Anomalies Detection in Wireless Sensor Networks Using Bayesian Changepoints

Rychelly Glenneson da S. Ramos and Paulo Ribeiro L. Junior
Inst. Fed. de Educação, Ciências e Tecnologia da Paraíba
Campus Campina Grande
Campina Grande, Paraíba, Brazil
Email: rychelly.g.ramos@ieee.org, paulo.ribeiro.lins.jr@gmail.com

José Vinícius de M. Cardoso
Universidade Federal de Campina Grande
Campina Grande, Paraíba, Brazil
Email: jvmirca@gmail.com

***Abstract*—Wireless Sensor Networks (WSN) have gained wide acceptance as a framework for telemetry and remote monitoring for applications such as telemedicine, precision agriculture, and climate monitoring. The complex and dynamic characteristics of such networks have made them vulnerable to anomalies, i.e., observations that do not correspond to the natural behavior of measurements. This paper evaluates the use of Bayesian Changepoints in the context of anomalies detection in WSNs, in order to determine under which conditions this technique leads to the minimization of false positives.**

***Index Terms*—Bayesian Changepoints; Wireless Sensor Networks; Anomalies Detection**

I. Introduction

Wireless Sensor Networks (WSN), core infrastructure of Internet of Things, may experience anomalies due to malfunction of either hardware or software, which cause failures in sensor nodes [1], [2], [3]. Such anomalies may lead a node to either shutdown, when it is related to power management, or to compromise the reliability of the transmitted data, resulting in errors on the information, which can even affect the network operation from a security point of view [4]. Therefore, network management tools must be able to detect not only shutdown conditions of sensor nodes, but also their malfunctioning.

Anomaly detection is a technique used to detect abnormal events of a system by comparing its current run-time profile to its reference profile. Nonetheless, anomaly detection can also be used to detect anomalies other than failures and malicious attacks. For instance, one can set a target behavior changing in a target-tracking system which can be detected using anomaly detection [5].

Hence, this paper evaluates the Bayesian Model Changepoints when applied to the detection of anomalies on the traffic of sensor nodes to the head node, i.e., the network manager.

The remainder of the paper is organized as follows. Section II describes the Bayesian Model Changepoints. Results and discussions are presented in Section III. Section IV presents the conclusions.

II. Bayesian Changepoints Model

The Bayesian Changepoints model used in this paper is an implementation of the Bayesian Online Changepoint Detection algorithm developed in [6].

This algorithm computes a probability distribution over a subset of transmitted data, where this subset refers to the number of observations since the last changepoint. When the probability of a 0-length run spikes, there is most likely a change point at the current data point [7]. Here, the length of this subset is refered to as “lag".

The process for detecting anomalies using Bayesian Changepoints Model consists in the observation of new data points x_t and in the evaluation of the likelihood of seeing this value for each possible run length. This is a probability vector with an element for all possible run lenghts. This algorithm assumes a Gaussian distribution between each pair of changepoints [7]

$$L(r) = P(x|x_r), \tag{1}$$

for each possible run length $r > 0$.

One of the most important choice for the operation of the algorithm is the lag size parameter which must be chosen *a priori*.

The anomaly detection probability can be described as [7]

$$P_t(L = r) = P_{t-1}(L = r - 1) \cdot L(r) \cdot (1 - 1/L_E), \tag{2}$$

in wich L is the lag length selected and L_E is a parameter describing the *a priori* best guess of run length. The larger L_E is the stronger evidence must be in the data to support a high changepoint probability.

The next step is calculate the probability of change, or $r = 0$, and, then, normalize it. For each incoming point, repeat this process. This per-point update is why the method is considered an online learning algorithm.

The output of this model is a probability distribution of run lengths for each point in the training data. A good example of what this might look like is presented in [6].

III. Results and Discussions

A labeled wireless sensor network data set was analyzed. Such data were collected from a multi-hop wireless sensor network deployed using TelosB motes, which is available in [8]. The data consist of humidity measurements collected during a period of 6 hours, with sampling frequency 0.2 Hertz. Also, the multi-hop data were collected on 10th May 2010. The label ‘0’ denotes normal data and label ‘1’ denotes

2155-6814/16 $31.00 © 2016 IEEE
DOI 10.1109/MASS.2016.28

an introduced event. In this case, steam from hot water is introduced to increase the humidity.

The Python programming language and Graphlab Create™were used to analyze this data. Graphlab Create™is an extensible machine learning framework that enables developers and data scientists to easily build intelligent applications and services at scale [9].

The analysis consisted in transforming the dataset in a time series and use it to create a model of anomalies detection based in Bayesian Changepoints. After creating such model, `changepoint_scores` is set, which consist in values varying between 0 and 1, which is the probability of change in value at some point on this data set. In Figure 1 the blue curve represents the data set, where it is possible to observe peaks on data points, the red curve, on the other hand, represents the values of `changepoint_scores` of the data point in the same time instant, indicating the probability that that data point is a possible anomaly.

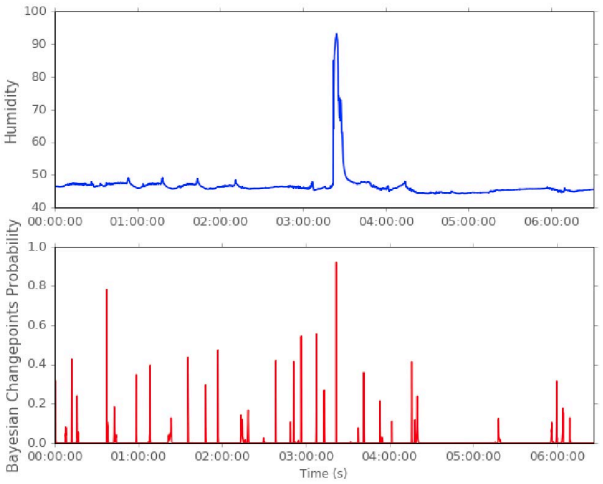

Figura 1. Time series representation of a humidity dataset (blue curve) and the Bayesian Changepoints Probability of that dataset (red curve).

In this data collection, 100 entries are incorrect, so that one can evaluate the optimal lag size determined a priori in order to reduce false alarms [8].

Figure 2 shows the number of anomalies detected as a function of the lag length.

This result shows that according to the measurements, the lag size is directly related with the occurrence of false alarm detections. Interestingly, both small and large sizes of the lag generate high rates of false alarms. According to the experiment, the correct identification number of anomalies (100) occurs exactly when the lag has a size of 51, 56 and 57 samples, however, lag sizes of approximately 50% of the number of anomalies group data show lower false alarm rates, as shown in the curve. This percentage might be an indirect measure of the amount of anomalies in a data set.

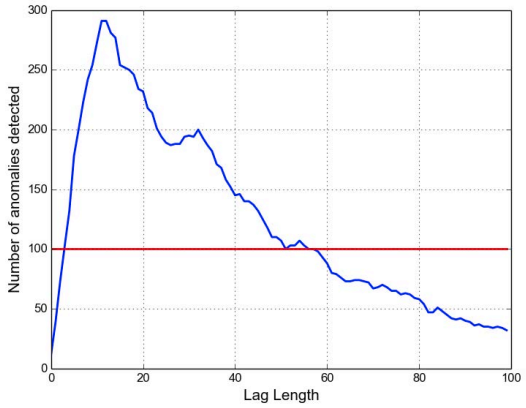

Figura 2. Number of anomalies detected versus the lag length. The red line indicates the amount of the incorrect values in the data set.

IV. Conclusion

In this study, we investigated the use of Bayesian Changepoints Method for detecting anomalies in wireless sensor network traffic. The technical efficiency of applying the algorithm to anomaly detection problem in wireless networks was tested on a set of data, being able to verify the influence of lag size on the efficiency of the approach, considering the incidence of false alarms as metric.

Acknowledgment

The authors would like to thank the IEEE Student Branch IFPB-CG and IFPB – *Campus* Campina Grande for the institutional support. This work was funded by CNPq.

Referências

[1] M. A. Rassam, A. Zainal, and M. A. Maarof, "Advancements of data anomaly detection research in wireless sensor networks: a survey and open issues," *Sensors*, vol. 13, no. 8, pp. 10 087–10 122, 2013.

[2] S. Rajasegarar, C. Leckie, M. Palaniswami, and J. C. Bezdek, "Quarter sphere based distributed anomaly detection in wireless sensor networks." in *ICC*, vol. 7, 2007, pp. 3864–3869.

[3] SBC, "Grandes desafios da pesquisa em computaÃğÃčo no brasil âĂŞ 2006-2016," Sociedade Brasileira de ComputaÃğÃčo, Tech. Rep., 2006.

[4] H. Sagha, J. d. R. Mill, R. Chavarriaga *et al.*, "Detecting and rectifying anomalies in body sensor networks," in *2011 International Conference on Body Sensor Networks*. IEEE, 2011, pp. 162–167.

[5] Q. Wang, "Traffic analysis & modeling in wireless sensor networks and their applications on network optimization and anomaly detection," *Network Protocols and Algorithms*, vol. 2, no. 1, pp. 74–92, 2010.

[6] R. P. Adams and D. J. MacKay, "Bayesian online changepoint detection," *arXiv preprint arXiv:0710.3742*, 2007.

[7] "Dato Machine Learning Platform User Guide - Bayesian Changepoints," https://dato.com/learn/userguide/anomaly_detection/, accessed: 2016-06-22.

[8] S. Suthaharan, M. Alzahrani, S. Rajasegarar, C. Leckie, and M. Palaniswami, "Labelled data collection for anomaly detection in wireless sensor networks," in *Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*. IEEE, 2010, pp. 269–274.

[9] "GraphLab Create," https://dato.com/products/create/, accessed: 2016-06-15.