GNN for Wireless Link Anomaly Detection

Blaž Bertalanič, Mihael Mohorčič and Carolina Fortuna
Department of Communication Systems, Jožef Stefan Institute, Slovenia
{blaz.bertalanic, miha.mohorcic@ijs.si, carolina.fortuna}@ijs.si

***Abstract*—In this paper, we present a new approach for detecting wireless link layer anomalies in large-scale IoT networks based on graph neural networks (GNN). We propose a method that transforms time series data into graphs with Markov Transition Field representation. The transformed data is then used to train a new GNN architecture that can successfully distinguish between 4 different link-layer anomalies and outperforms state-of-the-art shallow and deep learning methods.**

***Index Terms*—anomaly detection, wireless link, machine learning, graph neural network, graph transformation, time series**

I. Introduction

The exponential growth of the Internet of Things (IoT) [1] has been fuelled by the increasing affordability and availability of internet-connected devices, as well as the development of new wireless technologies and standards that make it easier to connect devices to the internet. This also enables the industry to upgrade its legacy infrastructure in the process of digitalization that increases the productivity and reliability, but it results in large complex IoT sub-networks that must be monitored and maintained. With a large number of IoT devices deployed in such sub-networks, it is crucial to make use of automatic monitoring of their status, especially in domains where wireless connectivity is mission-critical, such as for instance in healthcare, finance, and public safety. In this work we are focusing on a new approach for monitoring the status of wireless links and their anomalies, whereby we adopted four generic types of anomalies that can occur as defined in [2], namely sudden link degradation (SuddenD), sudden link degradation with recovery (SuddenR), instantaneous link degradation (InstaD) and slow link degradation (SlowD). In particular, we employ a machine learning-based method to detect and classify wireless link layer anomalies based on graph neural networks (GNN), which first transforms time series data into graphs and then classifies graphs using Graph Isomorphism Neural Networks

II. The proposed method

Let us assume there is a large IoT sub-network of devices where we want to provide high availability and anomaly-free data communication. To ensure this, we need to deploy an automatic monitoring system that collects and analyses time series data generated by the IoT devices within the sub-network. In case of anomalies, it sends a notification to specialised staff, who can then solve the problems accordingly.

We formulate the anomaly detection and classification part of the monitoring system as a classification problem that consists of three steps. The first step is the data collection part, the second is the transformation of time series data into graphs, and the third step is a classification of anomalies with GNN.

A. Time-series to graph transformation

The time-series to graph transformation G is done in two main steps: node determination and adjacency matrix computation. To determine the nodes of the graph we simply treat every point within the time series as a node. Given a time series $T_{1\times N} = \vec{t_N} = \{t_1, t_2, ..., t_N\}$ each point t_N is converted into a node or vertex in v_N with the value at t_N becoming the node's label (feature).

The second step is to compute the adjacency matrix corresponding to the set of nodes v_N. The adjacency matrix is also known as a connection matrix that is used to represent a connection between the nodes of the graph. To compute the adjacency matrix we adopted a Markov Transition Field (MTF) [3] representation of time series. MTF is a $N \times N$ square sparse matrix that encodes dynamic transition statistics from time series and represents Markov Transition probabilities in sequential order. To compute the MTF, the values of T are put into $B = \{b_1, b_2, ..., b_K\}$ bins where each value t_N is mapped to exactly one b_i. By counting the transitions between the bins b_i and normalizing the count we get the probability $w_{i,j}$ with which a point from bin b_i is succeeded in time by a point in the bin b_j. The MTF matrix is defined as follows:

$$\mathrm{M} = \begin{pmatrix} w_{ij|t_1\in b_i, t_1\in b_j}, \ldots, w_{ij|t_1\in b_i, t_N\in b_j} \\ w_{ij|t_2\in b_i, t_1\in b_j}, \ldots, w_{ij|t_2\in b_i, t_N\in b_j} \\ \vdots \quad \ddots \quad \vdots \\ w_{ij|t_N\in b_i, t_1\in b_j}, \ldots, w_{ij|t_N\in b_i, t_N\in b_j} \end{pmatrix} \tag{1}$$

Based on the calculated matrix M, the edges of G are constructed from M matrix for nodes v_N where every $w_{ij} > 0$ represents an edge between nodes i and j with w_{ij} being the edge weight. The resulting graph G contains a set of nodes v_N, edge indexes, and a set of edge weights.

An example of this process and transformation of anomalies defined in [2] are depicted in Fig. 1, where $T_{1\times 30}$ is transformed into a graph G with 30 nodes and their corresponding edges. On the left side of each of the sub-figures in Fig. 1 a raw time series representation of the anomaly is shown. The non-anomalous points in the time series are depicted in blue, while the points representing the anomalous part of the wireless link are in red. Both blue and red points correspond directly to the blue and red nodes depicted in the right part of the sub-figure that represents the transformed graph G.

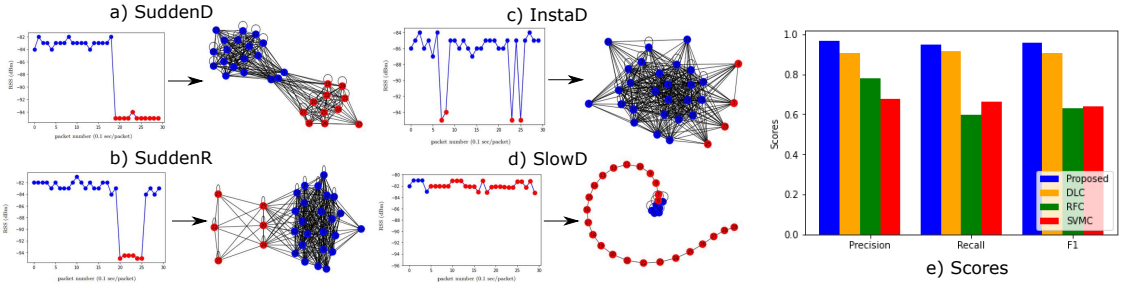

Fig. 1: Time-series and graph representations of the four types of anomalies and performance results of GNN model.

SuddenD is an anomaly where the RSS signal unexpectedly drops to a minimal value and never recovers. In the graph representation (Fig. 1a) nodes that are most similar to each other group together more and are also more densely connected.

SuddenR is similar to SuddenD only that the RSS signal recovers back to the normal value. Similar to SuddenD, the graph representation shows (Fig. 1b) that nodes group together with a higher number of connections between them.

InstaD shows itself as a spike within a RSS trace. Here the value drops to a minimum for an instantaneous amount of time and then recovers. In the corresponding graph representation (Fig. 1c) we can see that anomalies are mostly not connected between themselves but with non-anomalous nodes.

SlowD shows itself in a slightly decreasing slope within a RSS trace, where values slowly but gradually decrease. The graph representation (Fig. 1d), shows that anomalous nodes only connect to their neighbours within the time series, resulting in a snake-like graph representation.

III. GNN Anomaly Detection

A. Model architecture

For the anomaly detection part of the proposed method, we have designed a GNN architecture inspired by the Graph Isomorphism Neural Network (GIN) [4], specifically its variation named GINE [5] that takes advantage of graph edge features (weights). Based on the observations, graphs of similar-looking anomalies have similar nodes and edges connecting them, and thus graphs of the same anomalies are isomorphic. Our architecture consisted of 4 GINE layers and an output layer with softmax activation for the classification.

B. Dataset

To train the proposed GNN architecture, we utilized the Rutgers WiFi RSS dataset [6] as real-world measurement dataset. According to the guidelines, we synthetically inject four anomalies as per [2]. These traces were then transformed into graphs using the procedure described in Section II-A.

C. Model training and evaluation

The model was trained using a 10-fold stratified shuffle split technique in a ratio of 8:2. The performance of the proposed GNN mode was evaluated using precision, recall, and F1 score.

We evaluated the performance of the proposed method against the state-of-the-art methods from [7]. As shown in Fig. 1e compares the new proposed method with the time series based deep learning classifier (DLC), and two classical machine learning classifiers, namely Random Forest Classifier (RFC) and Support Vector Machine Classifier (SVMC), used as benchmarks also in [7]. As shown in Fig. 1e, the proposed method consistently outperforms the benchmark methods. In terms of the average F1 score, for instance, the GNN-based method outperforms the second best DLC by 0.05 (i.e., with 0.956 compared to 0.906), achieving this with only 0.026M trainable parameters compared to 0.333M parameters in case of DLC, yielding ≈12-fold reduction in complexity.

IV. Conclusions

Due to the complexity of the large IoT networks, it is essential to start using automatic systems for their status monitoring. This work investigated a new GNN-based approach for automatic anomaly detection in time series data describing the status of wireless links between IoT devices. The proposed method first transforms time series into graphs and then use a specific GNN-based architecture for their classification. We demonstrated that this proposed approach works on par with reference state-of-the-art methods at significantly reduced number of trainable parameters.

Acknowledgments

This work was supported by the Slovenian Research Agency under the grant P2-0016.

References

[1] J. Davies and C. Fortuna, *The Internet of Things: From Data to Insight*. John Wiley & Sons, 2020.

[2] G. Cerar, H. Yetgin, B. Bertalanic, and C. Fortuna, "Learning to detect anomalous wireless links in iot networks," *IEEE Access*, vol. 8, pp. 212 130–212 155, 2020.

[3] Z. Wang and T. Oates, "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks," in *Workshops at the 29th AAAI conference*, 2015.

[4] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *arXiv preprint arXiv:1810.00826*, 2018.

[5] W. Hu, B. Liu, J. Gomes, M. Zitnik, P. Liang, V. Pande, and J. Leskovec, "Strategies for pre-training graph neural networks," *arXiv preprint arXiv:1905.12265*, 2019.

[6] S. K. Kaul, I. Seskar, and M. Gruteser, "CRAWDAD dataset rutgers/noise (v. 2007-04-20)," Downloaded from https://crawdad.org/rutgers/noise/20070420/RSSI, Apr. 2007, traceset: RSSI.

[7] B. Bertalanic, H. Yetgin, G. Cerar, and C. Fortuna, "A deep learning model for anomalous wireless link detection," in *Proc. 17th WiMob Conference*, 2021, pp. 265–270.