A Survey of Cloud Service Events and Their Connections

Hangping Hu, Zhen Zhang, Weijian Qin, Yuan Wang and Xiaojian Li

School of Computer Science & Engineering Guangxi
Normal University, Guilin, China

ABSTRACT

Any unexpected service interruption or failure may cause customer dissatisfaction or economic losses. To distinguish the rights and interests or security disputes between cloud service providers and customers, explore the essence and rules of cloud service events and their various connections, such as: Normal contact of service scheduling, normal contact of service dependence, abnormal contact of resource competition, abnormal contact of service delay, abnormal contact of service dependence, etc., as well as their rules in time, resources, scheduling and other aspects, and the form of the rules; The purpose is to provide the above abnormal connections, as well as the rule and presentation form in terms of time, resources and load, for the study of violation determination and failure tracing in the cloud service accountability mechanism.

KEYWORDS

Cloud service, Event connections, Correlation, Label adaptation.

1. INTRODUCTION

After receiving user service requests, cloud service providers implement their requirements into one or more cloud service jobs, which may have sequence and dependency relationships. Each cloud service job consists of one or more tasks, and there are dependencies among the tasks, which are represented by the DAG (Directed Acyclic Graph) of the tasks. After the task is put into operation, one or more instances will be generated. Since there is a dependency relationship between tasks, instances of two tasks with a dependency relationship must have a dependency connection. It would be helpful to trace the source more accurately if we could point out which instances the connection is caused by, when it ends, and the type and strength of the connection.

This paper summarizes the research status of cloud service events and their connections, mainly including: firstly, this paper reviews the research status of cloud service events and their relationship. Secondly, summarizing and giving a classification of cloud service events and their connections from the current state of affairs. Thirdly, it points out that the common problem at present is interpretable label adaptive labeling. Finally, the research approaches of cloud service events and their correlation are suggested.

The following four sections are as follows: Section 2 summarizes the current situation of cloud service events and their connection; Section 3 gives the classification of current cloud service events and their connection; Section 4 summarizes the remaining problems in the current research and gives suggestions on research approaches; Section 5 is the summary of the full text.

2. RESEARCH STATUS

The relevant research status is summarized in two sections: Section 2.1 discusses cloud service events and Section 2.2 discusses the connection between cloud service events.

2.1. Cloud service events

For the lack of labeled data in abnormal detection, paper [2] presents an unsupervised abnormal detection method, which can identify three types of anomalies: Service timeout, network delay and data loss. This method ignores the dependency between events, and independently examines the abnormal conditions of events, and ignores the abnormal conditions such as the delay time of bad services and the dissatisfaction of resources.

Paper [3] makes an empirical analysis on abnormal events collected by 18 online service systems of Microsoft, identify accidental abnormal events, such as wrong procedures submitted by customers, so as to effectively deal with abnormal events, but this method relies on engineers to manually mark accidental abnormal events, which is costly.

Papers [4,6] only focus on the abnormal resource consumption event of the job. Paper[5] only focus on job scheduling failure events, papers [7-10] only focus on the abnormal consumption of container resources, and paper [11] only focus on instance authorization failure and Instantiate abnormal event. They all assume that each event is independent of other events, pay attention to the abnormal of events in terms of resources and scheduling, but ignore the abnormal caused by the dependency between events.

Papers [27-31] analyze the failure characteristics of nodes, but they separate the connection between nodes and independently investigate the failure characteristics of nodes, ignoring the abnormal connection between nodes.

It can be seen that there are still abnormal in the research of cloud service events, which ignore the delay time of bad services and the dependency between events, and rely on manual labeling, which is inefficient.

2.2. Cloud Service Event Connections

Papers [12-16] rely on manual annotation, Give the two types of connections between services "with-without", so as to trace the source of service interruption.

Paper [18] only focus on resource competition between jobs, papers [19-21] only focus on dependencies between containers. They all focus on the connection between events from a single aspect, without considering multiple abnormal connections between events from various aspects such as time, resources and scheduling.

Papers [22-24] can analyze whether there is a connection between events, and give the general characteristics of the connection, but they fail to distinguish the types of connections.

Papers [33, 34] based on subspace method, and papers [35-37] based on feature selection methods, extract or construct a new low-dimensional feature space in the high-dimensional space to detect the abnormal connection, but due to the unknown and heterogeneity of the abnormal connection, these methods are difficult to ensure that the newly constructed feature space already

contains the semantics required for the Multi-Classification and Multi-Label abnormal connections discovery [32].

It can be seen that the research on event connection still relies on manual annotation, which is inefficient, and does not integrate time, resources, scheduling, dependency and other aspects at the same time, and the problem of unclear understanding of multi-Classification connection characteristics.

3. CLOUD SERVICE EVENTS AND THEIR CONNECTIONS CLASSIFICATION

This section summarizes and classifies cloud service events and their connections from the current situation.

3.1. Cloud service events classification

According to the papers [2-11, 27-31], it is summarized that there are normal events, unknown events and abnormal events in cloud service events, this article focuses on abnormal events. The cloud service events classification diagram is shown in Figure 1.
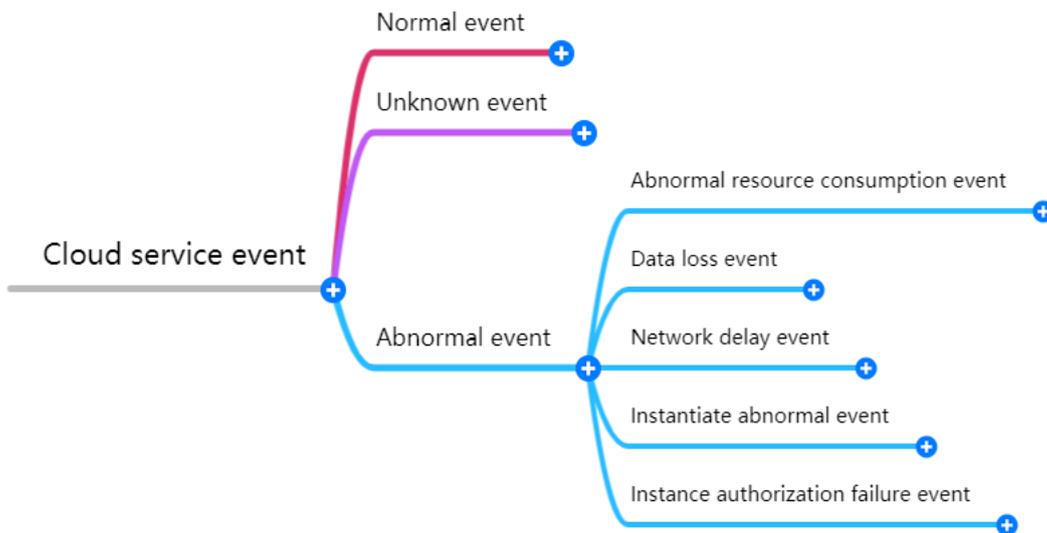

Figure 1. Cloud service events classification

We classify cloud service events into normal events, unknown events, and abnormal events based on their completion. Abnormal events include abnormal resource consumption events, data loss events, network delay events, instantiation abnormal events, and instance authorization failure events.

3.2. Cloud service event connections classification

According to the papers [12-16, 18-24, 32-37], it can be concluded that cloud service event connections includes normal connection, unknown connection and abnormal connection. This article focuses on abnormal connection. The cloud service event connections classification diagram is shown in Figure 2.
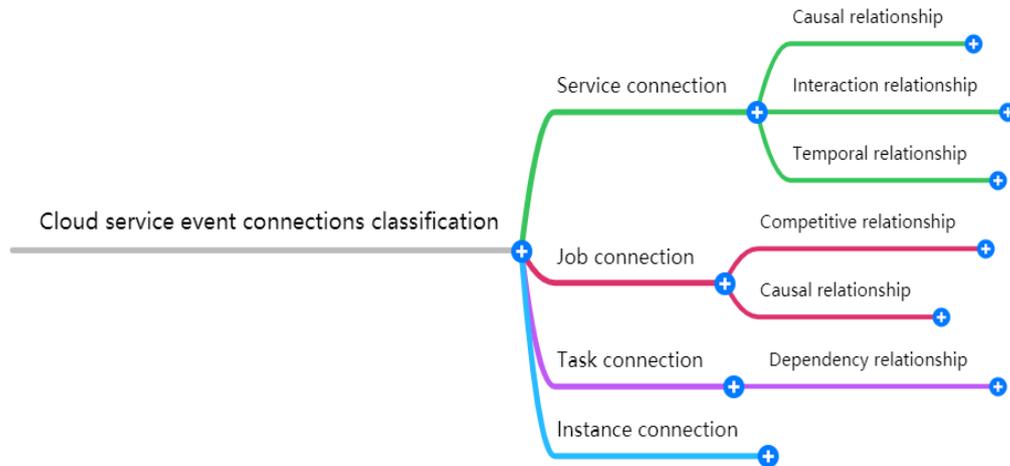
Figure 2. Cloud service event connections classification

According to the granularity of cloud service events, we classify event connections into service connection, job connection, task connection and instance connection. Among them, service connection includes causal relationship, interaction relationship and temporal relationship among services. Job connection includes competitive relationship and causal relationship between jobs. Task connection includes dependency relationship between tasks.

4. REMAINING PROBLEMS AND SUGGESTIONS FOR RESEARCH APPROACHES

4.1. Remaining problems

Problem 1: Adaptive labeling problems of interpretable labels.
Problem 2: Difficulties in understanding the nature of multi-classification connections in cloud service events.
Problem 3: The method challenge of multi-classification connection discovery of cloud service events.

4.2. Research approaches and suggestions

Based on the above urgent problems, the prospective system shown in Figure 3 is proposed.
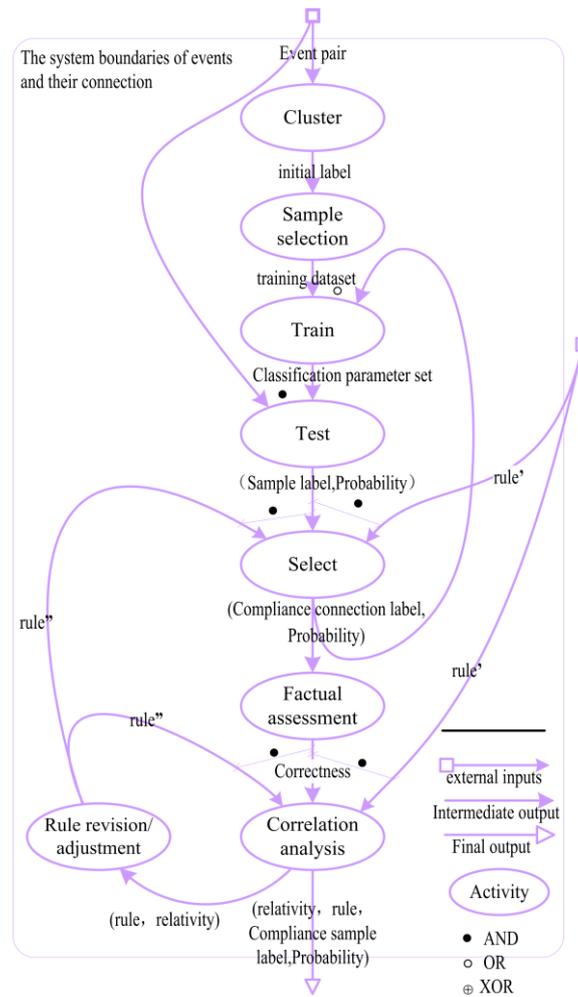
Figure 3. Prospective system

The ellipse in Figure 3 represents the activity, and the directed solid line represents the entity. Directed lines starting from hollow rectangles represent external inputs, and directed lines with solid arrows at the end represent intermediate outputs, and directed lines with hollow arrows at the end represent final outputs. The analysis of cloud service failure factors includes 8 activities: cluster, sample selection, train, test, select, factual assessment, correlation analysis and rule revision/adjustment.

According to the prospective system shown in Figure 3, we give the research scheme. Firstly, explore the adaptive annotation method of event connection labels, such as the initial label of the connection between events is given by clustering. Secondly, analyze the initial label meaning of the connections between events, and get the types of connections and some characteristics of each connection. Thirdly, part of the characteristics of all kinds of connections are regarded as semi-supervised classification rules, such as "do not accept the connection labels that start only after the subsequent events have ended", "do not accept the normal association labels that abnormal end of the subsequent events and abnormal end of the previous events", etc. Then, through semi-supervised classification, classify the types of connections between events.

Fourthly, according to the connections of events, get more characteristics and laws of the connections of various events, as well as the presentation forms of laws, and construct the event network. Finally, evaluate the above process, and give the evaluation results.

5. CONCLUSIONS

This article reviews the current research status of cloud service events and their connections, and points out the following three common problems:

1）Deep learning relies on labels, moreover, event connections label of cloud service is people's qualitative prior knowledge of event connections, it is usually manually annotation in advance and then classified, but manual annotation is labor intensive and inefficient.

2）The explanatory nature of the label depends on the recognition of multi-classification connection characteristics, however, the current cloud service event connection has not considered time, resources, scheduling, and dependencies at the same time, and the problem of unclear understanding of the nature of multi-classification of connection;

3）Machine adjustment labels depends on the understanding of the laws of multi-classification connection, however, the current research is not clear about the characteristics of multi-classification connections, and part of it contains the semantic problems required for multi-classification connections discovery.

For the above problems, we give the prospective system of cloud service events and their connections, which basically meet the requirements of cloud service events and their connections.

REFERENCES

[1]   Raju B K, Geethakumari G. Event correlation in cloud: a forensic perspective[J]. Computing, 2016, 98(11): 1203-1224.
[2]   Nedelkoski S, Cardoso J, Kao O. Anomaly detection and classification using distributed tracing and deep learning[C]//2019 19th IEEE/ACM International Symposium on Cluster, Cloud and   Grid Computing (CCGRID). IEEE, 2019: 241-250.
[3]   Chen J, Zhang S, He X, et al. How Incidental are the Incidents? Characterizing and Prioritizing Incidents for Large-Scale Online Service Systems[C]//2020 35th IEEE/ACM International Conference on Automated Software Engineering (ASE). IEEE, 2020: 373-384.
[4]   Lu C, Chen W, Ye K, et al. Understanding the Workload Characteristics in Alibaba: A View from Directed Acyclic Graph Analysis[C]//2020 International Conference on High Performance   Big Data and Intelligent Systems (HPBD&IS). IEEE, 2020: 1-8.
[5]   Jassas M S, Mahmoud Q H. Failure characterization and prediction of scheduling jobs in google cluster traces[C]//2019 IEEE 10th GCC Conference & Exhibition (GCC). IEEE, 2019: 1-7.
[6]   Chen X, Lu C D, Pattabiraman K. Failure analysis of jobs in compute clouds: A google cluster case study[C]//2014 IEEE 25th International Symposium on Software Reliability Engineering. IEEE, 2014: 167-177.
[7]   Ren R, Li J, Wang L, et al. Anomaly Analysis and Diagnosis for Co-located Datacenter Workloads in the Alibaba Cluster[C]//International Symposium on Benchmarking, Measuring and Optimization. Springer, Cham, 2019: 278-291.
[8]   Scheinert D, Acker A. Telesto: A graph neural network model for anomaly classification in cloud services[C]//International Conference on Service-Oriented Computing. Springer, Cham, 2020: 214-227.

[9]   Gulenko A, Schmidt F, Acker A, et al. Detecting anomalous behavior of black-box services modeled with distance-based online clustering[C]//2018 IEEE 11th International Conference on Cloud Computing (CLOUD). IEEE, 2018: 912-915.

[10] Acker A, Schmidt F, Gulenko A, et al. Online Density Grid Pattern Analysis to Classify Anomalies in Cloud and NFV Systems[C]//2018 IEEE International Conference on Cloud Computing Technology and Science (CloudCom). IEEE, 2018: 290-295.

[11] Vedurumudi P V V, Morusupalli P. System and method of providing post error analysis for instances of applications in cloud service environments on a per user basis: U.S. Patent 10,379,934[P]. 2019-8-13.

[12] Wang Y, Li G, Wang Z, et al. Fast Outage Analysis of Large-scale Production Clouds with Service Correlation Mining[C]//2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE). IEEE, 2021: 885-896.

[13] Meng Y, Zhang S, Sun Y, et al. Localizing failure root causes in a microservice through causality inference[C]//2020 IEEE/ACM 28th International Symposium on Quality of Service (IWQoS). IEEE, 2020: 1-10.

[14] Jia T, Chen P, Yang L, et al. An approach for anomaly diagnosis based on hybrid graph model with logs for distributed services[C]//2017 IEEE International Conference on Web Services (ICWS). IEEE, 2017: 25-32.

[15] Nedelkoski S, Cardoso J, Kao O. Anomaly detection from system tracing data using multimodal deep learning[C]//2019 IEEE 12th International Conference on Cloud Computing (CLOUD). IEEE, 2019: 179-186.

[16] Chen Y, Yang X, Lin Q, et al. Outage prediction and diagnosis for cloud service systems[C]//The World Wide Web Conference. 2019: 2659-2665.

[17] Reguieg H. Using mapreduce to scale event correlation discovery for process mining [D]. Université Blaise Pascal-Clermont-Ferrand II, 2014.

[18] Rosa A, Chen L Y, Binder W. Understanding unsuccessful executions in big-data systems[C]//2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. IEEE, 2015: 741-744.

[19] Weng J, Wang J H, Yang J, et al. Root cause analysis of anomalies of multitier services in public clouds [J]. IEEE/ACM Transactions on Networking, 2018, 26(4): 1646-1659.

[20] Raju B K, Geethakumari G. Event correlation in cloud: a forensic perspective [J]. Computing, 2016, 98(11): 1203-1224.

[21] Scheinert D, Acker A, Thamsen L, et al. Learning Dependencies in Distributed Cloud Applications to Identify and Localize Anomalies[J]. arXiv preprint arXiv:2103.05245, 2021.

[22] Khan S, Parkinson S. event log analysis: an association rule mining and automated planning approach[J]. Expert Systems with Applications, 2018, 113: 116-127.

[23] Fedorchenko A, Kotenko I, El Baz D. Correlation of security events based on the analysis of structures of event types[C]//2017 9th IEEE international conference on intelligent data acquisition and advanced computing systems: technology and applications (IDAACS). IEEE, 2017, 1: 270-276.

[24] Li G, De Carvalho R M, van der Aalst W M P. Configurable event correlation for process discovery from object-centric event data[C]//2018 IEEE International Conference on Web Services (ICWS). IEEE, 2018: 203-210.

[25] Jiang C, Han G, Lin J, et al. Characteristics of co-allocated online services and batch jobs in internet data centers: a case study from Alibaba cloud[J]. IEEE Access, 2019, 7: 22495-22508.

[26] Nguyen V, Dang T. CloudTraceViz: A Visualization Tool for Tracing Dynamic Usage of Cloud Computing Resources[C].2019 IEEE/ACM Industry/University Joint International Workshop on Data-center Automation, Analytics, and Control (DAAC). IEEE, 2019: 1-6.

[27] Bhattacharyya A, Singh H, Jandaghi S A J, et al. Online characterization of buggy applications running on the cloud[C]. 2016 12th International Conference on Network and Service Management (CNSM). IEEE, 2016: 282-286.

[28] Jassas M, Mahmoud Q H. Failure analysis and characterization of scheduling jobs in google cluster trace[C].IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society. IEEE, 2018: 3102-3107.

[29] Rosa A, Chen L Y, Binder W. Catching failures of failures at big-data clusters: A two-level neural network approach[C]. 2015 IEEE 23rd International Symposium on Quality of Service (IWQoS). IEEE, 2015: 231-236.

[30] El-Sayed N, Zhu H, Schroeder B. Learning from failure across multiple clusters: A trace-driven approach to understanding, predicting, and mitigating job terminations[C].2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). IEEE, 2017: 1333-1344.

[31] Tang H, Li Y, Jia T, et al. Analysis of Frequently Failing Tasks and Rescheduling Strategy in the Cloud System[J]. International Journal of Distributed Systems and Technologies (IJDST), 2018, 9(1): 16-38.

[32] Pang G, Shen C, Cao L, et al. Deep learning for anomaly detection: a review [J]. ACM Computing Surveys (CSUR), 2021, 54(2): 1-38.

[33] Keller F, Muller E, Bohm K. HiCS: High contrast subspaces for density-based outlier ranking[C]//2012 IEEE 28th international conference on data engineering. IEEE, 2012: 1037-1048.

[34] Lazarevic A, Kumar V. Feature bagging for outlier detection[C]//Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining. 2005: 157-166.

[35] Azmandian F, Yilmazer A, Dy J G, et al. GPU-accelerated feature selection for outlier detection using the local kernel density ratio[C]//2012 IEEE 12th International Conference on Data Mining. IEEE, 2012: 51-60.

[36] Pang G, Cao L, Chen L, et al. Sparse modeling-based sequential ensemble learning for effective outlier detection in high-dimensional numeric data[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018, 32(1).

[37] Pang G, Cao L, Chen L, et al. Learning Homophily Couplings from Non-IID Data for Joint Feature Selection and Noise-Resilient Outlier Detection[C]//IJCAI. 2017: 2585-2591.