I.J. Information Engineering and Electronic Business, 2018, 3, 15-22
Published Online May 2018 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2018.03.03

A Regression based Sensor Data Prediction Technique to Analyze Data Trustworthiness in Cyber-Physical System

Abdus Satter
Institute of Information Technology, University of Dhaka, Dhaka 1000, Bangladesh
Email: bit0401@iit.du.ac.bd

Nabil Ibtehaz
Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka 1000, Bangladesh
Email: nabilibtehaz2@gmail.com

Received: 10 November 2017; Accepted: 08 January 2018; Published: 08 May 2018

***Abstract*—**A Cyber-Physical System strongly depends on the sensor data to understand the current condition of the environment and act on that. Due to network faults, insufficient power supply, and rough environment, sensor data become noisy and the system may perform unwanted operations causing severe damage. In this paper, a technique has been proposed to analyze the trustworthiness of a sensor reading before performing operation based on the record. The technique employs regression analysis to select nearby sensors and develops a linear model for a target sensor. Using the linear model, target sensor reading is predicted in a particular time stamp with respect to each nearby sensor's reading. If the difference between the predicted and actual value is within a given limit, the reading is considered as trustworthy for the corresponding nearby sensor. At last, majority consensus is taken to consider the reading as trustworthy. To evaluate the proposed technique, a data set containing temperature reading of 8 sensors for 24 hours was used where first 90% data was used for nearby sensor selection and linear model construction, and rest 10% for testing. The result analysis shows that the proposed technique detects 19, 69, and 73 trustworthy data from 73 records with respect to 3%, 4% and 5% deviation from actual reading.

***Index Terms*—**Cyber Physical System, Sensor Data Trustworthiness.

I. Introduction

The demand of Cyber-Physical System (CPS) is increasing day by day [1]. Currently CPS is used to perform critical operations in various domains such as medical, battle ground, traffic monitoring, etc. Usually, CPS consists of two major parts, a physical process and a cyber system [2]. Physical part is responsible for gathering data from the deployed environment through sensors and executing operation by actuators as instructed by cyber part. On the other hand, cyber part collects data from the network transferred by sensors, analyzes the data, and delivers response in the form of instruction to actuators. Since all the operations are decided and executed by CPS based on the sensor data, it is important to ensure trust worthiness of the data before transmitting instructions to the actuator [3]. This increases the reliability of CPS.

In real world scenario, hundreds or even thousands of sensors are deployed on the physical environment which continuously capture data from the environment. Those data are transmitted to cyber part of a CPS through sensor network. Some part of the data can be noisy or untrustworthy during data collection by sensors, and propagation in the sensor network. Data can loose trustworthiness during data collection when corresponding sensors are damaged due to rough environment, inadequate power supply, or degradation of sensor power gradually [4]. Data can also lack validity during propagation in the network due to network errors or faults [5]. If those faulty data are processed and decision is made based on that, CPS will loose trustworthiness for doing inappropriate actions. Such actions may cause serious damage when CPS is employed for critical and sensitive operation [6].

Literature contains several methods to analyze trustworthiness of sensor data. Tang et. al. proposed a technique named TrueAlarm for trustworthiness analysis in sensor network [7]. In this technique, all the nearest neighbors' reading and their distances are utilized for validating the data of a particular sensor. However, in a real world environment, it is not possible at all to find the exact distance of all the sensors. Thus, incorrect calculation of distance may lead to false alarm in CPS. Another technique was proposed by Kebin et al. to identify non-faulty sensor in a sensor network [8]. The technique employs bayesian approach where assumption is that all the deployed sensors are in the same position and the environment is smooth. However, in a practical scenario, sensors are spread throughout the environment and it is not possible to have smooth environment. As a result, such assumption may cause the CPS recognizing faulty sensors as non-faulty. Lim et al. proposed a game-theoretic defense technique to protect sensor data from the attackers and increase trustworthiness [9]. Tang et al. also proposed a technique named Intrumine to detect intruders in a sensor network [10]. However, if a sensor is faulty, the technique will consider that the faulty reading of the sensor is correct.

In this paper, a technique has been proposed to analyze trustworthiness of the data in a CPS. The technique comprises two steps, deriving linear regression model for target sensor with its neighbors and predicting sensor data for trustworthiness analysis. In order to construct regression model, a set of target sensors for which data trustworthiness will be analyzed, is constructed. For each target sensor, a set of neighbor sensors is selected which acts as the basis for regression model. Later, a linear model is formulated for each sensor with its neighbor set using a training dataset. To analyze the data trustworthiness, deviation is measured between the actual value and the predicted value obtained from the linear model. If deviation is within the expected limit, the corresponding sensor data is considered as trustworthy with respect to corresponding nearby sensors. At last, a particular reading of a target sensor is considered as trustworthy if more than half of the nearby sensors support the reading.

In order evaluate the proposed technique, eight sensors were used in the experiment which sense temperature in the experimental environment. Experimental dataset contains temperature reading of all eight sensors for 24 hours with two minutes interval. Here, first 90% of the dataset was used as training data set for nearby sensors selection and linear model construction and rest 10% was used for testing. Besides, linear model is constructed and values are generated for a single sensor which is considered as target sensor in the experiment. The result analysis shows that nearby sensors' readings are linearly related with target sensor. Again, out of 73 target sensor reading, the proposed technique detects 19, 69, and 73 readings as trustworthy with 3%, 4%, and 5% deviation from actual reading respectively.

This paper makes the following contributions:

- A technique has been proposed based on Regression model to ensure trustworthiness of sensor data
- Intensive experimental analysis on a real life dataset has been discussed to perceive the effectiveness of the proposed approach
- The technique can be used with other approaches that ensure security during data propagation to increase trustworthiness of the sensor data

The rest of the paper has been organized as follows. Section II discusses significant works on the trustworthiness of sensor data in a CPS. Section III presents the proposed approach that predicts sensor data based on the linear regression model and finds the faulty sensor readings. Section IV describes an intensive experimental analysis of the proposed approach. Section V concludes the whole work.

II. Related Work

Several techniques have been proposed in the literature to increase trustworthiness of a CPS. Some techniques focus on the security of the sensor network. Some researchers concentrated on the intruder detection and prevention in a CPS. Few works have been carried out to identify faulty sensors in the sensor network. The most significant works on the trustworthiness of the sensor data have been described below.

Tang et al. proposed a technique named TrueAlarm which analyzes the trustworthiness of sensor data in a sensor network [7]. Usually, each sensor reading is important in CPS because instructions are given to actuators based on that reading. If the sensor reading is not trustworthy, this may lead to devastating situation. So, the technique, TrueAlarm employs distance among the sensors to detect which sensor data is trustworthy. More precisely, for trustworthiness analysis of a particular sensor reading, a number of nearby sensors of the target are selected and their values are compared with the target reading. The authors formulate equation which takes the distance between the target and nearby sensor, and predicts the value for the target sensor. If the difference between predicted value and actual value is within the threshold, the reading is marked as trustworthy. However, in real world scenario, it is difficult to find the exact location of a sensor which is required for distance calculation. If the distance cannot be calculated appropriately, the technique may provide wrong interpretation of sensor reading in terms of trustworthiness.

A Bayesian approach was proposed by Kebin et al. to select non-faulty sensor in a sensor network [8]. When sensors are deployed, some of the sensors are damaged due to critical environment or circumstances. Those sensors are needed to be detected and reading of these are required to be ignored when taking decision since such reading may cause catastrophic effect. The technique proposed in [8] separates the non-faulty sensors from the faulty sensors by employing a machine learning technique named Bayesian classification. Two assumptions are made for this technique such as the deployment environment is smooth and all the sensors are in the same location that is they are in a cluster. However, from practical point of view, it is difficult to have smooth environment and usually sensors are deployed throughout the environment. As real world scenarios do not abide by the assumptions, this technique cannot detect non-faulty sensor accurately. Even, it may provide false positive result due to wrong assumptions.

A technique named IntruMiner was proposed by Tang et al. to detect intruders in a sensor network [10]. The technique generates *monitoring graph* to model the relationships between sensors and intruders in large sensor network. Next, it calculates the signal strength and position of the intruders based on the link relationships in the generated *monitoring graph*. It calculates a confidence score for each detected potential intruder to reduce the false positive results. The technique was evaluated in synthetic and real datasets. The result analysis shows promising results in detecting intruders in an area that contains thousands of sensors. The technique also performs better than existing intruder detection techniques as per the experimental result analysis. However, the technique cannot recognize faulty sensors because it only detects the presence of intruders in a sensor network without considering the correctness of the sensor reading.

Literature contains some other works on secure data transportation in a sensor network. Perrig et al. proposed a secure protocol for sensor networks where sensors use a shared secret key for authentication [11]. Luk et al. proposed a secure sensor network communication architecture that uses block cipher mode [12]. Zhang et al. used Data Encryption Standard (DES) algorithm to provide security solutions for the network control systems [13]. They designed a Detection and Reaction technique based on the DES algorithm. The technique showed promising results against Denial of Service (DoS) attack in their experiment. All these techniques assume that all the sensors under study are non-faulty but such environment rarely exists in real life. It is required to identify faulty sensors and analyze the readings of non-faulty sensors to increase the trustworthiness in a Cyber-Physical System.

III. Proposed Approach

In this paper, a sensor data prediction technique has been proposed which employs linear regression model with nearest neighbor sensors. The technique comprises four steps which are *Target Sensor Selection*, *Nearest Neighbor Selection*, *Linear Regression Model Construction*, and *Data Prediction and Trustworthiness Analysis*. All these steps are described below.

A. Target Sensor Selection

A set of target sensors is selected for which data trustworthiness will be analyzed. Assume that the set of target sensors is as follows.

$T= \{s_{1,} s_{2,} s_{3, \ldots,} s_n\}$
$s_{i\,=}$ ith target sensor

B. Nearest Neighbor Selection

For each target sensor T_i, a set of K nearest neighbor sensors is selected. Assuming that more than half of the nearby sensors provide correct reading at a particular time. The nearest neighbor set for sensor T_i can be defined as follows.

$$N_i \subset S$$

C. Linear Regression Model Construction

A linear regression model is constructed for each sensor $T_i \in T$ by deriving linear equation with each nearby sensor N_i for T_i. The corresponding equation for linear regression model is shown as follows.

$$Y_{ij} = A_j * X_j + C_j$$

Y_{ij} = predicted value for sensor $s_i \in S$
A_i = coefficient that influence X_j
X_j= actual reading of sensor $t_j \in N_i$
C_j= bias

D. Data Prediction and Trustworthiness Analysis

For a given target sensor T_i, its reading can be predicted with respect to its neighbor sensor set N_i by employing the linear model. For sensor T_i, there are N_i number of linear equation, thus | N_i | number of reading will be predicted for T_i at a particular time period. For trustworthiness analysis, a deviation limit such as d is determined, that is the deviation between the actual value and predicted value will be within the limit in order to be trustworthy. For each predicted value, if the deviation within the limit, sensor T_i will earn a score. Here, higher score indicates higher trustworthiness of the data provided by T_i. If x is the actual value of T_i, y is the predicted value of T_i, *Limit* is the expected deviation, p_i is the predicted value of T_i with respect to ith nearby sensor in N_i, and n is the total number of nearby sensor, then following equations are used for score calculation.

$$devScore(x, y, Limit) = \begin{cases} 1 & abs(x-y) \leq Limit \\ 0 & otherwise \end{cases}$$

$$TotalScore = \sum_{i-1}^{n} devScore(x, p_i, Limit)$$

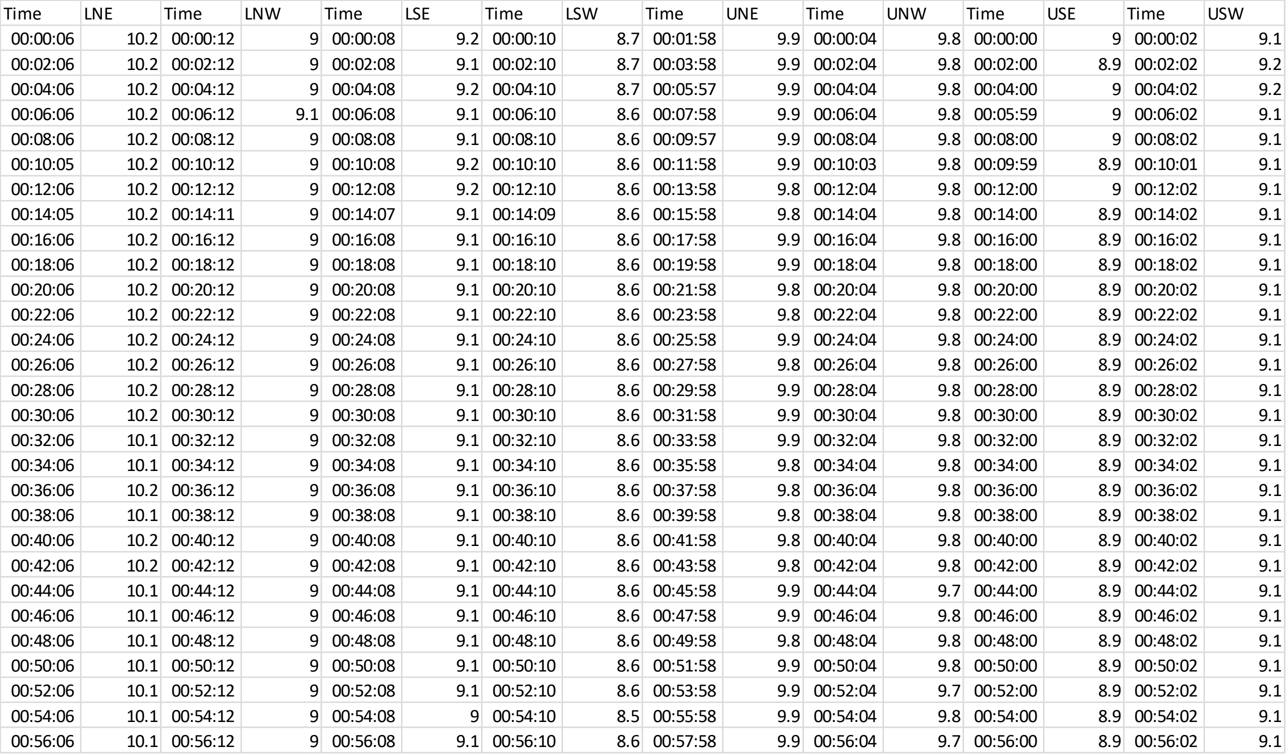
Sensor data of T_i will be considered as trustworthy if calculated score for T_i is more than half of the number of nearby sensors, that is,

$$TotalScore = \begin{cases} true & if\, a > \frac{b}{2} \\ false & otherwise \end{cases}$$

here,
a = Total Score
b = Number of Nearby Sensor

IV. Experimental Setup and Result Analysis

In order to evaluate the proposed approach, a control experiment was performed which includes a dataset of eight sensors. These sensors were used for sensing temperature in an environment. The dataset contains temperatures calculated by eight sensors in a single day. Temperature was sensed by each sensor with an interval of two minutes. The dataset is available in this link (https://tinyurl.com/yaywe7d2). A snapshot of the experimental dataset is shown in Fig. 1. In the figure, eight sensors are labeled as LNE, LNW, LSE, LSW, UNE, UNW, USE, and USW. In the experimental analysis, the proposed technique was used to predict the data of sensor LNE. Here, first 90% records was used for training and last 10% records was used for testing.



Time	LNE	Time	LNW	Time	LSE	Time	LSW	Time	UNE	Time	UNW	Time	USE	Time	USW
00:00:06	10.2	00:00:12	9	00:00:08	9.2	00:00:10	8.7	00:01:58	9.9	00:00:04	9.8	00:00:00	9	00:00:02	9.1
00:02:06	10.2	00:02:12	9	00:02:08	9.1	00:02:10	8.7	00:03:58	9.9	00:02:04	9.8	00:02:00	8.9	00:02:02	9.2
00:04:06	10.2	00:04:12	9	00:04:08	9.2	00:04:10	8.7	00:05:57	9.9	00:04:04	9.8	00:04:00	9	00:04:02	9.2
00:06:06	10.2	00:06:12	9.1	00:06:08	9.1	00:06:10	8.6	00:07:58	9.9	00:06:04	9.8	00:05:59	9	00:06:02	9.1
00:08:06	10.2	00:08:12	9	00:08:08	9.1	00:08:10	8.6	00:09:57	9.9	00:08:04	9.8	00:08:00	9	00:08:02	9.1
00:10:05	10.2	00:10:12	9	00:10:08	9.2	00:10:10	8.6	00:11:58	9.9	00:10:03	9.8	00:09:59	8.9	00:10:01	9.1
00:12:06	10.2	00:12:12	9	00:12:08	9.2	00:12:10	8.6	00:13:58	9.8	00:12:04	9.8	00:12:00	9	00:12:02	9.1
00:14:05	10.2	00:14:11	9	00:14:07	9.1	00:14:09	8.6	00:15:58	9.8	00:14:04	9.8	00:14:00	8.9	00:14:02	9.1
00:16:06	10.2	00:16:12	9	00:16:08	9.1	00:16:10	8.6	00:17:58	9.9	00:16:04	9.8	00:16:00	8.9	00:16:02	9.1
00:18:06	10.2	00:18:12	9	00:18:08	9.1	00:18:10	8.6	00:19:58	9.9	00:18:04	9.8	00:18:00	8.9	00:18:02	9.1
00:20:06	10.2	00:20:12	9	00:20:08	9.1	00:20:10	8.6	00:21:58	9.8	00:20:04	9.8	00:20:00	8.9	00:20:02	9.1
00:22:06	10.2	00:22:12	9	00:22:08	9.1	00:22:10	8.6	00:23:58	9.8	00:22:04	9.8	00:22:00	8.9	00:22:02	9.1
00:24:06	10.2	00:24:12	9	00:24:08	9.1	00:24:10	8.6	00:25:58	9.9	00:24:04	9.8	00:24:00	8.9	00:24:02	9.1
00:26:06	10.2	00:26:12	9	00:26:08	9.1	00:26:10	8.6	00:27:58	9.8	00:26:04	9.8	00:26:00	8.9	00:26:02	9.1
00:28:06	10.2	00:28:12	9	00:28:08	9.1	00:28:10	8.6	00:29:58	9.9	00:28:04	9.8	00:28:00	8.9	00:28:02	9.1
00:30:06	10.2	00:30:12	9	00:30:08	9.1	00:30:10	8.6	00:31:58	9.9	00:30:04	9.8	00:30:00	8.9	00:30:02	9.1
00:32:06	10.1	00:32:12	9	00:32:08	9.1	00:32:10	8.6	00:33:58	9.9	00:32:04	9.8	00:32:00	8.9	00:32:02	9.1
00:34:06	10.1	00:34:12	9	00:34:08	9.1	00:34:10	8.6	00:35:58	9.8	00:34:04	9.8	00:34:00	8.9	00:34:02	9.1
00:36:06	10.2	00:36:12	9	00:36:08	9.1	00:36:10	8.6	00:37:58	9.8	00:36:04	9.8	00:36:00	8.9	00:36:02	9.1
00:38:06	10.1	00:38:12	9	00:38:08	9.1	00:38:10	8.6	00:39:58	9.8	00:38:04	9.8	00:38:00	8.9	00:38:02	9.1
00:40:06	10.2	00:40:12	9	00:40:08	9.1	00:40:10	8.6	00:41:58	9.8	00:40:04	9.8	00:40:00	8.9	00:40:02	9.1
00:42:06	10.2	00:42:12	9	00:42:08	9.1	00:42:10	8.6	00:43:58	9.8	00:42:04	9.8	00:42:00	8.9	00:42:02	9.1
00:44:06	10.1	00:44:12	9	00:44:08	9.1	00:44:10	8.6	00:45:58	9.9	00:44:04	9.7	00:44:00	8.9	00:44:02	9.1
00:46:06	10.1	00:46:12	9	00:46:08	9.1	00:46:10	8.6	00:47:58	9.9	00:46:04	9.8	00:46:00	8.9	00:46:02	9.1
00:48:06	10.1	00:48:12	9	00:48:08	9.1	00:48:10	8.6	00:49:58	9.8	00:48:04	9.8	00:48:00	8.9	00:48:02	9.1
00:50:06	10.1	00:50:12	9	00:50:08	9.1	00:50:10	8.6	00:51:58	9.9	00:50:04	9.8	00:50:00	8.9	00:50:02	9.1
00:52:06	10.1	00:52:12	9	00:52:08	9.1	00:52:10	8.6	00:53:58	9.9	00:52:04	9.7	00:52:00	8.9	00:52:02	9.1
00:54:06	10.1	00:54:12	9	00:54:08	9	00:54:10	8.5	00:55:58	9.9	00:54:04	9.8	00:54:00	8.9	00:54:02	9.1
00:56:06	10.1	00:56:12	9	00:56:08	9.1	00:56:10	8.6	00:57:58	9.9	00:56:04	9.7	00:56:00	8.9	00:56:02	9.1

Fig.1. Snapshot of Experimental Dataset

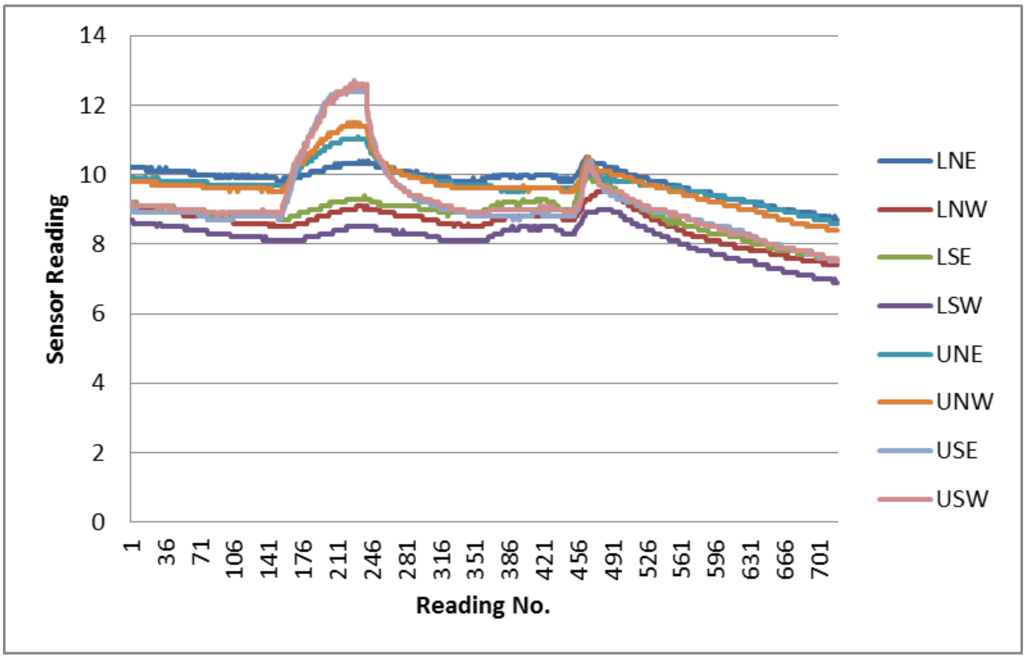


Fig.2. Trend Analysis of the Sensor Readings

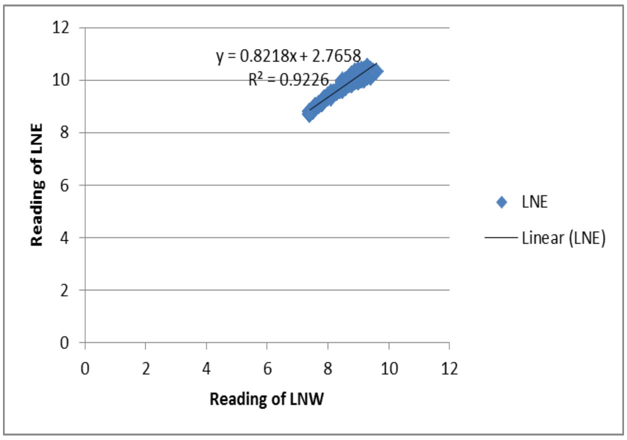


Fig.3. Linear Regression Model for LNE with respect to LNW

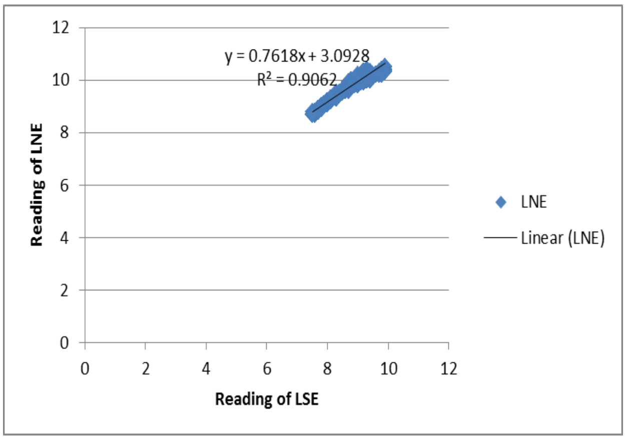


Fig.4. Linear Regression Model for LNE with respect to LSE

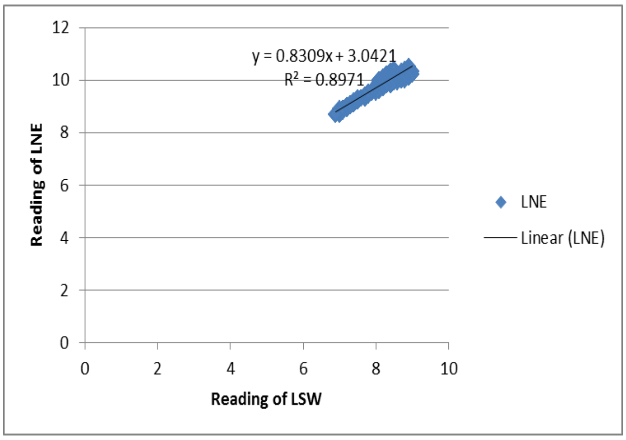


Fig.5. Linear Regression Model for LNE with respect to LSW

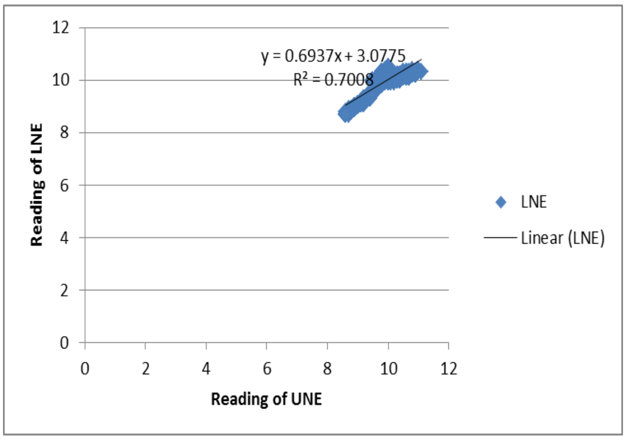


Fig. 6. Linear Regression Model for LNE with respect to UNE

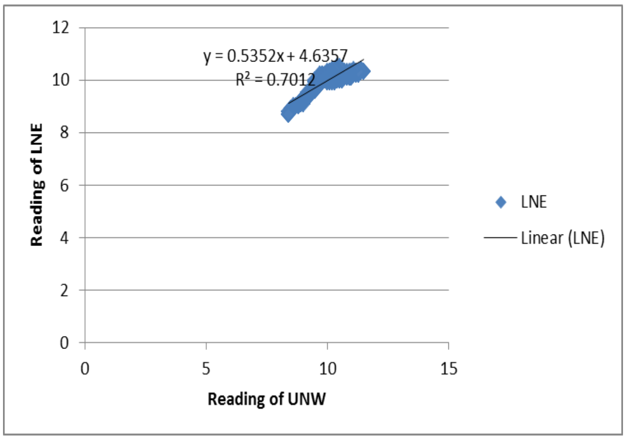


Fig. 7. Linear Regression Model for LNE with respect to UNW

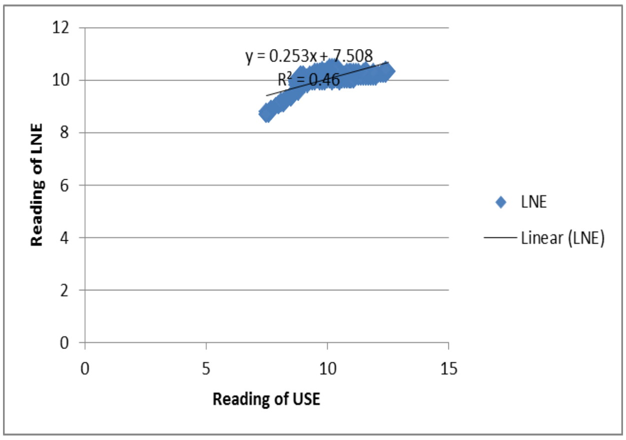


Fig. 8. Linear Regression Model for LNE with respect to USE

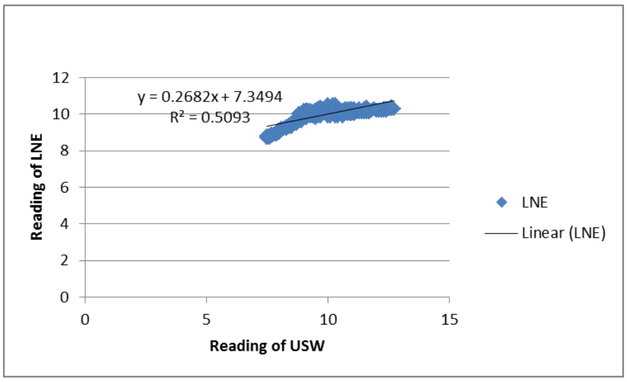


Fig. 9. Linear Regression Model for LNE with respect to USW

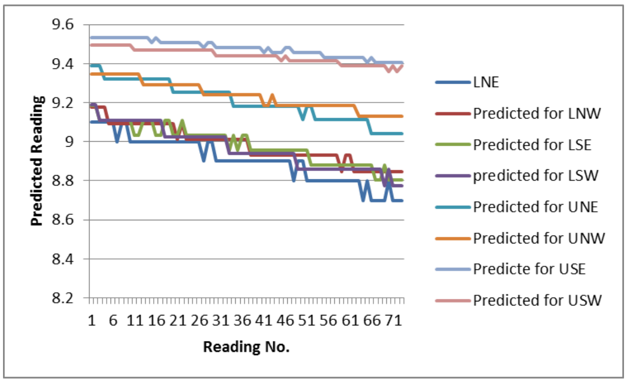


Fig.10. Actual Reading of LNE, and Predicted Reading of LNE with respect to Other Sensors

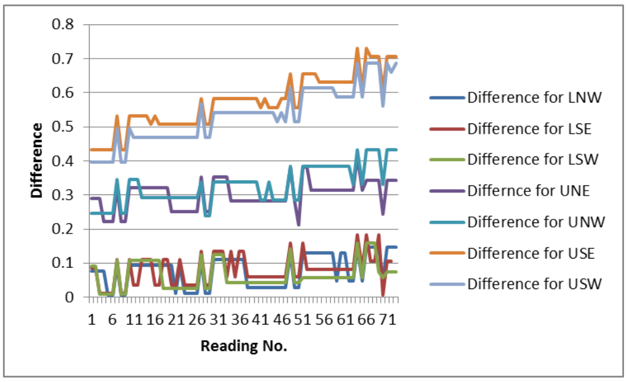


Fig.11. Trend Analysis for the Difference between Actual Reading and Predicted Reading of LNE with respect to Each Other Sensors

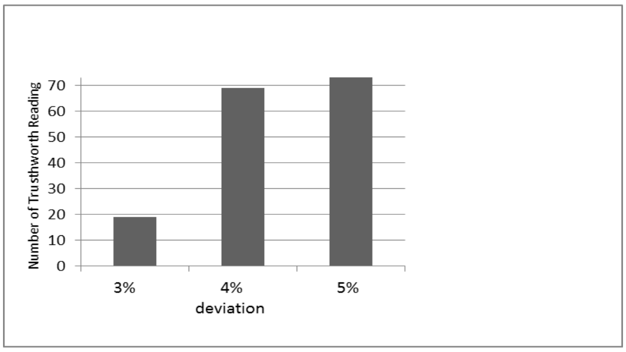


Fig. 12. Number of Trustworthy Data for Different Deviations

The technique starts with selection of nearby sensors as stated in the previous section. Usually sensors within the same region, provide data following a trend or pattern. For the selection of nearby sensors of LNE, a trend analysis is performed on the training dataset. The result is shown in Fig. 2. In this figure, it is seen that all the sensors are following approximately a common pattern or trend. So, LNW, LSE, LSW, UNE, UNW, USE and USW can be considered as nearby sensors of LNE.

The next part is to construct linear model for sensor LNE with respect to each of the rest 7 sensors. For this, a regression analysis is performed on the training dataset and linear equations are generated. Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig.7, Fig. 8, and Fig. 9 depict the regression analysis along with linear equation between LNE and other sensors, respectively.

After developing linear equations for the target sensor, LNE with respect to its nearby sensors, the testing phases was accomplished by predicting reading with these equations. Fig. 10 depicts the actual value of LNE and predicted value with respect to other sensors. In this figure, it is seen that the reading of LNE is very close to LSE, LSW and LNW. The reason is that these sensors are very close to LNE in comparison with other sensors like UNE, UNW, USE and USW. This figure provides evidence that the reading of LNE follows the trend with LSE, LSW and LNW. Fig. 11 illustrates the deviation between actual reading and predicted reading for LNE. The predicted value of LNE was calculated by using the linear equation with respect to the other sensors. The difference among the values with respect to LSE, LNE, LNW and LSW is maximum 0.18. That means predicted reading deviates maximum 2% with actual value when linear models of LNE with respect to LSE, LNE, LNW and LSW, respectively. Since close sensors sense the almost similar environment or object, LSE, LNE, LNW and LSW generate approximately similar data. On the other hand, deviation is between 0.2 to 0.75 (approximately 1.78% to 7%) for sensors UNE, UNW, USE, and USW due to the distance. However, still regression model application helps to predict the value of a sensor and in the experiment, the predicted value varies up to 7% from the actual value.

In order to consider whether a particular reading of a sensor is trustworthy or not, more than half of the nearby sensors need to support the value. In this experiment, sensor LNE has 7 nearby sensors and 7 predicted reading is generated against a particular reading of LNE. So, if at least 4 predicted values support the actual value with a given deviation, the actual value will be considered as trustworthy. Fig. 12 depicts the number of trustworthy data with different deviations. In this figure, it is seen that 19 reading is marked as trustworthy for deviation 3%, 69 for deviation 4% and 73 for 5% where the total number of test reading is 73. This provides evidence that the proposed technique outperforms with deviation 4% to 5%.

V. Conclusion

A faulty data may cause severe damage by a CPS because the cyber part of the system provides wrong instructions to actuator by analyzing the data [14]. So, data trustworthiness in a sensor network needs to be analyzed before processing and making decision by CPS. In this paper, a regression based technique is proposed for data trustworthiness analysis.

For a target sensor, the technique first selects the nearby sensors by employing regression analysis. Next, it constructs linear models with each of these sensors. Following the models, at a given time period, the target sensor reading is predicted with respect to each of the nearby sensors. The reading of the target sensor at that time period is considered as trustworthy if more than half of the nearby sensors predict the values that differs from the actual value within a given limit.

To evaluate the proposed technique, 8 sensors' temperature readings for 24 hours were used as experimental dataset. First 90% data was used for nearby sensor selection of particular sensor and linear model construction. The rest 10% was used for testing. The result analysis shows that the proposed technique detects 19, 69 and 73 trustworthy data from 73 sensor readings with respect to 3%, 4% and 5% deviation. In future, we have a plan to perform experiment on other types of sensor to observe the behavior of the technique. Besides, an experiment will be conducted on large dataset in future.

References

[1] Radhakisan Baheti and Helen Gill, “Cyber-physical systems.” The impact of control technology, 12:161–166, 2011.

[2] Edward A Lee, “Cyber physical systems: Design challenges.” In 11th IEEE International Symposium on Object oriented real-time distributed computing (ISORC), 2008, pages 363–369. IEEE, 2008.

[3] Xu Jin, Wassim M Haddad, and Tansel Yucelen, “An adaptive control architecture for mitigating sensor and actuator attacks in cyber-physical systems.” IEEE Transactions on Automatic Control, 2017.

[4] Xu Jin, Wassim M Haddad, and Tansel Yucelen, “Adaptive control architecture for mitigating sensor and actuator attacks in cyber-physical systems.” IEEE Transactions on Automatic Control, 2017.

[5] Wenjia Li, Pramod Jagtap, Laura Zavala, Anupam Joshi, and Tim Fin, “Care-cps: Context-aware trust evaluation for wireless networks in cyber-physical system using policies.” In IEEE International Symposium on Policies for Distributed Systems and Networks (POLICY), pages 171–172. IEEE, 2011.

[6] Yin Zhang, Meikang Qiu, Chun-Wei Tsai, Mohammad Mehedi Hassan, and Atif Alamri, “Health-cps: Healthcare cyber-physical system assisted by cloud and big data.” IEEE Systems Journal, 11(1):88–95, 2017.

[7] Lu-An Tang, Xiao Yu, Sangkyum Kim, Jiawei Han, Chih-Chieh Hung, and Wen-Chih Peng, “Tru-alarm: Trustworthiness analysis of sensor networks in cyber-physical systems.” In 10th International Conference on Data Mining (ICDM), pages 1079–1084. IEEE, 2010.

[8] Kevin Ni and Greg Pottie, “Bayesian selection of non-faulty sensors.” In International Symposium on Information Theory, pages 616–620. IEEE, 2007.

[9] Hyo-Sang Lim, Gabriel Ghinita, Elisa Bertino, and Murat Kantarcioglu, “A game-theoretic approach for high-assurance of data trustworthiness in sensor networks.” In 28th International Conference on Data Engineering (ICDE), pages 1192–1203. IEEE, 2012

[10] Lu-An Tang, Quanquan Gu, Xiao Yu, Jiawei Han, Thomas La Porta,Alice Leung, Tarek Abdelzaher, and Lance Kaplan, “Intrumine: Mining intruders in untrustworthy data of cyber-physical systems.” In 2012 SIAM International Conference on Data Mining, pages 600–611. SIAM, 2012

[11] Adrian Perrig, Robert Szewczyk, Justin Douglas Tygar, Victor Wen, and David E Culler, “Spins: Security protocols for sensor networks.” Wireless networks, 8(5):521–534, 2002.

[12] Mark Luk, Ghita Mezzour, Adrian Perrig, and Virgil Gligor, “Minisec: a secure sensor network communication architecture.” In 6th International Conference on Information Processing in Sensor Networks, pages 479–488. ACM, 2007.

[13] Liying Zhang, Lun Xie, Weize Li, and Zhiliang Wang, “Security solutions for networked control systems based on des algorithm and improved grey prediction model.” .International Journal of Computer Network and Information Security, 6(1):78, 2013.

[14] G. Shanmugasundaram, and G. Sankarikaarguzhali, “An Investigation on IoT Healthcare Analytics.” International Journal of Information Engineering and Electronic Business 9(2):11, 2017.

Authors’ Profiles

Abdus Satter is a graduate student at the Institute of Information Technology (IIT), University of Dhaka, Bangladesh. Currently, he is pursuing his Master of Science in Software Engineering (MSSE). He earned his Bachelor of Science in Software Engineering (BSSE) from the same institution with the top score in his class. His core areas of interest are data mining, machine learning, software engineering, web technologies, systems and security. He has numerous awards in various national and international software and programming competitions, hackathons & project showcasings.

Nabil Ibtehaz is a graduate student at the Department of Computer Science and Engineering (CSE), Bangladesh University of Engineering and Technology (BUET), Bangladesh. Currently, he is pursuing his Master of Science in Computer Science and Engineering (Msc CSE). He earned his Bachelor of Science in Electrical and Electronics Engineering (Bsc EEE) from the same institution. His core areas of interest are machine learning, natural language processing, computer vision, signal processing, evolutionary algorithms. He has participated in various national and international competitions. He has also achieved awards in various competitions.

How to cite this paper: Abdus Satter, Nabil Ibtehaz," A Regression based Sensor Data Prediction Technique to Analyze Data Trustworthiness in Cyber-Physical System", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.10, No.3, pp. 15-22, 2018. DOI: 10.5815/ijieeb.2018.03.03