Applying Time-Lapse Concepts onto Storage System for Long-Term System Trace Analysis: Technical Challenges and Blueprints

Woo-Jin Jeon
SysCore Lab.
Sejong University
Seoul, Republic of Korea
woojinjeon929@gmail.com

Daeseon Choi
Dept. of Medical Information
Kongju National Univ.
Kongju, Republic of Korea
sunchoi@kongju.ac.kr

Ki-Woong Park†
Dept. of Information Security
Sejong University
Seoul, Republic of Korea
woongbak@sejong.ac.kr

***Abstract*— New insights in biotechnology have been realized by analyzing time-series data accumulated over a long period of time in the field of biosystems. Similarly, a new visual insight that cannot be observed in the short-term accumulated time-series data has been provided by applying the time-lapse concept, which periodically captures images of a specific point over an extended period of time and replays it quickly. Considering these case studies, it can be concluded that analyzing computing systems based on time-series system traces accumulated over a long period of time can potentially provide a new insight for computing systems that cannot be extracted from short-term time-series data analysis. In this paper, we apply the time-lapse concept and data visualization for extremely long-term time-series system trace data to try and provide insights. For this purpose, we devise a rapid prototype for a large-capacity storage system as a proof of concept. In addition, we construct a testbed to test our concept and validate our assumptions.**

Keywords-Storage Visualization, Time-lapse, System Trace Analysis

I. INTRODUCTION

In the fields of biology and biotechnology, various studies have been conducted that deal with the analysis of time-series data accumulated over a long period of time [1], [2], [3]. New insights in the field of biotechnology have been derived by analyzing extreme data over an extended period of in biological systems. Based on academic achievements and case studies in the literature, it can be concluded that analyzing computing systems based on accumulated system traces accumulated over a long time period can potentially extract a new insight for computing systems that cannot be extracted from short-term time-series data analysis.

Figure 1 shows a comparison between short-term and long-term data, selected from the biotechnology and computer system storage fields. A considerable number of studies performed till date are based on short-term data. A typical example of short-term data analysis is genetic testing

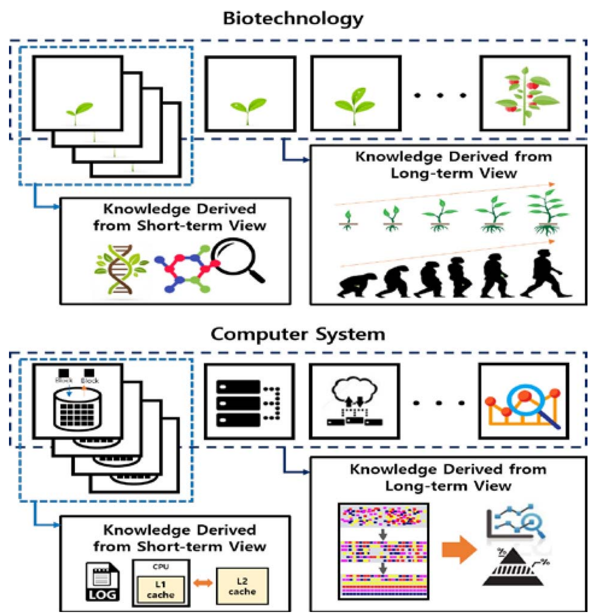

Figure 1. An Illustration Demonstrating our Proposed Concept

in biotechnology [4]. In computer systems, logging [5], [6] for timeline analysis and cache [7] for fast operation can be considered to be an example. Extracting new insights onto a computer system storage field requires a study involving analysis of long-term data. Therefore, in this paper, data visualization for long-term system trace data is applied. Moreover, a new visual insight that cannot be observed in the short-term accumulated time-series data has been provided by applying the time-lapse concept that periodically captures images of a specific point over an extended period of time and replays it quickly. For this purpose, we devise a rapid prototype for a large-capacity storage system as proof of concept.

† : Corresponding Author

978-1-5386-9555-5/18/$31.00 ©2018 IEEE
DOI 10.1109/AIKE.2018.00046

Authorized licensed use limited to: Sejong Univ. Downloaded on January 12,2024 at 13:49:22 UTC from IEEE Xplore. Restrictions apply.

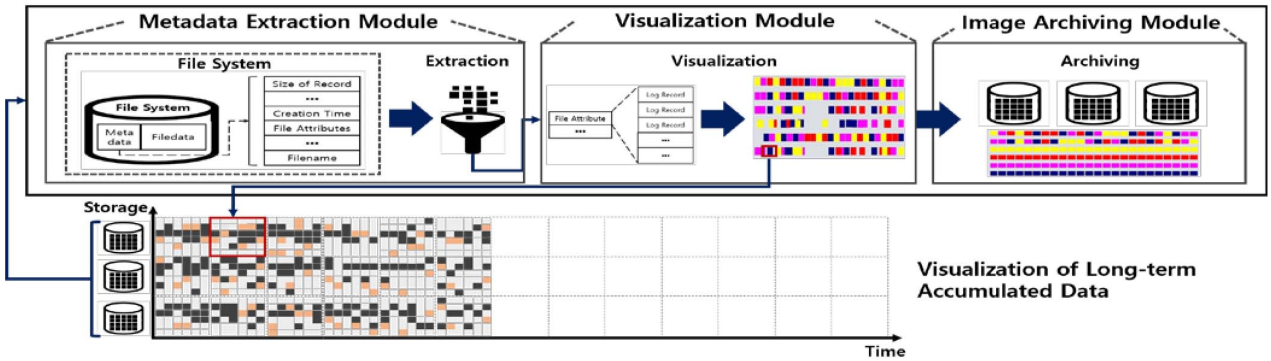

Figure 2. Prototype Architecture: Metadata Extraction Module, Visualization Module, and Image Archiving Module are the Major Components

Our contributions can be summarized as follows. First, to the best of our knowledge, this is the first attempt that involves analysis of long-term data for extracting insights. We attempted applying the time-lapse concept to the accumulated data. Second, we tried deriving the possibility through linkage with various other fields and the computer system field. Third, we developed a prototype to test our concept and constructed a testbed to verify our assumptions in a cloud computing environment.

The remainder of this paper is organized as follows: In Section 2, we apply the time-lapse concept to the storage system as a case study. In Section 3, we construct the testbed to test our concept and validate our assumptions in a cloud computing environment. Finally, in Section 4, we present our conclusions and suggest future works.

II. Case Study: Applying the Time-lapse Concept to a Cloud Storage System

In this section, we propose a rapid prototype to visualize long-term accumulated data.

Figure 2 shows the rapid prototype and its components. The three major components of our prototype are as follows:

- Metadata Extraction Module: This component extracts and collects the file system metadata from storage.
- Visualization Module: This component visualizes the collected metadata through the Image Conversion Module to generate an image stream.
- Image Archiving Module: This component collects and stores visualized data. It collects long-term visualized data and displays it as a single image stream using the time-lapse concept.

The *x*-axis of the image stream represents information regarding the extracted storage metadata, and this information gets accumulated over time. The *y*-axis represents information regarding each storage system from where metadata are being extracted from. We have attempted a computationally and spatially efficient scheme for the accumulated images, such as *FalconEye* [8], for the encoding schemes. This series of operations are performed according to the cycle set by the user. It is expected that there will be various cases involving analyses and utilization of accumulated long-term time-series data. For example, we can visualize metadata to observe changes over time and predict current trends and future trends while studying file systems [9], [10]. In addition, complex computations such as entropy measurement for ransomware detection can be simplified through data visualization. The storage system scenario tested herein is used as a case study, and we extrapolated that there are various use cases and expansion possibilities.

III. Testbed Construction in Progress

In this section, we discuss the construction of the testbed to test our concept and validate our assumptions.

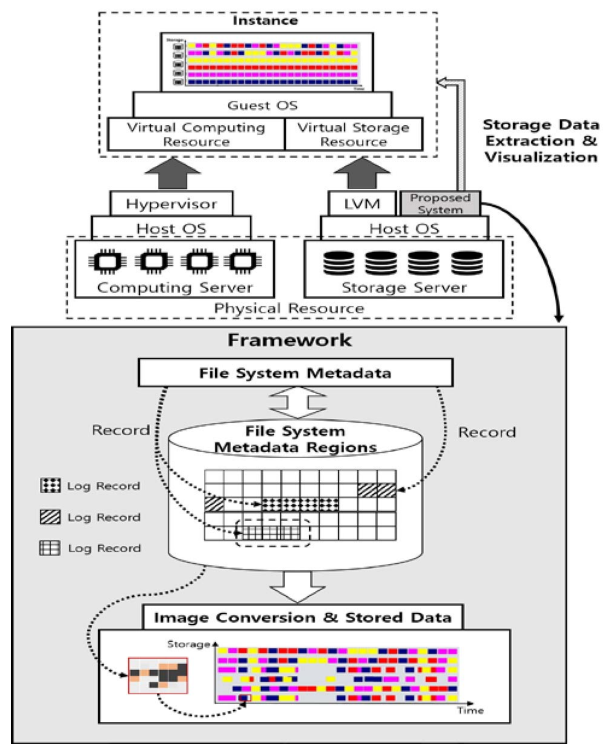

Figure 3. Framework of our Proposed Prototype in a Cloud Computing Environment

Authorized licensed use limited to: Sejong Univ. Downloaded on January 12,2024 at 13:49:22 UTC from IEEE Xplore. Restrictions apply.

Figure 3 shows the framework used for employing our proposed prototype with the cloud platform [11]. The cloud platform comprises two main types of nodes: a compute node with a hypervisor installed that provides a CPU and memory to the instance and a storage node that logically virtualizes physical storage through the Logical Volume Manager and provides it to the instance. The proposed system runs on the host machine and not the virtual machine on the cloud computing environment. Therefore, it does not affect the computing resources of the virtual machine.

IV. CONCLUSION & FURTHER WORK

This study aimed at providing new insights not observed in short-term analysis through extremely long-term trace data extraction and visualization. To accomplish this, we attempted a data visualization encoding scheme for a large-capacity storage system in a computationally and spatially efficient manner. To the best of our knowledge, this is the first study involving analysis of long-term data for extracting insights. We tried deriving the possibility through linkage with various other fields and the computer system field. The results were observed to be promising because we could implement a visualization scheme for large-capacity storage systems in a computationally and spatially efficient manner. This study will continue to place great value in the storage system field for the foreseeable future. The next step of this research would be to utilize specific examples of computer storage systems and work toward discovering new insights.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2017R1C1B2003957, NRF-2016R1A4A1 011761) and supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2018-0-00420, Development of Serverless Cloud Computing technology for resource allocation and usage measurement by API call unit).

REFERENCES

[1] Pereira, Inês, et al., "Incontinence surgery in obese women: comparative analysis of short-and long-term outcomes with a transobturator sling." International urogynecology journal 27.2 (2016): 247-253.

[2] Ryan R. Wick, Louise M. Judd, Claire L. Gorrie, and Kathryn E. Holt, "Unicycler: resolving bacterial genome assemblies from short and long sequencing reads." PLoS computational biology 13.6 (2017): e1005595.

[3] Bowman, John L., et al., "Insights into land plant evolution garnered from the Marchantia polymorpha genome." Cell 171.2 (2017): 287-304.

[4] Sati, Satish, and Giacomo Cavalli, "Chromosome conformation capture technologies and their impact in understanding genome function." Chromosoma 126.1 (2017): 33-44.

[5] Alan Wade Cohn, Gary Robert Fulkner, James Edward Kitchen, David Leon Proft, and Corey Wayne Quain, "Method and system for logging security event data." U.S. Patent Application No. 14/852,822.

[6] Michael S. Bittar, Martin D. Paulk, Gary Althoff, Paul F. Rodney, and Frode Hveding, "Systems and methods for displaying logging data." U.S. Patent No. 9,638,022. 2 May 2017.

[7] Boroumand, Amirali, et al., "LazyPIM: An efficient cache coherence mechanism for processing-in-memory." IEEE Computer Architecture Letters 16.1 (2017): 46-50.

[8] Ki-Woong Park, Woomin Hwang, and Kyu Ho Park, "FalconEye: Data center status extraction via vision transformation techniques." Computer Software and Applications Conference Workshops (COMPSACW), 2010 IEEE 34th Annual. IEEE, 2010.

[9] Nitin Agrawal, William J. Bolosky, John R. Douceur, and Jacob R. Lorch, "A five-year study of file-system metadata." ACM Transactions on Storage (TOS) 3.3 (2007): 9.

[10] Meyer, Dutch T., and William J. Bolosky, "A study of practical deduplication." ACM Transactions on Storage (TOS) 7.4 (2012): 14

[11] Hashem, Ibrahim Abaker Targio, et al., "The rise of "big data" on cloud computing: Review and open research issues." Information Systems 47 (2015): 98-115.

Authorized licensed use limited to: Sejong Univ. Downloaded on January 12,2024 at 13:49:22 UTC from IEEE Xplore. Restrictions apply.