IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 15 September 2023, accepted 21 October 2023, date of publication 25 October 2023, date of current version 2 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3327647

RESEARCH ARTICLE

Research of Deep Learning and Adaptive Threshold-Based Signaling Storm Prediction and Top Cause Tracking

DONGDONG FENG, SIYAO LI, ZHIMING YANG, YONG XIANG, JIAHUAN ZHENG, AND XIAOWU HE
China Telecom Research Institute, Guangzhou 510630, China

Corresponding author: Dongdong Feng (fengdd@chinatelecom.cn)

ABSTRACT Malicious software or misbehaving applications have the potential to trigger signaling storms on mobile networks, leading to battery drainage on devices and causing bandwidth overuse at the cell level. Additionally, these storms may result in an excessive signaling load within the mobile operator's infrastructure. This paper uses a combination of time series prediction, adaptive threshold, and anomaly detection algorithms to predict signaling storms. Whether a signaling storm will be triggered in the future can be determined based on the fluctuation pattern of the data. Our method enables us to identify the top cause of the signaling storm in advance, so that the network optimization team can address the issues that will arise in advance, maximizing the stop-loss. The time series prediction algorithm has significant advantages over the moving average and TFT(Temporal Fusion Transformers), with a WAPE(Weighted Absolute Percentage Error) of only 0.09. Adaptive threshold can avoid treating holiday data as abnormal data, and the accuracy of anomaly detection based on the automatic adaptive threshold is higher than the traditional fixed threshold. In addition, combining the signaling conduction chain can also perform top cause localization to identify the upstream network element instance that first encountered the problem. The entire algorithm not only performs well in the current network but also performs well in artificially generated signaling storm data, pioneering the field of signaling storm prediction.

INDEX TERMS Time series prediction, adaptive threshold, conduction chain, signaling storm prediction.

I. INTRODUCTION

The core network mainly consists of three parts: 4G core network, 5G core network, and IMS(IP Multimedia Subsystem) core network. The user's mobile terminal may initiate a call after completing registration on the core network. The core network is composed of multiple network elements with their corresponding ability to handle registration and call requests.

As the network receives more terminal signaling requests than it can handle, network congestion and the avalanche effect occur, resulting in network unavailability. This is referred to as a "signaling storm" by the media. For example, in the event of natural disasters such as earthquakes, a large number of users make phone calls, resulting in a call impact, excessive pressure on the PSBC (Proxy Call Session Control Function and Session Border Control) and DRA (Diameter Route Agent) network elements of the IMS network leads to a signaling storm. Another example is that batch user registration can cause overloading of HSS(Home Subscriber Server) or UDM(Unified Data Management) which are both core network elements, leading to a signaling storm. The widespread use of smartphones and the quick proliferation of mobile Internet have drawn attention to the signaling storm effect that communications networks could trigger. This trend will worsen rather than improve in the 5G age. As a result, it is critical to investigate the signaling storm issue in the core network.

The associate editor coordinating the review of this manuscript and approving it for publication was Bilal Khawaja.

Due to signaling strain, Telenor, the fourteenth-largest telecom provider in the world, went unavailable for 18 hours in June 2011. With 3 million people affected, this outage cost Telenor over USD 18 million [1], [2]. Similar to this, in April 2012, Verizon's Long Term Evolution (LTE) network malfunctioned and was out of commission for 24 hours due to signal overload, impacting hundreds of thousands of subscribers [3]. In December of the same year, Verizon's 4G Network collapsed three times [4]. Approximately 2.52 million users in Tokyo were impacted by the DoCoMo outage in January 2012, which was the seventh such incident in only eight months and was brought on by signaling load [5]. In the early morning of July 2, 2022, KDDI, the second largest operator in Japan, experienced a signaling storm due to a failed cutback operation of its core router, resulting in a 62-hour mobile network malfunction that affected 39.15 million users nationwide and a serious impact on industries such as finance, aviation, logistics, automotive, and power in Japan. It was the largest network system failure that KDDI had ever encountered [6]. These network failures demonstrate how critical it is to analyze signaling storm concerns.

The existing research on signaling storms in diameter networks lacks effective prediction, detection, and handling methods. To address recent episodes of signaling storms, commercial solutions have emerged, mainly falling into three categories:

(i) Anomaly detection and mitigation tools [7]: These approaches involve counting the number of successive signaling transitions that do not use allocated bandwidth. Mobile devices exceeding a certain threshold are temporarily blocked to prevent network overload.

(ii) Air interface optimization: This group aims to increase the number of simultaneously connected devices in the access network. The technologies in this category continually evolve with new standards, specifications, and proprietary admission/congestion control and scheduling algorithms. For example, one solution to address signaling issues over the s4 interface between SGSN (Service GPRS Support Node) and MME (Mobile Management Entity) is to internalize the s4 interface by co-locating the SGSN and MME in the same device [8].

(iii) Dedicated signaling infrastructure solutions: These solutions target the expected growth in core network signaling, especially concerning policies, charging, mobility management, and other new services introduced in LTE networks. Authors in [9] shed light on signaling attacks targeting the RRC (Radio Resource Control) protocol that generate signaling storms for 3G networks. Enabling dynamic resource scaling in response to network traffic requirements, congestion control, and load balancing in the core network is projected to be less problematic with the trend toward network function virtualization.

In terms of signaling storm prediction, some scholars use classification algorithms to predict signaling storms. Zhang et al. predicted signaling traffic and set traffic thresholds, and predicted signaling storms by comparing the predicted values with the thresholds [10]. Among those methods, some are still in the theoretical stage, while others have poor prediction performance in storm scenarios due to limited signaling storm data in the current network.

In our research, to predict the growth of signaling services, a dynamic signaling traffic prediction model is first built using historical traffic data and a prediction algorithm. The adaptive method is then merged, breaking the restriction that the original algorithm utilizes fixed thresholds, to dynamically configure and adjust the anomalous thresholds of the signaling traffic-related indicators of network elements. This can ensure that when holidays come, our traffic threshold is relatively high to avoid identifying holidays as anomalies. This study will also discover the aberrant top-cause network element that generates this signaling storm based on the justification of the entire signaling network conduction chain after detecting that a signaling storm is going to occur. Our contributions are summarized as follows:

- In order to forecast the growth of signaling services, this method builds a signaling traffic prediction model using a prediction algorithm and historical traffic data. This method's signaling model can more accurately and promptly predict traffic value since it can better capture traffic spikes.
- Not just the signaling storm-associated determination index can be predicted using this method. The technology makes advantage of adaptive dynamic traffic thresholds to effectively manage signaling storms. The signaling traffic thresholds for network element anomaly identification and evaluation were determined using the aforementioned prediction and evaluation results and the current adaptive algorithm.
- In addition, once a signaling storm is anticipated or confirmed, an early warning message is promptly sent out, and the top cause of the storm is identified using the network element conduction chain. This information is then given to the technical maintenance team so they can prepare for potential future signaling storms. As a result, this approach can effectively prevent and mitigate signaling storms.

The remaining part of the article is divided into four parts. Section II introduces relevant research in this field. Section III introduces our proposed method. Section IV is the experimental section including a discussion of the experimental results. Section V summarizes our work and provides some prospects.

II. RELATED WORKS

A. SIGNAL STORM IDENTIFICATION AND PREDICTION

Overall, there is a lack of effective solutions to predict or detect signaling storms. Yang Lin et al. have proposed a solution for ZTE's integrated signaling storm [11]. Usually, only one static, passive solution is used, which is to configure traffic thresholds in each link to handle signaling storms [12].

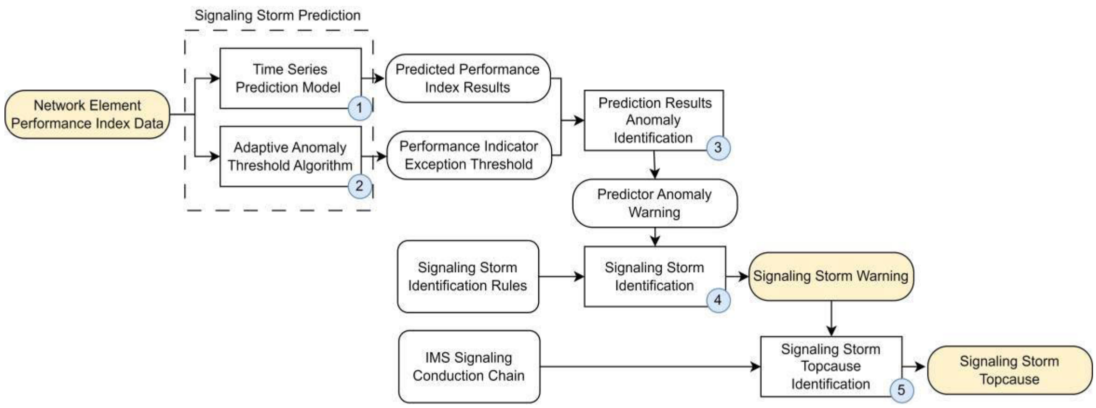


FIGURE 1. The architecture of the proposed framework.

Zhang Ning et al. used self affine fractal interpolation method [13], [14] to predict signaling traffic, and used adaptive thresholds to predict signaling storms by comparing predicted values with thresholds [10].

B. TIME SERIES PREDICTION

Many scholars use temporal prediction algorithms to predict temporal data. Zhang et al. use a self-affine fractal interpolation algorithm to predict signaling traffic [10]. Mu Wenxuan et al. used neural networks to analyze global warming [15]. The moving average model is adopted for the forecast of the trend in sales data of a confectionery baking industry [16]. In 2021, Google released a great time series prediction model TFT (Temporal fusion transformers [17]), which has strong explanatory power and is favored by many scholars. Shereen Elsayed et al. used ensemble decision trees for time series prediction, outperforming most deep neural networks in multiple tasks [18]. Haiyi Zhou et al. invented a time series prediction algorithm called Informer, which is a long series prediction artifact [19].

C. ANOMALY DETECTION

There are three main methods for anomaly detection:

- Distance based anomaly detection ([20], [21]);
- Density based anomaly detection ([22]);
- Clustering based anomaly detection ([23]).

Distance-based anomaly refers to that given distance d, if the number of other points in a circle with a radius of d around data point A is very small, then data point A is considered an outlier. Density based anomalies refer to that data objects in low density areas as outliers. Clustering based anomaly detection refers to grouping similar data points into a cluster, where outliers are points far from the cluster center.

One-class learning methods are also used for anomaly detection (e.g., [24]). Unlike the above methods, this method is trained on a normal dataset and then used to detect outliers.

Multiple outlier detection techniques, including conventional and one-class learning approaches, were examined by Swersky et al. [25], finally, it is concluded that SVDD(Support Vector Data Description) and LOF(Local Outlier Factor [22]) are superior to other anomaly detection algorithms, but these tests are implemented in a static context. [26] utilizes an autoencoder to detect outliers in data streams. Every time there is a notion drift, the model is retrained. Xuanhao Chen et al. proposed an unsupervised anomaly detection algorithm for multivariate time series.

The above methods all overlook the important issue of how to set the threshold between normal and abnormal values. Configuring traffic thresholds in each diameter link is highly challenging because a DRA device contains hundreds of diameter links, each with a different traffic value. As the network function workload changes, the traffic value of a diameter link changes and gets higher. As a result, the rising network demand cannot be supported by a fixed static traffic threshold. If the threshold for signaling traffic is set too high, the traffic control function will never reach its target value. The fixed static traffic threshold cannot be used in this circumstance since the signaling traffic value will dramatically grow over the vacation period. In conclusion, it is not possible to define fixed thresholds in signaling storm identification to effectively predict and manage signaling storms.

III. PROPSED FRAMEWORK

In this article, we invent an algorithm called DLAT to predict signaling storms. The algorithm mainly includes three parts: deep learning for time series prediction of traffic, adaptive threshold algorithm for providing dynamic thresholds of traffic, signaling storm discrimination and Top Cause identification module for determining whether a signaling storm is about to occur, and outputting the upstream network element that triggers the signaling storm. The algorithm architecture diagram is shown in Figure 1.

A. TIME SERIES PREDICTION

This paper uses a multiple regression neural network algorithm with linear layer and Relu activation function, to predict the future value of indicators by inputting past indicator values, time, holiday characteristics, etc. The calculation formula of the algorithm is as follows:

$$a_j^{[i]} = w_1^{[i-1]} * a_1^{[i-1]} + w_2^{[i-1]} * a_2^{[i-1]} + \ldots + w_n^{[i-1]} * a_n^{[i-1]} + b_n^{[i-1]} \quad (1)$$

$$a_j^{[i]} = Relu(a_j^{[i]}) \quad (2)$$

where $a_1^{[i-1]}, a_2^{[i-1]}, \ldots, a_n^{[i-1]}$ are the outputs from the previous layer, $(\boldsymbol{w}, \boldsymbol{b})$ are parameters to be learned.

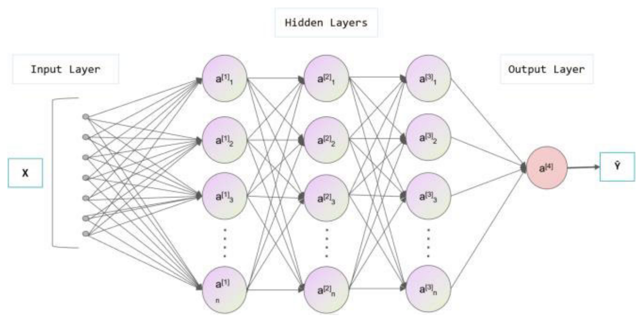


FIGURE 2. **Multiple regression diagram.**

```
Algorithm 1: Adaptive Threshold for Anomaly Detection
   Data: initial default data sets S_val, the window size for z test is
         T=288, the window size for updating the threshold is M=576,
         the Upper Bound is 5760, the α value for the z test is α=0.05
   Result: anomaly threshold threshold
 1 if detection_label == 0 then
 2     if cv ≤ 0.1 then
 3         default_threshold ← 3 * std(S_val) + mean(S_val);
 4     else if cv>0.1 then
 5         default_threshold ← 2 * mean(S_val);
 6     end
 7 else if detection_label == 1 then
 8     if cv ≤ 0.1 then
 9         default_threshold ← min(−1 * std(S_val) + mean(S_val), 0.95);
10     else if cv>0.1 then
11         default_threshold ← 0.9 * mean(S_val);
12     end
13 end
14 w1 ← S_val
15 while x_t is available do
16     w2 ← w2 ∪ x_t;
17     w ← w ∪ x_t;
18     if length(w1) ≤ T then
19         w1 ← w1 ∪ x_t;
20     if length(w2)>T then
21         if ztest(w1, w2) ≤ α then
22             threshold ← update_threshold(w);
23             w ← w[length(w) − T] to w[length(w)];
24             w1 ← w2;
25             w2 ← clear window;
26         else
27             w2.remove(w2[0])
28         end
29     if length(w) mod M=0 then  threshold ← update_threshold(w) ;
30     if length(w)>upperbound then  w.remove(w[0]) ;
31 end
```

FIGURE 3. **The pseudocode for adaptive threshold algorithm.**

B. ADAPTIVE THRESHOLD FOR ANOMALY DETECTION

In this research, we applied an automatic threshold updating algorithm to address the threshold setting process required for anomaly detection in non-static data streams efficiently, without the requirement to provide additional data with labels beyond the starter background or ordinary dataset required to train the anomaly detector. We have developed an intelligent adaptive approach that utilizes sliding windows and hypothesis testing. Hypothesis testing is employed to determine whether the thresholds should be updated. The algorithm, referred to as Algorithm 1, is divided into two parts and takes the history data (S) and the parameter values as inputs. It's important to note that the training data exclusively consists of normal data.

The algorithm is divided into two parts: one to calculate the threshold and one to assess whether the threshold needs to be updated based on the sliding window. For the part of threshold calculation, the detection label, which is defined to mark whether an upper threshold or a lower threshold is required for this indicator, is applied to calculate the threshold in two ways. Then the CV (Coefficient of Variation) is defined as:

$$c_v = \frac{\sigma}{\mu} \quad (3)$$

where the σ is the standard deviation of the input data, and the μ is the average of the input data.

The threshold updating section can be launched by applying the initial threshold for the whole algorithm, which is obtained by the judgment of the CV value. The algorithm utilizes three windows: w, w1, and w2. The purpose of window w is to calculate the threshold, while windows w1 and w2 are used to assess whether the means of scores in both windows are significantly different. The choice between the t-test and the z-test for determining significant mean differences depends on the sample size; the t-test is used when the sample size is less than 30, and the z-test is employed otherwise. Given that the number of values reviewed before making a decision is always larger than 30, the algorithm opts for the z-test.

When both windows w1 and w2 are full, the algorithm computes the p-value generated from the z-test on both windows. If the p-value is less than 0.05, the threshold is updated based on the historical data in w. The update_threshold function calculates the new threshold using the historical values stored in w, that's what line 1 to line 12 do. The assumption is that a concept drift has occurred when the mean of w1 significantly differs from that of w2. Therefore, the threshold is recalculated based on the new values stored in w. After the update, w is shrunk to retain only the most recent values. This allows the algorithm to forget earlier data in the data stream and prepares the window for detecting potential future drifts.

To ensure safety and handle scenarios where the conditions for updating the threshold are never met, the algorithm automatically updates the threshold when the size of w becomes divisible by 576. However, in such cases where no concept drift has taken place, the window is not shrunk to avoid discarding important past data. To control storage consumption, an upper bound is set on the size of w. When w reaches this upper bound, the oldest values are removed to ensure that the size of w remains within the specified limit.

When holiday comes, the threshold will be updated due to changes in data distribution. Taking traffic indicators as an example, the threshold for holidays will be relatively high compared to normal days, in order to avoid misjudging holidays as abnormal after an increase in holiday traffic.

C. STORM IDENTIFICATION & TOP CAUSE

(Storm Identification) In order to continue with the storm identification procedure, we need to collect the indicator anomaly detection result. We may retrieve the predicted values of signaling indicators by the timing prediction algorithm, and by comparing them with the adaptive thresholds-generated thresholds, we can further filter out the anomalous indicators. The forthcoming signaling storm is divided into three groups based on the relevant signaling storm disposal plan materials and professional principles of communication, which are the call impact signaling storm, the batch registration signaling storm, and the DRA overload signaling storm. For example, when the number of registration requests for PSBC exceeds its adaptive threshold and the registration success rate is less than its adaptive threshold, we conclude that a signaling storm caused by batch registration is about to occur in the future. Diverse indications that are irregularly triggered and satisfy a logical judgment requirement are classified as one of the three signaling storms, and additional signaling storm warnings are sent out.

(Top Cause) The top cause network element that generated the signaling storm is then identified by our algorithm based on the network element signaling conduction chain after the signaling storm alert has been activated. Three separate signaling chains for the discovered three different signaling storm situations can be summed up by analyzing the signaling chain of network elements, which are illustrated in Table 1. The 4/5G conduction chains are also shown in Figures 4 and 5 (mainly marked with red lines). For ease of understanding, we use a table to illustrate the network elements, as shown in Table 2.

TABLE 1. Conduction chain.

Scenario	Conduction chain
5G Registration	AMF-->UDM-->PSBC-->ICSCF-->SCSCF-->HDRA -->LDRA-->HSS
5G Sessions	AMF-->SMF-->UDM-->PSBC-->ICSCF -->ENUMDNS-->SCSCF-->HDRA-->LDRA-->HSS
4G Registration and Session	MME-->SAEGW -->PSBC-->ICSCF-->ENUMDNS -->SCSCF-->HDRA-->LDRA-->HSS

The top cause network element that triggered the entire signaling storm at the very beginning can be further tracked back by identifying the anomalous network element for the present anomaly indicator and the signaling conduction chain. By pre-locating top cause network elements, network optimizers can address anomalies and abnormalities in advance, preventing signaling storms to the maximum extent possible.

TABLE 2. Description of the network element.

Network element	Description
AMF	Access and Mobility Management Function
SMF	Session Management Function
UDM	Unified Data Management
PSBC	Proxy Call Session Control Function and Session Border Control
ICSCF	Interrogating-Call Session Control Function
SCSCF	Service-Call Session Control Function
ENUMDNS	Electronic NUMbering Domain Name System
DRA	Diameter Routing Agent
HSS	Home Subscriber Server
MME	Mobile Management Entity
SAEGW	System Architecture Evolution Gate Way

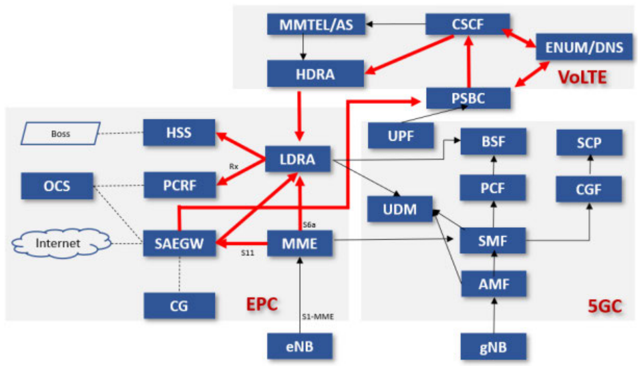


FIGURE 4. 4G conduction chain.

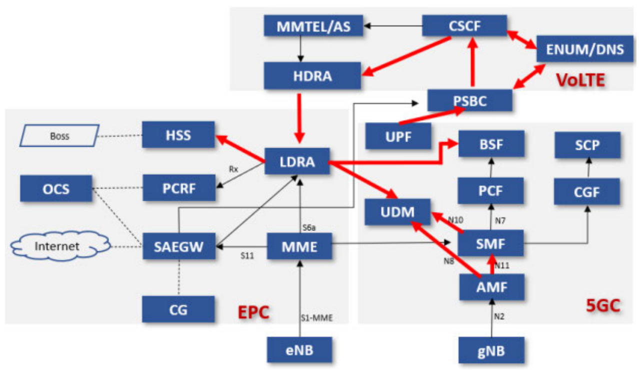


FIGURE 5. 5G conduction chain.

IV. EXPERIMENTS

A. DATASET

1) DATA SOURCE

The experimental data is generated and collected through experimental equipment. We connect experimental equip-

TABLE 3. The basic dataset statistics.

# element name	# element type	# sequence	# records
18	11	86	149K

TABLE 4. Part of indicators used and corresponding calculation formulas.

Network element	Indicator	Formula
PSBC	registration number	-
	registration success rate	$\frac{\text{registration success number}}{\text{registration number}}$
	call success rate	$\frac{\text{call success number}}{\text{call number}}$
AMF	registration number	-
	registration success rate	$\frac{\text{registration success number}}{\text{registration number}}$
UDM	active user number	-
ICSCF	called success rate	$\frac{\text{called success number}}{\text{called number}}$

TABLE 5. Comparison of the performance of three algorithms.

Algorithm	WAPE(storm scenario)	WAPE(normal conditions)
DLAT	0.091	0.083
Moving average	0.298	0.089
TFT	0.752	0.088

ment through NWDAF which is the network data analysis function of 5G core network to obtain PSBC, AMF and other network element data, and store the data in the hive database for analysis and prediction. The time span is from June 1, 2023 to June 18, 2023.

2) VARIABLE DESCRIPTION

Since the data is transmitted through a unified interface, the data has a standard format, including:

- Performance indicator measurement: counter ID, measResult (measure Result);
- Network element-related information: element name, element type, region code;
- Time information: start time, end time, duration.

3) DATA PROCESSING

Data is naturally processed, especially when calculating the total success rate, by dividing the total number of successful requests by the total number of requests. The basic dataset statistics are shown in Table 3 (A sequence consists of the element name and its corresponding counter ID). Table 4 is used to present some of the indicators and calculation formulas we use.

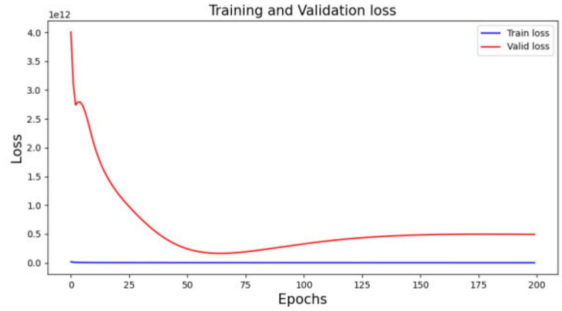


FIGURE 6. Training and validation loss of our prediction model.

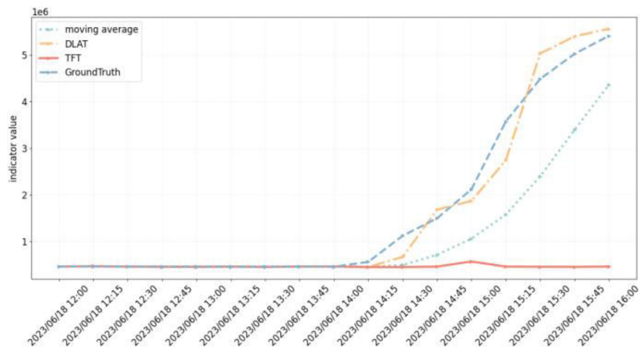


FIGURE 7. Line chart of three algorithms' prediction results (number of registration requests of a certain PSBC).

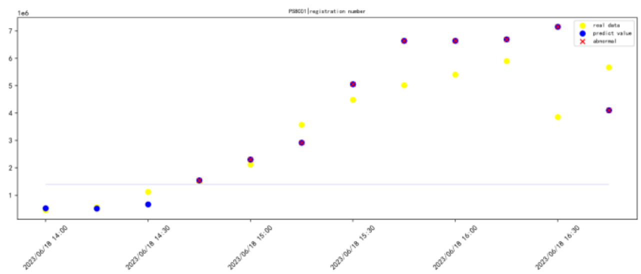


FIGURE 8. Prediction results and anomaly detection results of a certain PSBC's registration number.

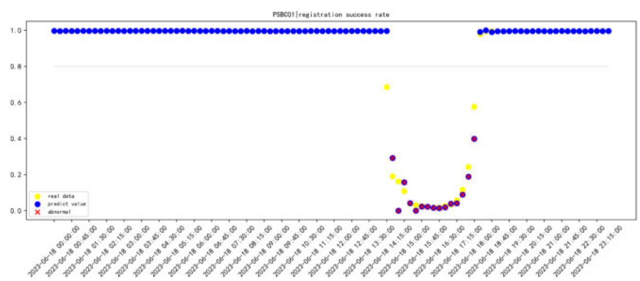


FIGURE 9. Prediction results and anomaly detection results of a certain PSBC's registration success rate.

B. COMPARISON OF PREDICTION ALGORITHMS

Figure 6 shows the trend of the loss function of the model on the training set and the validation set. As the number of

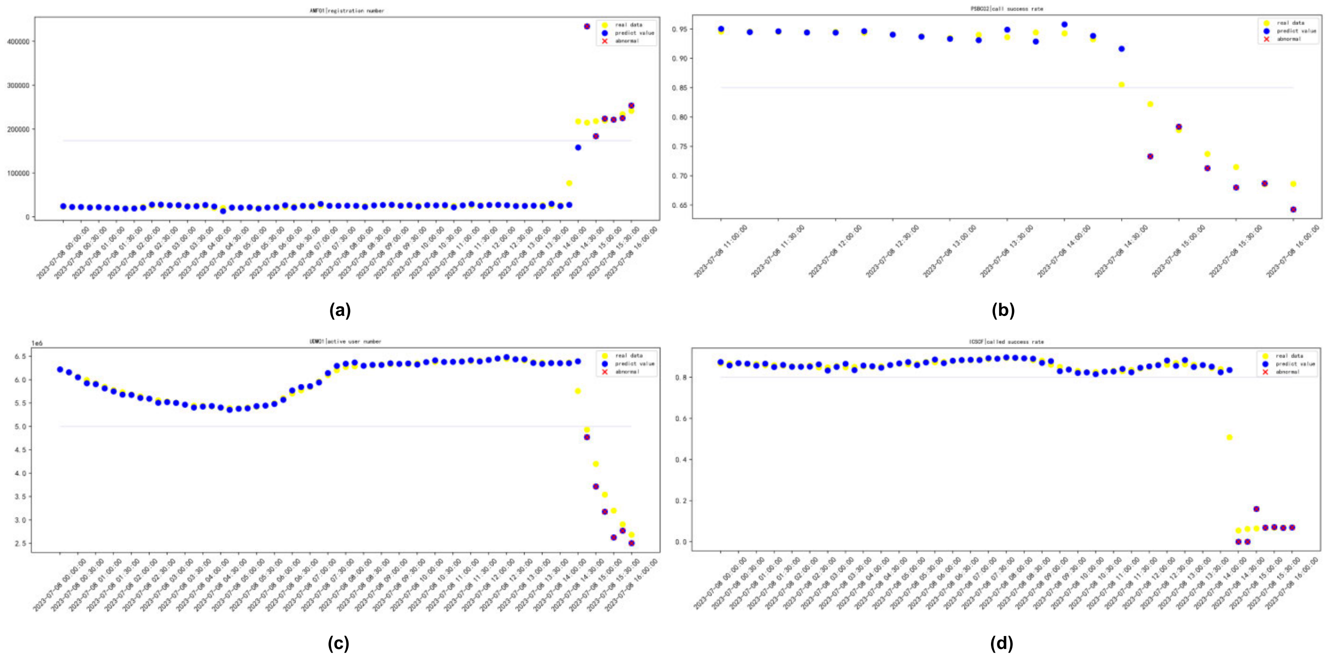


FIGURE 10. (a) Prediction results and anomaly detection results of AMF's registration number. (b) Prediction results and anomaly detection results of PSBC's call success rate. (c) Prediction results and anomaly detection results of UDM's active user number. (d) Prediction results and anomaly detection results of ICSCF's called success rate.

training steps increases, both the training loss and validation loss reach relatively reasonable values.

We compare the experimental results of the algorithm in this article with the moving average and TFT. Table 5 shows that our algorithm has significant advantages compared to the moving average and TFT. Under normal circumstances, the three algorithms have similar effects, but they appear high and low in storm scenarios.

Under normal circumstances, we visualized the true number of registration requests for a certain PSBC and the predicted values of the three algorithms and found that our algorithm had a good fitting effect, with the moving average having a certain lag, while TFT had poor performance. This may be because the training data of TFT did not include data from storm scenarios. Of course, there are very few signaling storm data in real network.

C. VISUALIZATION OF OVERALL ALGORITHM RESULTS

Figure 8 shows the prediction, threshold, and anomaly detection of the number of registration requests for a certain PSBC network element instance. The gray line represents the adaptive threshold, which is updated once over time. At 14:30, an abnormal indicator was predicted (predict a point every 15 minutes, and the value at 14:45 will exceed the threshold).

Similarly, Figure 9 shows the registration success rate of the network element instance. At 14:30, it is predicted that the indicator will soon fall below the threshold and an exception will occur.

Figures 10 shows the prediction and anomaly detection of relevant indicators for AMF, PSBC, UDM, and ICSCF network elements. The predicted value of AMF's registration number has suddenly increased at about 14:45, while the PSBC's called success rate, UDM's active user number, and ICSCF's called success rate have all experienced a sudden decrease.

D. ANALYSIS OF RESULTS

The blue dots in Figure 8 represent the predicted value of PSBC registration number, and the predicted value of 14:45 is predicted at 14:30. At the same time, the adaptive threshold algorithm provided an adaptive threshold (as shown by the gray line) at 14:30. Since the blue dot at 14:45 exceeded its adaptive threshold, it was judged as abnormal. That is to say, we predicted at 14:30 that the registration number for PSBC in the future will be abnormal.

Similarly, Figure 9 shows that we predicted at 14:30 that the future registration success rate of PSBC would be abnormal. Based on the signaling storm discrimination rules (as mentioned in the previous Section III Storm Identification module), since both the predicted value of registration number and registration success rate are abnormal, it can be determined that a signaling storm will occur in the future due to batch registration). That means we predicted at 14:30 that a signaling storm would occur in the future.

As this signaling storm is a signaling storm of registration type, we can search for the top cause along the 5G registration conduction chain and ultimately find the AMF (Figure 10 (a) indicates that we predicted at 14:30 that registration number for a certain AMF network element instance in the future will exceed its adaptive threshold, meaning that AMF is about to experience an exception). As AMF is at the upstream of the entire conduction chain, it is the top cause of this signaling storm.

At this point, we successfully predicted a signaling storm of registration type at 14:30, and the top cause of this signaling storm is AMF.

V. CONCLUSION

This article combines time prediction, adaptive threshold, and anomaly detection algorithms to predict signal storms. Firstly, this article differs from previous research on the control of signaling storms. We use time series prediction algorithms to predict signaling storm data, which facilitates the early detection and processing of signaling storms. Additionally, the multiple regression neural network algorithm used in this article is lighter and has better performance in terms of structured data (WAPE can reach 0.091 even in storm scenarios). This method, further combined with adaptive threshold-based anomaly detection, which is more effective and outperforms fixed anomaly thresholds to avoid considering holiday data as abnormal data. Finally, the combined signaling conduction chain of this method can also perform top cause localization to identify the upstream network element instance that encountered the problem first. The entire algorithm pioneered signal storm prediction, achieving prediction of signaling storm and top cause localization, which can prevent the occurrence of signaling storms in advance. In the future, more feature assisted prediction will be introduced, such as UE (User Equipment) information, to achieve earlier prediction of signaling storms.

REFERENCES

[1] (Jun. 2011). *Signal Storm Caused Telenor Outages*. [Online]. Available: http://www.newsinenglish.no/2011/06/16/signal-storm-caused-telenor-outages/

[2] (Jun. 2011). *Telenor Explains Mobile Outage*. [Online]. Available: http://www.lightreading.com/telenor-explains-mobile-outage/d/d-id/687713

[3] (Apr. 2011). *Verizon Reports LTE Network Crash*. [Online]. Available: http://www.wirelessweek.com/news/2011/04/verizon-reports-lte-network-crash

[4] (Apr. 2011). *Verizon 4G Outage is Third in Month: What's Causing 'Most Reliable' Network to Crash?*. [Online]. Available: http://www.ibtimes.com/verizon-4g-outage-third-month-whats-causing-most-reliable-network-crash-388920

[5] (Jan. 2012). *DoCoMo to Spend 164 Billion yen on Network After Service Outage*. [Online]. Available: http://www.bloomberg.com/news/2012-01-27/docomo-to-spend-164-billion-yen-on-network-after-service-outage.html

[6] (Jul. 2022). *The Most Serious Network Interruption Accident in Japan's History! What are the Inspirations for the Development of 5G in China?*. [Online]. Available: https://www.sohu.com/a/565707173_219833

[7] E. Gelenbe and O. H. Abdelrahman, "Countering mobile signaling storms with counters," *Internet Things*, vol. 169, pp. 199–209, 2016.

[8] (Jun. 2012). *The Evolution of the Signaling Challenge in 3G and 4G Networks (White Paper)*. [Online]. Available: http://nashvillelibrary.dyndns.org:823/YourFreeLibrary/lte/Heterogeneous%20Networks/HR Cisco LTE Signaling WP Final.pdf

[9] G. Gorbil, O. H. Abdelrahman, M. Pavloski, and E. Gelenbe, "Modeling and analysis of RRC-based signalling storms in 3G networks," *IEEE Trans. Emerg. Topics Comput.*, vol. 4, no. 1, pp. 113–127, Jan. 2016.

[10] N. Zhang, "Research on the method of handling diameter signaling storm based on active prediction," *Shandong Commun. Technol.*, vol. 42, no. 3, pp. 40–43, Sep. 2022.

[11] L. Yang, "ZTE integrated signaling storm solution," *Commun. World*, vol. 35, p. 29, Sep. 2012.

[12] Y. Ding, "Research on signal storm suppression scheme for diameter signaling network," Beijing Univ. Posts Telecommun., Beijing, China, vol. 3, no. 64.

[13] H. Gao and D. Zhang, "A traffic flow prediction model based on fractal and cubic exponential smoothing," *J. Nanjing Univ. Posts Telecommun. Natural Sci. Ed.*, vol. 38, no. 6, p. 5, 2018.

[14] F. Wang, S. Li, and H. Wang, "A short-term prediction method of air traffic flow based on fractal interpolation," CN Patent 1 11 785 093 A, Oct. 16, 2020.

[15] M. Wenxuan, M. Jiana, D. Xin, and M. Tengfei, "Multi-dimensional analysis of global warming based on neural network and grey system," *Academic J. Comput. Inf. Sci.*, vol. 6, no. 4, pp. 1–8, Apr. 2023.

[16] R. A. Kazeem, M. O. Petinrin, P. O. Akhigbe, T. C. Jen, E. T. Akinlabi, S. A. Akinlabi, and O. M. Ikumapayi, "Forecast of the trend in sales data of a confectionery baking industry using exponential smoothing and moving average models," *Math. Model. Eng. Problems*, vol. 10, no. 1, pp. 1–13, Feb. 2023.

[17] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecasting*, vol. 37, no. 4, pp. 1748–1764, Oct. 2021.

[18] S. Elsayed, D. Thyssens, A. Rashed, H. S. Jomaa, and L. Schmidt-Thieme, "Do we really need deep learning models for time series forecasting?" 2021, *arXiv:2101.02118*.

[19] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. 35th AAAI Conf. Artif. Intell.*, May 2021, vol. 35, no. 12, pp. 11106–11115.

[20] E. M. Knorr, R. T. Ng, and V. Tucakov, "Distance-based outliers: Algorithms and applications," *VLDB J. Int. J. Very Large Data Bases*, vol. 8, nos. 3–4, pp. 237–253, Feb. 2000.

[21] K. Zhang, M. Hutter, and H. Jin, "A new local distance-based outlier detection approach for scattered real-world data," 2009, *arXiv:0903.3257*.

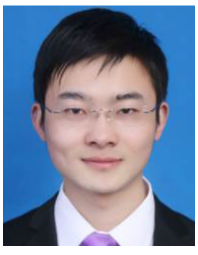
[22] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," *ACM SIGMOD Rec.*, vol. 29, no. 2, pp. 93–104, Jun. 2000.

[23] A. Loureiro, L. Torgo, and C. Soares, "Outlier detection using clustering methods: A data cleaning application," in *Proc. KDNet Symp. Knowl.-Based Syst. Public Sector*, Bonn, Germany, 2004, pp. 1–12.

[24] Y. Ma, P. Zhang, Y. Cao, and L. Guo, "Parallel auto-encoder for efficient outlier detection," in *Proc. IEEE Int. Conf. Big Data*, Silicon Valley, CA, USA, Oct. 2013, pp. 15–17.

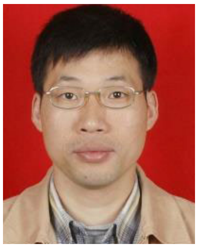
[25] L. Swersky, H. O. Marques, J. Sander, R. J. G. B. Campello, and A. Zimek, "On the evaluation of outlier detection and one-class classification methods," in *Proc. IEEE Int. Conf. Data Sci. Adv. Analytics (DSAA)*, Montreal, QC, Canada, Oct. 2016, pp. 1–10.

[26] Y. Dong and N. Japkowicz, "Threaded ensembles of autoencoders for stream learning," *Comput. Intell.*, vol. 34, no. 1, pp. 261–281, Feb. 2018.



DONGDONG FENG received the B.E. and M.S. degrees from Nanjing University, China, in 2012 and 2015, respectively. He is currently a Researcher with the China Telecom Research Institute, Guangzhou, China. His research interests include integrated tree models, fundamental models, trajectory prediction, time series forecasting, and optimization algorithm.

SIYAO LI received the M.S. degree from the King's College of London. She is currently a Researcher with the China Telecom Research Institute, Guangzhou, China. Her research interests include time series prediction, anomaly detection, reinforcement learning, and optimization algorithms.



ZHIMING YANG received the B.E. degree from the Huazhong University of Science and Technology, China, in 1995, and the M.S. degree from the South China University of Technology, China, in 1998. He has more than 20 years of experience in research and development, construction, and operation of large-scale intelligent operating systems. He is currently a Senior Researcher with the China Telecom Research Institute, Guangzhou, China. With Long-Term Research Development Experience, he has published a monograph "Technology and Practice of Operation and Maintenance Support System." His research interests include intelligent networks, intelligent operation and maintenance, and machine learning.

YONG XIANG received the M.S. degree from the University of Electronic Science and Technology, China. He is currently a Senior Expert with China Telecom and the Director of the Cloud Network Acquisition and Control Center. He was with China Telecom for 20 years and has rich experience in data processing.

JIAHUAN ZHENG received the M.Sc. degree in applied statistics, specializing in data analysis and data mining from Jinan University, China. She was with the Research Institute of China Telecom and she joined in a variety of projects, including the research and development of big data, such as data mining and AI algorithms in the field of communications, as an intermediate engineer.

XIAOWU HE received the bachelor's and master's degrees from the School of Information Science and Technology, Sun Yat-sen University, in 2010 and 2012, respectively. Since 2012, he has been with the China Telecom Research Institute as a Researcher and an Engineer. His research interests include 5G systems, the Internet of Things, software defined networks, and cloud computing.