Research on Tolerant Aggression Algorithms for Data Aggregation in Wireless Sensor Networks

Kai-jie REN*, Jing YU and Sen SUN

School of Computer Science & Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

*Corresponding author

Keywords: Wireless sensor network security, Time series model, Wavelet transform, Data injection attack.

Abstract. In view of the security issues in the data aggregation process of wireless sensor networks, this paper uses an auto regressive moving average model to characterize the spatio-temporal characteristics of the collected data, and combines the wavelet transform to improve the data prediction model to effectively identify malicious injection attacks in the process of data aggregation. And trusted prediction results replace wrong data. In this paper, a simulation platform based on MATLAB is used to evaluate the complexity and effectiveness of the algorithm. The simulation results show that the proposed scheme improves the relevant indicators for detecting malicious attacks in the wireless sensor network and corrects the wrong aggregated data in time.

Introduction

In recent years, wireless sensor networks have been widely used in intelligent transportation, environmental monitoring, security monitoring and other fields. In these scenarios, the user requests the wireless sensor network to detect abnormalities in the monitoring environment in time so that the user can make decisions. However, the security of the acquisition nodes in the wireless sensor network is very fragile, making it possible for an attacker to capture the acquisition nodes easily and control the captured nodes to inject incorrect data into the aggregation nodes, which ultimately defeats the purpose of destroying the authenticity and effectiveness of the aggregation results [1], causing users to make unreasonable judgments and handling.

At present, the research work on the security aggregation of wireless sensor networks in the academic community can be basically divided into two categories. The first category is the use of traditional cryptography methods to ensure the integrity and confidentiality of data in the aggregation process within the wireless sensor network, to a certain extent, can mitigate the impact of malicious data injection attacks on the normal operation of the wireless sensor network [2]. However, such methods cannot detect injection attacks from within the wireless sensor network. The second category is based on intrusion detection method, which detects malicious injection data during the aggregation process and enhances the robustness of the sink node’s malicious and data injection attacks [3]. However, the probability distribution of the data carried by the wireless sensor network and the probability distribution of the state transition determine the performance of such methods. In a complex outdoor environment, due to a variety of uncertainties, the wireless sensor network eventually causes malicious data injection. Attack detection efficiency is low and accuracy is low.

For the insufficiency of the existing work of the second method, this paper adopts a tolerant-passive algorithm based on the wavelet transform and time series model. The algorithm aims at improving the relevant indicators of the malicious attack detection in the wireless sensor network and correcting the wrong aggregated data in time. By using the temporal correlation of aggregated data, it analyzes and models historical aggregated data, makes reasonable predictions for future aggregated data, introduces a threshold judgment mechanism, and sets multi-level thresholds, if the differences between aggregated data and predicted aggregated data continues to be greater...
than the multi-level threshold, the wireless sensor network is judged to be maliciously attacked, and then the predicted aggregated data is used instead of the wrong aggregated data.

ARMA Tolerance Model Based on Wavelet Transform

ARMA Model

The Auto-Regressive and Moving Average Model (ARMA) [4] is a very effective tool for studying time series and can extract useful statistical attributes from time series. The model is an organic combination of an Auto-Regressive Model (AR) and a Moving Average Model (MA), and can simultaneously obtain the sliding average characteristics and auto-regressive characteristics of the studied time series. The AR component estimates the current value from the linear combination of past samples, while the MA component consists of a linear combination of random noises representing the noise in the data. This model is called ARMA (p,q) and is represented by the p and q parameters as:

\[ x(t) = \sum_{i=1}^{p} \varphi_i (k) x(t-k) + \sum_{i=1}^{q} \theta_i \epsilon(t-i) \]  

where \( x_t \) is time series, \( \epsilon_t \) is Gaussian white noise, \( \varphi_i \) and \( \theta_i \) are the parameters of the AR model and the MA model, respectively.

Hybrid Algorithm

Considering that wavelet transform has multi-scale resolution in digital signal processing, this paper builds a model based on wavelet transform. First, the raw data collected by the sensor is decomposed into high frequency component sequences, intermediate frequency component sequences, and low frequency component sequences by a wavelet decomposition algorithm. Then, because different frequency components have different characteristics, the ARMA models of different orders are established using the \((2n, 2n+1)\) modeling approach respectively. The model suitability test algorithm can be used to determine the order of the model. Commonly used inspection criteria including: F-information criterion, AIC information criterion, BIC information criterion, and then use each ARMA model to predict future data of different frequency component sequences. Finally, the future data predicted by each ARMA model is reconstructed through the inverse operation of the wavelet transform to generate the final prediction data. The technical road map is shown in Fig. 1.

**Figure 1. TECHNOLOGY ROAD MAP.**

\[ X_i = \{X_1, X_2, \ldots\} \] are the original time series, which are decomposed with wavelet- transformation and then reconstructed the decomposition signal of each layer.

\[ X = H_1 + H_2 + H_3 + \ldots + H_l + L_l \]  

Where \( H_l \): \{h_{l,1}, h_{l,2}, h_{l,3} \ldots \}, \ldots, \ H_N: \{h_{N,1}, h_{N,2}, h_{N,3} \} \) are the high frequency signal; \( L_N: \{l_{N,1}, l_{N,2} \ldots \} \) are the low frequency signal.

\[ X_i = h_{i,1} + h_{i,2} + h_{i,3} + \ldots + h_{i,N} + l_{i,N} \]  

If we know the \( X_i \) value of \{t_i | i \leq M\} time and use it to predict the value of k step, we need to solve \( X_{M+k} \), namely
\[ XM+k = h1,M+k + h2,M+k + h3,M+k + ... + hN,M+k + lN,M+k \]  \hspace{1cm} (4)

Since \( H_1, H_2, \ldots, H_N \) can be approximated as stationary time series, ARMA \((p, q)\) models with different orders can be established respectively, and \( h_{1,M+k}, h_{2,M+k}, \ldots, h_{1,M+k} \) are predicted.

The steps of the prediction are as follows:

1) Establish ARMA\((p, q)\) for \( H_j \) and estimate the parameters of the model according to the value of \( h_{j,M+k}^* \), where \( 1 \leq j \leq N, 1 \leq M; \)
2) Test the applicability of the model;
3) The best applicable ARMA \((p, q)\) was used to predict the \( h_{j,M+k}^* \).

**Simulation Results**

In this paper, MATLAB software is used as the simulation platform. The scheme proposed in this paper is simulated on this platform. This scheme is evaluated and analyzed based on the three indicators of data prediction accuracy, false positive rate, false negative rate. The simulation scenario in this article is to deploy 10 temperature sensors at random locations in a 20m*30m rectangular area. In this simulation, the raw data is a sample of temperature data collected by the Intel Berkeley Wireless Sensor Network Laboratory.

**Prediction Accuracy**

In order to objectively evaluate the prediction accuracy of various prediction models, Mean Absolute Deviation (MAE) and Mean Relative Deviation (MRE) are used to judge the proposed model and AR model, ARMA model and WT+AR model.

The following three typical time series-based prediction models are used to compare the prediction accuracy with the prediction model proposed in this paper:

Table 1. The Comparison of Prediction Accuracy with Various Models.

| Model          | MAE   | MRE  |
|----------------|-------|------|
| AR Model       | 2.835°C| 11.33%|
| ARMA Model     | 2.203°C| 10.19%|
| WT+AR Model    | 1.702°C| 8.41% |
| WT+ARMA Model  | 0.352°C| 1.45% |

From the analysis of the contents of Table 1, we can see that the time series model based on wavelet transform presented in this paper is superior to the other three prediction models on MAE and MRE indicators. The results show that the time series is decomposed into high frequency sequences and intermediate frequency sequences. The low-frequency sequence is characterized by different order ARMA models, and the prediction results are more accurate. This model is significant. The higher forecast accuracy also lays the foundation for the improvement of false alarm rate and false negative rate. Moreover, it also shows that using the ARMA model to represent the data set used in this paper is more accurate and suitable than the AR model.

**False Positive Rate**

According to studies, the quality of deployment environment communication, namely, packet loss rate \((p)\), detection threshold \((\tau)\), attack probability \((p_i)\), and attack strength \((D)\) have a decisive effect on the final performance of the solution. In this paper, the meaning of false positive rate \((FPR)\) is the probability that the model misjudges the correct data as illegally injected data, as shown in Fig.2, and the effect of detection threshold \( \tau \) on false positive rate is compared.
In Fig. 2, the false alarm rate $\alpha$ rapidly decreases with the increase of the detection threshold $\tau$ at the beginning, and then, as the detection threshold $\tau$ continues to increase, the false alarm rate $\alpha$ slowly decreases, and eventually converges and stabilizes. When the attack strength is the same, the environment communication quality is good, that is, the packet loss rate $p$ is low, and the false alarm rate $\alpha$ is also low, indicating that good communication quality is an important guarantee for the accuracy of state detection and can effectively reduce the false alarm rate $\alpha$.

**False Negative Rate**

In this paper, the meaning of False Negative Rate (FNR) is the probability that the model misjudges the data that was originally injected illegally as the correct data. Observe the effect of the detection threshold $\tau$ on the false negative rate. The relevant numerical results are shown in Fig. 3.

In Fig. 3, the missing rate $\beta$ at the beginning phase increases rapidly with the increase of the threshold $\tau$. Subsequently, as the detection threshold $\tau$ continues to increase, the false negative rate $\beta$ slows down and eventually converges to the attack probability, $p_F$.

**Summary**

Because the original data has multi-scale characteristics, this paper uses wavelet decomposition algorithm to decompose the original data into low frequency component sequences, intermediate frequency component sequences, high frequency component sequences, and use the ARMA model to model time series components of different frequencies, and then use wavelet reconstruct the final prediction result. The WT+ARMA has a better prediction effect on unstable raw data, which also lays a foundation for its secure aggregation in the wireless sensor network, resulting in better performance of the main indicators. This solution can effectively resist the malicious data injection attack in the wireless sensor network and achieve a reliable, stable, and efficient algorithm for intrusion-tolerant data aggregation.

**References**

[1] D. Estrin. Wireless sensor network tutorial part IV: Sensor network protocols. In proceedings of the ACM Mobile Computing and Networking (MobimCom), 2002, 23-28.
[2] Przydatek B, Song D, Perrig A. Sia: secure information aggregation in sensor networks [C]// Proceedings of the 1st ACM Conference on Embedded Networked Sensor Systems. 2013:255-265.

[3] Sun B, Shan X, Wu K, et al. Anomaly detection based secure in-network aggregation for wireless sensor networks [J]. IEEE Systems Journal, 2013, 7(1):13-25.

[4] Perrig, R. Szewczyk, J. D. Tygar, V. Wen and D. E. Culler. SPINs: security protocols for sensor networks. Wireless Networks, 2012, 8(5):521-34.

[5] Yang L, Ding C, Wu M. Rpida: recoverable riva cypreserving integrity-asserted data aggregation scheme for wireless sensor networks [J]. KSII Transactions on Internet and Information Systems, 2015, 9 (12):5189-5208.