Sādhanā (2020)45:212
https://doi.org/10.1007/s12046-020-01451-w

Graph based event measurement for analyzing distributed anomalies in sensor networks

P SHERUBHA[1,*], S P SASIREKHA[1], V MANIKANDAN[2], K GOWSIC[3] and N MOHANASUNDARAM[1]

[1]Department of Computer Science and Engineering, Karpagam Academy of Higher Education, Coimbatore, India
[2]Computer Networks, College of Engineering and Computer Science, Lebanese French University, Erbil, Iraq
[3]Department of Information Technology, KSR Institute for Engineering and Technology, Tiruchengode, India
e-mail: sherubha0106@gmail.com

MS received 20 April 2020; revised 13 May 2020; accepted 5 July 2020

Abstract. Wireless Sensor Network (WSN) has emerged drastically with numerous practical applications of considerable Engineering importance where privacy and security are of dominant influence. This paves the way for this investigation and present interest in the development of novel and innovative intrusion detection approach. This work anticipated a novel Intrusion detection framework by modeling sensor connectivity with a targeted graph and uses statistical graph properties by modeling intrusion detection. In anticipated graph-based detection, data capturing magnitude is modeled with the Gaussian model, and the corresponding correntropy is estimated by graph matrix with adaptive sensor measurements. Anticipated detection approach is modeled based on the Laplacian Matrix, and closed-form expressions are attained for statistical analysis. At last, temporal network analysis are characterized by evaluating sensor distance among measurement distributions in consecutive time. The results depict that the anticipated detection framework offers superior detection recital than compared to existing frameworks.

Keywords. Wireless Sensor Networks; intrusion detection; distributed sensor network; graph connectivity; corr-entropy computation; statistical distance measures.

1. Introduction

Wireless Sensor Networks (WSNs) are considered as the integration of communication, control, and computation technologies that are cast-off to manage and monitor infrastructures [1]. Improvements in sensor technologies and communication have offered deployment of massive nodes in WSN, those results in exception growth in the opportunistic implementation of an application to those systems [2]. The tremendous growth in WSNs leads to claims with an acute condition that has raised interest in security crisis [3]. In specific, concentration is towards network-based intrusion detection is distinctive and shows irregular changes in capturing data in nodes [4]. Authentic functionalities like transient variations in vapor detection from the smoke detector and criminal activities like worms and virus injection in power grids are two instances of such intrusions. In recent times, sensor networks are integrated with cyber-physical systems, which play a significant task in the advancement and management of economic and social infrastructures [5].

The current advancement in graph theory-based signal processing offers an opportunity to return conventional solutions for processing and widen applicability to the rising crisis with massive datasets. Graph-based processing has provided a novel framework for specifying model relation between data samples [6]. In data-based applications, similarities among data samples are evaluated by sensors that are defined by the weighted graph. The network is composed of a distributed sensor with spatial dependencies of unequal data measurements as nodes graph signal [7]. Sensor placements are fixed with changing measures. Graph theory-based matrix is generated, where graph edges reflect the similarity between closeness and signals of sensors [8]. In this scenario, the magnitude of data measured is considered to be randomly distributed variables with Gaussian distribution with correntropy and finite mean. The ratio test anticipates a novel statistical-based detection approaches. The entropy matrix of this model is evaluated with a graph matrix from empirical data used for computations of intrusion detections.

*For correspondence
Published online: 28 August 2020

2. System model

Here, analytical modeling of WSN is concentrated to validate that the anticipated model offers satisfactory resource allocation with reduced collisions in resource access. Contention and collision happen while multiple devices need requests for accessing nodes [9]. In all resource allocation slots, the number of nodes that participate in network connectivity is $'N'$ to establish random communication with neighborhood nodes with α resources. Probability of those requests between neighborhood nodes to develop a connection within the macro cell range is provided as M_c:

$$M_c = 1 - \left(1 - \frac{1}{\alpha_M}\right)^{N_M - 1} \tag{1}$$

Where, α_M and N_M specify resources and devices allocated to Macro-cell coverage of sensor nodes (SN), respectively. As well, SNs collision probability of devices in BS coverage is provided as:

$$M_s = 1 - \left(1 - \frac{1}{\alpha_s}\right)^{N_s - 1} \tag{2}$$

Where, α_s and N_s specifies resources and devices allocated to nodes under BS coverage, respectively.

Average collision probability is computed based on BS and Macro-cell collision probability and dividing it with $N+1$. Consider, the proposed model with BS located under Macrocell coverage, average collision probability if devices (M_p), that are competing with Number of small cell BS and Macrocell is given below:

$$M_p = \frac{N - \sum_{n=1}^{N} \left\{1 - \left(1 - \left(\frac{1}{\alpha_{s.n}}\right)\right)^{M_{s.n}-1}\right\} + 1 - \left(1 - \left(\frac{1}{\alpha_M}\right)\right)^{M_M - 1}}{N+1} \tag{3}$$

Therefore, the collision probability of SNs in this model is provided as:

$$M_P = \frac{N + 1 - \sum_{n=1}^{N} \left\{1 - \left(1 - \left(\frac{1}{\alpha_{s.n}}\right)\right)^{M_{s.n}-1}\right\} + 1 - \left(1 - \left(\frac{1}{\alpha_M}\right)\right)^{M_M - 1}}{N+1} \tag{4}$$

Consider, every small cell BS of node connectivity is allocated with equal resources and users for data transmission, that is, the number of devices that requests for support).

$$\left[\left(1 - \frac{1}{\alpha_M}\right)^{M_M - 1} - \left(1 - \frac{1}{\alpha_S}\right)^{M_s - 1}\right] < 0 \tag{5}$$

$$\frac{(1 - (1/\alpha_M))^{M_M - 1}}{\left(1 - \left(\frac{1}{\alpha_s}\right)\right)^{M_s - 1}} < 1 \tag{6}$$

Lemma 1 *Total available BS and Macro-cell resources are equal while the total Number of BS and users is also equal, that is, $\alpha_s = \alpha_m$ and $M_s = M_M$, then above Eq. is provided as:*

$$\frac{(1 - (1/\alpha_M))^{M_M - 1}}{(1 - (1/\alpha_M))^{M_M - 1}} < 1 \tag{7}$$

The above computation is pretended to be false while determining the functionality with the Equal amount of BS and Macro-cell, which is allocated with an equal number of users, and similar resources will improve outcomes in various collision constraints as in traditional systems.

Lemma 2 *Number of resources available for BS is lesser than Macro-cell resources and number of users in Macrocell, and small is Equal to, that is, $\alpha_s < \alpha_M$ and $M_s = M_M$, therefore, Eq. is provided as:*

$$\left\{\frac{\left(1 - \left(\frac{1}{\alpha_M}\right)\right)}{\left(1 - \left(\frac{1}{\alpha_s}\right)\right)}\right\} < 1 \tag{8}$$

Here, all $\alpha_s < \alpha_M = 1 - \left(\frac{1}{\alpha_s}\right) < 1 - \left(\frac{1}{\alpha_M}\right)$. Therefore the above computation is false, which shows that the proposed system model collision > traditional system model collision.

3. Graph-based node connectivity

To demonstrate similarity among network connectivity $'N'$ is sensors are considered as weighted graph $G = (w, v, e)$ comprising finite vertices and edges with appropriate weights. Weight is a similarity measure among vertices $'x'$ and $'y'$. Sensor $'x'$ and $'y'$ are similar if sensor lies within Macrocell coverage or under small cell BS, similarity weighted graph computation is provided based on the Gaussian kernel model as below:

$$D_{xy} = \sqrt{(B_x - B_y)^2 + (C_x - C_y)^2} \tag{9}$$

The above Equation is measured as geometric distance, $D_g^2 = (S_x - S_y)^2$ is a signal for transmitting data, σ_{xy} and σ_g is the level of similarity, S_x and S_y is graph signal on nodes $'x'$ and $'y'$. Graph similarity matrix $'W'$ for symmetric graph Matrix and the real-valued matrix is depicted as:

$$L = D - W \tag{10}$$

Graph properties are examined using eigenvectors and Eigenvalues related to graph matrix (common measurements set at every time instant). Eigen decomposition of Laplacian is provided as:

$$L = \sum_i \lambda_i u_i u_i^T \tag{11}$$

Where $'T'$ specifies transpose, $\lambda = \{\lambda_t\}_{(t=1,..,N\}}$ are eigenvalues set and $u = \{u_t\}_{t(1,..,N)}$ is a set of eigenvectors based on the Laplacian Matrix. Set $'u'$ provides basic functionality for signals defined by a graph, and λ is depicted as graph frequencies. Graph signal $g(s)$ is graph components specified by $\bar{g}(s)$ at every time with Eigenvectors $u_t(s)$, for $(1 \le i \le N)$ of Laplacian matrix at time $L(s)$ is provided:

$$\bar{g}(s) = U(s)^T g(s) \tag{12}$$

where $U(s)^T$ is a matrix made of Eigenvectors of $L(s)$. Graph model $g(s)$ is determined as graph smoothing concerning graph $'G'$, if the energy consumed is concentrated towards low frequencies. $\bar{g}(s)$ Coefficients are zero for larger values. Subsequently, smooth signal $g(s)$ provides rise to lesser value of graph regularization as provided below:

$$g^T(s)L(s)g(s) = \sum_{t=1}^{N} \lambda_t \left(u_t(s)^T, g(s)\right)^2 \tag{13}$$

Smoothness regularization is used for modeling intrusion detection.

4. Graph model for intrusion detection

In this detection strategy, the node receiver does not possess any graph constructed information and however, graph Laplacian Matrix for modeling graph, which is not known. More specifically, $g(s)$ is measured as instances of the Gaussian model to have a probability density function as below:

$$f(g(s)) = (2\pi)^{\frac{N}{2}} |Q(s)|^{\frac{1}{2}} * \exp\left(\frac{-1}{2}(g(s) - m(s))^T P_m(s)(g(s) - m(S))\right) \tag{14}$$

Where $m(s)$ is the mean vector and $P_m(s)$ is a symmetric matrix. Precision matrix $m(s)$ is considered as graph similarity matrix as in Equation below:

$$M_{xy} = \begin{cases} \sum\limits_y w_{xy} & x = y \\ -w_{xy} & x \neq y \end{cases} \tag{15}$$

There exists some constant relationship among Laplacian matrix $L(s)$ and precision matrix $m(s)$. In graph modeling, the node is connected to neighborhood nodes; therefore, the resulting graph matrix is spare.

5. Continuous event measures

The volume of intrusion detection environment and frequency in the monitoring environment is evaluated incessantly [10]. Here, the Gaussian model exemplifies continuous data event, where probability density function is provided by:

$$P(v|\sigma, \mu, k) = \sum_{i=1}^{k} w_i . N(v|\mu_i, \sigma_i) \tag{16}$$

Where $'v'$ specifies standard event value, $'k'$ are Gaussian components, $'m'$ and $'s'$ are mean and standard deviations, w_i is the component weight. Gaussian model is computed with Expectation Maximization. Probability value of v_j will be smaller or larger than maximum or minimal data instances. The distance measure is utilized, where the distance between v_j and k^* Gaussian element is evaluated by as:

$$h_k^*(v_j) = \frac{|v_j - \mu_k *|}{\sigma_k^*} \tag{17}$$

The last step is anomaly score computation of events to translate probability by:

$$\propto (\emptyset_j) = \min\left(\frac{-\ln(1-p)}{\tau}, 1\right) \tag{18}$$

Where τ is the highest anomaly score for normalization; however, the anticipated approach is extendedly inherent to assist the event with multi-variate values.

6. Node correlation

While monitoring data transmission, compute node correlation amongst two nodes which is specified as $N_i$and N_j by considering nodes event. Assume time $'T'$, where correlation is depicted as an average weighted correlation of nodes by the length of the overlapping period.

$$\rho_o(N_i, N_j, T) = \frac{1}{T} \sum_{\emptyset_a, \emptyset_b} \rho_s(\emptyset_a, \emptyset_b) ||t_a \cap t_b|| \tag{19}$$

where $N_i, N_j, T$are set of events for node movement of $\emptyset_a, \emptyset_b$at 'T'; $\emptyset_a, \emptyset_b$ are an event set where t_a and t_b are periods correspondingly. Nodes correlation between sensors is not captured. In intrusion detection, the node correlation of two sensor hosts is computed with average network traffic amongst them.

7. Correntropy computation

Correntropy computation is cast-off for evaluating similarities amongst feature vectors where it is a measure of second-order statistics and non-linear similarity for determining interactions for the provided feature set.

$$V(f_1, f_2) = E[k(f_1 - f_2)] \tag{20}$$

Where $E[.]$ specifies mathematical expectation, $k()$ is a Gaussian function, and $'k'$ determines kernel size, which is depicted as:

$$k(.) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(.)^2}{2\sigma^2}\right) \tag{21}$$

For using corr-entropy measures to distributed network data, it can be depicted for normal/abnormal feature vectors.

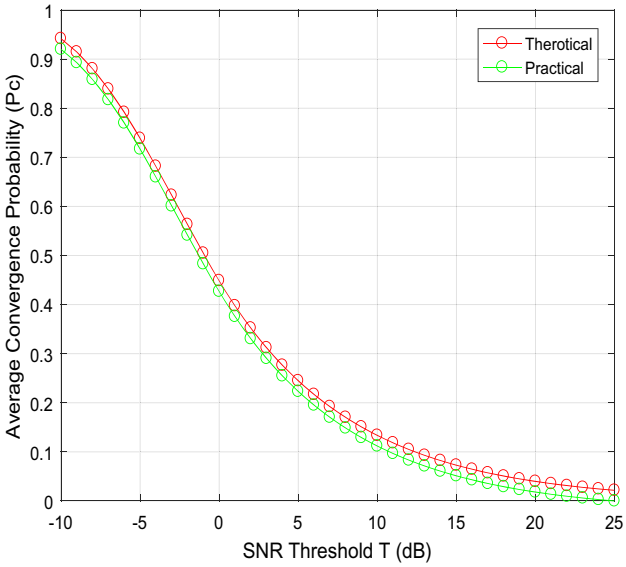

Figure 2. Coverage probability.

8. Analysis

Experiments were carried out in MATLAB to analyze the performance of an anticipated graph-based intrusion detection approach to evaluate its ability with other prevailing models. Figure 1 depicts some sample sensor measurements value over time when data collects univariate data. Figure 2 depicts the coverage probability to measure the SNR value. For time instant, the matrix is designed for a complete network, and a graph-based Laplacian matrix is attained. It is then used for detection strategy with a precision matrix. Matrix at every independently instant, no assumption is required for the anticipated model. It is identified from figure 3 that this model shows superior FAR performance compared to others, by offering the highest detection probability for provided false alarm probability.

To evaluate the computational complexity of the anticipated approach to other techniques, we compute essential CPU time with average runs, when experiments are reviewed in MATLAB on Intel Core with 4 GB RAM—average CPU times needed for diverse intrusion detection approaches.

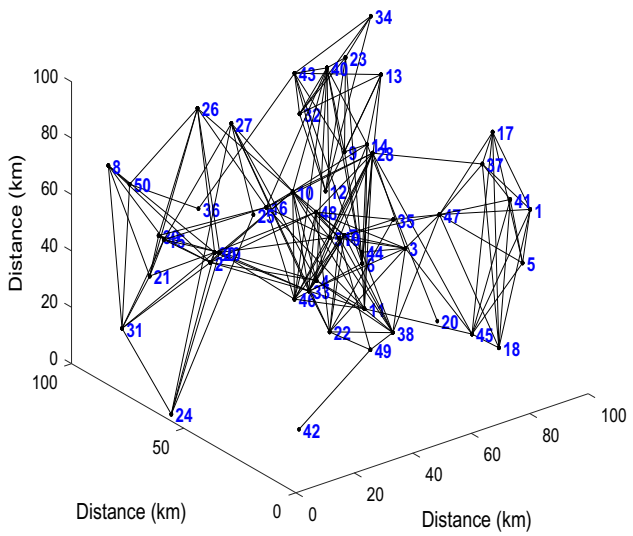

Figure 1. Graph connectivity.

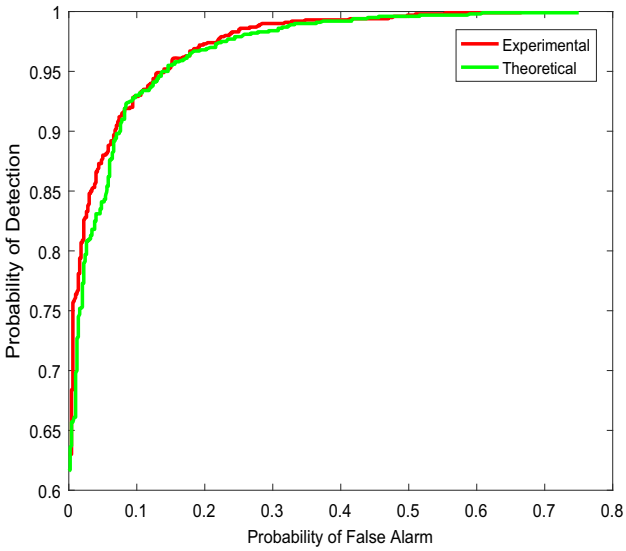

Figure 3. FAR computation.

9. Conclusion

Here, a novel statistical-based intrusion detection strategy for node distribution is anticipated. This model has been analyzed with graph modeling from both sensor placements and measurements; those results in similarity matrices and Laplacian matrices. This anticipated intrusion model is modeled with a Gaussian model based on hypothetical testing and employment of ratio criteria. The network expression for both test data and hypothesis data has been validated and derived experimentally. The anticipated model performance has been analyzed by performing numerous experiments. It has been confirmed that the anticipated intrusion detection model offers performance considerably fine than that of other approaches that is validated by the finest detection rate.

References

[1] Butun I, Morgera S and Sankar R 2013 A Survey of Intrusion Detection Systems in Wireless Sensor Networks. *IEEE Commun. Surv. Tutorials* 16: 266-282

[2] Riecker M, Biedermann S, Bansarkhani R and Hollick M 2015 Lightweight energy consumption based intrusion detection system for wireless sensor networks. *Int. J. Inf. Secur.* 14:155-167

[3] Leite R A, Gschwandtner T, Miksch C, KriglsteinS,Pohl M, Gstrein E and Kuntner J 2018 Eva: Visual analytics to identify fraudulent events. *IEEE Trans. Visual Comput. Graphics.* 1: 330–339

[4] Beck F, Burch M, Diehl S and Weiskopf D 2017 A taxonomy and survey of dynamic graph visualization. *Comput. Graphics Forum.*36: 133–159

[5] Zhao J, Cao N, Wen Z, Song Y, Lin Y R and Collins C 2014 Flux Flow:Visual analysis of anomalous information spreading on social media. *IEEE Trans. Visual Comput. Graphics* 20:1773–1782

[6] Liaskos C, Xeros A, Papadimitriou G I, Lestas M and Pitsillides A 2012 Balancing wireless data broadcasting and information hovering for efficient information dissemination. *IEEE Trans. Broadcast*. 58: 66–76

[7] Hu Y L 2012 Information sensing and interaction technology in internet of things. *Chin. J. Comp*. 35: 1147 – 1163

[8] Huang X, Cheng H B and Yang G 2009 Research of connectivity for wireless sensor networks. *J. Commun. Networks* 30: 129 – 135

[9] Sangwan A and Singh R P 2014 Survey on coverage problems in wireless sensor networks. *Wireless P. Commun*. 80: 1475 – 1500

[10] Wang L, Xing and Vokkarane V M 2014 Reliability and lifetime modeling of wireless sensor nodes. *Microelectron. Reliab*.54: 160 – 166