Research and Implementation of Knowledge-Enhanced Information Services

Bo Yang[1], Hao Wang[1], Liang Liu[1], Qian Ma[1], Ying Chen[1], and Hui Lei[2]

[1] IBM China Research Laboratory, Beijing, 100094, China
{yangbbo, wanghcrl, liuliang, maqian, yingch}@cn.ibm.com
[2] IBM T. J. Watson Research Center, Hawthorne, NY, USA
hlei@us.ibm.com

Abstract. Information isolation has been identified as a big challenge in IT Service Management (ITSM). Existing ITSM practices mostly rely on configuration information and are geared towards individual applications and processes. However, information available in complicated IT infrastructure goes beyond data from the configuration management domain. How to efficiently extract and integrate the hidden knowledge from a wide variety of information sources is a major pain point for ITSM. In this paper, a threading strategy (TS) with KPI mark and knowledge-enhanced information services is proposed to improve ITSM quality. The essential contribution of this work is to organize the highly complex IT service information with KPI mark and to build a knowledge repository for accumulateing and reusing experts' knowledge. In addition, a prototype called BIANCHIN is implemented to explore this knowledge-enhanced information services framework. Finally, a real business application of Cisco VoIP system is used as a case study for evaluating the effectiveness and efficiency of the knowledge-enhanced information services framework.

Keywords: IT service management, information services, configuration management database, knowledge database.

1   Introduction

Information services are critical to IT infrastructure management. They provide diverse information to users or other service components in an IT service management (ITSM) environment [1]. Information services constitute a new level of services that offer added value to information contained in data sources across an organization [2]. They integrate information to provide a unified view of information, add business context to raw information, and expose insightful relationships in information that in turn facilitate better decision-making.

Treating information as a service, organizations can improve the relevance and cost effectiveness of their information by reusing integration logic, making information available to people, processes and applications across the business, and improving the operational impact of information on driving innovation.

Many approaches have been proposed to improve the utility of information services. Jie, W. et al.[3] proposed an information service architecture model for information management in a Grid Virtual Organization, which is a hierarchical structure that consists of a VO layer, a site layer and a resource layer. In order to satisfy customers' individual demands based on their personal differences, Wang, J. et al. introduced a decision-tree approach of data mining to get special information demands, used agent technology to establish the model of an information service terminal and defined the functions of the components [4]. Zou, H. et al.[5] proposed a hybrid resource information service architecture based on the grid-monitoring architecture to promote the validity of the resource information service with low system cost. Lu, X.D. proposed a distributed information service system architecture [6]; Lu also defined the ratio of correlation and the degree of satisfaction and proposed the autonomous integration and optimal allocation of information services for heterogeneous Faded Information Fields. Zang, T. et al. presented an architecture of the information service and the models of information organization [7]. The main functionality of this information service is the provisioning of information essential to applications running in a distributed environment such as resource information, job status, resource workload, service meta-information, and queue status.

However, information management and organization in complicated ITSM environments with frequent changes is a challenging issue [8]. Information coming from different sources is characterized as diverse, dynamic, heterogeneous, and geographically distributed. In IT infrastructure management, change is much more accelerated, and what actually defines an enterprise is indeed morphing, becoming more fractured and distributed, engaging more third parties and stakeholders within their respective business value chains expanding across the globe. Effective collaboration within and beyond the various information has become both necessary and more difficult to manage. Information as a service needs to be more readily accommodated in an integrated and proactive fashion rather than via one-off efforts.

Our work focuses on a high-level information service which is enhanced by introducing key performance indicator (KPI) mark and knowledge database (KDB). It provides information essential to applications running in a complex IT infrastructure environment including resource information, service status and service dependence. This information is organized in relational models based on a threading strategy with time thread. The information service works with data capture and analysis systems to support resource discovery, job scheduling, and management visibility.

The rest of the paper is organized as follows. The problems confronted by information services are analyzed in Section 2. In Section 3, we illustrate the framework of our Threading Strategy for ITSM (TS-ITSM). Section 4 describes the implementation issues. An experiment with a real application is discussed in Section 5. Finally, the conclusions and future work are presented in Section 6.

2   Problem Analysis

ITSM has received growing attention from both the academia and the industry. An important and challenging subject in ITSM is information as a service. Information services are characterized by their wide distribution, high fault tolerance and dynamic functions as well as diversified forms.

A recent survey by IDC with corporate executives reveals that the executives require access to trusted and reliable information in a timely manner [8]. However, most enterprises are flooded with large scale data and content scattered in many systems and sources, and in multiple forms. The volume and variability of such information continues to increase, including application configurations, network configurations, OS configurations, service status, CPU usage, memory usage, transaction workload, transaction response time, etc. Sharing information and ensuring that the most appropriate views are discovered and used for their intended and changing purposes can be daunting given the many layers of hard-coded and semantic dependencies built within typical applications and systems. Furthermore, it is quite inefficient and disconcerting that different applications apply their respective approaches in a very fragmented, redundant, and inconsistent manner.
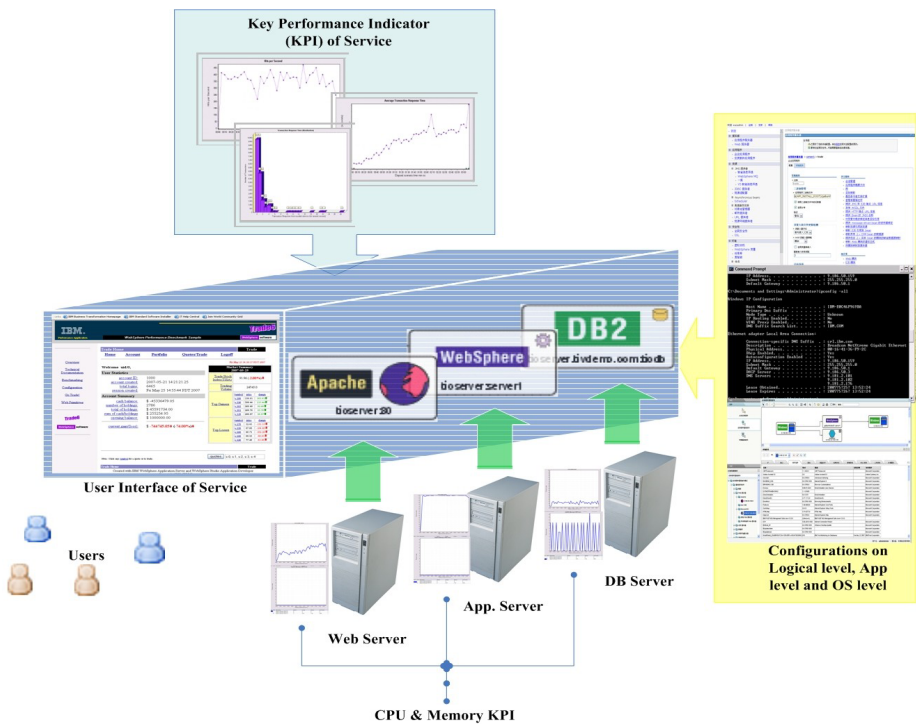

Fig. 1. Typical IT information architecture of distributed business applications

A critical step towards improving ITSM involves creating solutions geared for discrete applications or processes, this evaluation involves creating and instantiating the core elements and functions of the business in a fashion where they can be utilized in multiple ways. It allows one to view all of the components of the environment in a logical organization, as shown in Fig. 1, that is not constrained by any particular physical implementation or use scenario. To maintain an application, all related information need to be captured for performance analysis or problem diagnosis. Each information source is useful in context to its particular initial use case yet can

potentially provide tremendous added value when it is combined and utilized for multiple purposes.

For centralizing configurations control, the configuration management database (CMDB) [9-11] has been proposed that focuses on how organizations are positioned to extract value and raise competencies to address their unique information requirements. The concepts underlying CMDB include information governance, change management, as well as the development and maintenance of a flexible information infrastructure. CMDB is intended to be an infrastructure approach to coordinating data-oriented service and integration functions in a dedicated fashion. It provides connectivity to a vast amount of data and delivers relevant information, consolidating these functions in a unified fashion as shown in Fig. 2.
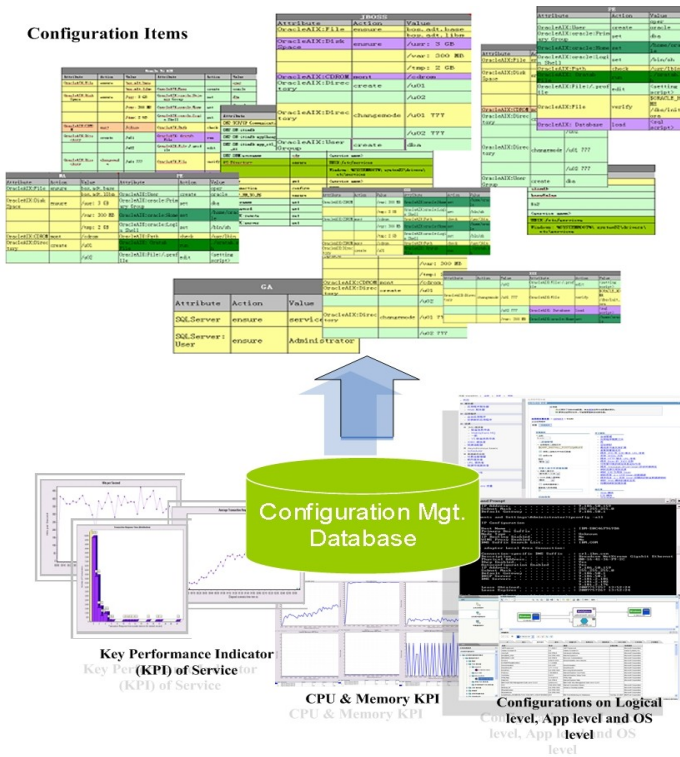

Fig. 2. Configurations management database for information service

However, applying a common form to all information sources within an enterprise is an impossible task, especially in large and changing environments [8]. Given uncertainties on the exact information that may be needed in the future, who will need it, and how it will be used, it is critical to build an information foundation that is open, flexible, and scalable. Furthermore, it is not sufficient simply to record all the information on systems and operations in CMDB. Without the assistance of domain knowledge, ITSM personnel can easily get confused and be lost in low-level redundant details.

3   Threading Strategy for ITSM

When studying how to provide real-time information, we must consider how to satisfy diverse demands from disparate applications. The concept of advanced information services has been put forward from the aspect of services. Without a common foundation of usable information, service-oriented architecture is just a loose confederation of abstract business processes.

   Since information from a complicated IT infrastructure is not the only data needed for configuration management, in this paper, we advocate a threading strategy (TS) with KPI mark and knowledge-enhanced information services. The KPI mark represents the performance status of system. To organize a large body of information coming from diverse sources, TS introduces time-thread based data management to coordinate all information in an open, flexible, and scalable style. Each source can be added or removed from the information foundation by changing it on the time thread. Every information chip is classified into a version according to its capture time. As shown in Fig. 3, all information around a time point is regarded as a version of related system description.
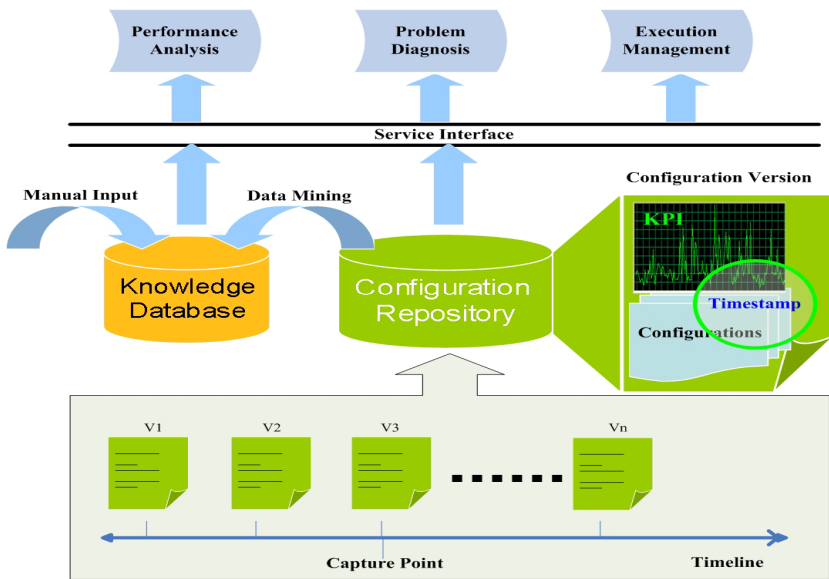

Fig. 3. Threading strategy with KPI mark and knowledge-enhanced information services

   In particular, for a version, the KPI mark of the system performance provides tremendous added value to other application services such as capability planning, performance analysis and problem diagnosis etc. It is almost impossible for a traditional CMDB containing only configuration information to provide such added value. KPI mark also provides the change trace of system performance, which can indicate what configurations should be used to meet special business requirements when the KPI mark is combined with configuration information.

Moreover, a knowledge database is introduced to store not only the KPI rank of system versions, but also extended patterns and rules that ITSM operators define in the course of their work. In a distributed environment, this strategy enables ITSM operators to share their domain knowledge for different application services. The reuse of experts' knowledge will be effective in reducing the labor cost for complicated IT service management.

4   Implementation of the Information Service

In this section, a proof of concept (POC) project BIANCHIN is developed to validate the Threading Strategy for information services. It is built on the Eclipse Toolkit with Java technology, and implements a container for information capture, data organization, pattern definition, rule and policy definition, and the GUI of using knowledge.
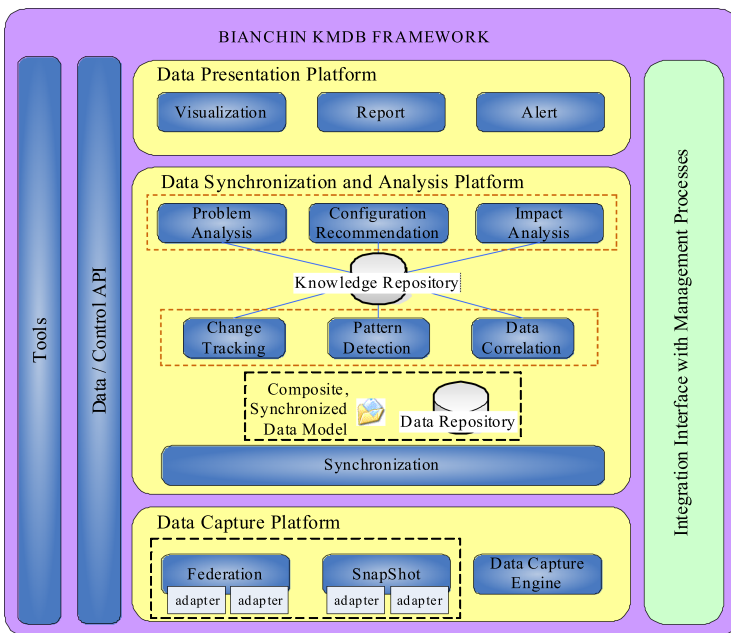

Fig. 4. Prototype framework about knowledge enhanced information service

In the following subsections, we will discuss some key issues involved in the implementation of our information service. These issues include the implementation of the information capture mechanism, the implementation of the data organization mechanism, the implementation of the knowledge accumulation and so on.

4.1   Threading Strategy for Information Coordination

The framework of BIANCHIN is shown in Fig. 4, in which a data capture platform, a data synchronization & analysis platform, and a data presentation platform are constructed to provide information services.

After data is captured from target systems, it needs to be coordinated to represent the target systems. For data coming from diverse sources with different capture schedule, a loose composite data model is proposed to synchronize data by timeline, compose the diverse data, and store the composed data in a data repository. A synchronized data model in xml is introduced in the prototype. All of the data in a version represents the system status at the time point. Thus it provides a comprehensive, synchronized view of the target business-IT infrastructure by aligning diverse information.

4.2 Knowledge-Enhanced ITSM

Moreover, and most importantly, the knowledge repository provides fundamental analysis functions upon the synchronization of data to facilitate knowledge generation. Frequently used data analysis functions such as comparison and change tracking, pattern detection and search, statistical data correlation, and KPI mark are provided as utility libraries. Comparison makes it possible to discover what is common among all the situations where the system can meet the SLA requirement, based on historical versions of system status. Change checking helps problem diagnosis when a specific service goes down from a healthy state. Pattern detection will improve problem analysis by recognizing the change in configuration items that cause a known problem. And statistical data correlation analysis computes the possibility that a change in configuration items (CIs) will impact the availability of the system.

When an expert diagnoses or solves a problem successfully, he can use "pattern definition" to file his solution in knowledge repository. Based on the xml data model in the BIANCHIN prototype, an XPath like the one below could be used in pattern definition:

//ServiceState:Adapter/ServiceState:Service_States/Service[Display_Name='Cisco CallManager' and State='Stopped']

where the XPath has two conditions: one is that the "Display_Name" element should equal a string "Cisco CallManager" and the other is that the "State" element should equal a string "Stopped". The conjunctions in pattern could be "and", "or", or "!". And the conditions could be any of "=", ">", "<", ">=", ">=", and "contains ()".

With fundamental analysis functions, the expert's knowledge of problem resolution is stored in the knowledge repository. This allows for effective knowledge reuse on performance analysis, configuration recommendation and problem diagnosis etc., and will add business value to other information services.

In our system, data analysis not only can track system status changes in the form of a configuration change, but also those in the form of KPI fluctuation at business level. That will be powerful to analyze the relationships between IT infrastructure and business requirement, and will be helpful to diagnose a problem's root cause.

5   Experiment Result and Analysis

The purpose of the experiments is to evaluate the efficiency of problem determination using Threading Strategy for ITSM, when TS is used for problem diagnosis on a real

business application of Cisco VoIP system. The VoIP system contains a complicated services topology, as shown in Fig. 5, where the system is supported by VoIP application services, database application services and OS services. And there are many relationships between services; for example, Cisco CDR Insert (CCI) service and Cisco Database Layer Monitor (CDLM) service depend on MSSQLSERVER service, Distributed Transaction Coordinator (DTC) service and Remote Procedure Call (RPC) service. Any related service that is blocked will influence the performance of the overall VoIP system. The information sources in our experiment include VoIP application configurations, OS configurations and services status.
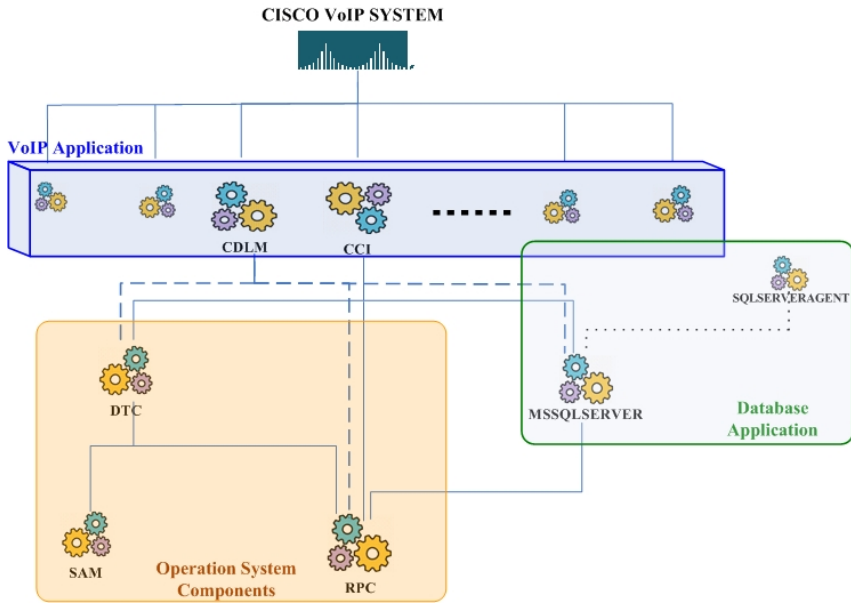

Fig. 5. Part of service relationships in the Cisco VoIP system

The experiment is designed to include 15 troubleshooting cases, Part of the cases are frequent cases cited in the Cisco CallManager trouble-shooting manual [12] and the others are summarized by experienced VoIP administrators. The cases covered both system problems, such as service operation error, and VoIP configuration problems such as improper device settings.

For each case, we examined the efficiency of Problem Determination and problem Remediation (PDR)[13-15] efficiency. The efficiency was measured in terms of the time cost of PDR. For comparative analysis, each troubleshooting case is performed using two methods: one is the traditional method which is purely manual diagnosis by a VoIP system administrator, and the other is diagnosis with assistance from our TS-ITSM prototype system, BIANCHIN. The experiments results are summarized in Table 1 and Figure 6.

Table 1. Summary of all the experiment results

	PURE MANUAL DIAGNOSIS	DIAGNOSIS WITH BIANCHIN
UNSOLVED PROBLEMS	3	0
SUCCESS RATES (%)	80	100
MAX PDR TIME (MIN)	55.63	18.1
MIN PDR TIME (MIN)	1.13	1.12
AVERAGE TIME (MIN)	20.46	7.82

Table 1 summarizes the respective experiment results of problem diagnosis by a VoIP administrator with and without the use of BIANCHIN. The columns represent the two diagnosis methods for the same problem, and the rows involve 2 main performance measures that are of interest to us: the success rate of PDR and the average time cost of PDR. The number of solved problems and the time cost for each case are recorded, as shown in Figure 6. For comparison, the number of solved problems as well as the maximum, minimum, and average of the time costs are presented in Table 1. We can observe that the percentage of problem diagnosis success rate is rather high when BIANCHIN is used by an administrator for PDR, and that the average time cost for PDR is lower than that for purely manual diagnosis. This indicates that information on the changes of CIs in the system is very helpful for an administrator to determine and resolve problems.

The MAX time costs summarized in Table 1 are very different across experiment cases, but the MIN time costs are very close to each other. This may be due to the fact that the MIN time costs correspond to situations where the problems can be very quickly diagnosed and the time costs are primarily time spent on fixing the problem, which is more or less the same for different cases. When problem determination is more complicated, the time cost will rise accordingly, resulting in disparities in the MAX time costs.

Figure 6 displays a diagram that compares the efficiency between purely manual problem diagnosis and BIANCHIN-assisted diagnosis. The diagram shows the distribution of the time cost for each of the 15 cases. The cases that cannot be solved in 1 hour (3600 seconds) were marked as unsolved problems in the experiments.

As shown in Figure 6 and summarized in Table 1, 3 problems cannot be resolved in purely manual diagnosis experiments, which are cases No. 11, No. 14 and No. 15. In comparison, the root cause of the problems was successfully determined and the problem resolved with BIANCHIN. Note that although the change management of CIs improves the success rate of problem determination (PD), it does not always improve the efficiency of PD. Cases 5, 8, 9, 12 and 13 indicate that the change management of CIs may increase the time cost of PD for an experienced administrator because he has to analyze a large amount of information about configuration changes to determine which change is the right root cause for the current problem. This process could be time-consuming. However, the cases can be solved efficiently when patterns of the problem has been accumulated in the knowledge repository, such as in case 2, 3 and 15. In those cases, BIANCHIN determined the problem root cause accurately according to pre-defined problem patterns, and provided refined information to the administrator leading to very efficient PDR.
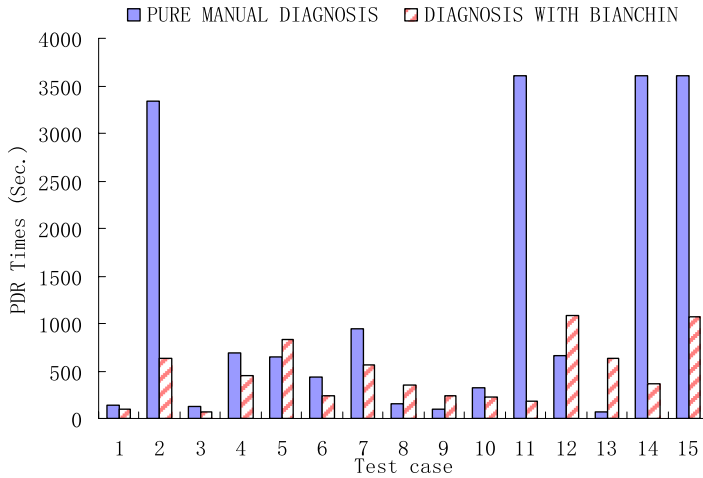
Fig. 6. PDR time cost for each tested case

Among the successfully resolved problems, case 2 is the most time-consuming one. It is a problem where services of VoIP such as CCM, CTI Manager, Extended Functions, and Voice Media Streaming have failed. When such a problem occurs, a VoIP phone user only observes that the IP phone is not working; and a VoIP administrator may observe that the administration page of VoIP does not display and some services cannot be started. In general, there are 3 possibles causes for such symptoms: the system user SQLSvc has been deleted unexpectedly, the password of user SQLSvc has been changed, or the hostname of the VoIP server has been changed.

In purely manual diagnosis, the administrator had to check every cause for PDR because there was no change management tool available to track changes of CIs. Unfortunately, testing SQLSvc password change was a long operation in the VoIP system. And it turned out that the real reason was that hostname of the VoIP server had been changed. It also took some time for the administrator to determine the changes on hostname as the difference between the names was subtle.

When there is no prior knowledge in the knowledge base as is the case with traditional ITSM, the administrator will be presented with excessive change information to make his decision. The information includes the states of services that are different, the change of application functions for the applications depending on the services, the error and warning events recorded in system and application log files, and of course, the configuration change of the hostname. Such information is useful to the administrator for PDR, but it is not time or cost efficient to manually extract the problem root cause from the large volume of information.

Given the same problem, BIANCHIN provides a more advanced approach to leveraging experts' knowledge. There are often repetitive use cases in the experiments. Patterns are defined in the knowledge base when the problem has been resolved successfully or when the case reveals critical insight on troubleshooting. In our experiment, the case about hostname change and its symptoms has been defined as a pattern in the BIANCHIN knowledge base. When the pattern about hostname change

occurs in case 2, BIANCHIN will list the applicable patterns for the problem in the control panel automatically. It highlights the pertinent information to help the administrator to determine the problem root cause quickly. If the unsolved problem is excluded, our knowledge-enhanced method has the largest improvement in case 2, an 81.2% improvement comparing to pure manual method. And the PDR time cost has also been improved through pattern matching, as in cases 2, 3 and 15 shown in Figure 6.

6   Conclusion and Future Work

IT is not just about providing computing technologies, but also about providing services to end users. Such recognition is drving the development of IT service management, which has become an ever important discipline. To materialize the notion of ITSM, concrete methods need to be defined and their effectiveness evaluated. Our investigation of an ITSM architecture contributes to this rapidly growing area and suggests new possible research directions.

In this paper, we have analyzed and evaluated the performance of ITSM when it is used in problem determination. From the problem-determination experiments on the VoIP system, we can conclude that knowledge-enhanced ITSM is effective and efficient, and provides significant improvent on experts knowledge reuse and problem root-cause determination. Once the information analysis phase has been completed, additional knowledge is available to suggest solutions for other services. When a special business requirement entails a change of configurations, a version of the configurations will be extracted from the configuration repository if the KPI is likely to meet the requirement. This method could also be used for system disaster recovery, where the system is restored to a specific state in history that is stable and controllable.

Moreover, for those more complicated problems that reference a large amount of configuration data, TS-ITSM may be suitable to filter out information irrelevant for system diagnosis and remediation. Obviously, knowledge-enhanced ITSM needs to be further developed and refined for applications in a wide variety of cases. This paper represents our initial effort. Further investigation will be conducted and reported in the future.

References

1. Information Technology Service Management (ITSM) (2005), Available: http://www.cce.umn.edu/professionalcertification/itil/
2. IBM Information On Demand - The Role of Information in a Service Oriented Architecture: IBM Global CFO Study (2006), Available: http://www-306.ibm.com/ software/data/information-on-demand

3. Jie, W., Hung, T., Turner, S.J., Cai, W.: Architecture Model for Information Service in Large Scale Grid Environments. In: Sixth IEEE International Symposium on Cluster Computing and the Grid, vol. 1, pp. 107–114. IEEE Computer Society Press, Los Alamitos (2006)

4. Wang, J., Ding, Z.F., An, S.: An agent-based study on personalized travel information service. In: 1st International Symposium on Systems and Control in Aerospace and Astronautics, p. 4 (2006)

5. Zou, H., Jin, H., Han, Z.F., Shi, X.H., Chen, H.H.: HRTC: hybrid resource information service architecture based on GMA. In: IEEE International Conference on e-Business Engineering, pp. 541–544. IEEE Computer Society Press, Los Alamitos (2005)

6. Lu, X.D., Mori, K.: Autonomous information services integration and allocation in agent-based information service system. In: IEEE/WIC International Conference on Intelligent Agent Technology, pp. 290–296 (2003)

7. Zang, T.Y., Jie, W., Hung, T., Lei, Z., Turner, S.J., Cai, W.T.: The design and implementation of an OGSA-based grid information service. In: IEEE International Conference on Web Services, pp. 566–573. IEEE Computer Society Press, Los Alamitos (2004)

8. Rogers, S.: Information as a Service to the Enterprise. White paper (December 2006)

9. Van Bon, J., Kemmerling, G., Pondman, D.: IT Service Management: An Introduction, Van Haren Publishing (September 1, 2002)

10. Berkhout, M., Harrow, R., Johnson, B., Lacy, S., Lloyd, V., Page, D., van Goethem, M., van den Bent, W.G.: Service Support: Service Desk and the Process of Incident Management, Problem Management, Configuration Management, Change Management and Release Management, London: The Stationery Office (2000)

11. Chen, P.Y., Kataria, G., Krishnan, R.: On Software Diversification, Correlated Failures and Risk Management. SSRN (April 8, 2006), Available: http://ssrn.com/abstract=906481

12. Troubleshooting Guide for Cisco CallManager, Release 4.2, Corporate Headquarters, Cisco Systems, Inc. (2006)

13. Chen, M.Y., Kiciman, E., Fratkin, E., Fox, A., Brewer, E.: Pinpoint: problem determination in large, dynamic Internet services. In: International Conference on Dependable Systems and Networks, pp. 595–604 (2002)

14. Hoi, C., Kwok, T.: An Autonomic Problem Determination and Remediation Agent for Ambiguous Situations Based on Singular Value Decomposition Technique: In: International Conference on Intelligent Agent Technology, pp. 270–275 (2006)

15. Agarwal, M.K., Gupta, V.M., Sachindran, N., Anerousis, N., Mummert, L.: Problem Determination in Enterprise Middleware Systems using Change Point Correlation of Time Series Data. In: Network Operations and Management Symposium, pp. 471–482 (2006)