities, nodes' geographic information and neighbor list are extracted. Then we construct the initial topologies of IoT. A novel MPGA-based algorithm is proposed to optimize the robustness of the network topology against malicious attacks.

We have designed two novel operators, the crossover operator and mutation operator. The initial degree of each node is unchanged during the process of these operators. Thus, the energy cost of adding edges will not be increased. Finally, we have simulated our algorithm and two existing algorithms. Their performance in improving the robustness of topologies is compared under different network sizes.

The experiment results show that the robustness of IoT topology against malicious attacks can be improved significantly by our algorithm. The values of *R* in the two existing algorithms reduce quickly with the size of networks increasing, but our algorithm still maintains the values of *R* at a high level. Thus, our algorithm can significantly improve the robustness of IoT in smart cities, especially against malicious attacks.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (Grant No. 61374154 and 61672131) and the Fundamental Research Funds for the Central Universities (DUT16QY27 and DUT17ZD216).

References

[1] J. A and Stankovic, "Research Directions for the Internet of Things," *IEEE Internet of Things J.*, vol. 1, no. 1, Feb. 2014, pp. 3–9.

[2] C. W. Tsai *et al.*, "Metaheuristics for the Deployment of 5G," *IEEE Wireless Commun.*, vol. 22, no. 6, Dec. 2015, pp. 40–46.

[3] T. Qiu *et al.*, "Heterogeneous Ad Hoc Networks: Architectures, Advances and Challenges," *Ad Hoc Networks*, vol. 55, no. 2, Feb. 2017, pp. 143–52.

[4] G. Han *et al.*, "Hysense: A Hybrid Mobile Crowdsensing Framework for Sensing Opportunities Compensation under Dynamic Coverage Constraint," *IEEE Commun. Mag.*, vol. 55, no. 3, Mar. 2017, pp. 93–99.

[5] G. Han *et al.*, "Green Routing Protocols for Wireless Multimedia Sensor Networks," *Sensors*, vol. 23, no. 6, June 2016, pp. 140–46.

[6] J. Wu *et al.*, "Spectral Measure of Structural Robustness in Complex Networks," *IEEE Trans. Systems, Man, and Cybernetics — Part A: Systems and Humans*, vol. 41, no. 6, Nov. 2011, pp. 1244–52.

[7] T. Qiu *et al.*, "Rose: Robustness Strategy for Scale-Free Wireless Sensor Networks," *IEEE/ACM Trans. Net.*, 2017. DOI: 10.1109/TNET.2017.2713530.

[8] R. H. Li *et al.*, "Measuring Robustness of Complex Networks Under MVC Attack," *Proc. 21st ACM Int'l. Conf. Info. and Knowledge Management*, Oct. 2012, pp. 1512–16.

[9] M. Zhou and J. Liu, "A Memetic Algorithm for Enhancing the Robustness of Scale-Free Networks Against Malicious Attacks," *Physica A: Statistical Mechanics and Its Applications*, vol. 410, no. 15, Sept. 2014, pp. 131–43.

[10] A. L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, Nov. 1999, pp. 509–12.

[11] W. Zhao, "Performance Optimization for State Machine Replication Based on Application Semantics: A Review," *J. Systems and Software*, vol. 112, Nov. 2016, pp. 96–109.

[12] W. Xiao, L. Lin, and G. Chen, "Vertex-Degree Sequences in Complex Networks: New Characteristics and Applications," *Physica A: Statistical Mechanics and Its Applications*, vol. 437, no. 11, Nov. 2015, pp. 437–41.

[13] C. Schneider *et al.*, "Mitigation of Malicious Attacks on Networks," *Proc. Nati'l. Academy of Sciences*, vol. 108, no. 10, July 2011, pp. 3838–41.
[14] H. J. Herrmann *et al.*, "Onion-Like Network Topology Enhances Robustness Against Malicious Attacks," *J. Statistical Mechanics: Theory and Experiment*, vol. 2011, no. 1, Jan. 2011, pp. 1–9.
[15] B. Pierre, D. Fabio, and T. Marco, "Optimizing the Robustness of Scale-Free Networks with Simulated Annealing," *Int'l. Conf. Adaptive and Natural Computing Algorithms*, Apr. 2011, pp. 167–76.

Biographies

Tie Qiu (qiutie@ieee.org)[M'11, SM'16] received his Ph.D. degree in computer science from Dalian University of Technology (DUT), China, in 2012. He is currently an associate professor in the School of Software, DUT. He is the founding director of the Smart Cyber-Physical Systems Laboratory (SmartCPS Lab). His research interests include the areas of embedded systems, cyber-physical systems, the Internet of Things, and mobile social networks.

Jie Liu (liujie.dut@gmail.com) received his B.E. from DUT in 2015. He is Master's student in the School of Software, DUT. He obtained an invention patent in 2016. He is a member of SmartCPS Lab. His research interests cover robustness optimization of the Internet of Things.

Weisheng Si (w.si@westernsydney.edu.au) is currently a lecturer in the School of Computing, Engineering and Mathematics, Western Sydney University. Prior to this, he was a postdoctoral researcher at National ICT Australia. He received his Ph.D., M.S., and B.S. degrees in computer science from the University of Sydney, the University of Virginia, and Peking University, respectively. His research interests include routing and topology control in wireless networks, graph theory, and data center networks.

Min Han (minhan@dlut.edu.cn) [M'95, A'03, SM'06] received her M.S. and Ph.D. degrees from Kyushu University, Fukuoka, Japan, in 1996 and 1999, respectively. Since 2003, she has been a professor in the Faculty of Electronic Information and Electrical Engineering, DUT. She is the author of four books and more than 200 articles. Her current research interests are neural networks, chaos, and their applications to control and identification.

Huansheng Ning (ninghuansheng@ustb.edu.cn) received his B.S. degree from Anhui University in 1996 and his Ph.D. degree from Beihang University in 2001. Now, he is a professor and vice dean of the School of Computer and Communication Engineering, University of Science and Technology Beijing, China. His current research focuses on the Internet of Things and cyber-physical modeling. He is the founder of the Cyberspace and Cybermatics International Science and Technology Cooperation Base.

Mohammed Atiquzzaman (atiq@ou.edu) is a professor of computer science at the University of Oklahoma. His research interests and publications are in next generation computer networks, wireless and mobile networks, switching and routing, optical communications, and multimedia over networks. He serves as the Editor-in-Chief of the *Journal of Network and Computer Applications* and the *Vehicular Communications* journal, and an Associate Editor of *IEEE Communications Magazine*, among others.