Postprint: Research on Assessing Sensor Data Credibility Based on BP Networks

Authors: Liu Xiaojiu, Yuan Ding, Liang Aiyun, Yan Qing

Date: 2018-05-24T00:00:00+00:00

Abstract

Traditional methods that utilize symmetric and asymmetric encryption for security assurance in sensor network systems require substantial encryption and decryption computations, and cannot accurately assess data trustworthiness after key compromise, thereby failing to effectively guarantee wireless sensor network system security. To secure wireless sensor network systems and address the issue of node information trustworthiness, this paper proposes a BP network-based approach for evaluating node information trustworthiness. The method employs a BP neural network on the border router to train on collected multi-feature data, subsequently using the training results to determine node trustworthiness and filter data accordingly. This approach exhibits low system overhead while providing high security assurance, enabling the identification of problematic nodes and ensuring secure operation of sensor networks. Experimental results demonstrate that the method achieves short authentication times and meets expected performance targets.

Full Text

Preamble

Research on Data Credibility of Sensors Based on BP Neural Networks

Liu Xiaojiu, Yuan Ding†, Liang Aiyun, Yan Qing
(School of Computer Science, Sichuan Normal University, Chengdu 610101, China)

Abstract: Traditional methods secure sensor network systems through symmetric and asymmetric encryption, which require extensive computational resources for encryption and decryption. Moreover, once keys are compromised, these methods cannot accurately assess data credibility, failing to effectively

guarantee wireless sensor network (WSN) security. To address this limitation and ensure WSN system security, this paper proposes a novel approach for evaluating node information credibility based on BP neural networks. The method deploys a BP neural network on border routers to train multi-feature data collected from sensor nodes, subsequently using the trained model to assess node credibility and filter data accordingly. This approach achieves low system overhead while providing high security assurance, enabling the identification of compromised nodes and ensuring safe WSN operation. Experimental results demonstrate that the method offers short authentication times and meets expected performance targets.

Keywords: BP network; wireless sensor networks; sensor nodes; credibility; security

0 Introduction

Wireless sensor networks (WSNs) represent a focal point in IoT research. These networks require no infrastructure support and enable rapid deployment of sensor nodes in designated areas for environmental monitoring, with data reported to base stations via wireless links. This flexible architecture suits hostile environments and large-scale geographic regions, finding widespread application in military and civilian domains such as target tracking, area reconnaissance, and field environmental monitoring. Specific scenarios include: (a) military applications, where sensor nodes monitor enemy vehicle traffic or track troop movements; and (b) environmental monitoring, such as forest fire surveillance or pollution detection in contaminated areas.

WSN nodes can be deployed through aerial dispersal or random scattering, typically independent of human intervention. However, in natural environments, magnetic interference, electric field interference, and malicious attacks can cause nodes to transmit fraudulent data. To ensure secure WSN operation, node information security emerges as a critical concern. Traditional security schemes employ symmetric encryption for information transmission and asymmetric encryption for node authentication. These approaches demand substantial computational resources, making them impractical for sensor nodes with limited processing power and energy. Furthermore, security cannot be adequately maintained once cryptographic keys are compromised.

To reduce node computational load, extend WSN lifecycle, and enhance system security performance, this paper proposes a method for assessing node data credibility using BP neural networks. This approach filters interference data while improving system operability by evaluating the trustworthiness of transmitted information.

1 Related Work

With the large-scale deployment of sensor network technology, node security issues have gained prominence. Based on the characteristics of WSN nodes and the BP neural network learning algorithm, this study employs BP neural networks to evaluate sensor node credibility.

---

2 Methodology

2.1 BP Neural Network Structure

The BP neural network consists of an input layer, hidden layer, and output layer. The input layer feeds training samples into the network, while the hidden layer trains the sample data. Theoretical proofs demonstrate that a three-layer BP neural network can approximate any complex nonlinear mapping provided the hidden layer contains a sufficient number of nodes [22]. This work adopts a three-layer BP neural network architecture, as illustrated in Figure 1 [Figure 1: see original paper].

The number of nodes in the input and output layers is determined by the problem domain, while the hidden layer node count is typically determined through trial-and-error methods. The empirical formula for determining the optimal number of hidden neurons is:

$$h = \sqrt{m+n} + a \quad (1)$$

where h represents the number of hidden layer nodes, m denotes input layer nodes, n indicates output layer nodes, and a is a tuning constant between [1,10].

2.1.1 Forward Propagation Process Let W_{ij} represent the weight between nodes i and j, b_j denote node j' s threshold, and x_j indicate the node' s output value. The output value is obtained through transformations of all upper-layer node outputs, weights, thresholds, and activation functions. The specific calculation method is as follows:

$$S_j = \sum w_{ij} x_i + b_j \quad (2)$$

$$x_j = f(S_j) \quad (3)$$

where f is the activation function, typically a Sigmoid function as shown in equation (4):

$$f(\theta) = 1 + e^{-}$$

where A and B are constants. This completes the BP network' s approximate mapping from n-dimensional to m-dimensional space vectors.

2.1.2 Backward Propagation Process The primary objective of BP neural networks is to iteratively adjust weights and thresholds to minimize error values. In the BP neural network, assuming the actual output of the j-th node in the output layer is d_j and the expected output is y_j, the error E is calculated using equation (5):

$$E(\omega, b) = \sum (d_j - y_j)$$

Weight and threshold adjustments between the hidden layer (input layer) and output layer (hidden layer) follow these update rules:

$$\omega_{ij} = \omega_{ij} - \eta_1 \times \frac{\partial \Lambda(\omega, b)}{\partial \omega_{ij}} = \omega_{ij} - \eta_1 \delta_{ij} x_i \quad (6)$$

$$b_j = b_j - \eta_2 \times \frac{\partial \Lambda(\omega, b)}{\partial b_j} = b_j - \eta_2 \delta_{ij} \quad (7)$$

where ω_{ij} represents the weight to be adjusted, b_j denotes the threshold to be adjusted, η_1 and η_2 are predefined learning rates, and δ_{ij} is the momentum factor from hidden layer (input layer) node i to output layer (hidden layer) node j.

2.1.3 Application Principle [Content minimal; section serves as transition to implementation]

2.2 Security Scheme Design

Wireless sensor nodes are typically deployed in public areas to collect environmental information. During system operation, malicious attackers can intercept WSN communication data through eavesdropping. After decryption, attackers may impersonate legitimate nodes to transmit fraudulent information, causing false alarms or misreports that disrupt system functionality. Traditional security methods cannot effectively defend against such attacks. Additionally, environmental factors such as magnetic or electric field interference can cause sensor nodes to transmit unstable data, also resulting in false alarms.

To prevent attackers from forging information and mitigate errors from environmental influences, this scheme proposes using BP neural networks to evaluate collected node information and filter out fraudulent and interference data.

2.2.1 Data Flow and System Operation Process Sensor node data reaches border routers after intra-region routing. The border router implements a BP neural network assessment method to evaluate node data credibility. Trusted node information is forwarded to the server, while untrusted information is discarded. In WSNs, transmitted information includes normal data, error data sent under environmental influence, and forged data from attackers. The data flow is illustrated in Figure 2 [Figure 2: see original paper].

System operation comprises three phases: data collection, data training, and normal operation.

Phase 1: Data Collection. The border router collects data transmitted by nodes for two rounds, with each round comprising data from N nodes. During this phase, data is only collected and forwarded without assessment, and initial trust values are assigned to nodes. The border router completely trusts data collected in the first two rounds.

Phase 2: Data Training. Training initiates at the beginning of the third data collection round. When training begins, the border router continues collecting and forwarding data while presetting initial trust values. No data assessment occurs during this phase. Training must start during the third collection round because time frequency serves as a feature of the training dataset and requires corresponding temporal values. After three rounds of data collection, the training process completes and the resulting model becomes available.

Phase 3: Normal Operation. After completing the third data collection round, the system enters normal operation. During this phase, when the border router receives data, it uses the trained model to assess incoming data and takes appropriate actions based on assessment results. While assessing current data, the system continues training on data from the previous round. The system can assess received data because training completes after the third collection round, yielding a trained model. The overall system flow is shown in Figure 3 [Figure 3: see original paper].

2.2.2 Judgment Process Sensor nodes transmit simple information such as temperature, illumination, humidity, ground vibration frequency, infrared intensity, environmental noise, and battery discharge curves. This work extracts meaningful feature information from node transmissions and trains a BP neural network to obtain a trained model. When nodes subsequently transmit information, the trained parameters are used to assess the credibility of sent data. If assessment results fall within the absolutely trusted range, data is forwarded and added to the training set. If results fall within the relatively trusted range, the information participates in subsequent training but is not forwarded. If results are absolutely untrusted, data is discarded directly. This dynamic authentication of node-transmitted information eliminates the need for encrypting node identity information, bypassing concerns about illegal node forgery while saving node computational time and extending node lifespan. Judgment results and

corresponding actions are shown in Table 1 , with the judgment flow illustrated in Figure 4 [Figure 4: see original paper].

The calculation formulas used during judgment are equations (2) through (4). The result set is derived from data trained in the previous iteration.

2.2.3 Learning Process Since node data requires no deletion or modification operations, an array structure can be used on the border router to store node data:

```
float M[10000][9];
```

This two-dimensional array stores node information, with the first dimension representing node IDs and the second dimension storing transmission frequency, temperature, illumination, humidity, ground vibration frequency, infrared intensity, environmental noise, battery discharge curve, and timestamp. The timestamp records when node data is received, while frequency data is calculated as the difference between current and previous timestamps. During the first data collection round, time frequency equals the timestamp.

2.2.4 Training Method Transmission frequency, temperature, illumination, humidity, ground vibration frequency, infrared intensity, environmental noise, and battery discharge curve serve as input layer data for the BP neural network. The BP neural network training process begins when the third round of information collection starts. After each subsequent collection round, the BP network initiates training.

---

3 Experimental Design and Results Analysis

The proposed method is a BP neural network-based security authentication approach that requires no asymmetric key encryption for node information and can assess information trustworthiness without extensive attention to node-specific conditions.

3.1 Experimental Environment

The experiment uses a computer with 2 GB memory and a Core single-core processor running Ubuntu 14.04 as the border router. The border router is configured with a runtime environment and the Workerman framework. A desktop computer serves as the border router to implement the BP neural network for input data credibility verification.

3.2 Experimental Method

Assume a WSN region contains N nodes, each transmitting collected environmental information at M time intervals. Nodes continuously gather environ-

mental information and send it to the border router, which records reception timestamps. Time frequency serves as a valid input feature for the BP neural network, alongside temperature, illumination, humidity, ground vibration frequency, infrared intensity, environmental noise, and battery discharge curve.

The system operates through three specific phases:

a) Data Collection Phase. The border router collects two rounds of node-transmitted data. During the first round, all node data is stored in an array. In the second round, node timestamps are subtracted from corresponding first-round timestamps to obtain transmission intervals. The border router only forwards data without assessment, assigning initial trust values randomly between 9.859 and 9.889. This range is set according to experimental requirements and can be arbitrarily defined. After two rounds, the data collection phase concludes.

b) Data Training Phase. Training initiates during the third data collection round. Starting training at this point is necessary because time frequency serves as a training dataset feature requiring corresponding values. After two collection rounds, time frequency values become available. When training begins, the border router continues collecting and forwarding data while presetting initial trust values, but performs no data assessment.

c) Normal Operation Phase. After completing the third data collection round, the system enters normal operation. Upon receiving data, the border router uses the trained model to assess incoming data and takes appropriate actions while simultaneously training on data from the previous round. The system can assess received data because training completes after the third collection round, yielding a trained model.

Forwarding rules after assessment are as follows: If assessed values fall within the initially defined range of 9.859–9.889, data is considered absolutely trusted, added to the training pool, and forwarded. If values fall between 9.0–9.859 or 9.889–10.4, data is relatively trusted, added to the training pool, but not forwarded. Values outside these ranges are considered absolutely untrusted—data is neither forwarded nor added to the training pool. Trained data is replaced with arbitrary values between 9.859 and 9.889. Intervals and assessment results are shown in Table 3 .

3.3 Experimental Results Analysis

Under the described experimental design and environment, the BP neural network-based WSN node data credibility assessment was validated. Data from N nodes underwent 1,000 test iterations. This number was selected because 1,000 tests adequately represent normal WSN operation, with additional tests providing negligible improvement. During the first three training rounds, all node-transmitted data was assumed legitimate. Thereafter, $N/30$ data items per round were introduced as interference. Experimental conditions and results are shown in Table 3. With node counts of 1,000, 2,000, 3,000, 5,000, and

10,333, operations were performed after 100 training iterations, with results presented in Table 4 .

Accuracy is calculated as:

$$\text{Accuracy} = \frac{\text{Average Accepted Data}}{\text{Transmitted Data} - \text{Interference Data}} \times 100\% \quad (8)$$

The proposed method demonstrates significantly improved accuracy compared to the D-S evidence theory method in reference [21]. Analysis reveals that reference [21] employs fixed correlation parameters for node transmission data, which cannot represent all environmental conditions given the random nature of environmental changes.

As node count increases and interference data grows, accuracy variation follows the pattern shown in Figure 5 [Figure 5: see original paper]. The BP neural network achieves credibility assessment accuracy fluctuating around 93%, consistently exceeding 90%.

4 Conclusion

The proposed method accurately identifies illegal data interference and eliminates fraudulent data. Compared with the D-S evidence theory method in reference [21], accuracy improves substantially. Several aspects warrant future enhancement: (a) improving data collection methods—currently training occurs after each data transmission round, but could be modified to train after specific time intervals for broader applicability; (b) improving training methods—this experiment uses fixed training iterations and parameters, but adaptive parameters could be implemented; (c) improving data storage methods—this experiment uses arrays for training data, but file-based storage could reduce memory overhead; and (d) integrating with traditional security methods to further enhance system security.

References

[1] Yuan Jianjun. An enhanced two-factor user authentication in wireless sensor networks [J]. Telecommunication Systems, 2014, 55 (1): 105-113.

[2] He Daojing, Chen Chun, Chan S, et al. Analysis and improvement of a secure and efficient handover authentication for wireless networks [J]. IEEE Communications Letters, 2012, 16 (8): 1270-1273.

[3] Wang Ding, He Debiao, Wang Ping, et al. Anonymous two-factor authentication in distributed systems: certain goals are beyond attainment [J]. IEEE Trans on Dependable & Secure Computing, 2015, 12 (4): 428-442.

[4] Wang Chenyu, Xu Guoai, Sun Jing. An enhanced three-factor user authentication scheme using elliptic curve cryptosystem for wireless sensor networks [J]. Sensors, 2017, 17 (12): 2946-2965.

[5] Li Xiong, Niu Jianwei, Kumari S, et al. A three-factor anonymous authentication scheme for wireless sensor networks in Internet of things environments [J]. Journal of Network & Computer Applications, 2018, 103 (2): 194-204.

[6] Wang Ding, Wang Ping. Two birds with one stone: two-factor authentication with security beyond conventional bound [J]. IEEE Trans on Dependable & Secure Computing, 2016, PP (99): 1.

[7] Chen Lei, Wei Fushan, Ma Chuangui. A provably-secure two-factor user authentication key exchange scheme for wireless sensor networks [J]. Application Research of Computers, 2016, 33 (5): 1514-1521.

[8] Fan Hong. Applications of digital signature in network communication security [J]. Journal of Graduate School of Chinese Academy of Social Sciences, 2001, 18 (2): 101-104.

[9] Sun Yahui, Li Feng. A mutual authentication mechanism based on the path [J]. Electronic Science and Technology, 2017, 30 (2): 177-179.

[10] Chen Jianjun. Study on trust model for wireless sensor networks based on trust chain [D]. Chengdu: Chengdu University of Information Technology, 2015.

[11] Cao Zheng. The research of nodes authentication protocol for wireless sensor networks [D]. Chengdu: Southwest Jiaotong University, 2015.

[12] Guo Ping, Fu Desheng, Cheng Yaping, et al. Design and proof of bilateral authentication protocol for wireless sensor network [J]. Computer Science, 2015, 42 (2): 100-103.

[13] Sun Erkun. The study on the security routing in the wireless sensor network [D]. Xi' an: Xidian University, 2013.

[14] Han Xiaona. Research on secure clustering protocol based on trust evaluation for wireless sensor networks [D]. Beijing: Beihang University, 2015.

[15] Yang Min. Research on cluster-based secure protocol for wireless sensor network [D]. Guilin: Guilin University Of Electronic Technology, 2014.

[16] Huang Bin, Liu Guangzhong, Xu Ming. Security authentication protocol for nodes in wireless sensor networks based on clusters [J]. Computer Engineering, 2016, 42 (7): 117-122.

[17] Qiu Gege, Wang Xueming, Zhang Yansheng. Research on WSN identity authentication protocol based on HECC [J]. Netinfo Security, 2015, 26 (12): 54-58.

[18] Huang Ling, Joseph A D, Tygar J D, et al. Adversarial machine learning [C]// Proc of ACM Workshop on Security and Artificial Intelligence. 2011: 43-58.

[19] Huang S, Papernot N, Goodfellow I, et al. Adversarial attacks on neural network policies [C]// Proc of the 5th International Conference on Learning Representations. 2017: 1-7.

[20] Suciu O, Mărginean R, Kaya Y, et al. When does machine learning FAIL? Generalized transferability for evasion and poisoning attacks [J]. arXiv: 1803.06975v1, 2018.

[21] Miao Chenglin. Research on trustworthiness problems in wireless sensor networks [D]. Hefei: University of Science and Technology of China, 2014.

[22] Bartkowiak A. Neural networks and pattern recognition [M]// Academic.

Note: Figure translations are in progress. See original paper for figures.