Evolution of Network Configurations: High-level Analysis of an Operational IP Backbone Network

Fuliang Li[1,2], Jiahai Yang[1,2,*], Huijing Zhang[3], Suogang Li[4], Xingwei Wang[5], Jianping Wu[1,2]
[1]Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing, 100084, China
[2]Tsinghua National Laboratory for Information Science and Technology (TNList), China
[3]Department of Computer Science, Beijing University of Posts and Telecommunications, China
[4]CERNET National Network Center, Beijing, 100084, China
[5]College of Information Science and Engineering, Northeastern University, Shenyang, China

Abstract—**In this paper, we gather the weekly reports of an operational IP backbone network from January 2006 to January 2013, according to which, we can restore the truth and uncover the evolution of network configurations of the studied network. Our high-level analyses illustrate that rate limiting and launching routes for new customers are most frequently configured. We can identify and construct configuration templates by correlating each task to a certain set of commands in configuration files, based on which, automated configuration provisioning for an operational backbone network is feasible. In addition, we can configure redundant links for those with higher rate of failures according to our detailed analyses of link failures, which will enhance the stability and reliability of data transmission.**

Keywords—network measurement; configuration evolution; high-level analysis; backbone network;

I. INTRODUCTION

Misconfigurations are main reasons for network interruption and anomalies [1-4]. Furthermore, network devices tend to support more functions and provide more services, which will result in more configuration errors. Therefore, configuration analysis and management have become important research topics.

Many analyses of network configurations have been conducted to address configuration properties and diagnose misconfigurations. H. Kim et al. [5] analyzed five-year data of router, switch and firewall configurations from two large campus networks, and presented a long-term longitudinal study on the evolution of network configuration, such as how configurations of each task evolve, which parts of configurations change more frequently, etc. Y. Sung et al. [6] investigated dynamics of configurations of five enterprise VPN customers and presented the sets of portions changing most frequently. N. Feamster et al. [7] conducted a statistical analysis of router configurations to identify misconfigurations in BGP. F. Le et al. [8] adopted data mining technologies to extract policies of router configurations, which would be utilized as test cases to detect configuration errors. Based on a snapshot of the configurations, G. Xie et al. [9] conducted a statistical analysis of network reachability, which can be used to troubleshoot reachability problems and perform *what-if* analysis. Related analyses have proved that misconfiguration is the main reason for network outage and anomalies [1-4]. T. Benson et al. proposed two models to quantify the complexity of configurations [10, 11]. They not only captured the difficulty of configuring control and data plane behaviors of routers, but also unraveled the complexity of configuring network-based ISP services.

Network configuration is a complicated and error-prone job that lays heavy burden on operators, thus extensive studies have been performed to achieve the goal of automated configuration provisioning. D. Caldwell et al. [12] developed a tool to identify configuration policies and templates, which can drive the migration of the network toward automated provisioning. J. Gottlieb et al. [13] put forward the template-driven approaches for ISPs to configure connections for new BGP-speaking customers automatically. W. Enck et al. [14] proposed a system called PRESTO, which can construct device-native configurations based on the composition of function templates. They also presented the experience of configuring VPN and VoIP services. X. Chen, et al. [15] put forward a mechanism for configuration management by taking advantage of database to abstract specific configurations and adopting declarative language to describe dependencies and restrictions among network components.

In this paper, we make a first-ever and high-level analysis on configuration evolution of an operational IP backbone network. According to the dataset gathered from January 2006 to January 2013, we can not only capture the configuration tasks that are more frequently configured, but also help to extract configuration templates for automatic configuration provisioning through associating high-level configuration tasks with specific configuration commands in configuration files. In summary, we gain a comparatively comprehensive understanding of configuration evolution of an operational IP backbone network, which we believe, can provide a practical basis for research on automated configuration.

The rest of the paper is organized as follows. We begin in section II by introducing data sources used in our study. Section III presents our methodology and high-level analysis of the configuration evolution. We conclude in section IV with a summary of our contributions.

II. DATA SOURCES

1) Weekly Reports. Our high-level analyses of configuration evolution are based on the weekly reports of China Education and Research Network (CERNET in short) [17]. The weekly reports contain abundant and valuable information of operating an IP backbone network and they are mainly comprised of two aspects. For one thing, they record the tasks that have been processed by operators. The tasks include

adding a new customer (such as a university, a research institute, etc) to access the Internet, stopping a customer from accessing the Internet, rate limiting, etc. For another, they record the failures of the backbone network. The failures include link failure, device failure, power failure, etc. We gather the weekly reports of the studied network from January 2006 to January 2013. According to the weekly reports, we can track what configuration behaviors dominate the configuration operations and which configuration tasks are performed most frequently, i.e., we can restore the truth and uncover the evolution of network configurations.

2) Configuration Files. The studied network consists of more than 100 backbone routers. Operators backup the configuration files of these routers through a self-developed system. Note that only the changed configuration files and the differences between two continuous versions are recorded. Once we make a mistake on configuration, we can quickly roll back the router to the most recent valid condition. In addition, we can simply identify and diagnose the misconfigurations based on the differences recorded. The configuration files are regarded as an auxiliary data source to the weekly reports, because each task can be associated with a certain set of commands that are manually implanted in configuration files.

III. METHODOLOGIES AND HIGH-LEVEL ANALYSIS RESULTS

A. Methodologies

Capturing the configuration evolution can shed direct light on which tasks are frequently performed on routers. Unlike common methodologies that extract changes by comparing two consecutive revisions or by analyzing the data generated by configuration management tools, we capture the evolution by tracking high-level configuration tasks recorded in the weekly reports. Table I shows the types of configuration tasks that can be tracked from the weekly reports. In addition, we correlate each high-level task to a certain set of commands that are manually implanted in configuration files, based on which, we associate the tasks with three types of basic operations (i.e., modification, addition and deletion) on configurations. Our method is more direct and efficient to investigate the evolution of network configurations. However, our methodology may be a more general and coarse-grained classification method. For example, modifying a routing policy may also involve adding or deleting operations.

TABLE I. TYPES OF CONFIGURATION BEHAVIORS THAT CAN BE TRACKED

Task	Introduction	Type
Routing related	Adjust the parameters of OSPF, BGP, etc.	Modification
Rate limiting	Increase or lower the access bandwidth for customers	
Suspending customer	Stop a customer from accessing the Internet	
Rebooting customer	Re-allow a customer to access the Internet	
Adding customer	Add a new customer to access the Internet	Addition
Canceling customer	Revoke the privilege of a customer to access the Internet	Deletion

B. High-level Analyses of Configuration Evolution

1) Analyze the distribution of configuration tasks

As depicted in Fig.1, rate limiting accounts for the largest percentage which is about 37.04% of the configuration changes. Task of rate limiting refers to increasing or decreasing the access bandwidth for customers. As the applications and services demand higher network performance and lower delay, customers often have the requirements of bandwidth upgrading. In addition, to maximize their profits, customers will access the Internet in a multi-homing way, so they may decrease the bandwidth of existing ISPs, which can save money for accessing a new ISP to satisfy special needs. Operators need to make adjustment of bandwidth in order to satisfy different requirements of customers at different times. Configurations of adding customer account for 36.78% of the total configuration changes. Adding customer mainly refers to adding a new customer to access the Internet through the studied network. Benefiting from the rapid development of the studied network, there are more and more customers who access the Internet through the studied network either for the high speed or for the abundant resources in the studied network. Furthermore, suspending customer refers to temporarily preventing a customer from accessing the network without paying in time, while rebooting customer refers to re-activating a customer to access the Internet through the studied network. These two types of configuration tasks account for 15.73% of the total configuration changes. Configurations of routing related account for about 7.81% of the total configuration changes. The routing related configurations mainly include adjusting the parameters of OSPF for intra-domain load balancing and adjusting the parameters of BGP for inter-domain load balancing. Routing related tasks also include adding new physical links, which will cause changes of network topology. Routing adjustments are complicated and error-prone jobs. Fortunately, configurations of routing related are not frequent. Tasks of canceling customer account for the smallest percentage which is only about 2.64% of the total configuration changes. With the increasing number of customers who are willing to access the Internet through the studied network and the fewer customers who are revoked the privilege to access the Internet, we can expect that the scale of the studied network is still expanding.

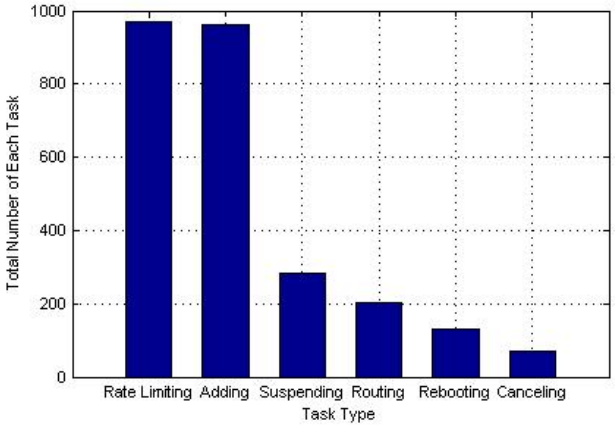

Fig. 1. Distribution of configuration tasks over seven years

2) Analyze the changes of configuration commands

We roughly correlate each high-level task with a basic configuration operation as shown in Table I. Then we correlate each high-level task with specific commands executed in configuration files. For example, operation of adding customer will add about eight configuration command lines. We list the total number of command lines of these three types of operations. As shown in Table II, operators perform adding operations in most cases, which is conformant with the conclusion we gain from Fig.1. As we have illustrated earlier in Table I, adding operation mainly refers to configurations of adding customers. Due to the rapid development of the studied network, more and more customers want to get the permission to access the Internet through it. We can gain a clear impression on this tendency through the data we collected. With the number of the customers increasing, more and more traffic is generated, which leads to congestion in some links. So operators will add more links or upgrade the existing links among the backbone routers, as well as adjust the routing policies. Adding new links will initiate a chain of adjustments on routing policies that could make the links more evenly used. Adding operation needs more command lines to be dealt with as reflected in Table II, thus it is a more complex configuration operation. Due to its complexity, it is quite crucial for operators to be careful when they conduct adding operations for the purpose to reduce the possibility of misconfiguration as much as possible. Furthermore, modifying operations also account for a considerable part among the basic operations. Modifying operation mainly refers to configurations of routing related and rate limiting, etc. Considering routing related tasks, operators need to make a better choice of route selected for both inter-domain and intra-domain routing in order to realize tradeoff between cost and performance, and simultaneously gain the maximal profit. Overall, routing related tasks are complicated, and they play a quite crucial role in a large scale network operation. In addition, tasks of rate limiting also account for a large percentage of modifying operations. Customers of the studied network are faced with the number of users in the customer networks growing, high performance and low delay of special applications or services demanding, maximal profits of the customer networks pursuing, etc. All these requirements need operations of rate limiting. Therefore, being aware of the importance of modifying operations which are complicated to be dealt with, much attention should be paid to them, so that configuration errors can be reduced as many as possible. Table II also shows that operators apply less deleting operations in our daliy administration due to few demands.

3) Longitudinal analysis of configuration tasks

Fig.2 shows longitudinal analysis of configuration tasks over seven years. Most configuration tasks follow the different patterns during these years. Taking rate limiting as an example, we can see that before the year of 2010 no rate limiting tasks were involved in our observation. However, a dramatic increase of this task seems to come into being after 2009. We can expect that due to the rapid development of the customer networks, requirements of high performance or choices of multi-homing lead to many changes on bandwidth selection. We also find that some configuration behaviors present a downward trend, such as adding customer and routing related. For one thing, the studied network was built to devote to an education and research network. Therefore, customers are mainly universities, research institutes, and government departments, etc. Customers of these types tend to be in a stable scale and new customers become less and less. However, since there are more tasks of adding customers than that of canceling customers, the size of the studied network is still increasing. For another thing, once an ISP is developed into a mature phase, the backbone network will be in a stable state, so configuration tasks related to routing adjustments decline. Note that operators did not record the tasks of routing related in weekly reports since 2010 because of the stylistic changes of the weekly report. In addition, when we observe Fig.2 along vertical axis, we can see that in different months of different years the amount of tasks varies. What's more, the amount of tasks can reach a climax in certain month each year. Therefore, operators need to be very careful and try their best to avoid misconfigurations while conducting operations during these months. Finally, Fig.2 shows that adding customer and rate limiting are most frequently configured. Operators can correlate a set of certain command lines in configuration files to these tasks in order to construct configuration templates. The templates can be transformed into executable scripts (including syntax of *expect*), which can configure the task on the corresponding devices automatically.

4) High-level analysis of link failure

As shown in Table III, network failures are primarily caused by link failures. Therefore, we must pay attention to link failures to guarantee the stability and reliability of data transmission. For a fault-tolerant network, once a link failure happens, redundant links for this link will be activated. In this section, we conduct a detail analysis of link failures, which is helpful to make intelligent routing policies that will be configured on routers for possible link failures.

TABLE II. CHANGE ANALYSIS OF CONFIGURATION COMMANDS OVER SEVEN YEARS

Basic configuration operation	Number of Command lines
Addition	7688
Deletion	552
Modification	1719

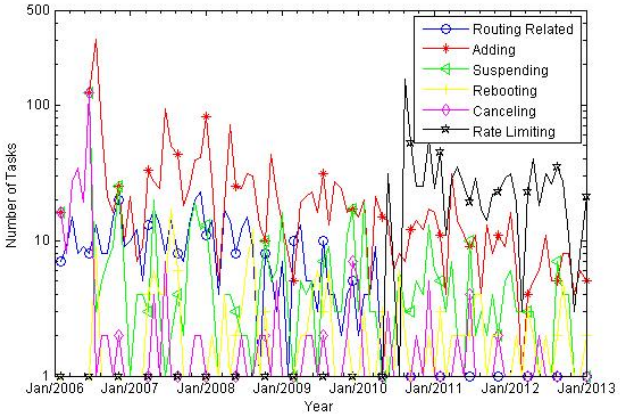

Fig. 2. Longitudinal analysis of configuration tasks over seven years

TABLE III. FAILURE TYPES AND THEIR RATIOS

Error Type	Ratio
Link failure	82.9%
Routing failure	8.1%
Power failure	4.6%
Device failure	3.4%
Others	1.0%

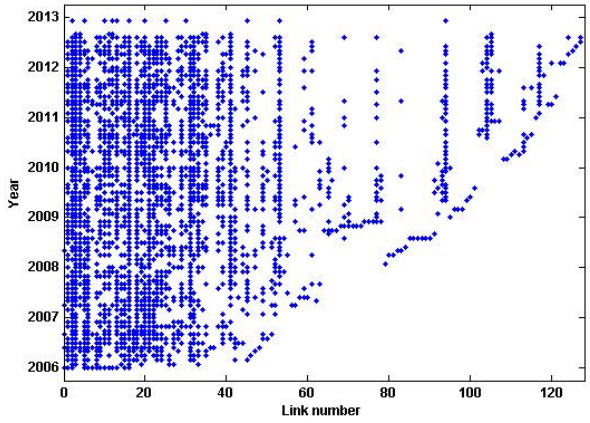

Fig. 3. Detailed analysis of link failures

As depicted in Fig.3, every point in the plot represents that a failure happened in the link during the past seven years. When we observe the figure along vertical axis, we can see that if the points are so dense that even a straight line is formed, we can infer that the corresponding link encounters a high rate of failures during our observation. Guaranteed continuous data transfer is crucial for ISPs because this will affect the credibility among the customers. Therefore, operators should configure redundant links for the links that are easy to break down in order to enhance the stability and reliability of data transmission. In addition, once we observe the figure along horizontal axis, it is obvious to come to a conclusion that the number of the links encountering link failures seems to show an increasing tendency during these years. Due to the rapid development of customer networks and the growth of new customers who access the Internet through the studied network, the traffic load of the links goes up dramatically. Some new links are added to share the traffic load. These new links may encounter failures. Operators also need to configure redundant links for these emerging links.

IV. CONCLUSIONS AND FUTURE WORKS

In this paper, we captured the configuration evolution of an IP backbone network in a high-level analysis way, i.e., correlating each high-level task recorded in the weekly report to specific command lines executed in the configuration files. The results showed that rate limiting and adding customers are most frequently configured. In addition, we conducted an analysis of network failures. We found that link failures are the main reasons for network failures and some links are easy to break down. Two key points can be gained from our observations. 1) we can identify and construct configuration templates by correlating each task that is frequently configured to a certain set of commands in configuration files, based on which, automated configuration provisioning for operational backbone network is feasible. 2) We can configure redundant links for the links with higher rate of failures according to our detailed analysis of link failures, which will enhance the fault tolerance of data transmission.

ACKNOWLEDGMENT

This work is supported by the National Basic Research Program of China under Grant No. 2012CB315806, the National Natural Science Foundation of China under Grant No. 61170211, 61202356, 61161140454, Specialized Research Fund for the Doctoral Program of Higher Education under Grant No. 20110002110056, 20130002110058, Tsinghua University Initiative Scientific Research Program under Grant No. 2012Z02151, Joint Research Fund of MOE-China Mobile under Grant No. MCM20123041, the National Science Foundation for Distinguished Young Scholars of China under Grant No. 61225012, 71325002.

REFERENCES

[1] R. Mahajan, D. Wetherall, and T. Anderson. Understanding BGP misconfiguration. In Proc. ACM SIGCOMM, pages 3–17, Pittsburgh, PA, Aug. 2002.

[2] Z. Kerravala. Configuration management delivers business resiliency. The Yankee Group, Nov. 2002.

[3] A. Markopoulou, G. Iannaccone, S. Bhattacharyya, et al. Characterization of Failures in an IP Backbone. In: Proc. of the Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies (INFOCOM), Hong Kong, 2004. 2307-2317.

[4] Oppenheimer D, Ganapathi A, Patterson D. Why do Internet services fail, and what can be done about it. In: Proc. of the 4th on USENIX Symposium on Internet Technologies and Systems (USITS), Seattle, WA, USA, 2003.1-15.

[5] H. Kim, T. Benson, A. Akella, and N. Feamster, "The evolution of network configuration: a tale of two campuses", in Proc. Internet Measurement Conference (IMC), 2011, pp.499-514.

[6] Y. Sung, S. Rao, S. Sen, and S. Leggett. Extracting Network-Wide Correlated Changes from Longitudinal Configuration Data. In Proc. PAM, Seoul, South Korea, Apr. 2009.

[7] N. Feamster and H. Balakrishnan. Detecting BGP Configuration Faults with Static Analysis. In Proc. 2nd USENIX NSDI, Boston, MA, May 2005.

[8] F. Le, S. Lee, T. Wong, H. Kim, and D. Newcomb. Minerals: using data mining to detect router misconfigurations. In Proc. MineNets'06, pages 293–298, Pisa, Italy, Sept. 2006.

[9] G. Xie, J. Zhan, D. Maltz, H. Zhang, A. Greenberg, G. Hjalmtysson. and J. Rexford. On static reachability analysis of IP networks. In IEEE INFOCOM, volume 3, pages 2170–2183, 2005.

[10] T. Benson, A. Akella, and D. Maltz. Unraveling Complexity in Network Management. In Proc. 6th USENIX NSDI, Boston, MA, Apr. 2009.

[11] T. Benson, A. Akella, and A. Shaikh, Demystifying configuration challenges and trade-offs in network-based ISP services. In Proceedings of SIGCOMM. 2011, 302-313.

[12] D. Caldwell, A. Gilbert, J. Gottlieb, A. Greenberg, G. Hjalmtysson, and J. Rexford. The cutting edge of IP router configuration. In Hotnets-II, Cambridge, MA, Nov. 2003.

[13] J. Gottlieb, A. Greenberg, J. Rexford, and J. Wang. Automated Provisioning of BGP Customers. IEEE Network, 2003.

[14] W. Enck, T. Moyer, P. McDaniel, S. Sen, P. Sebos, S. Spoerel, A. Greenberg, Y. Sung, S. Rao, and W. Aiello. Configuration management at massive scale: system design and experience. IEEE Journal on Selected Areas in Communications, 27(3):323–335, 2009.

[15] X. Chen, Y. Mao, ZM. Mao, and J. Van der Merwe, "Declarative configuration management for complex and dynamic networks", in Proc. of the conference on Emerging networking experiments and technologies (CoNext), 2010.

[16] CERNET - China Education and Research Network. http://www.edu.cn/english/.