Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 5099508, 11 pages
https://doi.org/10.1155/2022/5099508

Research Article

KL-Dection: An Approach to Detect Network Outages Based on Key Links

Ye Kuang, Dandan Li, and Xiaohong Huang

School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Xiaohong Huang; huangxh@bupt.edu.cn

Received 18 January 2022; Revised 23 January 2022; Accepted 18 February 2022; Published 7 March 2022

Academic Editor: Shalli Rani

Monitoring the states of network links is essential to detect network outages and improve Internet reliability. Currently, existing work detects network outages by monitoring all the links, which requires thousands of probes and large-scale measurements, resulting in high resource occupancy and cost. To solve this problem, this paper proposes the KL-Dection approach, which detects network outages via key links instead of all links. Firstly, we recognize the key links based on flow density, degree centrality, and probe-distance centrality. Next, based on the recognized key links, we give the critical value of their Round-Trip Time (RTT). Then, we detect the network outages by observing whether the RTT of the key link exceeds the critical value. Finally, we leverage two historical events to evaluate our approach, and the results demonstrate that our approach can detect the network outages effectively by only monitoring less than 0.06% of the links in detection area.

1. Introduction

The unprecedented growth of the Internet has resulted in an explosive increase in network security issues, such as network outages. Network outages inevitably degrade network connectivity and influence network performance [1–4]. For example, the network outage caused by censorship in 2011 blocked the Internet access of Libya [5]. Hence, the detection of network outages has become vital.

Over the years, several detection approaches have been proposed to detect network outages. These approaches are based on active probing, which deploy a large number of probes to monitor the changes in link performance (e.g., delay and connectivity). Specifically, Fontugne et al. detected the network outages by analyzing the Round-Trip Time (RTT) of all links in detection areas [6]. Quan et al. detected the network outages by deploying the probes to observe the connectivity of all links in edge networks [5]. Padmanabhan et al. detected the network outages by using ThunderPing [7] to measure the connectivity of all residential links in detection areas [7].

The above work detected the network outages by monitoring the performance of all links in detection area. However, they lead to high resource occupancy and cost in practice. The reasons lie in the following: (1) monitoring the network performance of all links needs to perform a bulk of measurement tasks. These tasks will inject extra traffic into the network, which may occupy the link bandwidth, reduce the network transmission speed, and increase the network burden [8, 9]. (2) Monitoring all links in detection area needs to deploy more probes, and managing the probes is costly for network operators (e.g., periodic maintenance and electricity costs). Thus, how to reduce the resource occupancy and cost by reducing the number of monitoring links without compromising the validity of outage detection is a challenge.

Previous work of traffic monitoring provided initial inroads to address this challenge [10, 11]. Their research results showed that a few key links that deliver larger traffic flows can well represent the traffic load information of all links in the detection area. However, focusing on these key links recognized by traffic load information is inadequate to detect network outages.

This is because detecting the network outages also needs to focus on the changes in link performance, e.g., RTT [4, 12]. In fact, due to the presence of noise and the interaction between the RTT of links, the fluctuation of RTT of each link is different. Only the link whose RTT fluctuation can notably and accurately reflect the deviations between the state of network outages and normal state can be regarded as the key link. Based on the above analysis, recognizing the key links for network outage detection should consider two aspects, one is the traffic load information, another is the factors affecting the RTT of link.

To achieve it, we use the number of flows going through the links to describe the traffic load information and use connected relation as well as the position of the link to describe the factors affecting the RTT of link. The reasons are as follows: (1) the number of flows going through the links describes the ability of links delivering traffic flows [13]. The links delivering larger flows can approximatively represent the traffic load information of all links in the detection area [12]. (2) The RTT of links will be affected by neighbor correlation [4]. The RTT of the links with poorer connected relations will not be significantly influenced by the neighbor correlation because these links have fewer adjacencies [4]. This causes that the RTT of these links may not fluctuate obviously in the case of a network outage. Hence, monitoring the links with poor connected relation may fail to detect the network outages. (3) With the distance between the probes and the nodes of links increases, the noise of measurements inflates [14]. Generally, the link is closer to the probe, and the RTT of the link we obtain is more accurate. Hence, monitoring the links close to the probes can accurately obtain the fluctuation of its RTT, which can detect network outages effectively.

After recognizing the key links, we monitor their RTT and detect the network outages finally. Our contributions are summarized as follows:

(i) For all we know, this is the first work that leverages the key links to detect network outages. Our approach can reduce the number of monitoring links notably without compromising the validity of outage detection

(ii) This paper proposes a key link recognition algorithm based on three metrics, i.e., flow density, degree centrality, and probe-distance centrality of links. Specifically, flow density describes the number of flows going through the links in unit time, degree centrality describes the connected relation, and probe-distance centrality describes the position of the link

(iii) This paper proposes a detection algorithm based on interquartile range, which detects the network outages by observing whether the RTT of any key link exceeds its critical value for a period of time. The experimental results demonstrate that our approach can detect network outages via key links rather than all links

The rest of this article is organized as follows. The following section provides a brief overview of network outage detection. Our approach and its architecture used for detecting the network outages have been explained in Section 3. The performance of the detection approach is discussed in Section 4. Finally, we draw the conclusions in Section 5.

2. Related Work

Several approaches have been proposed to detect network outages based on active probing. These approaches can be roughly divided into three categories according to different performance indexes they are based on, i.e., the approaches based on RTT, the approaches based on the number of probe responses, and the approaches based on the number of links change, respectively.

RTT-based outage detection approaches, such as [1, 4, 6, 15], have been proposed by utilizing different statistical models to characterize the RTT of all links to detect the network outages. Fontugne et al. [1] first obtained the differential RTT of all links. Then, they leveraged normal distribution to model the measurements and detected the network outages by applying the Wilson score. However, [6] rarely investigated the performance of the last-mile network, and the last-mile network is the centerpiece of broadband connectivity. Hence, Fontugne et al. [1] improved the previous work [6] and captured the RTT of all links in last-mile networks. Then, they used the Welch method to analyze the measurements and detected the network outages.

Since several studies [16–18] reported that normal distribution failed to characterize several distinct modes of the RTT distribution of links, Fontugne et al. [15] leveraged the log-normal distribution to model the RTT of all links and identified all the modes of RTT distribution. Then, they detected the network outages by observing the transitions between the different modes. However, [15] cannot precisely distinguish whether RTT changes are caused by network outage events or "normal" RTT fluctuations. In response to this fact, B. Hou et al. [4] collected the RTT measurements of all links and utilized the change-point detection algorithm twice to detect network outages. Their approach can effectively reduce the false positive rate.

Other outage detection approaches detected network outages by probing all links in the detection area and analyzing the number of probe responses [5, 7, 19, 20]. Heidemann et al. [20] and Dainotti et al. [19] used pings to probe all links in detection areas and detected the network outages by observing the apparent decrease in the number of probe responses. However, these approaches [19, 20] achieved low accuracy of detection. In order to improve this, Quan et al. [5] proposed a detection system named Trinocular. Specifically, they probed all links in detection area to capture the number of probe responses. Then, they used Bayesian inference to analyze these measurements and detected the network outages. However, [5] did not study the effect of weather on last-mile Internet performance, and the performance of last-mile networks affects the network connectivity of a large number of users. Hence, Padmanabhan et al. [7] used ThunderPing [21] to probe all residential links in the detection area and obtained the number of probe responses.

Then, they applied statistics to analyze the measurements and detected the network outages.

Recently, a novel approach [14] has been proposed to detect network outages by monitoring the paths of all links in the detection area and analyzing the number of link changes. The authors used traceroute to obtain the stable state of all links and leveraged the notion of empathy to aggregate the paths that changed similarly over time. Then, they detected the network outages by analyzing the number of link changes.

Note that the existing work detected the network outages by monitoring different performance metrics of all links in the detection area. Although they can detect network outages, they will lead to high resource occupancy and cost in practice. In detail, (1) active probing injects a mass of traffic into the network. Monitoring all the links may occupy the link bandwidth, reduce the network transmission speed, and increase the network burden [8, 9]. (2) Active probing is subject to its scalability. Monitoring all links prompts researchers to deploy more probes, and the deployment and operation of probes (e.g., periodic maintenance, fault analysis, and electricity costs) increase the costs in practice.

To address these challenges, we propose an approach to detect network outages by monitoring the RTT of key links. We first recognize the key links in the detection area in terms of three aspects. Then, we give the critical value of RTT for each key link. Finally, we detect the network outages by observing whether the RTT of the key link exceeds the critical value.

3. Network Outage Detection Approach

In this section, we propose an approach to detect network outages based on key links. The approach is called KL-Dection, which mainly consists of four parts: *data preprocessing*, *key link recognition*, *critical value calculation*, and *detection algorithm*. The architecture is depicted in Figure 1. Next, we describe each part in turn.

3.1. Data Processing. In order to monitor the network state and detect the network outages of the detection area, we need to obtain the performance measurements of the links in the detection area, i.e., the RTT of links. Hence, this paper obtains the performance measurements from two public datasets, i.e., RIPE Atlas Dataset [22] and Maxmind GeoIP City Dataset [23]. Specifically, we collect the traceroutes from RIPE Atlas and map each hop (node) in traceroutes to the geographic location using GeoIP City Dataset. For a certain detection area D, we obtain their corresponding traceroutes, denoted as dataset A. For each traceroute in dataset A, we extract the links formed by every adjacent node and focus on the RTT of each link.

3.2. Key Link Recognition. In this section, we propose a key link recognition algorithm to recognize the key links in the detection area. Previous work of traffic monitoring provided initial inroads to recognize the key links [10, 11]. These work defined the key link as the link delivering the larger traffic flows in the detection area. However, based on a basic observation, we find that monitoring the links that deliver larger traffic flows is inadequate to detect the network outages (see Section 4). This is because detecting the network outages also needs to monitor the link performance, e.g., RTT [4, 12].

In fact, the RTT of the link is affected by multiple aspects, including the neighbor correlation [4] and the presence of noise [2]. Specifically, the RTT of the links with poorer connected relation (fewer adjacencies) is less affected by the RTT of other links [4]. This causes that the RTT of the links with poor connected relation may not fluctuate obviously in the case of a network outage [24]. Hence, monitoring their RTT may fail to detect the network outages even though they deliver larger traffic flows. Moreover, the accuracy of the RTT is influenced by the distance between probes and the nodes of links. Generally, the links are closer to the probes, and the RTT of the link can be measured more accurately [14]. Hence, we may fail to detect the network outages by monitoring the RTT of the links far from the probes, even though these links deliver larger traffic flows.

In response to this fact, we recognize the key links by considering the traffic load information and the factors affecting the RTT of the link, consisting of three metrics, i.e., flow density, degree centrality, and probe-distance centrality. The flow density describes the number of flows going through the links in unit time. The degree centrality describes the connected relation. The probe-distance centrality describes the distance (hop) between the node of the link and probes. Next, we give the definitions of these three metrics and describe the process of key link recognition.

3.2.1. The Flow Density. The flow density represents the number of flows going through the links in unit time. We conduct the measurement in the detection area with short time intervals and over long timescales (days to weeks). The flow density of links at different time intervals is represented by a matrix $M \in R^{t*l}$ which is given as

$$M = \begin{pmatrix} f_{11} & \cdots & & f_{1l} & \\ \vdots & & & & \vdots \\ f_{i1} & \cdots & f_{ij} & \cdots & f_{il} \\ \vdots & & \vdots & & \\ f_{t1} & \cdots & f_{tl} & & \end{pmatrix}, \tag{1}$$

where f_{ij} denotes the flow density of j-th link during the time interval i, t denotes the number of consecutive time intervals (the number of rows), l denotes the number of total links in the network (the number of columns), and $t>>l$.

Next, we perform the singular value decomposition (SVD) [25] on M and illustrate how SVD can recognize a small set of links that can well represent the flow density of all links in the detection area. The decomposition of matrix M is given as

$$M = U\sum V^T, \tag{2}$$

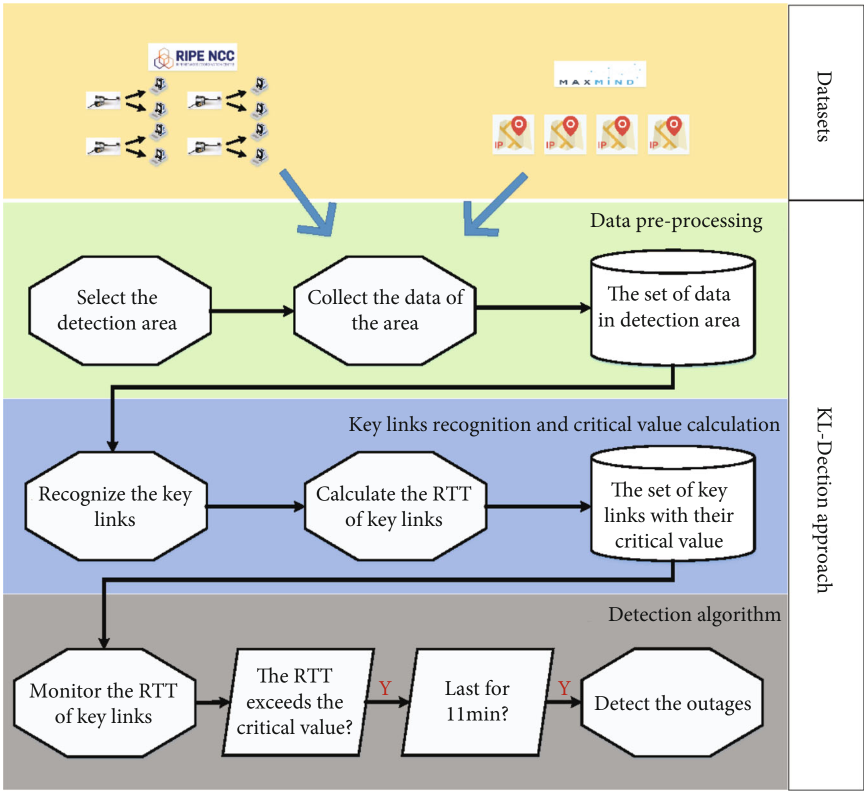

RIPE NCC
MAXMIND
Datasets
Data pre-processing
Select the detection area
Collect the data of the area
The set of data in detection area
Key links recognition and critical value calculation
Recognize the key links
Calculate the RTT of key links
The set of key links with their critical value
KL-Dection approach
Detection algorithm
Monitor the RTT of key links
The RTT exceeds the critical value?
Y
Last for 11min?
Y
Detect the outages

Figure 1: The architecture of KL-Dection.

where $U \in R^{t*t}$ and $U^T U = I$, $V \in R^{l*l}$ and $V^T V = I$, and $\sum$ is a $t \times l$ diagonal matrix, whose diagonal entry $\delta_i (\delta_i \geq 0)$ is known as the singular value and represents the importance of the link in the matrix M. The columns of U are denoted as $\{u_i | i \in \{1, \cdots, t\}\}$, and the columns of V are denoted as $\{v_j | j \in \{1, \cdots, l\}\}$. Hence, the matrix M can be calculated as

$$M = \delta_1 u_1 v_1^T + \delta_2 u_2 v_2^T + \cdots + \delta_l u_l v_l^T, \quad (3)$$

Suppose there are q positive singular values among all singular values δ_i; this means that we can find q columns in M to represent itself according to the crucial property of SVD [25]. Moreover, existing work [10] demonstrated that the singular value of flow density matrix M is sparsely distributed, and there are only r large singular values among q positive singular values ($r << q$). The r is called the effective rank of the matrix M, which means that the flow density of all links can be approximately represented by the flow density of r basic links in the detection area.

According to the analysis above, we first obtain the flow density matrix M of the detection area; then, we perform the SVD on matrix M; finally, we extract r basic links based on the singular values. For convenience, the set of these r basic links is denoted as B.

3.2.2. The Degree Centrality. The degree centrality of the node represents the number of adjacencies of the node. In view of this definition, we extend it to describe the degree centrality of the link. According to the bucket effect, the centrality of the link is constrained by the minimum of the centrality of its two nodes. Hence, we define the degree centrality of the link as the minimum of the degree centrality of its two nodes. Noting that the network is in a stable state during a period of time, we consider that the centrality of the link will not change over time. For convenience, the link formed by two adjacent nodes v_i and v_j is denoted as $l_{i,j}$. The degree centrality of the link $l_{i,j}$ (i.e., $d_{i,j}$) is calculated as follows:

$$d_{i,j} = \min \{d(v_i), d(v_j)\}, \quad (4)$$

$$d(v_i) = \{N_u | u \in V, (u, v_i) \in E\}, \quad (5)$$

where $d(v_i)$ is the degree centrality of the node v_i, V represents the set of nodes, E represents the set of links, and N_u represents the number of nodes that v_i connected with. Note that the larger value of $d_{i,j}$ represents that the link $l_{i,j}$ has richer connected relation, and its RTT can reflect the network state notably.

3.2.3. The Probe-Distance Density. In order to describe the distance between the probes and the node of the link, we propose a metric called probe-distance centrality. Since the distance is obtained by calculating the number of hops between two nodes, we calculate the probe-distance centrality of the link $l_{i,j}$ based on the following steps: (1) we extract the two nodes of $l_{i,j}$, and for each node, we calculate the average hops between it and all the probes in detection area;

(2) similar to the definition of the degree centrality of link, we select the minimum average hops as the probe-distance centrality of the link $l_{i,j}$.

The probe-distance centrality of the link $l_{i,j}$ (i.e., $p_{i,j}$) is formulated as

$$p_{i,j} = \min \left\{ p(v_i), p(v_j) \right\}, \tag{6}$$

$$p(v_i) = \frac{\sum_{k=1}^{K} \sigma_{(v_i,k)}}{K}, \tag{7}$$

where $p(v_i)$ represents the average hops between node v_i and all probes in the detection area, K represents the number of probes in the detection area, and $\sigma_{(v_i,k)}$ represents the number of hops between the node v_i and the k-th probe. Note that the lower value of $p_{i,j}$ indicates that the link is closer to all the probes, and its RTT can reflect the network state accurately.

3.2.4. Key Link. Based on the definitions of these three metrics, we define key link as follows. For each link $l_{i,j}$ in D, the set of key links K is defined as

$$K = \left\{ l_{i,j} \middle| P\left(l_{i,j} \subset C, l_{i,j} \subset E \middle| l_{i,j} \subset B \right) = 1 \right\}, \tag{8}$$

where C represents the set of links with $d_{i,j} \geq \Delta_1$ and E represents the set of links with $p_{i,j} \leq \Delta_2$. Δ_1 and Δ_2 are the critical value of the degree centrality and probe-distance centrality, respectively. As can be seen from Equation (8), if a link $l_{i,j}$ is a basic link ($l_{i,j} \subset B$) and it meets the conditions of $d_{i,j} \geq \Delta_1$ and $p_{i,j} \leq \Delta_2$, it can be regarded as the key link.

Note that the process of key link recognition is not limited by network topology and detection area. For convenience, we take Colorado as an example to illustrate how to obtain B, C, and E in turn.

First, we extract the traceroutes from July 31 to August 30, 2020, in Colorado. Then, based on the traceroutes, we obtain the flow density matrix in this area, a 11424×556 matrix, consisting of the flow density of 556 links during 11424 time intervals. Finally, we get 556 singular values and sort them in a descending order manner. We note that the *30*-th singular value is already close to zero. Therefore, we only present the first 30 singular values of the matrix in Figure 2(a).

From Figure 2(a), we can observe that the singular value decreases rapidly from the first element to the *17*-th element, and after the *17*-th element, the singular value decreases slowly and eventually stabilizes. We note that the singular value becomes very small and is almost closed to zero after the *17*-th element. Considering the fact is that the larger singular value, the more important the link is; hence, we conclude that the effective rank of the matrix is 17, which means that only 17 basic links are enough to represent the traffic load information of Colorado. Next, we use the QR factorization with column pivoting [25] to obtain these 17 basic links, which constitute the set B.

Then, based on Equation (4) and Equation (5), we give the distribution of the degree centrality of the links in Colorado. Figure 2(b) reveals that the percentage increases rapidly when the degree centrality is below 4; then, the percentage increases slowly between the degree centrality is 5 and 8; finally, the percentage stabilizes when the degree centrality is above 9. From Figure 2(b), we note that only about 10% of the links have a higher degree centrality (≥8). As mentioned before, during the network outages, the RTT of links with high degree centrality may fluctuate substantially due to the influence of neighbor correlation, which can reflect the network state notably. Monitoring the RTT of these links can help the network managers to detect network outages. Therefore, we define the set of links with $d_{i,j} \geq 8$ as C.

In addition, we also give the distribution of probe-distance centrality for all links according to Equation (6) and Equation (7). In Figure 2(c), the value of probe-distance centrality is divided into six bins. We note that a large share of links has the probe-distance centrality above 7 (first two bins). On the contrary, the links with lower probe-distance centrality (≤7) only account for 26.52%, which indicates that they are closer to the probes in Colorado. As mentioned before, the RTT of the links with lower probe-distance centrality can reflect the network state accurately. Monitoring the RTT of these links can help the network managers to detect network outages. As a result, we define the set of links with $p_{i,j} \leq 7$ as E.

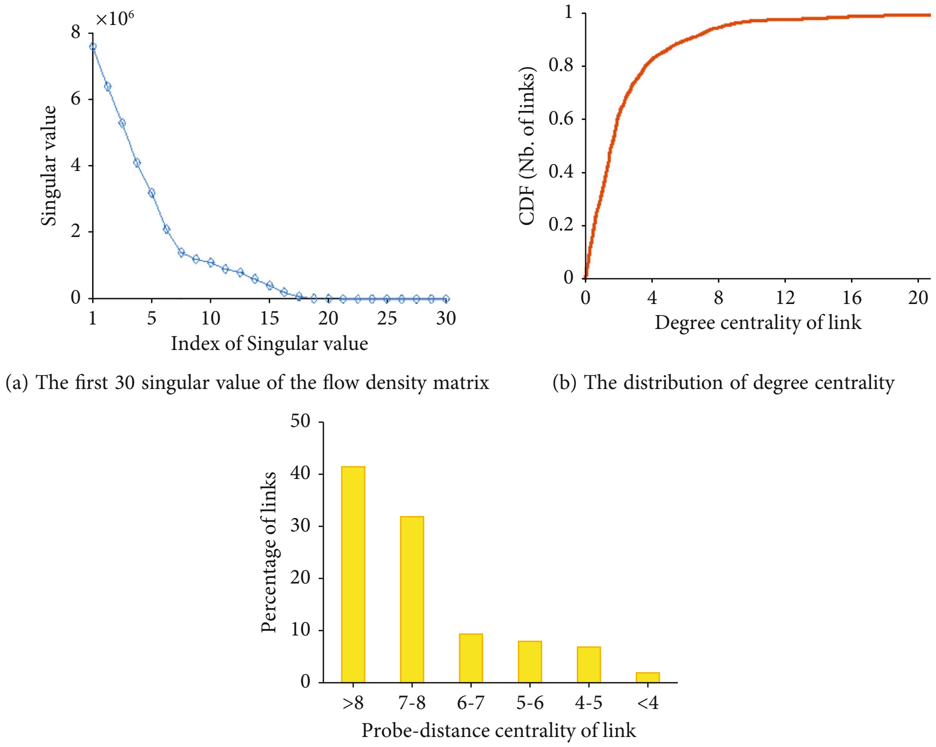
In conclusion, for each link $l_{i,j}$ in Colorado, if it satisfies the conditions,

$$K = \left\{ l_{i,j} \middle| P\left(p_{i,j} \leq 7, d_{i,j} \geq 8 \middle| l_{i,j} \subset B \right) = 1 \right\}, \tag{9}$$

it can be regarded as the key link. The process of the key link recognition is summarized in Algorithm 1. Specifically, we first select the detection area D and obtain the flow density matrix M of D. Then, we apply SVD on M and acquire r basic links. The set of these basic links are denoted as B. For all the links in B, we extract the links belonging to sets C and E. The results are the set of key links K.

3.3. Critical Value Calculation. RTT is demonstrated as a key metric to gain insights into the performance of links [4, 7]. Moreover, the critical value of RTT can distinguish whether network outages occur [24]. Hence, we propose a critical value calculation algorithm based on interquartile and give the critical value of RTT for the recognized key links. Specifically, we first extract the raw RTT of each key link over a period of time and sort them in an ascending order manner. Then, for each key link $l_{i,j}$, we define $R_{uq}(l_{i,j})$ as its upper quartile (the value located at 75% of the data range), $R_{lq}(l_{i,j}$) as its lower quartile (the value located at 25% of the data range), and $R_d(l_{i,j})$ as the difference between the upper and lower quartiles. The critical value $R_{cv}(l_{i,j})$ is calculated as follows:

$$R_{cv}(l_{i,j}) = R_{uq}(l_{i,j}) + kR_d(l_{i,j}), \tag{10}$$



(a) The first 30 singular value of the flow density matrix

(b) The distribution of degree centrality

(c) The distribution of probe-distance centrality

Figure 2: The distribution of three metrics.

```
Input: $D$, $r$, $\Delta_1$, $\Delta_2$;
Output: $K$;
1: Calculate the matrix $M$ in $D$ and apply SVD over it,
2: Acquire $r$ basic links, and denote them as the set $B$,
3: According to $\Delta_1$, $\Delta_2$, extract the links that satisfy the conditions in $B$, and denote them as $K$,
4: return The set of key links $K$.
```

Algorithm 1: The recognition of key links.

where k is the regular factor. Finally, according to [26], we set k as 1.5 and obtain the critical value of RTT for each key link.

3.4. KL-Dection Algorithm. Note that network outages are different from network congestion since they will persistently influence the state of the network [27–29]. Therefore, we add a constraint of duration to the definition of network outages. Existing work [5, 7] defined network outages as no response or missing a set of pings from any vantage point in 11 minutes. As a reference, in this paper, we define the network outage as the phenomenon that the RTT of any key link exceeds its critical value and lasts for more than 11 minutes.

Next, we summarize the process of the KL-Dection approach and give its pseudocode in Algorithm 2. Firstly, we select the detection area D and obtain their corresponding traceroutes, denoted as dataset A. Then, we recognize the key links from A using Algorithm 1. Next, we calculate the critical value $R_{cv}(l_{i,j})$ of each key link $l_{i,j}$, respectively. Finally, we detect the network outages by observing whether the RTT of any key link exceeds the critical value for more than 11 minutes. It is worth noting that the KL-Dection algorithm can be used in any network topology and detection area.

4. Results and Discussion

Although our approach is applicable for the network outage detection in any detection area, due to space limitations, we take California as the detection area for presentation. We first present the visualization results of key links in this area. Then, we leverage one outage event in California to demonstrate the validity of the definition of key link. Next, we leverage another outage event to demonstrate the validity of the KL-Dection approach. All outage events we considered occurred in the Internet. Finally, we compare the existing approach [14] and our approach in terms of the number of monitoring links in three detection areas.

```
Input: $D$, $r$, $\Delta_1$, $\Delta_2$;
Output: Network outage;
Extract the traceroutes in $D$, and denote them as dataset $A$,
Recognize the key links from dataset $A$ using Algorithm 1,
Obtain $R_{cv}(l_{a,b})$ for each key link $l_{a,b}$,
Monitor the RTT of $l_{a,b}$ in parallel,
if The RTT of any $l_{a,b}$ exceeds its corresponding critical value $R_{cv}(l_{a,b})$ and lasts for more than 11 minutes then
    return Network outage,
end if
```

Algorithm 2: KL-Dection algorithm.

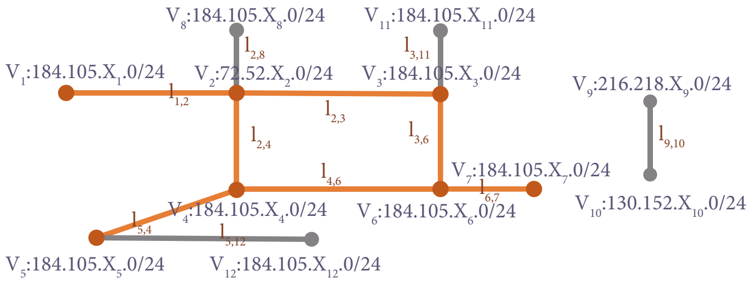


Figure 3: A part of visualization results of links. The topology of links is consistent with the practical environment. The orange lines represent the key links, which are used to detect two outage events. The gray lines represent the links, which are used to detect the outage event occurred on October 21, 2016.

Table 1: The critical value of RTT.

Key link	$l_{1,2}$	$l_{2,3}$	$l_{2,4}$	$l_{5,4}$	$l_{4,6}$	$l_{3,6}$	$l_{6,7}$
Critical value (ms)	26.0	81.9	25.4	19.3	27.6	21.5	18.0

4.1. The Visualization Results of Key Links. We first collect the traceroutes going through California from July 31 to November 30, 2020, in RIPE, corresponding to 2.7 T data. Based on the data, we analyze 11936 links and leverage Algorithm 1 to recognize seven key links. The visualization results of these seven key links are shown in Figure 3. To protect the privacy of address information, we describe the nodes of key links in the form of prefixes.

Based on the key links, we extract their raw RTT from the 2.7 T data and obtain their corresponding critical value. The critical value (ms) is shown in Table 1.

From Table 1, we can infer that the network outage is happening if the RTT of any key link exceeds its critical value and lasts for 11 minutes.

4.2. The Validity of the Definition of Key Link. After obtaining the critical value, we consider an outage event that occurred on October 21, 2016 [30], in California to evaluate the validity of the definition of key link. In this case study, we extract the RTT from 2016-10-21 11 : 00 UTC to 17 : 30 UTC for analysis.

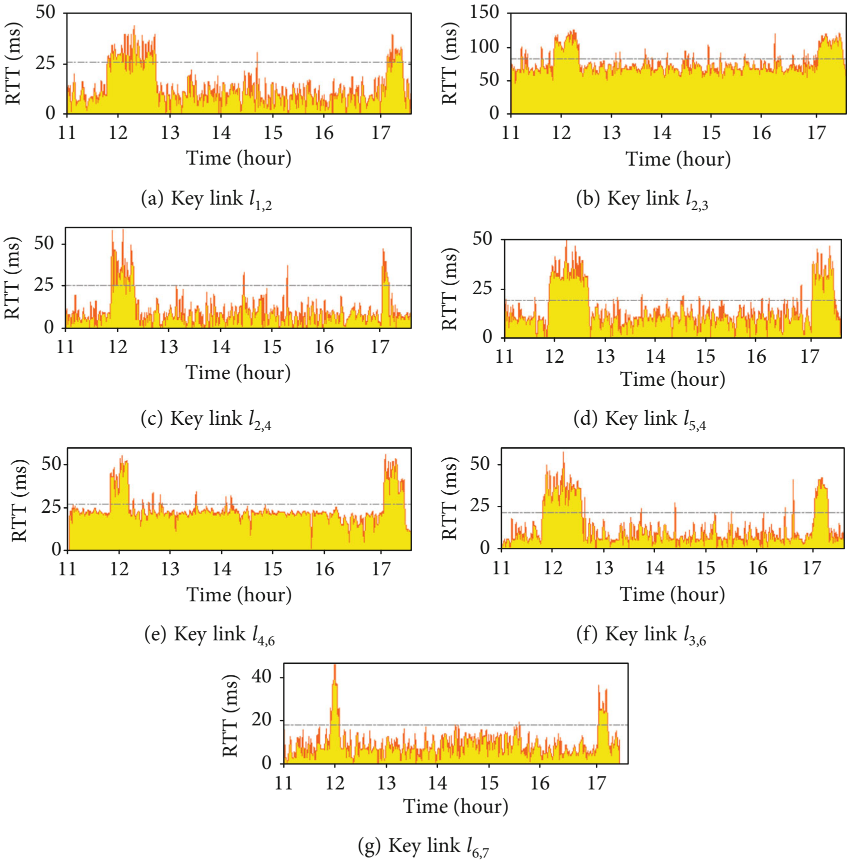
The RTT of each key link during 2016-10-21 11 : 00 UTC to 17 : 30 UTC is shown in Figure 4. In each figure, the x-axis is the time (in hours), the y-axis is the RTT (ms), and the dotted line is the critical value of the RTT. We take Figure 4(a) as an example to illustrate. In Figure 4(a), the RTT is below the critical value before 11 : 53; then it increases rapidly and exceeds the critical value during 11 : 53 to 12 : 45; afterward, the RTT is gradually back to normal during 12 : 45 to 17 : 10; next, the RTT increases again and exceeds the critical value during 17 : 10 to 17 : 30. From Figure 4(a), we can observe that the outage occurred from 11 : 53 to 12 : 45 and 17 : 10 to 17 : 30, respectively, which is in good agreement with the time reported in [31].

As can be seen from Figure 4, we can detect the network outages effectively by monitoring the RTT of any key link. It is worth noting that the time consumption of our approach is low because it monitors the state of all key links in parallel.

Next, in order to evaluate the validity of the definition of key link, we randomly select four links (except the key links), which represent different types of links in California. We present the three metrics of these four links in Table 2.

As can be seen from Table 2, only one metric of the first three links does not satisfy Equation (8), and the last link only satisfies the condition of flow density. Then, we leverage these four links to detect the outage event mentioned above, and the RTT of these four links is given in Figure 5. It can be seen from Figure 5 that the RTT of these four links has different fluctuations, but they do not exceed the critical value. Hence, we conclude that no network outage occurs, which is inconsistent with the ground truth. The results demonstrate that focusing on the links recognized by any one or two metrics alone is inadequate to detect network outages. Moreover, based on the comparison of the results between Figures 4 and 5, we can conclude that the definition of the key links proposed in this paper is effective in network outage detection.

4.3. The Validity of KL-Dection Approach. In order to evaluate the validity of KL-Dection approach, we consider an



(a) Key link $l_{1,2}$

(b) Key link $l_{2,3}$

(c) Key link $l_{2,4}$

(d) Key link $l_{5,4}$

(e) Key link $l_{4,6}$

(f) Key link $l_{3,6}$

(g) Key link $l_{6,7}$

FIGURE 4: RTT of the key links.

TABLE 2: The three metrics of links.

Link	$l_{i,j} \subset B$	$l_{i,j} \subset C$	$l_{i,j} \subset E$
Link $l_{2,8}$	✓	☒	✓
Link $l_{9,10}$	✓	✓	☒
Link $l_{3,11}$	☒	✓	✓
Link $l_{5,12}$	✓	☒	☒

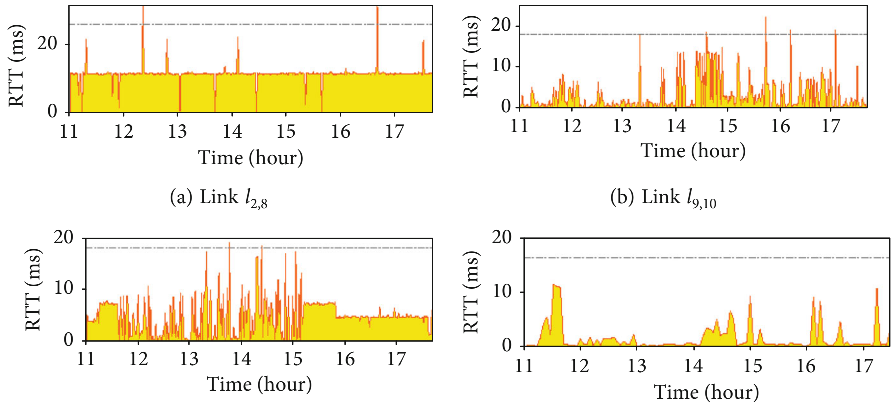
outage event that occurred on May 24, 2019, in California [32]. Since the outage event lasted from 21 : 47 to 23 : 58, we extract the RTT from 2019-05-24 20 : 00 UTC to 24 : 00 UTC for analysis.

From Figure 6, we can observe that the key links $l_{2,3}$, $l_{5,4}$, $l_{3,6}$, and $l_{6,7}$ can detect the outage effectively. In detail, we can detect the outage from Figure 6(d) because the RTT exceeds the critical value and lasts from 23 : 30 to 23 : 50. Similarly, we can also detect the outage from Figure 6(g) because the RTT exceeds the critical value and lasts from 21 : 47 to 23 : 58. However, the duration of the outage event inferred from these two key links are different. This phenomenon can be explained by the fact that because part or parts of the power grid remain operational, the links in some areas of California still maintain the normal network state. This phenomenon is verified in electric disturbance events' annual summaries [32].

In addition, we found that the RTT of key links $l_{1,2}$, $l_{2,4}$, and $l_{4,6}$ is stable over time. This phenomenon can be explained by the fact that the outage event occurred far away from the location of these key links, and it did not affect the performance of these key links. The results of Figure 6 demonstrate that our approach can detect the outage event effectively by observing whether the RTT of any key link exceeds the critical value for more than 11 minutes.

4.4. Performance Comparison. In this section, we aim to compare the KL-Dection approach with the existing approaches in terms of the number of monitoring links when both approaches can detect the network outages successfully. Consider that the latest approach [14] detected the network outage by collecting the traceroute and monitoring the performance of all links in the detection area, which is similar to the dataset and detection mode adopted in this paper. As a consequence, we compare our approach with the latest approach [14] in terms of the number of monitoring links in three detection areas. Specifically, our approach and the existing work can successfully detect the outage event that occurred in these three detection areas [30], and the results of the comparison are shown in Table 3.

As can be seen from Table 3, the number of monitoring links our approach needed is notably smaller than the existing work [14]. This is especially true when the number of links in the detection area is large. Specifically, in California, our approach only needs to monitor less than 0.06% of the links



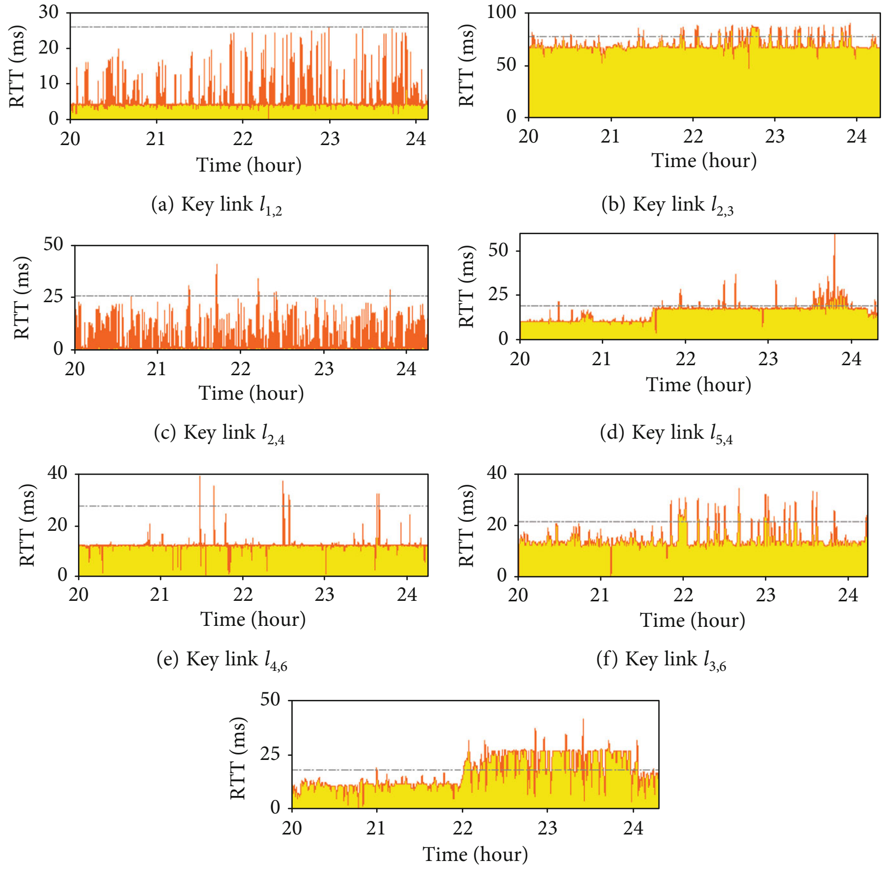
(a) Link $l_{2,8}$

(b) Link $l_{9,10}$

(c) Link $l_{3,11}$

(d) Link $l_{5,12}$

Figure 5: RTT of the links.



(a) Key link $l_{1,2}$

(b) Key link $l_{2,3}$

(c) Key link $l_{2,4}$

(d) Key link $l_{5,4}$

(e) Key link $l_{4,6}$

(f) Key link $l_{3,6}$

(g) Key link $l_{6,7}$

Figure 6: RTT of the key links.

Table 3: Existing approach [14] and KL-Dection approach all can detect the outage event that occurred on October 21, 2016 [30]. Under this case, we give the results of the comparison of the number of monitoring links in three detection areas.

Detection area	KL-Dection (ours)	Existing approach [14]
Colorado	5	5234
District of Columbia	6	7420
California	7	11936

for network outage detection, and the runtime of our approach is 2 seconds.

Monitoring a large number of links will prompt the researchers to deploy more probes and take continuous measurements, which may occupy the link bandwidth, reduce the network transmission speed, and increase the network burden [9]. Besides this, the operation of probes increases the costs (e.g., periodic maintenance, fault analysis, and electricity costs). Thus, the KL-Dection approach can obviously reduce resource occupancy and cost without compromising the validity of outage detection. We believe that our approach can provide better scalability and is more acceptable in practice than existing work.

5. Discussion

The results presented in this paper have several implications for the networking community. Because our approach is lightweight and effective, the network managers can leverage our approach to understand the network performance of their customers with less cost. Similarly, in the scenario of the Internet of Things, managers can also effectively understand the performance of the network by monitoring the connectivity of links that connect the key devices.

However, several limitations should be considered when leveraging our approach. First, the key links are recognized from vantage points, but the vantage points may not be representative of the detection area, especially when the number of Atlas probes is low. Hence, the results are prone to the bias of Atlas deployment. Second, in the scenario of the data center, the result of our approach is not satisfactory. This is because compared with edge network, the network topology of the data center is small, and a majority of links in data center have high flow density, degree centrality, and probe-distance centrality. Hence, our approach does not work well in this scenario. The solutions to these limitations are left as a future work.

6. Conclusion

In this paper, we propose KL-Dection approach, which detects network outages via key links instead of all links. Specifically, we recognize the key links in terms of three metrics, including flow density, degree centrality, and probe-distance centrality of links. Then, based on recognized key links, we give a critical value calculation algorithm on RTT that distinguishes whether network outages occur. Finally, we leverage two historical events to demonstrate that our approach can detect the network outages effectively.

Data Availability

The data included in this paper are available without any restriction.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Key Research and Development Program of China under Grant No. 2020YFE0200500.

References

[1] R. Fontugne, A. Shah, and K. Cho, “Persistent last-mile congestion:not so uncommon,” in *Proceedings of the Internet Measurement Conference (IMC)*, pp. 420–427, New York, 2020.

[2] J. Kučera, R. B. Basat, M. Kuka, G. Antichi, M. Yu, and M. Mitzenmacher, “Detecting routing loops in the data plane,” in *Proceedings of the 2020 ACM CoNEXT Conference*, pp. 466–473, Barcelona, 2020.

[3] G. Kumar, N. Dukkipati, K. Jang et al., “Swift: delay is simple and effective for congestion control in the datacenter,” in *Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM)*, pp. 514–528, New York, 2020.

[4] B. Hou, C. Hou, T. Zhou, Z. Cai, and F. Liu, “Detection and characterization of network anomalies in large-scale RTT time series,” *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 793–806, 2021.

[5] L. Quan, J. Heidemann, and Y. Pradkin, “Trinocular: understanding internet reliability through adaptive probing,” in *Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM)*, pp. 255–266, Hong Kong, 2013.

[6] R. Fontugne, C. Pelsser, E. Aben, and R. Bush, “Pinpointing delay and forwarding anomalies using large-scale traceroute measurements,” in *Proceedings of the Internet Measurement Conference (IMC)*, pp. 15–28, London, 2017.

[7] R. Padmanabhan, A. Schulman, D. Levin, and N. Spring, “Residential links under the weather,” in *Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM)*, pp. 145–158, Beijing, 2019.

[8] N. Gaur, A. Chakraborty, and B. S. Manoj, “Delay optimized small-world networks,” *IEEE Communications Letters*, vol. 18, no. 11, pp. 1939–1942, 2014.

[9] M. Hasib and J. A. Schormans, “Limitations of passive & active measurement methods in packet networks,” in *London Communi- cations Symposium (LCS)*, London, UK, 2003.

[10] J. Badshah, M. Alhaisoni, N. Shah, and M. Kamran, “Cache servers placement based on important switches for SDN-based ICN,” *Electronics*, vol. 9, no. 1, pp. 39–65, 2020.

[11] T. Yingying Cheng and X. Jia, “Compressive traffic monitoring in hybrid SDN,” *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 12, pp. 2731–2743, 2018.

[12] D. Perdices, D. Muelas, I. Prieto, L. de Pedro, and L. de Vergara, “On the modeling of multi-point RTT passive measurements for network delay monitoring,” *IEEE Transactions on*

Network and Service Management, vol. 16, no. 3, pp. 1157–1169, 2019.

[13] M. S. Kang, S. B. Lee, and V. D. Gligor, "The crossfire attack," in *2013 IEEE Symposium on Security and Privacy*, pp. 127–141, Berkeley, CA, 2013.

[14] M. Di Bartolomeo, V. Di Donato, M. Pizzonia, C. Squarcella, and M. Rimondini, "Extracting routing events from traceroutes: a matter of empathy," *IEEE/ACM Transactions on Networking*, vol. 27, no. 3, pp. 1000–1012, 2019.

[15] R. Fontugne, J. Mazel, and K. Fukuda, "An empirical mixture model for large-scale RTT measurements," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, pp. 2470–2478, Hong Kong, China, 2015.

[16] B.-Y. Choi, S. Moon, Z.-L. Zhang, K. Papagiannaki, and C. Diot, "Analysis of point-to-point packet delay in an operational network," *Computer Networks*, vol. 51, no. 13, pp. 3812–3827, 2007.

[17] S. Shakkottai, N. Brownlee, A. Broido, and K. Claffy, "The RTT distribution of TCP flows on the internet and its impact on TCP based flow control," Technical report, Cooperative Association for Internet Data Analysis (CAIDA), 2004.

[18] G. Maier, A. Feldmann, V. Paxson, and M. Allman, "On dominant characteristics of residential broadband internet traffic," in *Proceedings of the Internet Measurement Conference (IMC)*, pp. 90–102, Chicago, 2009.

[19] A. Dainotti, C. Squarcella, E. Aben et al., "Analysis of country-wide internet outages caused by censorship," *IEEE/ACM Transactions on Networking*, vol. 22, no. 6, pp. 1964–1977, 2014.

[20] J. Heidemann, L. Quan, and Y. Pradkin, "A preliminary analysis of network outages during hurricane sandy," Tech. Rep, ISI-TR-685b, University of Southern California, Information Sciences Institute, 2012.

[21] A. Schulman and N. Spring, "Pingin' in the rain," in *Proceedings of the Internet Measurement Conference (IMC)*, pp. 19–28, New York, 2011.

[22] A. Milolidakis, R. Fontugne, and X. Dimitropoulos, "Detecting network disruptions at colocation facilities," in *2019 IEEE Conference on Com- puter Communications (INFOCOM)*, pp. 2161–2169, Paris, France, 2019.

[23] G. Aceto and A. Pescapé, "Internet censorship detection: a survey," *Com- puter Networks*, vol. 83, pp. 381–421, 2015.

[24] J. H. Wang and C. An, "A study on geographic properties of internet routing," *Computer Networks*, vol. 133, pp. 183–194, 2018.

[25] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3, The Johns Hopkins Univ. Press, Baltimore, MD, USA, 2012.

[26] K. L. Spafford, J. S. Meredith, and J. S. Vetter, "Quartile and outlier detection on heterogeneous clusters using distributed radix sort," in *2011 IEEE International Conference on Cluster Computing*, pp. 412–419, Austin, TX, USA, 2011.

[27] R. Zhao, Z. Li, Z. Xue, T. Ohtsuki, and G. Gui, "A novel approach based on lightweight deep neural network for network intrusion detection," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, Nanjing, China, 2021.

[28] P. Thorat, N. K. Dubey, K. Khetan, and R. Challa, "SDN-based predictive alarm manager for security attacks detection at the IoT gateways," in *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*, pp. 1-2, Las Vegas, NV, USA, 2021.

[29] N. Leslie, "An unsupervised learning approach for in-vehicle network intrusion detection," in *2021 55th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–4, Baltimore, MD, USA, 2021.

[30] G. C. M. Moura, J. Heidemann, M. Müller, R. O. de Schmidt, and M. Davids, "When the dike breaks: dissecting DNS defenses during DDoS," in *Proceedings of the Internet Measurement Conference (IMC)*, pp. 8–21, Boston, MA, USA, 2018.

[31] S. Mansfield-Devine, "DoS goes mainstream: how headline-grabbing attacks could make this threat an organisation's biggest nightmare," *Network Security*, vol. 2016, no. 11, pp. 7–13, 2016.

[32] "Electric disturbance events (oe-417) annual summaries," https://www.oe. http://netl.doe.gov//OE417annualsummary.aspx.