Mo et al. Journal of Cloud Computing: Advances, Systems and Applications (2020) 9:51
https://doi.org/10.1186/s13677-020-00202-w

Journal of Cloud Computing: Advances, Systems and Applications

RESEARCH Open Access

An active and verifiable trust evaluation approach for edge computing

Wen Mo[1], Tian Wang[2], Shaobo Zhang[3] and Jinhuan Zhang[1*]

Abstract

Billions of Internet of Thing (IoT) devices are deployed in edge network. They are used to monitor specific event, process and to collect huge data to control center with smart decision based on the collected data. However, some malicious IoT devices may interrupt and interfere with normal nodes in data collection, causing damage to edge network. Due to the open character of the edge network, how to identify the credibility of these nodes, thereby identifying malicious IoT devices, and ensure reliable data collection in the edge network is a great challenge. In this paper, an Active and Verifiable Trust Evaluation (AVTE) approach is proposed to identify the credibility of IoT devices, so to ensure reliable data collection for Edge Computing with low cost. The main innovations of the AVTE approach compared with the existing work are as follows: (1) In AVTE approach, the trust of the device is obtained by an actively initiated trusted detection routing method. It is fast, accurate and targeted. (2) The acquisition of trust in the AVTE approach is based on a verifiable method and it ensures that the trust degree has higher reliability. (3) The trust acquisition method proposed in this paper is low-cost. An encoding returned verification method is applied to obtain verification messages at a very low cost. This paper proposes an encoding returned verification method, which can obtain verification messages at a very low cost. In addition, the strategy of this paper adopts initiation and verification of adaptive active trust detection according to the different energy consumption of IoT devices, so as to reliably obtain the trust of device under the premise of ensuring network lifetime. Theoretical analysis shows that AVTE approach can improve the data collection rate by 0.5 ~ 23.16% while ensuring long network lifetime compared with the existing scheme.

Keywords: Edge computing, Data collection, Active trust, Verifiable trust, Trust evolution

Introduction

With the development of Internet of Things (IoT), there will be more than 20 billion IoT devices by 2020 [1–3]. Most of these IoT devices are deployed at the edge of the network [4–6]. Due to the huge number of these IoT devices and the development of micro-processing technology, their computing and storage capabilities have greatly improved. For example, the computing and storage capacity of mobile phones now exceeds that of personal computers more than 10 years ago [7–9]. These huge changes have led to extremely huge computing and storage capacity of the edge network, so the current network computing center is transferred from the network center to the edge of the network, forming an emerging computing model such as Edge Computing (EC) or Fog Computing [1, 10–12]. Because most applications are based on the data sensed and acquired by IoT devices [13, 14], many emerging applications sensitive to latency and bandwidth, such as virtual reality, augmented reality and infrastructure for smart cities, benefit from edge computing. In the past, data needs to be sent to the cloud over a long distance, and the results of the cloud center are returned to the user through a long path. The edge computing is closer to the data source, which can reduce long delay, packet loss and large energy consumption [15–18].

* Correspondence: jinhuan_zhang@csu.edu.cn
[1]School of Computer Science and Engineering, Central South University, Changsha 410083, China
Full list of author information is available at the end of the article

However, the Edge network is an open network with various IoT devices in various forms connected [19–21]. For example, Luo et al. [4] gave a typical application of smart city as shown in Fig. 1. In smart city, a large number of IoT devices, most of which are wireless sensing devices, are deployed in various applications in the city to realize the perception of the surrounding environment [4, 9, 22, 23]. These sensing devices are relatively simple and do not have the ability to communicate directly with the Internet [4, 9, 24]. They are often self-organized into networks [25–27]. Sensing devices on the roadside act as gateway to collect the data of the entire network [4], and then send the collected data to the passing mobile vehicles [28–30]. Due to the advanced hardware of mobile vehicles, they can communicate directly with the Internet. Thus, the low-cost opportunistic routing method can establish data communication with the edge network [4, 9, 31, 32]. Such a data collection method has been widely applied in the current edge network due to its low cost [4, 9].

In such a network, many IoT devices are generally self-organized into a network [4, 33]. One or more nodes are selected as data collection nodes. The data collection nodes have different names in different networks [4]. In wireless sensor networks (WSNs), the data collection node is called sink [34, 35], and the data collection node is gateway in the application shown in Fig. 1 [4]. A common feature of these networks is that the data collection node is the center. All other nodes transmit their data to the data collection node through multi-hop routing [36, 37]. However, an important issue is the security of data collection [36–39]. Because of the network openness, many IoT devices can be added to the network autonomously [40, 41]. Thus, the malicious IoT devices will maliciously prevent the normal data collection. The most common is an attack called black hole [36, 42]. In such an attack, malicious nodes drop all packets forwarded by themselves to destroy the data collection [42]. The other is called selective forwarding attack (SFA), which is a smart attack [43]. In SFA attacks, malicious nodes are not as simple as black hole dropping data packets, but selectively dropping packets of some nodes [43]. Therefore, there is a certain packet loss rate in the wireless network. The malicious nodes can selectively drop some packets to effectively protect themselves from being discovered, so that they can initiate attacks at a critical time to cause longer-term and worse damage [43]. The data-based applications rely heavily on obtaining data consistency. The insecure behaviors such as the interception of data by malicious nodes can cause the loss of packets, which can cause the control center to make wrong decision in case of lack of data [43], further causing serious losses. Therefore, it is an important issue in the edge network to identify malicious nodes so as to clear them from the network or do not forward the data through them to avoid packet loss.

There are currently some related studies on protecting data collection. The studies are divided into two categories. The first type is targeted strategies to resist attacks. The main idea of the strategy is to take corresponding actions against attacks based on the characteristics of malicious nodes attacks, thereby invalidating the attack. For example, Liu et al. [42] proposed a Security and Energy-efficient Disjoint route (SEDR) strategy against black attack. In the SEDR scheme, data packets are divided into T shares through (T, M) -threshold secret-sharing algorithm, which use the same hop routing method to route to as far apart as possible, and then to the sink. As long as the M shares in the T shares can

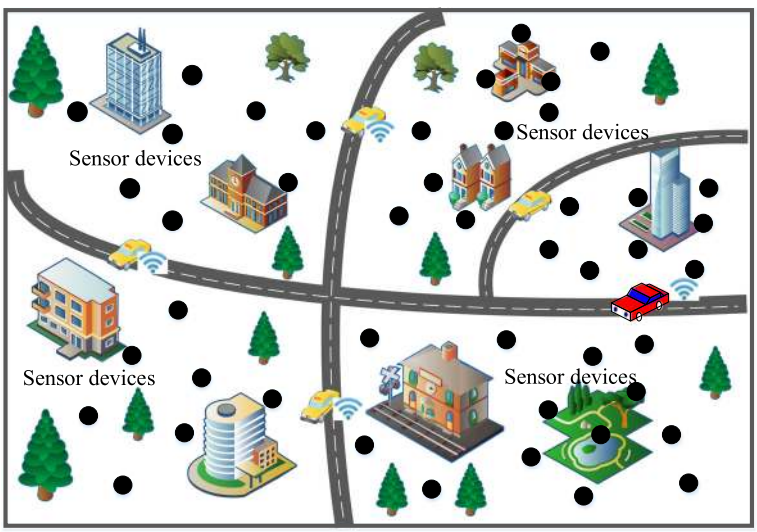

Fig. 1 The IoT devices in edge network for smart city

reach the sink safely, the packet can be successfully recovered. In this way, SEDR uses multiple separate routes to reduce the probability of being attacked by malicious nodes and to increase the probability of data successfully transmitted to the sink. The method is designed for a specific attack, so it is highly targeted and effective. However, the disadvantage is that the adaptability is narrow and it is generally ineffective for other attacks. Another shortcoming is the high cost of the strategy. There are two reasons for the high cost. First, the implementation cost of the strategy is high with the packet divided into T shares. Due to the redundancy among the T shares, the sink only needs to receive the M shares to recover the entire data packet. Obviously, there is at least T-M shares of redundancy. Because the main energy consumption of wireless nodes is caused by the data transmission. The strategy pays for additional energy consumption, which affects its lifetime. The other is that the data will be lost after the network attacked. Then, t the network will pay a huge price for wrong decisions.

The other type is a generally applicable method of defending against attacks. The most important method of defending against attacks is a strategy based on trust [5, 6, 11, 14, 30, 34, 36]. This type of strategy is not designed for a specific attack behavior, but adopts the corresponding data routing strategy to obtain the trust status of the node. If the behavior of the node conforms to the expected behavior, the node is considered to be trustworthy, and its trust is increased. Otherwise, its trust is reduced [36]. After obtaining the node's trust status, a node with high trust is selected as relay node to avoid the selection of malicious nodes when nodes transmit data, which can increase the probability of successfully data transmission to the sink [36]. Obviously, the behavior of trusted nodes is in line with expectations, and the behavior of dropping packets is not in line with expectations, thus the trust level is reduced. The strategy of using trust-based defense against attacks is not aimed at a specific attack, and it is effective for most network attacks. Therefore, it is a generally applicable method to resist attacks [36]. Relatively speaking, the cost of the method of obtaining trust is low.

The premise of trust-based attack defense methods is to obtain the trust of the node. However, how to effectively obtain the trust of the node is a challenge issue [36]. The main difficulties are as follows: (1) Observing the behavior of nodes is actually very difficult in the edge network. Because IoT devices have limited energy and hardware, their communication radius is small. Besides, they are deployed in some specific areas. It is difficult to observe the behavior of nodes forwarding data. Obviously, it is difficult to judge the trust degree of a node without observing the behavior of the node; (2) The node energy is very limited and the low lost trust system is needed to be designed.

Some strategies based on trust to resist attacks have been proposed. These studies are divided into two categories according to the way they gain trust: (1) Strategies for passive trust acquisition methods [6, 11, 30]. In this trust-based strategy, the system only observes the interaction behavior of nodes, but the system itself does not take action. These studies obtain the trust degree of nodes by observing the behavior of nodes, trust reasoning and evolution methods, and then take appropriate data collection strategies to avoid attacks according to the trust degree of nodes; (2) The other is the strategy of active trust acquisition [5, 34, 36]. Liu et al. first proposed an active trust acquisition method called ActiveTrust for WSN [36]. In the proposed method, the node actively initiates a detection route and the data packet of the detection route is empty. However, the attacker does not know that it is a detection route. When the attacker initiates an attack, it will be exposed, thereby reducing trust degree. Obviously, this method can be initiated actively according to the needs of the application, which accelerates the speed and accuracy of trust acquisition, and thus has higher efficiency [36].

Although active trust acquisition has some advantages, there is still room for further research. In this paper, an Active and Verifiable Trust Evaluation (AVTE) approach is proposed to identify the credibility of IoT devices, so to ensure reliable data collection for edge computing with low cost. The main innovations of the AVTE approach compared with the existing work are as follows:

(1) The AVTE method proposes an active trust detection strategy with active detection and feedback. The existing active trust detection strategy only initiates active detection without feedback. Thus, most of the trust acquisition occurs between neighboring nodes, which seriously affects the effectiveness of the strategy. The AVTE approach proposed in this paper requires the receiving node to return the information of the transmitted packet to confirm whether the packet received or not. Therefore, the trust acquisition of the AVTE approach is a verifiable method with higher reliability of trust acquisition ensured. It is different from the previous trust acquisition method which are not verifiable.
(2) The AVTE method proposes a method to return verification information of multiple data packets at one time to reduce the node energy cost of returning feedback. Previous methods do not return a verification message or the cost of returning verification is high. In the method proposed in this paper, the node returns the encoding of the

verification. We can know which packets successfully received by explaining the encoding. At the same time, the length of the encoding is the same as the information length of a packet, so the feedback information can obtain the verification message at a very low cost. In addition, the strategy of this article also uses initiation and verification of adaptive active trust detection according to the different energy consumption of IoT devices, so as to reliably obtain the trust of devices under the premise of ensuring network lifetime.

The rest of the paper is organized as follows. Related works are reviewed in Section 2. In Section 3, we describe the system model and formulate the problem of AVTE scheme. Sections 4 presents the detailed design of AVTE scheme. The proposed AVTE scheme is evaluated in Section 5. We conclude in Section 6.

Related works

Due to the rapid improvement of the computing, storage, and perception capabilities of IoT devices, IoT devices are widely deployed in various applications [44–46]. Although the storage and computing power of a single IoT device is relatively small [47], the massive amount of IoT devices deployed to the edge of the network have huge computing and storage capabilities. With more and more services are deployed to the network edge, current network is transferred from the cloud of the network to the edge of the network. Many applications are calculated on the edge of the network. Because the edge of the network is close to the data source, and the calculated result is also close to the user, so the data and results returned to the user do not need to go through a long transmission path like cloud computing, which brings users a higher Quality of Experiment (QoE) [48, 49]. Due to the current development of artificial intelligence technology [12, 31, 38, 39], the development of edge networks and edge computing is in the ascendant [48, 50]. Secure data collection is an important guarantee for the applications development [27, 34, 37, 51, 52]. Therefore, how to ensure the safety of data collection is an important research issue and there have been quite a lot of researchers focusing on it [53–55]. Some related research results are given below in this section.

This article divides the studies into two categories according to the targets of defending security attacks in data collection. One is the security strategies for specific attack behaviors [36, 37, 41, 43]. The other is the trust-based security data collection strategies [5, 6, 11, 14, 30, 34, 36]. Trust-based strategies are divided into active trust acquisition strategies and passive trust acquisition strategies.

Defense strategies for specific attacks

This type of data collection strategy for specific attack behavior based on the characteristics of the attack adopts corresponding defense strategies. The following gives several important types of security strategies proposed for specific security attacks.

(1) Strategies to resist dropping packet attacks

Black hole attack and SFA are the most widespread of these types of attacks [36, 42, 43]. Black hole attack is such an attack that malicious nodes drop all received data packets, which is similar to all data packets sent to malicious nodes as if entering black hole, so it is called black hole attack [36, 42]. SFA is a more difficult attack to resist. If all data packets are dropped, it is easy to find, so the survival time of malicious nodes is not long. In SFA, malicious nodes only drop a part of the packets selectively, even are the same as the normal node for a long period of time, but they drop important data in critical periods. In this way, malicious nodes not only protect themselves, but also bring harm to the network. It is also known as gray hole attack [43]. For the black hole attack, the SEDR strategy proposed by Liu et al. [42] is mentioned earlier. In fact, the strategy is to send packets to the sink through multi-hop routing at the same time. Even if some routes are attacked, as long as a route successfully reaches the sink, it can guarantee the packets collection. In short, the basic principle of this type of method is to send redundant packets. Whether it is the method of sending multiple slices or the method of multi-routes, the key idea is exploiting multiple routes. Even if one or more routes are blocked, it is still possible to guarantee the security collection of packets. Relatively speaking, SFA is more difficult to resist. Xiao et al. [43] proposed a more classic strategy named CHEMAS (checkpoint-based multi-hop acknowledgement scheme) to resist selective forwarding attacks. The method of the CHEMAS mechanism is to select a certain number of nodes as checkpoint nodes in the routing path from the source node to the sink. Once the checkpoint node receives the data packet, it returns an ACK packet to the upstream of the data source. The ACK packet contains a time to live (TTL) that the ACK can survive. Each time the ACK packet passes a detection node, its TTL number is reduced by 1. If the TTL is 0, ACK packet is discarded. After the node forwards the data, it waits for the arrival of ACK packets. If the node does not receive the expected number of ACK packets, it sends a warning message to the source node [43].

(2) Defend against Sinkhole attack [37]

Sinkhole attack is such an attack behavior. In WSNs, most data routing strategies rely on hop-based routing. Each node selects a node with a smaller hop count as the next hop node for data routing. In response to this, malicious nodes claim that they have a smaller number of hops to the sink, and broadcast their own hops to the sink. Therefore, neighbor nodes will choose malicious nodes as the next hop. Neighbor nodes then spread routing messages outward, so that the data of nodes within the scope of Sinkhole will be routed to malicious nodes, then malicious nodes drop the data collected by these nodes [37]. This causes the data of nodes within a certain range cannot be collected by sink. Liu et al. [37] proposed a better strategy for detecting and avoiding sinkholes. The main idea is to send a detection packet to sink every other detection cycle with a certain probability while the node normally sends data, and ask sink to return a confirmation packet. The confirmation packet must contain the sink's digital signature, confirming that its identity is true. If the node does not receive the return message of the true sink, it indicates that there is a sinkhole. Then take the far-sink routing method to avoid the influence of the sinkhole and to find the true sink. Then, notify the true sink to take measures to clear the sinkhole.

There are many security strategies for specific attack behaviors, such as how to resist clone attacks [48], and injection attacks [49], etc. Due to space limitations, they are not discussed one by one.

Trust-based security strategy

The main idea of this type of research is to evaluate the trust of nodes. The nodes that faithfully fulfill their commitments have a higher trust, while the nodes that behave badly are given a low trust evaluation, thereby avoiding the participation of low-trust nodes in the data collection process.

The total trust degree of a node is based on the integration of trust degree in the recent period. The simplest method used for the synthesis of trust degree is the average value of trust degree evaluation [36]. In practice, most studies use the principle of prioritization of recent information. The trust value closest to the current time is more important, and the further away from the current time the trust value becomes less important. Therefore, the trust value closer to the current time is given greater weight, and the trust value further away from the current time is given less weight. Integrating the trust degree weighting for a period of time forms a comprehensive trust degree evaluation [36].

Expanding the trust relationship is a method to enrich the trust relationship. The most commonly used mechanism is the reasoning and evolution of trust [5, 11, 14, 30]. Infer the unknown trust relationship through the existing direct trust relationship, which can enrich the trust relationship and make the trust evaluation scope wider. The main idea of this type of research is that trust is divided into two types. One type is direct trust [36], that is, there is direct interaction between nodes, and the trust evaluation of the other party is obtained according to the result of direct interaction. The other type is indirect trust [36]. Although there is no direct interaction between the two nodes, there are indirectly interacted nodes, so an indirect trust relationship can be established. The calculation of indirect trust generally adopts the principle of trust multiplication. The product of the trust of the nodes on the transfer path is the indirect trust. In this way, indirect trust is actually a way of decay. Nodes can evaluate the trust of many nodes in the network according to the reasoning and evolution mechanism of trust. Finally, select the collaborators based on the nodes of the trust evaluation. But the biggest disadvantage of this kind of method is that the accuracy of indirect trust is difficult to verify. In addition, in such a trust inference mechanism, the number of nodes that a node directly interacts with is small, and the reliability of trust reasoning decreases sharply as the level of reasoning increases, making the results unreliable. Therefore, nodes often only get few the trust status of some nodes, while the trust status of most nodes may not be obtained.

In traditional research, the acquisition of trust is a passive acquisition method [6, 11, 30]. This is because the trusted evaluator does not take any special actions to gain trust, but only observe the interaction behavior of the evaluated object [6, 11, 30]. Then, the trust evaluation value is given based on the interaction behavior of the evaluated object. The disadvantage of this passive trust evaluation method is that the obtained trust relationship is relatively small, it takes a long time to obtain the trust value, and it cannot be applied to the network with strong dynamic change. Moreover, if the acquisition speed of the trust value is less than the speed of the dynamic change of the network, the acquisition of the trust value becomes very limited. At the same time, this method cannot perform on-demand and timely trust evaluation of key evaluated objects.

At present, there are not many studies on active trust acquisition. Liu et al. [36] are the first researchers to propose active trust. The main idea of the proposed strategy is that the node actively initiates some probe routes with no data content, and if a node launches an attack on it, its trust is reduced, so that it can obtain the trust status of more nodes in a shorter time. However, in the strategy proposed by Liu et al. [36], data routing does not return information. Therefore, non-neighbor nodes do not know which node the data route has reached and the forwarding situation of other nodes, which leads to certain difficulties in its accurate and rapid

determination of the trust value. Based on the above analysis, this paper proposes an active trust acquisition strategy based on encoded return data routing information, which has better meaning.

System model and problem statement

The network model

The network model used in this article is similar to the network model of Ref. [36]. A certain number of wireless sensor nodes are deployed in a certain area, and there is a special node called sink to collect data of the entire network. The data of other nodes is transmitted to sink through multi-hop routing. But the network model in this paper is not only applicable to the typical planar wireless sensor network of active trust scheme, but also applicable to the network sites in many edge networks such as the linear network formed by many wireless sensor nodes deployed along oil pipelines in industrial applications, and the strip network deployed along river channels. At the same time, it is also suitable for the network formed by many IoT devices. For example, in the smart city proposed by Luo et al. [4], many IoT devices are deployed in various communities, and these IoT devices are self-organized into a network. As shown in Fig. 2, the nodes located on the side of the road act as gateways or sinks and are responsible for data collection for the entire network. Other IoT devices route their perceived data to the gateway node through multi-hop routing. When the mobile vehicles pass through the gateway communication range, the gateway node sends the collected data to the mobile vehicles, and the mobile vehicles send the received data to the edge network server to complete the data collection (Fig. 2).

The energy consumption model and related definitions use a typical energy consumption model. Eq. (1) represents the transmission energy consumption, Eq. (2) represents the reception energy consumption. E_{elec} represents the transmission circuit loss. The model uses free space and multipath fading (d^4 power loss) according to the distance between the transmitter and the receiver (d^2 power loss). ε_{f_s} and ε_{amp} are respectively the energy required for power amplification in the two models. The energy consumption for receiving l-bit packets is shown in Eq. (2). The above parameters are shown in Table 1.

$$\begin{cases} E_{member} = lE_{elec} + l\varepsilon_{f_s} d^2 \ if \ d \le d_0 \\ E_{member} = lE_{elec} + l\varepsilon_{amp} d^4 \ if \ d > d_0 \end{cases} \tag{1}$$

$$E_r(l) = lE_{elec} \tag{2}$$

The problem statements

The main objective of this paper is to design a feedback strategy based on data routing to improve the data collection rate. The purpose of returning the data routing information is to verify the correctness of the received data, thereby identifying malicious nodes, ensuring the safety of data collection. There are three main problems to be solved:

(1) Maximize the data collection rate

The data collection rate refers to the ratio of the number of data packets received to the total number of data packets sent. Malicious nodes in the network will drop the data packets, thus affecting the data collection. Considering that the number of data packets sent is M, and

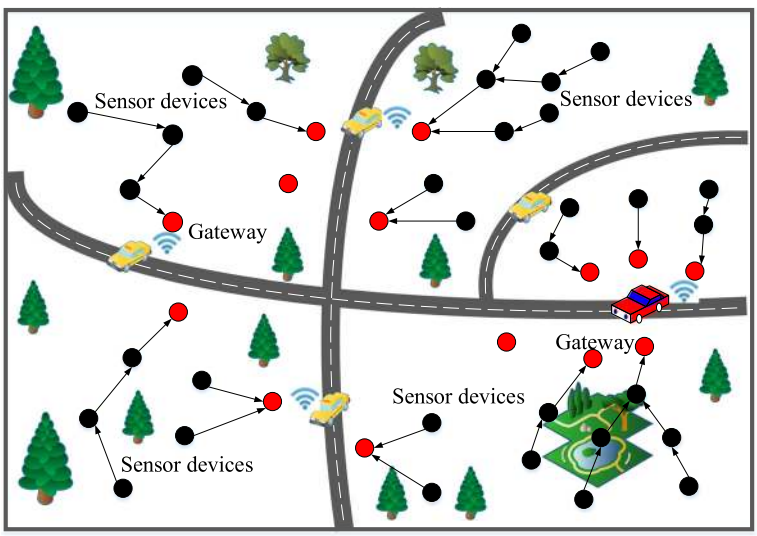

Fig. 2 The network model

Table 1 Network parameters

Parameter	Value
Threshold distance(d_0)(m)	87
Sensing range r_s (m)	15
E_{elec}(nJ/bit)	50
e_{f_s}(pJ/bit/m2)	10
e_{amp}(pJ/bit/m4)	0.0013
Initial energy(J)	0.5

the number of data packets that successfully reach the sink is N, the formula for maximizing the data collection rate μ can be expressed as:

$$\max(\mu) = \max\frac{N}{M} \tag{3}$$

(2) Trust acquisition

The AVTE strategy can accurately and quickly obtain the trust value of a node. The given trust value depends on the behavior of the node. For the trusted node, the trust value is high, while for the untrusted node, the trust value is low. When the trusted nodes generally show high trust, it means that the AVTE strategy can effectively identify the trust of the nodes. The average trust degree of trusted nodes is defined as ξ, which is used to reflect the ability of AVTE method to identify the trust degree of nodes. The average trust value is higher, the ability to obtain the trust value is stronger.

(3) Network lifetime is maximized

Network lifetime is related to energy. Reducing the energy consumption of nodes can extend the lifetime of the network. We define the death time of the first node in the network as the network lifetime. If the energy consumption of node i is E_i, the longest lifetime is expressed as

$$\max(\mathrm{T}) = \min\max(E_i) \tag{4}$$

In summary, the research goal of this paper is as follows:

$$\begin{cases} \max(\mu) = \max\frac{N}{M} \\ \max(\xi) \\ \max(\mathrm{T}) = \min\max(E_i) \end{cases} \tag{5}$$

The design of AVTE scheme

Research motivation

ActiveTrust scheme is one of the methods to effectively detect black hole attacks and obtain node trust. Nodes actively initiate detection routes and the data packets in the routes are empty. Empty packets routing will cause black hole attacks to expose the location of damaged nodes without causing loss or damage to data packets. However, there is room for improvement in this active trust scheme. This paper has improvements mainly in the following three aspects:

(1) Add a feedback mechanism between the source node and the routing nodes to detect the reliability of the data received by the node. By verifying whether the two packets are consistent, determine whether there are data packets loss, so as to evaluate whether the trust value is increased or decreased.
(2) The source node broadcasts an inquiry signal and sends it to all routing nodes, requiring the routing node to return feedback information after receiving all the data, instead of feeding back every time a packet received. Thus, the system pays a low price.
(3) Using the k data packets as feedback signals consumes more energy, we consider encoding k data packets to a packet as a feedback signal after k data encoding and XOR, thereby reducing energy consumption.

AVTE design

Figure 3 shows an overview of the AVTE scheme that includes detection routing, data routing, and feedback signals.

The source node sends m packets at time t, and the data is sent in binary encoding. Assuming that m data are respectively encoded as n-bit binary data, the node in the network receives the data. At time $t + a$, the source node broadcasts an inquiry signal, requiring all routing nodes to return a feedback signal to the source node to confirm whether the node has received the packet sent by the source node. Nodes encode and XOR the received k $(k \le m)$ packets, and the obtained result is fed back to the source node. The source node compares the feedback data routing information with the source data to determine whether they are consistent. If they are consistent, it means that the behavior of transmitting data is credible, which improves the trust of nodes, and vice versa. Through feedback routing, the trust of the nodes can be further improved. Select nodes with high trust for data transmission, thus ensuring the security of data collection and at the same time feeding back the results of k data XOR, which can reduce energy cost.

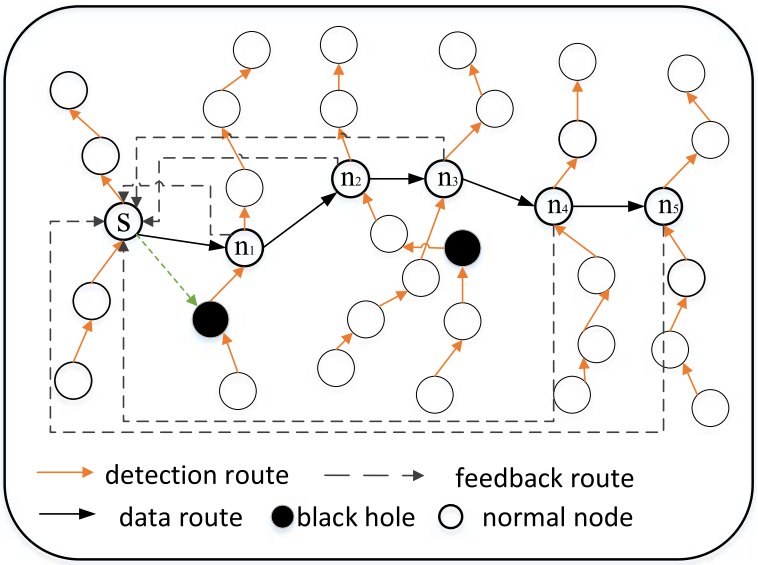

Fig. 3 Illustration of AVTE scheme

a. Encoding rules for source data: After receiving the XOR result, we must determine the unique composition of the target data and ensure that the result of the XOR is unique after taking any k of the m data.
b. The composition form of the feedback signal: the feedback signal includes not only k data encoding and XOR results, but also binary encoding of the received data by the routing node, which is then fed back to the source node as the prefix of XOR results This can increase the uniqueness of the target data determination.

Encoding method of feedback signal

Assuming that ε data is uniformly encoded as a ζ bit binary number $b_1b_2b_3...b_\zeta$, b_ζ takes the value 0 or 1. Each data is uniquely determined after encoding, and the data sent is $b_1b_2b_3...b_\zeta$, $b_3b_1b_\zeta...b_4$,..., $b_2b_3b_\zeta...b_4$, $b_\zeta b_1b_3...b_2$. Node receives k data. After the operation of XOR, we get $a_na_{n-1}a_0b_\zeta b_1b_3...b_2$, where $a_n = 0$ or 1. $a_na_{n-1}a_0$ is the binary encoding of the amount of data received. Assuming that $b_1b_2b_3...b_\zeta \oplus b_3b_1b_\zeta...b_4 \oplus ... \oplus b_2b_3b_\zeta...b_4 = b_\zeta b_1b_3...b_2$, it means that the node is reliable. If it does not exist, the following two reasons are discussed.

Case 1: If $a_n2^n + a_{n-1}2^{n-1} + ...a_02^0 < k$, it shows that the received data is less than k, there is data lost during the data transmission, and the trust value needs to be reduced.

Case 2: If $a_n2^n + a_{n-1}2^{n-1} + ...a_02^0 = k$, the node may indeed receive k data or may not. If a node receives k data, but the feedback signal or the code bit of the received data is lost when it is returned to the source node, this will cause the source node to make an error in the judgment of the amount of data received. Therefore, it cannot be verified successfully and the trust value also is reduced. If not, it is similar to case 1.

Theorem 1: Assuming that ε source data is represented by a binary digit of ζ, there is the most suitable value for ζ

$$\zeta = \lceil \log_2(\varepsilon + 1) - 1 \rceil \tag{6}$$

Proof: ζ bits are needed to be able to represent ε data completely. And ζ bits can represent $2^0 + 2^1 + 2^2 + ... + 2^\zeta = 2^{\zeta+1} - 1$ numbers, so $2^{\zeta+1} - 1 \geq \varepsilon$. Considering energy consumption, the length of ζ is not as long as possible, because the more digits the more energy is consumed. Therefore, ζ is rounded up and $\zeta = \lceil \log_2(\varepsilon + 1) - 1 \rceil$.

It should be noted that if the result is not unique when x data of ζ-bit XOR, the length of the code needs to be increased.

Theorem 2: Assuming that m data is sent, n bits represent the amount of data received. The amount of data received by each node is different, and the value of n is also different. The range of n is:

$$0 \le n \le \lceil \log_2(m+1) - 1 \rceil \tag{7}$$

Proof: n bits can represent $2^0 + 2^1 + 2^2 + \ldots + 2^n = 2^{n+1} - 1$ data, and a node can receive up to m data. $2^{n+1} - 1 \le m$, that is $n \le \log_2(m+1) - 1$, and n is an integer, so the maximum value of n is:

$$\max(n) = \lceil \log_2(m+1) - 1 \rceil \tag{8}$$

Therefore, the feedback signal is $n + \zeta$ bit. Once ζ is determined, each source data is ζ bits and ζ-bit data XOR is still ζ bits. Thus, the feedback signal of each node is fixed with ζ bits. By observing the previous n bits, we can know the amount of data received by the node. If all the sent data are lost, there is no XOR result, then $n = 0$.

We assume that the source node sends such a set of data: 001001,010111,011100,101011,111000, any k from the 5 data to form a group for XOR, the results are listed in Table 2.

Considering a situation in Fig. 4, we assume that the source node sends 2, 2, 3, 4, and 4 data to the nodes n_1- n_5, respectively, and the nodes n_1, n_2, n_3, n_4, and n_5 send 10010101, 10111100, 11001011, 11001111, 100001101 as feedback signals to source node.

Taking the verification node n_1 as an example, the signal 10010101 indicates that n_1 receives two data, and the result of the XOR of the two data is 010101. As shown in Table 3, we can know that 001001⊕011100 = 010101, from which we can know that node n_1 does not lose data, improving its trust. The verification status of the remaining 3 nodes is also shown in Table 3.

For node n_3, we can see that the feedback signal is abnormal, which shows that node receives 3 data, but there is no set of data that matches it. This is contradictory. Since 010111⊕011100=001011, we can conclude that node n_3 is likely to lose a data. Here we will not discuss the exact reason, but we can conclude that the n_3 node drops the data and the trust value will be reduced. Algorithm 1 is the detailed description of AVTE strategy.

Table 2 Results of XOR of k Data

k	Results of XOR of k Data
1	001001, 010111, 011100, 101011, 111000
2	011110,010101,100010,110001,001011,111100,101111,110111,100100, 010011
3	000010,110101,100110,100000,110011,001111,111110,101101,011010, 000100
4	101001,111010,001101,000110,011000
5	010001

Algorithm 1: Algorithm of AVTE Strategy

```
For node A receives k packets and generates a detection packet P Do
    Select the node B closest to the sink as the next hop
    Send packet P to B
    Assign a value to the length of the detection path
    For node A Do
        Binary encode the k data
        XOR the binary encoded information
        Send the feedback signal F_S to the source node
        If There are such k data XOR equal to F_S then
            improve the trust of node A
        Else
            reduce the trust of node A
        End if
    End for
    For node B receiving the detection packet Do
        length=length-1
        If length=0 then
            generate a feedback package q
            send q to the source node,restore the initial value
        End if
        If length≠0 then
            node B continues to select the next hop in the same way
        End if
    End for
End for
```

Calculation of trust value

When receiving a feedback signal, we can know whether the data received by the node is consistent with the data sent, so as to obtain the trust of the node. If the trust value is lower than the threshold, it is regarded as a malicious node, and this node will not be selected in the future routing. The neighbor node with high reliability will be selected to participate in the data routing to improve the security of data collection for the entire route. Algorithm 2 gives the data routing scheme.

Theorem 3 (change of trust value of a single node): the initial trust value of the node is Φ, and the change degree of the trust value is ϕ, then the trust value after transmitting data is:

$$\Gamma = \Phi \pm (1 - \Phi) \times \phi \tag{9}$$

Proof: 1-Φ represents the gap with full reliability, and multiplied by ϕ indicates the change of this gap. "+" means that the gap is reduced and the trust value is

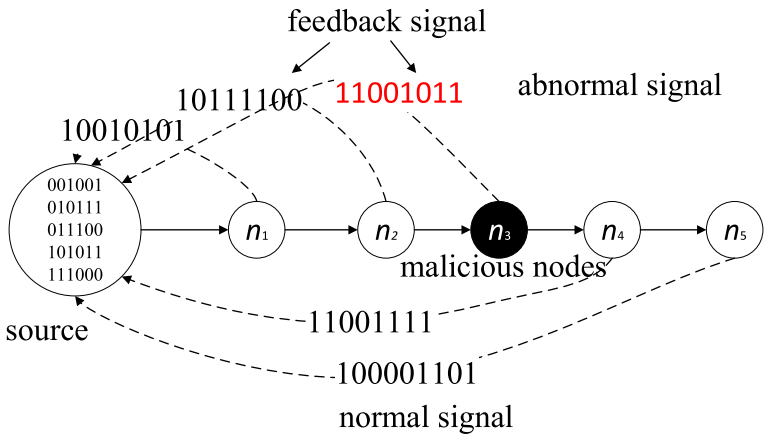

Fig. 4 The feedback signal verifies the consistency between the received data and the transmitted data

increased; "-" means that the gap is enlarged and the trust value is reduced. Therefore, "+" is taken when the verification is successful, and "-" is taken when the verification fails. We can know whether the verification is successful based on the feedback data, and then use Eq. (9) to calculate the change of node trust.

Algorithm2 : Data Routing Algorithms

1：**For** node A that generates or receives data packets **Do**
2： Select node B as the next hop
//node B has not been selected during the routing process, and has the highest trust value, the closest to the sink
3： **If** node A finds node B that meets the requirements as the next hop **then**
4： send the packet to node B
5： **If** node B is the sink **then**
6： this data route is complete
7： **End if**
8: **Else**
9: send a failed feedback signal to the upper node C
10: **End for**
11：**For** node C receiving the failed feedback signal **Do**
12： repeat operations 2-7 until finding a node that meets the requirements
13：**End for**

Algorithm 3: Algorithms for Calculating Trust

1：**For** node A **Do**
2： Assign an initial value to the initial trust value Φ of node A and the degree of change φ of the trust value
3: **If** node A is trusted **then**
4: $\Gamma = \Phi + (1 - \Phi) \times \varphi$
5: **Else**
6： $\Gamma = \Phi - (1 - \Phi) \times \varphi$
7: **If** Γ<Θ **then**
8: node A is no longer selected as a routing node
9： **End if**
10: **End if**
11：**End for**

Theorem 4 (trust value of sink): Assuming that there are l nodes on routing path, the trust value of the first node can be calculated by Eq. (9) and use the trust value of the first node as the initial value of the next node. So, the trust value of the second node can be calculated. By analogy, the trust value of the third, the l-1th, and the lth nodes, that is, the trust value of sink can be calculated.

$$\Gamma_1 = \Phi_1 \pm (1 - \Phi_1) \times \phi_1 \tag{10}$$

$$\Gamma_2 = \Gamma_1 \pm (1 - \Gamma_1) \times \phi_2 \tag{11}$$

$$\Gamma_n = \Gamma_{n-1} \pm (1 - \Gamma_{n-1}) \times \phi_n \tag{12}$$

Γ_n refers to the trust value of the node, Φ_1 is the initial trust value of the first node, and ϕ is the degree of change of the trust value of each node. Algorithm 3 shows calculation of the trust value.

The following is a specific example. Considering a routing path, the model is simplified as shown in Fig. 5.

Table 3 Data verification of n_1- n_5 nodes

node	n_1	n_2	n_3	n_4	n_5
feedback signal	10010101	10111100	11110100	11001111	100001101
k	2	2	3	3	4
source data	001001 011100	010111 101011	-- -- --	011100 101011 111000	001001 010111 101011 111000

We discuss the calculation of trust in three situations: successful data verification, failed data verification and random data verification. Finally, the trust value of the node to the sink is discussed.

Case 1 (data verification is successful): Assuming that the initial trust value of node 1 is $\Phi_1 = 0.5$, the degree of change of trust value is $\phi = 10\%$. First feedback signal shows that the data received by the node and the data sent by the source node are normal, then the trust value of node 1 becomes $0.5+ (1–0.5) \times 10\% = 0.55$, and the trust value increases by 0.05. If the data verification is successful in the second time, the trust value of node 1 becomes $0.55+ (1–0.55) \times 10\% = 0.595$, and the trust value increases by 0.045 on the original basis.

If each feedback signal indicates that the data received by node 1 is consistent with that of the source node, its trust value will continue to increase, and the trust value is close to 80% after 10 verifications. After multiple verifications, the trust value will be close to 1, and the reliability of the node is very high. When ϕ is changed, the trust value changes accordingly. When $\phi = 15\%$ or $\phi = 20\%$, the change of node 1 trust value is shown in Table 4. Figure 6 demonstrates the results of change of node trust value when the initial value is the same and the ϕ is variable.

Case 2 (data verification is failed): Assuming that the initial trust value of node 2 is 0.5 and $\phi = 10\%$. The first feedback signal shows that the data received by the node and the data sent by the source node are inconsistent, then the trust value of node 1 becomes $0.5- (1–0.5) \times 10\% = 0.45$, the trust value is reduced by 0.05. If the second feedback signal still cannot find the source data corresponding to the received data, the trust value will be reduced again to $(1–0.45) \times 10\% = 0.055$, the trust value will drop to 0.0256 after 7 times, which is very close to 0. At this point, the node is already very unreliable. If we set the threshold $\Theta = 0.2$, the trust value has fallen below 0.2 after 5 rounds of data transmission, and node will not be selected in the future routing process. Considering its neighbor node, calculate the trust value of node 2 to decide whether it can be selected as the next hop node.

It can be seen from Table 5 and Fig. 7 that changing the initial value and ϕ, trust value is different. When $\phi = 15\%$, the trust value of node 2 is reduced from 0.8 to 0.1909(< 0.2). If the 11th data verification fails, the trust value will be reduced to 0.0695, so we can get the trust value of the node through 10 rounds of data transmission, providing a basis for the selection of routing nodes. By comparing with the threshold, we can also know that the node can participate in several rounds of data transmission at most. For example, when $\phi = 15\%$, the trust value of the first four rounds of data transmission is greater than 0.7, so node 2 can still participate in data transmission with high reliability.

Case 3 (data verification is random): Assuming the initial trust value of node 3 is $\Phi_1 = 0.2$, $\phi = 10\%$. The first

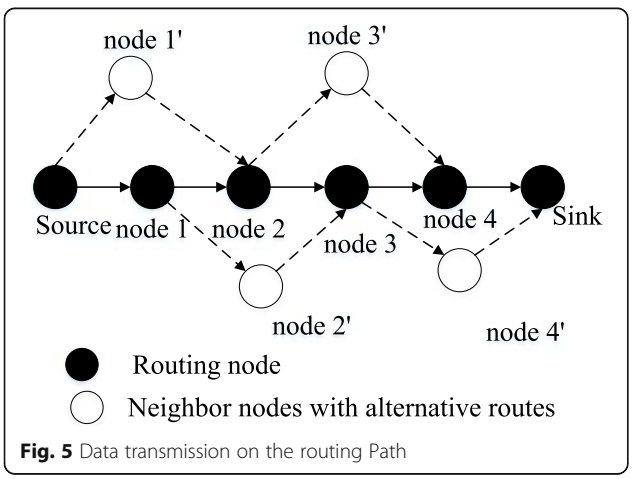

Fig. 5 Data transmission on the routing Path

Table 4 Change of node 1 trust value in Case 1

n	$\varphi = 10\%$	$\varphi = 15\%$	$\varphi = 20\%$
0	0.5	0.5	0.5
1	0.55	0.575	0.6
2	0.595	0.6388	0.68
3	0.6355	0.6929	0.744
4	0.6720	0.7390	0.7952
5	0.7047	0.7781	0.8362
6	0.7343	0.8114	0.8689
7	0.7608	0.8397	0.8951
8	0.7847	0.8638	0.9161
9	0.8063	0.8842	0.9329
10	0.8256	0.9016	0.9463

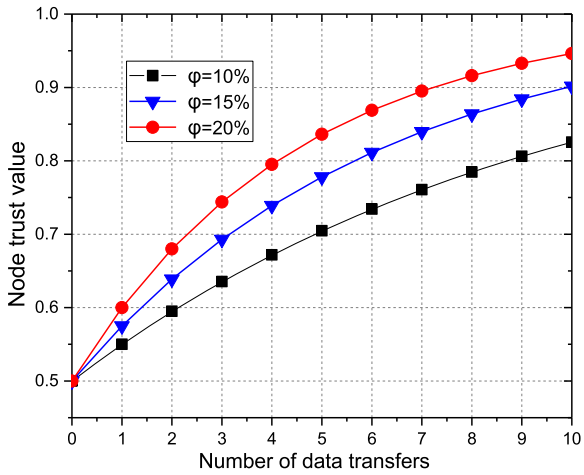

Fig. 6 Change of trust value if verification is successful

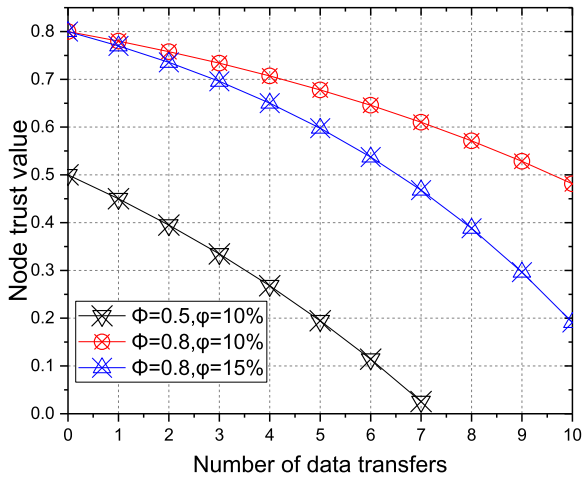

Fig. 7 Change of trust value if verification failed

feedback signal shows that the data received by the node and the data sent by the source node are normal, then the trust value of node 1 becomes 0.5+ (1–0.5) × 10% = 0.55, and the trust value increases by 0.05. When the second data transmission fails verification, the trust value of node 3 becomes 0.55- (1–0.55) × 10% = 0.505, the probability of successful verification each time is random, and the trust value after 10 rounds of data transmission is shown in Table 6. Figure 8 presents the change in node trust value after different rounds of data transmission and if verification is random.

Case 4 (consider all the nodes on the routing path in Fig. 5 and calculate the trust value of the sink after a round of data transmission): Assuming that the initial value of node 1 is 0.5 and ϕ = 90%. The probability of each data verification success is random, we assume that the data received by node 1 during the first data transmission is correct and passes to the neighbor node, the initial trust value of node 2 becomes 0.5+ (1–0.5) × 90% = 0.95. If node 2 data verification fails and the data continues to be passed to node 3, it will affect the trust value of node 3, and so on. We can get the reliability of the data when it reaches the sink. The trust value of sink is shown in Table 7.

Therefore, the trust value of the data transmitted through this routing path to the sink becomes 0.9123. This is the result of a round of data transmission. After each round of data transmission is completed, the trust of the data that reaches the sink will change. We calculate the results after ten rounds of data transmission as shown in Table 8, and the trust of the data arriving at the sink is shown in Fig. 9.

Performance analysis of AVTE scheme

Trust acquisition

The AVTE method can not only identify malicious nodes, but also accurately obtain the node trust degree. In the active trust scheme, the node that does not attack the detection route is evaluated as a good node, and the trust state of the node is qualitatively obtained. However,

Table 5 Change of node 2 trust value in Case 2

n	φ = 10%	φ = 10%	φ = 15%
0	0.5	0.8	0.8
1	0.45	0.78	0.77
2	0.395	0.758	0.7355
3	0.3345	0.7338	0.6958
4	0.2680	0.7072	0.6502
5	0.1947	0.6778	0.5977
6	0.1142	0.6457	0.5374
7	0.0256	0.6103	0.4680
8	–	0.5713	0.3882
9	–	0.5284	0.2964
10	–	0.4813	0.1909

Table 6 Change of node 3 trust value in Case 3

n	Φ = 0.5φ = 10%	Φ = 0.8 φ = 8%	Φ = 0.4 φ = 12%
0	0.5	0.8	0.4
1	0.55	0.784	0.472
2	0.505	0.76672	0.5354
3	0.5545	0.7854	0.4796
4	0.50995	0.8026	0.5385
5	0.4609	0.7868	0.4831
6	0.4070	0.8038	0.4211
7	0.4663	0.7881	0.4905
8	0.5197	0.7712	0.5517
9	0.4717	0.7895	0.4979
10	0.5245	0.8063	0.5581

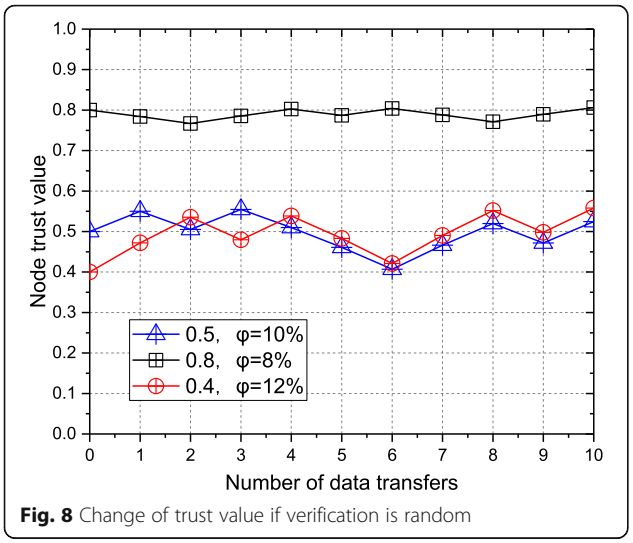

Fig. 8 Change of trust value if verification is random

the AVTE method gives the trust degree of the node quantitatively, which provides a more accurate standard for selecting the next hop.

According to the criterion of high trust given by trusted nodes and low trust given by untrusted nodes, the average trust degree of trusted nodes is used to evaluate the ability of AVTE scheme to identify trusted nodes. The trust degree of these trusted nodes is generally high, so the average trust degree is also high. Based on the calculation of the trust degree in Section 4.5, we consider that there are n_1, n_2, ..., n_m nodes in total, among which h nodes of n_2, n_5, ..., n_d, n_q are not malicious nodes. The given trust degrees are T_2, T_5, ..., T_d, T_q, then the average trust degree is:

$$\xi = \frac{(T_2 + T_5 + \ldots + T_d + T_q)}{h} \tag{13}$$

If the average trust level is below 0.5, it means that the AVTE strategy has an error in identifying the trusted node. The higher the average trust degree, the higher the security of trusted nodes, and the stronger the ability of AVTE to identify the credibility of nodes.

With the number of detections increases, the trust value of good nodes will continue to increase, the average trust value will also increase, and the trust value of bad nodes will decline. Figure 10 shows that after 11 data verifications, the average trust degree of trusted

Table 7 Trust values of sink

node	n_1	n_2	n_3	n_4	n_5
results	S	F	S	F	S
Φ	0.5	0.95	0.91	0.9045	0.9026
φ	90%	80%	50%	20%	10%
Γ	0.95	0.91	0.9045	0.9026	0.9123

Table 8 Trust value of sink after 10 rounds of data transmission

rounds	n_1	n_2	n_3	n_4	n_5
1	0.95	0.91	0.9045	0.9026	0.9123
2	0.905	0.981	0.9715	0.9772	0.9749
3	0.8195	0.9639	0.9820	0.9729	0.9702
4	0.9820	0.9675	0.9513	0.9415	0.9473
5	0.9657	0.9383	0.9691	0.9753	0.9728
6	0.9348	0.8827	0.9414	0.9296	0.9367
7	0.9935	0.9987	0.9980	0.9977	0.9974
8	0.9876	0.9777	0.9666	0.9733	0.9759
9	0.9765	0.9953	0.9976	0.9972	0.9969
10	0.9976	0.9958	0.9936	0.9924	0.9916

nodes is continuously rising and higher than 0.5, indicating that the trust value obtained by the AVTE method is effective.

Data collection rate

There are detection routes, data routes, and feedback routes in the edge network at the same time. The detection routes are responsible for identifying malicious IoT devices, the data routes are responsible for sending data to the sink, and the feedback routes are responsible for returning data routing information to the source node. When the data route transfers the data of node n_i to node n_j, the detection route detects two nodes are not black nodes. The trust of node n_i to node n_j is called direction trust. Direction trust refers to the trust relationship established by two nodes directly transmitting data. At the same time, there is also indirect communication between nodes in the network. Data is transmitted to another node through an intermediate node. The trust relationship established at this time is called indirect trust. In the edge network, nodes mostly establish direct

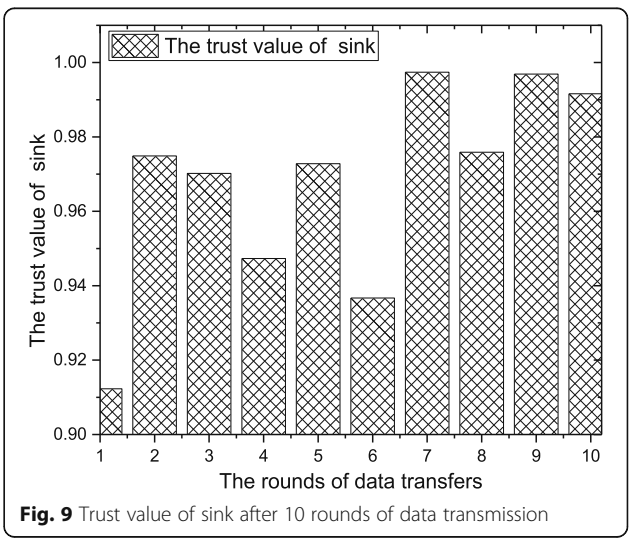

Fig. 9 Trust value of sink after 10 rounds of data transmission

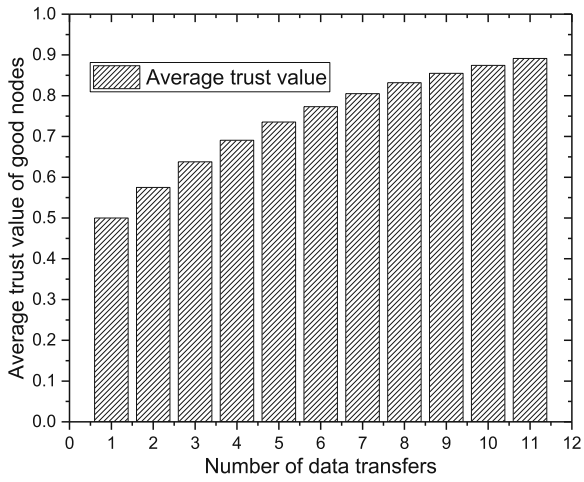

Fig. 10 Average trust value of trusted nodes

trust through direct interaction, and the direct trust obtained is more accurate and reliable. Therefore, the network relationship is simplified in our manuscript, and Fig. 3 only shows the relationship between nodes directly transmitting data to nodes. Therefore, Theorem 5 and Theorem 6 in our manuscript only consider direction trust, which has no effect on the result of data collection rate.

After receiving the data, the node will choose the neighbor node closer to the sink as the next-hop node. Therefore, detection routes need to detect whether the neighbor node is a black node. When all its neighbor nodes are black nodes, it means that the transmission fails.

Theorem 5: Considering the number of hops from the source node n_i to sink is ω, the number of nodes is α, and the ratio of malicious nodes is λ. If only direction trust is considered, the number of nodes with direction trust is β, then the data collection rate of the sink is

$$\begin{cases} \Psi_i = \left(1-\lambda^{\alpha/3}\right)^{\omega-1} & \beta \geq \alpha \\ \Psi_i = \left(1-\lambda^{\beta/3+1}\right)^{\omega-1} & \beta < \alpha \end{cases} \tag{14}$$

Proof: First, calculate the success rate of single hop transmission of any node A. The failed transmission means that node A finds that all of the detected nodes whose hop number smaller than its own are black holes. The detected nodes cannot be selected, and A must select from the undetected nodes. If the selected undetected node is a black hole, the transmission fails.

Therefore, the failure probability is as follows. There are three states for node A, that is, more than, less than, and equal to the hop count of node A. For the number of nodes α, the number of nodes whose hops are smaller than A's is $\alpha/3$. If the number of nodes with direction trust is β, there are β detections in total. For nodes with hop count less than node A, the total detections is $\beta/3$.

If $\beta \geq \alpha$, all neighbors of node A can be detected. The proportion of malicious nodes is λ, and the probability that all detected nodes are malicious nodes is $\lambda^{\alpha/3}$, so the probability of transmission failure is $\lambda^{\alpha/3}$.

If $\beta<\alpha$, all neighbor nodes cannot be detected. The probability that the detected nodes are malicious nodes is $\lambda^{\beta/3}$. The probability of being a malicious node at the next hop node is λ, so the failure probability is $\lambda\lambda^{\beta/3} = \lambda^{\beta/3+1}$.

The source node n_i has ω hops to sink. Considering that the last hop is not a malicious node, the probability that the sink successfully collects data is

$$\begin{cases} \Psi_i = \left(1-\lambda^{\alpha/3}\right)^{\omega-1} & \beta \geq \alpha \\ \Psi_i = \left(1-\lambda^{\beta/3+1}\right)^{\omega-1} & \beta < \alpha \end{cases}$$

Theorem 6: The number of hops from source node n_i to sink is ω, the number of nodes is α, and the ratio of malicious nodes is λ. Only the direction trust is considered, the number of nodes with direction trust is β, and the ratio of malicious nodes changes to γ, then the probability that the sink successfully collects the data packet is

$$\begin{cases} P_i = \left(1-\lambda(1-\gamma)^{\alpha/3}\right)^{\omega-1} & \beta \geq \alpha \\ P_i = \left(1-\lambda(1-\gamma)^{\beta/3+1}\right)^{\omega-1} & \beta < \alpha \end{cases} \tag{15}$$

Proof: The feedback mechanism verifies the data received and the data sent. If the packets are inconsistent, the trust value is reduced. If the trust value drops below the threshold, the node is a malicious node. The active trust scheme can detect whether neighbor node is a malicious node, our scheme can further detect malicious nodes on the data route, so our scheme makes the ratio of malicious nodes change from the previous λ to λ (1-γ). The rest of the proofs are as in Theorem 5.

Theorem 7: The source node sends the data to the sink through ω hops. When using the shortest path protocol, the data collection rate is

$$\eta = (1-\lambda)^{\omega-1} \tag{16}$$

Proof: When using the shortest path protocol, the nodes are randomly selected, and the probability that these selected nodes are black nodes is λ. The last hop is not a black node, so the probability of choosing a non-black node after ω hop $\eta = (1-\lambda)^{\omega-1}$.

Figure 11 shows the data collection rate when the ratio of malicious nodes is different in the AVTE scheme and the ActiveTrust scheme, where α =6, β =8, ω =15. Figure 12 is the data collection rate comparison of the

AVTE scheme and the ActiveTrust scheme under different hop counts, where $\gamma = 0.2$, $\alpha = 6$, $\beta = 8$, $\lambda = 0.2$. Figure 13 presents the data collection rate of the two schemes when the number of nodes with direction trust is less than the number of nodes, where $\gamma = 0.2$, $\alpha = 9$, $\beta = 6$, $\omega = 10$. From the above three figures, we can see that the data collection rate of the AVTE scheme is higher than that of the ActiveTrust scheme. The performance of the AVTE scheme is improved comparing with the active trust scheme.

Figure 14 shows the probability of successful data collection of ActiveTrust scheme, AVTE scheme and the shortest path. The shortest path has the lowest data collection rate. When $\omega = 15$, $\lambda = 0.2$, the total data collection rate drops below 0.1. The ActiveTrust scheme and the AVTE scheme have maintained a high success rate (> 60%). And the AVTE scheme is superior to the ActiveTrust scheme, because the active trust can identify all black nodes. The next hop only needs to select a good non-black node for routing, but the reliability of all non-black nodes is different. Adding a feedback mechanism can compare the reliability of neighbor nodes that are not black nodes, so as to provide non-black next-hop routing nodes with higher trust for data routing. Therefore, the AVTE scheme has higher data collection rate than the ActiveTrust scheme.

Figure 15 shows the ratio of the data collection rate of AVTE, the ActiveTrust scheme and the shortest route scheme when the number of malicious nodes is different. It can be seen that the AVTE scheme and the ActiveTrust scheme have a significant improvement over the shortest route. With the increase of malicious nodes, the AVTE scheme has improved the performance by more than 8 times and the ActiveTrust scheme has improved by more than 6 times.

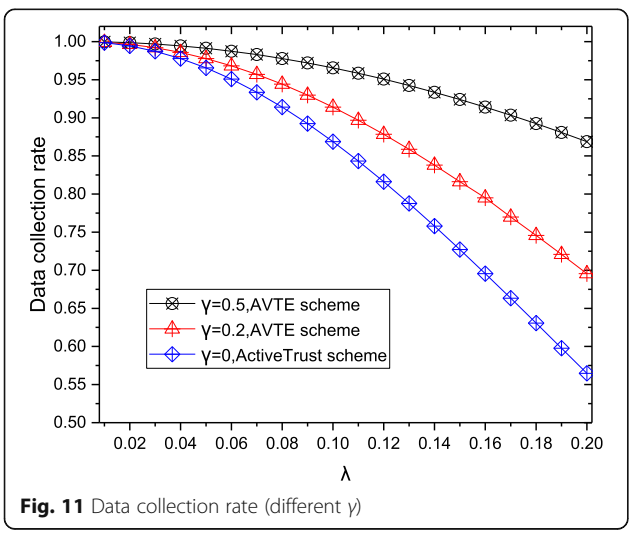

Fig. 11 Data collection rate (different γ)

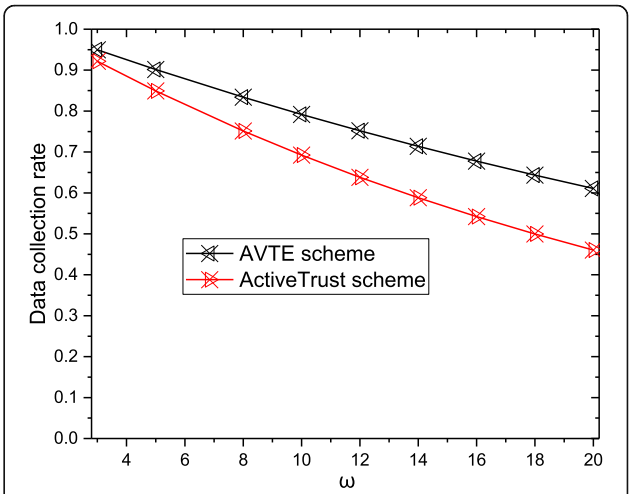

Fig. 12 Data collection rate (different ω)

Compared with the ActiveTrust scheme, the ratio of data collection rate of AVTE scheme remains above 1 and increases slightly. This is because the greater the ratio of malicious nodes, the ActiveTrust scheme can effectively detect the location of malicious nodes and avoid their application in data routing, so that the success rate of data packets to sink is greatly increased. On the basis of ActiveTrust scheme, the AVTE scheme detects black nodes that have been used as routing nodes in data routing and selects nodes with higher trust as next hop routing nodes, once again increasing the data collection rate.

Energy consumption and network lifetime

The energy of nodes is mainly consumed in sending and receiving, detecting and confirming packets. The network lifetime is defined as the death time of the first node. When each node needs to send confirmation

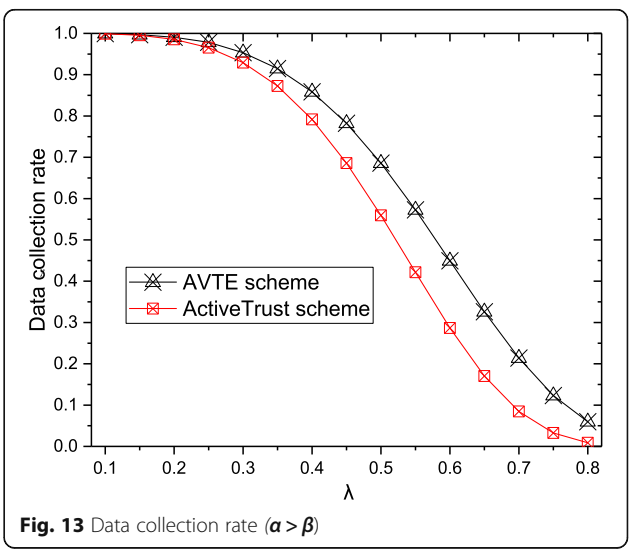

Fig. 13 Data collection rate ($\alpha > \beta$)

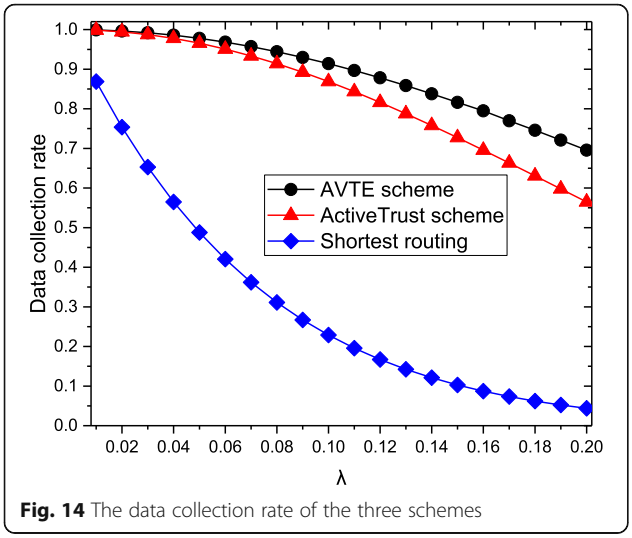

Fig. 14 The data collection rate of the three schemes

packet to the source node in the presence of data routing and detection routing, it will consume part of the energy. We need to calculate the energy consumption in the network and analyze the impact of the feedback mechanism on the network lifetime. The composition of the confirmation packet is the same as the data packet, so the energy consumed is equivalent to the energy consumption of the unit data packet.

The data packet contains the m-bit binary code of k data. The feedback data packet is a binary code after k data XOR. Because the amount of data has fewer coded bits, and the energy consumption of the encoded feedback signal itself is low, the energy consumption of the coded bits of the number of received data can be ignored, which is equivalent to m-bit encoding. Therefore, the composition of the feedback data packet is the same as that of the data packet. The energy consumption per unit data packet is e_p. The feedback data packet contains one piece of data so the energy consumed by the feedback signal is also e_p.

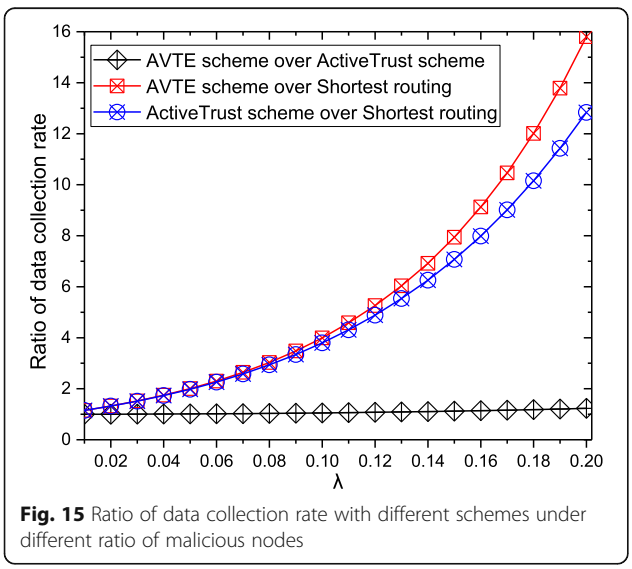

Fig. 15 Ratio of data collection rate with different schemes under different ratio of malicious nodes

We analyze whether the remaining energy in the network can establish detection routes after the data packets and confirmation packets consume some energy. The energy consumption is related to the number of packets carried by the node. Consider the network radius is R, the transmission radius of the node is r, the event occurrence rate is v, and the distance from the node to the sink is l [36]. According to the Ref. [36], we can get the number of data packets loaded by the node is

$$d_l = \left((z+1) + \left(\frac{z(z+1)r}{2l}\right)\right)v \tag{17}$$

z is an integer and satisfies $l + zr < R$.

Eq. (17) shows that the energy consumption depends on the amount of data and the lifetime of the network depends on the node with the highest energy consumption. We consider that the maximum data load of the node is d_{max} and the energy consumption is $d_{max}e_u$. The node with the data loads less than d_{max} has residual energy. The remaining energy can be used to send feedback packets to the source node and construct detection routes. For a node with a distance l to the sink, the remaining energy is $(d_{max} - d_l)e_u$. If the distance of the active probe route is measured by the hops and after sending a confirmation packet, the available hops of the active probe route are as follows.

Theorem 8: If the distance from the node to the sink is l, the maximum number of detection hops that the remaining energy can reach is

$$\chi = \frac{(d_{max} - d_l - 1)(1 + k_2)}{1 + k_2/k_1} \tag{18}$$

Where k_1 is the ratio of the length of the data packet to the detection packet, and k_2 is the ratio of the body length of the data packet to the packet header of the detection packet.

Proof: According to Eq. (17), for the node with distance l from the sink, the data load is $d_l = ((z+1) + (z(z+1)r/2l))v$. Therefore, the node closest to the sink has the largest data load $d_{max} = ((z+1) + (z(z+1)r/2l_{min}))v$. e_p represents the energy consumption for sending and receiving a unit data packet. Each node sends a confirmation packet to the source node, so the energy consumption of the node for the feedback signal is also e_p. The remaining energy of the node is $(d_{max} - d_l)e_p - e_p = (d_{max} - d_l - 1)e_p$.

Considering that the energy consumed by sending and receiving one bit data is e_u, $e_p = xe_u$, $x = x_1 + x_2$, where x

is the unit packet length, x_1 is the packet header length, and x_2 is the packet body length. The available remaining energy is $(d_{max} - d_l - 1)xe_u = (d_{max} - d_l - 1)e_u(x_1 + x_2)$. The energy consumption of sending and receiving a detection packet is $e_q = ye_u = e_u(y_1 + y_2)$, where y is the packet length of the detection packet, y_1 is the packet header length, and y_2 is the packet body length. Considering $x_1 = y_1$, $x_2 = k_1y_2$, $x_2 = k_2y_1$, the hop counts of the active detection routes that can be achieved by the remaining energy of the node are

$$\chi = \frac{(d_{max} - d_l - 1)e_u(x_1 + x_2)}{e_u(y_1 + y_2)} \tag{19}$$

$$\Rightarrow \chi = \frac{(d_{max} - d_l - 1)(1 + k_2)}{1 + k_2/k_1}.$$

Assuming $R = 500$, $r = 50$, $v = 0.8$, $l_{min} = 10$, Fig. 16 shows the maximum detection hops that can be provided by the remaining energy of nodes at different distances to the sink. When k_1 and k_2 are different, the number of detection hops can reach hundreds. The closer the node is to sink, the greater the data load of the node, the less energy left, and the fewer hops available for detecting the path. The amount of data of nodes far away from the sink is small. Thus, there is more residual energy, and the number of hops available for the detection path is up to 500. In addition to the energy consumed by the node sending and receiving packets and confirming packets, the node has enough energy to build the detection route for monitoring. Figure 17 shows that the number of hops of the detection route decreases when the network radius and the transmission radius are increased, but it can still reach 200 hops.

Consider $R = 500$, $v = 0.8$, $l_{min} = 50$, $k_1 = 5$, $k_2 = 5$, Fig. 18 shows the number of hops of the detection path at different distances from sink under different transmission radius. It can be seen that the number of hops of the detection path gradually increases. The initial growth is relatively large and detection hops grow slowly near the sink. The maximum data load is smaller with the larger transmission radius. The detection hops afforded by remaining energy of nodes decreases, but the maximum hops remain above 200. The maximum detection hops can be up to 500, which shows that the remaining energy in the network is sufficient to support the establishment of the detection path. The ActiveTrust mechanism verifies that the lifetime of the network is the same as other solutions without any security strategy. Compared with the ActiveTrust scheme, our scheme can establish up to hundreds of hops for the active detection path. Therefore, we can conclude that the network lifetime of our scheme is the same as that of the ActiveTrust scheme and the scheme without any security strategy.

Conclusion

In this paper, an Active and Verifiable Trust Evaluation (AVTE) approach is proposed to identify the credibility of IoT devices, so to ensure reliable data collection for Edge Computing with low cost. The main innovation of the AVTE approach is the use of an encoding-based feedback mechanism for data routing. This method can more accurately and directly obtain the trust status of more nodes, so that a richer trust relationship can be obtained more quickly, making evolution and reasoning of trust is more credible and richer. Our theoretical analysis proves its effectiveness. The conclusion of this article is: (1) Compared with the traditional passive trust mechanism, the AVTE approach is an active trust mechanism and it has a great advantage in obtaining trust, which can improve the data collection rate very well.

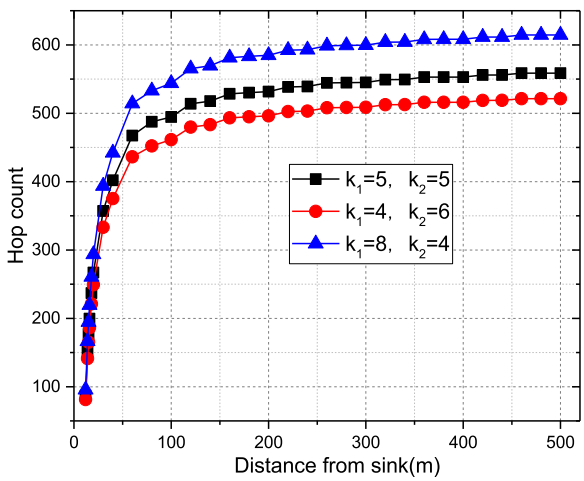

Fig. 16 Number of detection hops created by the remaining energy (different k_1, k_2)

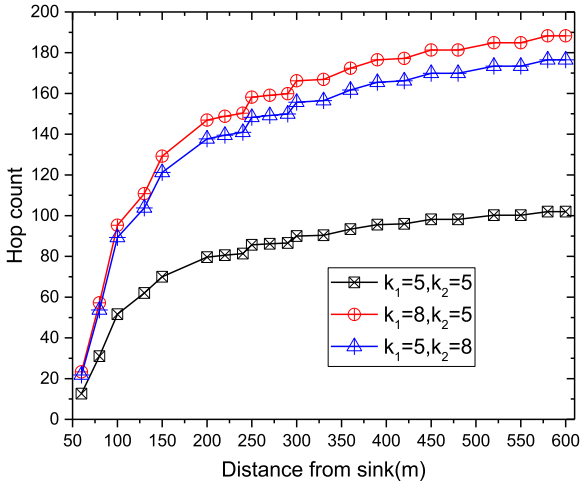

Fig. 17 Number of detection hops that created by the remaining energy($r = 50$)

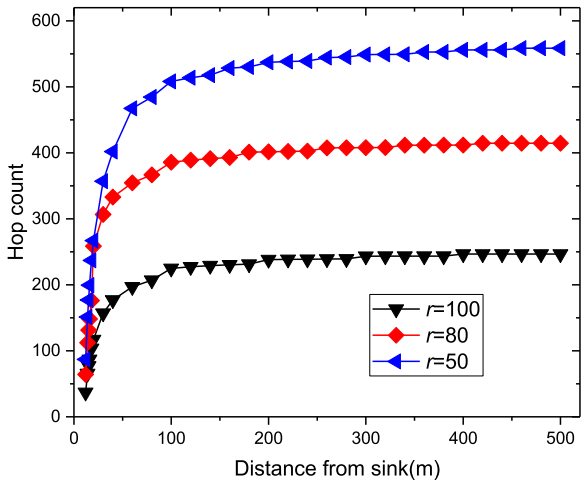

Fig. 18 Number of detection hops created by remaining energy (different *r*)

And it is effective for all kinds of attacks known and unknown in the network. (2) It has a feedback mechanism for data routing, which can greatly increase the available trust relationships and make trust evaluation more rapid and accurate. The encoding-based feedback mechanism makes the cost of feedback information is less, which helps to improve network performance. There are some problems need our further research. First, this paper studies a network scenario of a self-organizing network. The next step is to expand it to a more complex and heterogeneous edge network. Second, we accurately combine the current emerging data acquisition and perception technologies and tools to further proactively and quickly obtain the trust relationship of IoT devices to promote the further development of IoT.

Abbreviations

IoT: Internet of Things; WSNs: Wireless sensor networks; EC: Edge Computing; SFA: Selective forwarding attack; SEDR: Security and Energy-efficient Disjoint Route; AVTE: Active and Verifiable Trust Evaluation; QoE: Quality of Experiment; CHEMAS: Check-point-based multi-hop acknowledgement scheme; TTL: Time to live

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (61772554, 61902432).

Authors' contributions

Wen Mo is the main author of the current paper. Tian Wang, Shaobo Zhang commented the work. Jinhuan Zhang contributed to the conception. All authors read and approved the final manuscript.

Funding

This work was supported in part by the National Natural Science Foundation of China (61772554, 61902432).

Availability of data and materials

Not applicable.

Competing interests

The authors declare that there are no competing interests regarding the publication of this paper.

Author details

[1]School of Computer Science and Engineering, Central South University, Changsha 410083, China. [2]College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China. [3]School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China.

Received: 18 May 2020 Accepted: 10 September 2020
Published online: 21 September 2020

References

1. Sarkar S, Chatterjee S, Misra S (2015) Assessment of the suitability of fog computing in the context of internet of things. IEEE Trans Cloud Comput 6(1):46–59
2. Huang M, Liu A, Xiong NN, Wang T, Vasilakos AV (2020) An effective service-oriented networking management architecture for 5G-enabled internet of things. Comput Netw 173:107208. https://doi.org/10.1016/j.comnet.2020.107208
3. Qi L, Yu J, Zhou Z (2017) An invocation cost optimization method for web services in cloud environment. Sci Program https://doi.org/10.1155/2017/4358536
4. Luo Y, Zhu X, Long J (2019) Data collection through Mobile vehicles in edge network of Smart City. IEEE Access 7:168467–168483
5. Jiang B, Huang G, Wang T, Gui J, Zhu X (2020) Trust based energy efficient data collection with unmanned aerial vehicle in edge network. Trans Emerg Telecommunications Technol https://doi.org/10.1002/ett.3942
6. Wang T, Wang P, Cai S, Ma Y, Liu A, Xie M (2020) A unified trustworthy environment based on edge computing in industrial IoT. IEEE Trans Ind Inform 16(9):6083–6091
7. Wang T, Qiu L, Sangaiah AK, Liu A, Md B, Ma Y (2020) Edge computing based trustworthy data collection model in the internet of things. IEEE Internet Things J 7(5):4218–4227
8. Qi L, Zhang X, Dou W, Hu C, Yang C, Chen J (2018) A two-stage locality-sensitive hashing based approach for privacy-preserving mobile service recommendation in cross-platform edge environment. Futur Gener Comput Syst 88:636–643
9. Bonola M, Bracciale L, Loreti P, Amici R, Rabuffi A, Bianchi G (2016) Opportunistic communication in smart city: experimental insight with small-scale taxi fleets as data carriers. J Ad Hoc Netw 43:43–55
10. Tan J, Liu W, Wang T, Zhao M, Liu A, Zhang S (2020) A high-accurate content popularity prediction computational Modelling for Mobile edge computing by using matrix completion technology. Trans Emerg Telecommun Technol. https://doi.org/10.1002/ett.3871
11. Qi L, Zhang X, Dou W, Ni Q (2017) A distributed locality-sensitive hashing-based approach for cloud service recommendation from multi-source data. IEEE J Selected Areas Commun 35(11):2616–2624
12. Zeng D, Gu L, Pan S, Cai J, Guo S (2019) Resource Management at the Network Edge: A deep reinforcement learning approach. IEEE Netw 33(3):26–33
13. Sánchez-Casado L, Maciá-Fernández G, García-Teodoro P, Magán-Carrión R (2015) A model of data forwarding in MANETs for lightweight detection of malicious packet dropping. Comput Netw 87:44–58
14. Ren Y, Zeng Z, Wang T, Zhang S, Zhi G (2020) A trust-based minimum cost and quality aware data collection scheme in P2P network. Peer-to-Peer Netw Appl https://doi.org/10.1007/s12083-020-00898-2
15. Qi L, Chen Y, Yuan Y, Fu S, Zhang X, Xu X (2020) A QoS-aware virtual machine scheduling method for energy conservation in cloud-based cyber-physical systems. World Wide Web 23:1275–1297
16. Xie K, Li X, Wang X, Cao J, Xie G, Wen J, Qin Z (2018) On-line anomaly detection with high accuracy. IEEE/ACM Trans Networking 26(3):1222–1235
17. Huang M, Zhang K, Zeng Z, Wang T, Liu Y (2020) An AUV-assisted data gathering scheme based on clustering and matrix completion for Smart Ocean. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2020.2988035
18. Liu X, Qiu T, Dai B, Yang L, Liu A, Wang J (2020) Swarm intelligence-based rendezvous selection via edge computing for mobile sensor networks. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2020.2966870

19. Chen Y, Liu W, Wang T, Deng Q, Liu A, Song H (2019) An adaptive retransmit mechanism for delay differentiated services in industrial WSNs. EURASIP J Wireless Communications and Networking https://doi.org/10.1186/s13638-019-1566-2
20. Deng X, Jiang Y, Yang LT, Lin M, Yi L, Wang M (2019) Data fusion based coverage optimization in heterogeneous sensor networks: A survey. Inform Fusion 52:90–105
21. Teng H, Ota K, Liu A, Wang T, Zhang S (2020) Vehicles joint UAVs to acquire and analyze data for topology discovery in large-scale IoT systems. J Peer-to-Peer Netw Appl https://doi.org/10.1007/s12083-020-00879-5
22. Liu Q, Hou P, Wang G, Peng T, Zhang S (2019) IntelligentRoute planning on large road networks with efficiency and privacy. J Parallel Distributed Comput 133:93–106
23. Hafeez KA, Zhao L, Ma B, Mark J (2013) Performance analysis and enhancement of the DSRC for VANET's safety application. IEEE Tran Vehicular Technol 62(7):3069–3083
24. Zhao Y, Wang T, Zhang S, Wang Y (2020) Towards mini-mum code dissemination delay through UAV joint vehicles for smart city. IET Commun. https://doi.org/10.1049/iet-com.2019.1205
25. Abdelhamid S, Hassanein HS, Takahara G (2017) Reputation-aware, trajectory-based recruitment of smart vehicles for public sensing. IEEE Trans Intell Transportation Syst 19(5):1387–1400
26. Jiang J, Han G, Wang F, Shu L, Guizani M (2014) An efficient distributed trust model for wireless sensor networks. IEEE Trans Parallel Distributed Syst 26(5):1228–1237
27. Han G, Shen W, Duong TQ, Guizani M, Hara T (2014) A proposed security scheme against denial of service attacks in cluster-based wireless sensor networks. J Secur Commun Netw 7(12):2542–2554
28. Li T, Zhao M, Won K (2020) Machine learning based code dissemination by selection of reliability Mobile vehicles in 5G networks. Comput Commun 152:109–118
29. Chen M, Wang T, Ota K, Dong M, Zhao M, Liu A (2020) Intelligent resource allocation Management for Vehicles Network: an A3C learning approach. Comput Commun 151:485–494
30. Morra L, Lamberti F, Pratticó FG, La Rosa S, Montuschi P (2019) Building Trust in Autonomous Vehicles: role of virtual reality driving simulators in HMI design. IEEE Trans Vehicular Technol 68(10):9438–9450
31. Wang T, Liang Y, Yang Y, Xu G, Peng H, Liu A, Jia W (2020) An intelligent edge-computing-based method to counter coupling problems in cyber-physical systems. IEEE Netw 34(3):16–22
32. Zhang N, Yang P, Ren J, Chen D, Li Y, Shen X (2018) Synergy of big data and 5G wireless networks: opportunities, approaches, and challenges. IEEE Wirel Commun 25(1):12–18
33. Deng X, Yang LT, Yi L, Wang M, Zhu Z (2018) Detecting confident information coverage hole in industrial internet of things: an energy-efficient perspective. IEEE Commun Mag 56(9):68–73
34. Liu Y, Liu X, Liu A, Xiong N, Liu F (2019) A trust computing based security routing scheme for cyber physical systems. ACM Trans Intell Syst Technol 10(6):1–27
35. Peng M, Liu W, Wang T, Zeng Z (2020) Relay selection joint consecutive packet routing scheme to improve performance for wake-up radio-enabled WSNs. Wirel Commun Mob Comput 2020:7230565
36. Liu Y, Dong M, Ota K, Liu A (2016) ActiveTrust: secure and trustable routing in wireless sensor networks. IEEE Trans Inform Forensics Secur 11(9):2013–2027
37. Wang T, Luo H, Zeng X, Yu Z, Liu A. Sangaiah A. (2020) Mobility based trust evaluation for heterogeneous electric vehicles network in smart cities, IEEE Trans Intell Transp Syst, DoI: https://doi.org/10.1109/TITS.2020.2997377
38. Wang T, Cao Z, Wang S, Wang J, Qi I, Liu A, Xie M, Li X. (2020) Privacy-enhanced data collection based on deep learning for internet of vehicles. IEEE Trans Ind Inform 16(10):6663–6672
39. Wang H, Ma S, Dai H. N, Imran M, Wang T. (2020). Blockchain-based data privacy management with nudge theory in open banking. Futur Gener Comput Syst, 110, 812–823
40. Li T, Liu W, Wang T, Zhao M, Li X, Ma M (2020) Trust data collections via vehicles joint with unmanned aerial vehicles in the smart internet of things. Trans Emerg Telecommun Technol. https://doi.org/10.1002/ett.3956
41. Zhuo C, Luo S, Gan H, Hu J, Shi Z (2019) Noise-Aware DVFS for Efficient Transitions on Battery-Powered IoT Devices. IEEE Trans Comput Aided Des Integr Circuits Syst 2019. https://doi.org/10.1109/TCAD.2019.2917844
42. Liu A, Zheng Z, Zhang C, Chen Z, Shen X (2012) Secure and energy-efficient disjoint multi-path routing for WSNs. IEEE Trans Veh Technol 61(7):3255–3265
43. Xiao B, Yu B, Gao C (2007) CHEMAS: identify suspect nodes in selective forwarding attacks. J Parallel Distributed Comput 67(11):1218–1230
44. Wang J, Wang F, Wang Y, Wang L, Qiu Z, Zhang D et al (2019) HyTasker: hybrid task allocation in Mobile crowd sensing. IEEE Trans Mob Comput. https://doi.org/10.1109/TMC.2019.2898950
45. Xie K, Ning X, Wang X, Xie D, Cao J, Xie G, Wen J (2017) Recover corrupted data in sensor networks: A matrix completion solution. IEEE Trans Mob Comput 16(5):1434–1448
46. Liu Q, Tian Y, Wu J, Peng T, Wang G (2019) Enabling verifiable and dynamic ranked search over outsourced data. Trans Serv Comput. https://doi.org/10.1109/TSC.2019.2922177
47. Zhao W (2016) Performance optimization for state machine replication based on application semantics: a review. J Syst Softw 112:96–109
48. Zheng Z, Liu A, Cai L. X, Chen Z, Shen X. (2016) Energy and memory efficient clone detection in wireless sensor networks. IEEE Trans Mob Comput 15(5):1130–1143
49. Kuang Z, Li G, Zhang L, Zhou H, Li C, Liu A (2020) Energy efficient mode selection, Base Station selection and resource allocation algorithm in D2D heterogeneous networks. Peer-to-Peer Netw Appl. https://doi.org/10.1007/s12083-020-00915-4
50. Lu R, Lin X, Zhu H, Liang X, Shen X (2012) BECAN: A Bandwidth-Efficient Cooperative Authentication Scheme for Filtering Injected False Data in Wireless Sensor Networks. IEEE Trans Parall Distr 23(1):32–43
51. Liu X, Lin P, Liu T, Wang T, Liu A, Xu W (2020) Objective-variable tour planning for Mobile data collection in partitioned sensor networks. IEEE Trans Mob Comput. https://doi.org/10.1109/TMC.2020.3003004
52. Wang T, Zhao D, Cai S, Jia W, Liu A (2020) Bidirectional prediction based underwater data collection protocol for end-edge-cloud orchestrated system. IEEE Trans Ind Inform 16(7):4791–4799
53. Huang M, Liu W, Wang T, Liu A, Zhang S (2020) A cloud-MEC collaborative task offloading scheme with service orchestration. IEEE Internet Things J 7(7):5792–5805
54. Aazam M, Harras KA, Zeadally S (2019) Fog computing for 5G tactile industrial internet of things: QoE-aware resource allocation model. IEEE Trans Ind Inform 15(5):3085–3092
55. Dong M, Liu X, Qian Z, Liu A, Wang T (2015) QoE ensured Price competition model for emerging Mobile networks. IEEE Wirel Commun 22(4):50–57

Publisher's Note