Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 6, June 2013, pg.28 – 31

RESEARCH ARTICLE

RECOGNITION AND PREPROCESSING OF INTRUSION DATA IN WIRELESS SENSOR NETWORK

Manjot Singh[1], Himanshu Sharma[2]
[1]School of Engineering & Technology, IFTM University, Moradabad, U.P, India
[2]School of Engineering & Technology, IFTM University, Moradabad, U.P, India

[1] *manjot.singh96@gmail.com*

Abstract— In wireless sensor networks (WSNs), the significant deviation of measurements from the normal pattern of sensed data are considered as intrusion. The main sources of intrusion are noise and errors, events, and malicious attacks on the network. In this work, we focused on the problem of detecting the intrusion by proposing a model that is based on distribution of sensor data stream approximately over the data space. With the set of data collected from Intel Berkley lab, we processed this data by our proposed scheme and evaluated it. We find that experimental evaluation of our proposed scheme can achieve high precision rating for detecting intrusion.

Key Terms: - Intrusion; Statistic; Kernel; Sensitive; Probability

I. INTRODUCTION

Today, as technology involvement is increasing day by day in human based activity, this has lead to urgency of Sensor usage in many real world situations which ultimately are useful for detecting interesting events and effective monitoring of the physical phenomenon. Due to its small size, low cost and power efficiency wireless sensor network which is composed of huge numbers of tiny sensor is replacing the work done by humans into sensor based work.

Replacement of Sensor no doubt has improved processing time, but have also lead to various concerns, as they produce uncertain, discrete and unfaithful data streams. This uncertainty, occur due to the presence of intrusion in the data streams collected by sensors. For our convenience in our work we replace intrusion with its alternative term outlier. Where, Outlier is simply an unexpected or unwanted set of values which has very low probability of occurrence in our system or deviate from other values in the system.

The various characteristics of a data streams collected by sensors are discussed below:

A. Correlated Data

Data samples have temporal correlation among and within data streams i.e. between inter-stream and intra-stream correlated, since these data streams are only temporal observations of the physical world. Suppose sensors are deployed in traffic monitoring system and they are detecting the vehicle position say at time T0 and T1 it's obvious that these two readings are correlated through the vehicle velocity and time difference (T1-T0).

B. Stream Unfaithfulness

Data streams generated by sensor readings are discrete in nature for a continuous physical phenomenon. These samples of data describe a state of physical world at a particular time instant. Data collection, Communication of data and Computation are major reasons for unfaithfulness in the given data streams.

C. **Energy consumption sensitiveness:**

A sensor has inbuilt low power battery within it and since the size of sensors is a big question of concern here, hence it usually get discharged by unnecessary transmission. So, how and when frequently data samples are transmitted, this question now becomes a big issue while determining the lifetime of sensors in wireless sensor network.

II. MOTIVATION

Deriving useful information from data gathered by sensor is a challenging task because of Limited CPU capability, Limited Network bandwidth and Short life time of sensors. In this process, instead of sending all data to central base station we will first preprocess data at sensor level to discard those values which have very less relevancy in our work. It enables us to save the battery life and network bandwidth and CPU capability. So this motivates us to propose such a model which would transform the raw sensor readings into meaningful information of observing physical world.

III. PROBLEM STATEMENT

A sensor network generates huge amount of data rapidly in the form of various streams and these data streams are discrete, uncertain and unfaithful ,due to this processing of these data now becomes much more difficult and tedious. So, the uncertainty generated by outliers in the data collected by sensor can be eased down if we estimate distribution of data on a data space at a particular time .In this, estimation of data distribution for sensor readings in our hand we can know the density of the data space around each value, which further eases our problem of detecting outliers. Now a particular point or value which lies in low density region will be flagged as an outlier here. So, the main problem of concern is to choose appropriate algorithms or methodologies which can solve the above mentioned problems and use them for detecting outliers.

IV. PROPOSED APPROACH

The proposed model is based on Statistical modeling technique to transform raw sensor readings to meaningful information with very little amount of information loss and also minimize the cost of processing the responses of respected long queries. The proposed approach work as follow:

A. **Statistical technique**

In this technique the work of a sensor network is to captures samples of data if each sensor is measuring a single real valued attribute Xi, at each time instant. Then, taking into consideration this situation we have to model the set of attributes X 1, ... , X n, as an n-dimensional random variable X. After the statistical modeling technique is implied we generate uniform random samples. With random samples generated, we distribute them in data space with the help of probability density function. As data is distributed over the data space we apply kernel density estimation over it.

B. **Estimating density by kernel**

In this there are various kernel function in kernel density estimation which distribute the weight age to the area near to the value or point to which we are processing at a particular time .To choose from various kernel function we select the only kernel function for density estimation which produce better accuracy and increase efficiency in kernel based density estimation .The method which we use for assigning weight age to the area near to the point are: Arithmetic based weighting method and Exponential based weighting method .

After, this to calculate the kernel based density estimation for whole data set we combine all the kernel function and retrieve the estimation for whole data set.

C. **Probability Density Function**

After the kernel density is estimated by kernel density technique, we apply probability density function over the distributed data space to estimate density for a particular point which will be useful in determining various factors like deviation factor and normalized deviation factor which we will implement in our algorithm.

To detect outlier we can implement above data in algorithm and can detect outlier in a given data streams of values produces by sensor in wireless sensor network. In our algorithm for a given point y we will find r-neighborhood of point y in the data space. The r-neighborhood also known as sampling neighborhood are the local neighborhood for the point y which fall in the given radius r. After knowing the sampling neighborhood for the point in the interval 2αr, we determine the counting neighborhood for the point y.

When the sampling and counting neighborhood for the point y is known then we calculate the deviation and normalized deviation factor for the specific point y. Point y is detected as outlier if the deviation factor of y is greater than normalized deviation factor for the point y.

V. RESULT AND ANALYSIS

For our proposed method, we need a dataset. And, we got the real data set from Intel Berkeley Research lab and downloaded it. This dataset is freely available on their website and have information about data collected from 54 sensors deployed in the lab between February 28th and April 5th, 2004. There is a log of about 2.3 million readings collected from these sensors in this file.

Table 1: Statistical characteristics for sensor1

Dataset	Min	Max	Mean	Median	Std Dev
Temperature	18.1954	125.1530	25.8824	27.1444	36.6511

We feed up the above data collected from the Intel lab into our algorithm. For implementing our algorithm we make use of the software known by the name massive online analysis. After the implementation of data and algorithm into the massive online analysis software we get output of our data in the form of graphical representation as follow.

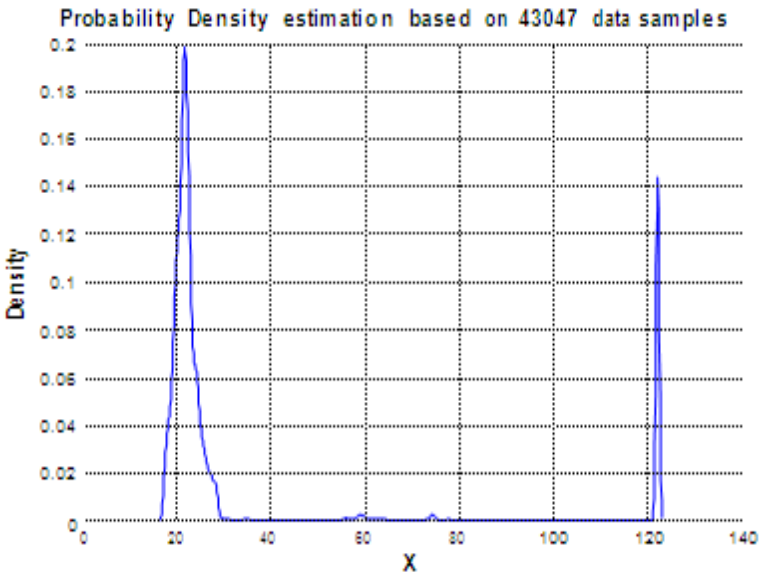

Fig1. Performance evaluation graph for above data

From above we can easily analyse the presence of outlier in the given data. Suppose we have normalized deviation factor for above data shown in graph is 0.10 as shown in above fig.1. And at the same time deviation factor for particular point is computed as 0.16. Then above point will be detected as outlier because it has more deviation factor than its normalized deviation factor as our algorithm. Now from above graph we can perceive the variance in the above data and conclude that for a particular point, if there is more deviation then the normal deviation then, that particular point can be considered as outlier.

VI. CONCLUSION AND FUTURE WORK

In above work, we focused on the main problems of outlier detection in wireless sensor networks (WSNs). With the help of a set of experiments with datasets taken from Intel Berkeley research lab we have processed and evaluated our proposed scheme. Similarly, we have offered a model that is based on the calculation of the sensor data distribution. This method focuses various characteristics and features of streaming sensor data. The experiments proves that our algorithm can achieve very high precision and recall rates for finding outliers, and demonstrate the success of the proposed approach. At present we are working for single attribute sensors but we will try to extend our scheme for multi-attribute sensors in future. For outlier detection for multi-attribute sensors we will focus on some other idea if required. We will focus on other density estimation techniques like orthogonal series expansion (wavelet density estimation) as our future work. The main idea of this technique is to calculate the distribution of measurement by determining the coefficients of its Fourier transform and then implementing it on multi-attribute sensors.

REFERENCES

[1] E.Elnahrawy, B.Nath : Cleaning and Querying Noisy Sensors. In: Proc. of WSNA. (2003)
[2] Silverman, B Chapman and Hall (1986): Density Estimation for Statistics and Data Analysis.
[3] S. Subramaniam T. Palpanas D. Papadopoulos, V.Kalogeraki, D. Gunopulos: Online Outlier Detection in Sensor Data Using NonParametric Models.
[4] Victoria J Hodge & Jim Austin(2004), A survey of outlier detection methodologies Artificial Intelligence
[5] Tan, P. N. (2006) “Knowledge Discovery from Sensor Data”.
[6] Morgan Kaufmann ,J.Han, M.Kamber and San Francisco., (2006) 'Data Mining: Concepts and Techniques',
[7] Gaber, M. M. (2007) 'Data Stream Processing in Sensor Networks', Springer.
[8] Sheng, B., Li, Q., Mao, W. and Jin, (2007) 'Outlier detection in sensor networks', Proceedings of MobiHoc.
[9] Keinosuke Fukunaga, “Introduction to Statistical Pattern Recognition”, Academic press
[10] A. Faradjian, J. Gehrke and P. Bonnet, “GADT: A Probability Space ADT for Representing and Querying the Physical World”
[11] Yang Zhang,Nirvana Meratnia, Paul Havinga(2008), A survey :Outlier detection Techniques for wireless sensor networks

