Big Data Research 3 (2016) 10–23

Contents lists available at ScienceDirect

Big Data Research

www.elsevier.com/locate/bdr

A Cloud Computing Based Network Monitoring and Threat Detection System for Critical Infrastructures ☆

Zhijiang Chen, Guobin Xu, Vivek Mahalingam, Linqiang Ge, James Nguyen, Wei Yu *, Chao Lu

Department of Computer and Information Sciences, Towson University, Towson, MD 21252, United States

ARTICLE INFO

Article history:
Received 2 June 2015
Received in revised form 15 September 2015
Accepted 2 November 2015
Available online 26 November 2015

Keywords:
Network monitoring
Threat detection
Cloud computing

ABSTRACT

Critical infrastructure systems perform functions and missions that are essential for our national economy, health, and security. These functions are vital to commerce, government, and society and are closely interrelated with people's lives. To provide highly secured critical infrastructure systems, a scalable, reliable and robust threat monitoring and detection system should be developed to efficiently mitigate cyber threats. In addition, big data from threat monitoring systems pose serious challenges for cyber operations because an ever growing number of devices in the system and the amount of complex monitoring data collected from critical infrastructure systems require scalable methods to capture, store, manage, and process the big data. To address these challenges, in this paper, we propose a cloud computing based network monitoring and threat detection system to make critical infrastructure systems secure. Our proposed system consists of three main components: monitoring agents, cloud infrastructure, and an operation center. To build our proposed system, we use both Hadoop MapReduce and Spark to speed up data processing by separating and processing data streams concurrently. With a real-world data set, we conducted real-world experiments to evaluate the effectiveness of our developed network monitoring and threat detection system in terms of network monitoring, threat detection, and system performance. Our empirical data indicates that the proposed system can efficiently monitor network activities, find abnormal behaviors, and detect network threats to protect critical infrastructure systems.

1. Introduction

A critical infrastructure system, as a typical cyber–physical system (CPS), is a system that integrates computation, networking, and physical elements together to support different applications [1]. It covers smart transportation, smart electrical power grids, smart medical systems, smart manufacturing systems, etc. In a critical infrastructure system, a huge amount of data will be collected from physical and cyber components and transmitted to the computing core through communication networks. The collected real time data leads to efficient and secured operations of a critical infrastructure system [2,3]. For example, in the smart grid, renewable energy sources, distributed energy storage, and generation need to be efficiently integrated and managed through complex and computationally intense models, real-time analysis, and visualization. Then, massive amount of data will be generated from the power grid and transmitted to the energy management system (EMS) in order to enable efficient system operations [4].

Similarly, in a safe and reliable transportation system, various sensors will be installed on vehicles and deployed on roadsides to collect information and transmit collected data to the operation center. With a large number of vehicles dynamically running in a transportation system, huge volumes of streaming data (big data) are generated by monitoring variations in traffic characteristics (e.g., traffic densities, speeds, vehicles, etc.), over time for timely processing and analysis [5]. For example, real-world SHRP2 dataset is over a petabyte in size [6]. Thus, the mounting volume of stored and processed data, along with the continuously increasing requirements of storage and processing capacity pose significant challenges, which hinders the effectiveness of critical infrastructure systems.

To support highly secured critical infrastructure systems, a generic threat monitoring and detection system will be developed to efficiently mitigate cyber threats. Effectively monitoring data from both physical and cyber components will facilitate the detection of cyber threats and help security administrators respond to cyber-

☆ This article belongs to Big Data Networking.

* Corresponding author.
E-mail address: wyu@towson.edu (W. Yu).

http://dx.doi.org/10.1016/j.bdr.2015.11.002
2214-5796/

threats in a timely manner. Nonetheless, developing a scalable, reliable and robust defense system for protecting critical infrastructure systems is a challenging issue. First, it is challenging to quantify the impact of threats as they may come from various sources. Second, it is difficult to detect threats because the detection system has to inspect various data sources, which are always in large-scale with different formats and semantics. Monitoring different applications and detecting threats can be characterized as high volume data streams and real-time processing requirements. In addition, resources in critical infrastructure systems (e.g., bandwidth, storage, etc.) are also limited. Thus, how to efficiently store and process such big data to ensure security and efficient operations of critical infrastructure systems become a challenging and urgent issue.

To address the aforementioned challenges, in this paper, we developed a cloud-computing based network monitoring and threat detection system for critical infrastructures, which can efficiently enhance the security of critical infrastructures. The proposed system consists of monitoring agents, cloud infrastructure, and an operation center. Monitoring agents can be deployed on devices in a critical infrastructure system to collect the threat monitoring information and transmit the information to the cloud infrastructure. A cloud infrastructure is a distributed system that is deployed with a number of servers, providing both storage and computation resources to efficiently process the large scale of collected data. In the cloud infrastructure, we used both Hadoop MapReduce [44] and Spark [45] to speed up data processing by separating and analyzing the data streams concurrently. The operation center plays the intelligence role that dynamically updates system operation policies and configuration, and monitors the system security.

To demonstrate the effectiveness of our developed network threat monitoring and detection system, we conducted experiments from three aspects: network monitoring, threat detection, and system performance. In network monitoring, we designed several scenarios to find the outgoing traffic volume from each server, the traffic volumes based on source and destination IP address, the incoming traffic volume to each server, and the port access count on each server in a given time duration. We implemented k-mean clustering algorithm in our proposed system to fast classify data into different groups based on their similarities. With dynamic thresholds, we conducted threat detection on emulated distributed denial-of-service (DDoS) network traffic and measured both detection rate and false positive rate. In addition, we compared the efficiency of Hadoop and Spark for the data processing in network monitoring and threat detection. Through our extensive evaluations, our result shows that our proposed system can efficiently help the system administrator to monitor network activities and identify abnormal behaviors. Moreover, our proposed system can accurately and dynamically detect network threats. Our experimental data shows that, with the implementation by Spark, the system performance is almost 30 times better than that of the implementation by Hadoop.

The remainder of the paper is organized as follows: In Section 2, we conduct a literature review of big data in critical infrastructure and cloud computing to assist network monitoring and threat detection. In Section 3, we introduce our system architecture, components, and features. In Section 4, we propose to present the system implementation in detail. In Section 5, we demonstrate the experimental results to validate the effectiveness of the developed threat monitoring system. In Section 6, we show the extension of our proposed system. Finally, we conclude the paper in Section 7.

2. Related work

In this section, we conduct the literature review on big data issues in critical infrastructure systems, cloud computing to assist network monitoring and threat detection.

2.1. Big data issue in critical infrastructure system

Governments, research communities, and enterprises can all make use of the overwhelming amounts of digital data, which is available, evidently creating new opportunities and nurturing powerful business intelligence for decision support [7,8]. Big data can be used by the organizations in creating real-time solutions to the challenges put forth by healthcare, agriculture, transportation, and more. The relentless growth of data will not only challenge information technology engineers, but also researchers in various fields. In order to process massive amounts of data that have been collected, there have been a number of studies on critical infrastructure systems in the past [7,8,20–26,36–40]. For example, Nakazato et al. [9] designed a prototypical system related to big data, in which the driver can use his smartphone to view traffic conditions on an expressway in real time. Chen et al. [10] developed a system called DiabeticLink, which offers electronic health record search, diabetic health indicator tracking, Q&A forums, diabetic medication side effect reporting, and diet recommendations for diabetic patients. The researchers utilized the modern data, text, and web mining algorithms that are relevant to healthcare decision support. In addition, the rapid growth of smart phones has apparently increased the usage of low cost sensors that detect environment and user interaction. For example, Billen et al. [11] presented a framework that stores, fuses and processes smartphone sensor data. Smart phone sensor data can be used in various areas, notably to detect the drunkenness of the driver [12], and pothole detection [13]. Detecting potholes (road surface defects) on real-time can avoid accidents and higher costs.

2.2. Cloud computing to assist network monitoring

In critical infrastructure systems, network traffic has always been the primary resource for network security enthusiasts [15–19,26–28,35,38,42,43]. There have been several MapReduce based algorithms implemented to monitor network traffic [16,17]. Albeit there are several programming models for parallel processing, MapReduce is a generic mechanism to perform challenging computing tasks [14]. As indicated in [14,17], the key features of MapReduce include cost effectiveness, extreme scalability, high throughput and high performance. Vieira et al. [15] evaluated the efficiency of MapReduce in packet level analysis and DPI (Deep Packet Inspection) and verified that packet level analysis and DPI are Map-intensive. Their study implied that block, input and cluster sizes play a vital role in defining the job completion time and efficiency of MapReduce [15]. Lee and Lee [16] analyzed multi-terabytes of network traffic in a scalable manner. They have devised TCP, IP and HTTP traffic analysis using MapReduce algorithms to improve the scalability of analysis. In addition, a web-based interface to execute Hive queries on NetFlow data was developed.

2.3. Cloud computing to assist threat detection

In order to quickly and efficiently detect threats, integrating detection algorithms with cloud computing has been one of the most active research areas [17–19,28–31,33–35,41–43]. For example, Aljarah and Ludwig [17] proposed an intrusion detection system that uses MapReduce to analyze network traffic. They proposed an algorithm IDS-MRCPSO, which is based on the particle swarm optimization method and clustering based on MapReduce