Association Mining Analysis of Alarm Root-Causes in Power System with Topological Constraints

Chen Su
National Computer Network
Emergency Response Technical
Team/Coordination Center of
China, Beijing, P. R. China
(86)+13581540057
chens@cert.org.cn

Zhu Hailong
National Computer Network
Emergency Response Technical
Team/Coordination Center of
China, Beijing, P. R. China
(86)+13911900439
zhl@cert.org.cn

Xu Junbiao
National Computer Network
Emergency Response Technical
Team/Coordination Center of
China, Beijing, P. R. China
(86)+13810685840
xjb2015mailbox@163.com

ABSTRACT

Power supply system is extremely important in data center since it is the basis for the operation of infrastructure such as servers, switches, etc.. Power supply alarm management is essential because alarm flooding always occurs when a power supply device is cut off or other equipment is damaged, causing the high pressure of dealing with a large quantity of burst data and the risk of alarm omission. In this paper, we propose a novel alarm root-cause association analysis method based on actual topological constraints, for alarm diagnosis in power supply system. First, a new alarm clustering algorithm, namely DTIBFS, is proposed to cluster the alarm nodes and figure out the root-cause node. In this way, the system can handle over 90% of alarm records by dealing with only 4 alarm clusters in the alarm flooding period, or handle over 60% alarm records by dealing with only 20 alarm clusters on the average in a long time period, both of which contribute to a remarkable improvement of efficiency and reduce of operation workload compared with the one-by-one alarm record addressing. Furthermore, an improved FP-growth association analysis based on DTIBFS above is introduced. Experiments on the actual alarm records of the data center indicate that a multitude of meaningful rules can be obtained. Due to the consideration of supply system topology, our method can mine the associations from both statistical point and topology point, which is helpful to detect whether there are missing alarms during the alarm floods or provide tips for alarm diagnosis in data center operation.

CCS Concepts

• Information systems → Enterprise information systems • Information systems →Association rules

Keywords

Power supply system; Root-cause alarm; Topological constraints; Association analysis; FP-growth

ICIT 2017, December 27–29, 2017, Singapore, Singapore
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-6351-8/17/12…$15.00
DOI: https://doi.org/10.1145/3176653.3176710

1. INTRODUCTION

Power supply system is an extremely important system in data center[1], which is essential for the operation of infrastructure such as servers, switches, routers, etc.. Therefore the power supply alarm management becomes very important in data center operation[2][3]. When a main power device is cut off, alarm flooding always occurs consequently, which results in the high pressure of dealing with a large quantity of burst data and the risk of alarm omission. Hence, it is necessary to locate the root cause of the massive flooding alarms as well as mine the association rules among device alarms[4][5].

Association rule mining is to analyze interesting associated relationships between records[6], which is also known as Market Basket Transaction (MBA) analysis[7]. Many Big-Data based analysis methods have been widely applied for this realm, among which Sequential Pattern Mining method, which deals with data represented as sequences[8], is a common one. There are numerous researching algorithms for Sequential Pattern Mining, but the most two popular algorithms are Apriori-like algorithm[9] and FP-growth-like algorithm[10], proposed by Agrawal[in 1994] and Han[in 2000]. The Apriori-like algorithm is realized based on an anti-monotone heuristic, while the FP-growth-like algorithm is realized based on a frequent pattern tree (FP-tree) structure. The procedure of Apriori algorithm and FP-growth algorithm are detailed in Reference [9] and [10], respectively.

The research on Sequential Pattern Mining is heated and is widely applied in Intrusion Detection System[11][12], Telecommunication Network alarm analysis[4][5], etc. For the temporal records like alarm records, the typical way can be briefly summarized that a series of alarm transactions are preprocessed by setting an appropriate sliding-time-window[13-15] and then be used for association analysis based on Apriori algorithm or FP-growth algorithm, the most two popular sequential pattern mining algorithms introduced above.

However, those methods above do not take the power system network topology into consideration, which contains a wealth of information useful for alarm diagnosis. In this work, we propose a novel root-cause alarm association analysis method with actual topological constraints in power supply system. On one hand, a network-structure-aware clustering is realized. As a result, the root devices of alarms can be located, which will help figure out the root cause of failure in power supply management and furthermore significantly reduce the operation workload by uncoupling the alarm records. On the other hand, unlike previous alarm transaction generation by sliding-time-window segmentation, we adopt a new strategy based on above mentioned clustering to mine the association rules of alarms. By considering both temporal correlation and faulty device position in network topology, this paper shows a more comprehensive description on the power supply status. Therefore, for the first time, this paper takes the topological constraints into consideration for association analysis of alarm root-causes, which is different from that by sliding-time-window above[4-5][13-15].

The rest of this paper is organized as follows. In section 2, the scenario of power system topology model is described. In section 3, our DTIBFS clustering algorithm is introduced to generate the alarm clusters and the root-cause node of each cluster is figured out. Then in section 4, a topology-constrained association analysis method based on FP-growth is proposed to analyze the association rules. Numerical results are shown in section 5. Finally, in section 6, the conclusions are drawn.

2. SYSTEM SCENARIO

For convenience, the relevant terminologies are listed below:

a) Alarm set: The whole alarm record sets of power device, where each record contains the fields of alarm ID, alarm content, etc..

b) The topology node: A topology node represents a power supply device in power topology, where each node contains the fields of node (device) ID, node name, upstream node, and downstream node, etc.. (shown in Table 2).

c) Alarms clusters: An alarm cluster is a series of alarm records that are clustered by specific rules defined by the algorithm.

d) Root-cause: A root-cause node is the root node in the topology that results to a series of alarm records occurring in the same time.

Before association analysis, it is necessary to extract alarm transactions out of the alarm records in advance. This will have significant influences directly on the final performance. Instead of employing sliding-time-window model as mentioned above, we take network topology into consideration and generate a series of alarm clusters for alarm devices. It will be applied for the establishment of alarm transactions. This section describes the modeling of power system topology, which will be used for the association analysis algorithm introduced afterwards.

As is known, the supply system can be treated as network and each device is a network node. Each upstream backbone supply device node can mount several downstream supply device nodes and grow layer by layer until reaching the terminal loads. To meet the requirement of reliability and robustness in actual system, redundant design usually exists.

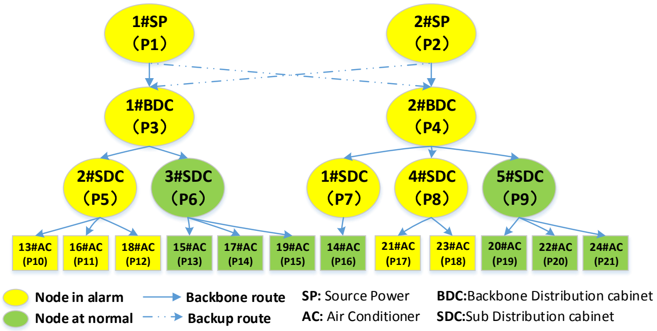

Figure 1. Different clustering results under a sample topology.

Take an actual application for example. Figure 1 shows how data center provides power supply for air conditioners. It's a 4-layer topology model, where P1-P2 are in the first layer (the source power, namely SP), and P3-P4 are in the second layer (the Backbone Distribution cabinet, namely BDC), then P5-P9 are in the third layer (Sub Distribution cabinet, namely SDC) and the terminal nodes P10-P21 are in the fourth layer (Air Conditioner, namely AC). The backbone route and the Backup route are the two routes provided for the power supply on the terminal air conditioners. So for each AC, there are two power sources provided, i.e. one for primary use and the other for backup. For example, the routes P1→P3→P5→P10 and P2→P3→P5→P10 both provide power supply for P10.

The existence of backup branch will change the final clustering result of alarm devices in different conditions. In Figure 1, supposing the following nodes (devices) have alarms: {P1,P2,P3,P4,P5,P7,P8,P10,P11,P12,P17,P18} (the yellow nodes represent those in alarm while the green ones represent normal), we can obtain all the alarm clusters and the root causes in each cluster, but the clusters vary with the relationship changing between P1,P2, and P3,P4, as described in Table 1.

Table 1. Clustering results in different conditions

Conditions	Clustering Results	Root-cause Node
P3 is the child of P1, P4 is the child of P2.	{P1,P3,P5,P10,P11,P12}	P1
	{P2,P4,P7,P8,P17,P18}	P2
P3,P4 are the children of P1	{P1,P3,P4,P5,P7,P8,P10, P11,P12,P17,P18}	P1
	{P2}	P2
P3,P4 are the children of P2	{P1}	P1
	{P2,P3,P4,P5,P7,P8,P10, P11,P12,P17,P18}	P2
P3 is the child of P2, P4 is the child of P1	{P2,P3,P5,P10,P11,P12}	P2
	{P1P4,P7,P8,P17,P18}	P1

In mathematics and computer science, a directed acyclic graph structure (DAG)[16] is defined as a finite directed graph with no cycles, namely there is no way to start at any vertex v and follow a sequence of edges that eventually loops back to v again. So the topology above can be modeled as a typical DAG since the current flow direction is irreversible. The DAG model makes it difficult to find out the root-cause node of a cluster. In reality, although there are usually two-upstream nodes for most of the devices, only one upstream node exists at a specific working time, because a device can only accept one main supply provision(called the main) while the other one is just backup and does not work(called the backup). In addition, the main and the backup switch their roles since the power supply of the upper layer changes from time to time.

Considering these characteristics above, we take the topology as a dynamic multi-tree (DMT) structure. For one thing, in mathematics and computer science, a tree structure is defined recursively as a collection of nodes, a root node and multiple children nodes, where there is only one parent node for each child node. Similarly, in this power system, for a specific working time, one downstream node (the child) only connects with only one single upstream node (the parent), which is the same as the tree structure, so we can regard it as a multi-tree naturally. For another, the role-switching between the main and the backup over time results in the change of the tree-structure, so we regard it as a dynamic tree. That's why we call it a dynamic multi-tree structure. In this way, we transform the problem of finding out the root-cause of clusters on a DAG to the one on a dynamic multi-tree.

For convenience, we define the node topology attributes as shown in Table 2. Based on this, we make a comb for all the supply system device nodes for our further study.

Table 2. Description of the node attributes

Name	Description
ID	Device identity
NAME	Device name
ROOM	Room number where the Device is located
PARENT1	The parent ID (primary branch)
PARENT2	The parent ID (backup branch), note that PARENT2=0 means there is no backup branch
FLAG	If FLAG = 0, PARENT1 is the parent ID ; If FLAG = 1, PARENT2 is the parent ID
CHILD$_i$	i_{th} child of the device($i = 0,1,2...$)

3. ROOT CAUSE ALARM CLUSTERING MODEL

Via above analysis, we can cluster the alarm devices and figure out the root-cause of a cluster by solving the root node of a dynamic multi-tree. For convenience, Table 3 shows the notations.

Table 3. Notations

Symbol	Description
T	Time period. In addition, ΔT means the maximum time interval between two temporally associated alarms.
S	Alarm sets. $S = \{s_i\}$, where s_i is the i_{th} alarm record.
C	Clusters of alarm device. $C=\{C_j\}_{1\le j\le n}$, where C_j is the j_{th} segmented clusters. $C_j=\{c_{ij}\}_{1\le i\le m}$, where c_{ij} is the i_{th} root-cause alarm cluster among C_j .
R	Root-cause ID sets. $R = \{r_{ij}\}_{1\le i\le m,\, 1\le j\le n}$, where r_{ij} is the root-cause device ID of cluster c_{ij} .
X	Device (node) IDs. $X = \{x_i\}$, where x_i is device ID of the alarm record s_i .
I	Alarm types. $I=\{I_k\}$, where I_k is the k_{th} alarm type.

Assuming a sequence of alarm records $s_1,...,s_i,s_{i+1}...$ (whose device ID is $x_1,...,x_i,x_{i+1}...$ correspondingly) occur at time $t_1,...,t_i,t_{i+1}...$ during $[t_0, t_0+T]$, we consider that only when $|t_{i+1}-t_i| \le \Delta T$ does there exist temporal correlation between the two records. On the contrary, there will be an alarm node segmentation between x_i and x_{i+1} if $|t_{i+1}-t_i| > \Delta T$, which divides them into two segmentations. Suppose x_i belongs to the j_{th} segmented part, i.e. $x_i \in C_j$, $C_j \subseteq C$, then:

$$x_{i+1} \in \begin{cases} C_j, & if\ |t_{i+1}\text{-}t_i| \le \Delta T \\ C_{j+1}, & otherwise \end{cases} \tag{1}$$

where $C_{j+1} \subseteq C$ is the $j+1_{th}$ alarm segmentation.

In this way, a series of alarm segmentations $\{C_j\}_{1\le j\le n}$ is derived.

Moreover, the alarms in each segmented alarm cluster C_j can be further clustered into root-cause clusters c_{ij} because several root-cause alarms may occur during this same time period. Figure 2 shows the relationship of alarm sets S, segmented clusters C_j, root-cause clusters c_{ij} and root-cause nodes r_{ij} .

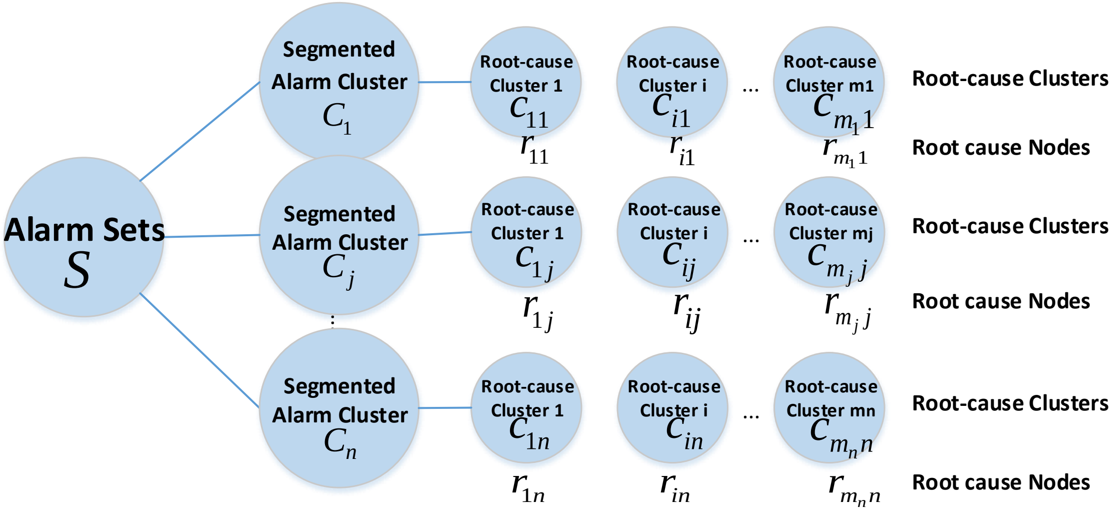
Alarm Sets
S
Segmented Alarm Cluster C_1
Segmented Alarm Cluster C_j
Segmented Alarm Cluster C_n
Root-cause Cluster 1 C_{11}
Root-cause Cluster i C_{i1}
...
Root-cause Cluster m1 $C_{m_1 1}$
r_{11}
r_{i1}
$r_{m_1 1}$
Root-cause Clusters
Root cause Nodes
Root-cause Cluster 1 C_{1j}
Root-cause Cluster i C_{ij}
Root-cause Cluster mj $C_{m_j j}$
r_{1j}
r_{ij}
$r_{m_j j}$
Root-cause Clusters
Root cause Nodes
Root-cause Cluster 1 C_{1n}
Root-cause Cluster i C_{in}
Root-cause Cluster mn $C_{m_n n}$
r_{1n}
r_{in}
$r_{m_n n}$
Root-cause Clusters
Root cause Nodes

Figure 2. The relationship of alarm sets, segmented alarm clusters, root-cause clusters and root-cause nodes.

According to the processing above, the problem is transformed into finding out the alarm clusters c_{ij} and the corresponding root-cause nodes r_{ij} in each segmented alarm cluster C_j .We think it as a problem of graph traverse. Breadth-first search (BFS) is widely used for graph traversal, but the traditional BFS method is non-directional and unordered, which is not suitable in this condition. Therefore, our algorithm called improved BFS on dynamic tree (DTIBFS) is proposed to cluster the alarm nodes and figure out root-cause node effectively. The algorithm is detailed in Figure 3.

In DTIBFS algorithm in Figure 3, the inputs are alarm sets S obtained during $[t_0, t_0 + T]$ and the topology model where the nodes described in Table 2. There are three main steps. Step I initialize the parameter ΔT . Step II generates the segmentations $\{C_j\}_{1\le j\le n}$. Step III realizes the function to cluster c_{ij} and find out its root r_{ij} for each C_j . Procedure 1 in Step III is an initialization for clustering, in which c_{ij} , Q , C_j , c_{ij} are initialized respectively. To mention, the parameter Q is queue-structured which follows the FIFO principle. Procedure 2 in step III is the main process for clustering. Here, we define a head property $\mathrm{H}|c_{ij}$ for cluster set c_{ij} , which stores the local result of alarm root in each cluster c_{ij} . In Procedure 2.1.1 and 2.1.2, when Q 's head node x pops out, its parent node $\hat{x}$ and children nodes $\breve{x}$ are figured out according to the Table 2, where the attribute FLAG is used to select which parent is available, so as to keep it consistent with the real-time power supply system. From Procedure 2.1.3-2.1.5, the parent node $\hat{x}$ enqueues ahead of children nodes $\{\breve{x}\}$ to guarantee $\mathrm{H}|c_{ij}$ always updates to the local root node during the iteration. Thus $\mathrm{H}|c_{ij}$ is the root cause node r_{ij} globally in Procedure 2.3 after finishing traversing the entire cluster c_{ij} . After a specific root-cause alarm cluster c_{ij} is done, it moves to another c_{ij} (in Procedure 2.4 and 2.5), then repeat Procedure 2.1-2.5 until $C_j = \varnothing$. After clustering for C_j is done, it moves to step III again and repeat Procedure 1-3 until each $\{C_j\}_{1\le j\le n}$ is clustered. In the end, we obtain the outputs: root-cause clusters c_{ij} , root-cause device node ID r_{ij} in each c_{ij} during time $[t_0, t_0 + T]$.

```
DTIBFS Algorithm
INPUT: Alarm sets $S$ obtained during $[t_0, t_0 + T]$, topology model where
       the nodes described as Table 2.
I. Initialization: $\Delta T = 60s$;
II. Alarm Segmentation on $S$ based on formula (1) to generate segmentations
    $C = \{C_j\}_{1\le j\le n}$;
III. Realize the clustering $c_{ij}$ and find out its root $r_{ij}$ for $C_j$ :
  For $\forall C_j \subseteq C$ do
    1. Initialization:
      1.1 Set $i = 0$, root-cause cluster set $c_{ij} = \varnothing, \forall i \in [1, m]$,visit queue
          $Q = \varnothing$;
      1.2 Select a node $x$ randomly from $C_j$, $C_j = C_j \setminus \{x\}$;
      1.3 $c_{ij} = c_{ij}\ \mathrm{U}\{x\}$; $enqueue\ x \to Q$;
    2. While $C_j \neq \varnothing$ do
      2.1 While $Q \neq \varnothing$ do
        2.1.1 $x \leftarrow Q.dequeue$;
        2.1.2 Find $\hat{x} \leftarrow parent\langle x\rangle$, $\breve{x} \leftarrow child\langle x\rangle$;
        2.1.3 $enqueue\ \hat{x} \to Q$ in priority;
        2.1.4 $\forall \breve{x}, enqueue\ \breve{x} \to Q$;
        2.1.5 $\mathrm{H}|c_{ij} \leftarrow \hat{x}$;
        2.1.6 $\forall \breve{x}, c_{ij} = c_{ij}\ \mathrm{U}\{\breve{x}\}$;
        2.1.7 $C_j = C_j \setminus \{\hat{x}\}$ , $C_j = C_j \setminus \{\breve{x}\}, \forall \breve{x}$;
      2.2 End while
      2.3 $r_{ij} \leftarrow \mathrm{H}|c_{ij}$;
      2.4 $i++$;
      2.5. Select another node $x$ randomly in $C_j$;
    3. End while
  End for
OUTPUT: Root-cause clusters $c_{ij}$ , $\forall i \in [1, m], j \in [1, n]$; root-cause device
        node ID $r_{ij}$ in each $c_{ij}$ during $[t_0, t_0 + T]$
```

Figure 3. Improved BFS algorithm on Dynamic Tree for root cause clustering (DTIBFS algorithm).

The main idea of DTIBFS algorithm is to cluster the alarm nodes c_{ij} for each $C_j \subseteq C$, and figure out root-cause node r_{ij} of each cluster c_{ij} effectively. The key is, regarding a given device node, the parent node will be visited in priority, and then children nodes afterwards, guaranteeing the root-cause node r_{ij} always updates to the local root node during the traverse and stays in the head of visited list. The DTIBFS algorithm derives from the traditional Breadth-First (BFS) strategy, while there exists a difference between DTIBFS and BFS, where in BFS, there is no priority between the nodes when traversing. The DTIBFS algorithm will be applied in the association analysis in section 4.

4. ALARM ASSOCIATION ANALYSIS ON FP-Growth METHOD

According to the model in section 3, we can obtain the clusters by DTIBFS algorithm and the root cause node of each cluster. Nevertheless, we have no prior knowledge on the association of the alarm records. Therefore, we propose association analysis between root-cause clusters for further research. Considering the complexity of the two widely used association analysis algorithms, Apriori based and FP-growth based algorithm, in this paper we adopt FP-growth method due to its $O(n)$ complexity performance. The algorithm is detailed in Figure 4. The cluster $c_{ij} \subseteq C$ is obtained by DTIBFS algorithm in section 3.

Because raw alarm record $s \in S$ is literally different from each other and is ill-conditioned for statistic-sensitive association analysis, a mapping process is adopted from original raw alarm content to typical alarm information types, which have been defined in advance according to the actual root-cause alarms. In addition, the generation of transactions for association analysis is based on the alarm clusters mentioned in section 3, i.e. all alarm records that belong to the same cluster will be processed as one transaction and furthermore become the input of FP-Growth analysis. Compared with the traditional sliding-time-window transaction method, our cluster-ware method can better utilize the topology structure of actual network, which will lead to a more reasonable association representation of alarm information. The method in Figure 4 differs from the common ways, since traditionally they are generated by a sliding window [13-15].

```
INPUT: Alarm sets $S$ , root-cause alarm clusters $c_{ij} \subseteq C$ ,
          alarm type $I$ , $minSupport$ , $minConfidence$ .
I. Alarm content mapping:
   1. For $\forall c_{ij}, i \in [1, m], j \in [1, n]$ do
      1.1 Map its relevant $s \in S$ to corresponding alarm type $I_k \in I$ ;
      1.2 Generate alarm type transaction based on $I_k$
   2. End for
II. Frequent item set generation based on FP-growth:
   3. FP-tree construction with transactions generated in 1.2;
   4. Recursively generate frequent item sets by using conditional
      pattern base of each $I_k$ under $minSupport$ constraints;
III. Association rule analysis:
   5. Generate all possible subset $A$ and different set $B$ for each
      frequent item set;
   6. Obtain rule $A \rightarrow B$ if $minConfidence$ requirement is satisfied.
OUTPUT: Association rule sets.
```

Figure 4. Power supply system alarm association analysis method based on FP-Growth.

5. NUMERICAL RESULTS

5.1 The Results of DTIBFS Model

We make a comb of 1158 devices (nodes) according to Table 2 in section 2, and extract 30,972 alarm records occurring in April, 2017 from the backend management system database of the government's data center(Ministry of industry and information technology of the People's Republic of China) for experiment. Part of the records (3,287 records on April 23 and 24) are shown in Figure 5. It shows burst characteristics, meaning that a great many alarms occur in a remarkably short time while they seldom occur at other moments.

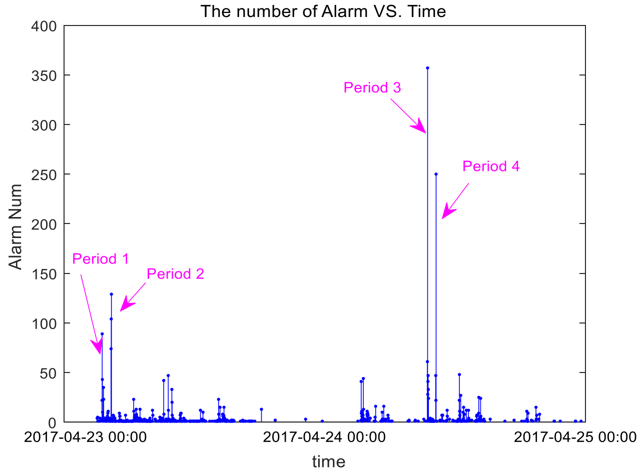

Figure 5. The number of Alarm VS. Time.

To better describe the results, Accumulative Percentage of the TOP-K alarm clusters ($AP\_TOP(K)$) is defined:

$$AP\_TOP(K)=\frac{\sum_{k}^{K} Number\ of\ alarm\ records\ in\ c_{ij}^{k}}{\sum_{j}\sum_{i} Number\ of\ alarm\ records\ in\ c_{ij}} \qquad (2)$$

where c_{ij} is the clusters obtained by DTIBFS algorithm in section 3 and c_{ij}^{k} is the k_{th} cluster listed by the number of alarm records in c_{ij} in descending order.

To test the effectiveness of DTIBFS, four periods are chosen when the most alarms occur in Figure 5. The experimental results are shown in Figure 6, where we set $\Delta T = 60s$ .We conclude that once we deal with the top 4 alarm clusters, we will address more than 90% of the alarms, which will greatly decrease the pressure of system operation workload. Note that the alarms occurred in an extremely short time during these four periods.

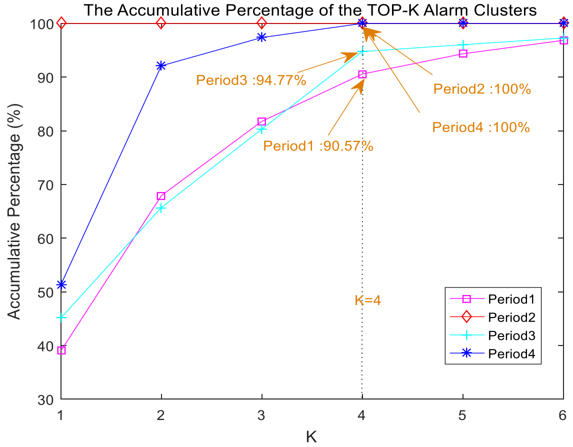

Figure 6. The accumulative percentage of the TOP-K alarm clusters.

In order to generalize the conclusion, the average Accumulative Percentage of TOP-K alarm clusters ($AP\_TOP(K)$) according to the entire history alarm records is researched and indicated in Figure 7. It turns out that choosing top 20 alarm root-causes(clusters) suffices for addressing more than 60% of the total alarms, which is also a significant improvement compared with the inefficiency of handling the records one by one.

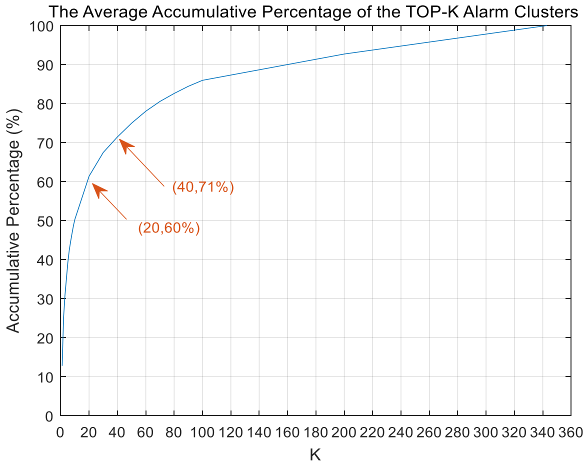

Figure 7. The average accumulative percentage of the TOP-K alarm clusters.

Furthermore, to explore the effect of ΔT , a comparison is made in Figure 8. We find that the $AP\_TOP(K)$ increases with ΔT , but different ΔT has different influences. When ΔT changes from 10s to 30s, $AP\_TOP(K)$ increases remarkably. Then the influence reduces slowly with increasing ΔT . Especially, the $AP\_TOP(K)$ curves are nearly overlapping when ΔT changes from 40s to 80s. When ΔT exceeds 100s, we consider that the time interval for two successive alarms is so big that there is no association between these two alarms.

As is indicated in Figure 8, the bigger ΔT is, the faster the accumulative speed of $AP\_TOP(K)$ will be, because we can put more alarm records in the same alarm set as a cluster. However, the smaller ΔT is, the stronger association between the alarm records inside one cluster will be. Based on the above two contrary influences, it's reasonable that ΔT be set between 40s and 80s.

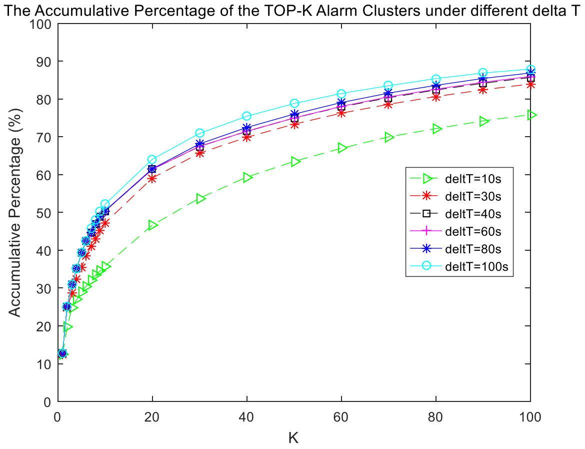

Figure 8. The accumulative percentage of the TOP-K alarm clusters under different delta T.

Generally, in the actual data center, alarm flooding occurs from time to time, especially the alarm caused by power supply system faults. Based on the results above, we see that the root cause model of DTIBFS algorithm can help eliminate the high pressure of a large quantity of burst alarm data. By choosing an appropriate K (where K is chosen according to the practical situation) and an appropriate ΔT (where ΔT is set 60s usually) , and working on the DTIBFS algorithm, we can handle the TOP-K alarm clusters when alarm flooding occurs, and find the root cause of each cluster. In the actual case above, we can handle over 90% of alarm records by dealing with only 4 alarm clusters in the alarm flooding period, or handle over 60% alarm records by dealing with only 20 alarm clusters on the average, so as to eliminate the workload of operation considerably in data center.

5.2 Association Analysis on Root-cause Model

We have defined 39 alarm types for the supply system. Based on the FP-growth method in section 4, the association analysis can be realized. With different condition of minSupport and minConfidence constraints at ΔT =60s, we obtain different association rules, whose numbers are shown in Figure 9.

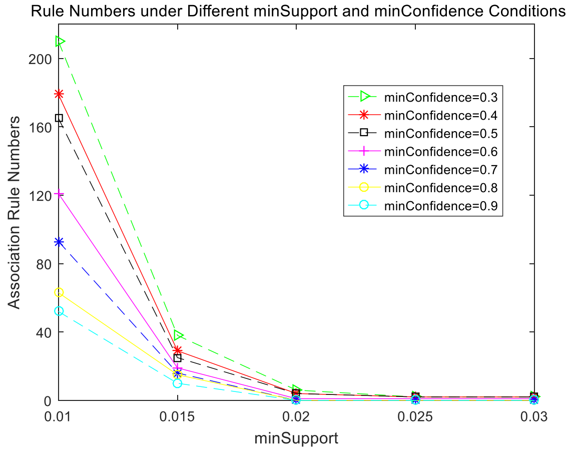

Figure 9. Association rule numbers with different minSupport and minConfidence constraints.

With the increase of minSupport and minConfidence, the number of association rules decrease sharply, which is consistent with the FP-growth algorithm. Taking minSupport=0.015 , minConfidence=0.7 as an example, the 16 rules are listed in detail in Table 4.

Table 4 Association rules under a specific condition of minSupport and minConfidence

A→B (minSupport=0.015, minConfidence=0.7, ΔT =60s)			
1	UPS fault	→	UPS battery mode or fault
2	UPS fault	→	Low (average) Line voltage
3	UPS fault	→	UPS battery mode or fault & Low (average) Line voltage

4	UPS fault & UPS battery mode or fault	→	Low (average) Line voltage
5	UPS fault & UPS battery mode or fault	→	UPS battery mode or fault
6	UPS battery mode or fault & UPS battery mode or fault	→	UPS fault
7	Low frequency	→	Low (average) Phase voltage
8	Abnormal Input of UPS electric	→	UPS battery mode or fault
9	Abnormal Input of UPS electric	→	Low (average) Line voltage
10	Abnormal Input of UPS electric	→	UPS battery mode or fault & Low (average) Line voltage
11	UPS battery mode or fault & Abnormal Input of UPS electric	→	Low (average) Line voltage
12	Low (average) Line voltage & abnormal Input of UPS electric	→	UPS battery mode or fault
13	UPS battery mode or fault & UPS Ballast fault	→	Low (average) Line voltage
14	Low (average) Line voltage & UPS Ballast fault	→	UPS battery mode or fault
15	Low (average) Phase voltage & air conditioner fan Shut down	→	Low (average) Line voltage
16	air conditioner fan Shut down &Low (average) Line voltage	→	Low (average) Phase voltage

In Table 4, each line represents an association rule from set A on the left to set B on the right under the condition of minSupport=0.015, minConfidence=0.7, and $\Delta T = 60s$. The sign "&" represents the condition "and". From Table 4, among the 16 rules, the main alarm types are focused on UPS system (e.g. phase voltage, line voltage and frequency), battery system (e.g. battery mode, voltage), and air conditioner system (e.g. fans shutting down). It turns out that the alarm types of UPS system, battery system, and air conditioner system are strongly associated with each other since they occur simultaneously (as an association rule). That means, if a UPS alarm occurs, we can also figure out the potential alarms that will occur subsequently on air conditioner system or battery system, etc.. In this way, we can find out the related missing alarms in time.

In a word, due to the consideration of supply system topology, the rules above can dig the association from both statistical point and topology point, which is helpful to detect whether there are missing rules during the alarm floods or provide tips for alarm diagnosis and handling.

6. CONCLUSION

There are two main contributions in this paper. First, a novel root-cause alarm association analysis method based on actual topological constraints for alarm diagnosis is proposed. Unlike employing sliding-time-window model as previous, we take network topology into consideration and generate a series of alarm clusters for alarm devices. In order to achieve above goal, an algorithm named DTIBFS is introduced, which can cluster the alarm nodes and figure out root-cause node efficiently. In this way, over 90% alarm records can be addressed by dealing with only 4 alarm clusters in the alarm flooding period and over 60% be handled by focusing on only 20 alarm clusters averagely in a long time period, both of which contributes to a remarkable improvement of efficiency and reduce of operation workload compared with the inefficiency of handling the records one by one as the traditional operation. Second, based on the transactions generalized according to the DTIBFS algorithm above, an improved association analysis on FP-growth algorithm is made. Lots of meaningful rules under different association requirements are obtained. By considering the supply system topology, it will be easier to locate the critical alarm device quickly and meanwhile analyze the alarm cause in a more comprehensive aspect by mining the history data.

7. REFERENCE

[1] Chen, M., Mao, S., & Liu, Y. 2014. Big data: A survey. Mobile Networks and Applications, 19(2), 171-209.

[2] Harris, M., & Geng, H. 2015. Data center infrastructure management. Data center handbook, 1st edn. Wiley, Hoboken, NJ, 601-618.

[3] Jennings, B., & Stadler, R. 2015. Resource management in clouds: Survey and research challenges. *Journal of Network and Systems Management*, 23(3), 567-619.

[4] Solmaz, S. E., Gedik, B., Ferhatosmanoğlu, H., Sözüer, S., Zeydan, E., & Etemoğlu, Ç. Ö. 2017. ALACA: A platform for dynamic alarm collection and alert notification in network management systems. *International Journal of Network Management*.

[5] Costa, R., Cachulo, N., & Cortez, P. 2009, October. An intelligent alarm management system for large-scale telecommunication companies. In Portuguese Conference on Artificial Intelligence (pp. 386-399). Springer, Berlin, Heidelberg.

[6] Solanki, S. K., & Patel, J. T. 2015, February. A survey on association rule mining. In Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on (pp. 212-216). IEEE.

[7] Kaur, M., & Kang, S. 2016. Market Basket Analysis: Identify the changing trends of market data using association rule mining. Procedia Computer Science, 85, 78-85.

[8] Mooney, C. H., & Roddick, J. F. 2013. Sequential pattern mining--approaches and algorithms. ACM Computing Surveys (CSUR), 45(2), 19.

[9] Agrawal, R., & Srikant, R. 1994, September. Fast algorithms for mining association rules. In *Proceedings. 20th int. conf. very large data bases*, VLDB (vol. 1215, pp. 487-499).

[10] Han, J., Pei, J., & Yin, Y. 2000, May. Mining frequent patterns without candidate generation. In ACM sigmod record (Vol. 29, No. 2, pp. 1-12). ACM.

[11] Julisch, K. 2001, December. Mining alarm clusters to improve alarm handling efficiency. In Computer Security Applications Conference, 2001. ACSAC 2001. P*roceedings 17th Annual* (pp. 12-21). IEEE.

[12] Buczak, A. L., & Guven, E. 2016. A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153-1176.

[13] Wang, J., Li, H., Huang, J., & Su, C. 2016. Association rules mining based analysis of consequential alarm sequences in chemical processes. *Journal of Loss Prevention in the Process Industries*, 41, 178-185.

[14] Deypir, M., Sadreddini, M. H., & Hashemi, S. 2012. Towards a variable size sliding window model for frequent itemset mining over data streams. Computers & industrial engineering, 63(1), 161-172.

[15] Wrench, C., Stahl, F., Le, T., Di Fatta, G., Karthikeyan, V., & Nauck, D. 2016. A Method of Rule Induction for Predicting and Describing Future Alarms in a Telecommunication Network. In Research and Development in Intelligent Systems XXXIII: Incorporating Applications and Innovations in Intelligent Systems XXIV 33 (pp. 309-323). Springer International Publishing.

[16] Wang, L. 2013. Directed acyclic graph. In Encyclopedia of Systems Biology (pp. 574-574). Springer New York