ISSN:0975-9646

Real Time Data Mining Using Cyber Physical System

Pallavi .P.Pophale , Dr.M.S.Ali
Dept. of Computer Science and engineering,
Amravati University ,
P.R.M.C.E.A.M,Badnera
Amravati(MH),India

***Abstract*: The cyber physical systems are the systems which hold tight combinations of physical systems with computational systems. The physical elements are the real words entities and computation of entities is software systems. The cyber systems does not applied over more complex and critical systems as there is little chance of human loss.**

The cyber systems applied in developments like energy, medical, constructions and cloud data centres. The task of cyber system is to take run time values from physical devices and perform data mining and analytics to give results to operator in loop so that system can improve in its own way. We are proposing the scheme of ARIS – automatic reliability improvement System for cyber physical system in real world environment, where we perform the data quality analysis using latest data mining technique incremental local Outlier factor (LOF) to analyse the data points to generate the results.

***Keywords*: cyber-physical system; system reliability; reliability engineering; data analysis; machine learning; data mining;runtime environment**

I.INTRODUCTION

This paper describes a data-centric runtime monitoring platform named ARIS (Autonomic Reliability Improvement System) along with real time data mining. ARIS works in parallel with the cyber-physical system, continuously conducting automated online evaluation at multiple stages along the system workflow and providing operator-in-the-loop feedback for reliability improvement. One technique used by ARIS is data quality analysis, wherein computational intelligence is given to evaluate data quality in an automated and efficient way. ARIS also makes use of self-tuning, autonomically self- managing and self-configuring the evaluation system to ensure that it adapts itself to both changes in the system and feedback from the operator. This self-tuning continuously adapts the evaluation system to ensure proper function, which leads to a more robust evaluation system and improved system reliability. Autonomic computing meansthe self-managing characteristics of distributed computing resources, adapting to unpredictable changes while hiding intrinsic complexity to operators and users.

The initial evaluation checks to see if the input data meets the quality specifications pre-defined by the application developer and the system operator. Examples of data quality specification include data existence, up-to-date, conforming to certain distribution, time-synchronization across different sources, variation and pattern. The output data evaluation checks the quality of the results of the application. For example, for a machine learning based prediction system, data output quality relates to the accuracy or confidence level of the prediction. For a non machine learning-based system, such as a building energy management system, the quality of the data output relates to the extent to which results can be used to guide subsequent action.

II. LITERATURE SURVEY

Data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

In 2001 researcher have proposed in terms of feature construction for detection models[1], DC-1 (Detector Constructor) [6], first it involves a sequence of operations for constructing features (indicators) before constructing a cellular phone fraud detector (a classifier). There they faced with a more difficult problem because there is no standard record format for connection or session records (we had to invent our own). They also need to construct temporal and statistical features not just for individual records, but also over different connections and services. That is, we are modeling different logical entities that take on different roles and whose behavior is recorded in great detail. Extracting that from a vast and overwhelming stream of data adds considerable complexity to the problem. The work most similar to unsupervised model generation is a technique developed at SRI in the Emerald system [10]. Emerald uses historical records to build normal detection models and compares distributions of new instances to historical distributions. Discrepancies between the distributions signify an intrusion. One problem with this approach is that intrusions present in the historical distributions may cause the system to not detect similar intrusions in unseen data. Related to automatic model generation is adaptive intrusion detection. Teng et al. [18] perform adaptive real time anomaly detection by using inductively generated sequential patterns. Also relevant is Sobirey's work on adaptive intrusion detection using an expert system to collect data from audit sources [16]. Many different approaches to building

anomaly detection models have been proposed. A survey and comparison of anomaly detection techniques is given in

[19]. Stephanie Forrest presents an approach for modeling normal sequences using look ahead pairs [7] and contiguous sequences [8]. Helman and Bhangoo [12] present a statistical method to determine sequences which occur more frequently in intrusion data as opposed to normal data. Lee et al. [15, 14] uses a prediction model trained by a decision tree applied over the normal data. Ghosh and Schwartzbard [8] use neural networks to model normal data. Lane and Brodley [11, 12, 13] examine unlabeled data for anomaly detection by looking at user profiles and comparing the activity during an intrusion to the activity under normal use. Cost-sensitive modeling is an active research area in the data mining and machine learning communities because of the demand from application domains such as medical diagnosis and fraud and intrusion detection. Several techniques have been proposed for building models optimized for given cost metrics. In our research we study the principles behind these general techniques and develop new approaches according to the cost models specific to IDSs. In intrusion data representation, related work is the IETF Intrusion Detection Exchange Format project [9] and the CIDF effort [17]

In 2009 reseacher proposed data mining methods were first used for knowledge discovery from telecommunication event logs more than a decade ago [5]. In the context of IDS alert log mining[2], a number of approaches have been suggested. Clifton and Gengo [3] have investigated the detection of frequent alert sequences, in order to use this knowledge for creating IDS alert filters. Long et al. [4] have suggested a supervised clustering algorithm for distinguishing Snort IDS true alerts from false positives. Julisch and Dacier have proposed a conceptual clustering technique for IDS alert logs, so that clusters correspond to alert descriptions, and a human expert can use them for developing filtering and correlation rules for future IDS alerts. During their experiments, Julisch and Dacier found that these hand written rules reduced the number of alerts by an average of 75% . This work was later extended by Julisch who reported the reduction of alerts by 87% . Al-Mamory et al. [9, 10] have proposed clustering algorithms for finding generalized alarms which help the human analyst to write filters. During the experiments, the number of alarms decreased by 93% [9] and 74% [10].

III.PROPOSED MODEL

Autonomic system works in parallel with the cyber-physical system to perform continuous assessment at multiple stages along the system workflow and provide operator-in-the-loop feedback for reliability improvement.

This makes real time evaluation of data from cyber-physical systems. For example, abnormal input and output data can be detected and flagged based on data quality analysis. Hence, alerts can be sent out that enable the operator-in-the-loop to take actions and make changes to the system in order to minimize system downtime and maximize system reliability.

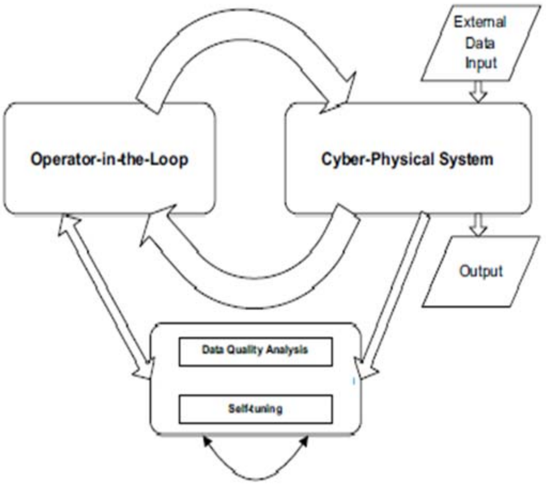

Figure 1: Workflow of proposed model

IV. CONCLUSION

The cyber physical system plays the important role in real time data mining. As most of the latest applications are using these to track the development of their project. We have generated the alerts based on the overall development of the website through ARIS system. The ARIS system is self tuned and changed the threshold values based on the decision of classifier.

REFERENCES

[1] Wenke Lee; S. J. Stolfo; P. K. Chan; E. Eskin; Wei Fan; M.Miller; S. Hershkop;Junxin Zhang DARPA *Information Survivability Conference & Exposition II, 2001. DISCEX '01.*

[2] R. Vaarandi "Real-time classification of IDS alerts with data mining techniques *” Military Communications Conference, 2009. MILCOM 2009. IEEE*

[3] C. Clifton and G. Gengo. "Developing Custom Intrusion Detection Filters Using Data Mining," *in Proc. of 2000 MILCOM Symposium, pp. 440-443.*

[4] J. Long, D. Schwartz, and S. Stoecklin. "Distinguishing False from True Alerts in Snort by Data Mining Patterns of Alerts," *in Proc. of 2006 SPIE Defense and Security Symposium*, pp. 62410B-1--62410B-10..

[5] K. Hätönen, M. Klemettinen, H. Mannila, P. Ronkainen, and H. Toivonen. "Knowledge Discovery from Telecommunication Network Alarm Databases," *in Proc. of 1996 International Conference on Data Engineering*, pp. 115-122.

[6] T. Fawcett and F. Provost. Adaptive fraud detection. *Data Mining and Knowledge Discovery*, 1:291–316, 1997.

[7] S. Forrest, S. A. Hofmeyr, A. Somayaji, and T. A. Longstaff. A sense of self for unix processes. *In 1996 IEEE Symposium on Security and Privacy, pages 120–128. IEEE Computer Society, 1996.*

[8] A. Ghosh and A. Schwartzbard. A study in using neural networks for anomaly and misuse detection. *In Proceedings of the Eighth USENIX Security Symposium, 1999.*

[9] Internet Engineering Task Force. Intrusion detection exchange format .In *http://www.ietf.org/html.charters/idwgcharter.html, 2000.*

[10] H. S. Javitz and A. Valdes. The nides statistical component: description and justification. *In Technical Report, Computer Science Labratory, SRI International,* 1993.

[11] T. Lane and C. E. Brodley. Sequence matching and learning in anomaly detection for computer security. *In Proceedings of the AAAI-97 Workshop on AI Approaches to Fraud Detection and Risk Management,* pages 43–49. Menlo Park, CA: AAAI Press, 1997.

[12] T. Lane and C. E. Brodley. Temporal sequence learning and data reduction for anomaly detection. *In Proceedings of the Fifth ACM Conference on Computer and Communications Security,* pages 150–158, 1998.

[13] T. Lane and C. E. Brodley. Temporal sequence learning and data reduction for anomaly detection. *ACM Transactions on Information and System Security*, 2:295–331, 1999.

[14] W. Lee and S. J. Stolfo. Data mining approaches for intrusion detection. *In Proceedings of the 1998 USENIX Security Symposium*, 1998.

[15] W. Lee, S. J. Stolfo, and P. K. Chan. Learning patterns from unix processes execution traces for intrusion detection. *In AAAI Workshop on AI Approaches to Fraud Detection and Risk Management*, pages 50–56. AAAI Press, 1997.

[16] M. Sobirey, B. Richter, and M. Konig. The intrusion detection system aid. architecture, and experiences in automated audit analysis. In Proc. of the IFIP TC6 / *TC11 International Conference on Communications and Multimedia Security*, pages 278 – 290, Essen, Germany, 1996.

[17] S. Staniford-Chen, B. Tung, and D. Schnackenberg. The common intrusion detection framework (cidf*). In Proceedings of the Information Survivability Workshop*, October 1998.

[18] H. S. Teng, K. Chen, and S. C. Lu. Adaptive real-time anomaly detection using inductively generated sequential patterns. *In Proceedings of the IEEE Symposium on Research in Security and Privacy*, pages 278–284, Oakland CA, May 1990.

[19] C. Warrender, S. Forrest, and B. Pearlmutter. Detecting intrusions using system calls: alternative data models. *In 1999 IEEE Symposium on Security and Privacy, pages 133– 145. IEEE Computer Society, 1999*