When Big Data Meets Software-Defined Networking: SDN for Big Data and Big Data for SDN

Laizhong Cui, F. Richard Yu, and Qiao Yan

Abstract

Both big data and software-defined networking (SDN) have attracted great interests from both academia and industry. These two important areas have traditionally been addressed separately in the most of previous works. However, on the one hand, the good features of SDN can greatly facilitate big data acquisition, transmission, storage, and processing. On the other hand, big data will have profound impacts on the design and operation of SDN. In this paper, we present the good features of SDN in solving several issues prevailing with big data applications, including big data processing in cloud data centers, data delivery, joint optimization, scientific big data architectures and scheduling issues. We show that SDN can manage the network efficiently for improving the performance of big data applications. In addition, we show that big data can benefit SDN as well, including traffic engineering, cross-layer design, defeating security attacks, and SDN-based intra and inter data center networks. Moreover, we discuss a number of open issues that need to be addressed to jointly consider big data and SDN in future research.

Big data has become one of the hottest topics in both academia and industry. Big data represents data sets so large and complex that traditional data management tools or processing methods are inadequate to deal with it. Big data is popularly characterized by "5Vs" (initially it was referred to as "3Vs"; two have been added recently): volume (size of data set), variety (range of data type and source), velocity (speed of data in and out), value (how useful the data is), and veracity (quality of data) [1]. The amount of data from different sources, such as the Internet of Things (IoT), social networking websites (Facebook, Twitter, Flicker, etc.), and scientific research, is increasing at an exponential rate. An International Data Corporation (IDC) report predicts that the global data volume will grow by a factor of 300, from 130 EB in 2005 to 40,000 EB in 2020, representing double growth every two years.

Big data applications (especially real-time or near-real-time applications) would not be possible without the underlying support of *networking* due to their extremely large volume and computing complexity [2]. Recently, software-defined networking (SDN) has attracted great interest as a new paradigm in networking [3]. The main idea of SDN is to detach the control plane from the forwarding plane, to break vertical integration, and to introduce the ability to program the network. SDN allows logical centralization of feedback control, and decisions are made by the "network brain" with a global network view, which eases network optimization. In SDN, data plane elements become highly efficient and programmable packet forwarding devices, while the control plane elements are represented by a single entity, the controller. Compared to traditional networks, it is much easier to develop and deploy applications in SDN. In addition, with the global view in SDN, it is straightforward to enforce the consistency of network policies. SDN represents a major paradigm shift in the evolution of networks, introducing a new pace of innovations in networking infrastructure.

While some excellent work has been done on big data and SDN, these two important areas have traditionally been addressed separately in most previous works. However, on one hand, SDN, as an important networking paradigm, will have significant impact on big data applications. In particular, several good features (e.g., separation of the control and data planes, logically centralized control, global view of the network, ability to program the network) can greatly facilitate big data acquisition, transmission, storage, and processing. For example, big data is usually processed in cloud data centers. Compared to traditional data centers, SDN-based data centers can have better performance by dynamically allocating resources in data centers to different big data applications to meet the service level agreements (SLAs) of these big data applications [4, 5].

On the other hand, big data, as an important network application, will have profound impact on the design and operation of SDN. Specifically, with the global view of the network, the logically centralized controller in SDN can obtain big data from all the different layers (i.e., from physical to application layers) with arbitrary granularity. From past experience

Laizhong Cui and Qiao Yan (corresponding author) are with Shenzhen University.

F. Richard Yu is with Carleton University.

This work is jointly supported by the National Natural Science Foundation of China (Grant Nos. 61402294, 61170283, and 61572330), the Major Fundamental Research Project in the Science and Technology Plan of Shenzhen (Grant No. JCYJ20140828163633977), and the Natural Sciences and Engineering Research Council of Canada.

0890-8044/16/$25.00 © 2016 IEEE

in cross-layer design, we have learned that although sharing information among different layers can improve network performance, the network becomes so complex that traditional approaches are inadequate to design and optimize such networks. Fortunately, big data analytics, which leverages analytical methods to obtain insights from data to guide decisions, can help the design and operation of SDN. For example, with big traffic data analytics, it is easier for the controller to perform traffic engineering to improve the performance of SDN.

In this article, we first discuss the basic characteristics and the new trends of big data and SDN. Then we present the good features of SDN in solving several prevailing issues with big data applications, including big data processing in cloud data centers, data delivery, joint optimization, scientific big data architectures, and scheduling issues. We show that SDN can manage the network efficiently to improve the performance of big data applications. In addition, we show that big data can benefit SDN as well, including traffic engineering, cross-layer design, defeating security attacks, and SDN-based intra- and inter-data-center networks. Moreover, we discuss a number of open issues that need to be addressed to jointly consider big data and SDN in future research.

To the best of our knowledge, the interrelationship between big data and SDN has not been well addressed in previous work. In essence, the unique characteristics associated with big data and SDN present unique challenges beyond the existing works. We believe that the initial steps we take here help in understanding how to make full use of SDN's advantages to improve the performance of big data applications, and how to make SDN work better and more effectively through big data.

The rest of the article is organized as follows. We present an overview of big data and SDN. We show that SDN can benefit big data application. We show that big data can benefit SDN. Some open research issues are presented. Finally, we conclude this study.

Overview of Big Data and Software-Defined Networking

In this section, we discuss the basic concepts and features of big data and SDN, as well as the brief interrelationship between big data and SDN.

Big Data

Big data has become a popular research topic in recent years. The history of big data is tied closely to the capability of storing and processing data sets. With the capability improvement over the years, new technologies have been developed to store and process larger and larger datasets, from megabyte to gigabyte to terabyte to petabyte to exabyte. There are different types of definitions for big data in the literature [6]. In the attributive definition, big data is defined according to five salient features: volume, variety, velocity, value, and veracity. In the comparative definition, big data can be defined as "data sets the size of which is beyond the ability of typical database tools to capture, store, manage, and analyze."

The sources of big data are coupled with its generating domains. Some examples of the generating domains include business (e.g., business-to-business and business-to-consumer transactions), networking (e.g., the Internet, cellular networks, and the Internet of Things [IoT]), and scientific research (e.g., computational biology, astronomy, and high-energy physics). Big data can be very useful in a variety of applications, such as enterprise applications, IoT-based applications, healthcare applications, and scientific research applications. Almost all the dominating companies, including Google, Amazon, Facebook, Microsoft, and Oracle, have begun to develop big data projects. A number of U.S. federal agencies, including the Defense Advanced Research Projects Agency (DARPA), the National Science Foundation, and the National Institutes of Health, have invested heavily in big data research.

Many research challenges remain to be addressed for big data, including analytics platforms, underlying support of networking, data processing models, data staging, distributed storage, and data privacy and security. In this article, we focus on the challenges related to the underlying support of networking for big data applications.

Software-Defined Networking

With the increasing amount of data of various applications over the Internet, SDN has been considered as one of the most promising solutions to revolutionize the cyber world with proper network management [3]. OpenFlow is regarded as the first standardized protocol in SDN, and is also known as the enabler of SDN. The term SDN was originally coined to represent the ideas and work around OpenFlow at Stanford University. According to the original definition, SDN refers to a network architecture where the forwarding state in the data plane is managed by a remote control plane decoupled from the former. More and more SDN protocol standards have been developed in real applications. Nowadays, SDN is not limited to any one implementation, but is a general term for the platform.

The Open Networking Foundation (ONF), which is a nonprofit consortium dedicated to development and standardization of SDN, gives the definition of SDN as follows: "In the SDN architecture, the control and data planes are decoupled, network intelligence and state are logically centralized, and the underlying network infrastructure is abstracted from the applications." The following four pillars can be identified for SDN:

1. The control and data planes are decoupled.
2. Forwarding decisions are flow-based instead of destination-based.
3. Control logic is centralized, moved to the so-called SDN controller or network operating system (NOS).
4. The network is programmable through software applications running on top of the NOS, which interacts with the underlying data plane devices.

These good features can greatly facilitate big data acquisition, transmission, storage, and processing.

In general, SDN can be divided into three layers in its reference model: the infrastructure layer, control layer, and application layer. In the SDN architecture, various applications involving SDN can be observed in the application layer. With the help of the northbound application programming interfaces, the application layer communicates with the SDN control plane regarding the status of the network and its particular requirements. The SDN control plane in the control layer is regarded as the primary component of the SDN model, which controls the overall functioning of the network. Starting with NOX (the first designed SDN controller), various types of SDN controllers have been designed, such as Onix, POX, Beacon, Ryu, Trema, OpenDaylight, Contrail, and Floodlight.

Recently, SDN has been used to design data center network architectures to provide better performance with low complexity and energy consumption [4]. Through the collection of traffic data, SDN can dynamically allocate resources to improve application performance. Moreover, there is a new trend to leverage the capabilities of SDN for application-aware networking in data center applications and network architectures. As big data is usually processed in data centers, SDN-based data centers can benefit big data applications by dynamically

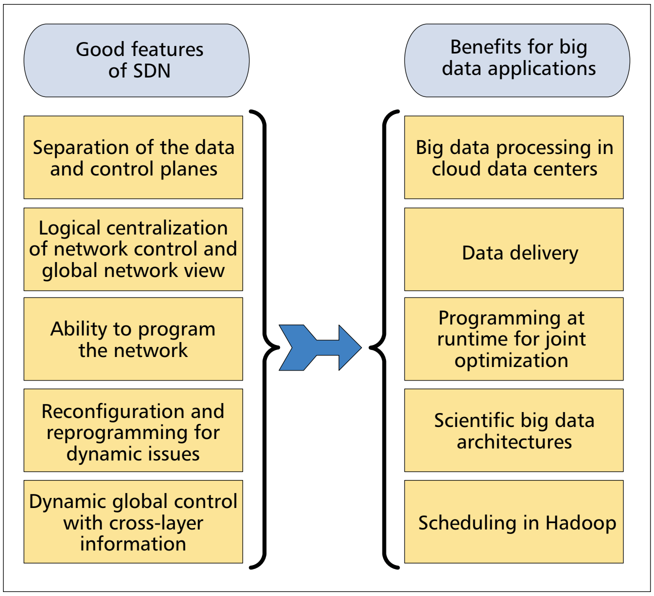

Figure 1. Good features of SDN that can benefit big data applications.

allocating resources according to the SLAs of different big data applications.

Software-Defined Networking Can Benefit Big Data Applications

In this section, we show that the good features of SDN can benefit big data applications in various aspects, including big data processing in cloud data centers, data delivery, programming at runtime for optimizing big data applications, scientific big data architectures, and scheduling in Hadoop, as shown in Fig. 1.

SDN Can Benefit Big Data Processing in Cloud Data Centers

Big data is usually processed in cloud data centers. Because the resource requirements of big data applications change dynamically in cloud data centers, it is important to assign and manage resources of cloud data centers efficiently to meet the SLAs of different big data applications. The SLA of a big data application is the agreement negotiated between a big data service provider and its users. It defines the characteristics of the provided big data service, including service level objectives, expected quality of service (QoS), and penalties if these objectives are not met by the big data service provider. Some common SLAs for big data applications performance are response time, processing time, rate of failure, trust, security, negotiation time, maintenance time, cost estimation, amount of data processed, amount of data received, amount of data emitted, data quality of input and output, data persistence time, data movement latency, and data losses.

The technology trend of cloud computing is putting pressure on data center providers to create more flexible services with better performance and security. With SDN-enabled data centers, the networking and storage infrastructure is delivered as a service and automated by SDN. From the perspective of infrastructure, the cloud platform is built on the basis of the data centers, which can be called cloud data centers. Cloud service providers can host this infrastructure for big data applications.

In [4], an SDN-based cloud data center is studied for big data application. Specifically, an SDN-based OpenFlow network with combined input and crosspoint queued (CICQ) switches is proposed to schedule packets for different big data applications. In this approach, the controller maintains a bandwidth provisioning table for different types of big data applications and sends it to CICQ switches. Then the switches decide packet scheduling priorities based on the bandwidth provisioning table from the controller. The resource is allocated efficiently, and the power consumption is also reduced for different big data applications in cloud data centers. In [5], the authors present SDN-based data centers with optimal topology composition and traffic load balancing. The optimal topology composition finds a subset data center network topology that can accommodate the expected traffic demands from different big data applications. To handle network congestion, the traffic load balancing distributes ever changing traffic demands over the found optimal subset topology.

SDN Can Benefit Data Delivery for Big Data Applications

Due to the large volume, data delivery in big data applications is a great challenge. The authors of [7] propose an SDN-based optical network to accelerate and support "*-cast" traffic delivery for big data applications. Big data applications have led to massive volumes of traffic featuring diverse communication patterns denoted as "*-cast" combining unicast, multicast, incast, and all-to-all cast. In [7], a hybrid (optical and electrical) approach is presented that leverages physical layer optics to accelerate traffic delivery for each pattern. The application-driven control plane is compatible with SDN, enabling the flexible and dynamic runtime configuration of photonic devices to support complex traffic patterns. To handle the bursting data in big data applications, the authors of [8] present an SDN-enabled topical transport architecture that meshes seamlessly with the deployment of SDN. In this programmable architecture, a core transport node is abstracted into a programmable virtual switch that leverages the OpenFlow protocol for control. With a prototype demonstration of big data applications, it is shown that the programmability and flexibility that SDN brings can greatly benefit data delivery for big data applications.

Moreover, ultra-high-definition video streaming and video conference are examples of big data applications that require large bandwidth for efficient end-to-end delivery. These aggregated bandwidth demands for delivery over core networks can be met with high-capacity wavelength-division multiplexing (WDM) circuit switched optical networks. The authors of [9] propose a system that integrates SDN and generalized multiprotocol label switching (GMPLS) control planes, and demonstrate that such an architecture can support various big data applications. In this approach, OpenFlow can be referred to as the "network brain," and GMPLS is used to support multiple types of switching like time-division multiplexing (TDM), Lambda, wave-band, and fiber switching.

Programming at Runtime for Optimizing Big Data Applications

Due to ever changing environments, many big data applications require frequent reconfigurations. SDN's capability of programming at runtime is very useful for big data applications that require frequent reconfigurations. To study the integration between big data applications and network control, the authors of [10] introduce a cross-layer structure of an SDN-based network for big data applications. The programming of SDN is explored across the layers of a network, from physical topology to routing and flow level traffic engineering. The authors combine the SDN controller and optical switching to realize close cooperation of network control and big data applications. The joint optimization of application performance and network utilization is also studied. Specifically, the authors consider a system that consists of a collection of optical circuit switches with OpenFlow enabled on top-of-rack (TOR) Ethernet switches that have multiple optical uplinks. These switches are controlled by an SDN controller, which manages the physical connection between the switches by configuring them. In the integrated network control plane, the topology construction and routing mechanisms for a number of communication patterns are proposed to improve application performance. All big data applications in the application layer are treated as a single master component. The job scheduling strategies for big data applications to accommodate dynamic network configuration must also be well designed. With a relatively small overhead of configurations, this integrated network control architecture to program the network at runtime through an SDN controller can give better performance for big data applications.

SDN Can Benefit Scientific Big Data Architectures

Scientific research is highly data driven. Some scientific research instruments produce enormous amounts of data. Moreover, due to political and technical reasons, some scientific big data cannot be processed and analyzed at the same location where they are generated. The big data has to be sent to various data centers and researchers at different universities or institutes for analysis. However, current campus networks are not capable of handling such a huge amount of data. As the amount of data increases, there is a greater need for simple, scalable end-to-end network architectures and implementations that enable applications to use the network most efficiently [11].

To bypass traditional performance hotspots in typical campus networks for scientific big data applications, SDN has some advantageous to solve this problem. Monga *et al.* introduce SDN to scientific big data architectural models [11]. They propose a new architecture model for campus networks that use SDN for multi-science disciplines within the same campus network. The model also creates a virtual network over wide area networks (WANs) for scientific research collaborations. In addition, an SDN/OpenFlow-based end-site architecture is introduced to support multiple science disciplines. This model proposes a data transfer node (DTN) redirector. Each DTN can send data transfer requests to the DTN redirector. When the DTN redirector receives a data transfer request, data is transferred to the requesting DTN using OpenFlow rules. It can encapsulate the data flow in a pre-approved virtual local area network, and hence bypass the firewall function. The SDN/OpenFlow controller manages all the policies regarding the campus network. The effectiveness of this new architecture model is demonstrated for scientific big data applications.

SDN Can Benefit Scheduling in Hadoop for Big Data Applications

As a software framework that supports big data storage and processing, Hadoop has attracted substantial attention from both industry and scholars. Hadoop integrates data storage, data processing, system management, and other modules to form a powerful system-level solution, which is becoming the mainstay in handling big data challenges. Despite its popularity, an important issue to be solved in Hadoop is the NP-complete minimum make span problem [12], which has significant impact on the performance of Hadoop systems. This problem is to find solutions on how to minimize the job completion time in Hadoop systems. Most existing solutions to this problem are not optimal due to the lack of a global view for allocating tasks [12].

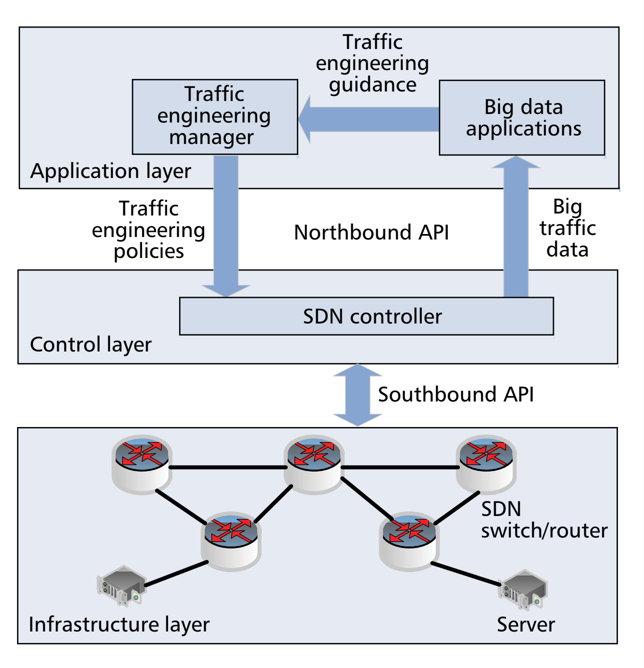

Figure 2. Dynamic traffic engineering system architecture with SDN and big data.

SDN is adopted in [12] to solve this problem in Hadoop. Specifically, an SDN-based and bandwidth-aware scheduler is proposed, which can flexibly assign tasks in an optimal manner and guarantee data locality from a global view. It first utilizes SDN to manage the network bandwidth and allocates it in a time slot manner; then it decides whether to assign a task locally or remotely depending on the completion time. Therefore, this approach can guarantee data locality from a global view; meanwhile, it can efficiently assign tasks. The key point of this approach is that the scarce network bandwidth from an SDN/OpenFlow controller is not only taken into account but also regarded as a vital parameter for task scheduling. Real-world experiments show that it can improve the performance of scheduling in Hadoop for big data applications.

Big Data Can Benefit Software-Defined Networking

In this section, we show that big data can benefit SDN, including traffic engineering, cross-layer design, defeating security attacks, and SDN-based intra- and inter-data-center networks.

Big Data Can Benefit Traffic Engineering in SDN

Traffic engineering is an important method of optimizing the performance of a network by dynamically analyzing, predicting, and regulating the behavior of data transmitted over that network. Typical objectives of traffic engineering include balancing network load and maximizing network utilization.

SDN and big data analytics provide a convenient and effective way to perform traffic engineering and improve network performance on a large scale. Typically, an SDN-based network consists of more than thousands of hosts with significant bandwidth requirements. Traffic engineering in such networks is very challenging. The combination of big data and SDN for traffic engineering would be an apt solution due to the following reasons:

1. It is relatively easier to obtain big data traffic and failure information via a logically centralized SDN controller.

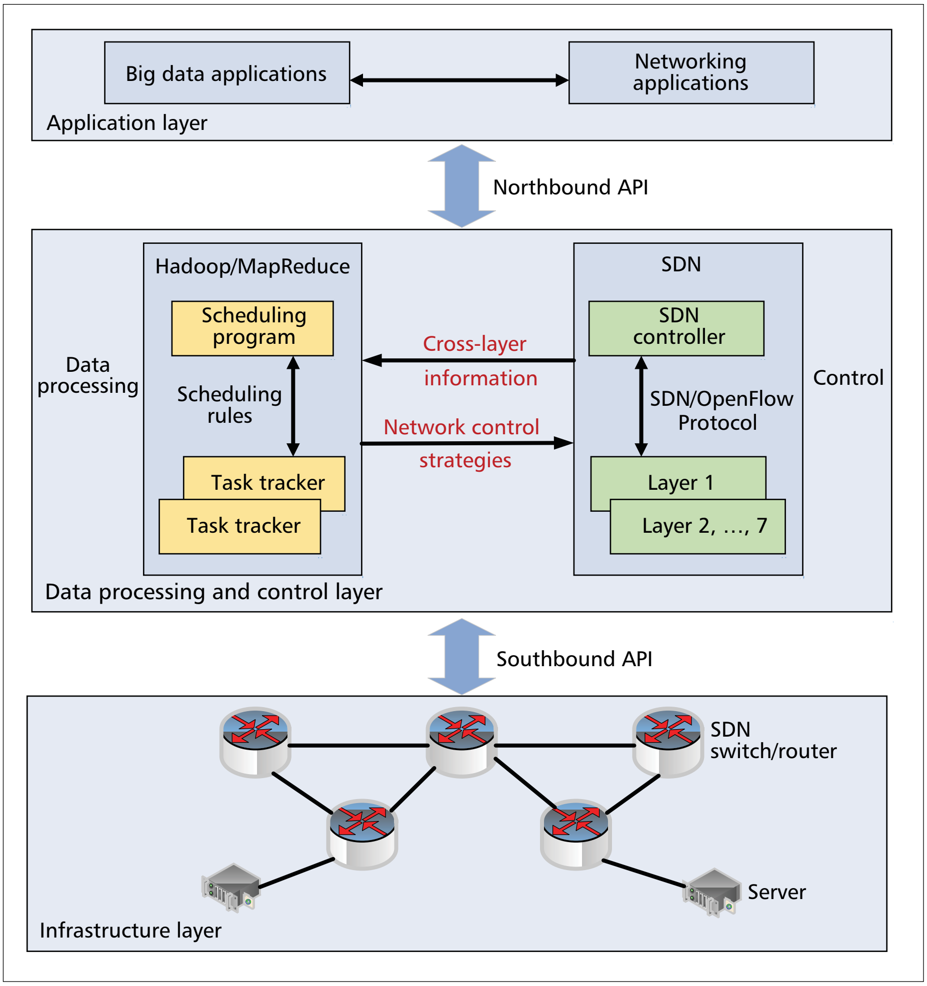

Figure 3. Cross-layer design with SDN and big data.

2. Any flow format of big traffic data with arbitrary granularity can be exploited for traffic engineering.
3. It is relatively easier to apply traffic engineering results to switches in a data center network by modifying flow tables within the switches.

In Fig. 2, we describe a dynamic traffic engineering system architecture with SDN and big data, which consists of four components: a data center network, an SDN controller, a traffic engineering manager, and big data applications. In the data center network, there are many servers and SDN switches/routers, which is considered to be a target network of the traffic engineering system. The SDN switches/routers in the data center network report their big traffic data and failure status to the SDN controller through the control/data plane interface. The SDN controller aggregates and summarizes the collected big traffic data information, and sends it to the big data applications. Big data analytics, which leverages analytical methods to obtain insights from the big traffic data, then gives guidance to the traffic engineering manager, which derives the traffic engineering policies. According to these traffic engineering policies, the SDN controller changes switching behavior of the SDN devices by updating their flow tables, and turns on/off devices and links in the data center network to minimize power consumption and link congestion.

Big Data Can Benefit Cross-Layer Design in SDN

Big data can benefit not only traffic engineering, but also other layers in SDN. Traditionally, networking is divided into different layers, and a set of protocols are used for communications between adjacent layers. In traditional layered design, it is forbidden to have direct communications between non-adjacent layers. However, recent advances in cross-layer designs show that non-adjacent layers can share informations during run-time, which will result in new algorithms and significantly improved performance in networking systems.

Although sharing information among different layers can improve the performance, the principle of modularity is broken, and the network will become so complex that traditional approaches are inadequate to design and optimize such networks. Fortunately, big data can benefit cross-layer design in SDN. The logically centralized controller in SDN has a global view of the network, which enables it to obtain big data from all the different layers with arbitrary granularity, such as channel state information at the physical layer, packet information at the data link/network layers, and application information at application layer. Applying the big data technologies to network control and management can significantly improve the network control and management processes. %Therefore, cross-layer design in SDN will be challenging.

Here we present an architecture of combining big data and SDN, which can facilitate the cross-layer design in SDN with the help of big data. There are three layers in the architecture: the infrastructure layer, data processing and control layer, and application layer, as shown in Fig. 3. The infrastructure layer consists of switches/routers, servers, and data center devices. The switching or routing devices transfer data packets to the next hop in accordance with forwarding rules stored in local memory. The servers in the data center store the big data and run the tasks. In the data processing and control layer, the SDN controller and Hadoop will have close cooperation with processing the big data

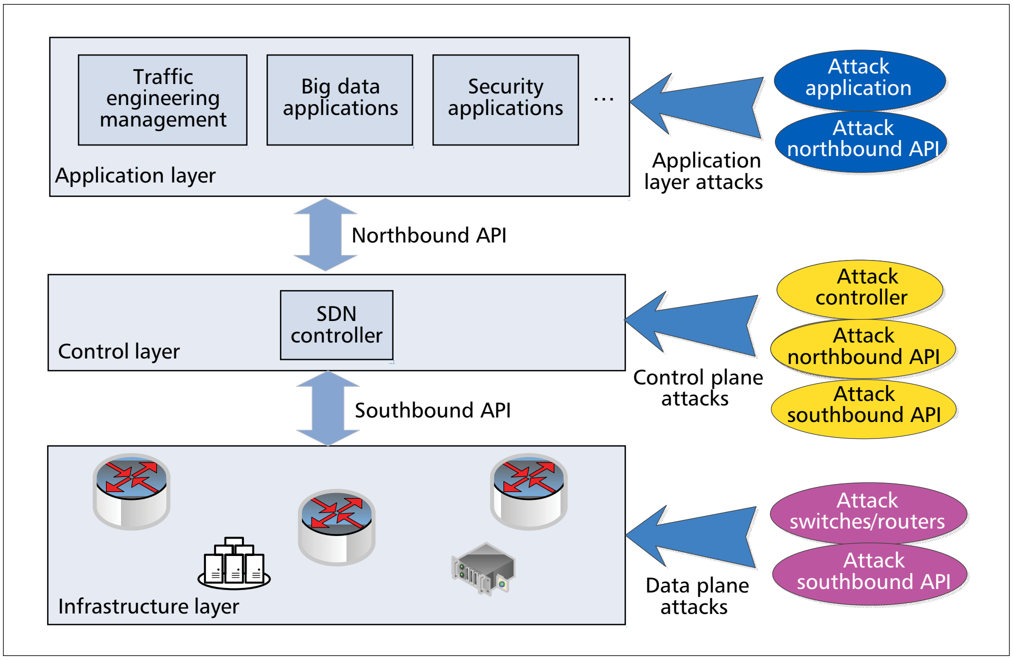

Figure 4. Potential attacks can be launched on the three layers of SDN.

and making the decisions together. The SDN controller provides Hadoop with the cross-layer information from all the different layers, while Hadoop provides an SDN controller with network control strategies (physical layer parameters adaptation, resource allocation, topology construction, routing mechanisms, congestion control, etc.) to operate and optimize the performance of SDN. Both big data applications and networking applications run on the top of the data processing and control layer.

Big Data can Benefit SDN in Defeating Security Attacks

The security of SDN is a big concern. Since SDN is vertically split into three main functional layers, potential malicious attacks can be launched on these three layers of SDN's architecture. Based on the possible targets, we can classify the attacks launching on SDN into three categories: application layer attacks, control layer attacks, and infrastructure layer attacks, as shown in Fig. 4 [13]. There are two methods to launch application layer attacks. One is to attack some applications; the other is to attack a northbound application programming interface (API). The controller could potentially be seen as a single point of failure risk for the network, so it is a particularly attractive target for attacks in the SDN architecture. The following methods can launch control plane attacks: attacking a controller, a northbound API, or a southbound API. There are two methods to launch infrastructure plane attacks. One is to attack some switches/routers; the other is to attack a southbound API. Some attackers, such as distributed denial of service (DDoS) attackers, take advantage of botnets and other high-speed Internet access technologies, and the size of attacks has grown dramatically. For example, the size was as high as 300 Gb/s in the 2013 DDoS attack on Spamhaus. Therefore, traditional data analysis methods have many difficulties in defeating these attacks.

The application of big data analytics to mitigate security attack problems is becoming more and more attractive. The ability of big data analytics enables us to comprehensively analyze large volumes of disparate and complex data from different sources in different formats. We can compare these data, perform anomaly detection, and combat cyber threats in real time. Multi-dimensional to ultra-high-dimensional data models can be built to accurately profile the data streams online, which allows detecting and even predicting security attacks in real time. Big data analysis can also provide correlation methods among heterogeneous security data. Furthermore, the machine learning methods for big data analytics have potential to successfully defend against future attackers and detect anomalies.

SDN-Based Intra- and Inter-Data-Center Networks with Big Data

Data centers have gained popularity among service providers and network operators. This has been a cause of concern for data center networks because these applications require high-bandwidth low-latency low-energy-consumption networks. SDN can be a promising solution for maximal resource utilization in data centers. For the intra-data-center network, the authors of [14] propose an architecture called time-aware SDN (TASDN) in OpenFlow based optical networks. Depending on the arrival of requests to the data center, TASDN can coordinate big data applications by time factors and their response factors. The two key aspects of TASDN are:

1. It schedules the data center services with different delay requirements and optimizes the big data application stratum resources.
2. It can deliver the quick burst service provision of big data applications through centralized TASDN control and procedure.

For the inter-data-center network, Zhang *et al.* present an enhanced SDN (eSDN) architecture for big data applications [15]. In the proposed architecture, eSDN performs not only path computing and flow table management, but also big data application service management and network information management. eSDN uses a transport controller instead of a network controller (NC) to meet the burst requirement of IP services from big data applications.

Both TSDN and eSDN for intra- and inter-data-center networks are typical examples of combining big data and SDN. Based on these works, in Fig. 5, we present an architecture of intra- and inter-data-center networks with big data and SDN, which also includes three layers: the infrastructure layer, data processing and control layer, and application layer. In the infrastructure layer, there are SDN switches and servers in the intra-data-center network. The different data centers are

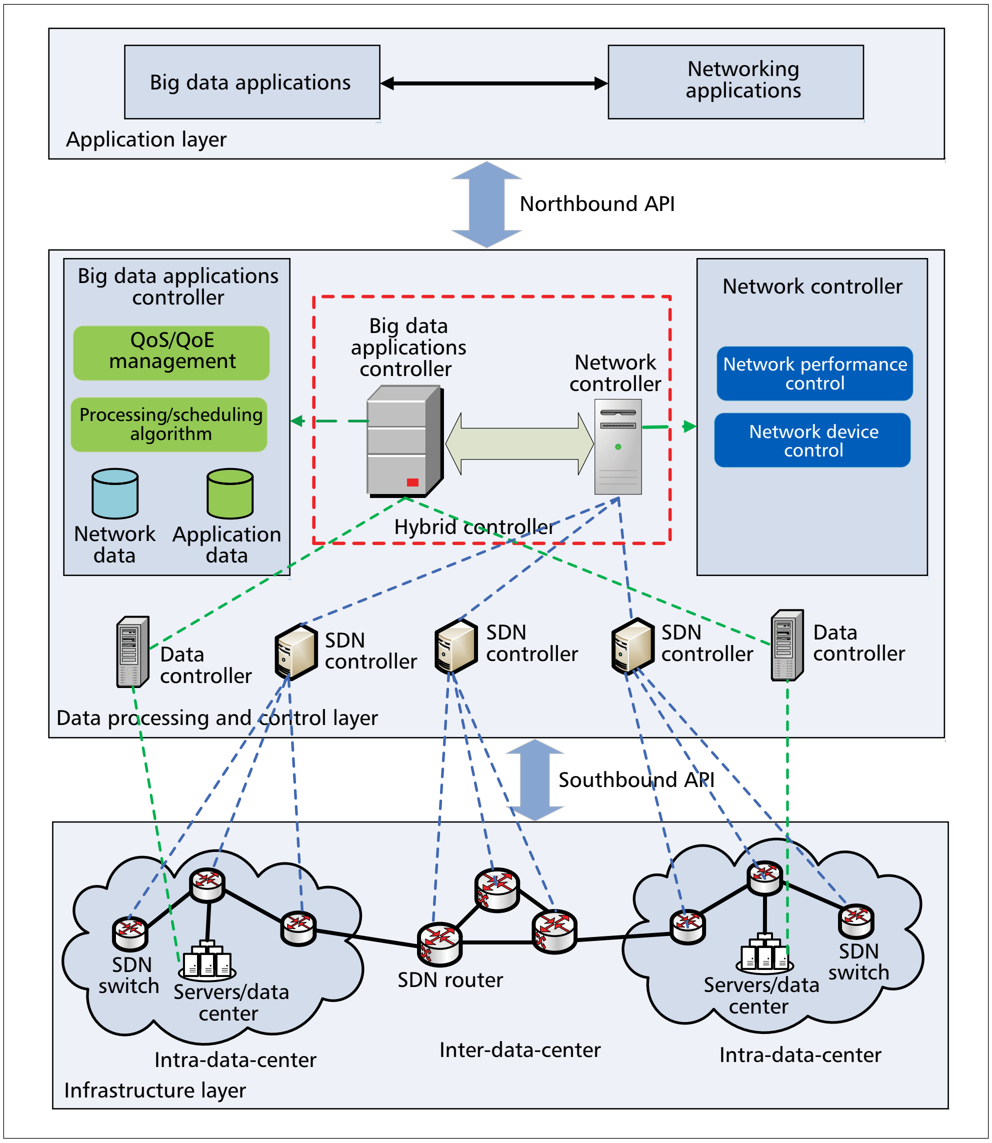

Figure 5. SDN-based intra- and inter-data-center networks with big data.

connected through SDN routers over an IP network, which form the intra-data-center network. In the control layer, there are several SDN controllers to manage the switches intra-data-center or the routers inter-data-center. The data controllers are in charge of scheduling data processing. All the SDN controllers are controlled by a network controller, while all the data controllers are controlled by a big data application controller. The network controller and data controller are regarded as a hybrid controller, which is the "brain" of the control layer. The application layer can provide various big data applications and networking applications through the programming interface.

Open Issues

There are many open issues that are still not well studied and need to be tackled by future research efforts. In this section, we discuss some of these open issues for joint design of big data and SDN.

Scalable Controller Management

When the network controller in SDN is used for big data applications, its performance could be degraded due to the rapid and frequent flow table update requests as well as big data transmission and processing. Accordingly, the network controller should be designed in such a way that it can accommodate larger flow tables and big data. Current network controller flow tables have around 8000 entries, which are insufficient for big data operations [3]. The design of a scalable network controller for big data applications needs further research. One possible solution is that a cluster of controllers can be used to solve the scalability in SDN. Another way to alleviate the scalability challenge is to employ parallel computation in multicore systems and improve I/O performance. In addition, the control plane elements can be distributed physically, but the centralized network-wide view can still be maintained.

Intelligent Flow Table/Rule Management

The main motivation of designing SDN/OpenFlow is to simplify the switching devices by separating the control plane from the forwarding plane, and transferring the intelligence to the control plane. In the joint design of big data and SDN, forwarding tables are sent to the switches in the network, and the switches utilize these forwarding rules to transfer data packets of big data applications. However, since these networking switches do not have any intelligence, they just send raw data packets to the network controller. In other words, no data preprocessing is done in the switches, which results in heavy load in the controller. Therefore, some intelligence in the forwarding plane of SDN is of great importance for big data applications.

High Flexible Language Abstraction

A common high-level programming language in SDN is still missing for big data application development. Currently, different controllers use different programming languages. For example, POX uses Python, whereas Beacon uses Java. SDN requires a common programming language that should be able to adapt to frequent flow rule changes caused by big data applications.

Wireless Mobile Big Data

SDN has been studied for not only wired networks, but also wireless mobile networks, such as vehicle networks, cellular networks (e.g., 4G and 5G networks), and sensor networks. SDN-enabled wireless mobile networks will bring some promising features, including easier deployment of new services, reduced management and operational costs of heterogeneous technologies, efficient operation of multi-vendor infrastructures, increased accountability and service differentiation, and continuous and transparent enhancement of network operations. With the rapid development of wireless mobile networks, more data are now collected from mobile devices and networks. Therefore, more research should be conducted to study the benefits of SDN for big data applications in wireless mobile networks.

Conclusion

In this article, we first discuss the basic characteristics and new trends of big data and SDN. Then we present some good features of SDN (separation of the control and data planes, logically centralized control, global view of the network, ability of programming the network, etc.) that can benefit big data applications, including big data processing in cloud data centers, data delivery, joint optimization, scientific big data architectures, and scheduling issues. In addition, we show that big data can benefit various aspects of SDN, including traffic engineering, cross-layer design, defeating security attacks, and SDN-based intra- and inter-data-center networks. Moreover, we discuss some open issues for future research, such as scalable controller management, intelligent flow table/rule management, high flexible language abstraction, and wireless mobile big data.

To sum up, the joint design of big data and SDN can become a promising solution for big data networking. How to make full use of SDN's advantages to improve the performance of big data applications and how to utilize big data to make SDN work better and more effectively are urgent problems that need to be addressed. This article attempts to briefly explore the technologies related to the joint design of big data and SDN, and discuss future research that may be beneficial on these issues.

Acknowledgment

We thank the reviewers for their detailed reviews and constructive comments, which have helped to improve the quality of this article.

References

[1] V. N. Gudivada, R. Baeza-Yates, and V. V. Raghavan, "Big Data: Promises and Problems," *Computer*, vol. 48, no. 3, Mar. 2015, pp. 20–23.

[2] S. Yu, X. Lin, and J. Misic, "Networking for Big Data," *IEEE Network*, vol. 28, no. 4, July 2014, p. 4.

[3] D. Kreutz *et al.*, "Software-Defined Networking: A Comprehensive Survey," *Proc. IEEE*, vol. 103, Jan. 2015, pp. 14–76.

[4] W. Hong, K. Wang, and Y.-H. Hsu, "Application-Aware Resource Allocation for SDN-Based Cloud Datacenters," *Proc. Int'l. Conf. Cloud Computing and Big Data 2013*, Fuzhou, China, Dec. 2013.

[5] Y. Han *et al.*, "Software Defined Networking-Based Traffic Engineering for Data Center Networks," *Proc. 16th Asia-Pacific Network Operations and Management Symp.*, Taiwan, Sept. 2014.

[6] H. Hu *et al.*, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial," *IEEE Access*, vol. 2, 2014, pp. 652–87.

[7] P. Samadi *et al.*, "Accelerating Cast Traffic Delivery in Data Centers Leveraging Physical Layer Optics and SDN," *Proc. Int'l Conf. Optical Network Design and Modeling*, Stockholm, Sweden, May 2014.

[8] A. Sadasivarao *et al.*, "Bursting Data Between Data Centers: Case for Transport SDN," *Proc. IEEE 21st Symp. High-Performance Interconnects*, Mountain View, CA, Aug. 2013.

[9] S. Azodolmolky *et al.*, "Integrated Openflow — GMPLS Control Plane: An Overlay Model for Software Defined Packet Over Optical Networks," *Proc. 37th Euro. Conf. Optical Commun.*, Geneva, Switzerland, Sept. 2011.

[10] G. Wang, T. E. Ng, and A. Shaikh, "Programming Your Network at Runtime for Big Data Applications," *Proc. 1st Wksp. Hot Topics in Software Defined Networks*, Helsinki, Finland, Aug. 2012.

[11] I. Monga, E. Pouyoul, and C. Guok, "Software-Defined Networking for Big-Data Science — Architectural Models from Campus to the WAN," *Proc. High Performance Computing, Networking, Storage and Analysis*, 2012 SC Companion, Nov. 2012, pp. 1629–35.

[12] P. Qin *et al.*, "Bandwidth-Aware Scheduling with SDN in Hadoop: A New Trend for Big Data," arXiv, vol. abs/1403.2800, 2014.

[13] Q. Yan *et al.*, "Software-Defined Networking (SDN) and Distributed Denial of Service (DDOS) Attacks in Cloud Computing Environments: A Survey, Some Research Issues, and Challenges," *IEEE Commun. Survey & Tutorials*, 2015, DOI:10.1109/COMST.2015.2487361.

[14] Y. Zhao *et al.*, "Data Center Optical Networks (DCON) with Openflow Based Software Defined Networking (SDN)," *Proc. 8th Int'l CHINACOM*, Guilin, China, Aug. 2013.

[15] J. Zhang *et al.*, "First Demonstration of Enhanced Software Defined Networking (eSDN) over Elastic Grid (eGrid) Optical Networks for Data Center Service Migration," *Proc. OFC/NFOEC*, Anaheim, CA, Mar. 2013.

Biographies

LAIZHONG CUI [M] (cuilz@szu.edu.cn) is a lecturer in the College of Computer Science and Software Engineering at Shenzhen University, China. He received his Ph.D. degree in computer science and technology from Tsinghua University, Beijing, China, in 2012. His research interests are in peer-to-peer networks, content distribution, cloud computing, and social networks. His current focus is on the research and development of software defined networking and big data.

F. RICHARD YU [SM] (richard.yu@carleton.ca) is an associate professor at Carleton University, Canada. He received the IEEE Outstanding Leadership Award in 2013, the Carleton Research Achievement Award in 2012, the Ontario Early Researcher Award (formerly Premier's Research Excellence Award) in 2011, the Excellent Contribution Award at IEEE/IFIP TrustCom 2010, the Leadership Opportunity Fund Award from Canada Foundation of Innovation in 2009, and Best Paper Awards at IEEE ICC 2014, IEEE GLOBECOM 2012, IEEE/IFIP TrustCom 2009, and the International Conference on Networking 2005. His research interests include cross-layer design, security, green IT, and QoS provisioning in wireless networks. He serves on the editorial boards of several journals, including Co-Editor-in-Chief for *Ad Hoc & Sensor Wireless Networks*, Lead Series Editor for *IEEE Transactions on Vehicular Technology*, and Editor for *IEEE Communications Surveys & Tutorials*. He has served on the Technical Program Committee (TPC) of numerous conferences, and as the TPC Co-Chair of IEEE GreenCom '15, INFOCOM-MCV '15, GLOBECOM '14, INFOCOM-MCC '14, GLOBECOM '13, GreenCom '13, CCNC '13, INFOCOM-CCSES '12, ICC-GCN '12, VTC '12, GLOBECOM '11, INFOCOM-GCN '11, INFOCOM-CWCN '10, IEEE IWCMC '09, VTC-Fall '08, and WiN-ITS '07.

YAN QIAO (yanq@szu.edu.cn) is a professor at the College of Computer Science and Software Engineering at Shenzhen University. She received her Ph.D. degree in information and communication engineering from Xidian University, Xioan, China, in 2003. From 2013 to 2014 she worked at Carleton University, Ottawa, Canada, as a visiting scholar. Her research interests are in network security, cloud computing, and software-defined networking. Her current focus is research and development of security in software defined networking.