LETTER

A Statistical Reputation Approach for Reliable Packet Routing in Ad-Hoc Sensor Networks

Fang WANG^{†a)}, Member and Zhe WEI[†], Nonmember

SUMMARY In this study, we propose a statistical reputation approach for constructing a reliable packet route in ad-hoc sensor networks. The proposed method uses reputation as a measurement for router node selection through which a reliable data route is constructed for packet delivery. To refine the reputation, a transaction density is defined here to showcase the influence of node transaction frequency over the reputation. And to balance the energy consumption and avoid choosing repetitively the same node with high reputation, node remaining energy is also considered as a reputation factor in the selection process. Further, a shortest-path-tree routing protocol is designed so that data packets can reach the base station through the minimum intermediate nodes. Simulation tests illustrate the improvements in the packet delivery ratio and the energy utilization.

key words: statistical reputation, reliable packet routing, transaction density, shortest-path-tree routing

1. Introduction

Ad-hoc sensor networks are usually data oriented and individual sensor nodes in the network are equal peers serving as both data senders as well as data routers. There are normally two ways for data of interest to reach a base station (BS): 1) a sensor node senses and collects the data and sends them directly through one hop to the BS, provided that they are within the communication range; 2) or such data have to be indirectly passed or routed by the intermediate nodes and finally reach the BS. The indirect or multi-hop data transmission is a typical working mode for ad-hoc sensor networks in that many sensor nodes are placed far away from the BS and the router nodes have to be applied.

Although being collaborative, these (router) sensor nodes are autonomous and some of them may not always well-behaved, e.g., nodes with selfish behavior could deliberately drop some of the data packets so as to save their energy, or nodes of malicious nature route the data packets to a longer or nonexistent path to exhaust the system resources. In addition, sensor nodes are also easy to be captured and later become compromised and hostile. Thus to improve the packet delivery in ad-hoc sensor networks, node selection for reliable routing is an important issue to be addressed.

Due to the complex algorithm, traditional routing protocols relying on cryptography and authentication are not suitable for computing-constrained sensors. Further these

protocols mainly tackle the external attacks while malicious behaviors from the internal nodes cannot be effectively countered. Reputation as a soft security mechanism provides a solution to detect misbehaving nodes [1]. In this study, we use reputation as a measurement for router node selection through which a reliable data route is constructed for packet delivery. A transaction density is also defined to properly describe the influence of node transaction frequency over the reputation. To avoid the repetitive usage of the same node resulting in faster battery drainage, the residual energy is considered as a reputation factor in the selection process to balance the energy consumption. Additionally, a shortest-path-tree routing protocol is designed so that data packets can reach the BS through the minimum intermediate nodes.

2. System Model

2.1 Reputation Engine

Among reputation systems in sensor networks, statistics based reputation relying on historical behavior records of participating nodes has received wide attention. RFSN [2] is a classical representative of such a reputation method and many well-known literatures like [3], [4] and even some latest ones such as [5], [6] either directly or indirectly extend and modify the method used in RFSN.

Assume that in a packet routing cooperation, a sensor node i has the probability Θ to route a packet to another node, let α and β denote the historical number of successful and unsuccessful routing cooperations respectively, then according to RFSN, $P(\Theta)$ is defined by

$$P(\Theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \Theta^{\alpha-1} (1 - \Theta)^{\beta-1} \quad (1)$$

where $0 \leq \Theta \leq 1$, $\alpha \geq 0$, and $\beta \geq 0$. Θ can be used as the success probability in Bernoulli observations, and let $T \in [0, 1]$ denote the outcome of the next cooperation, then $P(T|\Theta)$ is defined by

$$P(T|\Theta) = \Theta^T (1 - \Theta)^{1-T} \quad (2)$$

After the cooperation is complete, the posterior of Θ is defined by

$$P(\Theta|T) = \frac{P(T|\Theta)P(\Theta)}{\int P(T|\Theta)P(\Theta)d\Theta} \quad (3)$$

Manuscript received July 14, 2018.

Manuscript revised October 11, 2018.

Manuscript publicized November 6, 2018.

[†]The authors are with the School of Computer Science, Civil Aviation Flight University of China, China.

a) E-mail: cafuc_wf@foxmail.com

DOI: 10.1587/transinf.2018EDL8152

and the mathematical expectation of Θ is defined by

$$E(\Theta) = \frac{\alpha + T}{\alpha + T + \beta + 1 - T} \quad (4)$$

In Eq. (4), $E(\Theta)$ is regarded as the reputation value of node i , and the shape parameter (α, β) can be interpreted as the historically observed number of positive outcomes (good behaviors) and negative outcomes (bad behaviors) respectively. When a sensor node's cooperation or transaction with other nodes is observed as a positive outcome, its parameter α will be incremented, otherwise parameter β is incremented. The reputation method used in RFSN is easy to be implemented, however, the main drawback is that due to the usage of only two reputation parameters, the cooperation results between nodes are only classified into either *good* or *bad*, *successful* or *unsuccessful*. Therefore, such a binary state instead of more states potentially limits sensor networks in many applications.

In our previous work [7], we proposed a negative multinomial reputation for self-organized networks. Its mathematical expression is as follows. In a sequence of independent repeated trials, $(A_1, \dots, A_t)$ are t different observations in each trail with the probability $P(A_i) = \xi_i$. Let τ_i denote the number of occurrences of A_i and X denote the total number of trials until the τ_i th occurrence of A_i . Then X follows the negative multinomial distribution and its probability function is defined by

$$P(X = n) = \frac{(\tau_1 + \dots + \tau_{t-1})(n-1)! \xi_1^{\tau_1} \dots \xi_t^{\tau_t}}{\tau_1! \dots \tau_t!} \quad (5)$$

$$f(\xi) = \frac{\Gamma(\alpha_1 + \dots + \alpha_t)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_t)} \xi_1^{\alpha_1-1} \dots \xi_t^{\alpha_t-1} \quad (6)$$

where α_i is the prior count of A_i and Γ is the gamma function. Then the posterior of ξ is defined by

$$P(\xi|X) = \frac{P(X = n)f(\xi_1, \dots, \xi_t)}{\int P(X = n)f(\xi_1, \dots, \xi_t)d\xi} \quad (7)$$

Thus the expectation of ξ is defined by

$$E(\xi) = \left(\frac{\alpha_1 + \tau_1}{\sum_{i=1}^t (\alpha_i + \tau_i)}, \dots, \frac{\alpha_t + \tau_t}{\sum_{i=1}^t (\alpha_i + \tau_i)} \right) \quad (8)$$

where $\tau_i = 0, 1, 2, \dots, t$ and $\tau_1 + \dots + \tau_t = n$.

In ad-hoc sensor networks, $E(\xi_i)$ can be regarded as the reputation of a certain transaction, α_i is the historical count of the transaction outcome, and τ_i is the current count of the transaction outcome. For example, in the observation about the data routing of node j , let R_1, R_2, R_3 , and R_4 respectively denote the reputation of j about the successful routing to the next predefined node, successful routing to the next node other than the predefined one, unsuccessful routing, and unknown routing resulted from the environmental interference. According to Eq. (8), R_k ($k = 1, 2, 3, 4$) is defined as

$$R_k = E(\xi_k) = \frac{\alpha_k + \tau_k}{\sum_{i=1}^4 (\alpha_i + \tau_i)} \quad (9)$$

In practice, R_1 is the favorable reputation to measure the trustworthiness of a sensor node when working as a router. But in view of the communication interference, R_4 representing the reputation of the unknown routing is incorporated so as to accurately describe R_1 . R_1 is redefined by

$$R = R_1 = \frac{\alpha_1 + \tau_1}{\sum_{i=1}^4 (\alpha_i + \tau_i)} + \sigma \frac{\alpha_4 + \tau_4}{\sum_{i=1}^4 (\alpha_i + \tau_i)} \quad (10)$$

where $\sigma \in (0, 1)$ is the certainty factor denoting the degree that R_4 should be added and it is tunable under different circumstances where sensor nodes are deployed.

2.2 Transaction Density

Some nodes can obtain proper reputation only through several transactions. For example, j and k are partners of i , $\langle j, i \rangle$ used to have 10 transactions with 6 positive outcomes and $\langle k, i \rangle$ had 100 transactions with 60 positive ones, thus to i , j and k have the same reputation 0.6 but with different number of transactions. Malicious nodes may take advantage of a few transactions to earn reputation and then launch selective forwarding attack in which all received packets are forwarded in certain transactions followed by forwarding some or no packets in one transaction. In case this happens, the transaction density is defined here as follows.

Definition 2.1 (Transaction Density): The transaction density TD is used to describe the degree of transaction frequency with a certain node. Suppose that node j is one of node i 's partners and the TD between i and j is defined by

$$TD\langle i, j \rangle = 1 - \lambda^{\frac{n}{m}}, \lambda \in (0, 1), n \leq m \quad (11)$$

where λ is a tunable constant, n is the number of transactions between i and j , m is the total number of transactions between i and the others. The aim of TD is to reasonably showcase the influence of transaction frequency over the reputation of sensor nodes and the greater the TD is, the more trusted the reputation becomes. Equation (10) is redefined by

$$R^{TD\langle i, j \rangle} = (1 - \lambda^{\frac{n}{m}}) \cdot \frac{\alpha_1 + \tau_1 + \sigma(\alpha_4 + \tau_4)}{\sum_{i=1}^4 (\alpha_i + \tau_i)} \quad (12)$$

2.3 Energy Factor

Under ideal circumstances, nodes of high reputation should be selected for task execution such as packet routing, but in reality the frequent use of the same node would result in faster battery drainage and later these nodes are usually discarded. To balance the energy consumption, the residual energy should be considered as part of the reputation.

According to [8], when transmitting k -bit data packet within distance d in ad-hoc sensor networks, the transmitter energy consumption $E_T(k, d)$ is defined by

$$E_T(k, d) = \begin{cases} kE_{elec} + k\epsilon_{FS}d^2 & d < d_0 \\ kE_{elec} + k\epsilon_{MP}d^4 & d \geq d_0 \end{cases} \quad (13)$$

where E_{elec} is the electronics energy such as signal coding and spreading, $\varepsilon_{FS}d^2$ and $\varepsilon_{MP}d^4$ are the amplifier energy in the free space fading channel (d^2 power loss) and multi path fading channel (d^4 power loss) respectively. If the distance d is less than the predefined threshold d_0 , power loss can be modeled as the Free Space model; or else, if d is greater than or equal to d_0 , power loss is modeled as the Multi Path model. And when receiving this k -bit data packet, the receiver energy consumption $E_R(k)$ is defined by

$$E_R(k) = kE_{elec} \quad (14)$$

Definition 2.2 (Energy Factor): The energy factor ψ is the ratio of the residual energy ER to the initial energy EI . Let ψ_j denote the residual energy of node j then

$$\psi_j = \frac{ER}{EI} = \frac{EI - E_T - E_R}{EI} \quad (15)$$

ψ can be treated as the energy reputation and is exchanged periodically among one-hop neighbors. A threshold T_ψ is set in the simulation tests so as to firstly check whether the node has enough energy reputation to get its job done. Further, with the transaction density and the energy factor, the reputation of node j held by node i is defined by

$$\mathcal{R}(i, j) = \omega_1 \cdot R^{TD(i, j)} + \omega_2 \cdot \psi_j \quad (16)$$

where $0 < \omega_1, \omega_2 < 1$ and $\omega_1 + \omega_2 = 1$

2.4 Shortest-Path-Tree Routing

A shortest-path-tree routing (SPTR) protocol is designed so that data packets can reach the BS through the minimum intermediate nodes. In SPTR, the BS periodically broadcasts a special message called beacon which mainly consists of three fields: $\langle$ BS ID, intermediate node ID, path counter $\rangle$. The path counter is a positive number set by the BS and the intermediate node ID is initially void. Any node that can receive the beacon is the one-hop neighbor to the BS and it fills in its ID, executes counter++, and rebroadcasts this beacon. Further, any node that receives the rebroadcasted beacon is considered as the two-hop neighbor to the BS and it copies down the second field of the received beacon, replaces this field with its own ID, executes counter++, and rebroadcasts this beacon again. In case a node receives multi rebroadcasted beacons, it compares the counters and selects the beacons with the minimum counter. This iteration process continues until all the nodes know its closest neighbor(s) to the BS.

2.5 Reliable Packet Routing

In our proposed method for reliable packet routing, suppose that sensor nodes are capable of bidirectional communication on every link and they work in a promiscuous mode so that nodes can overhear the ongoing packets from its neighbors, and that a sensor node only calculates the reputation

of its one-hop neighbors based on the perceived routing outcomes. The procedure of reliable packet routing is described as follows.

- Upon the packet request from the BS, a packet from a sensor node is either directly transmitted, or mostly indirectly routed towards the BS.
- Before a source node needs the help of packet routing, it will use its previous kept results of Eq.(16) to select a reputation qualified router node from its closest neighbors that are generated by the SPTR protocol.
- After that, the source node sends the packets to the selected router node, overhears or observes how many packets has been retransmitted or dropped by the selected router node, updates the reputation parameters about the selected router node accordingly, and uses Eq.(16) again to reevaluate the reputation of the selected router node.
- The selected router node will use the same procedure to choose its next router until the packet reaches the BS.

In this way, a reliable packet routing is formed by the reputation qualified router nodes.

3. Simulation

3.1 Assumptions and Settings

Suppose that the Free Space model is used, $EI = 0.5J$, $E_{elec} = 50nJ/bit$, $d = 1m$, and $\varepsilon_{fs}d^2 = 10pJ/bit/m^2$. The channel bandwidth is set to 1 Mb/s and a packet size is of 48 bytes including 36-byte data and 12-byte header. Each packet header also contains a hop counter recording the hop numbers from the source to the BS. 100 sensor nodes are evenly deployed in a circular area with a BS at the center. The BS has enough power to directly send request to every node and it launches 1000 queries to collect the data packets from these nodes over a fixed period of time. It is also assumed that there exist 30 evenly placed malicious nodes in the network that either launch selective forwarding attack by randomly dropping some or all the received packets, or deliberately misroute the received packets to a longer path against the SPTR protocol. The topology diagram is presented in Fig. 1 where the black nodes denote the malicious ones.

Further if a (router) node's reputation R^{TD} is less than 0.3, it will be no longer selected. For the tests, the certainty factor σ has a neutral value 0.5, the initial reputation of each node is 0.45, the energy threshold T_ψ is 0.00005J, and the NS-2 simulator is applied.

3.2 Packet Delivery Ratio

Normally, the BS cannot receive all the packets from the network mostly due to the packet selective forwarding by the malicious router nodes. Although some literatures like RFSN [2] successfully used reputation methods to stimulate nodes' cooperation such as packet forwarding, they did not

consider both the influence of transaction frequency over the reputation and the energy balance during the router node selection. In this part, we first define the packet delivery ratio and then select RFSN as a baseline for comparison with our proposed method. Test results are shown in Fig. 2.

Definition 3.1 (packet delivery ratio): The packet delivery ratio is the ratio of the actual received packets by the BS to the total packets generated by the network.

It can be noticed in Fig. 2 that as the queries from the BS increase, the packet delivery ratio, or PDR in RFSN always keeps fluctuating around 0.65. This is because in RFSN, malicious router nodes are well disguised by selectively forwarding and dropping some or all the received packets. In this way they maintain their reputation, survive in the network, and continue to attack the network without being easily detected, thus the BS in RFSN cannot receive all the packets from the network resulting in lower PDR. By contrast, the proposed method keeps a relatively higher PDR, because the transaction density is based on the transaction frequency with a certain node and the malicious router nodes have less successful transactions com-

pared with the ordinary router nodes. Thus the reputation of malicious router nodes will be weakened and later some of them are *isolated* from the network making more packets reach the BS.

In addition, Fig. 2 also presents the effects of the transaction density and the energy factor. To this end, four groups of parameters are set and compared in the proposed method. As can be seen in Fig. 2, when the parameter λ in the transaction density is fixed such as $\lambda = 0.1$, the more weight put on ω_1 and less on ω_2 in Eq. (16), e.g., $(0.1, 0.6, 0.4)$ vs $(0.1, 0.4, 0.6)$ result in lower PDR. It is because malicious router nodes in the proposed method can still survive for some time by owning a larger ω_1 at the expense of having a smaller ω_2 . It also means that these malicious nodes have to exhaust more energy so as to be reputation qualified and survive in the network. Then this is also the reason that the compared group $(0.2, 0.6, 0.4)$ has a lower PDR than $(0.2, 0.4, 0.6)$. Further, it can be noticed that when ω_1 and ω_2 are fixed, the compared group $(0.2, 0.4, 0.6)$ has a higher PDR than $(0.1, 0.4, 0.6)$ as is in the case of $(0.2, 0.6, 0.4)$ and $(0.1, 0.6, 0.4)$. This is because the larger the λ becomes, the more effective the transaction density is on the reputation of malicious router nodes, and the less the packets are dropped. But the high PDR comes at a cost, which will be analyzed in the following test.

3.3 Average Hop Count

Ideally, every data packet should reach the BS with the minimum hops or through an optimal path under the SPTR protocol. But in practice, some packets would undergo more hops due to the fact that they may detour around the low-reputation (malicious) router nodes or they have no choice but to be routed by the undetected malicious routers which then would route these packets to a longer or nonexistent path. This is the reason that in Fig. 3, the average hop count, or AHC in the proposed method is always higher than that in RFSN. It also helps to explain that a higher PDR comes at the expense of certain system resources like a higher AHC. The same result can further be confirmed by testing the four

Fig. 1 Network topology diagram 1

Fig. 2 Packet delivery ratio

Fig. 3 Average hop count

compared groups in the proposed method. For example, the compared group (0.2, 0.4, 0.6) not only has a higher ACH in Fig. 3 but also has a higher PDR in Fig. 2 than (0.1, 0.4, 0.6).

Thus to ensure the reliable routing, a tradeoff must be made between the PDR and the system performance such as AHC. A higher ACH might deteriorate the system by exhausting the energy of some nodes, but it does not necessarily mean that a higher ACH in the proposed method will damage the system energy utilization.

3.4 Energy Utilization

The energy utilization is the overall energy usage of a system and it usually involves the energy balancing among the system components. But to node selection for reliable routing, nodes with distinctive features such as high reputation will be repetitively selected for job completion leading to their faster battery drainage, while other nodes remain idle. To investigate the energy utilization of the compared methods, the number of viable or working nodes is tested and results are shown in Fig. 4.

In Fig. 4, the proposed method has a better viable node number as the query continues. This is because besides considering the node transaction reputation, node residual energy is also treated as a reputation factor during the selection process, e.g., a node with relatively lower energy would be selected for packet routing in later transactions, although this node has enough power to complete the current routing. Thus the energy consumption can be balanced among the sensor nodes, which means that more nodes have enough energy to survive and serve the network in the proposed method. Additionally, through the above three tests, it can be noticed that by properly setting the parameters in the proposed method, e.g. (0.1, 0.6, 0.4), we can get a high PDR, a low AHC, and a better number of viable nodes.

To further showcase the performance of the proposed method under different topology, the above three tests are carried out in a square area with the BS on one of the corners. The rest assumptions and settings about the test remain unchanged. The diagram of this topology is presented

in Fig. 5 and the test results are shown in Fig. 6, Fig. 7, and Fig. 8 respectively.

In Fig. 5, when the BS is placed far from most of the sensor nodes, data packets from those sensors have to be routed by many intermediate nodes so as to reach the BS,

Fig. 5 Network topology diagram 2

Fig. 6 Packet delivery ratio under topology 2

Fig. 4 Number of viable nodes

Fig. 7 Average hop count under topology 2

Fig. 8 Number of viable nodes under topology 2

which means that those packets have a higher probability of encountering more malicious router nodes. In Fig. 6, the test groups of the proposed method still remain a higher PDR than RFSN, but due to the BS location and higher probability of meeting more malicious router nodes, the PDR in both the compared methods is lower than that in Fig. 2. It can also be noticed that the PDR in RFSN is affected most under this scenario, e.g., in most cases, the PDR in RFSN is close to 60 in Fig. 6. For the same reason, in Fig. 7, the AHC in both the compared methods is higher than that in Fig. 3, which is around 10 in most of the time.

Further, the BS location in Fig. 5 also has an impact on the number of viable nodes. In Fig. 8, the test groups of the proposed method always keep a higher number of viable nodes than RFSN throughout the queries, but compared with Fig. 4, the number of viable nodes in both the compared methods is lower. This is because nodes close to the BS will get a heavy load of packet transmission, which results in their fast energy consumption.

4. Conclusion

Node selection is important for reliable and effective data routing in ad-hoc sensor networks. In this research, we propose a reputation method to help select the trusted next-hop router nodes through which packets can be reliably

forwarded to the BS. Simulation results show that the proposed method can achieve a better PDR and a finer energy utilization. But the PDR in the proposed method cannot reach 1 meaning that some malicious nodes can still hide themselves by selective packet dropping and forwarding. Such on-off attack countermeasures will be our future research. Further, the BS location affects the test performance of both the compared methods, thus the BS placement will also be studied later in detail.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China under Grant No. 61602066, the Scientific Research Program of CAFUC under grant Nos. J2018-03, F2017KF02, F2018KF07 and A201834.

References

- [1] V.K. Verma, S. Singh, and N.P. Pathak, "Towards comparative evaluation of trust and reputation models over static, dynamic and oscillating wireless sensor networks" *Wireless Networks*, vol.23, no.2, pp.335–343, Feb. 2017.
- [2] S. Ganeriwal, L.K. Balzano, and M.B. Srivastava, "Reputation-based framework for high integrity sensor networks," *ACM Trans. Sensor Networks*, vol.4, no.3, pp.15–37, May 2008.
- [3] S. Ozdemir, "Functional reputation based reliable data aggregation and transmission for wireless sensor networks," *Computer and Communications*, vol.31, no.17, pp.3941–3953, Nov. 2008.
- [4] P. Mukherjee and S. Sen, "Comparing reputation schemes for detecting malicious nodes in sensor networks," *The Computer Journal*, vol.54, no.3, pp.482–498, March 2011.
- [5] B.A. Ali, H.M. Abdulsalam, and A. Alghemlas, "Trust based scheme for IoT enabled wireless sensor networks," *Wireless Pers. Commun.*, vol.99, no.2, pp.1061–1080, March 2018.
- [6] T. Nguyen, D. Hoang, D. Nguyen, and A. Seneviratne, "Initial trust establishment for personal space IoT systems," *IEEE Conference on Computer Communications Workshops*, pp.784–789, 2017.
- [7] F. Wang and Z. Wei, "A statistics-based data fusion for ad-hoc sensor networks," *IEICE Trans. Fundamentals*, vol.E97-A, no.12, pp.2675–2679, Dec. 2014.
- [8] W.B. Heinzelman, A.P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol.1, no.4, pp.660–670, Oct. 2002.