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

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.

© 2013 Elsevier B.V. All rights reserved.

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).

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.
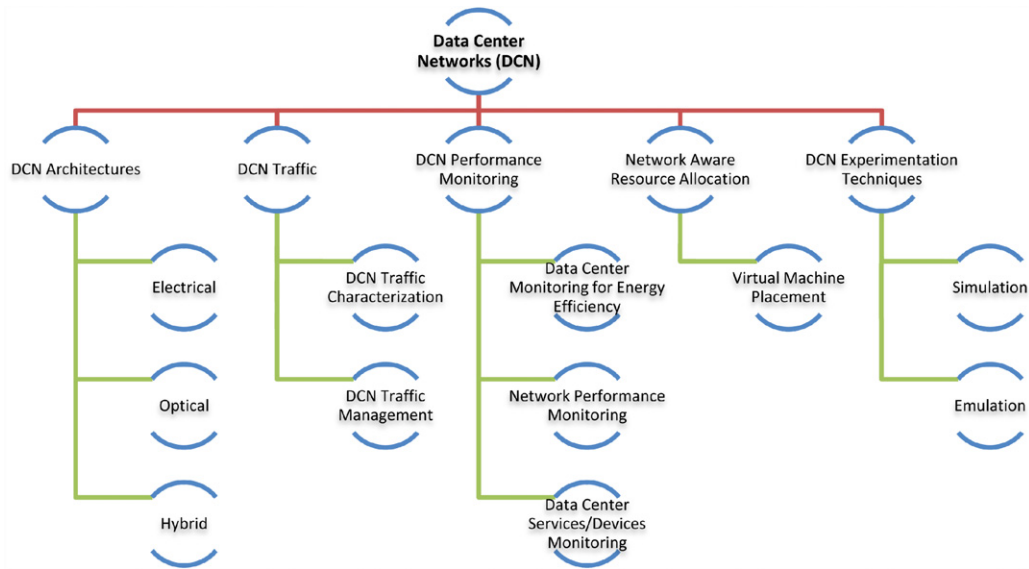
Fig. 1. Data center networks' taxonomy.

2. Data center network architectures

The DCN architecture lays the communication infrastructure within the data center and requires extreme consideration [1]. It plays a pivotal role in the data center scalability and performance bounds. End-to-end aggregate bandwidth is a major bottleneck to the data center network performance [42]. Typical data centers use three-tier hierarchical network design and enterprise-class equipment at the higher layers of the network [43]. The enterprise-class network equipment is expensive and power hungry, consuming excess of energy [1]. To accommodate the growing demands of the data center communication and handle the problems faced by the legacy DCN, new DCN architectures are required to be designed. The research community addressed the challenges faced by the DCN architectures using: (a) commodity electrical network elements, (b) optical technology, and (c) hybrid network technologies. Commodity network elements consume much less power as compared to the enterprise-level network elements. Similarly, optical DCNs consume much less energy as compared to the electrical networks. Moreover, optical and hybrid network technologies can be used to design energy efficient DCNs and augment the existing network infrastructure. This section elaborates the state-of-the-art DCN architectures that use the aforementioned technologies. In addition, Section 2.4 details the state-of-the-art energy efficiency practices within the data centers.

2.1. Commodity network elements based DCNs

Three-tier architecture is the most widely deployed DCN architecture [44,43]. The three-tier architecture follows a layered approach to arrange the network switches in three layers. The network elements (switches and routers) are arranged in three layers namely: (a) access layer, (b) aggregation layer, and (c) core layer in the three-tier DCN architecture (see Fig. 2). Power-hungry enterprise-level equipment is used at the higher layers of the three-tier architecture. The legacy three-tier DCN architecture lacks the capability to meet the current data center bandwidth and growth trend [1]. The major shortcomings of the legacy DCN architecture can be expressed in terms of: (a) scalability, (b) cost, (c) energy consumption, (d) cross-section bandwidth, and (e) agility [1]. To accommodate the shortcomings of the legacy DCN architecture, new architectures are proposed by the research community. Some

of the major DCN architectures are discussed in the following paragraphs.

The fat-tree based DCN architecture is proposed by Al-Fares et al. (see Fig. 2) [42]. The fat-tree architecture aims to maximize the end-to-end bisection bandwidth. Moreover, the architecture is highly cost effective and energy efficient as compared to the legacy three-tier DCN, as it uses commodity switches at each network level. The architecture is (a) scalable, (b) cost effective, (c) energy efficient, (d) fault tolerant, and (e) backward compatible. The fat-tree architecture possesses great potential for green communication by exploiting the link overprovisioning and multiple paths between the end nodes. The authors have also discussed a customized IP addressing scheme and a customized multipath routing algorithm. The major drawbacks of the algorithm are: (a) manual and location dependent addressing scheme, (b) use of large number of network switches, (c) requirement to change the routing logic in switches, and (d) increased cabling cost.

VL2 is a hierarchical fat-tree based DCN architecture [45]. The architecture uses a flat automated addressing scheme that facilitates the placement of servers anywhere in the network without configuring the address manually. VL2 also uses commodity network switches for cost and energy efficiency. It mainly focuses on: (a) automated addressing, (b) potential for transparent service migration, (c) load balanced traffic flow for high cross-section bandwidth, and (d) end system based address resolution. The drawbacks of the architecture include (a) virtual overlays, (b) virtual network shim layer, and (c) centralized management (VL2 Directory services).

A recursively defined DCell architecture is proposed by Guo et al. [46]. The architecture is extremely scalable and can easily scale to millions of servers in the data center. DCell is defined recursively by building higher level *dcells* using *dcell*$_0$ as the basic building block. DCell is hybrid architecture and places network traffic routing logic on servers instead of network switches. The servers act as routing devices besides performing computation. Each server has multiple network interfaces and each server is connected to other servers in different *dcells*. The architecture uses commodity equipment and defines a customized routing scheme. The architecture: (a) is highly scalable, (b) provides efficient one-to-one and all-to-all communication, and (c) is cost effective. The drawbacks of the architecture include: (a) high oversubscription ratio, (b) low cross-section bandwidth, (c) long communication
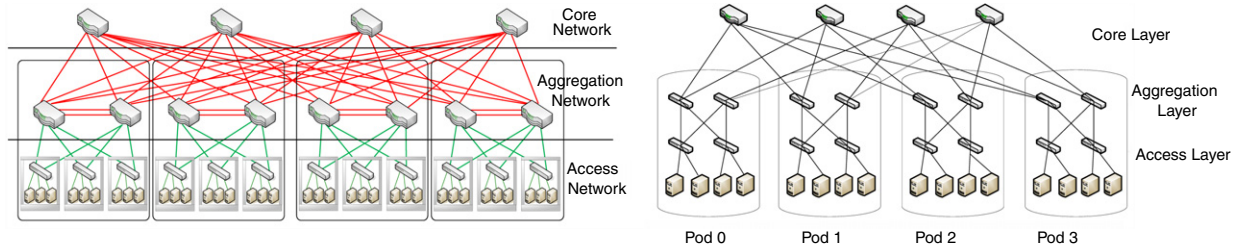
Fig. 2. Data center network architectures. Three-tier DCN (left) and fat-tree DCN (right).

paths (more than fifteen hops in a level 3 DCell), (d) additional network interfaces, and (e) increased cabling cost.

The BCube DCN architecture is a recursively defined DCN architecture. The architecture adopts some of the characteristics from the DCell architecture [47]. BCube is designed for container data centers, and accordingly the scalability is low (limited to some thousands of servers). It uses network switches to interconnect different cells instead of using server-to-server connection for inter-cell communication. It provides high cross-section bandwidth and fault tolerance. BCube is: (a) cost effective, (b) efficient for one-to-one and many-to-many communication, (c) provides high cross-section bandwidth, and (d) fault tolerant. The drawbacks of BCube include: (a) poor scalability (some thousands of servers as it is designed for container data centers), (b) additional network interfaces, and (c) increased cabling cost.

CamCube is a pure server-centric DCN architecture designed for container data centers [48]. It uses torus topology to connect a single server with six other servers using network interfaces. Network traffic routing is performed by servers. Network switches are not used at all in the CamCube DCN architecture. The CamCube uses different routing schemes for different services and provides a load balanced flow assignment. It eliminates the use of similar routing logic for all services, and instead facilitates the use of customized routing logic for individual services.

Singla et al. [49] proposed a DCN architecture named Jellyfish. The key focus of the proposed architecture is to enable incremental growth in the data center with respect to: (a) servers and (b) bandwidth. Jellyfish offers: (a) incremental growth, (b) random connectivity, and (c) network equipment heterogeneity. The Jellyfish architecture promises high cross-section bandwidth. The authors used a random graph model for interconnectivity at the switch level. N servers are connected to a switch in random, and each switch is connected to K other switches in random. The advantages of the discussed approach using random graph connectivity include scalability, incremental growth, and high bandwidth. On the downside, the authors have not discussed any routing approach and presented no discussion related to the handling of the major issues at the switch layer. We believe that the broadcast handling, loop-free forwarding, and large forwarding table size at the network switches are the most obvious problems with the proposed approach. Moreover, Jellyfish defines no network structure that may present a serious challenge in future upgrades and network traffic routing issues.

The DCN architecture demands high end-to-end bandwidth with lower cost and fault tolerance. Scalability is also one of the most desired features of the DCN architectures. Table 1 compares the major DCN architectures in terms of scalability, cost effectiveness, fault tolerance, energy efficiency, network architecture, flexibility for server placement, and network topology.

The energy efficiency column in Table 1 highlights the energy consumption and potential of the DCN architectures to employ network energy efficiency techniques. The three-tier architecture uses power-hungry enterprise-level network equipment and high oversubscription ratio. The oversubscribed network limits the employment of energy efficiency techniques. On the contrary, the fat-tree, VL2, and Jellyfish switch-centric network architectures use commodity network equipment that consumes much less power than the enterprise-class network. Moreover, the aforementioned DCNs offer low oversubscription ratio, high network overprovisioning, and multiple end-to-end network paths, thereby offering the opportunity to use network energy efficiency techniques. Heller et al. [19] demonstrated the feasibility of around 50% energy savings using the fat-tree architecture. The torus based CamCube server-centric architecture exhibits high path diversity and low average path length. The network switches are not used at all in the CamCube. Service-centric routing can be used to employ energy efficiency techniques. However, using the computational servers as network devices may limit the transition of servers to sleep or power-off mode for energy efficiency. The DCell and BCube architectures use a hybrid routing mechanism and possess high average path length. The servers constitute the intermediate network hops for packet forwarding. Moreover, as demonstrated in [1], the DCell architecture exhibits performance degradation in case of high network utilization. Therefore, employment of the energy efficiency techniques is infeasible.

2.2. Optical technologies in data centers

Optical technology is an active research topic in data center networks. Several solutions have been proposed to accommodate the future network demands. Photonic systems provide high speeds and low power consumption. The optical approach can provide a flexibility that is required in new generation data centers. Traditional electronic data networks in data centers no longer meet the current demand of speed and throughput demands [50]. Electronic switch architectures result in poor: (a) throughput, (b) latency, (c) scalability, and (d) efficiency [51]. Moreover, traditional electronic solutions incur a complicated management and higher cost [52]. Ref. [52] discussed the benefits of incorporating the Optical Circuit Switching (OCS) in the data center. The OCSs are extremely energy efficient, because there is no per-packet processing and no added latency. Data centers and computer clusters consist of tens of thousands of nodes that are capped to 10 Gbs. Increasing the number of servers in the data center results in increased network nodes, increased stress on the communication links, performance degradation, high cabling complexity, and elevated power consumption [53]. The speed rate bottleneck has been changed from the server to the network [5]. Therefore, the growing demands of bandwidth and computing capacity required the inclusion of alternative technologies in DCNs [52]. Significant improvements in data centers have been made using photonic elements. Photonic solutions increase capacity and bandwidth and decrease power consumption [50]. Photonics may be effective in providing efficient and congestion-free solutions to improve the network performance, energy efficiency, and throughput [51].

Some of the proposed network solutions, such as [50,5,54,55], permit the configuration of routing traffic based on packets or circuits simultaneously. Ref. [53] presented a new architecture,

Table 1
DCN architectures' comparison.

Reference	Scalability	Cross-section bandwidth	Cost effectiveness	Fault tolerance	Energy efficiency	Network architecture	Server placement and addressing flexibility	Network topology
Three-tier [44]	Medium	Low	Low	Medium	Low	Switch-centric	Low	Hierarchical
Fat-tree [42]	High	High	High	High	High	Switch-centric	Manual addressing and poor flexibility for migration	Clos based hierarchical topology
DCell [46]	Very high	Medium	High	Medium	Low	Hybrid (switch +server)	Manual addressing and poor flexibility for migration	Recursively defined topology, based on DCell0
BCube [47]	Low	High	High	Medium	Low	Hybrid (switch+server)	Manual addressing and poor flexibility for migration	Recursively defined topology, based on BCube0
VL2 [45]	High	High (due to load balanced flow assignment)	High	High	High	Switch-centric	Flat addressing and servers can be placed anywhere in network	Clos based hierarchical topology
CamCube [48]	Low	High	High	High	Medium	Server-centric	Key based addressing	Torus
Jellyfish [49]	High	High	High	High	High	Switch-centric	Random	Flat

termed Proteus, that uses Wavelength Division Multiplexing (WDM) and Optical Wavelength Switching (OWS) to obtain run-time network topology re-configurability according to the network traffic demands. Proteus includes three units: (a) server racks, (b) optical (de)multiplex unit, and (c) optical switching matrix. Moreover, a centralized network manager that allows run-time topology configurability is also used. The system provided higher bandwidth, simplified cabling, high network scalability, high flexibility, low power consumption, and low network latency.

The authors in [54] adopted hybrid networks to develop a combination of a conventional router and algorithm that uses optical bypasses. Moreover, the proposed hybrid system permeates the networks to optimize the energy efficiency. The authors experimentally demonstrated reconfiguration in an on-the-fly router-failure scenario and proposed the system as an experimental test-bed for new routing algorithms.

Proietti et al. [56] proposed a 40 Gbps 8 × 8 optical switch with hybrid approach to avoid packet dropping and allow low communication latencies. The addressed switch exploits optical parallelism and high speed switching and includes a loopback buffer that prevents packet drop. The authors presented satisfactory experimental results for various small sized packets with a switching latency as low as 118.2 ns. Ref. [55] analyzed the advantages of the electrical and optical interconnects. The authors proposed a system, named as OpenFlow, that combines optical and electrical elements to obtain high bandwidth, lower energy consumption, high burst tolerance, and many-to-many interconnections. The integration of the optical and electrical systems is quite challenging. The hardware challenges include: (a) switching time and (b) link setup, and the software challenges are: (a) flow duration, (b) correlations in demand, (c) interference between applications, and (d) priority across flows within and across applications. OpenFlow is a hybrid packet/circuit switched data center with circuit controllers and real integrated optics. The authors incorporated the properties of modern data centers such as fundamental heterogeneity and multi-tenancy in OpenFlow. The system was satisfactorily tested on real hardware over a year.

Xu et al. [57] presented a hybrid optical packet and wavelength switching platform that allows packets and data stream with correct logics. The proposed solution made use of silicon microrings and Semiconductor Optical Amplifier (SOA) based switches, combining the power to select a wavelength through the microring and the broadband band behavior given by the SOA. The authors implemented a hardware test-bed, where they achieved error-free low-power-consumption communications with up to 40 Gbps in Amplitude-Shift-Keyed (ASK) signals and 10 Gb/s in Differential

Phase-Shift Keying (DPSK) signals, with packet and circuit switching. The proposed architecture is highly energy efficient and potentially integratable. Alternatively, the authors in [51], instead of combining electrical and optical components in data centers, proposed the idea of all-optical switching technology including a new class of switches with tunable wavelength converters, a waveguide router, and fixed wavelength converters. According to the authors, an electronic switch architecture degrades the power efficiency by increasing the amount of energy consumed in the interconnection networks. Therefore, the authors presented optical interconnection with low latency that consisted of tunable lasers, an Arrayed Waveguide Grating Router (AWGR), and a set of electronic memory queue switches, allowing multiple packets to travel on different wavelengths to the same port simultaneously. The system provides a solution for input–output limitations faced in current data centers.

Wang et al. [50] developed a 4 × 4 SOA based optical switching node, where each switching node can be configured for packet or circuit mode simultaneously. Error-free routing and transmission of 10 Gbps communication was successfully demonstrated using a custom test-bed with four sub-modules, each assigned to an individual input port. To verify the switch, error-free communication was tested at 10 Gbps by sending data from different sources. According to Ref. [50], the data centers' application classes generate various traffic patterns, such as large extended flows and short message with arbitrary destinations. An interconnection of both the traffic patterns may produce efficient solutions in terms of energy and cost. The authors proposed the fully optical system as an experimental platform for varying traffic demand within data centers. Wang et al. [58] presented a truly bidirectional, cost-effective, and power-efficient photonic switch. The authors utilized six SOA gate devices by leveraging the inherent bidirectional transparency of SOAs. The discussed approach incurs a 63% power savings in the number of devices necessary in the fat-tree [42] architecture.

Ref. [59] used Orthogonal Frequency Division Multiplexing (OFDM) instead of WDM. The authors studied the power consumption and energy efficiency of optical interconnects of a large-scale parallel system, specifically with an adaptively modulated OFDM and an increase in the capacity using MultiMode Fiber (MMF). Adaptive modulation was performed by negotiations between transmitter and receiver. The authors demonstrated using a Field Programmable Gate Array (FPGA) test-bed that the proposed system achieved a decrease of power consumption by 70% compared to the traditional non-adaptive OFDM modulation with speeds up to 26 Gbps and link(s) of 300 m length.

Table 2
Optical technologies in DCN.

Reference	Hybrid node	Packet and circuit traffic	Hardware tested (test-bed)	FPGA model	Software tested	Traffic adaptation	Switch architecture
[50]	✓	✓	✓	×	✓	✓	×
[53]	✓	×	×	×	✓	×	×
[5]	✓	✓	✓	✓	×	×	×
[52]	✓	×	×	×	×	×	×
[51]	×	×	✓	×	×	×	✓
[58]	×	×	✓	×	×	×	✓
[54]	×	✓	✓	✓	×	×	✓
[59]	×	×	✓	✓	×	×	×
[56]	✓	×	✓	✓	×	×	×
[57]	✓	×	✓	×	×	×	✓
[55]	✓	✓	✓	×	×	✓	×

Table 2 summarizes the addressed papers in this section. The cited references are displayed in rows and the work characteristics are presented in columns. Hybrid node refers to the nodes that combine electrical and optical components. Packet and circuit traffic relate to those references that included a discussion of a system capable of handling both types of traffic simultaneously. Most of the proposed architectures were validated using either a hardware (such as FPGA based), or software based test-bed. The traffic adaptation column indicates whether the referenced system includes a traffic adaptation mechanism. The last columns relate to the references that focus their attention to the design of a switch. Hybrid DCN architectures are discussed in the next section.

2.3. Hybrid data center architecture

A hierarchical tree of electrical network switches is used in the conventional method of interconnecting the server racks in a data center. To support the aggregate traffic, the links at the core layer of the network need to operate at a much higher bandwidth. Switches that provide higher bandwidth at the core level are prohibitively expensive [42]. As all-to-all communication is rare in the network, the links are usually oversubscribed [60]. Data-intensive computing in data centers has increased rapidly [61]. Data-intensive computing jobs are prone to forming hotspots and congestion in an oversubscribed data center network [62]. Many researchers have addressed the issues related to job scheduling, routing, and network topology formation in an oversubscribed and congested network. Network topologies like fat-tree [42] have been proposed to provide (or deliver) full bisectional bandwidth throughout the network. Because fat-tree is constructed using commodity switches, the network is less expensive.

However, a major challenge in adopting the fat-tree architecture is the inability to augment the existing data center network architectures. To install a fat-tree network, the existing network has to be torn down completely [62]. Statistics indicate that full bisectional bandwidth is not necessary for most of the applications running on production data centers [62]. Only a few special applications tend to create network congestion. Therefore, research is being performed to provide extra bandwidth to the links on demand without making drastic changes to the existing network. Consequently, hybrid networks are proposed. Moreover, hybrid networks offer opportunities for energy efficiency by using energy efficient equipment and workload offloading. This section provides an analysis of the work related to such hybrid networks.

Helios is a hybrid optical/electrical switch architecture proposed in [60]. According to the authors, hybrid networks reduce the power consumption of the data center and improve the energy proportionality with minimum overheads. Pods are modules that contain 250 to 1000 servers with the interconnecting switch fabric. Organizations, such as Google, use pods that are readily available from vendors to build modular data centers [60]. The proposed architecture is a two-level multi-root tree that is intended to be used in a modular data center to interconnect the pods. There is an electrical packet switch for each pod at the pod level. The core level consists of a mixture of packet switches and optical circuit switches. The links between the pod layer and the core layer are optical links. Moreover, the uplinks of the pod layer switches and all the links of the core layer packet switches have optical transceivers. The circuit switches in the core layer use mirrors mounted on a micro-electromechanical system to route the optical signals. The links from pod layer switches to the core layer circuit switches are laid through passive optical multiplexers. Passive optical multiplexers use coarse wavelength division multiplexing to achieve a higher bandwidth. The network traffic demands at the pod level are stable and suitable for a hybrid network [60]. Helios identifies sturdy large flows (*elephants*) at the core layer and directs them through the optical core circuit switch for better performance. Moreover, Helios shows better performance as compared with a non-blocking electrical switch in terms of less complexity and energy consumption.

Ref. [61] proposed the hybrid optical/electrical network architecture named HyPac (Hybrid Packet and Circuit). HyPac connects all the Top-of-Rack (ToR) switches via an optical switch, such that there is only one hop between any two ToR switches. The optical network is separate from the hierarchical packet switch network. In HyPac, the applications' queue, the data in the socket buffer, and the HyPac Optical Manager (OM) decide the allocation of optical paths based on the lengths of queues at each socket and destination. OM sends the information of the topology to the server racks once the optical paths are allocated. Like Helios, a ToR switch can send data only to one ToR switch using the optical path. At the ToR switch level, some application data appears as a burst at socket buffer. HyPac tries to identify such bursts of data and allocates it to the optical links. Various cloud applications, such as large data transfer, witnessed significant performance improvements in terms of lower cost and less energy consumption.

Emerging radio technologies, such as 60 GHz, are foreseen as feasible solutions to the network extension and connectivity in data centers [63]. Energy efficiency is among the major design requirements of the 60 GHz wireless devices. Commercial 60 GHz devices use around 10 mW transmit power as compared to 50 mW in typical 802.11 devices [62]. 60 GHz technology can be used to replace or augment the data center wired network. Halperin et al. [62] used 60 GHz wireless links "flyways" to augment the conventional data center network. One or more wireless devices with directional antennas are placed on each ToR. Contrary to optical links, wireless links that are proposed for flyways are slower than electrical links. The authors found that the network congestion is caused by high fan in/out rather than the elephant flows between two nodes. Provided that the traffic demand is readily available, the system identifies the congested downlinks/uplinks between aggregate layer and ToR switch caused by high fan in/out. Consequently, some of the traffic from the congested link is detoured through a nearby non-congested link to a ToR switch, which makes a flyway to the destination ToR.

Table 3
Hybrid DCN architectures.

Reference	Augmenting link type	Level of connectivity	Simulated data sets	On-line traffic demand calculation	Hardware simulation
Helios [60]	Optical	Pod	Synthetic	✓	✓
Halperin et al. [62]	Wireless	ToR	Real	×	×
HyPac [61]	Optical	ToR	Synthetic	✓	×

Data centers are gaining popularity and a wide range of applications including the data-intensive applications run on data centers. Data-intensive programs congest the oversubscribed networks. It is demonstrated that the elephant flows between two nodes cause the congestion. Section 3 details the traffic management and elephant flows in the data center. The authors in [61,60] proposed methods to offload the elephant flows from the network. The authors in [60] claim that the Helios, a hybrid electrical and optical switch based architecture, presents a cost effective and energy efficient data center architecture. Moreover, the authors in [62] claim that high fan in/out causes the congestion based on the information conceived by analyzing four real data center traffic traces. However, all of the proposed methods seem to relieve congestion in the network. To decongest the data center networks more efficiently, further research is required to better understand the traffic patterns on the data center network. Table 3 compares different characteristics of the aforementioned hybrid DCN architectures.

2.4. Energy efficiency in data center architectures

DCN being the communication backbone of the data center plays a key role in the data center performance and energy consumption. The legacy three-tier architecture use power-hungry high-end enterprise-class network equipment at higher layers of the network topology. Exponential growth in the data processing and storage stimulates the growth of data centers. Due to uncertain energy supplies and volatile energy prices, energy efficient data centers are required. Energy-efficiency improvements in the data center can be achieved in three areas: (a) software, (b) power supply chain, and (c) cooling. The aforementioned areas are highly interdependent. For instance, the energy efficiency gain achieved in software may decrease the power demand and demand of cooling [64].

Various research efforts have been made to use the cost and energy efficient commodity network equipment instead of costly and power-hungry enterprise-class equipment. Al-Fares et al. [42] proposed the fat-tree based DCN architecture using commodity network switches for reduced cost and energy consumption and lower heat dissipation. The authors build a prototype of the forwarding algorithms using NetFPGAs and Click to carry out large-scale evaluation of the proposed architecture. The authors reported that a fat-tree based data center consumes around 56.6% less energy and dissipates around 56.5% less heat [42].

The DCN architectures are overprovisioned for peak loads and fault tolerance. On average, the DCNs remain highly underutilized with an average load of around 5%–25% [65]. Network overprovisioning and underutilization can be exploited for energy efficiency. Numerous efforts have been made to exploit the underutilization and resource overprovisioning for the energy efficient DCN. Heller et al. [19] proposed ElasticTree, to consolidate the workload on a subset of network resources to save energy. The authors estimated a feasibility of around 50% energy savings using simulation and hardware prototype.

Mahadevan et al. [66] simulated the effects of Network Traffic Consolidation (NTC) and Server Load Consolidations (SLC) combined with the NTC scheme for e-commerce traffic trace and data center topology. The author reported an energy savings of around 74% of the total network power.

Table 4
Energy efficiency in data center architecture.

Reference	Techniques used	Energy efficiency
[42]	NetFPGAs and click	56.6%
[66]	NTC and SLC	74%
[25]	Energy-aware traffic engineering	50%
[26]	Heuristics base approach	24%
[67]	Optical amplifiers	20%
[19]	ElasticTree	50%

Abts et al. evaluated an energy-proportional Flattened butterfly (FBFLY) network using an in-house event driven network simulator. Moreover, the authors used Adaptive Link Rate (ALR) techniques to scale down the communication link data rate using network traffic prediction, and reported an energy savings of 30%–45% [9].

Carrega et al. [65] used the idea of traffic aggregation for energy efficiency by merging traffic to a subset of links. The authors reported 22% energy savings. Moreover, Bolla et al. [25] used energy-aware traffic engineering to redirect network traffic from underutilized links for energy savings. The authors reported 27%–42% energy savings when the link load is less than 50%.

A heuristics base approach to turn off network links is used by Chiaraviglio et al. [26]. The authors considered a wide area network scenario inspired by a real network of an ISP. The simulation results showed up to 24% energy savings.

Optical interconnects offer high bandwidth at low energy consumption compared to electrical networks. Optical networks consume significantly less energy as compared to electrical networks [67]. It has been reported that the energy efficiency of optical networks is improving at the rate of 15%–20% per annum [68]. As presented in Table 4, Ref. [66] revealed better result and presented an energy efficient approach by reporting 74% energy saving of around 74% of the total network power.

Optical amplifiers used in the second generation of optical systems have reduced the energy per bit to around 10 nJ/b/1000 km [68]. The energy consumption of a high-end electrical router is approximately 20 nJ/b, while an optical network transmitter consumes around 0.5 nJ/b [69]. A complete optical DCN is estimated to provide around 75% energy savings [64].

The hybrid data center attempts to relieve the highly congested links (hotspots) in the oversubscribed data center networks, which is one of the foremost requirements to build fine-grained energy-proportional data centers [9]. Wireless networks can be used for augmenting oversubscribed DCNs in case of high traffic demands and congestions. According to Ref. [62], the wireless devices consume less than 10 mW each, and therefore the overhead of augmenting wireless links on Top of the Racks (ToRs) is small.

The 60 GHz RF technology based fully wireless DCNs are also under consideration in the literature [70,71]. Wireless networks offer better candidatures for flexible and adaptive topology for energy savings on the fly [71]. Network traffic from the underutilized links can be migrated to fewer wireless links and idle devices can be placed in sleep mode for energy savings.

3. Data center network traffic management and characterization

Most of the future Internet communication is expected to occur within data centers [72]. Many communication intensive

applications hosted within the data centers use one-to-all or all-to-all communication pattern [73]. Traffic management is the process of managing, prioritizing, controlling, and reducing the traffic that is running over the network [74]. The objective of the traffic management is to reduce the (a) congestion, (b) latency, and (c) packet loss to improve the performance of the whole network [75]. This section elaborates the network traffic management techniques and data center traffic characterization.

3.1. Data center traffic characterization

For the past few years, data centers have been increasingly employed in universities, enterprises, and consumer settings to run a wide range of applications. The applications are mainly but not limited to web search, Internet Messaging (IM), gaming, and mail services [76]. For proper functioning of a wide range of applications for customer satisfaction, a deeper look and analysis of the traffic characteristics of the data centers are required [76]. For example, a congested data center network, where internal traffic is routinely subjected to losses and poor throughput, might lead to delayed search queries, interrupted instant messaging, and hanged email services [76]. Elephant flows are one of the major causes of the aforementioned issues. Elephant flows are also one of the main characterizations of traffic in data centers. Moreover, elephant flows can affect the performance of the architecture by causing congestion in the network. Therefore, a lot of research has been conducted to effectively mitigate the effect of elephant flows on the performance of the architecture (more discussion in Section 3.2). The hybrid data centers are proposed to diminish the effect of elephant flows in data centers, as discussed in Section 2.3.

Benson et al. [76] investigated traffic characteristics in data centers. The study is based on the Simple Network Management Protocol (SNMP) data collected from 19 corporate and enterprise data centers. The SNMP data provides aggregate traffic statistics of every network device of five minutes' intervals. The authors analyzed the data for conducting the macroscopic study of temporal and spatial variations in traffic volumes and loss rates. The analysis revealed that the average link loads are high in the network core than the network edge. Alternatively, average link losses were higher at the edge than core of the network.

The second set of collected data comprised of packet traces from five network switches in one of the 19 data centers. The traces were collected by attaching a dedicated packet sniffer to a Switched Port ANalyzer (SPAN) port of each switch [76]. The packets were recorded at a granularity of ten milliseconds over the span of a fortnight for fine-grained traffic information, such as packets' inter-arrival times. The authors analyzed the fine-grained data to perform a microscopic study of traffic behavior in data centers. An on–off traffic pattern was observed at the network edge. The on and off periods are packet inter-arrival times, where the on period follows the log normal distribution. The authors developed a straw man framework for reverse engineering the fine-grained characteristics from coarse-grained information. The framework can be used to: (a) optimize load balancing and traffic engineering mechanisms and (b) design workload generators for data centers. A study on ten different data centers (three university campuses, two private enterprises, and five Clouds) is demonstrated in [77]. The authors shed light on the network design and usage. Amongst the studied data centers, three data centers run a variety of internet facing applications, whereas the remaining predominantly run the MapReduce workloads. The data collected includes SNMP link statistics for all data centers and fine-grained packet traces from selected switches. The authors performed a top-down analysis, starting with the applications (running in each data center) and then digging down to the application's send and receive patterns, and their network impact level. The types of applications running

in each data center were examined and their relative contribution to the network traffic is measured. The fine-grained sending patterns at the packet and flow level were examined. Finally, SNMP traces were used to examine the network-level impact in terms of link utilization, congestion, and packet drops.

The study covers most of the recently proposed data center network traffic management strategies that are used to avoid congestion and to increase the performance of the network efficiently. The comparison and the study will be helpful in understanding the breadth of techniques that are deployed to handle the problem of network traffic management. Some common features, such as scalability, load balancing, and energy efficiency, are selected amongst the approaches to compare. The features listed are some of the most common concerns required for the successful implementation of data center networks, as presented in the Gartner list [78]. Table 6 summarizes and provides a comparison of the techniques discussed in this section.

The key findings include the following.

(a) A wide variety of applications ranging from customer-facing applications, enterprise applications, and data-intensive applications were studied. The authors observed that the application placement is non-uniform across racks.
(b) Most flows in the data centers were small (<10 kB); a significant fraction of flows last under a few hundreds of milliseconds. The number of active flows per second was under 10,000 per rack.
(c) Despite the differences in the size and usage of the data centers, traffic originating from a DC rack was on/off in nature.
(d) In Cloud data centers, a majority of traffic originated by servers (80%) stayed within the rack. For university and private enterprise data centers, most of the traffic (40%–90%) left the rack and traversed the network interconnect.

Chen et al. [79] performed an analysis of inter-data center (D2D) traffic characteristics using Yahoo data sets. Analyses of traffic characteristics within the data centers reveal the: (a) workload distribution and (b) congestion-occurring areas, and help in better design and management of data centers. Therefore, a detailed understanding of the traffic characteristics among multiple data centers (of a single service provider) and their interactions with client triggered traffic is critical for effective operations and management of multiple data centers. Traffic characteristics help in service deployment, and improve caching and load-balancing strategies. In [79], it is demonstrated that Yahoo organizes their data centers in a hierarchy. In satellite data centers, D2D traffic is highly correlated with the client traffic. D2D traffic in the backbone data centers can be classified into: (a) client triggered D2D traffic and (b) background D2D traffic. The background D2D traffic demonstrates little variance over the day, whereas the client triggered D2D traffic shows a higher degree of variance. The authors studied Yahoo data sets; however the methodology can be applied to understanding D2D and Data center to Client (D2C) traffic characteristics for any Cloud-service provider [79]. Table 5 shows different types of data centers that were analyzed.

3.2. Traffic management in data centers

Different approaches, such as routing algorithms, transmission protocols, flow detection, and congestion control strategies, have been proposed by different researchers to control the traffic over the network. In this section, we will study various network traffic management strategies and protocols. The purpose of the study is to cover the breadth of approaches that are used to solve the traffic management problems (to avoid congestion and to balance the load efficiently) in data centers.

Table 5
Data center network traffic analysis.

Reference	Numbers of data centers analyzed	Data center types	Data sets used	Applications
[76]	19	Corporate and enterprise	SNMP data logs, packet traces from five switches	Data center traffic engineering, traffic generators for data center
[77]	10	University, Cloud, enterprise	SNMP polls, packet traces, network topology	Flow level communication characteristics, packet level communication characteristics
[79]	5	Yahoo!	Net flow data sets from 5 data centers, D2D traffic, D2C traffic	Data interference, flow classification, traffic correlation

Table 6
Comparison of traffic management approaches in the data center networks.

Reference	Features					Topology	Goal
	Scalability	Cost effective	Energy efficient	Load balancing	Traffic pattern/distribution		
Wu et al. [81]	✓	×	×	✓	Many-to-one	Fat-Tree	TCP performance for incast congestion
Curtis et al. [71]	✓	✓	✓	✓	Map-Reduce Traffic	Clos Network Topology	Throughput Improvements and Low overheads
Shang et al. [89]	×	✓	✓	×	TCP Traffic	Fat-Tree, BCube	Energy Efficient Routing
Mahaptra et al. [88]	✓	×	×	✓	Random-uniform, Random non-uniform	Fat-Tree	Flow Level Load Balancing, Compare Flow and Packet level VLDB
Alizadeh et al. [80]	✓	✓	×	✓	Many-to-one, All-to-All	Multihop Topology	High burst tolerance, Low latency, and High throughput,
Abu-Libdeh et al. [48]	×	✓	×	✓	NA	3D torus topology,	Application specific routing
Zats et al. [84]	×	✓	×	✓	Sequential queries, and Partition/Aggregate queries	Fat-Tree	Reducing the flow completion time
Al Fares et al. [86]	✓	×	×	✓	Random Bijective traffic pattern	Fat-Tree	Maximize bisectional bandwidth
Li et al. [85]	✓	✓	✓	✓	Uniform Distribution, Power Law Distribution	BCube	Reduced network traffic, Increased throughput
Benson et al. [91]	✓	×	×	✓	Bursty Traffic Patterns	Fat-Tree	Traffic scheduling

Alizadeh et al. [80] proposed a variant of the Transmission Control Protocol (TCP) known as Data Center TCP (DCTCP). The diverse nature of cloud applications that require small predictable latency and large sustainable throughput makes the TCP inefficient [80]. The authors proposed DCTCP that leverages Explicit Congestion Notification (ECN) and a simple multi-bit feedback mechanism at the host side. The authors analyzed the production of data center traffic, which uses commodity switches to identify the impairments that affect the performance of data centers. The goal of the DCTCP is to achieve high burst tolerance, low latency, and high throughput using commodity switches. The DCTCP algorithm works by reacting in proportion to the congestion. The aforementioned algorithm used a simple marking scheme that sets the Congestion Experienced (CE) code-point of packet as soon as the congestion is noticed. The congestion is indicated when a buffer exceeds a small fixed threshold. The DCTCP reacts by reducing the window by a factor that depends on the fraction of marked packets over the total packets sent. The DCTCP uses ECN-echo at the receiver and the receiver has to send the ACK messages until it receives a CWR confirmation from the sender. The said overhead may have some effect on the computation and may affect the power consumption of the network. Alizadeh et al. analyzed that the DCTCP delivers the same or better throughput compared to the TCP using 90% less buffer space. Moreover, the DCTCP also provides high burst tolerance and low latency for short lived flows.

Wu et al. [81] proposed the Incast Congestion Control Protocol for TCP (ICTCP) to improve the performance under incast congestion. The data center networks are well structured and layered to achieve high bandwidth and low latency. The buffer size of ToR Ethernet switches is usually small [81]. Moreover, recent studies show that the barrier synchronized many-to-one traffic pattern is common in data center networks, caused by MapReduce like applications [82,83]. Furthermore, the transmission data volume for the traffic pattern is usually small. Highly bursty traffic of multiple TCP connections overflows the Ethernet switch buffer in a short period of time that causes intense packet losses resulting in TCP incast collapse [81]. Therefore, the TCP retransmission and timeout occurs. Earlier solutions in the literature focused on either reducing the waiting time for packet loss recovery by faster retransmissions, or controlling switch buffer occupation to avoid overflow by using ECN and modified TCP at both sender and receiver sides. Ref. [81] performed incast congestion avoidance at the receiver side. The receiver side is a natural choice since the throughput of all TCP connections and the available bandwidth is known to the receiver. The receiver side adjusts the receive window size for each TCP connection, so that the aggregate burstiness of all the synchronized senders remain controlled [81]. The aim of the ICTCP is to improve the performance by reducing congestions. However, energy efficient aspects are not discussed by the authors in the paper.

Zats et al. [84] proposed DeTail, an in-network, multipath-aware congestion management mechanism. The purpose of De-Tail is to focus on reducing the flow completion time tail in data center networks. Three mechanisms that are being used by De-Tail to effectively manage congestion are: (a) in-network traffic management for quick detection and response to the congestion, (b) multipath data transfers for alternate traffic routes to avoid congestion hotspots, and (c) traffic differentiation for allocating resources based on flow requirements. DeTail incorporates link-layer flow control, adaptive load balancing, and priority to perform in-network, multipath-aware congestion management. To avoid head-of-line-blocking every switch propagates multiple messages from the bottleneck nodes to other nodes. The aforementioned requires routers to perform more computation and ultimately increase the power utilization of the network. The authors evaluated DeTail through implementation and simulation, demonstrating the ability to reduce the flow completion time across a variety of workloads.

Li et al. [85] proposed an Efficient and Scalable Multicast (ESM) routing scheme for data center networks in [85]. ESM exploits the feature of modern data center networks to achieve multicasting based on a regular topology. The source-to-receiver expansion approach is used to build the multicast trees. ESM excludes the unnecessary intermediate switches used in normal receiver driven multicast routing. The authors built the multicast tree by leveraging the multistage graph feature of the data center. ESM uses in-packet bloom filters (to eliminate the necessity of in-switch routing entries) and in-switch entries to make the scheme scalable. Through experiments and simulation, the authors demonstrated that ESM saves around 40%–50% network traffic, reduces the number of in-switch entries, and doubles the throughput.

Al-fares et al. [86] presented Hedera, a dynamic flow scheduling system for multi-stage switch topologies in data centers. Hedera collects flow information from constituent switches, computes non-conflicting paths for flows, and instructs switches to re-route traffic accordingly. The goal of Hedera is to (a) maximize aggregate network utilization, (b) bisection bandwidth, and (c) minimal scheduler overhead or impact on active flows. The authors took a global view of routing and traffic demands, and enabled the scheduling system to observe bottlenecks that switch local schedulers cannot anticipate. The authors supported the idea by large-scale simulations and implementing the Hedera on the PortLand test-bed. The algorithm delivers near to optimal performance using a hypothetical non-blocking switch for numerous interesting and realistic communication patterns, and delivers up to 4 times more bandwidth than the state-of-the-art ECMP techniques. Hedera has three control loops that include switches and scheduler. The said loops have some communication and computation overheads that increase the energy requirements for Hedera. However, the authors performed experiments to prove that Hedera delivers bandwidth improvements with modest control and computation overhead.

Curtis et al. [87] discussed the timely detection of elephant flows to effectively utilize the bisection bandwidth provided by the topologies of traffic management. The authors stated that the existing elephant flow detection methods have several limitations that make them unsuitable for data center networks. Mainly three techniques are used in the literature to identify elephant flows in a network: (a) periodic polling of statistics from switches, (b) streaming techniques like sampling or window based algorithms, and (c) application-level modifications. The authors claim that modifying the application is probably not an acceptable solution because of the Quality of Services (QoS). Moreover, the other two approaches (a and b) have high monitoring overheads, significant switch resource consumption, and long detection times. In the said perspective, the authors proposed to detect elephant flows at the end hosts by observing the end hosts' socket buffers. According to the authors, the end hosts can provide better, more efficient visibility of flow behavior. Furthermore, the proposed approach reduces the monitoring overheads and switches resource utilization that ultimately reduces the overall energy consumption of the network. Mahout, a low-overhead effective traffic management system that follows a central controller approach for network management, is proposed by the authors in [87]. Once an elephant flow is detected, an end host signals the network controller using in-band signaling rather than monitoring the whole network. At the switches, any flow not signaled as an elephant is routed using a static load-balancing scheme. Only elephant flows are monitored and managed by the central controller. The aforementioned helps reduce the energy consumption required in monitoring. The authors concluded that the combination of end host elephant detection and in-band signaling eliminates the need for per-flow monitoring in the switches. Therefore, the aforementioned incurs low overhead and requires fewer switch resources.

The fat-tree topology is well suited for Valiant Load Balancing (VLB) [71]. VLB can be applied in a TCP/IP network at the flow level to avoid the out-of-order packet issue that can cause performance degradation in the TCP. However, flow-level VLB can have serious performance problems in presence of large flows. Moreover, a fairly large percentage of flows in data center traffics are large [71], which may be a significant problem. Mahapatra et al. [88] demonstrated that flow-level VLB can be significantly worse than packet-level VLB for non-uniform traffics in data center networks. Two alternate load-balancing mechanisms that utilize the real-time network state information to achieve load are proposed in [88]. Queue-Length Directed Adaptive routing selects the output port with the smallest queue length that can reach the destination. The first packet of a flow (e.g., the SYN packet of a TCP flow) adaptively constructs the path towards the destination based on the queue-lengths of output ports. Once the path is determined, all other packets follow the same path to avoid the out-of-order packet problem. Probing Based Adaptive routing sends a number of probe packets following different paths to the destination. The number of paths to be probed is a parameter of the scheme. The receiving node replies with an acknowledgment packet for the first probe packet received and drops the other probed packets for the flow. The acknowledgment packet carries the path information, and the source will then use the selected path for the communication. Probing the network information for large flows ensures that large flows are not routed over congested links, which ultimately achieves load balancing. The scheme allows the current network state to be probed before a large flow is routed and decreases the chance of network congestion. As stated in Section 3.3, there is a tradeoff between performance and energy consumption. The proposed strategies improve the performance but aggravate the energy consumption. This is due to the fact that more switch resources are utilized by processing large probe messages and then sending the acknowledgments back to the sender. However, the tradeoff between performance and energy consumption can always be balanced in regard to the SLA requirements.

Shang et al. [89] discussed how to save energy in high-density data center networks in routing perspective. The authors proposed a routing scheme to turn off the switches that are not being used. The key idea is to use as few network devices as possible to provide the routing service, with no or little sacrifice in the network performance. The idle network devices can be shut down or placed into sleep mode for energy savings. The authors designed a heuristic algorithm to achieve the energy efficient routing. Initially, routing and the corresponding network throughput are calculated taking all the switches into consideration. Afterwards, the switches involved in basic routing are eliminated based on the workload. When the network throughput falls down to the threshold level, the elimination process is stopped. Finally, the switches that are not involved in the routing are powered off or placed into sleep mode. The heuristic routing algorithm consists of three modules: (a) Route Generation (RG), (b) Throughput Computation (TC), and (c) Switch Elimination (SE). The proposed technique may not be practical for large-scale networks. Therefore, a tradeoff between the computation complexity and the efficiency of the results exists. The authors supported the idea by performing simulation using Fat-Tree and BCube architectures to demonstrate that energy-aware routing can effectively save the power consumed by network devices.

Abu-Libdeh et al. [48] explored the benefits and feasibility of using multiple service-specific routing protocols. CamCube [90] is used to find out the feasibility of the aforementioned idea. Ref. [90] uses 3D torus topology in which every server is connected to six other servers. The authors of [48] utilized low level link oriented API from CamCube that provides flexibility to the services to implement customized routing protocols. The authors demonstrated

that the applications are more efficient and generate less additional traffic control overheads. Moreover, an extended routing service that allows easy implementation of application-specific routing protocol on CamCube is also proposed. The proposed strategy maintains a separate queue per service that affects the energy utilization of the network. However, the authors proved through experiments that the additional overheads of using extended routing are very low.

Benson et al. [91] developed Micro Traffic Engineering (MicroTE) that leverages the short term and partial predictability of the traffic matrix, to adapt the traffic variations. According to the authors, the existing approaches of traffic engineering perform 15%–20% worse than the optimal solution. The reasons are: (a) failure to utilize multipath, (b) failure to adapt to change in traffic load, and (c) lack of global view of traffic. Motivated by the above factors, Ref. [91] mitigated the impact of congestion due to unpredictable traffic. MicroTE relies on a central controller to track ToR pairs that have predictable traffic and then optimally route the traffic. Moreover, the remaining unpredictable traffic is routed to the weighted equal-cost multipath routes. The weights in the aforementioned routes reflect the available capacity after routing the predictable traffic. The MicroTE requires firmware upgrades to the switches and designating a server as a controller. The authors in [91] performed experiments and simulations to show that MicroTE offers close to optimal performance in predictable traffic and degenerates to ECMP when traffic is not predictable. However, the energy utilization may aggravate due to the network and monitoring component of MicroTE.

3.3. Energy efficiency using network traffic management

Traffic management strategies, such as [92,93,23], can help in making data centers green and energy efficient. Network traffic management and characterization provide the basic foundation to employ the energy efficiency techniques in the data center. Most of the energy efficiency techniques discussed in Section 2.4 heavily rely on network traffic characterization and communication pattern. DCNs typically experience an average network load of less than 25% of the peak load. It has been observed that a considerable amount of network links remain idle for 70% of the time in data centers. Several strategies are used to achieve energy efficiency and proportionality, such as discussed in [93], to redirect the network traffic from the links of one router to another router to place the network interfaces and entire chassis of router in sleep mode. Similarly, in [23], a network-level solution 'Green Traffic Engineering (GreenTE)' is proposed that combines the idea of workload consolidation and network optimization. GreenTE finds a routing solution that maximizes the power saving by turning off line-cards while satisfying performance constraints including link utilization and packet delay. The aforementioned strategies can improve the performance at a certain level, because there is always a tradeoff between the performance and the amount of energy consumption. The strategies discussed in the above sections are subject to graceful performance degradation. Therefore, a thorough analysis is always required to decide which one is more desired.

4. Data center performance monitoring

To ensure 24 × 7 smooth operation of DCNs, it is of extreme importance to monitor and manage various critical services and hardware running at data centers. With the evolution in communication technology and exponential growth in DCN subscribers, there has been a continuous research in the development of smart applications and hardware to improve the monitoring and management capabilities of data centers. The basic goals of monitoring

applications are to: (a) optimize the data center energy consumption through energy audits and smarter distribution, (b) optimize the network performance through real-time traffic monitoring to enable congestion-free routing of network data, and (c) proactively control the failure of a particular device or a service by forecasting the future behavior, based on the historical logs of events.

In the following subsections, we discuss various methodologies in practice to monitor data centers for achieving the aforementioned optimizations in terms of energy, network performance, and devices'/services' performance.

4.1. Data center monitoring for energy efficiency

A vital role of the data center's monitoring architecture is to observe and optimize the energy usage. With the growing demands of energy efficient solutions, significant progress has been made in the research and development of energy monitoring services. Monitoring services facilitate organizations with a live tracking of power usage, thermal map, and intelligent power distribution to individual devices. Monitoring services also provide thermal and power signature and traces of the data center. The information can be used by optimizer modules and schedulers for efficient and optimized resource allocation for energy savings. The cooling infrastructure is one of the major energy consumers within the data center. Smart monitoring services enable the intelligent cooling mechanisms resulting in considerable energy savings [94,95].

Liang et al. [96] presented a *Power Manager* energy monitoring service in their *GreenCloud* data center solution. Power Manager collects the energy usage information from Intelligent Power Distribution Units (iPDUs). The iPDUs are the devices installed on physical machines of the data center to perform real-time energy monitoring. The monitoring information is stored into a database for visualization, analysis, and mining. The authors analyzed and extracted the temporal patterns of energy usage of different devices from the database. The power usage patterns allowed the authors to devise efficient energy distribution models for various machines used in the *GreenCloud*.

Cisco Systems reported an annual saving of nearly 1 million dollars by employing an energy management and monitoring services technology named *EnergyWise* [97]. Cisco's *EnergyWise* is a cross-platform energy management architecture that automates the measurement, monitoring, and management of power consumption of various IT equipment, including Ethernet devices, IP phones, wireless access points, and lighting controllers. The intelligent energy consumption policies and practices adopted by the *EnergyWise* provide efficient power distribution mechanisms to networked devices. *EnergyWise* can manage around 18,000 devices resulting in estimated savings of around 1.3 million dollars annually.

Hewlett Packard's (HP) Energy and Sustainability Management (ESM) program is intended to provide a monitoring architecture that includes services to observe and report energy usage and carbon emissions in buildings, data centers, and supply chains [98]. The ESM will enable organizations to determine intelligent ways of energy control and distributions, with a goal of reducing overall energy consumption.

Intel installed a control and monitoring service named Building Management System (BMS) in one of the data centers in India. The BMS is comprised of various power usage meters for networking devices and thermal and air flow sensors. Based on the monitored data, the BMS plans efficient power distribution within the data center, resulting in significant energy savings.

4.2. Network performance monitoring

DCNs are required to deliver high bandwidth with lower latency to meet the growing demands of users. Network efficiency

Table 7
Data center network performance characteristics and tools.

Reference	Network performance monitoring	Data collection	Congestion control	Network troubleshooting	Tool	Decision making
[100]	✓	TCP statistics	✓	✓	Web100	Centralized
[101]	✓	Socket log, TCP statistics	N/A	✓	SNAP	Distributed
[99]	✓	Congestion feedback and network width of each entity	✓	✓	Seawall	Centralized
[102]	✓	Packets, bytes, and flow duration	✓	N/A	DevoFlow	Threshold on per-flow counters
[103]	N/A	N/A	✓	N/A	MPTCP	Established multiple subflows

increases the optimum utilization of the available network resource regardless of the number of flows, protocols, and destinations in the application traffic mix [99]. To address the scalability requirement of the network, different policies are required at the controller and switch level to avoid congestion. The reliability of the DCN is based on the ability of the network to deal with real-time problems in a lower cost. Network performance of a data center can be optimized with the right choice of servers and the routing path to respond the group of clients. The network load monitoring also helps in the precise scheduling of auto shutdown of under-utilized network equipment, which results in the significant savings of energy from cooling processes and power supply to devices. In the following paragraphs we discuss the various techniques practically utilized to monitor the network performance. Table 7 summarizes the network performance monitoring techniques.

Web100 is a measurement tool that helps in the collection of statistics for proper identification of performance limiting factors [100]. Web100 accesses the network condition at the Transport layer to segregate network congestion from other factors to produce more accurate estimates of the path performance. The collected data is classified into 4 categories based on performance limitation characteristics: (a) Network path, (b) Server network stack, (c) Clients, and (d) DCN server applications [100]. A percentage of connections having TCP performance problems are obtained by calculating the connections' time fraction for each performance bottleneck. The network graphs generated through Web100 help in finding out the problematic links and allow network administrators to efficiently distribute the network traffic without congestion.

Yu et al. [101] presented SNAP (Scalable Network Application Profiler), a tool used to guide the developers to identify and fix the performance problems. SNAP collects TCP statistics, socket-level logs of application, and read/write operations in real-time to classify and correlate the data to pinpoint the problematic links with low computation and storage overheads. SNAP only logs the TCP statistics on problem arrival and can dynamically tune the polling rate for different connections. Moreover, SNAP focuses on profiling the interactions between applications and the network, and diagnosing network performance problems. Having the ability to precisely detect the performance degradation of individual connections, SNAP can be a good resource for administering the link errors in a DCN.

Seawall is a network bandwidth allocation tool that divides network capacity based on administrator defined policy [99]. Seawall analyzes the traffic flow using a logical tunnel with the allocated traffic rate. Seawall receives congestion feedback from the tunnel. Bandwidth allocator, which is a component of Seawall, utilizes the feedback from all of the tunnels of the DCN to redistribute unused traffic shares proportionally to active sources in the network. The bandwidth allocator is responsible for regenerating the allowed rate on each entity tunnel in the network. Seawall is a useful tool to monitor the network traffic and generate feedback for congestion control to maximize the utilization of the network resources.

DevoFlow is a network traffic manager presented by Curtis et al. in [102]. DevoFlow is a simplified approach of the OpenFlow model.

Although, OpenFlow supports fine-grained flow-level control for optimal management of network traffic, the model incurs large monitoring overheads. Instead, DevoFlow allows the switches to make local routing decisions and forward flows where controller is not required. DevoFlow creates wild-card rules to pre-authorize certain sets of flows and install these rules to all switches. The controller in DevoFlow can define flow categories that demand pre-flow vetting. DevoFlow devolves the control to a switch by rule cloning (such that all packets matching a given rule are treated as one flow) and local actions (rules that provide autonomy to the switch to perform local routing actions without invoking the controller). The DevoFlow statistics collection mechanism helps to find significant flow in a particular domain, allowing scalable implementation with reduced number of flows. The controller installs flow-table entries at the switches on the least-congested path between flow endpoints after the controller detects elephant flow. The DevoFlow scheduler re-routes the elephant flows. DevoFlow reduces the burden of the controller and establishes flexibility in the policy implementation.

Multipath TCP (MPTCP) is considered effective for utilization of available bandwidth, improved throughput, and better fairness for different data center topologies. Each MPTCP end system paired with congestion controller is used to redirect the traffic from heavily congested path to lightly congested path. The approach of MPTCP can effectively utilize the dense parallel network topologies, significantly improving throughput and fairness compared to single-path TCP. Compared to single-path TCP that utilizes Randomized Load Balancing (RLB) to randomly choose a path for each flow, MPTCP can establish multiple sub-flows across different paths between the same pair of endpoints for single TCP connections.

4.3. Data center services/devices monitoring

In recent years, various models have been proposed (e.g., [104–106]) to perform the real-time monitoring of network devices and services. Two basic mechanisms utilized in most of the proposed schemes are Complex Event Processing (CEP) and machine learning algorithms. In CEP, special monitoring queries are utilized to extract various patterns of information from a continuous stream of events, generated by various devices and applications. Analysis of such patterns helps in finding the root cause of a problem and forecasting the problems that are more likely to occur in future. Moreover, such information can also be useful for smart power distribution to data center equipment to achieve energy efficiency. Machine learning techniques are applied to classify various types of network anomalies and to autonomously generate routing decisions on real-time traffic to avoid congestion problems.

Narayanan et al. [104] proposed a data center monitoring architecture that provides an integrated view of the real-time monitoring information from various independent devices. The architecture has three basic components namely: (a) monitoring tools, (b) event collector engine, and (c) CEP engine. Monitoring tools gather statistics from devices. Event collector engines collect events from monitoring tools. The events are converted into

Table 8
A comparison of various data center monitoring techniques.

Reference	CEP	Machine learning	Monitoring software	SNMP	Robots/sensors	Thermal sensor	Devices
[104]	✓	✗	Cayuga	✗	✗	✗	SAN/SAN switches
[105]	✗	✗	N/A	✗	✗	✗	VM servers
[106]	✓	✓	N/A	✓	✓	✗	Servers
[108]	✗	✗	MEO	✓	✓	✓	Servers
[110]	✗	✗	N/A	✗	✓	✓	Servers

a format with specific syntax and semantics and sent to CEP engine. The CEP engine analyzes the received events and finds the correlation among various patterns to find the root cause of the problem. The authors used *Cayuga* [107], a general purpose event monitoring tool to perform evaluations using scenarios from real data center, and suggested that the proposed architecture helps in isolating faults in a data center's complex network.

Kutare et al. [105] proposed an online system for analysis and monitoring of large-scale data centers, specifically targeting the virtualized system environments. The authors [105] presented a hierarchical data center monitoring architecture. The three main components of the proposed architecture are: (a) agents, (b) brokers, and (c) zones. The agents are small software/components that are integrated in an application/hardware, and locally collect the monitoring data from the entity. In hardware, the agents are the specialized/customized device drivers that provide the required monitoring information. The agents may also act as sensors to monitor power levels of a device. In hierarchy, on top of the agents are brokers that analyze collected information (by agents) for possible anomaly detection. The brokers provide the generation of various graphs after filtering out the information on specific parameters from the aggregated data. The brokers either run on VMs, dedicated host, or both. Moreover, the zones are the physical machines that are the host virtual machines. The authors presented an overlay monitoring service that manages various brokers and perform the underlying communication and filtering of information provided by the brokers. The basic purpose of the monitoring service is to ensure that the applications are fulfilling the SLA requirements and operate within the set threshold of the quality of service. Furthermore, the service also initiates the corrective actions in case of anomaly detection.

Ref. [106] applied machine learning algorithms in combination with CEP to perform the analysis of streams of real-time monitoring data to detect anomalies. The proposed monitoring architecture has three major components: (a) CEP engine, (b) machine learning module, and (c) visualization and storage. The CEP engine processes the streams of monitoring data to distinguish various patterns (such as alerts) to identify anomalies in the device/application. The machine learning module takes as input, the CEP processed information in the form of time series data and works in an unsupervised manner to detect anomalies. Various monitoring reports are displayed through visualization and storage component. The authors employed a customized version of SNMP to collect monitoring information from routers, servers, switches, hubs, bridges, temperature sensors, and host computers depending on a predefined set of parameters. The proposed system was tested to monitor a frequently accessed TV show website hosted by the largest ISP of Brazil (Globo.com), and the results indicated that the proposed system is successful in anticipating failures.

The authors in [108] utilized sensors, such as infrared cameras, temperature sensors, and humidity sensors. The sensors generate thermal maps of various locations within the data centers and send reading to energy optimization software "IBM Tivoli Maximo for Energy Optimization (MEO)" [109]. The live monitoring of temperature/moisture data from various locations helps to automate the cooling process in data centers, which prevents the risk of damage to devices due to overheating. Similar approaches were proposed

by Liang et al. [110] that utilize cyber physical systems to monitor and automate the cooling process, to achieve energy efficiency.

Table 8 provides a brief comparison of the techniques discussed above. A few parameters selected for comparison are: (a) CEP, (b) machine learning, (c) monitoring software, (d) SNMP, (e) robots, (f) thermal/humidity sensing, and (g) monitored devices.

5. Energy-aware resource allocation

Underutilization of servers in a data center is a major cause of higher power consumption. Higher number of running servers result in higher power consumption [111]. Therefore, optimal utilization of servers will result in lesser number of turned on servers and high power efficiency. Virtual Machine (VM) placement is an important characteristic of data center that consolidates the servers resulting in cutback of the amount of hardware usage. The general approach for handling the VM placement problem is to have a mathematical representation or a metric of resource utilization [112]. Mapping a VM correctly to a PM is based on the capacity of the PM and the resource requirements of the VM [113]. VM placement is an important research domain in data centers where provisioning is performed manually [113–115]. This section explains network based VM placement for increased throughput, low latency, and energy efficiency.

5.1. Virtual machine placement in data centers

Recently, theoretical server virtualization approaches have been applied in production environments, because server virtualization ensures massive savings in hardware, power consumption, and cooling [112]. In this section, different VM placement approaches and a comparison of identified features are discussed.

Meng and Watson [116] designed a two-tier estimated technique called TVMPP accompanied by a proportional analysis for performance increase of VM placement based on different traffic patterns and network architectures. To evaluate the proposed technique, different traffic traces of a real data center are used. The inputs to TVMPP algorithm are traffic and cost matrix. Because of the complexity of TVMPP algorithm, the authors utilized Local Optimal Pairwise Interchange (LOPI) [116] and Simulated Annealing [116] heuristic algorithms. The partition step splits VMs and hosts into separate groups. After the partitioning step, TVMPP matches VMs and hosts at cluster and individual level. To verify the effectiveness of TVMPP algorithm, the authors considered four different architectures with diverse traffic patterns. Moreover, based on the traffic patterns, the authors used the data and impact study to estimate the results of the proposed algorithm.

The problem of VM placement and routing for data center network are addressed separately and the benefits of a joint design are hardly explored. In this regard, Ref. [117] addressed a VM placement and route selection problem by exploiting multipath routing capability and dynamic VM migration. The authors propose: (a) an offline algorithm and (b) an online solution for the management of network resources and VM placement in a dynamic environment with changing traffic. Based on the Markov approximation, the proposed algorithm is tailored to the data center architecture taking VM dynamics into consideration and requires a very small number

of VM migrations. Synthesized data center traffic traces are used for performance analysis. The analysis on various topologies under a spectrum of elephant and mice workloads demonstrates a consistent and significant improvement over the benchmark achieved by common heuristics used in today's data centers.

Given a heterogeneous hardware infrastructure, assigning VMs to PMs is a challenging task. Ref. [118] proposed a service that realizes a two-phase VM-to-PM placement scheme. The first phase identifies a promising group of PMs, named Cohort, that hosts the virtual infrastructure of the user request. The second phase determines the final VM-to-PM mapping considering all low-level constraints arising from: (a) particular user requests and (b) special characteristics of the selected Cohort. The evaluation demonstrates a significant reduction to the VM placement plan production time and improved plan quality in large heterogeneous physical infrastructures. Compared to a heavyweight monolithic approach, the scheme can scale up to several hundreds of physical nodes. Moreover, the quality of the deployment plans remains largely unaffected by the size of the physical infrastructure. The overall plan quality is improved because latency amongst deployed VMs is reduced and the average bandwidth increases dramatically.

Mihailescu et al. [119] proposed OX that allows application owners to specify groups of highly available VMs, following component roles and replication semantics. To discover application topologies, OX transparently monitors network traffic among VMs. Moreover, OX builds online topology graphs for applications and incrementally partitions the graphs across the infrastructure to enforce availability constraints and optimize communication between VMs. OX uses the information to dynamically implement VM placement optimizations to enforce application availability constraints and reduce and/or alleviate application exposure to network communication anomalies, such as traffic bottlenecks. The authors evaluate OX in a realistic cloud setting using a mix of Hadoop and YCSB/Cassandra workloads and show how OX increases application robustness, by protecting applications from network interference effects and rack-level failures. The results demonstrate that OX enhances application resilience to infrastructure anomalies, by reducing the impact of inter-rack network interference, and by shielding applications from rack-level failures.

Balaji et al. [120] improved MapReduce job performance by (a) refining data locality, (b) coupling information, and (c) VM placement. Purlieus require virtual MapReduce clusters for VMs' admission and intermediate information from local or nearby physical machines as input. The locality-awareness feature improved the runtime performance of individual jobs and had an additional advantage of reducing network traffic. The demerits of conventional architectures result in job execution delay and data duplication. In contrast, Purlieus store the information in a dedicated MapReduce cloud and execute job requests on the same machines that stored the data as well as storage cloud. The authors conducted a detailed evaluation of Purlieus and confirmed considerable improvements in network traffic with 50% decrease in job execution times for varying workloads. Purlieus solve the locality problem by combining data placement with VM placement. The experiment results show significant performance improvements with some scenarios showing up to 50% reduction in job execution time and around 70% reduction in the cross-rack network traffic.

Ref. [91] presented the design, implementation, and evaluation of Cloud Networking-as-a-Service (CloudNaaS). CloudNaaS is a networking framework that extends the self-service provisioning model of the cloud beyond virtual servers and storage to include a rich set of accompanying network services. One of the key optimizations in CloudNaaS is the joint placement of VMs with virtual network segment provisioning. The programmable network devices used in the design provide the fine-grained control required

for per-customer network services. CloudNaaS introduces network controller component into the cloud management system. The network controller is responsible for configuring virtual network segments throughout the cloud by mapping the logical requirements in the communication matrix onto the physical network resources. Moreover, the network controller also controls resources by determining VM placement and performing re-mapping when available resources change, to ensure that tenant requirements are consistently satisfied in an efficient manner.

McGeer et al. [121] proposed techniques such as traffic aggregation and VM placement to save significant network energy by powering off unused switching elements. The authors formulated minimization of network power consumption for any general data center topology as a combinatorial optimization problem and demonstrate that the general problem is NP-complete. To realize the network power savings, the authors visualized a centralized network power controller program in the data center. The controller gathers traffic data and server statistics from all the switches and servers in the data center, and uses the information to perform traffic aggregation and VM placement and migration. The traces are collected from a production data center hosting an e-commerce application. Moreover, the System Activity Reporter toolkit is used to find the number of bytes transmitted and received from 292 servers. The traces are used to: (a) determine the best placement for the VMs, and (b) compute bandwidth savings.

Zhang et al. [122] utilized locality property for avoiding traffic propagation to higher-level networks. Moreover, Ref. [122] performed experimental evaluation of two DCN architectures, FiConn [123] and fat-tree [42], in a three-tier transaction system with a cluster based virtualized implementation.

Moreover, the authors proposed a VM placement scheme called ideal placement. Similar to FiConn, the ideal placement for fat-tree is defined as a setting where servers belonging to the same service tier (i.e., Web, application, and database) are placed in the same Pod. The two-server placement scheme does not impose a significant impact on application performance in the fat-tree architecture. In an ideal placement, fat-tree yields lower throughput under 512 and 1024 clients than FiConn. The aforementioned process can be attributed to the overheads of Click router's user-level emulation of fat-tree's two-level routing algorithm.

In this section, different approaches have been discussed to address the problem of VM placement in modern data centers. Under the category of VM placement techniques, the aforementioned approaches focus on various criteria. The criteria range from the consideration of topology and the bandwidth to the type of workload. A number of features are identified and a brief comparison is performed among the approaches implemented in the data center. Table 9 summarizes the different approaches with the comparison of common features identified among the different discussed approaches. The selected features are: (a) topology, (b) bandwidth, (c) latency, (d) scalability, and (e) workload.

Virtualization is one of the major techniques for energy efficiency in data centers [20,124]. It has been reported in a survey that 50% of the data center professional survey respondents reported the energy saving using virtualization [124]. In a data center, with the efficient utilization of VMs and server consolidation, the total energy consumption for servicing the clients can be reduced with negligible performance degradation [122]. From an energy efficiency viewpoint, it is practical to have as few servers as possible turned on, with each powered on server being highly utilized and energy proportional. Moreover, virtualization technology offers the feasibility of application and server independence.

System-wide resource allocation and server consolidation are deemed necessary to minimize the energy consumption [121]. The integration of virtualization and energy management is critical for overall system design because existing energy optimization techniques require modifications to be effective in a VM environment [121].

Table 9
VM placement features' comparison.

Reference	Topology	Bandwidth	Latency	Scalability	Workload
[116]	Tree,VL2, Fat-Tree, BCube	VM to VM	Average Traffic Latency	Optimize VM Placement	Enterprise and MapReduce
[117]	Clique, Fat-Tree, HyperX, and BCube	Bisection Bandwidth	×	Fewer Switches and Smaller Hop counts	Elephant and mice
[118]	LAN and Star Network	Average Network Bandwidth	Packet hop count	Dynamically-Formed Cohorts	LIGO Inspiral Analysis Generator
[119]	Application Topology and Online Topology Graphs	Inter-rack Link and Network Traffic	Application	Per-application Incremental Adaptation	Hadoop and YCSB/Cassandra
[120]	Large Cloud Scale Data center	Inter-rack Link	×	Size of Data center Network, Number of VMs, Workload Type	MapReduce
[91]	Canonical 3-tier network, Fat-Tree	Link Bandwidth and Bandwidth along a Path	Average Network	Multi-tier Interactive and Batch Application Models	Interactive Multi-tiered and Batch Multi-tiered Applications
[121]	n-level Fat-Tree	Core-switch, Bisection bandwidth, Outbound Bandwidth	×	×	×
[122]	Fat-Tree	×	Response Latency	×	RUBiS Workload Generator

Table 10
Experimentation techniques' comparison.

Reference	Network validation tool	Software based	Hardware based	Dedicated machines	Slow down time	Reproduce traffic characteristics	Target energy efficiency	Number of nodes
[126]	✓	✓	×	✓	✓	✓	×	1000
[127]	✓	×	✓	×	✓	✓	✓	10 000
[139]	×	✓	✓	✓	×	×	×	×
[128]	✓	✓	×	✓	×	✓	×	1000
[140]	✓	✓	×	×	×	✓	×	×
[129]	×	✓	×	×	×	×	✓	×
[1]	✓	✓	×	×	×	✓	×	4096
[137]	×	✓	×	×	×	×	✓	×
[138]	×	✓	×	×	×	×	✓	×
[141]	×	✓	×	×	×	×	✓	×
[142]	×	✓	×	×	×	×	✓	×
[43]	×	✓	×	×	×	×	✓	×
[143]	×	✓	×	×	×	×	✓	×
[144]	×	✓	×	×	×	×	✓	×

6. Experimentation techniques in data centers

During the design and development phase of policies for the Data Center Networks (DCNs), the experimentation and simulation serve as the keys to find and understand the critical balance between energy efficiency and performance. Experimentation is also critical to prove a new developed design. Experimentation and simulation techniques save lots of time and efforts that could have been used in the setup of real DCN. In the literature, many platforms have been proposed and are available for evaluating the energy efficiency of a DCN. Each technique differs from the other in many aspects [125]. Emulation systems aid in the experimentation of real networks, providing a tool to test new protocols and applications. Simulators reproduce the behavior of a network [126]. Some emulators are software based that reproduces the behavior of the network, while others create hardware tools that physically generate conditions found in certain networks. The hardware approach allows the experimenter to control different features in the network, such as buffer sizes, line rates, and network topologies providing a more realistic environment than software based emulators [127]. In this section, experimentation techniques to analyze DCNs are discussed in detail. The comparisons of the experimentation techniques are presented in Table 10.

Simulation environments facilitate the experimentation of various network features, allowing cost effective future network technologies' research. The performance of the emulator is reliable and provides a powerful tool for energy modeling, resource consumption, and timing modeling [127]. Although, the experimentation tools allow the emulation of future network environments, the tools are still simple models and are limited to short path routing via static routes [126].

Ref. [126] presented an emulation environment, named as ModelNet. The environment was based on software and took advantage of dedicated machines to mimic both user-specified network topologies and end-user applications with transport protocols. The model was used to develop various Transport and Application level protocols including: (a) Distributed Hash Tables (DHTs), (b) overlays, (c) content distribution protocols, (d) smart switch architectures, and (e) bandwidth estimation tools. ModelNet uses two main applications: (a) DieCast and (b) Swing. DieCast leverages time dilatation that allows exploring the impact of a potential upgrade to 10 Giga bit Ethernet (GigE) equipment only using 1 GigE technology. Moreover, Swing generates traffic that allows emulating time dilatation and other simulation features, such as traffic across a range of timescales and tunable traffic. ModelNet is still an approximation of a real network with limited number of nodes, limited packet processing, and shortest path routing via static routes. The energy model is not incorporated in the emulator.

In [128], the authors demonstrated DCR, a Linux based software approach that served as a debugging system for data center applications. The proposed system reproduced the regular system behavior amalgamated with the realistic non-deterministic failures and interferences, such as node failures, network partitions, unforeseen maintenance, or program misbehaviors. DCR allows practical replay-debugging of large-scale, data-intensive distributed systems. Energy efficiency was not the main target of the proposed approach.

Ganapathi [129] studied the state-of-the-art workload generators, such as SURGE [130], SPECweb [131], TPC-W [132], SLAMD [133], Harpoon [134], Optixia [135], and Hammer [136]. The work in [129] revealed that these workload generators do

not perform well in a black-box system. The authors presented a framework capable of generating custom workloads using an algorithm that combined metrics of interest, such as clustering, static workload, stratified sampling, dynamic workload, future behavior, and synthetic workload generation. The analysis presented in the aforementioned research works was limited to data collected in a single day, to prevent memory exhaustion due to the large scale of collected data, and excessive tweaking of the used data. Workload models can be generated specific to the metric of interest, such as performance, power, throughput, user-response latency, or any combination of the above.

Lim et al. [137] presented a simulator, named MDCSim, to evaluate the energy efficiency of data centers. The simulator evaluates the power consumption on a per-server basis based on the response time of data center jobs. Both the performance and energy consumption can be evaluated on MDCSim, using different scheduling policies. However, power saving policies such as DVFS are not supported by MDCSim.

Meisner and Wenisch [138] improve the MDCSim [137] by estimating the average of data center job response times and named the new model as Stochastic queuing simulation (SQS). The authors developed the model on the basis of the assumption that an average expected response time should be equal to an average expected service time as data center servers normally operate at low utilization level. Although the power model used in SQS is the same as MDCSim, the estimation of DVFS is included in the model. The drawback of the proposed model lies in the absence of connection between power model and performance model.

Ellithorpe et al. [127] presented an FPGA based emulation tool. The solution provided a more accurate emulation framework, however at a higher computational complexity. Individual components such as the node and switch were modeled. The system allowed: (a) monitoring resource consumption, (b) slowing down the time clock (similarly to software based emulators), (c) easy scalability with multiple FPGAs, and (d) energy modeling. The simulation environment is proposed as a Computer-Aided Design (CAD) tool to explore and research custom networking architectures. The work reported in [139] implemented a data center using on-chip solutions. On-chip environments offer several advantages, such as better power efficiency, better interconnection, better utilization, better performance of parallel applications, and reduced heterogeneity.

Ersoz et al. [140] discussed the mathematical parameters needed for the behavior modeling in a data center. The authors described the characteristics of the network behavior within a clustered, multi-tiered data center, and focused the inter-tier network traffic. The authors developed a three-tier prototype to study some of the common workloads in multi-tier data centers. The front-end tier is a cluster of web servers, the middle-tier is a cluster of application servers responsible for handling dynamic web contents, and the back-end tier handles database transactions from the middle-tier. The authors concluded that, in most cases, the message rates between tiers and message sizes follow a log-normal distribution, and service times can be modeled following the Pareto distribution. Energy efficiency is not the main target of the model.

Bilal et al. [1] simulated the state-of-the-art DCN architectures using the ns-3 simulator. The authors presented a comprehensive analysis of: (a) fat-tree, (b) recursively defined DCell, and (c) legacy three-tier DCN architectures. The authors compared the aforementioned DCN architectures under different network configurations and traffic patterns. The analysis revealed that the fat-tree architecture outperforms the DCell and three-tier architectures in terms of network throughput and latency. The analysis also demonstrated that the performance of the DCell architecture is heavily dependent on the network size [1]. Energy efficiency was not the main goal in [1].

Calheiros et al. [141] developed the simulator, named as CloudSim, that not only supports different power consumption models but also supports power saving policies. The main energy consumption model of CloudSim is similar to MDCSim [137] and SQS [138]. CloudSim is a virtual machine level model and has the drawback of not being able to assess the conflicts and degradation of performance or more energy requirement in case of running multiple virtual machines on the same server.

A Virtual Machine based model, named EEFSim, was proposed by Julia et al. [142] in 2010. The model aims at finding the effects of scheduling and migration algorithms for VMs on power consumption. EEFSim constructs a power model using CPU utilization only and cannot estimate the performance model. EEFSim is also not scalable and performs well with more than 5000 servers.

Kliazovich et al. [43] proposed the model for data centers, named GreenCloud, that can simulate different power saving mechanisms for network elements. GreenCloud not only estimates the total energy consumption of the data center but it also estimates the energy consumption of different individual devices such as switches in the network. The drawback of GreenCloud is the lack of description about performance model.

Gupta et al. [143] presented the simulator, named GDCSim, which estimates the energy efficiency of data centers. Scheduling algorithms in GDCSim are user defined. Support for power management like CPU sleep state transition and DVFS are included in GDCSim. The thermal model of GDCSim controls the cooling units and contributes in significant energy saving. The major drawbacks of GDCSim include its simple scheduler and lack of ability to estimate the interference effect of different jobs.

Aksanli et al. [144] presented the simulator, named GENSim, that estimates the impact of both services and batch job on a single job. GENSim estimates the power usage and energy consumption as CPU and memory resource usage.

Ref. [125] presented the comparison of most of the simulators for green data centers. The authors have included only software based simulators. We have included not only the software but also hardware based simulators and experimentation techniques for data center networks.

Table 10 presents a summary of the addressed papers in this section. The references are displayed as rows and the characteristics are presented in columns. Network validation tool refers to those references that proposed emulators as a tool to validate a certain data network with a fixed number of features such as node speeds and number of tiers. Then, the emulators are divided into hardware based or software based. However, there are cases that the reference addresses a model that includes both hardware and software components. The dedicated machines' columns indicated if the work's emulator specified the use of only dedicated machines to simulate network data centers. The slowdown time, presented in [126,127], indicates that the system can manipulate clock time to emulate a slowing down action in simulation time.

Experimentation environments are powerful and useful tool to reproduce the behavior and estimate the energy consumption of a data center network and applications without having to spend on a great amount of resources. Most of the experimentation environments are software based and are developed especially to optimize energy distribution in data centers. The primary role of such simulators is to bifurcate the various sections of a data center to observe the energy flow at the fine-grained level. This helps in the identification and isolation of devices and equipments that consume the energy despite being idle most of the time. For example, the servers and the switches that remains in ON state at times of lesser loads, such as during night time. The reports generated by such simulations will help in proper planning and designing of DCNs to ensure the fair distribution of workload and energy.

7. Conclusions

Data center networks play a pivotal role in the data center scalability and performance bounds. Legacy data center networks lack the capability to fulfill current user and application demands. DCNs are facing numerous challenges in the network domain. Extensive research is being conducted to deal with the future data center networking challenges. This survey presented a comprehensive review on the state-of-the-art research in the domain of data center networks. The survey delivered significant insights in different areas of DCN research. Various research areas, such as DCN architectures, DCN monitoring, network management, network traffic engineering, and experimentation techniques within the DCN domain, were highlighted in the survey. Summarized comparisons of the aforementioned areas were presented and research challenges are highlighted in various sections. The survey will significantly help the researchers to understand the broad range of research focus and challenges in the DCN. Several research streams are expected to emerge from the comprehensive discussion in the survey. The researchers working in the areas but not limited to: (a) DCN architecture, (b) DCN management, (c) DCN monitoring, (d) DCN algorithm developers, (e) Green networks, (f) DCN traffic engineers, and (g) Green DCN may greatly benefit from the presented discussion.

Acknowledgments

References

[1] K. Bilal, S.U. Khan, L. Zhang, H. Li, K. Hayat, S.A. Madani, N. Min-Allah, L. Wang, D. Chen, Quantitative comparisons of the state of the art data center architectures, Concurrency and Computation: Practice and Experience 25 (12) (2013) 1771–1783. http://dx.doi.org/10.1002/cpe.2963.

[2] J. Shuja, S.A. Madani, K. Bilal, K. Hayat, S.U. Khan, S. Sarwar, Energy-efficient data centers, Computing 94 (12) (2012) 973–994.

[3] D. Kliazovich, P. Bouvry, S.U. Khan, DENS: data center energy-efficient network-aware scheduling, Cluster Computing 16 (1) (2013) 65–75.

[4] Hewlett-Packard, ExaScale data center, ExaScale Computing Lab. http://www.hp.com/hpinfo/newsroom/press_kits/2008/cloudresearch/fs_exascaledatacenter.pdf (accessed 15.09.2012).

[5] K. Bergman, Optically interconnected high performance data centers, in: 2010 36th European Conference and Exhibition on Optical Communication, ECOC, September 2010, pp. 1–3.

[6] A. Carter, Do it Green: media interview with michael manos. http://edge.technet.com/Media/Doing-IT-Green/ (accessed 20.02.2012).

[7] L. Rabbe, Powering the Yahoo! Network. http://yodel.yahoo.com/2006/11/27/powering-the-yahoo-network/ (accessed 20.02.2012).

[8] J. Koomey, Growth in Data Center Electricity Use 2005 to 2010, Analytics Press, 2011.

[9] D. Abts, M. Marty, P. Wells, P. Klausler, H. Liu, Energy proportional datacenter networks, in: Proceedings of the 37th International Symposium on Computer Architecture, Saint Malo, France, June 2010, pp. 338–347.

[10] J. Baliga, K. Hinton, R.S. Tucker, Energy consumption of the Internet, in: Joint International Conference on Optical Internet, 2007 and the 32nd Australian Conference on optical FibreTechnology, COIN-ACOFT'07, June 2007, pp. 1–3.

[11] Global Action Plan, An Inefficient Truth, Global Action Plan Report, 2007, http://globalactionplan.org.uk (accessed 7.11.2011).

[12] Gartner, Gartner Says Data Centers Account for 23% of Global ICT CO2 Emissions, Press Release, 11 October 2007. http://www.gartner.com/it/page.jsp?id=530912 (accessed 7.11.2011).

[13] Molly Webb, Smart 2020: enabling the Low Carbon Economy in the Information Age, the Climate Group and Global e-Sustainability Initiative, 2008.

[14] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, B. Maggs, Cutting the electric bill for internet-scale systems, in: Proceedings of the Conference on Data Communication, SIGCOMM'09, August 2009, pp. 123–134.

[15] M. Gupta, S. Singh, Greening of the Internet, in: Proceedings of the ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, SIGCOMM'03, August 2003, pp. 19–26.

[16] W. Fisher, M. Suchara, J. Rexford, Greening backbone networks: reducing energy consumption by shutting off cables in bundled links, in: Proceedings of 1st ACM SIGCOMM Workshop on Green Networking, SIGCOMM'10, August 2010.

[17] A. Bianzino, C. Chaudet, D. Rossi, J.-L. Rougier, Energy-awareness in network dimensioning: a fixed charge network flow formulation, in: International Conference on Energy-Efficient Computing and Networking, e-Energy'10, April 2010.

[18] P. Mahadevan, P. Sharma, S. Banerjee, P. Ranganathan, Energy aware network operations, in: Proceedings of the IEEE Global Internet Symposium, April 2009.

[19] B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, N. McKeown, ElasticTree: saving energy in data center networks, in: Proceedings of the 7th USENIX Symposium on Networked System Design and Implementation, NSDI'10, April 2010, pp. 249–264.

[20] K. Bilal, S.U. Khan, S.A. Madani, K. Hayat, M.I. Khan, N. Min-Allah, J. Kolodziej, L. Wang, S. Zeadally, D. Chen, A survey on green communications using adaptive link rate, Cluster Computing 16 (3) (2013) 575–589. http://dx.doi.org/10.1007/s10586-012-0225-8.

[21] Energy Efficient Ethernet IEEE 802.3az. http://standards.ieee.org/getieee802/download/802.3az-2010.pdf (accessed 07.11.2011).

[22] K. Christensen, P. Reviriego, B. Nordman, M. Bennett, M. Mostowfi, J.A. Maestro, IEEE 802.3az: the road to energy efficient Ethernet, IEEE Communications Magazine 48 (11) (2010) 50–56.

[23] M. Zhang, C. Yi, B. Liu, B. Zhang, GreenTE: power-aware traffic engineering, in: IEEE ICNP 2010, Kyoto, Japan.

[24] C. Panarello, A. Lombardo, G. Schembra, L. Chiaraviglio, M. Mellia, Energy saving and network performance: a trade-off approach, in: Proceedings of e-Energy 2010, e-Energy'10, April 2010.

[25] R. Bolla, R. Bruschi, A. Cianfrani, M. Listanti, Enabling backbone networks to sleep, IEEE Network 25 (2) (2011) 26–31.

[26] L. Chiaraviglio, M. Mellia, F. Neri, Reducing power consumption in backbone networks, in: IEEE International Conference on Communications, ICC'09, June 2009, pp. 1–6.

[27] G.L. Valentini, S.U. Khan, P. Bouvry, Energy-efficient resource utilization in cloud computing, in: A.Y. Zomaya, H. Sarbazi-Azad (Eds.), Large Scale Network-Centric Computing Systems, John Wiley & Sons, Hoboken, NJ, USA, September, 2013.

[28] A.N. Khan, M.L.M. Kiah, S.U. Khan, S.A. Madani, Towards a secure mobile cloud computing: a survey, Future Generation Computer Systems 29 (5) (2013) 1278–1299.

[29] D. Zinn, Q. Hart, B. Ludascher, Y. Simmhan, Streaming satellite data to cloud workflows for on-demand computing of environmental data products, in: IEEE Conference on Workflows in Support of Large-Scale Science, 2010, pp. 1–8.

[30] P. Watson, P. Lord, F. Gibson, Cloud computing for e-science with CARMEN, in: Proceedings of IBER GRID Conference, Portugal, May 12–14, 2008, pp. 1–5.

[31] S. Kumar, S.E. Sarma, J.L. Willers, G.A. Milliken, Infrastructure for data-driven agriculture: identifying management zones for cotton using statistical modeling and machine learning techniques, in: IEEE Conference on Emerging Technologies for a Smarter World, November 2011, pp. 1–6.

[32] Miguel Delgado, The evolution of health care IT: are current US privacy policies ready for the clouds? in: IEEE World Congress on Services, 2011, pp. 371–378.

[33] Srisakdi Charmonman, Data Center for Ecommerce Business Continuity. http://inceb2009.su.ac.th/pdf/KeynoteSpeaker2.pdf (accessed 18.09.2012).

[34] A. Ipakchi, F. Albuyeh, Grid of the future, IEEE Power and Energy Magazine 7 (2) (2009) 52–62.

[35] X. Fang, S. Misra, G. Xue, D. Yang, Managing smart grid information in the cloud: opportunities, model, and applications, IEEE Network 26 (4) (2012) 32–38.

[36] N. Otuka, S. Dunaeva, The role of the nuclear reaction data centres in experimental nuclear data knowledge sharing, Journal of the Korean Physical Society 59 (2) (2011) 1292–1297.

[37] O. Schwerer, V. McLane, H. Henriksson, S. Maev, Nuclear reaction data centre network: a success story, in: Proceedings of the International Conference on Nuclear Data for Science and Technology, ND'04, Santa Fe, USA, September 2004.

[38] R. Bolla, R. Bruschi, F. Davoli, F. Cucchietti, Energy efficiency in the future Internet: a survey of existing approaches and trends in energy-aware fixed network infrastructures, IEEE Communications Surveys & Tutorials 13 (2) (2011) 223–244.

[39] Y. Zhang, P. Chowdhury, M. Tornatore, B. Mukherjee, Energy efficiency in telecom optical networks, IEEE Communications Surveys & Tutorials 12 (4) (2010) 441–458.

[40] P. Bianzino, C. Chaudet, D. Rossi, J.-L. Rougier, A survey of green networking research, IEEE Communications Surveys & Tutorials 14 (1) (2012) 3–20.

[41] S. Zeadally, S.U. Khan, N. Chilamkurti, Energy-efficient networking: past, present, and future, The Journal of Supercomputing 62 (3) (2012) 1093–1118.

[42] M. Al-Fares, A. Loukissas, A. Vahdat, A scalable, commodity data center network architecture, in: ACM SIGCOMM 2008 Conference on Data Communication, Seattle, WA, 2008, pp. 63–74.

[43] D. Kliazovich, P. Bouvry, S.U. Khan, GreenCloud: a packet-level simulator for energy-aware cloud computing data centers, Journal of Supercomputing 62 (3) (2012) 1263–1283.

[44] Cisco Data Center Infrastructure 2.5 Design Guide, Cisco Validated Design, March 9, 2010.

[45] A. Greenberg, J. Hamilton, D.A. Maltz, P. Patel, The cost of a cloud: research problems in datacenter networks, SIGCOMM Computer Communication Review 39 (1) (2008) 68–73.

[46] C. Guo, H. Wu, K. Tan, L. Shi, Y. Zhang, S. Lu, DCell: a scalable and fault-tolerant network structure for data centers, ACM SIGCOMM Computer Communication Review 38 (4) (2008) 75–86.

[47] C. Guo, G. Lu, D. Li, H. Wu, X. Zhang, Y. Shi, C. Tian, Y. Zhang, S. Lu, BCube: a high performance, server-centric network architecture for modular data centers, in: ACM SIGCOMM 2009 Conference, Barcelona, Spain, pp. 63–74.

[48] H. Abu-Libdeh, P. Costa, A. Rowstron, G. O'Shea, A. Donnelly, Symbiotic routing in future data centers, in: ACM SIGCOMM 2010 Conference, New Delhi, India, 2010, pp. 51–62.

[49] A. Singla, C. Hong, L. Popa, P. Godfrey, Jellyfish: networking data centers, randomly, in: 3rd USENIX conference on Hot topics in cloud computing, Portland, OR, pp. 12–17.

[50] H. Wang, A. Garg, Keren Bergman, Design and demonstration of an all-optical hybrid packet and circuit switched network platform for next generation data centers, in: OFC 2010, OTuP3, March 2010, pp. 1–3.

[51] X. Ye, Y. Yin, D. Ding, S. Johnson, V. Akella, S.J.B. Yoo, Assessment of optical switching in data center networks, in: OFC/NFOEC 2010, pp. 1–3.

[52] A. Vahdat, H. Liu, X. Zhao, C. Johnson, The emerging optical data center, in: Optical Fiber Communication Conference, Optical Society of America, 2011, pp. 1–3.

[53] A. Singla, A. Singh, K. Ramachandran, L. Xu, Y. Zhang, Feasibility study on topology malleable data center networks (DCN) using optical switching technologies, in: Optical Fiber Communication Conference and Exposition (OFC) and the National Fiber Optic Engineers Conference, NFOEC, 2011, pp. 1–3.

[54] C. Ware, C. Lai, D. Brunina, W. Zhang, A. Garg, B. Bathula, K. Bergman, Cross-layer reconfigurable optical network: fast failure recovery in testbed for routing algorithms in ICTON, we.B3.3, 2011, pp. 1–4.

[55] H.H. Bazzaz, M. Tewari, G. Wang, G. Porter, T.S.E. Ng, D.G. Andersen, M. Kaminsky, M.A. Kozuch, A. Vahdat, Switching the optical divide: fundamental challenges for hybrid electrical/optical datacenter networks, in: Procedings 2nd ACM Symposium on Cloud Computing, SOCC, October 2011, pp. 1–8.

[56] R. Proietti, X. Ye, Y. Yin, A. Potter, R. Yu, J. Kurumida, V. Akella, S.J.B. Yoo, 40 Gb/s 8 × 8 low-latency optical switch for data centers, in: Optical Fiber Communication Conference, OFC/NFOEC, Paper OMV4, 2011, pp. 1–3.

[57] L. Xu, W. Zhang, H. Lira, M. Lipson, K. Bergman, A hybrid optical packet and wavelength selective switching platform for high-performance data center networks, Optics Express 19 (24) (2011) 24258–24267.

[58] H. Wang, K. Bergman, A bidirectional 2 × 2 photonic network building-block for high-performance data centers, in: Optical Fiber Communication Conference and Exposition, OFC/NFOEC, 2011 and the National Fiber Optic Engineers Conference, March 2011, pp. 1–3.

[59] C. Kachris, E. Giacoumidis, I. Tomkos, Energy-efficiency study of optical OFDM in data centers, in: Optical Fiber Communication Conference and Exposition, OFC/NFOEC, March 2011, pp. 1–3.

[60] N. Farrington, G. Porter, S. Radhakrishnan, H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, A. Vahdat, Helios: a hybrid electrical/optical switch architecture for modular data centers, in: ACM SIGCOMM 2010 Conference, vol. 40, I. 4, October 2010, pp. 339–350.

[61] G. Wang, D. Andersen, M. Kaminsky, K. Papagiannaki, T. Ng, M. Kozuch, M. Ryan, c-through: part-time optics in data centers, in: ACM SIGCOMM'10, 40, I. 4, October 2010, pp. 327–338.

[62] D. Halperin, S. Kandula, J. Padhye, P. Bahl, D. Wetherall, Augmenting data center networks with multi-gigabit wireless links, in: ACM SIGCOMM 2011 Conference, vol. 41, I. 4, August 2011, pp. 38–49.

[63] K. Ranachandran, R. Kokku, R. Mahindra, S. Rangarajan, 60 GHz data-center networking: wireless ≥ worryless, in: NEC Laboratories America, Inc., Tech. Rep., July 2008.

[64] Vision and roadmap: routing telecom and data centers toward efficient energy use, in: Vision and Roadmap Workshop on Routing Telecom and Data Centers, 2009. http://www1.eere.energy.gov/manufacturing/datacenters/pdfs/vision_and_roadmap.pdf.

[65] A. Carrega, S. Singh, R. Bruschi, R. Bolla, Traffic merging for energy-efficient datacenter networks, in: International Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS'12, Genoa, Italy, July 8–11, 2012.

[66] P. Mahadevan, S. Banerjee, P. Sharma, A. Shah, P. Ranganathan, On energy efficiency for enterprise and data center networks, IEEE Communications Magazine 49 (8) (2011) 94–100.

[67] S. Figuerola, M. Lemay, V. Reijs, M. Savoie, B. Arnaud, Converged optical network infrastructures in support of future Internet and grid services using iaas to reduce GHG emissions, Journal of Lightwave Technology 27 (12) (2009) 1941–1946.

[68] R.S. Tucker, Green optical communications—part I: energy limitations in transport, IEEE Journal on Selected Topics in Quantum Electronics 17 (2) (2011) 245–260.

[69] R.S. Tucker, Green optical communications—part II: energy limitations in networks, IEEE Journal on Selected Topics in Quantum Electronics 17 (2) (2011) 261–274.

[70] J.Y. Shin, E.G. Sirer, H. Weatherspoon, D. Kirovski, On the Feasibility of Completely Wireless Data Centers, Technical Reports, Cornell University, 2011.

[71] Y. Cui, H. Wang, X. Cheng, B. Chen, Wireless data center networking, IEEE Wireless Communications 18 (6) (2011) 46–53. http://dx.doi.org/10.1109/MWC.2011.6108333.

[72] R. Mysore, A. Pamboris, N. Farrington, N. Huang, P. Miri, S. Radhakrishnan, V. Subramanya, A. Vahdat, Portland: a scalable fault-tolerant layer 2 data center network fabric, in: Proceedings of the ACM SIGCOMM 2009 Conference, Spain, 2009, pp. 39–50.

[73] K. Chen, C. Hu, X. Zhang, K. Zheng, Y. Chen, A.V. Vasilakos, Survey on routing in data centers: insights and future directions, IEEE Network 25 (4) (2011) 6–10.

[74] NetVeda LLC, Policy.Net: policy-based Network Traffic management, 2003. http://www.netveda.com/downloads/PolicyNet_Whitepaper.pdf (accessed 16.04.2012).

[75] T. Ye, S. Kalyanaraman, D. Harrison, B. Sikdar, B. Mo, H.T. Kaur, K. Vastola, B. Szymanski, Network management and control using collaborative on-line simulation, in: CNDSMS, 2001, pp. 1–8.

[76] T. Benson, A. Anand, A. Akella, M. Zhang, Understanding data center traffic characteristics, SIGCOMM Computer Communication Review 40 (1) (2010) 92–99.

[77] T. Benson, A. Akella, D. Maltz, Network traffic characteristics of data centers in the wild, in: Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, Australia, 2010, pp. 267–280.

[78] A. Gonsalves, Gartner: data center problems ahead, InformationWeek, November 2009. http://www.informationweek.com/news/hardware/data_centers/221901180 (accessed 28.03.2012).

[79] Y. Chen, S. Jain, V. Adhikari, Z. Zhang, K. Xu, A first look at inter-data center traffic characteristics via Yahoo! datasets, in: Proceedings of IEEE INFOCOM, 2011, pp. 1620–1628.

[80] M. Alizadeh, A. Greenberg, D. Maltz, J. Padhye, P. Patel, B. Prabhakar, S. Sengupta, M. Sridharan, Data center TCP (DCTCP), in: ACM SIGCOMM 2010, October 2010, pp. 63–74.

[81] H. Wu, Z. Feng, C. Guo, Y. Zhang, ICTCP: incast congestion control for tcp in data center networks, in: ACM CoNEXT 2010, November 30–December 3, pp. 1377–1385.

[82] S. Kandula, S. Sengupta, A. Greenberg, P. Patel, R. Chaiken, The nature of datacenter traffic: measurements & analysis, in: 9th ACM SIGCOMM conference on Internet Measurement Conference, 2009, pp. 202–208.

[83] J. Dean, S. Ghemawat, MapReduce: simplified data processing on large clusters, in: 6th Conference on Symposium on Opearting Systems Design & Implementation, 2004, pp. 10–10.

[84] D. Zats, T. Das, P. Mohan, R.H. Katz, DeTail: reducing the flow completion time tail in datacenter networks, in: Technical Report, UCB/EECS-2011-113, EECS Department, University of California, Berkeley, 2011.

[85] D. Li, Y. Li, J. Wu, S. Su, J. Yu, Efficient and scalable data center multicast routing, IEEE/ACM Transactions on Networking 20 (3) (2012) 944–955.

[86] M. Al-Fares, S. Radhakrishnan, B. Raghavan, N. Huang, A. Vahdat, Hedera: dynamic flow scheduling for data center networks, 7th ACM/USENIX—NSDI, San Jose, CA, April 2010, pp. 19–19.

[87] A.R. Curtis, K. Wonho, P. Yalagandula, Mahout: low-overhead datacenter traffic management using end-host-based elephant detection, in: IEEE, INFOCOM'11, 2011, pp. 1629–1637.

[88] S. Mahapatra, X. Yuan, Load balancing mechanisms in data center networks, in: 7th CEWIT, September 2010, pp. 1–6.

[89] Y. Shang, D. Li, M. Xu, Energy-aware routing in data center network, in: ACM SIGCOMM Workshop Green Networking'10, 2010, pp. 1–8.

[90] P. Costa, A. Donnelly, G. O'Shea, A. Rowstron, CamCube: a key-based data center, in: Technical Report MSR TR-2010-74, Microsoft Research, 2010.

[91] T. Benson, A. Anand, A. Akella, M. Zhang, The case for Fine-grained traffic engineering in data centers, in: ACM CoNEXT 2011, December 2011, pp. 1–12.

[92] J. Baliga, R.W.A. Ayre, K. Hinton, R.S. Tucker, Green cloud computing: balancing energy in processing, storage, and transport, Proceedings of the IEEE 99 (1) (2011) 149–167.

[93] N. Vasic, D. Kostic, Energy-aware traffic engineering, in: 1st International Conference on Energy-Efficient Computing and Networking, 2010.

[94] J. Moore, J. Chase, P. Ranganathan, R. Sharma, Making scheduling 'cool': temperature-aware workload placement in data centers, in: USENIX Annual Technical Conference, April 2005, pp. 61–75.

[95] Q. Tang, S.K. Gupta, G. Varsamopoulos, Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: a cyber-physical approach, IEEE Transactions on Parallel and Distributed Systems 19 (11) (2008) 1458–1472. http://dx.doi.org/10.1109/TPDS.2008.111.

[96] L. Liu, H. Wang, X. Liu, X. Jin, W. He, Q. Wang, Y. Chen, GreenCloud: a new architecture for green data center, in: 6th International Conference Industry Session on Autonomic Computing and Communications Industry, 2009, pp. 29–38.

[97] White paper: cisco energy management: a case study in implementing energy as a service. www.cisco.com/en/US/prod/collateral/.../CiscoEMSWhitePaper.pdf (accessed on 27.10.2012).

[98] HP's ESM. http://www.networkworld.com/news/2011/031011-hp-expands-energy-and-carbon.html (accessed on 27.10.2012).

[99] A. Shieh, S. Kandula, A. Greenberg, C. Kim, B. Saha, Sharing the data center network, in: 8th USENIX conference on Networked systems design and implementation, NSDI'11, April 2012, pp. 23–23.

[100] P. Sun, M. Yu, M.J. Freedman, J. Rexford, Identifying performance bottlenecks in CDNs through TCP-level monitoring, in: ACM SIGCOMM Workshop on Measurements Up the Stack, August 2011, pp. 49–54.

[101] M. Yu, A. Greenberg, D. Maltz, J. Rexford, L. Yuan, S. Kandula, C. Kim, Profiling network performance for multi-tier data center applications, in: 8th USENIX conference on Networked systems design and implementation, NSDI'11, April 2012, pp. 5–5.

[102] A.R. Curtis, J.C. Mogul, J. Tourrilhes, P. Yalagandula, P. Sharma, S. Banerjee, DevoFlow: scaling flow management for high performance networks, SIGCOMM Computer Communication Review 41 (4) (2011) 254–265.

[103] C. Raiciu, C. Pluntek, S. Barre, A. Greenhalgh, D. Wischik, M. Handley, Data center networking with multipath TCP, in: Ninth ACM SIGCOMM Workshop on Hot Topics in Networks, Hotnets'10, Article 10, October 2010, pp. 1–6.

[104] K. Narayanan, S.K. Bose, S. Rao, Towards 'Integrated' monitoring and management of data-centers using complex event processing techniques, in: Fourth Annual ACM Bangalore Conference, 2011, pp. 133–147.

[105] M. Kutare, G. Eisenhauer, C. Wang, K. Schwan, V. Talwar, M. Wolf, Monalytics: online monitoring and analytics for managing large scale data centers, in: 7th International Conference on Autonomic Computing ACM, New York, USA, 2010, pp. 141–150.

[106] P.H. Dos, R. Gomes, R.A. Kaiser, D.A. Vieira, HOLMES: an event-driven solution to monitor data centers through continuous queries and machine learning, in: Fourth ACM International Conference on Distributed Event-Based Systems, 2010, pp. 216–221.

[107] L. Brenna, A. Demers, J. Gehrke, M. Hong, J. Ossher, B. Panda, M. Riedewald, M. Thatte, W. White, Cayuga: a high-performance event processing engine, in: ACM SIGMOD international conference on Management of data, SIGMOD'07, 2007, pp. 1100–1102.

[108] J. Lenchner, C. Isci, J.O. Kephart, C. Mansley, J. Connell, S. McIntosh, Towards data center self-diagnosis using a mobile robot, in: 8th ACM International Conference on Autonomic Computing ACM, New York, USA, 2011, pp. 81–90.

[109] IBM, Maximo Asset Management. http://www-01.ibm.com/software/tivoli/products/maximo-asset-mgmt/ (accessed 16.05.2012).

[110] C.M. Liang, J. Liu, L. Luo, A. Terzis, F. Zhao, RACNet: a high-fidelity data center sensing network, in: 7th ACM Conference on Embedded Networked Sensor Systems ACM, New York, USA, 2009, pp. 15–28.

[111] H. Goudarzi, M. Pedram, Energy-efficient virtual machine replication and placement in a cloud computing system, in: Proceedings of IEEE 5th International Conference on Cloud Computing, 2012, pp. 750–757.

[112] CDW Reference Guide, Data Center Optimization Bringing efficiency and improved services to the infrastructure, February 2012. http://www.biztechmagazine.com/sites/default/files/rg_c_datacenter_012512_3.pdf (accessed 7.04.2012).

[113] Hewlett Packard Web Site, Introduction to integrity virtual machines white paper, 2007. http://docs.hp.com/en/9987/IntroVM2.1.pdf (accessed 7.04.2012).

[114] Wmware, Server Consolidation and Containment with Virtual Infrastructure, Solutions Brief, VMware Web Site, 2007. http://www.vmware.com/pdf/server_consolidation.pdf (accessed 1.04.2012).

[115] S. Cabuk, C. Dalton, H. Ramasamy, M. Schunter, Trusted Systems Laboratory HP Laboratories Bristol Towards Automated Provisioning of Secure Virtualized Networks, HP Labs Technical Report HPL-2007-139, 2007.

[116] X. Meng, T. Watson, Improving the scalability of data center networks with traffic-aware virtual machine placement, in: INFOCOM 2010 IEEE, Mar. 2010, pp. 1–9.

[117] J. Jiang, T. Lan, S. Ha, M. Chen, M. Chiang, Joint VM placement and routing for data center traffic engineering, in: IEEE INFOCOM Mini Conference, March 2012, pp. 2876–2880.

[118] K. Tsakalozos, M. Roussopoulos, A. Delis, VM placement in non-homogeneous IaaS-Clouds, in: 9th International Conference on Service Oriented Computing, ICSOC 2011, December 2011, pp. 172–187.

[119] M. Mihailescu, A. Rodriguez, C. Amza, D. Palcikovs, G. Iszlai, A. Trossman, J. Ng, Enhancing application robustness in cloud data centers, in: 2011 Conference of the Center for Advanced Studies on Collaborative Research, CASCON'11, November 2011, pp. 133–137.

[120] P. Balaji, A. Singh, L. Liu, B. Jain, Purlieus: locality-aware resource allocation for MapReduce in a cloud, in: 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, 2011, pp. 1–11.

[121] R. McGeer, P. Mahadevan, S. Banerjee, On the complexity of power minimization schemes in data center networks, in: 2010 IEEE Global Telecommunications Conference, GLOBECOM 2010, December 2010, pp. 1–5.

[122] Y. Zhang, A. Su, G. Jiang, Evaluating the impact of datacenter network architectures on application performance in virtualized environments, in: 18th IEEE International Workshop on Quality of Service, IWQoS, June 2010, pp. 1–5.

[123] D. Li, C. Guo, H. Wu, K. Tan, Y. Zhang, S. Lu, FiConn: using backup port for server interconnection in data centers, in: IEEE INFOCOM, April 2009, pp. 2276–2285.

[124] S. Murugesan, Harnessing green IT: principles and practices, IEEE IT Professional 48 (1) (2008) 24–33.

[125] B. Aksanli, J. Venkatesh, T. Rosing, Using datacenter simulation to evaluate green energy integration, Computer 45 (9) (2012) 56–64.

[126] K.V. Vishwanath, D. Gupta, A. Vahdat, K. Yocum, Modelnet: towards a datacenter emulation environment, in: 9th IEEE International Conference on Peer-to-Peer Computing, September 2009, pp. 81–82.

[127] J.D. Ellithorpe, Z. Tan, R.H. Katz, Internet-in-a-box: emulating datacenter network architectures using FPGAs, in: 46th Annual Design Automation Conference, DAC'09, 2009, pp. 880–883.

[128] G. Altekar, I. Stoica, DCR: replay debugging for the datacenter, Technical Report in UCB/EECS, 74, UC Berkeley, 2010.

[129] A. Ganapathi, Towards web server workload characterization and generation, Technical Report, UC Berkeley, 2006.

[130] P. Barford, M. Crovella, Generating representative web workloads for network and server performance evaluation, in: Joint International Conference on Measurement and Modeling of Computer Systems—Performance Evaluation Review, SIGMETRICS'98/PERFORMANCE'98, 1998, pp. 151–160.

[131] SPECweb2005, Standard Performance Evaluation Corporation. http://www.spec.org/web2005/ (accessed 12.04.2012).

[132] TPC, Transaction Processing Performance Council. www.tpc.org/tpcw/default.asp (accessed 12.04.2012).

[133] SLAMD, Distributed Load Generation Engine. www.slamd.com (accessed 12.04.2012).

[134] J. Sommers, P. Barford, Self-configuring network traffic generation, in: 4th ACM SIGCOMM Internet Measurement Conference, 2004, pp. 68–81.

[135] Optixia. http://www.ixiacom.com/products/display?skey=ch_optixia_x16 (accessed 12.04.2012).

[136] Hammer, End-to-end pre-deployment testing. www.empirix.com/ (accessed 12.04.2012).

[137] S.H. Lim, B. Sharma, G. Nam, E.K. Kim, C.R. Das, MDCSim: a multi-tier data center simulation, platform, in: IEEE International Conference on Cluster Computing and Workshops, 2009, CLUSTER'09, IEEE, 2009, pp. 1–9.

[138] D. Meisner, T. Wenisch, Stochastic Queuing Simulation for Data Center Workloads, in: Proc. Workshop Exascale Evaluation and Research Techniques, EXERT 10, ACM, 2010.

[139] M. Kas, Towards on-chip datacenters: a perspective on general trends and on-chip particulars, The Journal of Supercomputing (2011) 1–13.

[140] D. Ersoz, M.S. Yousif, C.R. Das, Characterizing network traffic in a cluster-based, multi-tier data center, in: 27th Intl. Conference on Distributed Computing Systems, ICDCS'07, 2007, pp. 59.

[141] R.N. Calheiros, R. Ranjan, A. Beloglazov, C.A.D. Rose, R. Buyya, CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms, Software: Practice and Experience 41 (1) (2011) 23–50.

[142] F. Julia, J. Roldan, R. Nou, J.O. Fitó, A. Vaqué, Í Goiri, J.L. Berral, EEFSim: energy efficiency simulator, in: Research Report No. UPC-DAC-RR-CAP-2010–15, Dept. of Computer Architecture, BarcelonaTech, 2010.

[143] S.K.S. Gupta, R.R. Gilbert, A. Banerjee, Z. Abbasi, T. Mukherjee, G. Varsamopoulos, GDCSim: a tool for analyzing Green Data Center design and resource management techniques, in: Green Computing Conference and Workshops, IGCC, 2011 International, 25–28 July 2011, pp. 1–8.

[144] B. Aksanli, J. Venkatesh, L. Zhang, T. Rosing, Utilizing green energy prediction to schedule mixed batch and service jobs in data centers, ACM SIGOPS Operating Systems Review 45 (3) (2012) 53–57.

Kashif Bilal received his B.S. and M.S. degree in Computer Science from COMSATS Institute of Information Technology, Pakistan. He received a campus medal in his B.S. He served as a lecturer at COMSATS Institute of Information Technology from 2004 to 2011. Currently, he is pursuing his Ph.D. in Electrical and Computer Engineering at North Dakota State University, Fargo, ND, USA.

Kashif's research domain encompasses topics related to data center networks, distributed computing, wireless networks, and expert systems. He has published 15 research papers in peer reviewed conferences and journals. He is also a co-author of two book chapters.

Saif Ur Rehman Malik is currently a Ph.D. student at North Dakota State University, Fargo, ND, USA. He has completed his M.S. degree from COMSATS Institute of Information Technology, Islamabad, Pakistan, in 2009. He is working in the field of Data Intensive Cloud Computational Systems and his research interest includes Data Management, Computational Systems, Formal Verification of Routing Protocols in Data Centers, Data Warehousing and Mining. He has 15 international journal and conference publications.

Osman Khalid did Masters in Computer Engineering from Center for Advanced Studies in Engineering (CASE), Pakistan. Currently, he is pursuing Ph.D. at the North Dakota State University, Fargo, USA. Osman's area of research include: (a) opportunistic communications, (b) Delay Tolerant Networks, (c) Trust and Security in Wireless Sensor Networks.

Abdul Hameed is a graduate student and pursuing his Ph.D. degree in Electrical and Computer Engineering department at NDSU. He received bachelors in Computer Science from COMSATS in 2006 and masters in Computer Software Engineering from NUST in 2008 respectively. He is a member of the IEEE Computer Society and Association of Computer Machinery. His broad research interests include design and analysis of algorithms, video analysis and processing, computer organization and microprocessor architecture.

Vidura Wijaysekara is a graduate student in the Electrical and Computer Engineering department, North Dakota State University. He received his bachelor's degree in Electrical Engineering from the same institute in 2011. His area of interest is formal verification. His current research is developing highly automatable algorithms for verifying latency insensitive systems.

Rizwana Irfan received a Masters of Science in Computer Sciences (MS) degree from COMSATS Institute of Information Technology (CIIT), Pakistan. She did Masters in Computer Science (MCS) from University of Peshawar, Pakistan. She is pursuing Ph.D. at the North Dakota State University, Fargo, USA. Rizwana's research domain encompasses topics related to Context-awareness in Social Networking, Trust and Security in Social Networking, Intelligent Knowledge Management systems, and Expert Systems. She has published 5 research papers in peer reviewed conferences and journals. She has supervised more than 8 undergraduate students. Rizwana's future targets include development of Context-aware and secure applications for social computing.

Sarjan Shrestha is a graduate student in the Electrical and Computer Engineering department at North Dakota State University, Fargo, ND. He is working in Asynchronous Circuit Design. He completed his B.E. in Electronics and Communication engineering from Institute of Engineering, Lalitpur, Nepal, in 2008.

Debjyoti Dwivedy is a graduate student in the Electrical and Computer Engineering department, North Dakota State University. He received his bachelor's degree in Electronics & Communication Engineering, from Haldia Institute of Technology, India in 2008.

Mazhar Ali received his B.S. degree in computer science from COMSATS Institute of Information Technology Abbottabad, Pakistan, in 2004, and M.S. degree in information security from KTH Sweden in 2009. He now is a Ph.D. student at North Dakota State University, USA. His research interests include cloud computing and information security.

Usman Shahid Khan is a Ph.D. student in the department of Electrical and Computer Engineering, North Dakota State University (NDSU), USA. Previously, he did his masters from National University of Science and Technology (NUST), Pakistan. His research interests are wireless networks, cognitive radio networks, smart grids, and data center networks.

Assad Abbas completed his Master of Science in Informatics from University of Skovde, Sweden, in 2010. Currently he is pursuing his Ph.D. at North Dakota State University, USA. He is affiliated with COMSATS Institute of Information Technology, Pakistan, since 2004. His research interests are mainly but not limited to Knowledge Engineering and representation, and Information Systems.

Nauman Jalil is a Ph.D. Student in the Department of Electrical and Computer Engineering, North Dakota State University (NDSU), USA. Previously, he did his M.S. from Center of Advance Studies in Engineering (CASE) Institute, Pakistan. His current research activities are in Cloud computing and Networks-on-Chip (NoC).

Samee U. Khan received a B.S. degree from GIK Institute of Engineering Sciences and Technology, Pakistan, and a Ph.D. degree from the University of Texas. Currently, he is Assistant Professor at the North Dakota State University. Professor Khan's research expertise encompasses topics such as sustainable computing, social networking, and reliability. In these areas, he has published over 140 papers. He is the recipient of the Sudhir Mehta Memorial International Faculty Award, Chinese Academy of Sciences Young International Scientist Fellowship, CEA Researcher of the Year Award, ACM/IEEE GreenCom Best Paper Award, Nortel Outstanding Doctoral Dissertation Award, and John Schuchman Memorial Outstanding Doctoral Student Award.