EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

Network anomaly detection for railway critical infrastructure based on autoregressive fractional integrated moving average

Tomasz Andrysiak[1], Łukasz Saganowski[1] and Wojciech Mazurczyk[2*]

Abstract

The article proposes a novel two-stage network traffic anomaly detection method for the railway transportation critical infrastructure monitored using wireless sensor networks (WSN). The first step of the proposed solution is to find and eliminate any outlying observations in the analyzed parameters of the WSN traffic using a simple and fast one-dimensional quartile criterion. In the second step, the remaining data is used to estimate autoregressive fractional integrated moving average (ARFIMA) statistical models describing variability of the tested WSN parameters. The paper also introduces an effective method for the ARFIMA model parameters estimation and identification using Haslett and Raftery estimator and Hyndman and Khandakar technique. The choice of the "economically" parameterized form of the model was based on the compromise between the conciseness of representation and the estimation of the error size. To detect anomalous behavior, i.e., a potential network attack, the proposed detection method uses statistical relations between the estimated traffic model and its actual variability. The obtained experimental results prove the effectiveness of the presented approach and aptness of selection of the statistical models.

Keywords: Anomaly detection, Statistical model, Network traffic prediction, Critical infrastructure, Transportation system management

1 Introduction

Intelligent transportation systems (ITS) are currently a key technology that is identified as an answer to the growing need for mobility of goods and people. Owing to the use of ITS, it is possible to establish a fully functioning, accurate, real-time, and efficient transportation management system. It can be achieved by combining information systems and technologies like wireless networks and sensors, computing/networking devices, Global Positioning System (GPS), mobile telephony, and camera recognition systems. Thanks to ITS, it is possible to improve the level of services and capacity of transportation systems. In particular, ITS can help to enhance the transportation infrastructures, its overall safety, and security of critical information for different transportation means. It must be noted that currently, the main focus of academics and industry is on vehicular networks and more precisely on developing inter-vehicle and vehicle-to-infrastructure networks [1–3]. However, ITS are not limited only to manage vehicular traffic but they can also provide services and they can be successfully implemented in air, water transport, and rail systems [4, 5] as well.

An important aspect of any ITS is to correctly address potential security and privacy issues. Aijaz et al. [4] define the following vital attack aspects depending on what the target is. Authors identify attacks on the following: (i) wireless interface; (ii) sensor inputs to different processing units; (iii) software and hardware parts of the systems; and (iv) security infrastructure behind wireless access networks (e.g., certification and traffic authorities, transportation vehicle manufacturers). On the other hand, various security solutions have been proposed to tackle these problems and they can be classified [6] as *proactive* (e.g., tamper resistant hardware, proprietary

* Correspondence: wmazurczyk@tele.pw.edu.pl
[2]Institute of Telecommunications, Warsaw University of Technology, Warsaw, Poland
Full list of author information is available at the end of the article

system design, and digitally signed messages) or *reactive* (e.g., anomaly-based, context-based, and signature-based approaches). Especially anomaly-based detection systems constitute an important part of every ITS-based management system. They allow to assess the imminent emergence of any incidents, i.e., to detect deviations from normal patterns (events, situations). Therefore, identifying anomalous events is essential as they can lead to critical conditions where immediate actions must be taken.

In the existing literature, several anomaly detection approaches for ITS have been proposed. However, they have been mostly proposed for the vehicular networks, see e.g., [7–9]. However, very few solutions so far have been proposed for railway systems and they are mentioned below.

Rabatel et al. [10] focused on the field of train maintenance. Monitoring of trains is provided using sensors positioned on the main train components, e.g., motors wheels, to transmit information regarding, e.g., the temperature, acceleration, and velocity. Then an automatic detection system is introduced to identify anomalies in order to predict potential failures in advance. The proposed approach considers also the contextual criteria associated to railway data like weather conditions and itinerary.

Holst et al. [11] developed a statistical anomaly detection method which has been deployed in a tool which aim is to monitor train fleets and that allows inspecting and visualizing the occurrence of event messages generated on the trains. The designed anomaly detection component is based on the Bayesian Principal Anomaly [12] and aids operators to quickly find significant deviations from normal behavior and to detect early indications of potential problems.

Anomaly detection system that is able to indicate degraded condition of track and rolling stock has been proposed by Goodman et al. [13]. Authors utilized a sensor system installed on one of the 110 boxcars on a train on a high-tonnage loop test track. The data from the sensors was sent to a specialized collection gateway hub which was mounted inside the boxcar. The main goal was to discover abnormalities in railroad tracks, rolling stock, bearings, rotating shafts, and gears. The obtained results confirmed that such a detection system is efficient enough to identify, locate, and characterize such types of anomalies.

Considering the above, in this article we introduce another type of a novel anomaly detection system dedicated for wireless sensor networks (WSN) traffic that is based on clustering and statistical model with long memory, i.e., autoregressive fractional integrated moving average (ARFIMA). The main idea of the proposed approach is to analyze the deviations between parameters of the real network traffic and the estimated statistical models of that traffic. We develop a two-stage anomaly detection method. The first step is to find and eliminate possible observations outlying from the analyzed WSN traffic parameters. This step is performed by means of a simple yet effective one-dimensional quartile criterion. In the second step, the remaining data are used to estimate the ARFIMA statistical models describing variability of the tested WSN parameters.

The proposed anomaly detection method is used as a security measure for railroad gates tracking system that is based on WSN sensors. It is a part of a more comprehensive system responsible for supervising and visualization of the critical infrastructure, i.e., railroad crossings. The feasibility and effectiveness of the introduced method is proved based on the abovementioned real-life railroad crossings monitoring system; however, it must be noted that it can be conveniently ported to any other transportation system as well.

The rest of this paper is structured as follows. The next section focuses on the main security issues related to WSN. Then in Section 3, the assumed scenario as well as details of the proposed anomaly detection approach is outlined. Section 4 presents the real-life setup as well as obtained experimental results. Finally, Section 5 concludes our work.

2 Overview of security in WSN

Ensuring security of the WSNs is an important factor for their correct operation due to the fact that they are distinctly sensitive to hazards emerging from human intentional actions that include illegal use or incapacitation (e.g., impersonating or eavesdropping the user, terrorist attacks) [14], or from environmental influence, for instance fire, electromagnetic signal, etc. [15]. When comparing cable networks to WSNs, the latter offer restricted computational abilities and limited energy resources [16].Taking this into account, maintaining safe operation of WSN is difficult, nevertheless necessary. It is because the networks must be able to uphold their basic functioning which consists in collecting data from the sensors and transmitting it to both the monitoring unit and the WSN infrastructure management centers.

Wireless sensor networks might be subjected to different types of attacks (either passive or active). Passive attacks happen when an intruder does not utilize signal emissions aimed at disturbing the proper operation of WSN in order to, e.g., access its data, infrastructure, or modify transmitted messages. On the other hand, active attacks rely on utilization of such emissions of signals or actions that may be detected [15, 17], while trying to obtain an unauthorized access or a possibility to alter the messages.

During passive attacks on WSN, the intruder aims at passive interception of the exchanged network traffic to implicitly acquire the transmitted data. An example of such action is eavesdropping of the data that is transmitted between the nodes. The WSN radio medium, because of its specific features, is relatively vulnerable to such attacks. Another instance may be the network traffic analysis which intends to examine and disclose the WSN topology. A characteristic aspect of wireless sensor networks is a great load of information transmitted through a part of their nodes. If this data transmission is increased on these nodes, the neighboring nodes counteract by retransmitting the information to the base station. Due to the network traffic analysis, the attacker may obtain knowledge of how much workload the sensor network critical nodes are burdened with [18, 19].

Contrary to the passive methods, active attacks enable the intruder to directly or indirectly influence data content transmitted in the WSN. Moreover, the active attacks can be easily detected because their impact on WSN performance is direct, i.e., they may degrade the WSN quality, or even deny access to some services or completely disenable control over the network. For the network critical infrastructure, direct attacks on WSN hardware are especially dangerous. Such attacks can cause diminishing of the monitoring area of the sensor network, or entire disposal of the WSN [19, 20].

The aim of manipulating WSN nodes is to distract the operator of the sensor network from the main origin of the threat. e.g.. from spoofing or distributed denial of service ((D)DoS) attacks. Moreover, if the attacker uses a short-term high-energy electromagnetic pulse then an annihilation of either the given sensor network or any electronic device within the EMP destruction field [21] is possible.

The attacks aimed at the data's confidentiality or integrity constitute an immense threat because they let the intruder enter the network without authorization to transmit data. The *Sybil Attack* is an example of a masking technique which consists in spoofing by transmitting numerous identifiers through a harmful confluence, or framing a legal confluence and taking over its specification to obtain access to the WSN infrastructure [22].

The (D)DoS attacks in WSN aim at excessive charging of the attacked sensor networks features in order to disenable data gathering from the attacked nodes or to restrain the efficiency of tools provided by the victim WSN. The (D)DoS invasions can be directed onto every level of the network model, i.e., ISO/OSI [21].

Because the WSN are vulnerable to a large number of dangers, limiting the possibility of a successful attack requires the use of advanced methods and algorithms. These methods are spread spectrum techniques (hindering the successful interfering with radio transmission); methods using cryptographic algorithms (securing the confidentiality and integrity of the transferred information); a proper nodes' device construction to deny access to their internal systems (for instance, information about cryptographic algorithms, or keeping secret keys); utilizing related protocols (documents guarding the transfer of information); and ways of monitoring and identifying abnormalities in the WSN data transfer [23–25].

In this paper, we suggest the two stages of analysis of abnormal behavior detection for WSN traffic. The first step prepares proper data by removing the outlier values. In the second step, the parameters of ARFIMA statistical model are used for detecting anomalies. In the course of stages, performed according to particular scenarios, the satisfying results were obtained. The following scenarios of anomaly attacks are analyzed and calculated for the sake of efficient protection of railroad crossings: (i) electromagnetic distortion; (ii) intentional damage of selected infrastructure; and (iii) attacks performed by means of the important WSN component, i.e., the WSN IP gateway.

3 Network anomaly detection: the proposed approach

In the rest of the paper, we assume the scenario depicted in Fig. 1 in which there is a management system that is utilized for visualizing and controlling railroad crossings critical infrastructure. The required infrastructure for monitoring a single railroad crossing consists of WSN sensors used for analyzing the state and position of the railroad crossing separate gates, WSN IP gateway which aggregates the traffic from sensors, firewall, and multi-WAN router for providing different links to the WAN network.

As mentioned in the introduction, an anomaly detection system is a vital component of any ITS management solution. That is why the Intrusion Detection/Prevention Systems (IDS/IPS) for detecting attacks and/or intrusions are currently utilized as one of the main components to provide security of the critical infrastructure. Their main function is to accurately identify, detect, and respond to an unauthorized activity directed against the protected network resources [26].

Generally, we may classify IDS/IPS based on the utilized threat identification technique to signature-based or anomaly detection systems. The first consists in the detection of intrusions using the signature of previously known attacks [27]. Comparatively, the latter relies on monitoring the defended system and to detect any abnormalities. Thus, any deviation from a defined model or profile of legitimate activities reflected in the WSN network traffic parameters is treated as a symptom of the attack. Such a deviation from normal reference is called an anomaly [14, 28].

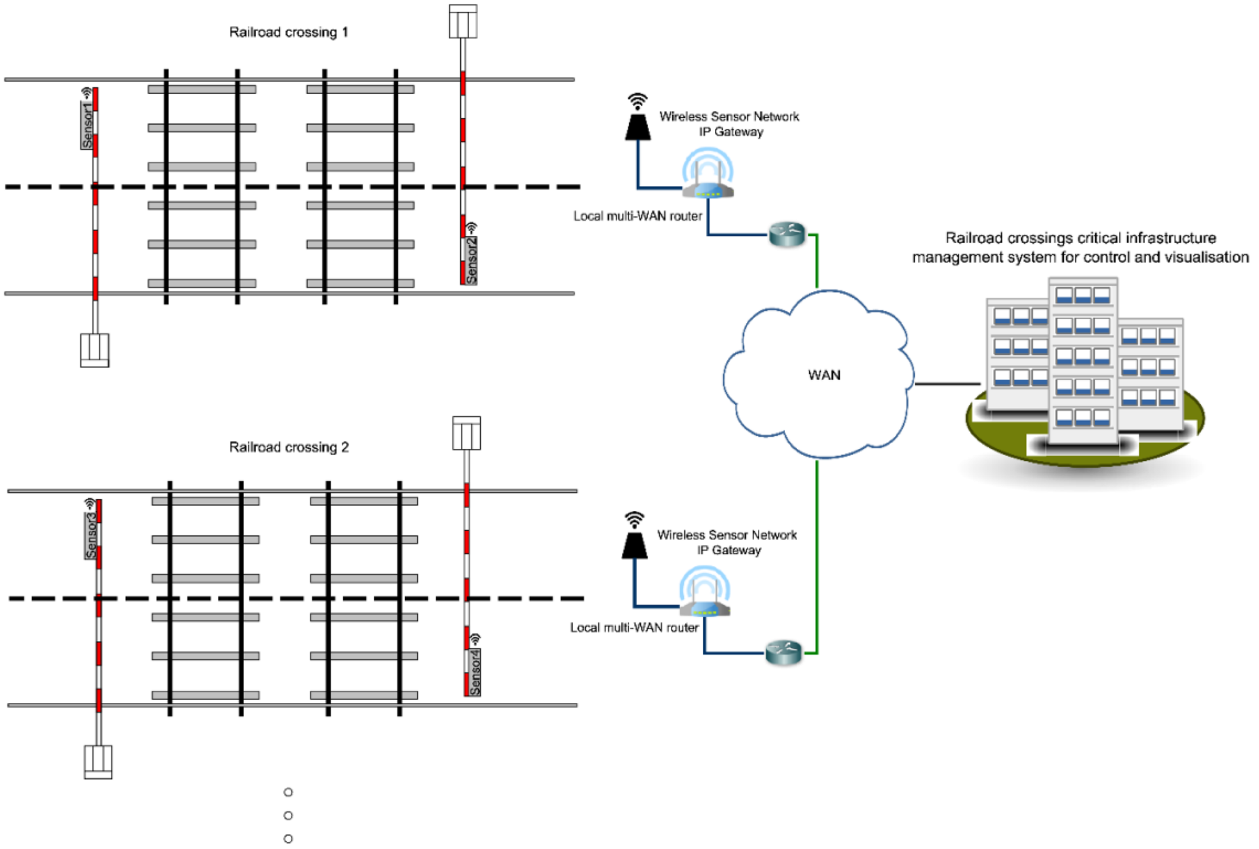

Railroad crossing 1
Sensor1
Sensor2
Wireless Sensor Network IP Gateway
Local multi-WAN router
Railroad crossings critical infrastructure management system for control and visualisation
WAN
Railroad crossing 2
Sensor3
Sensor4
Wireless Sensor Network IP Gateway
Local multi-WAN router

Fig. 1 Railroad crossings critical infrastructure management system—an overview

The outstanding profit of abnormal behavior noticing solutions is the fact that they are able to detect unknown intrusions interrupting correct network traffic parameters. Therefore, ID/PS relying on anomaly detection are (if properly configured) more effective than signature-based ones [29].

Considering the above, in this article the recognition of abnormal behavior is applied. This approach is based on the idea of analyzing the deviations of parameters of the real network traffic from the estimated statistical models of that traffic (see Fig. 2). We suggest a two-stage anomaly detection method. In the first step, ARFIMA model base for the analyzed WSN network traffic parameters is built. This is realized on the formerly selected and calculated features of the network traffic. In the following steps, outlier observations are eliminated and estimation of ARFIMA models parameters of the analyzed WSN network traffic features is performed. In result, statistical models base is created and serves as a basis for an anomaly detection system. The second step is a normal operation of an anomaly detection system (ADS), i.e., selection and calculation of the relevant network traffic features, and assessment of the difference between the actually transmitted data (i.e., network traffic) and the calculated ARFIMA representation of the traffic for the chosen WSN network parameters.

The motivation for choosing the ARFIMA statistical model was based on the results of previous authors' research, i.e., on the use of autoregressive models and heteroscedastic and regression models with variable sampling resolution of the dataset for anomaly detection in the LAN/WAN networks. Findings included in [30–32] clearly indicate the superiority of the ARFIMA model for modeling network traffic parameters' variability for the purpose of anomaly detection.

It must be emphasized that the first stage is always initiated after performing any change in WSN network infrastructure or topology. It can also be performed periodically to update the statistical models base, which is a basis for the anomaly detection system. However, the elimination procedure of outliers' observations (realized at this stage) disables degradation of ARFIMA models by rejecting non-standard parameters of the analyzed network traffic.

Below we present and discuss main components of the proposed approach in details.

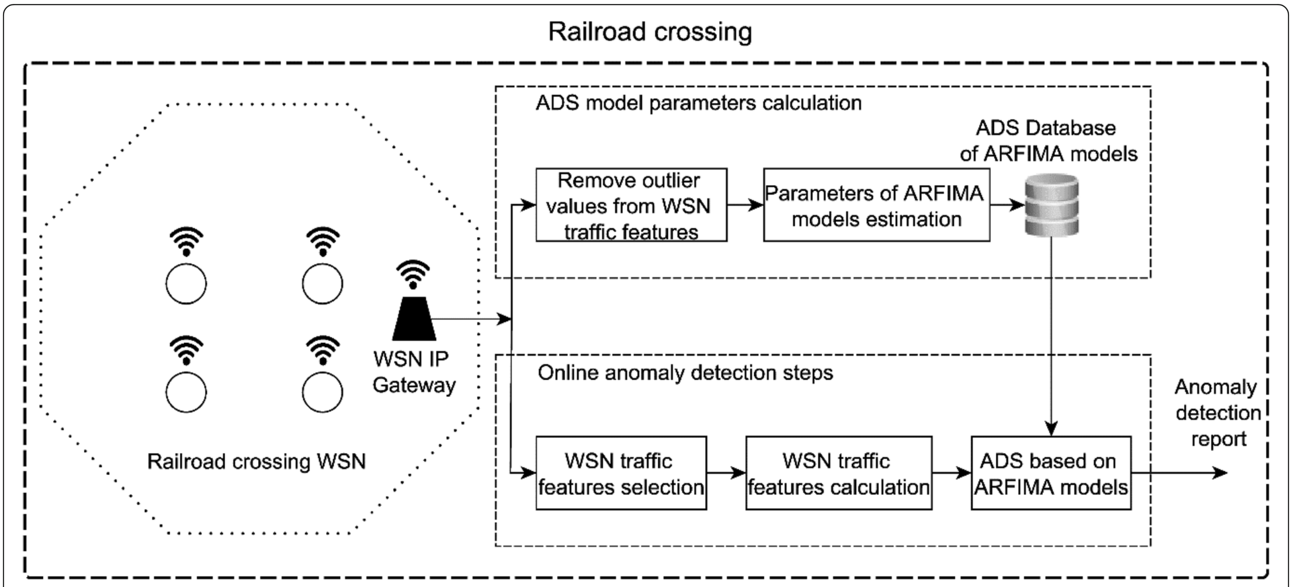


Fig. 2 Anomaly detection system for WSN traffic

3.1 Detection of outlying observations—one-dimensional quartile criterion

Due to the nature of the transportation critical infrastructure and its monitoring using WSN, there is a real hazard of fluctuation of the analyzed network traffic parameters, i.e., possibility of emerging outlying observations (outliers). These fluctuations may have diverse sources, for instance (i) environmental—connected with interference of radio wave propagation, (ii) technical—related to changes in the infrastructure, (iii) devices' damage; or (iv) they can be a consequence of network attack.

In our approach, identification of the outliers of the analyzed WSN traffic parameters is performed by means of one-dimensional quartile criterion introduced by Tukey [33], which is used for the construction of box plots. For every parameter, we calculate the first (Q1) and third (Q3) quartile and interquartile range IRQ = Q3 – Q1. Quartiles divide all our observations into four equal-number groups (Fig. 3, left).

The first quartile (Q1) divides observations in respect of 25–75 %, which means that 25 % are lower or equal to Q1, and 75 % of observations are equal or greater that Q1. The second quartile (Q2), otherwise known as the median, divides observations into 50–50 % proportion. The third quartile (Q3) divides the observations in respect of 75–25 %, which means that 75 % of observations are lower or equal to Q3, and 25 % are equal or greater that Q1. Observations which can be considered as outliers are those whose values exceed the range (Q1 – 1.5IRQ, Q3 + 1.5IRQ). In contrast, observations of extreme outliers (see Fig. 3, right) are identified as those for which the attributes are outside the range (Q1 – 3IRQ, Q3 + 3IRQ).

3.2 The ARFIMA model—estimation of the WSN traffic features variability

Grange, Joyeux [34], and Hosking [35] introduced a model called the autoregressive fractional integrated moving average (ARFIMA) which is composed of the two different processes, i.e., fractional differenced noise and auto regressive moving average. ARFIMA's aim is to examine the attribute of long memory, and for data presented as time series $\left\{y_{t_n}\right\}$, it is:

$$\Phi(B)(1-B)^d y_{t_n} = \Theta(B)\epsilon_{t_n}, \quad t_n = 1, 2, \ldots \mathrm{T}_n, \tag{1}$$

where $\epsilon_{t_n} \tilde{} (0, \sigma^2)$ is the statistic process (white noise process) with zero mean and variance σ^2, $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is the autoregressive polynomial, $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is the moving average polynomial, B is the backward shift operator, and $(1 - B)^d$ is the fractional differencing operator explained by the particular binomial expansion $(1-B)^d = \sum_{l=0}^{\infty} \begin{pmatrix} d \\ l \end{pmatrix} (-1)^l B^l$ and $\begin{pmatrix} d \\ l \end{pmatrix} (-1)^l = \frac{\Gamma(d+1)(-1)^l}{\Gamma(d-l+1)\Gamma(l+1)} = \frac{\Gamma(-d+l)}{\Gamma(-d)\Gamma(l+1)}$. The gamma function is marked by $\Gamma(*)$ and the number of differences necessary to present the stationary series is marked with d. The d^{th} power of the differencing operator, included in Eq. (1) is marked with $(1 - B)^d$.

When the value of the differencing parameter is in range (−0.5, 0.5), the ARFIMA model can be described as stationary, and if the value of the differencing parameter belongs to (0, 0.5), the process is characterized as a long-memory behavior. If there are suitable k differences, it is possible to transform many non-stationary processes into stationary ones by fulfilling condition (1). Consequently,

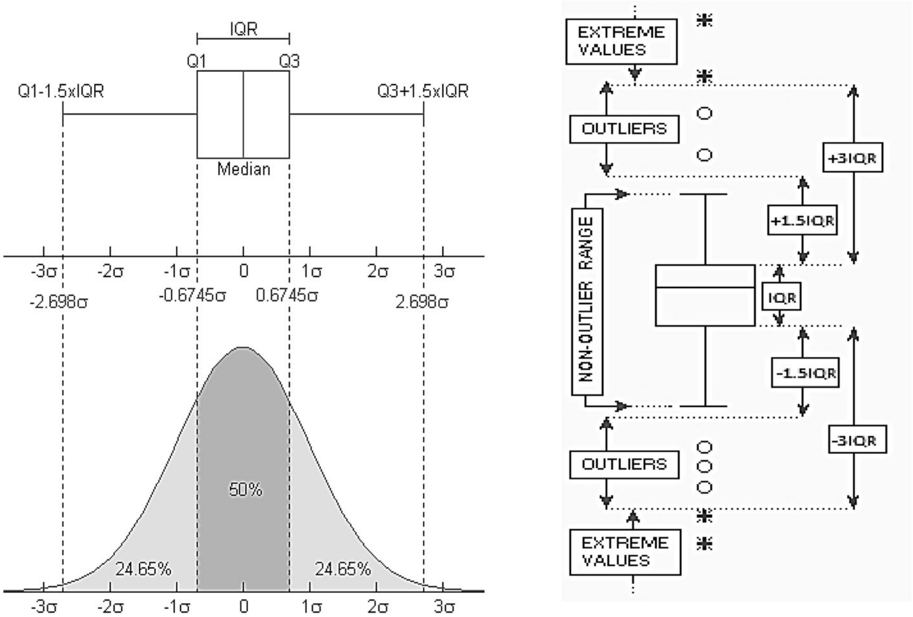


Fig. 3 Identification of outliers: (*left*) box plot and the normal distribution of observation, (*right*) ranges of outliers and extreme values

the non-stationary processes obtain the long-memory attribute [36].

It is possible to predict the ARFIMA processes by means of an infinite autoregressive representation of formula (1), recorded as $\Pi(B)y_{t_n} = \epsilon_{t_n}$, also

$$y_{t_n} = \sum_{i=1}^{\infty} \pi_i \, y_{t_n - i} + \epsilon_{t_n} , \quad (2)$$

where $\Pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \ldots = \Phi(B)(1 - B)^d \Theta(B)^{-1}$.

From the perspective of numerical realization, the above equation requires truncation after k lags; nevertheless, it is not easy to obtain. The difficulty in truncation will influence the forecast horizon included in predictions (see [36]). Formula (2) explains that the predicting rule absorbs the impact of the remote lags, by which it captures their persistent impact. However, if shifts appear in the process, the pre-shift lags will also influence the prediction, and in consequence, the post-shift horizons may have some biases [37, 38].

3.2.1 Estimation and selection of parameters of the ARFIMA model

To find a proper prognostic model, contrary to using the highest number of precise parameters that describe the variability of the analyzed data presented as time series, it is of crucial importance to understand that too large adjustment of series may provide either the description of the signal itself or the random noise (that may show accidental regularity in a definite number of attempts). Therefore, the main aim is to find a model which, with the use of a limited number of statistically important parameters, will be able to describe the essential features of the analyzed time series.

There are two relatively simple and effective methods for calculation of the autoregressive models' parameters: maximum likelihood estimation (MLE) and quasi-maximum likelihood estimation (QMLE) [39, 40]. For the MLE, the basic computational problem is finding the solution of the following equation:

$$\frac{\partial \log(L_T(\vartheta))}{\partial \theta} = 0, \quad (3)$$

where θ is the calculated data set, $L_T(\vartheta)$ is the likelihood function, and T is the quantity of modeled parameters controls. For many cases, it is impossible to find analytical solution to Eq. (3) for the defined form of the model; thus, the numerical estimation is used. Using the maximum likelihood method requires establishing the complete model, hence the formed estimator's sensitivity to possible mistakes in procedure of the auto regressive (AR) and moving average (MA) polynomials that define the dynamics of the process.

A universal criterion for selecting model's form does not exist. The common practice is that mapping the model onto the data is most optimal when the model's likelihood function and level of complexity increase concurrently. Nevertheless, there is a bigger possibility of an error occurrence when a greater number of parameters is being estimated. Therefore, one should seek an optimization of the quantity of parameters that appears in the described model along with wanted likelihood function value. Akaike (AIC), Schwarz (SIC) or Hannan-Quinn (HQC) propose choosing the "economical" model's form on the grounds of informational criterion, i.e., they suggest selecting the form that possesses the minimal value of information criterion [41].

Below we present results of parameter estimation obtained by means of MLE and QMLE methods, and the ARFIMA form of the model. The differentiation parameter d value is calculated with the use of the mentioned techniques and the HR estimator, which is developed by Haslett and Raftery [42]. Furthermore, we calculated the selection of the row of the analyzed model using exponential smoothing in the state space and such information factors as Akaike (AIC) (see Hyndman and Khandakar [43]). Owing to the above approach, we were able to obtain satisfactory computational efficiency and automatic realization of the used algorithms.

3.2.2 MLE method in estimation of parameter d

The analysis of an ARFIMA process Z_t from the perspective of the Gaussian Log-likelihood presented by formula (1) that refers to

$$\log G_L(z;\varrho) = -0.5\left[n\log(2\pi) + \log|\Sigma(\varrho)| + z^t\Sigma^{-1}(\varrho)z\right], \tag{4}$$

with $z = (z_1, z_2, \ldots, z_n)^t$ being the vector described by parameter, and $\varrho = (\sigma^2, H_e)$, Σ describes the $n\ x\ n$ covariance matrix of Z relying on z and ϱ, H_e denotes the Hurst exponent and where the determinant of Σ is described by $|\Sigma|$. The MLE of $\hat{\varrho}$ may be calculated by the maximum value of $\log G_L(z;\varrho)$ respectively to ϱ.

The calculation of first partial derivative of formula (4) has been described by

$$G_L^{'}(z;\varrho) = -0.5\left[\frac{\partial}{\partial\varrho}\log|\Sigma(\varrho)| + z^t\left[\frac{\partial}{\partial\varrho}\Sigma^{-1}(\varrho)\right]z\right]. \tag{5}$$

The maximum likelihood estimation $\hat{\varrho}$ creates the result of the $G_L^{'}(z;\varrho) = 0$. Provided that the parameters present high dimension, or there is a long time series, it is difficult to compute the exact MLE due to its numerical instability, for the formula (5) stimulates the estimation of the determinant and the elements of matrix Σ are inversed [36, 44, 45].

Out of numerous analogous MLE methods that can be conveyed by calculation of approximation of the likelihood function, we decided to use the HR estimator based on a quick and precise Haslett and Raftery's method [42], whose heuristic idea is to achieve autoregressive approximations. Such autoregressive, infinite order process may represent a Gaussian ARFIMA. However, since the quantity of samples is definite, the truncated model is obtained in accordance with $m < t \le n$,

$$Z_t - \varrho_1 Z_{t-1} - \cdots - \varrho_m Z_{t-m} = \epsilon_t, \tag{6}$$

with ϱ being the coefficients of the formula $\Phi(B)\Theta(B)(1-B)^d$. Since approximating as well as refining are performed, a QMLE of $\hat{\varrho}$ is brought about by the operation of maximization

$$G_L^*(z;\varrho) = C - 0.5n\log\left(\hat{\sigma}_\epsilon^2(\varrho)\right), \tag{7}$$

where C is a constant, $\hat{\sigma}_\epsilon^2(\varrho) = 0.5\sum_{t=1}^{n}\frac{\left(Z_t - \hat{Z}_t\right)}{\overline{\nu_t}}$, $\nu_t = var(Z_t - \hat{Z}_t)$, $\hat{Z}_t = \Phi(B)\Theta(B)\sum_{i=1}^{t-1}\omega_{ti}\,Z_{t-i}$ and $\omega_{ti} = -\begin{pmatrix} t \\ i \end{pmatrix}\frac{\Gamma(i-d)\Gamma(t-d-i+1)}{\Gamma(-d)\Gamma(t-d+1)}$. A more extensive study on this approximation method can be found in [42].

3.2.3 The calculation and selection of model features

The ways of exponential smoothing of the models of state space are obtained as follows:

$$b_t = W(a_{t-1}) + R(a_{t-1})\epsilon_t, \tag{8a}$$

$$a_t = F(a_{t-1}) + G(a_{t-1})\epsilon_t, \tag{8b}$$

where $\{\epsilon_t\}$ is a Gaussian white noise process with zero mean and variance σ^2, and $\mu_t = W(a_{t-1})$. The sample containing additional mistakes has $R(a_{t-1}) = 1$, and consequently, $b_t = \mu_t + \epsilon_t$. The analyzed model with multiplicative mistakes has $R(a_{t-1}) = \mu_t$, and consequently, $b_t = \mu_t(1 + \epsilon_t)$. Hence, $\epsilon_t = (b_t - \mu_t)/\mu_t$ creates the mistake related to the multiplicative model. The models created as a result of this action are not distinctive. Apparently, each value of $R(a_{t-1})$ results in creating identical prediction points for b_t.

The values of a_0 and the parameter ϑ are necessary for these models to be useful in terms of forecasting. Hence, calculating the likelihood of the improvements of models of state space (8) does not create difficulties, and so is achieving the maximum likelihood estimates:

$$G_L^*(\vartheta;a_0) = n\log\left(\sum_{t=1}^{n}\epsilon_t\right) + 2\sum_{t=1}^{n}\log|R(a_{t-1})|. \tag{9}$$

It is easy to calculate (9) with the recursive equations in [43, 46]. As far as multiple sources of mistakes of the state space models are concerned, it is necessary to apply the Kalman filter to estimate likelihood; our calculations are free of that requirement. The sets of the parameters ϑ and the initial states a_0 are realized by operation of computing of minimization of G_L^*.

The method for choosing the present model is based on the Akaike criterium

$$V_{AIC} = G_L^*\left(\hat{\vartheta}; \hat{a}_0\right) + 2n_p, \tag{10}$$

where V_{AIC} is the value of AIC, n_p creates the quantity of parameters in ϑ along with the quantity of free states in a_0, and $\hat{\vartheta}$ and $\hat{a}_0$ define the calculations of ϑ together with a_0. From the models applicable for the data, we selected the one that minimizes the AIC.

On the basis of the mentioned ideas we achieve an efficient and commonly appropriate algorithm for automatic predicting. To summarize, the stages of the undertaken performances are as follows [43]:

- Stage 1: optimize the parameters, i.e., smooth them and the initial state variable, for every series use all the matching models
- Stage 2: choose the most effective model in terms of AIC
- Stage 3: create point forecasts with the use of most applicable model (with parameters that are created in the optimization process) for any stages in advance as necessary

A thorough explanation of the proposed method is described in Hyndman and Khandakar work [43].

4 Experimental installation and results: railway crossings critical infrastructure management system

In this section, we describe our experimental setup which has been implemented on a real-life railroad transportation system. Using this installation, a set of experiments has been performed to prove that the approach proposed in this paper is feasible and effective.

4.1 Experimental setup

As mentioned above, experimental results presented in this paper have been obtained using real-world installation placed on the active railroad crossings. Railroad gates tracking component that is based on WSN sensors is a part of the more comprehensive system for supervising and visualization railroad crossings. The presented system is an original solution for supervising critical infrastructure of railroad crossings.

In Fig. 4, main components of the control and visualization system for railroad crossings is illustrated.

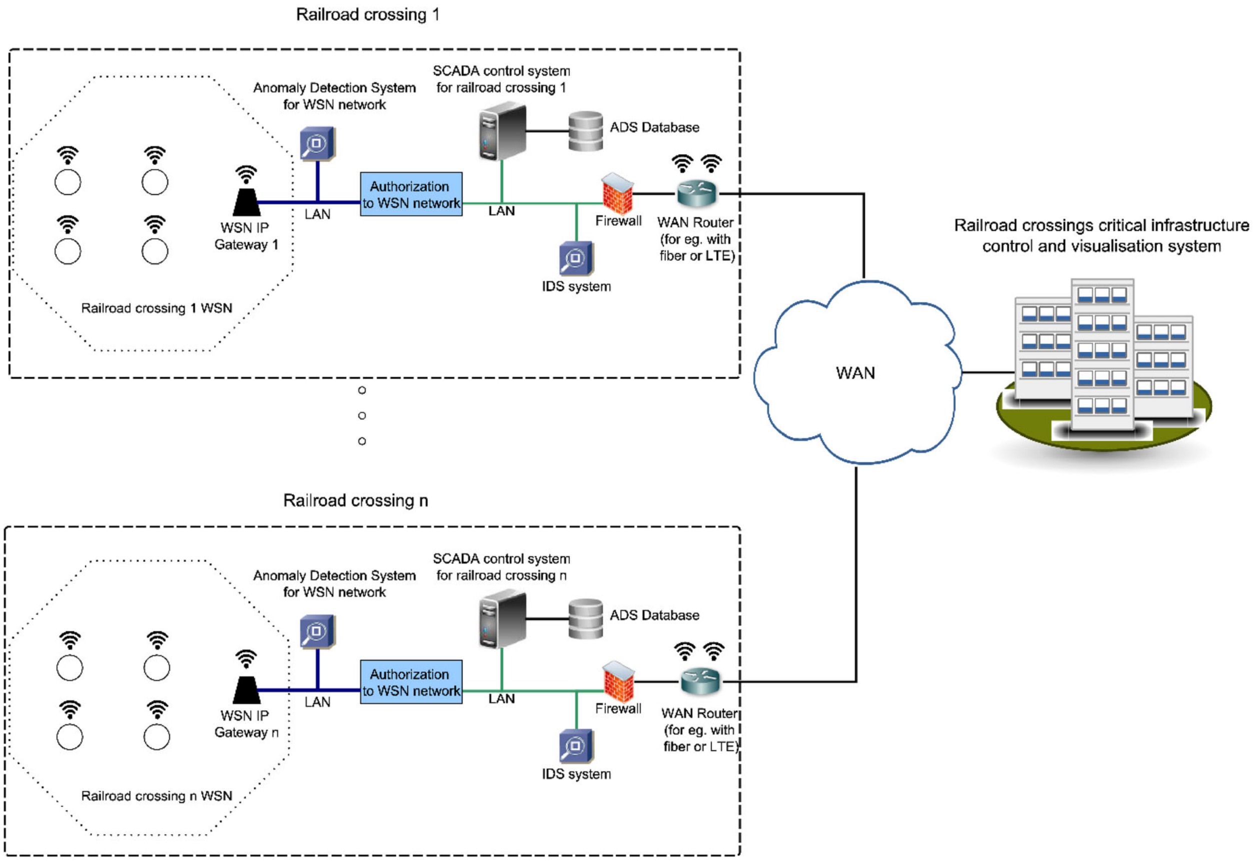

Railroad crossing 1
Anomaly Detection System for WSN network
SCADA control system for railroad crossing 1
ADS Database
Authorization to WSN network
LAN
LAN
WSN IP Gateway 1
Firewall
WAN Router (for eg. with fiber or LTE)
IDS system
Railroad crossing 1 WSN
Railroad crossings critical infrastructure control and visualisation system
WAN
Railroad crossing n
Anomaly Detection System for WSN network
SCADA control system for railroad crossing n
ADS Database
Authorization to WSN network
LAN
LAN
WSN IP Gateway n
Firewall
WAN Router (for eg. with fiber or LTE)
IDS system
Railroad crossing n WSN

Fig. 4 Railroad crossings critical infrastructure management system with the use of WSN for the railroad gates state and position supervising

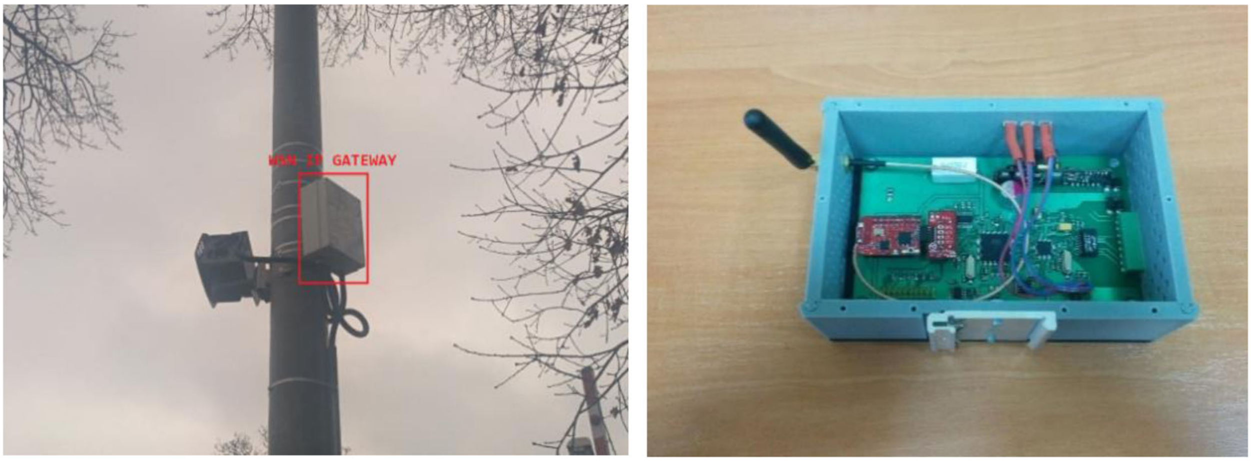


Fig. 5 WSN IP gateway (*right*) installed on the railroad crossing column (*left*)

Telecommunication infrastructure for one railroad crossing consists of WSN sensors used for analyzing the state and position of the railroad crossing separate gates, industrial SCADA computer, classic Intrusion Detection System (IDS) using previously known attacks signatures database, firewall, and multi-WAN router for providing different links to the WAN network. In our solution, we propose an anomaly detection system (ADS) for the WSN part that is complementary to the implemented classic IDS. ADS system obtains the WSN traffic from Ethernet link provided by the WSN IP gateway. As already mentioned proposed in this paper, a novel detection approach is based on the statistical model with long memory—ARFIMA. Parameters of the ARFIMA models obtained for different traffic features are stored in ADS database. For every railroad crossing, a separate ADS instance exists and the same telecommunication infrastructure. WAN routers are used for communication with control and visualization management application. Railroad crossings were situated on the same rail link and connected to the control and visualization

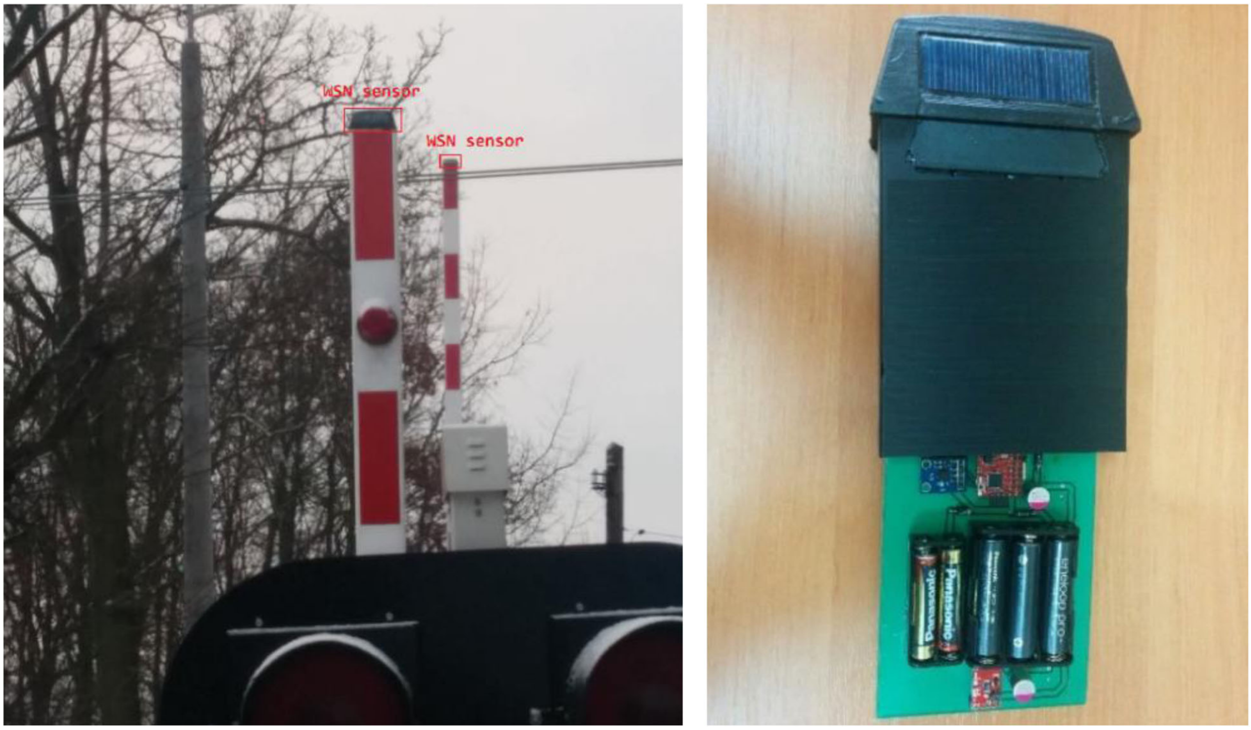
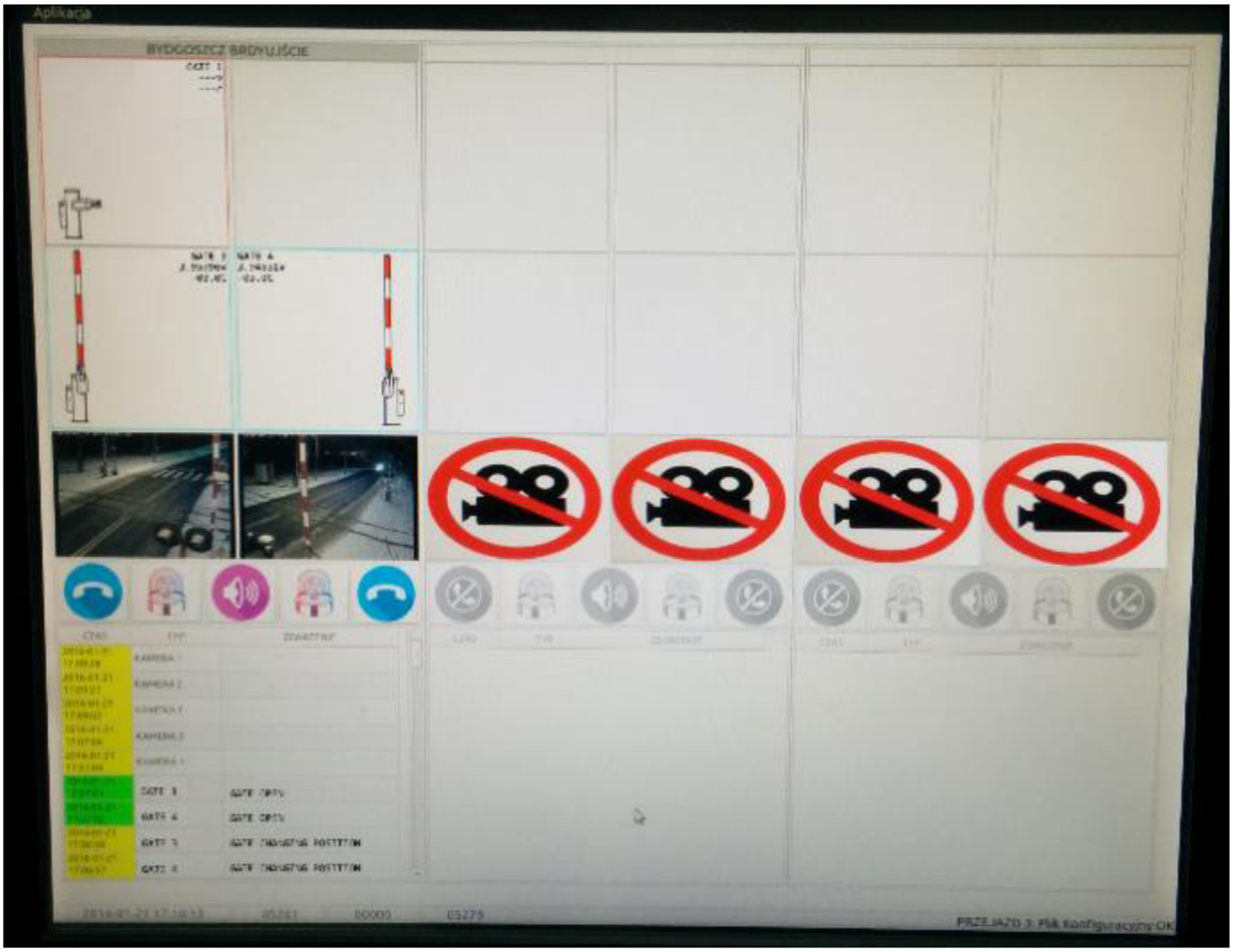


Fig. 6 WSN sensors (*right*) installed on railroad gates (*left*)



CZAS	TYP	ZDARZENIE
[illegible]	KAMERA 1	
[illegible]	KAMERA 2	
[illegible]	KAMERA 1	
[illegible]	KAMERA 2	
[illegible]	KAMERA 1	
[illegible]	GATE 3	GATE OPEN
[illegible]	GATE 4	GATE OPEN
[illegible]	GATE 3	GATE CHANGING POSITION
[illegible]	GATE 4	GATE CHANGING POSITION

2016-01-21 17:18:13 05261 00000 05275

PRZEJAZD 3: Plik konfiguracyjny OK

Fig. 7 Part of the railroad crossings management application with railroad gates visualization

system by fiber or radio WAN connection, e.g., by means of long-term evolution (LTE).

Practical realization of WSN IP gateway is depicted in Fig. 5 where a printed circuit board and the gateway installed on one of the railroad crossing column are presented.

WSN sensors are installed on the top of railroad gates. Installed sensors on the railroad crossing and a sensor printed circuit board are presented in Fig. 6.The sensor is powered by battery banks and additionally supported by a small solar panel. Static position/tilt of the railroad gate is measured by a Microelectromechanical Systems (MEMS) sensor which provides position in three dimensions—x, y, and z. WSN sensors transmit packets in an idle state (gates are not moving) in approximately constant periods of time. In the idle state, we control physical presence of railroad gates and battery health, signal strength (RSSI), ambient temperature, and gates' three dimensional position. In the idle state, insignificant railroad gates' movements resulting from, e.g., wind or vibrations caused by heavy vehicles are not taken into account. For the sake of reliability, packets from sensors are received by two redundant IP gateways. Every sensor transmits packets to the WSN IP gateway when triggered by railroad gates' movements.

In Fig. 7, part of the railroad crossings management application with railroad gates visualization is depicted. One

Table 1 WSN traffic features captured from railroad crossings through the WSN IP gateway

Feature	WSN traffic feature description
F1	Battery supply voltage
F2	MEMS sensor position x
F3	MEMS sensor position y
F4	MEMS sensor position z
F5	RSSI of WSN sensor [dBm]
F6	The number of WSN packets per time period
F7	The number of WSN packets per time period during idle state of mems x, y, z position

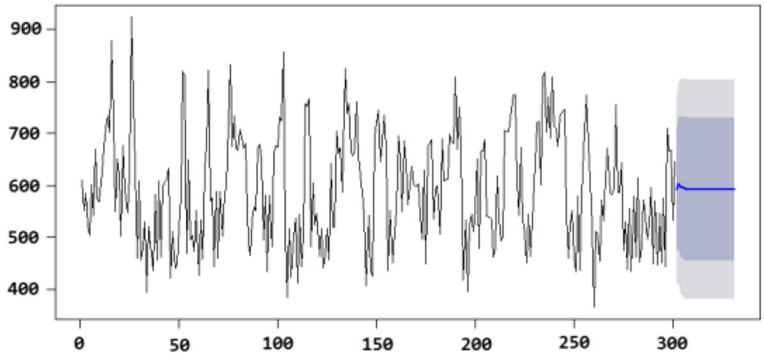


Fig. 8 30 samples forecasting horizon for F6 WSN traffic feature (the number of WSN packets per time period)

active railroad crossing with railroad gates visualization based on the data taken from WSN sensors is visible (left side of Fig. 7). The list below button ribbons provides information, for example, about the state and the position of separate railroad gates (e.g., gate open, closed, changing position, broken gate). As mentioned, presented in Fig. 7 screenshot is a part of the more comprehensive system for analyzing railroad crossings critical infrastructure which considers also, for example, video images.

4.2 Experimental results

As mentioned before, the proposed anomaly detection system comprises two main steps. In the first step, we remove outlier values (see Section 3.1) for every observed traffic feature (see Table 1). This step prepares data for the next step where parameters of the statistical model with long memory dependence are calculated. The second step is based on the calculation of parameters of ARFIMA statistical model (see Section 3.2).

We selected seven features which are related to the most important functionalities of railroad crossing critical infrastructure (see Table 1). For every WSN traffic feature, we achieve forecasting interval (30 samples forecasting horizon) based on ARFIMA model (see Figs. 8 and 9). Prediction intervals are described by mean value (line in the center of prediction interval), 80 % prediction interval (narrower interval), and 95 % prediction range (wider interval). Examples of prediction intervals for WSN traffic feature F6 (the number of WSN packets per time period) and F5 (received signal strength indication (RSSI) [dBm] of WSN sensor) are presented in Figs. 8 and 9, respectively.

WSN traffic features presented in Table 1 are captured from the Ethernet link of the WSN IP gateway. WSN sensors transmit packets to WSN IP gateways placed on both sides of the railroad crossing. WSN gateway converts received packets to IP packets. Then, packets converted by WSN IP gateway are captured in the next step by software sensor installed on railroad industrial computer. Every WSN traffic feature presented in Table 1 is extracted from packets captured by the IP network sensor.

In a subsequent step, WSN traffic features are in real time processed by the proposed anomaly detection solution that indicates possible anomaly/attack when the value of the online calculated traffic feature is outside an interval determined by two prediction intervals. When values for a given traffic feature are inside 80 % prediction intervals, we assume that there is no anomaly/attack for a given traffic feature. When WSN traffic features lie

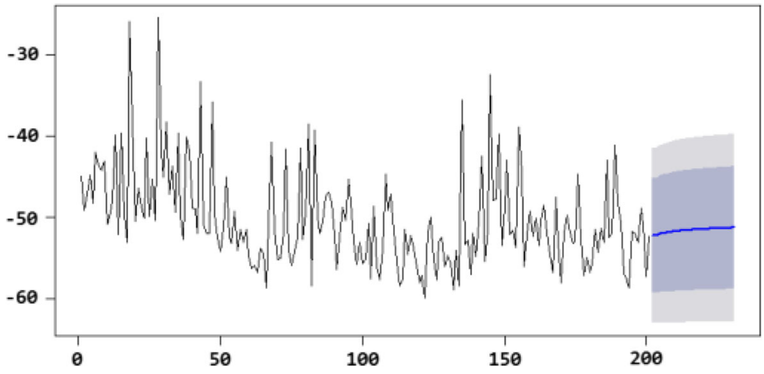


Fig. 9 30 samples forecasting horizon for F5 WSN traffic feature (RSSI of WSN sensor)

inside interval described by 80 to 95 % of prediction intervals, we treated this traffic as suspicious where an anomaly or attack can be present. Traffic features with values outside 95 % prediction intervals triggers anomaly/attack by anomaly detection algorithm.

The proposed anomaly detection method for WSN traffic has been tested with different anomaly/attack scenarios assumed. Because railroad crossing is a critical railway infrastructure anomaly/attacks had to be simulated and carefully controlled in order to preserve safety on active testing railroad crossings. In this paper, we evaluated the following anomaly/attack scenarios:

Scenario 1: electromagnetic distortion
Scenario 2: intentional damage of the selected infrastructure
Scenario 3: attacks performed by means of the WSN IP gateway

The proposed anomaly/attack scenarios have impact on different sets of traffic features from Table 1. Taking into account scenario 1, i.e., simulated electromagnetic distortions, will have impact on RSSI values—F5, and the number of packets which successfully reaches the IP gateway in a certain period of time—features F6 and F7. Partial results of detection rate (DR) [%] and false positives (FP) [%] for scenario 1 are presented in Table 2:

Scenario 2 can be understood as a situation where, for example, railroad gates will be hit or bent by a vehicle (but sensors are still able to communicate with the IP gateway). In this scenario, the most noticeable impact will be seen for features responsible for measurement of three dimensional positions of railroad gates—F2, F3, and F4. Results for this scenario are presented in Table 3.

A different variant of scenario 2 covers situations where railroad gates will be moved outside the railroad crossing area, but the sensor will not be damaged or sensors and railroad gates will be completely damaged. In these cases, we can observe an impact on features F2, F3, F4, F5, F6, and F7. Results can be observed in Table 4.

Table 2 DR[%] and FP[%] for attacks performed on WSN network with scenario 1

Feature	DR[%]	FP[%]	Description
F1	–	–	No impact for scenario 1
F2	–	–	No impact for scenario 1
F3	–	–	No impact for scenario 1
F4	–	–	No impact for scenario 1
F5	94.20	5.20	
F6	94.00	8.40	
F7	93.00	8.20	

Table 3 DR[%] and FP[%] for attacks performed on WSN network with scenario 2

Feature	DR[%]	FP[%]	Description
F1	–	–	No impact for scenario 2
F2	98.00	1.6	
F3	97.00	2.8	
F4	97.00	3.0	
F5	–	–	Negligible impact for scenario 2
F6	–	–	Negligible impact for scenario 2
F7	–	–	Negligible impact for scenario 2

In scenario 3, WSN IP gateway was used to perform an attack. The aim of this attack was to drain batteries or delay packets from sensors. This attack requires the knowledge of the specific communication protocol between sensors and the IP gateway. F1, F6, and F7 are features influenced by this attack scenario (see Table 5).

Experimental results presented in Table 6 contains aggregated detection rate together with false positive for all seven WSN traffic features (see Table 1). Based on the obtained results, an overall performance of the proposed ADS solution can be observed taking into account all anomaly/attack scenarios described earlier. Most of the anomaly/attacks have been successfully identified. Detection rates varied between 93 and 98 %, while false positive rates were below 9 %. The best results have been achieved for WSN features: F2, F3, and F4 calculated based on readings from MEMS sensors and feature F1 (but only for scenario 3).

5 Conclusions

Ensuring a proper level of security for resources and systems of critical infrastructure, particularly transportation ones, realized as sensor radio networks is currently an intensively explored research topic. It is apparent that WSN, due to their nature, are vulnerable to a substantial number of threats originating both from the outside and inside of their own infrastructure. Therefore, these networks require ensuring the integrity and confidentiality

Table 4 DR[%] and FP[%] for attacks performed on WSN network with scenario 2

Feature	DR[%]	FP[%]	Description
F1	–	–	No impact for scenario 2
F2	97.00	2.1	
F3	96.20	2.4	
F4	96.00	3.1	
F5	95.20	4.8	
F6	92.40	9.1	
F7	93.50	8.6	

Table 5 DR[%] and FP[%] for attacks performed on WSN network with scenario 3

Feature	DR[%]	FP[%]	Description
F1	98.00	2.0	
F2	–	–	No impact for scenario 3
F3	–	–	No impact for scenario 3
F4	–	–	No impact for scenario 3
F5	–	–	No impact for scenario 3
F6	93.40	8.4	
F7	94.40	9.4	

of the transmission, as well as protection of nodes and data transferred with their use. While developing mechanisms, algorithms, or protocols that increase transmission security in WSN, one also needs to consider the restrictions imposed by the unique characteristics of WSN, such as self-organization, dislocation, equipment limitations, and ease of fiasco of nodes and protocols. The increasing number of novel attacks, their global scope, and complexity level enforce dynamic development of network security systems. The most often implemented solution aiming at ensuring security are detection and classification methods that allow to identify abnormal behaviors reflected in the analyzed network traffic.

The advantage of such an approach is the protection against so far unknown attacks, developed specially (targeted attacks) in order to realize attacks onto particular resources of network infrastructures or simply constituting so called zero-day exploits. Anomaly detection systems may play a crucial role in those environments. Their purpose is to detect (for auto-response) unusual traffic behavior representing symptoms of unauthorized actions directed against protected critical infrastructure resources, implemented as WSN networks.

That is why in this paper we introduced a novel network traffic anomaly detection method for a critical railway transportation infrastructure which is utilizing sensor radio network. In order to detect anomalies, differences between the actual network traffic and the estimated ARFIMA model of that traffic for the analyzed WSN network parameters were used. For the purpose of suitable preparation of data for statistical modeling, observations outlying in the analyzed WSN network parameters with the use of a simple and fast one-dimensional quartile criterion were found and eliminated. Parameter estimation and identification of the row of the ARFIMA statistical models were realized as a compromise between the model's coherence and the size of its estimation error. The obtained experimental results performed on the real-life railway crossings infrastructure confirm efficacy and accuracy of the presented anomaly detection method. We achieved overall detection rates varied between 93 and 98 %, while false positive rates were below 9 %. Most valuable WSN features for anomaly detection purposes were F1, F2, and F3 and they were based on reading from MEMS sensor. In case of sensor failure, we took into account other undamaged sensors readings in order to ensure system successful operation. To conclude, we consider utilization of an efficient IDS system as a must in every railway critical infrastructure management system.

Table 6 Overall DR[%] and FP[%] for attacks performed on WSN network

Feature	DR[%]	FP[%]
F1	98.00	2.00
F2	97.50	1.85
F3	96.60	2.60
F4	96.50	3.05
F5	94.70	5.00
F6	93.27	8.63
F7	93.63	8.73

Future work will be related to further exploring the most efficient set of parameters used for the proposed network traffic anomaly detection method. Moreover, we also plan experiments on a greater scale and during a longer time period which will allow to further tune the proposed solution and to model the abnormal behaviors even better.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Institute of Telecommunications and Computer Science, University of Technology and Life Sciences, Bydgoszcz, Poland. [2]Institute of Telecommunications, Warsaw University of Technology, Warsaw, Poland.

Received: 5 May 2016 Accepted: 2 October 2016
Published online: 13 October 2016

References

1. R Naja, A survey of communications for intelligent transportation systems. Wireless Vehicular Netw. Car Collision Avoidance. (2013), pp. 3–35. http://dx.doi.org/10.1007/978-1-4419-9563-6_1.
2. S An, B Lee, D Shin, *A Survey of Intelligent Transportation Systems* (Third International Conference on Computational Intelligence, Communication Systems and Networks, Bali, 2011), pp. 332–337
3. G Karagiannis, O Altintas, E Ekici, G Heijenk, B Jarupan, K Lin, T Weil, Vehicular networking: a survey and tutorial on requirements, architectures, challenges, standards and solutions. IEEE Commun. Surv. Tutorials. **13**(4), 584–616 (2011)
4. K Qureshi, A Abdullah, A survey on intelligent transportation systems. Middle-East J. Sci. Res. **15**(5), 629–642 (2013)
5. AB Bochow, F Dotzer, A Festag, M Gerlach, R Kroh, T Leinmuller, *Attacks on Inter Vehicle Communication Systems—an Analysis. 3rd International Workshop on Intelligent Transportation*, 2006
6. T Leinmuller, E Schoch, C Maihofer, *Security Requirements and Solution Concepts in Vehicular Ad Hoc Networks. 4th Annual Conference on Wireless On demand Network Systems and Services*, 2007

7. J Fadlil, HK Pao, YJ Lee, Anomaly detection on ITS data via view association. ACM SIGKDD Workshop. Outlier. Detect. Description. (2013), pp. 22–30. http://doi.acm.org/10.1145/2500853.2500859.
8. E Kwon, S Noh, M Jeon, D Shim, *Scene Modeling-based Anomaly Detection for Intelligent Transport System. 4th International Conference on Intelligent Systems, Modelling and Simulation*, 2013, pp. 252–257
9. J Raiyn, T Toledo, Real-time road traffic anomaly detection. J. Transp. Technol. **4**, 256–266 (2014)
10. J Rabatel, S Bringay, P Poncelet, Anomaly detection in monitoring sensor data for preventive maintenance. Expert Syst. Appl. **38**, 7003–7015 (2011)
11. A Holst, M Bohlin, J Ekman, O Sellin, B Lindström, S Larsen, Statistical anomaly detection for train fleets. AI Magazine. **34**(1), 33–42 (2012)
12. A Holst, J Ekman, in *Proc. of the 11th Scandinavian Conference on Artificial Intelligence*, ed. Kofod-Petersen A., HF, and H., L. Incremental Stream Clustering for Anomaly Detection and Classification (2011), p. 100-107
13. DL Goodman, J Hofmeister, R Wagoner, *Advanced Diagnostics and Anomaly Detection for Railroad Safety Applications Using a Wireless, IoT-Enabled Measurement System. Ridgetop Inc. Whitepaper*, 2011
14. A Perrig, J Stankovic, D Wagner, Security in wireless sensor networks. Commun. ACM **47**(6), 53–57 (2004)
15. J Undercoffer, S Avancha, A Joshi, J Pinkston, Security for sensor networks. CADIP Res. Symp. (2002).
16. B Strulo, J Farr, A Smith, Securing mobile ad hoc networks—a motivational approach. BT Technol. J. **21**(3), 81–89 (2003)
17. E Cayirci, C Rong, *Security in Wireless Ad Hoc and Sensor Networks* (John Wiley & Sons, Ltd., 2009)
18. D Culler, A Perrig, R Szewczyk, JD Tygar, V Wen, *SPINS: Security Protocols for Sensor Networks* (Proceedings of Seventh Annual International Conference on Mobile Computing and Networks, 2001)
19. S Slijepcevic, M Potkonjak, V Tsiatsis, S Zimbeck, MB Srivastava, *On communication security in wireless ad-hoc sensor networks, 11th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 2002, pp. 139–144
20. HKD Sarma, A Kar, *Security Threats in Wireless Sensor Networks* (Elsevier, 2006)
21. AD Wood, JA Stankovic, Denial of service in sensor networks. IEEE Comput. **35**(10), 54–62 (2002)
22. J Newsome, E Shi, D Song, A Perrig, *The Sybil attack in sensor networks: analysis & defences. Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks*, (ACM Press 2004) pp. 259–268
23. YC Hu, A Perrig, DB Johnson, Packet leashes: A defense against wormhole attacks in wireless networks. In Proceedings - IEEE INFOCOM. **3**, 1976–1986 (2003).
24. C Karlof, D Wagner, Secure routing in wireless sensor networks, attacks and countermeasures. Ad Hoc Sensor Netw. **1**, 293–315 (2003)
25. J Deng, R Han, S Mishra, *INSENS: Intrusion-Tolerant Routing in Wireless Sensor University of Colorado, Department of Computer Science Technical Report CU-CS-939-02*, 2002
26. V Chondola, A Banerjee, V Kumar, Anomaly detection: a survey. ACM Comput. Surv. **41**(3), 1–72 (2009)
27. M Esposito, C Mazzariello, F Oliviero, SP Romano, C Sansone, Evaluating pattern recognition techniques in intrusion detection systems. Proceedings of the 5th International Workshop on Pattern Recognition in Information Systems PRIS, 144–153, (2005)
28. L Caviglione, M Gaggero, J-F Lalande, M Urbanski, W Mazurczyk, Seeing the unseen: revealing mobile malware hidden communications via energy consumption and artificial intelligence. IEEE Trans. Inf. Forensics Secur. **11**(4), 799–810 (2016)
29. W Dargie, C Poellabauer, Fundamentals of Wireless Sensor Networks (John Wiley & Sons, Ltd., 2010)
30. T Andrysiak, Ł Saganowski, Network anomaly detection based on ARFIMA model, image processing & communications, challenges 6. Adv. Intell. Syst. Comput. **313**, 255–261 (2014)
31. T Andrysiak, Ł Saganowski, A Marchewka, A comparative study of statistical models with long and short-memory dependence for network anomaly detection, image processing & communications, challenges 7. Adv. Intell. Syst. Comput. **389**, 255–265 (2015)
32. T Andrysiak, Ł Saganowski, *Network Anomaly Detection Based on Statistical Models with Long-Memory Dependence, Proceedings of the Tenth International Conference on Dependability and Complex Systems DepCoS-RELCOMEX*, 2015, pp. 1–10
33. JW Tukey, *Exploratory Data Analysis* (Addison-Wesley, Boston, 1977)
34. CWJ Granger, R Joyeux, An introduction to long-memory time series models and fractional differencing. J. Time Ser. Anal. **1**, 15–29 (1980)
35. J Hosking, Fractional differencing. Biometrika **68**, 165–176 (1981)
36. N Crato, BK Ray, Model selection and forecasting for long-range dependent processes. J. Forecast. **15**, 107–125 (1996)
37. P Brockwell, R Davis, Introduction to time series and forecasting. (Springer Verlag, 2002)
38. VJ Gabriel, LF Martins, On the forecasting ability of ARFIMA models when infrequent breaks occur. Econometrics J. **7**, 455–475 (2004)
39. F Sowell, Maximum likelihood estimation of stationary univariate fractionally integrated time series models. J. Econometrics **53**, 165–188 (1992)
40. Y Hosoya, The quasi-likelihood approach to statistical inference on multiple time series with long-range dependence. J. Econometrics **73**, 217–236 (1996)
41. HJ Bierens, Information Criteria and Model Selection (Pennsylvania State University, 2006)
42. J Haslett, AE Raftery, Space-time modelling with long-memory dependence: assessing Ireland's wind power resource (with Discussion). Appl. Stat. **38**(1), 1–50 (1989)
43. RJ Hyndman, Y Khandakar, Automatic time series forecasting: the forecast Package for R. J. Stat. Softw. **27**(3), 1–22 (2008)
44. JA Beran, Statistics for Long-Memory Processes (Chapman and Hall, 1994)
45. GE Box, MG Jenkins, *Time series analysis forecasting and control* (Holden-Day, San Francisco, 1976)
46. GE Box, MG Jenkins, G Reinsel, *Time series analysis* (Holden-Day, San Francisco, 1970)