Performance Evaluation of a Power-Efficient and Robust 60 GHz Wireless Server-to-Server Datacenter Network

Sayed Ashraf Mamun, Senior Member, IEEE, Sree Gowrishankar Umamaheswaran, Student Member, IEEE, Amlan Ganguly, Member, IEEE, Minseok Kwon, Member, IEEE, and Andres Kwasinski, Senior Member, IEEE

Abstract—Datacenters have become the digital backbone of the modern world. To support the growing demands on bandwidth, datacenter networks (DCNs) consume an increasing amount of power. Additionally, the complex cabling in traditional datacenters poses design and maintenance challenges and increases the energy cost of the cooling infrastructure by obstructing the flow of chilled air. In this paper, we present a wireless server-to-server datacenter network architecture using millimeter-wave links to eliminate the need for power-hungry switching fabric of traditional fat-tree based datacenter networks. The server-to-server wireless datacenter network (S2S-WiDCN) architecture requires line-of-sight (LoS) between servers to establish direct communication links. However, in the presence of an obstruction such as an IT technician, the LoS may be blocked. To address this issue, we propose a novel obstruction-aware adaptive routing algorithm for S2S-WiDCN. We evaluate the performance of S2S-WiDCN in terms of traffic flow completion duration and throughput, with different kinds of datacenter traffic as well as in presence of obstruction to LoS links. We also compare the performance and power efficiency of S2S-WiDCN with the traditional fat tree DCN as well as a top-of-rack (ToR)-to-ToR wireless DCN.

Index Terms—Datacenter network, millimeter wave communication, OFDM, server-to-server, wireless communication.

I. INTRODUCTION

Datacenters have become an essential part of computing resources as they provide large storage banks and processing power for cloud services. According to a 2016 report from Lawrence Berkeley National Laboratory, it is projected that the power consumption of all datacenters in the USA is going to be 73 billion KWh per year by 2020 and the power consumption of the Datacenter Network (DCN) will continuously increase to support the increasing demand for bandwidth [1]. Keeping this high-fidelity network active, often constructed from legacy switching fabrics, consumes 10 to 50% of the total IT power depending on server utilization [2]. This high-power consumption warrants immediate attention to improve efficiency and power consumption of datacenters in general and DCNs in particular.

Traditionally, DCNs are interconnected with fat tree topologies using wired links over multiple hierarchical levels of aggregation. These tree-based networks exhibit inherent limitations in scalability and oversubscription as they rely on copper and optical cable-based links and a hierarchical topology [3]. Moreover, wired network technologies require power-hungry switches and create large bundles of cables, causing design overheads and maintenance challenges and obstructions to the flow of chilled air for cooling [4]. Inefficient cooling only exacerbates the energy efficiency problem especially for small and mid-size datacenters with hundred to a few thousands of individual servers in educational institutions and private enterprises. This is because these are designed and deployed in an ad-hoc manner often leading to structural and functional heterogeneity making regular systematic design impossible. To address these common design issues faced by wired datacenters, wireless datacenter architectures are being investigated as a promising alternative. The capability of the unlicensed 60GHz wireless band to deliver very high communication rates has led to the development and approval of the IEEE 802.11ad wireless local area network (WLAN) standard [5]. Therefore, recently proposed designs leverage newly developed technologies in the unlicensed 60GHz wireless band for wireless DCNs [6], [7].

Advancements in the 60GHz technologies allow the transceivers to consume low power, in the range of a few Watts [8], [9] and establish multi-gigabit communication channels [10]. Directional horn antennas, which have been used in wireless DCNs [11] as well as more recently developed phased arrays of antennas in the 60GHz bands [12] are capable of providing high directional gains and beam steering capability between wireless transceivers. Using such antennas, the 60GHz channels can exhibit spatial reusability, allowing multiple concurrent links reusing frequency bands to be formed within the same datacenter. The low power consumption combined with the ability to form concurrent
multigigabit channels makes these transceivers ideal for use in power-efficient wireless DCNs.

In this paper, we present a server-to-server wireless DCN architecture called S2S-WiDCN based on the 60GHz wireless technology for a small or medium-size datacenter. Through direct server-to-server wireless links using directional antennas, the power-hungry switching fabric of traditional DCNs is eliminated, resulting in significant power savings in data transfer [13]. The communication between servers in the wireless DCN are achieved along horizontal lines and vertical planes as shown in Fig. 1, which shows only a few horizontal red lines and a single yellow plane for clarity. The horizontal communication lines are used for data transfer between rows, whereas, the yellow planes are used for transmissions within the same row. However, the presence of any obstruction in the datacenter aisles such as an IT technician may result in blocking of the horizontal Line-of-Sight (LoS) lines causing a failure in data transmission. Therefore, in this paper we propose a novel adaptive routing mechanism to recover from such obstructions. We compare the performance as well as power savings of the proposed server-to-server wireless DCN (S2S-WiDCN) to that of a conventional hierarchical fat tree-based DCN. We have considered various kinds of datacenter traffic for these evaluations, which are typically encountered in index-search/query-response and multimedia/video applications. In this way, we demonstrate that S2S-WiDCN is able to sustain and provide performance comparable to the conventional counterpart at five to seventeen times lower power consumption. The novel contributions of this paper are:

- We design the S2S-WiDCN architecture with an obstruction-avoidance adaptive routing for server-to-server LoS communication using 60GHz wireless links.
- We evaluate the performance of S2S-WiDCN with different kinds of traffic patterns depending on different types of applications.
- We evaluate the performance of S2S-WiDCN in presence of obstructions to LoS paths.
- We modeling and estimate the power consumption of S2S-WiDCN and compare it to traditional tree-based DCNs.

The rest of the paper is organized as follows. Section II presents the related works regarding both novel wired and wireless datacenter networks. Section III discusses the topology, communication protocol and an obstruction-avoidance routing. In Section IV we present the evaluation of S2S-WiDCN. The conclusions are discussed in Section V.

II. RELATED WORK

Various designs have been proposed to address DCN design issues such as energy consumption, cabling complexity, scalability, and over-subscription. One popular topology used today in datacenter networks is a fat-tree topology [14]. In this topology, servers are connected through a hierarchy of access, aggregate and core layer switches. The drawbacks of fat-tree networks include congestion or oversubscription at the upper levels of the hierarchy [11]. Several alternative DCN architectures have been proposed such as, BCube [15], Dccll [16], DOS [17], VL2 [18], FireFly [19] and Helios [20]. Optical interconnects were explored as well, to improve the performance even further [21]. However, these innovations still rely on copper or optical cables and do not mitigate the challenges due to high power consumption, design and maintenance of a DCN with physical links.

To alleviate the issues of DCNs with power hungry switching fabrics and bundles of cables wireless datacenters with mm-wave inter-rack links are envisioned in [6], [11], [22], and [23]. Most of the recent works on wireless datacenters propose interconnecting entire racks of servers as units with 60GHz wireless links primarily in order to utilize the commodity Ethernet switching between servers inside individual racks [6]. Phased antenna arrays or directional horn antennas are used to establish wireless links between Top-of-Racks (ToRs) in the entire datacenter [23], [24], and [25]. LoS communication paths are necessary between the antennas for reliable communication in a wireless datacenter [23]. Paths through metal frames and racks will have increased losses due to obstructions. Hence, reflectors on ceilings in the form of metallic mirrors or signal relays can be mounted to form paths where direct LoS does not exist [26]. Cylindrical [27] or polygonal [28] arrangements of servers are proposed to create LoS wireless links between servers. This however, requires non-traditional cylindrical arrangement of servers having implications on cooling, server density and scalability of the DCN which are not well known at this point. Our work is fundamentally different from the existing wireless DCN architectures as we propose LoS based server-to-server wireless connectivity while maintaining regular rectangular arrangement of server racks.

III. WIRELESS DATACENTER NETWORK ARCHITECTURE

In this section, we discuss the architecture of S2S-WiDCN. We describe the design methodologies, the adopted antenna technology, and finally its communication protocols.

A. Wireless Datacenter Network Topology

In S2S-WiDCN, the datacenter racks are laid out in the traditional rectangular pattern, adjacent to one another with aisles
running between rows of racks. In order to avoid obstruction to
the wireless communication links, we establish wireless links
only along horizontal lines and vertical planes to communica-
tion between any two servers in the three-dimensional space as
shown in Fig. 1. To achieve this, each server will be equipped
with two high gain 60 GHz antenna arrays [12]. We propose
attaching one of the antennas on the top of the server to enable
the communication in the horizontal direction and other one
on the back or front of the server projecting out from the rack
as shown in Fig. 2 to enable communication in the vertical
plane. To avoid interference and obstructions from the rack
frames, communication in the horizontal plane are restricted
only to a single line between horizontally aligned servers.

Datacenters are typically arranged in hot aisle/cold aisle lay-
out where servers in adjacent rows are either face-to-face or
back-to-back [29]. To minimize the interference between two
 neighboring rows of racks, servers will have the provision to
connect the antenna for vertical plane communication either
on the back or on the front side. This will ensure that no
two separate vertical planes will exist in a single aisle and
hence eliminate interference between vertical planes. Using
the beam-steering capability of the antenna array, LoS links
between communicating servers can be established with the
help of a control interface discussed in Section III-B. Each
server is assigned a unique ID according to its geometric
location in order to help determine the beam-steering angles.

The angles are precomputed depending on the location of the
communicating servers. All the metallic surfaces and walls of
the datacenter should be coated with anti-reflecting material
cover [30] to eliminate the multipath propagation of a signal.
Such anti-reflection material with low reflection coefficients
are relatively easily available and only add another additional
layer to the building infrastructure without significant change
in building design.

The proposed design can be adapted for racks having doors
with few minor modifications. Traditional doors for datacen-
ter racks comes in two varieties- perforated metal sheet with
breathable mesh design or acrylic glass with metal frame. For
the perforated metal doors, we propose a series of rectan-
gular opening areas as shown in Fig. 3(a), aligned with the
top antenna in all the racks. This will ensure horizontal LoS
between top antennas of the servers. The metal doors are per-
forated to aid in cooling air circulation, and the rectangular
openings foster air conditioning further. By contrast, for the
doors with acrylic glass, the doors are designed to contain
the chilled air implementing air conditioning which is differ-
ent compared to datacenters with metal doors. This material
is relatively transparent to 60GHz wireless band (compared to
actual glass) with only 1.02 dB/cm of path loss through it [31].
While each door’s glass is thinner than 1cm, in case of paths
through many doors, the link-budget analysis should take into
account this loss while designing the wireless datacenter with
this type of door. Therefore, in case, the number of rows in
the datacenter is high, this type of rack is not recommended
for the S2S-WiDCN architecture. Furthermore, to use racks
with glass door, the top antennas can be mounted on the side
panel of the racks with a high quality, low loss 60GHz cable
or waveguide rather than on top of the servers. To create LoS
between the servers in different racks across multiple rows,
a narrow open space is required between adjacent racks in the
same row to accommodate the horizontal LoS lines through
the sides of the racks. To prevent hot and cold air contam-
nation, thermal ducts individually deployed in each rack as
envisioned in [32], must be used in such racks to remove the
hot air completely from the floor through the racks.

To create LoS in the vertical plane in presence of doors
in the rack, we propose mounting the antenna arrays on an
extended panel as shown in Fig. 3(a) and (b) for metal and
acrylic door respectively. The antennas will be connected to
the corresponding servers with high-quality, low-loss 60GHz
cable or waveguide fitted to the frame which can be coupled
to the servers. This will ensure that movements of the doors
will not affect the antenna alignment.

B. Antenna Technology for the Wireless Datacenter

Each server in S2S-WiDCN is equipped with a wireless
module consisting of a transceiver and two accompany-
ing antenna arrays [12]. This particular array [12] is fabri-
cated using semiconductor lithography techniques on a single
wafer and hence, is extremely compact with a size of only
10.5mm×3.3mm. As the radiation pattern suggests, the array
provides high directional gain of 9dBi in the forward and back-
ward directions. Moreover, by adjusting the relative phase of

![Fig. 2. Single server showing two antenna arrays and the WiFi control module. Inset: each antenna array.](image)

![Fig. 3. Creating LoS between servers in presence of (a) metal rack doors in horizontal and vertical planes and (b) acrylic glass door in vertical plane.](image)
the antenna elements by activating various feed paths, beam-steering can be accomplished over an angle of 60 degrees. As horizontal communication happens in a single straight line, no beam-steering is required in the antenna arrays on top of the servers. However, as the range of beam steering angle is 60 degrees for this particular array, 6 antenna arrays are required to cover the entire 360 degree panorama in the vertical plane. Only one out of the 6 arrays will need to be signaled at any given point of time to establish a single link involving that server. Electronic beam-steering for the antenna array has negligible latency compared to mechanically steered horn antennas used in earlier wireless DCNs [25], [26]. Moreover, the antenna array being extremely compact requires very tiny space on top of each server to enable LoS communication in the horizontal direction. The effect of these spaces on the vertical server density in the datacenter racks is discussed and quantified in the next section.

This beam-steering of the transmitting and receiving antennas is controlled by using a separate control interface using IEEE 802.11 2.4/5 GHz ISM bands. Although the data rates sustained by the IEEE 802.11 2.4/5 GHz bands are much lower than the 60GHz bands, it is sufficient for the short control packets. Moreover, the isotropic antennas in the IEEE 802.11 2.4/5 GHz modules do not require any antenna steering before the control messages can be transmitted. When a traffic flow between a pair of servers is created, a short control or header packet for the flow will be sent over the IEEE 802.11 2.4/5 GHz ISM band to enable communicating servers to steer their antenna beams towards each other when required. The details of the steering are discussed in Section III-D.

C. Wireless Communication Protocols

Establishing connections between servers require reliable wireless 60GHz physical and Medium Access Control (MAC) layer protocols. The IEEE 802.11ad standard [5] is designed for 60GHz wireless LANs. This standard defines a physical layer protocol that supports beam-forming, and also supports extremely high data rates in both a single carrier (SC) and Orthogonal Frequency Division Multiplexing (OFDM) mode of operation with maximum achievable data rates are 4.62Gbps and 6.76Gbps respectively. Motivated by these high data rates, we adopt IEEE 802.11ad as the 60GHz physical layer protocol for wireless datacenters. IEEE 802.11ad MAC layer protocol incorporates a Carrier Sense Multiple Access (CSMA) mechanism for on-demand establishment of wireless links depending upon the traffic flow requirements. The MAC layer protocol establishes as many non-interfering links as possible, greedily on a first-come first-serve basis until all traffic flow demands are met or all the available OFDM channels are exhausted. The IEEE 802.11ad standard only allows wireless links to be established where a bit error rate (BER) of $3 \times 10^{-7}$ or lower can be achieved considering the signal to interference plus noise ratio (SINR) and the corresponding data rates to be sustained by the wireless link. Once a flow is found not to be feasible due to interference with already-existing flows in any of the OFDM channels, the flow is no longer considered serviced and that demand is left incomplete. We evaluate the performance of the MAC layer protocol through a comparison and analysis against similar-sized wired networks in our case studies in the next section. We use TCP as the transport layer protocol for reliable packet delivery for its widespread use and well-known characteristics in datacenter networks as well as in the Internet.

The feasibility of 60GHz communication in datacenters is established with physical channel measurements in [11] and [23]. However, the effect of temperature on the 60GHz channel has not been considered. The 60GHz channel is known to be affected by molecular absorption, which in turn, is affected by the temperature. Due to variation in temperature in a datacenter between hot and cold airflows, the wireless path loss may vary. However, the recommended range of temperature variation in a datacenter according to ASHRAE thermal guideline is between 18 to 27 degree Celsius [33]. The path loss varies by roughly 2dB/km for this range of temperature [34]. Therefore, for typical data center dimensions in the range of few meters, the variation is negligible.

D. Adaptive Routing Protocol for S2S-WiDCN

In this section, we propose an adaptive routing protocol for S2S-WiDCN, which is capable of routing traffic flows even in the presence of obstruction of the LoS between two servers due to reasons such as presence of human beings along the datacenter aisles. First, we describe the default routing mechanism followed by our proposed method to make the default routing adaptive for robustness against obstruction of LoS.

1) Default Routing Mechanism: For the default routing mechanism, we develop a Horizontal-First routing described in this subsection. The server arrangement plays a vital role in the design of this routing protocol. For the purpose of the Horizontal-First routing algorithm the servers are considered to be arranged in a 3D Cartesian coordinate system

![Fig. 4. Possible communication paths between servers situated in (a) same rack, (b) same vertical plane, (c) same horizontal line, and (d) different horizontal lines and vertical planes.](image)
with each server having a unique 3D coordinate as shown in Fig. 4. In this coordinate system, X-axis runs along rows, Y-axis runs along columns and Z-axis runs along racks as shown in Fig. 4(a). Server-to-server communication in a datacenter can be broadly classified into two types, i.e., inter-rack and intra-rack communication based on the location of the source and destination servers. All the intra-rack communications are completed in one hop in the vertical plane as shown in Fig. 4(a) whereas inter-rack communication depends on the relative position of the source and destination servers. There are three possible scenarios for inter-rack communication:

- Both the source and destination servers are located in the same vertical plane (the same row or the same Y coordinate) as shown in Fig. 4(b). In this case a direct single hop link will be established between the source and destination for data transfer.
- Both the source and destination servers are in the same column with same height above the ground (the same X and the same Z coordinates, but different Y coordinates). In this case a single hop direct link along a horizontal line will be established between the source and destination for communication as shown in Fig. 4(c).
- The source and destination servers are in different row and column (different X and Y coordinates, may or may not have the same Z coordinate). In this case a 2-hop link will be established for communication using an intermediate server as shown in Fig. 4(d). The intermediate server is the one that is in the same column and height from the source server, but in the row of the final destination (the same X and Z coordinates as that of the source and the same Y coordinate as the destination). As the data travels along the horizontal line first, the adopted routing protocol is referred to as Horizontal-First routing. In the proposed topology, every server is capable of working as a potential intermediate node.

The pseudocode of the Horizontal-First routing strategy for these various conditions is shown in Fig. 5. Control information in the form of a control packet with instructions for intermediate and destination servers to steer their antennas in the correct directions is sent over a separate IEEE 802.11 2.4/5 GHz ISM band. Each server is equipped with an IEEE 802.11 2.4/5 GHz transceiver. As the radiation pattern has main lobes in both forward and backward directions, steering is not required for the horizontal linear communication as shown in Fig. 2. For communications in the vertical planes, the server, which is ready to send data, first sends a control packet to the receiving server while simultaneously steering its antenna array towards the receiver. Upon receipt of this control message, wireless module at the receiver chooses the antenna array in the correct sector out of the set of 6 and steers that array towards the sending server by activating the correct phase differences (paths connecting the elements). The IEEE 802.11 2.4/5 GHz ISM band is also used for sending the acknowledgments to enable the CSMA-based MAC for the 60GHz links using the IEEE802.11ad protocols. In order to provide access to the Internet with necessary bandwidth, we envision gateway functionalities to be hosted at multiple server locations within the rectangular arrangement in the wireless DCN. These gateways will therefore be connected directly or indirectly, to all the servers and will also need to run firewall and security functionalities as per the requirement of the datacenter.

2) Obstruction-Avoidance Routing Mechanism: In some scenarios, the LoS necessary for the Horizontal-First routing can be obstructed. For example, when a human technician or any other obstacle is in front of an aisle it can potentially obstruct the horizontal server-to-server LoS communication between all servers in the aligned racks as shown in Fig. 6. This will not only affect servers of the rows directly adjacent to the human obstruction but also servers in racks of all rows that use those horizontal paths for inter-row communication. We propose an Obstruction-Avoidance adaptive routing mechanism to address this failure model and to successfully route traffic flows in presence of such obstructions between specific racks. In the adaptive routing, all servers start sending packets following the default Horizontal-First routing strategy outlined earlier. We utilize the CSMA acknowledgment mechanism to detect a failed transmission after several trials.
Fig. 7. Pseudocode for routing protocol if obstruction is detected.

```
if Obstruction detected in the Horizontal line← false
    Default Horizontal-First Routing
else
    if servers are in different row and same column
        choose a random server in source plane in a different rack
        as 1st intermediate node
        select the corresponding server in destination plane (same
        height and same column as of 1st intermediate node) as the
        2nd intermediate node
        if Obstruction detected in the Horizontal line ← true
            go to: 5
        else
            route the flow in 3 hop using Vertical-First routing
    end if
end if
if servers are in different row and different column
    choose the server in the source plane situated in the same
    height and same column of the destination server as
    intermediate node
    if Obstruction detected in the Horizontal line ← true
        go to: 14
    else
        route the flow in 2 hop using Vertical-First Routing
    end if
end if
```

According to the IEEE 802.11ad MAC. Then a retransmis-
sion is attempted again using the adaptive routing algorithm
as described in Fig. 7.

In this adaptive routing strategy, after detecting a failed
transmission, the sender determines the route of the next trans-
mision attempt. If the destination server is in another rack
in the same row, the sender retransmits the flow using the
default Horizontal-First routing algorithm. This is because
the failed transmission did not happen because of the hori-
zontal LoS obstruction from the technician as that LoS link
was not used in the first transmission attempt. The transmis-
sion happens over the back/front vertical plane, which is not
obstructed by the failure model under consideration. However,
if the destination is in another row, instead of adopting the
Horizontal-First approach, a Vertical-First routing approach
is adopted where, a server in the same row but a differ-
ent rack is chosen and the path is established to
that server using the back/front vertical plane. Control packets
are sent over the IEEE 802.11 2.4/5GHz ISM control plane
to establish the links using beam-steering. From that other
server, again the default Horizontal-First routing is adopted to
reach the final destination. Such a path is shown in Fig. 6.
If the randomly chosen intermediate server for Obstruction-
Avoidance routing is also obstructed by another technician,
the Obstruction-Avoidance routing approach can be repeated
again till the Horizontal-First routing is successful to transfer
packets to the destination row. In this way, this adaptive rout-
ingen mechanism can be extended to an obstruction model with
multiple technicians obstructing multiple racks in the datacen-
ter. The performance of S2S-WiDCN in presence of such an
obstruction will be degraded for the obstructed flows. In the
configuration of a 20×8 array of racks as shown in
next section, we describe the performance of S2S-WiDCN in
presence of various flow traffic patterns and obstructions.

IV. MODELING, RESULTS AND ANALYSIS

In this section, we present our modeling, results and corre-
responding analysis of S2S-WiDCN. We first demonstrate that
it can sustain comparable performance compared to that of
cventional DCNs with network-level simulations in terms of
communications between servers within a datacenter. Next,
we present the estimates of power consumption to highlight
the main benefit of S2S-WiDCN.

A. Simulation Platform

We use the Network Simulator-3 (NS-3) suite [35] to evalu-
ate S2S-WiDCN. NS-3 supports the characteristics of wireless
propagation as well as network-level communications. It is
important to simulate both the propagation and network-level
communication characteristics accurately in order to obtain
credible performance results. We use a modified version of
NS-3 extended with features of wireless datacenter including
the 60GHz band and the IEEE 802.11ad standard as discussed
in [11]. This extension incorporates interference modeling, bit
error rates, and directional antenna modeling. The accuracy of
these parameters is verified with physical layer measurements
of prototype 60GHz hardware [11]. Additionally, we introduce
criteria for wireless link selection to enable many concurrent
links and modify the IEEE 802.11ad physical layer to allow
multiple OFDM channels.

We have considered two datacenter sizes in this paper to
represent datacenters belonging to two different classes. The
first one, with a total of 800 servers, is a small-sized datacen-
ter representing those in an educational institution. The second
one is a mid-sized datacenter and has 1600 servers represent-
ing those in private enterprises [36], [37]. In both cases, the
servers are arranged in a 20×8 array of racks as shown in
Fig. 8. There are 10 racks arranged in a single row and two
columns of 8 rows, totaling 160 racks. Each rack occupies an
area of 0.6m×0.9m and is 2m high. Adjacent rows are separated by 1m and the width of the central aisle is 2m. Each rack contains 5 and 10 servers for the 800 and 1600 server datacenters respectively. In our simulations, the racks are assumed to be without any front or back side door.

To account for the latency required to set up the 60GHz communication links using the exchange of control information over the IEEE 802.11 2.4/5 GHz ISM band and beam-steering, we run a conservative simulation using NS-3 with the packet size of 200 bytes representing control packets of 60GHz S2S-WiDCN datacenter with beam-steering information. Each new flow according to the flow arrival process discussed in Section IV-B (1) in the DCN is considered to generate a control packet. The flow arrival process is considered to be the same as described in Section IV-B (1). The simulation showed that a single wireless access-point can sustain the demand for the control packet traffic with an average latency of 266µs in a system with 240 servers. As the maximum number of servers in a single vertical plane is 200, a single access point per row should be enough to server the requirement. This latency overhead is considered in the evaluation of S2S-WiDCN next.

B. Performance Evaluation and Analysis

Here we present the simulation results of S2S-WiDCN along with a comparative analysis with respect to existing DCNs in terms of flow completion duration and throughput. The throughput is defined as the average rate of bit transferred per second over the DCN. We compare S2S-WiDCN with a traditional wired fat-tree based DCN and a ToR-ToR wireless DCN. In the traditional wired fat-tree based DCN, we have considered 3 hierarchical layers consisting of 160 access, 2 aggregate and 2 core layer switches similar to the architecture evaluated in [22]. Each traditional DCN link between the access, aggregate and core layer switches is considered to have a channel bandwidth of 10Gbps. The intra-rack communication in the fat-tree based DCN occurs through the ToR switch, which has 1Gbps direct links to each server in its rack. Although the proposed wireless DCN is a direct server-to-server network, we have not compared it with a wired server-to-server all-to-all DCN because such a network is not practical and will have extremely high degree of connectivity at each server. In the ToR-to-ToR wireless DCN (ToR-WiDCN) the intra-rack communication is managed in a traditional manner, same as in the conventional wired DCN. The inter-rack communication is done with ToR-to-ToR 60GHz wireless links using the same physical and MAC layers as in S2S-WiDCN. We use the simulation platform described in Section IV-A and the datacenter traffic model discussed below to evaluate these DCNs.

1) Datacenter Traffic Model: S2S-WiDCN is at first evaluated with a set of traffic flows based on application demands. These demands reflect real traffic within the network over a period of time. The application demands include information specifying the flow arrival time, identity of the source and destination, the flow size, and the data rate at which the traffic is generated. Real datacenter traffic for different classes of datacenters such as educational (small), private (medium) and corporate (large) running typical query/response based applications like map-reduce and index-search are measured in [36]. Using these measured traffic flows, a Poisson shot-noise based model to synthesize datacenter traffic is proposed and verified in [38]. According to [38], the new flow arrival time, the flow duration and the injection rate for each application follow a Poisson, Pareto and, Gaussian distribution respectively. The new flow arrival time is generated using a Poisson distribution with an average flow arrival rate. The average flow arrival rate is considered to be 1000 flows/second for the small-sized and 3000 flows/second for the mid-sized datacenter [36]. The flow injection rate and the flow duration are independent parameters. The flow duration refers to the time required to inject the flow into the DCN and is different from the flow completion duration which is a performance metric. The flow size is a product of the injection rate and duration and therefore depends on both. In our evaluations, we have considered a skewed Gaussian distribution for the injection rate to have a mean of 1.0kBps such that 90% of the traffic rates are less than 10kBps and the remaining 10% can be as high as several MBps as per the traffic model from [38]. Application flow duration is generated following an independent Pareto distribution having a minimum duration of 10 microseconds [36]. The characteristic parameters for these distributions are summarized in Table I. As customary for TCP traffic, we have considered the size of all packets in the generated flows equal to the maximum transmission unit (1500 bytes). The CDFs of the transmission rates of both of these two traffics are shown in Fig. 9 (a) and Fig. 9 (b).

According to [36] and [38], in a DCN, around 80% of the total traffic stays within the same rack. Only 20% communication takes place between the servers situated in different racks. For each new flow in our simulations, a random destination was chosen such that 80% of the destinations belonged to the same rack as the flow source. The simulations were conducted such that no new flows were allowed to be injected after

<table>
<thead>
<tr>
<th>Name</th>
<th>Model</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow arrival rate, $\lambda(t)$</td>
<td>Poisson for small DCN, $6000$, for medium DCN</td>
<td></td>
</tr>
<tr>
<td>Flow duration, $\Delta t$</td>
<td>Pareto $\alpha = 1.504$, $\mu = 1.0091$</td>
<td></td>
</tr>
<tr>
<td>Flow transmission rate, $Y$</td>
<td>Gaussian $\mu = 8.606$ Kbps, $\sigma = 69.936$ Kbps</td>
<td></td>
</tr>
<tr>
<td>Flow size, $S$</td>
<td>$Y$, $\Delta t$, $E[S] = 1.0647$ KB</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 9. CDF of flow transmission rates of the (a) index/query based traffic for small size (b) index/query based traffic for medium size (c) multimedia traffic for small size DCN.
20 seconds but the simulations were run until the completion of all the established flows. Next, we analyze the performance of the DCNs in the presence of this traffic pattern.

2) Flow Completion Duration: Here, we estimate the flow completion duration of the applications in the different DCNs. The flow completion duration for both small and medium sizes of all 3 different DCNs are shown in Fig. 10. It can be seen that the average completion duration of S2S-WiDCN is lower than that of the wired network for both sizes because of fewer number of hops involved, resulting in lower time of flight and switching overheads. For the wired network, two servers even in a single rack need to go through the access layer switch to communicate, requiring at least two hops. On the other hand, in the wireless architecture, those two servers can communicate directly using a single hop. As the major portion of communications in datacenters is intra-rack [38], the reduction in delay of intra-rack flows in the wireless DCN is likely to reduce the overall average flow completion duration. The beam-steering latency is 266µs for exchange of control information over the control plane is considered while computing the flow completion duration of S2S-WiDCN as shown in Fig. 10.

Fig. 10 also captures the minimum and maximum flow completion durations for all the DCNs. The minimum flow completion duration in case of S2S-WiDCN is higher than that of the fat-tree based DCN due to the beam-steering latency. This effect is also seen in the ToR-WiDCN. However, as seen in [38] only a very small fraction of the flows have such short flow durations.

3) Throughput: Fig. 11 shows the average throughput along with the minimum and maximum for all flows in each DCN. As we can see for both sizes S2S-WiDCN provides the same throughput as that of the wired traditional fat-tree based DCN. Even the ToR-WiDCN is capable of achieving a similar performance. All the DCNs achieve throughputs which closely match the injection rate of the flows. This is because, in the traditional wired network, the available bandwidth per link is 1.06Gbps and that available for OFDM wireless channels is 0.563Gbps. We have considered 3 sub-carriers in the 60GHz band each with maximum OFDM rates of 6.76Gbps, which are in turn, split into 12 sub-channels each, to cater to all the application flows injected into the wireless DCNs. So the physical bandwidths of both wired and wireless channels are much higher than the average injection rates encountered in these scenarios. Moreover, we find that in S2S-WiDCN the throughput is higher than that in both the traditional wired DCN and ToR-WiDCN. This is because the lower number of hops in S2S-WiDCN implies that the flow will encounter fewer intermediate nodes resulting in a reduced likelihood of being congested. Therefore, for the type of application considered here, S2S-WiDCN performs better than the fat-tree based DCN. Although the flow arrival rate increases with the number of servers, its impact on performance is marginal as the flow transmission rates of most of the flows are less than the 60GHz OFDM channel capacity.

4) Evaluation With Different Traffic Patterns: In this subsection, we further evaluate S2S-WiDCN with a different traffic scenario that can be encountered in multimedia or video hosting/streaming servers. This kind of application is significantly different from query/index search applications primarily from two perspectives. First, this kind of traffic generally has a higher data rate requirement. Multimedia/video streaming servers hosting applications are seen to have average data rates of 100Mbps [39]. Second, these applications typically experience bursty flow arrivals [40]. In order to evaluate S2S-WiDCN with multimedia/video, traffic we adopt a bursty flow arrival rate with a high average data rate of 100Mbps based on [39]. Unlike the query based traffic where the flow arrival process is assumed to be a Poisson process, the bursty flow arrival is modeled as a fractal process. The entire simulation duration is divided into windows of 30ms and each window is randomly chosen to be either in ON or OFF phase. New flow arrivals are allowed only in the ON phase. The new flows have an arrival rate such that the overall average flow arrival rate is the same as that of the query-based traffic for the simulation duration. The details of this traffic are listed in Table II. The CDF of number of concurrent flow arrivals within a window size of 30ms for both the query-based traffic and the bursty multi-media traffic is shown in Fig. 12.
than 50 whereas, it can be as high as 250 in the bursty flow arrival pattern. The bursty traffic pattern coupled with the high flow rate requirements of this traffic type is therefore, expected to stress the DCN more compared to the query/response type traffic.

Fig. 13 shows the average flow completion duration and average throughput of a small sized DCN with 800 servers for this multimedia/video traffic. We have compared the performance of S2S-WiDCN with a fat-tree based wired DCN with this traffic. Similar to query-based traffic, a few small flows with very low flow durations incur a higher flow completion duration in S2S-WiDCN as can be seen in the minimum of the flow duration range in Fig. 13 (a). This is because the beam-steering latency of S2S-WiDCN is higher than the flow duration of these very small flows. Moreover, we can see that some of the high data rate flows achieve lower throughputs in S2S-WiDCN compared to the fat-tree wired DCN as shown by the maximum value of the range of throughput of S2S-WiDCN in Fig. 13 (b). This is because the maximum data rate per OFDM channel that can be supported in S2S-WiDCN is 0.563Gbps. As the flow rates follow a Gaussian distribution with a mean of 100Mbps, a few flows require a data rate higher than 0.563Gbps. The effective throughput of these flows are reduced in S2S-WiDCN. However, as can be seen from the CDF of flow rates in Fig. 9(c), these flows with data rates higher than 0.563Gbps are few in number and therefore do not affect the average throughput significantly.

5) Evaluation in the Presence of LoS Obstruction: In this subsection, we evaluate the performance of S2S-WiDCN in the presence of an LoS obstruction. For this purpose, a scenario is assumed where an IT technician is present in front of the column 3 of row 2 in the floorplan shown in Fig. 8. We have assumed this obstruction to be stationary within the observed window of 20s, which is reasonable as it is a human obstruction. Due to the presence of the obstruction, the traffic flows from all servers in all rows corresponding to the obstructed column, which were supposed to be routed through the horizontal lines in column 3 with the default Horizontal-First routing protocol, need to follow the alternate Obstruction-Avoidance routing mechanism.

Fig. 14 shows the flow completion duration of the small-scale S2S-WiDCN with 800 servers with adaptive routing, in the presence of the obstruction. We have evaluated the impact of the obstruction on the overall flow completion duration as well as that of the affected traffic flows only. As 80% of the traffic generated from each server is intra-rack, they use the vertical plane for communication and are unaffected by the presence of a technician. Among the 20% inter-rack traffic, a smaller percentage requires inter-row paths going in the horizontal direction as inter-rack traffic in the same row also uses the vertical plane. The percentage of traffic flows from all rows, whose inter-row traffic is obstructed, is 1.87% of all the flows in S2S-WiDCN. Hence, the LoS obstruction has very little impact on the overall average flow completion duration due to the small percentage of traffic flows that are affected. We have further investigated the impact of the obstruction as a function of the number of retransmission trials made before rerouting using the Vertical-First path of the adaptive routing method. As can be seen, in case of a higher number of allowed retransmission attempts before rerouting the flow, the impact of the obstruction is higher. Hence, the number of retransmission attempts can be customized based on the performance demands of the applications.

### TABLE II

| Table II: Parameters for Video/Multimedia Traffic Generation |
|-----------------|-----------------|-----------------|
| Name            | Model           | Parameters      |
| Flow arrival time, $\lambda(t)$ | $N$ | $\lambda(t) = 1000$ |
| Flow duration, $D_x$ | Pareto          | $a_\lambda = 1.504$ s, $M_x = 1.0001$ s |
| Flow transmission rate, $Y_x$ | Gaussian        | $E[X_i] = 100.0$ Mbps, $\sigma[X_i] = 114.153$ Mbps |
| Flow size, $S_x$ | $Y_x \cdot D_x$ | $E[S_x] = 101.025$ MB |

Fig. 12. Distribution of number of concurrent connections for (a) query-based traffic (b) bursty multimedia traffic.
C. Power Consumption Analysis

From the previous section it is seen that S2S-WiDCN can sustain comparable performance as that of a traditional fat-tree DCN. In this section we evaluate its most important benefit, which is the reduction in power consumption of S2S-WiDCN with respect to the traditional DCN. We discuss the model and parameters used in the power estimation followed by the results.

1) Power Model: It is a complex task to estimate the actual electrical power consumed by a datacenter. The power consumption depends on several internal factors such as utilization of computing power, the cooling mechanism, and datacenter networks. Datacenter power consumption is also affected by external parameters like the geographical location, weather, temperature, and humidity. Our focus is solely on networking, and we only analyze the power consumption involved in networking. In this regard, we assume that the power consumption other than networking is identical in all the cases. We estimate power consumption for wired DCNs using commercially available data from Cisco network switches [41], [42], [43]. Specifically, we use Cisco 7702 for the core-level switches, Cisco 9508 at the aggregation level, and Cisco 9372 for access-level switches. We also use the data from Silicom PE2G2135 for the power consumption of the network interface cards (NIC) [44]. The power consumption of each core and aggregate switches are as follows:

\[
P_{\text{Core}} = P_{I/O \text{ Cards}} + P_{\text{Fan Tray}} + P_{\text{Sys Ctrl}}.
\]

\[
P_{\text{Agg}} = P_{I/O \text{Cards}} + P_{\text{Fan Tray}} + P_{\text{Sys Ctrl}} + P_{\text{Fabric}}.
\]

where \( P_{I/O \text{ Cards}} \), \( P_{\text{Fan Tray}} \), \( P_{\text{Sys Ctrl}} \), \( P_{\text{Fabric}} \) represent the power consumption of the input/output card, fanout ports, supervisor controller, cables and system controller respectively.

Then the total power is calculated as follows:

\[
P_{\text{Total}} = N_{\text{Core}} P_{\text{Core}} + N_{\text{Agg}} P_{\text{Agg}} + N_{\text{Acc}} P_{\text{Acc}} + N_S P_{\text{NIC}}.
\]

where, \( N_{\text{Core}} \), \( N_{\text{Agg}} \), \( N_{\text{Acc}} \), \( N_S \) are the number of core, aggregation, access switches, and the total number of servers, respectively; \( P_{\text{Core}} \), \( P_{\text{Agg}} \), \( P_{\text{Acc}} \), \( P_{\text{NIC}} \) are the power consumptions of an individual core, aggregation, access switches and network interface cards, respectively.

In S2S-WiDCN, however, no core, aggregate or access layer switches are needed, but only antennas, transceivers and NICs are required for wireless communication. The power consumption of the wireless 60GHz transceiver is modeled based upon the assessment of emerging 60GHz transceivers such as [8] and [9]. The NICs of S2S-WiDCN are equipped with two transceivers for horizontal and vertical communication.

In the traditional DCN, external connections are established via the two Cisco7702 switches. To provide equivalent connectivity in S2S-WiDCN, we employ two servers to work as gateways, and their power consumption is modeled as that of the Cisco 7702 switch. The power consumption of communication per server in S2S-WiDCN is calculated as:

\[
P_{\text{Wireless}} = P_{60 \text{GHz Tran}} + P_{\text{Antenna}} + P_{\text{WifiControl}} + P_{\text{NIC}}.
\]

where \( P_{60 \text{GHz Tran}} \) is the power consumption of the 60GHz transceiver, \( P_{\text{WifiControl}} \) is the power consumption of the 802.11 2.4/5 GHz ISM adapter for the control channel, and \( P_{\text{Antenna}} \) is the power consumed by the antennas for beamsteering. We conservatively adopt \( P_{60 \text{GHz Tran}} \) and \( P_{\text{Antenna}} \) to be 1.7W [8], [9]. We consider \( P_{\text{WifiControl}} \) to be 220mW from the datasheet of D-link DWA-171 2.4GHz ISM adapter.

Finally, the total power consumption in S2S-WiDCN becomes:

\[
P_{\text{Total WiDCN}} = N_{\text{Core}} P_{\text{Core}} + N_S P_{\text{Wireless}}.
\]

The power consumption of all the off-the-shelf switching components used in our model is shown in Table III.

2) Comparative Analysis of Power Consumption: We posit that the primary advantage of S2S-WiDCN is lower power consumption. To study this more deeply, the total power consumption estimated for the typical and maximum cases for all the DCNs is shown in Fig. 15. In the “typical” scenario, the average power consumption of every device is used, while their maximum power consumption is considered in the

<table>
<thead>
<tr>
<th>Device</th>
<th>Model</th>
<th>Used in</th>
<th>Power Consumption (Watt)</th>
<th>Typical</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access Layer Switch</td>
<td>Cisco 9372</td>
<td>Fat-Tree, ToRWiDCN</td>
<td>210.0</td>
<td>650.0</td>
<td></td>
</tr>
<tr>
<td>Aggregate Layer Switch</td>
<td>Cisco 9508</td>
<td>Fat-Tree</td>
<td>2522.0</td>
<td>3324.0</td>
<td></td>
</tr>
<tr>
<td>Core Layer Switch</td>
<td>Cisco 7702</td>
<td>Fat-Tree, ToRWiDCN, S2S-WiDCN</td>
<td>837.0</td>
<td>1305.0</td>
<td></td>
</tr>
<tr>
<td>Network Interface Card</td>
<td>Silicom PE2G2135</td>
<td>Fat-Tree, ToRWiDCN, S2S-WiDCN</td>
<td>2.64</td>
<td>3.36</td>
<td></td>
</tr>
<tr>
<td>IEEE802.11 2.4/5GHz ISM</td>
<td>D-link DWA-171</td>
<td>S2S-WiDCN</td>
<td>0.22</td>
<td>0.24</td>
<td></td>
</tr>
</tbody>
</table>

![Fig. 15. Total power consumption of various DCN architectures.](image-url)
“maximum” scenario. For small-sized and mid-sized DCNs, the result shows eight-fold and five-fold reduction in power consumption of S2S-WiDCN compared to the traditional fat-tree based DCN topology in the “typical” case. The maximum improvement in power consumption is observed to be seventeen-fold for the small-sized DCN in the “maximum” utilization scenario. The complete elimination of power-hungry aggregate and access-layer switches contribute to this drastic reduction primarily. Since access-level switches are needed per rack in the ToR-WiDCN, its reduction in power consumption is not as significant as that of S2S-WiDCN. Therefore, by establishing direct links between servers S2S-WiDCN reduces the power consumption significantly compared to all the DCNs that require higher level switches.

D. Estimate of the Overheads

Vertically adjacent servers need to have space between them to accommodate the antenna arrays at the top of the servers. With the compact size of the antenna of only 10.5mm × 3.3mm, we envision a separation of 30mm between the servers should be enough. As the typical height of a server is 90mm, we can reduce the vertical server density to be around 33%. In other words, this reduction in server density can increase in rack height by 33% to accommodate the same number of servers per rack. We anticipate that, this does not have significant impact on infrastructure cost. In fact, this type of server arrangement will aid in cooling by enabling better chilled air circulation around each server.

As a narrow LoS exists between the servers communicating along the same horizontal line, the possibility of interference with adjacent receivers decreases as the top and bottom server structures do not allow the antenna radiation lobe to reach other vertically adjacent receivers as depicted in Fig. 16 (a). While multipath transmission may still be caused by diffraction around rack structures even though non-reflective coating is used on all reflective surfaces, the narrow aperture to establish the LoS between the antennas makes it unlikely that the multipath non-LoS signals will have significant power. Similarly, the wireless communication in a vertical plane also does not interfere with the wireless communication with its adjacent vertical plane as row of racks will act as a shield against the wireless signal of one plane interfering with a different plane as shown in Fig. 16 (b). Moreover, the half-angle of the main radiation lobe is narrow enough to prevent the transmission from one vertical plane in reaching another.

However, in both cases, receivers in the LoS of an active communication cannot use the same channel to receive data from another sender. A different OFDM channel is used in such a case to avoid interference.

V. CONCLUSION AND FUTURE WORK

The challenges in current DCN’s are high design and maintenance cost, huge power consumption, high cabling complexity, hard to keep accurate per-cable information and inefficient cooling. Structured cabling bundle incur significant initial effort and cost to setup and still may cause airflow blockage. All these challenges can be overcome by using the proposed completely wireless server-to-server DCN architecture. We observe that S2S-WiDCN provides comparable flow completion duration and throughput to a conventional fat-tree based DCN for query/response and multimedia/video based applications, while reducing the power consumption by five to seventeen times.

REFERENCES


Sayed Ashraf Mamun (S’XX) received the B.Sc. and M.Sc. Engineering degrees from the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka. He is currently pursuing the Ph.D. degree with the GCEES Department, Rochester Institute of Technology. His current research interests include wireless networks and communications.

Sree Gowrishankar Umamaheswaran (S’XX) received the B.S. degree from the Sri Venkateswara College of Engineering and the M.S. degree from the Computer Engineering Department, Rochester Institute of Technology. His research interests include wired and wireless networking.

Minseok Kwon received the B.S. degree in computer science and statistics from Seoul National University, Seoul, South Korea, in 1994, and the M.S. and Ph.D. degrees in computer science from Purdue University in 2001 and 2004, respectively. He is currently an Associate Professor with the Department of Computer Science, Rochester Institute of Technology, Rochester, NY, USA. His research interests are in the area of computer networking, wireless networks, and network security.

Andres Kwasinski (S’98–M’04–SM’11) received the Diploma degree in electrical engineering from the Buenos Aires Institute of Technology, Buenos Aires, Argentina, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, MD, USA, in 2000 and 2004, respectively. He is currently a Professor with the Department of Computer Engineering, Rochester Institute of Technology, Rochester, NY, USA. His current areas of research include cognitive radios and wireless networks, cross-layer techniques in wireless communications, smart infrastructures and networking, and operation of the Internet-of-Things.
AUTHOR QUERIES
AUTHOR PLEASE ANSWER ALL QUERIES

PLEASE NOTE: We cannot accept new source files as corrections for your paper. If possible, please annotate the PDF proof we have sent you with your corrections and upload it via the Author Gateway. Alternatively, you may send us your corrections in list format. You may also upload revised graphics via the Author Gateway.

AQ1: Note that if you require corrections/changes to tables or figures, you must supply the revised files, as these items are not edited for you.

AQ2: Please provide the report number for Reference [7].

AQ3: Please provide the organization location for References [29] and [33].

AQ4: Please provide the missing IEEE membership years for the authors “S. A. Mamun,” “S. G. Umamaheswaran,” and “A. Ganguly.”