Future Generation Computer Systems 36 (2014) 189–208

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

A taxonomy and survey on Green Data Center Networks

Kashif Bilal *, Saif Ur Rehman Malik, Osman Khalid, Abdul Hameed, Enrique Alvarez, Vidura Wijaysekara, Rizwana Irfan, Sarjan Shrestha, Debjyoti Dwivedy, Mazhar Ali, Usman Shahid Khan, Assad Abbas, Nauman Jalil, Samee U. Khan

Department of Electrical and Computer Engineering, North Dakota State University, Fargo, ND, USA

HIGHLIGHTS

- We provide an overview of the research within Data Center Networks (DCNs).
- We present the state-of-the-art energy efficiency techniques for a DCN.
- The survey elaborates on the DCN architectures (electrical, optical, and hybrid).
- We also focus on traffic management, characterization, and performance monitoring.
- We present a comparative analysis of the aforementioned within the DCN domain.

ARTICLE INFO

Article history:
Received 4 November 2012
Received in revised form
16 May 2013
Accepted 17 July 2013
Available online 2 August 2013

Keywords:
Data center
Data center networks
Network architectures
Network performance
Network management
Network experimentation

ABSTRACT

Data centers are growing exponentially (in number and size) to accommodate the escalating user and application demands. Likewise, the concerns about the environmental impacts, energy needs, and electricity cost of data centers are also growing. Network infrastructure being the communication backbone of the data center plays a pivotal role in the data center's scalability, performance, energy consumption, and cost. Research community is endeavoring hard to overcome the challenges faced by the legacy Data Center Networks (DCNs). Serious efforts have been made to handle the problems in various DCN areas. This survey presents significant insights to the state-of-the-art research conducted pertaining to the DCN domain along with a detailed discussion of the energy efficiency aspects of the DCNs. The authors explored: (a) DCN architectures (electrical, optical, and hybrid), (b) network traffic management and characterization, (c) DCN performance monitoring, (d) network-aware resource allocation, (e) DCN experimentation techniques, and (f) energy efficiency. The survey presents an overview of the ongoing research in the broad domain of DCNs and highlights the challenges faced by the DCN research community.

1. Introduction

A data center is a facility for hosting computing resources networked together using the communication infrastructure for data storage and application hosting [1–3]. The present era marks the beginning of the Exascale computing [4]. The Exascale data centers are expected to operate at a computing power of 10^{18} floating-point operations (flops) per second (one million trillion flops per second). Consequently, data centers are growing exponentially in the number of hosted servers, thereby escalating the role of Data Center Network (DCN) to connect hundreds of thousands of servers. Today's data centers are constrained by the interconnection networks instead of the computational power [5], consequently marking the DCNs as the critical scalability bottleneck. Data centers belonging to Yahoo, Microsoft, and Google already host hundreds of thousands of servers [6,7].

The major Information and Communication Technology (ICT) components within the data centers are: (a) servers, (b) storage, and (c) interconnection network. DCN being the communication backbone is one of the foremost design concerns in the data center [1]. The DCN infrastructure plays a vital role in ascertaining the performance aspects and initial capital investment in the data center. Exponential growth in the number of servers poses critical challenges in terms of: (a) scalability, (b) fault tolerance, (c) energy efficiency, and (d) cross-section bandwidth in the DCN [1]. Tremendous efforts are laid by the research community to overcome the challenges faced by the DCNs.

* Corresponding author. Tel.: +1 7013690896.
E-mail addresses: kashif.bilal@ndsu.edu, eishaalkashif@gmail.com (K. Bilal), Saif.Rehmanmalik@ndsu.edu (S.U.R. Malik), Osman.Khalid@ndsu.edu (O. Khalid), Abdul.Hameed@ndsu.edu (A. Hameed), Enrique.Alvarezvazquez@ndsu.edu (E. Alvarez), Vidura.Wijaysekara@ndsu.edu (V. Wijaysekara), Rizwana.Irfan@ndsu.edu (R. Irfan), Sarjan.Shrestha@ndsu.edu (S. Shrestha), Debjyoti.Dwivedy@ndsu.edu (D. Dwivedy), Mazhar.Ali@ndsu.edu (M. Ali), Ushahid.Khan@ndsu.edu (U. Shahid Khan), Assad.Abbas@ndsu.edu (A. Abbas), Nauman.Jalil@ndsu.edu (N. Jalil), Samee.Khan@ndsu.edu (S.U. Khan).

0167-739X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.future.2013.07.006

DCN being an integral part of the data center, acting as a communication backbone, requires extreme consideration and plays a pivotal role in greening the data center. Network components are one of the major energy consumers within data centers besides servers and cooling infrastructure. In 2010, the network infrastructure was estimated to consume around 15.6 billion kWh of energy in data centers [8]. The cooling infrastructure and servers are becoming more energy efficient as a result of the ample research efforts. While considering energy-proportional servers, the energy consumption share of the network equipment is expected to increase up to 50% [9]. Moreover, emerging technologies, such as network virtualization, Virtual Desktop Interface (VDI), cloud gaming, and mobile cloud, demand high bandwidth and network utilization. Consequently, the energy consumption of the networks is anticipated to rise amply. Electrical energy consumption contributes in Green House Gases (GHG) emissions [10]. The Information and Communication Technology (ICT) sector is identified to contribute around 2% of the total GHG emission, equivalent to the GHG emission of the aviation industry worldwide [11]. The networking components are projected to contribute around 37% of the total ICT GHG emissions [12,13]. The aforementioned discussion ratifies the apparent need and impetus for the energy efficient networking in data centers.

The idleness of the network links and devices can be exploited to employ network energy efficiency techniques. Network energy efficiency can be achieved by: (a) consolidating the network traffic on fewer links and devices to power off the idle devices and (b) scaling down the network link data rate to save energy. The resource consolidation approach exploits the resource overprovisioning of the networking components to consolidate the workload on a set of active network components to switch off the underutilized networking equipment [14–18]. Resource consolidation approaches: (a) compute the required subset of network links and devices to satisfy the workload demand, (b) redirect the traffic to the calculated subset, and (c) power off the idle devices and links [19]. Some of the proposals that use resource consolidation for the network energy efficiency are discussed in Section 2.4.

Individual network devices and links can be exploited for energy saving by employing the *proportional computing* technique. Proportional computing refers to the concept of energy consumption in proportion to the resource utilization [20]. Adaptive Link Rate (ALR) is a proportional computing technique that is applied on network links to reduce the energy consumption by: (a) scaling down the communication link data rate for underutilized links and (b) placing the idle link to sleep mode. Data rate switching is controlled by the ALR policy to decide the data rate to fulfill the workload demand. Various ALR policies have been proposed in the literature. A detailed survey of the ALR is presented in [20]. IEEE Energy Efficient Ethernet (EEE) task force standardized the ALR (IEEE 802.3az standard) in 2010. Energy Efficient Ethernet (IEEE 802.3az standard) provides a mechanism for green Ethernet using the ALR [21,22]. The IEEE 802.3az introduces Low Power Idle (LPI) mode to place the link in low power mode to save energy. It has been estimated that around five TWh of energy saving can be achieved by using IEEE 802.3az enabled devices [22]. Several energy efficient network solutions have been contributed by the research community using the aforementioned green networking techniques, such as [19,23–26].

For the past few years, data centers have been increasingly employed to run a wide range of applications in various domains, such as scientific applications, healthcare, e-commerce, smart grids, and nuclear science. Cloud computing [27,28] has emerged as a feasible platform for the execution of such scientific applications. Applications, such as weather forecasting require data streams from satellites and ground instruments, such as radars and weather stations are fed to the Cloud to compute EvapoTranspiration (ET) coefficient [29]. The CARMEN e-science project describes the working of brain and allows neuroscientists to share and analyze data [30]. The yield monitor sensors mounted on harvest machines equipped with Global Positioning System (GPS) produce intensive data in the agriculture domain. Heuristic algorithms are used to identify key management zones for cotton field [31]. In the healthcare domain, data centers are used to provide services to various clinics and hospitals [32].

Many e-commerce applications make use of data centers and support customers by accessing data [33]. For example, eBay is one of the popular auction websites. To increase the range of operations, eBay acquired the Switch-X data center in 2006 and purchased more land at South Jordan and Utah, USA, in 2008 [33]. The main data warehouse of eBay has around two petabytes of user data, millions of queries per day, and tens of thousands of users. The classical power grids have advanced to smart grids by distributing electrical power with an additional control over the appliances [34]. A smart grid information management paradigm for smart grids is presented in [35]. Nuclear reactions form the basis of energy production, environmental monitoring, radiotherapy, disease diagnosis, and material analysis. Moreover, data centers can be useful for exchange, distribution, and collection of information related to nuclear reactions [36,37]. The Nuclear Reaction Data Centers (NRDC) is a group of fourteen data centers having origin in fourteen countries and two global organizations. Data centers in NRDC exchange experimental information according to data types and message formats defined in [36,37].

This survey presents a comprehensive overview of the state-of-the-art research conducted in the domain of data center networks. Major DCN areas focused on in the paper are: (a) DCN architectures, (b) network traffic management and characterization, (c) performance analysis, (d) network-aware resource allocation and experimentation strategies, and (e) greening the DCNs. The taxonomy of the highlighted research areas is presented in Fig. 1.

Various surveys encompassing the state-of-the-art in Green Data Center Networks exist. For example, Bolla et al. [38] discussed the contemporary approaches and trends in energy-aware fixed network infrastructures. The survey conducted by Zhang et al. [39] focused on energy efficiency in optical networks mainly with certain discussion on energy efficiency in data centers. Bianzino et al. [40] discussed energy efficiency for wired networks, whereas the solutions to minimize the energy consumption in communication devices, protocols, networks, end-user systems, and data centers are presented in [41]. We believe that our survey is more versatile and covers a broad domain in the state-of-the-art network research. Our survey details various DCN architectures, such as electrical, optical, and hybrid along with the potentials of the energy efficiency that can be achieved using various DCN architectures and technologies. Moreover, to enlighten the feasibility of employing green networking techniques, we present a detailed analysis on network traffic management and network monitoring that comprise the key considerations for implementing the network efficiency techniques. Furthermore, we present the details of energy-aware resource scheduling within the data center and discuss various experimentation platforms that can be used to simulate and determine the applicability of the network energy efficiency techniques.

The paper is organized as follows. Section 2 provides a detailed overview of the DCN architectures using: (a) electrical, (b) optical, and (c) hybrid network elements and their energy efficiency related discussion at the end. Section 3 presents a detailed overview of the data center traffic analysis and management covering details of network traffic management strategies, protocols, and data center traffic characteristics. Data center performance monitoring techniques are characterized in Section 4. Network-aware resource allocation and experimentation strategies (simulations and emulation test-beds) are discussed in Sections 5 and 6, respectively. Every section is supported by comprehensive discussion of the state-of-the-art green practices in the area. Section 7 concludes the discussion.