Efficient Changing Pattern Detection on High Bandwidth Network Measurements

Wucherl Yoo and Alex Sim

Computational Research Division
Lawrence Berkeley National Laboratory
{wyoo, asim}@lbl.gov

Abstract. Since the distributed scientific collaborations and data volume grow, it has become more challenging to discover network usage patterns on a high-bandwidth, high-speed networks. It is computationally challenging due to the high frequency, large volume in the network measurements. To detect the patterns in the network measurements, we have selected and applied change point detection methods. These methods are computationally efficient and have adjustable parameters to decide the frequencies of resulted pattern changes. The detected changes can provide insights about the network usage patterns about many large data transfers or hardware/software failures.

Keywords: Network, Change Detection, Change Point Detection

1 Introduction

Network usage patterns in high bandwidth networks provide insights about network characteristics and traffic trends. Analyzing changing patterns in the network utilization can also give useful information about large data transfers, hardware/software failures, and anomaly behavior, as well as long/short term bandwidth prediction models.

While change point detection methods have been extensively studied previously, it is challenging to directly use them to discover usage patterns in high bandwidth networks. This is because of the high frequency, large volume in today's high speed network measurements. It requires high computation to compute change point detection methods for this measurement data. Therefore, the change point detection methods need to be computationally efficient. We select three change point detection methods: Bayesian Change Point (BCP) [7] [3], Sequential Change Point (SCP) [11], and Pruned Exact Linear Time (PELT) method [8]. These methods include an adjustable parameter such as posterior probability, penalty, or threshold. The adjustable parameter is essential to detect changing patterns so that the we can decide the frequency of pattern changes. Other change detection methods are not selected for this application due to the lack of adjustable parameter or high computational requirements.

In our experiments, we detected changing patterns in the network measurements on Oct/1/2014 (GMT) from Energy Sciences Network (ESnet) [1]. By adjusting the parameter in each selected method, the number of detected change points were similar among different methods. We compared the computation time and results from the change point detection methods. The rest of paper is organized as follows. Sec. 2 presents related work. Sec. 3 demonstrates how to find changing patterns from change point detection methods. Sec. 4 presents experimental evaluation of the changes detection model, and Sec. 5 concludes.

2 Related Work

Change detection has been studied in time series model for outlier detection and prediction model [13] [5]. There have been several proposed works to detect anomalies by using change point detection methods. Brutlag et al. propose to use time series forecast model to focuses on short-term anomalies [4]. Barford et al. [2] propose wavelet-based signal analysis to focus on high-frequency components of the network traffic signal. These methods are difficult to be directly applied to large size data from massive data flows due to their required computation. Sketch-based approach [9] [12] were proposed to overcome the computation challenge. However, sketch-based approach is difficult to backtrack the information from detected anomalies due to hashing operations in sketch.

Several previous works were proposed to use change detection in other applications. Orion [6] uses change detection in network traffic delay to automatically discover the dependencies of network application. Mercury [10] uses rank-baed change detection to compare performance after network changes. We have applied the change point detection methods to discover usage patterns in high bandwidth network. It is more computationally efficient to overcome the challenges of analyzing large size data.

3 Changing Pattern Detection

Change point detection is the process to identify single or multiple points within a dataset where the statistical properties such as mean and variance change over time. From the previously studied change point methods, we select to use Bayesian Change Point (BCP) [7] [3], Sequential Change Point (SCP) [11], and Pruned Exact Linear Time (PELT) method [8]. The criteria of this selection was 1) the method is computationally efficient on a large data volume; 2) the method has an adjustable parameter to control the number of change points depending on the characteristics of data and the detected changing frequencies.

BCP uses Markov Chain Monte Carlo method to find the probability of a change point at each location in a data sequence. The posterior probability of each point can be an adjustable parameter to control the number of detected change points. Setting higher posterior probability can decrease the number of detected change points. SCP is a nonparametric sequential change point detection method. Mood's median test is used for comparing changes in scale.

The average run length is the average number of observations before a false positive detection occurs in a certain probability. Setting higher threshold can decrease the number of detected change points. We use PELT to detect changes in both mean and variance. The asymptotic penalty (type 1 error) in PELT is an adjustable parameter to control the number of detected change points. Setting a lower penalty value can decrease the number.

After detecting change points, changing usage patterns can be discovered. For instance, with a large data transfer, higher throughput measurements can be found for a significantly long time, and the event can be discovered with the changing point detection method. Denoting the *change window* as a pair of change points, we can calculate the average bandwidth utilization within a change window. By comparing the difference between the average of the previous change window and that of the current change window, we can identify changing usage patterns. By setting a threshold for the differences in the averages of change windows, we can adjust the frequency of detected changing patterns.

4 Experimental Results

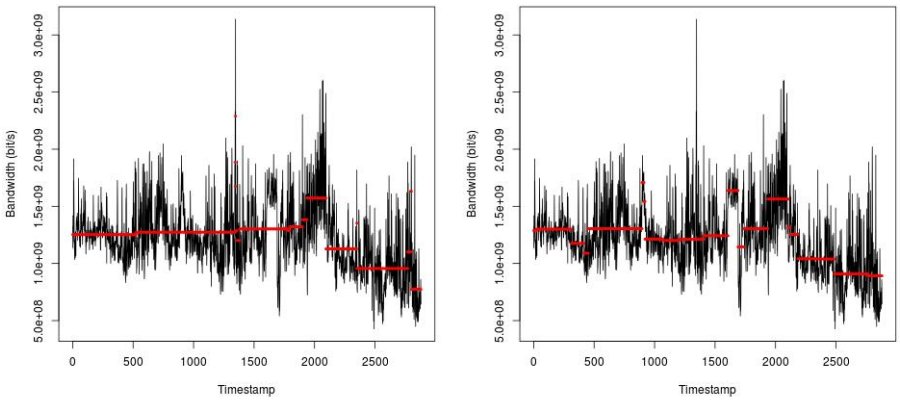
4.1 Experimental Setup

In the experiments, we used 6 directional paths connecting two sites on ESnet [1] in the USA. The paths include 6 or 7 links in the wide area network among National Energy Research Scientific Computing Center (NERSC) in California, Oak Ridge National Laboratory (ORNL) in Tennessee and Argonne National Laboratory (ANL) in Illinois. We denote *PID* as the path identification: `P1` and `P2`, directional paths between NERSC and ORNL, `P3` and `P4`, those between NERSC and ANL, and `P5` and `P6` between ORNL and ANL. The measured Simple Network Management Protocol (SNMP) data shows the aggregated bandwidth utilization for 30 second interval at the routers. The bandwidth utilization time series data were constructed by selecting the maximum value in the links of each path during the interval. This 30 second interval represents one timestamp in the time series. We conducted the experiments on a machine with AMD Opteron 6128 CPU and 64 GB memory.

4.2 Evaluation

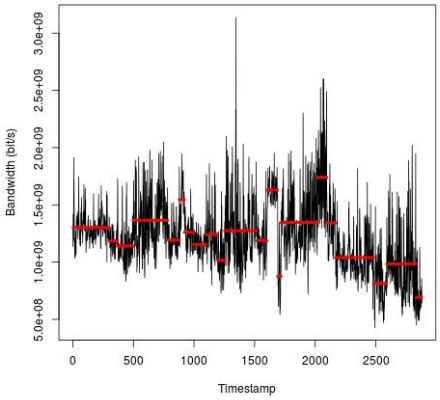
Fig. 1 shows the detected change points from the aforementioned methods (BCP, SCP, and PELT). The data includes 2880 timestamps of 30 second interval on Oct. 1 GMT 2014 for the `P5` (from ORNL to ANL). To compare these methods, we adjust the parameter for the posterior probability of BCP, the penalty of SCP and the threshold of PELT so that the number of detected change points are similar to each other. These parameters were selected to show smaller number of change points but sufficiently sensitive to discover the changes with the higher and lower bandwidth usages.

As shown in Fig. 1a, BCP detects several short-lived spikes as change points in five instances, which is not desirable for the detection of large data transfers.



(a) BCP

(b) SCP



(c) PELT

Fig. 1: Change Point Detection: The change points are detected from different methods (BCP, SCP, and PELT). The x axis is the timestamp for each 30 second interval on Oct. 1, GMT 2014. (2880 timestamps) The y axis is the bandwidth utilization (bit/s). The red colored line represents the average bandwidth utilization within the change window.

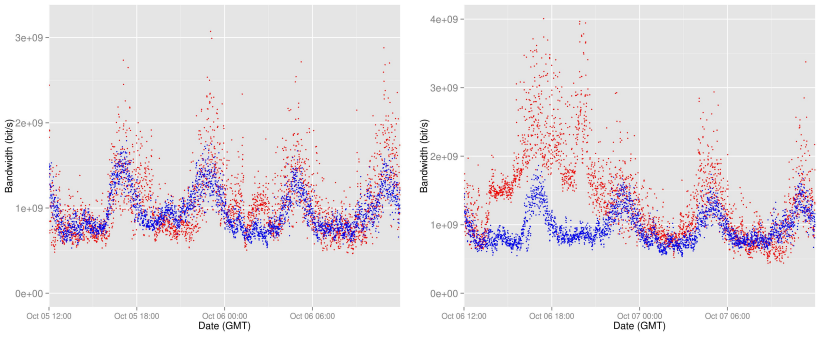
As shown in the first half of timestamps in Fig. 1a, BCP was less sensitive to detect the usage changes. If we adjust the penalty parameter of BCP, it can match the sensitivity of SCP and PELT. However, it will detect more change points, and more spikes will be detected. As shown in Fig. 1b and Fig. 1c, SCP also detects spikes as change points in two instances. This result is better than BCP but worse than PELT (0 instance) to detect large data transfers.

Table 1: The number of detected change points.

	P1	P2	P3	P4	P5	P6
BCP	20	18	75	58	18	20
SCP	18	20	27	27	21	21
PELT	19	19	75	18	19	20

The average computation times for one-day measurements on 6 paths were respectively 4.833, 0.287, 0.054 (s) for BCP, SCP, and PELT. The computational efficiency is crucial for handling large high-frequency measurement data. Tab. 1 shows the number of detected change points from the three methods. BCP and PELT detected 75 change points in P3, however SCP only detected 27. Our examination confirmed more variances and more changes in P3, which showed the less sensitivity from SCP. In addition, BCP detected 58 change points in P4, however our examination confirmed that BCP unnecessarily detected short-lived spikes as change points. We observed that the PELT method showed the least computation time with similar or better detection results from SCP. Although SCP showed higher computation time in order of magnitude than PELT, it is applicable to online sequential detection for streaming data because PELT requires entire dataset for the computation. In other words, PELT can be used in offline batch computation when full dataset is already stored and available.

4.3 Discussion



(a) Oct/5/2014, 12pm GMT (b) Oct/6/2014, 12pm GMT

Fig. 2: Bandwidth Utilization Predictions: Blue is the prediction of one day ahead from the starting time. Red is the observed data of the same duration.

Fig. 2 shows the prediction of bandwidth utilization for `P2` on Oct/5/2014, 12pm GMT in Fig 2a and on Oct/6/2014, 12pm GMT in Fig 2b [14]. Compared to Fig 2a, Fig 2b shows more prediction errors due to the large data transfers started around on Oct/6, 1pm GMT. With the detected change point on the large data transfer, we can adjust the prediction by adding the difference of the average bandwidth utilization of change windows due to the large data transfer. This adjustment of the prediction will decrease the prediction error.

Fig. 2 also shows that there is a strong daily seasonality, e.g. Fig 2a and Fig 2b show peaks and troughs in the similar times. We observed that these periodic changes were detected with the change point detection methods. Since these periodic changes are normal usage patterns, the seasonal adjustment for the changing pattern detection is necessary to avoid detecting the periodic changes as change points. We leave this as future work.

5 Conclusions

We apply the change point detection methods to discover usage patterns in high bandwidth network. We select BCP, SCP, and PELT methods because they have an adjustable parameter to control the number of change points and show computational efficiency. Among them, PELT shows the least computation time, resilient to sudden spikes. While PELT can do batch processing, SCP can do sequential (streaming) processing with the order of magnitude higher computation time. We believe that we can detect usage change patterns in our streaming measurement dataset by setting appropriate threshold between the differences of change windows. In addition, we believe that the usage pattern detection can be applicable to other measurement datasets such as system performance profiling data. As future work, we plan to investigate finding an optimal threshold to detect changing patterns and how to use detected changing patterns to improve the other analysis such as the bandwidth utilization prediction model. We also plan to incorporate seasonality adjustment in the changing pattern detection.

6 Acknowledgments

This work was supported by the Office of Advanced Scientific Computing Research, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The authors would like to thank Chris Tracy, Jon Dugan, Brian Tierney, Inder Monga and Gregory Bell at ESnet; Arie Shoshani, K. John Wu, and Jay Krous at LBNL; Richard Carlson at Dept. of Energy.

References

1. Energy Sciences Network (ESnet). http://www.es.net/ (2014)

2. Barford, P., Kline, J., Plonka, D., Ron, A.: A signal analysis of network traffic anomalies. In: Proceedings of the 2Nd ACM SIGCOMM Workshop on Internet Measurment. pp. 71–82. IMW '02, ACM, http://doi.acm.org/10.1145/637201.637210
3. Barry, D., Hartigan, J.A.: A bayesian analysis for change point problems 88(421), 309–319, http://www.jstor.org/stable/2290726
4. Brutlag, J.D.: Aberrant behavior detection in time series for network monitoring. In: LISA'00: Proceedings of the 14th USENIX conference on System administration. pp. 139–146. USENIX
5. Chen, C., Liu, L.M.: Joint estimation of model parameters and outlier effects in time series 88(421), 284–297, http://www.jstor.org/stable/2290724
6. Chen, X., Zhang, M., Mao, Z.M., Bahl, P.: Automating network application dependency discovery: Experiences, limitations, and new solutions. In: Proceedings of the 8th USENIX Conference on Operating Systems Design and Implementation. pp. 117–130. OSDI'08, USENIX Association, http://dl.acm.org/citation.cfm?id=1855741.1855750
7. Erdman, C., Emerson, J.W.: A fast bayesian change point analysis for the segmentation of microarray data 24(19), 2143–2148, http://bioinformatics.oxfordjournals.org/content/24/19/2143
8. Killick, R., Fearnhead, P., Eckley, I.A.: Optimal detection of changepoints with a linear computational cost 107(500), 1590–1598, http://dx.doi.org/10.1080/01621459.2012.737745
9. Krishnamurthy, B., Sen, S., Zhang, Y., Chen, Y.: Sketch-based change detection: Methods, evaluation, and applications. In: Proceedings of the 3rd ACM SIGCOMM Conference on Internet Measurement. pp. 234–247. IMC '03, ACM, http://doi.acm.org/10.1145/948205.948236
10. Mahimkar, A.A., Song, H.H., Ge, Z., Shaikh, A., Wang, J., Yates, J., Zhang, Y., Emmons, J.: Detecting the performance impact of upgrades in large operational networks. In: Proceedings of the ACM SIGCOMM 2010 Conference. pp. 303–314. SIGCOMM '10, ACM, http://doi.acm.org/10.1145/1851182.1851219
11. Ross, G.J.: Sequential change detection in the presence of unknown parameters 24(6), 1017–1030, http://link.springer.com/article/10.1007/s11222-013-9417-1
12. Schweller, R., Li, Z., Chen, Y., Gao, Y., Gupta, A., Zhang, Y., Dinda, P.A., Kao, M.Y., Memik, G.: Reversible sketches: Enabling monitoring and analysis over high-speed data streams 15(5), 1059–1072, http://dl.acm.org/citation.cfm?id=1322413.1322420
13. Tsay, R.S.: Time series model specification in the presence of outliers 81(393), 132–141, http://www.jstor.org/stable/2287980
14. Yoo, W., Sim, A.: Network bandwidth utilization forecast model on high bandwidth network, http://www.osti.gov/scitech/servlets/purl/1136782