



HAL
open science

Leveraging utilization as performance metric for CDN enabled energy efficient internet of things

Saif Ul Islam, Hasan Ali Khattak, Jean-Marc Pierson, Ikram Ud Din, Ahmad Almogren, Mohsen Guisani, Mansour Zuair

► **To cite this version:**

Saif Ul Islam, Hasan Ali Khattak, Jean-Marc Pierson, Ikram Ud Din, Ahmad Almogren, et al.. Leveraging utilization as performance metric for CDN enabled energy efficient internet of things. Measurement - Journal of the International Measurement Confederation (IMEKO), 2019, 147, pp.106814. 10.1016/j.measurement.2019.07.042 . hal-02878976

HAL Id: hal-02878976

<https://hal.science/hal-02878976v1>

Submitted on 23 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive Toulouse Archive Ouverte


OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in: <https://oatao.univ-toulouse.fr/26261>

Official URL:

<https://doi.org/10.1016/j.measurement.2019.07.042>

To cite this version:

Islam, Saif Ul and Khattak, Hasan Ali and Pierson, Jean-Marc  and Din, Ikram Ud and Almogren, Ahmad and Guisani, Mohsen and Zuair, Mansour *Leveraging utilization as performance metric for CDN enabled energy efficient internet of things*. (2019) *Measurement*, 147. 106814. ISSN 0263-2241.

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Leveraging utilization as performance metric for CDN enabled energy efficient internet of things

Saif Ul Islam ^a, Hasan Ali Khattak ^b, Jean-Marc Pierson ^c, Ikram Ud Din ^d, Ahmad Almogren ^{e,*}, Mohsen Guizani ^f, Mansour Zuair ^g

^a Department of Computer Science, Dr. A. Q. Khan Institute of Computer Science and Information Technology, Rawalpindi, Pakistan

^b Department of Computer Science, COMSATS University Islamabad, Islamabad, Pakistan

^c University Paul Sabatier – Toulouse III, Toulouse, France

^d Department of Information Technology, The University of Haripur, Pakistan

^e Chair of Cyber Security, Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11633, Saudi Arabia

^f Department of Computer Science & Engineering, Qatar University, Doha 2713, Qatar

^g Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

A B S T R A C T

The exponential increase in the Internet of Things (IoT) has drastically augmented the need for computing and storage resources. IoT devices generate massive data that requires to be efficiently stored and disseminated for further processing and decision making. Given this, Content Distribution Networks (CDN) play a vital role to store and serve contents through geographically distributed surrogate servers in an IoT environment. Thanks to the explosive rise in IoT devices, CDN installations are also significantly increasing, hence, consuming a substantial amount of energy. The dynamic energy consumption of CDN resources is proportional to their utilization. Therefore, it is very crucial to design and develop utilization aware policies and strategies for sustainable content distribution. In the literature, different approaches have been proposed to improve the CDN performance by improving CDN utility in terms of security, authentication, and confidentiality, etc. Whereas, these works largely neglect the efficient utilization of CDN resources. In this perspective, this paper investigates the fact that CDN utility does not reflect the utilization of CDN resources. For this purpose, a comprehensive analysis is performed by comparing both metrics – CDN utility and CDN utilization, under various CDN scenarios. The results show that CDN utilization exhibits lower values even if the CDN utility is showing promising results. Hence, we recommend to consider utilization as a major metric to evaluate the performance of a CDN policy.

Keywords:

Content Distribution Networks
CDN utility
CDN utilization
CDN Security
Confidentiality
Authentication

1. Introduction

The rapid increase in Internet usage has led to significant developments in the large scale distributed computing paradigm. In this context, CDN is one of the popular and major large scale content providing distributed systems [1] intended to provide efficient content delivery to geographically dispersed Internet users through multiple distributed surrogate servers. Initially, envisioned as a reverse proxy mechanism to push content to the end users, CDNs have been an ideal paradigm for the newly emerging domain of

IoT. IoT is basically used as a source for collecting massive amount of smart cities data which are used for data analysis through various machine learning techniques [2,3]. Recently, numerous applications have emerged where the users consume large amount of data for various purposes, for instance, traffic management [4,5], information systems, smart agriculture [6] urban management [7] and health monitoring for precision medicine.

The massive increase is expected in the number of IoT devices by year 2020 as predicted by Gartner [8]. This has attracted a vast majority of the service providers towards the CDN based solutions either through Mobile CDN (M CDN) or Network CDN (N CDN) [9]. The possibilities to scale networks and their respective performance by leveraging CDN has been highly demanding [10]. Several applications have been envisioned through CDN based infrastructures to improve the overall performance of the network. Consequently, according to the latest Cisco Visual Networking Index [11], the CDNs are responsible to handle 52% of the overall Internet

* Corresponding author at: Chair of Cyber Security, Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11633, Saudi Arabia.

E-mail addresses: saiflu2004@gmail.com (S. Ul Islam), hasan.alikhattak@gmail.com (H.A. Khattak), pierson@irit.fr (J.-M. Pierson), ikramuddin205@yahoo.com (I.U. Din), ahalmogren@KSU.EDU.SA (A. Almogren), mguizani@ieee.org (M. Guizani), zuair@ksu.edu.sa (M. Zuair).

traffic in year 2018 that is predicted to be augmented up to 71% in year 2021. The reason behind this is the fact that the CDN improves the content distribution by bringing content to user's vicinity that results in optimized bandwidth utilization and improved latency [12]. Moreover, service availability and fault tolerance are also ensured by deploying redundant CDN resources.

Despite all the above mentioned advantages, the CDNs face several serious issues that need to be addressed. The increase in heterogeneity and richness of content have made timely and efficient content delivery a quite challenging task. Moreover, the continuously changing trends of content types and high demand in streaming media are also consuming considerable bandwidth. Additionally, sophisticated IoT applications require low latency and high speed. To fulfill these requirements, the service providers are installing powerful and energy hungry devices that ultimately increase the network and the overall system cost. Consequently, the energy consumption of such systems is exponentially increasing. Therefore, efficient resource management in the CDN is crucial to optimize resource utilization. Moreover, there is a dire need to renew the evaluation standards and metrics to meet the modern system requirements.

It is observed that the surrogate servers have lower average utilization throughout the year. Majority of the servers remain underutilized for a considerable amount of time. To ensure the service availability, content providers keep the servers available to satisfy customers' demands that result in high wastage of resources. Consequently, the service cost is substantially augmented. To evaluate the performance of such systems, CDN utility (also called net utility) is used as a metric. In fact, the net utility is the ratio of content uploaded to the content downloaded by a surrogate server [14]. Whereas, the net utility does not take into consideration the overall utilization of resources. It is thus elaborated that utilization presents itself as a significant metric while considering the performance of an IoT based content delivery architecture.

In this paper, a rigorous evaluation of net utility and CDN utilization is presented. The results present that even if the surrogate servers have lower utilization, CDN utility exhibits prominent values. It proves that evaluating CDN performance only based on net utility is inadequate to efficient resource utilization. CDN utilization itself is the metric which determines a surrogate server utilization. It is used to show the ratio of active connections to the total number of connections any surrogate server can handle simultaneously. This paper aims at evaluating and comparing the CDN utility and CDN utilization under different IoT scenarios. The main contributions of this paper are listed below:

- CDN utility and CDN utilization are detailed along with their models
- The differences between both metrics (CDN utility and CDN utilization) have been highlighted
- We exhibit that the CDN utilization has a direct impact on the energy consumption of resources
- We compare different policies to demonstrate that the CDN utilization is an important metric for performance evaluation of a CDN
- Experiments are conducted using a state of the art simulator named CDNSim considering various IoT based CDN scenarios
- We define the CDN utility as a supporting metric to analyze the effectiveness of a CDN implementation in an IoT environment
- An extensive comparison of net utility and CDN utilization is carried out
- We evaluate the CDN utility and the CDN utilization under different scenarios depicting IoT demand and demonstrate that the CDN utility does not reflect the utilization of surrogate servers

Rest of the paper is organized as follows: Section 2 presents the background and overview of state of the art. The paper presents and differentiates CDN Utility and CDN Utilization in Sections 2.1 and 2.2, respectively. Next, the paper presents the elements influencing the CDN Utilization in Section 3. In Section 4.1, the paper presents the dataset we have used for this experimental work along with configurations needed to set up the experimental environment for performance evaluations. Finally, we present results and discussions related to the CDN experiment in Section 5, where the paper has analyzed and compared CDN utility and CDN utilization. Section 6 identifies the trends of IoT traffic and shows usefulness of surrogate servers to circulate IoT data in a content distribution network.

2. Background

CDNs have gained considerable recognition due to their highly scalable distributed nature. These systems bring contents near the end users that result in higher bandwidth and lower delays. Moreover, CDNs are used to balance the cost and Quality of Service (QoS) for content distribution [15]. CDN providers contact with content providers for some particular period of time and for a particular set of users from specific geographical locations under specific Service Level Agreements (SLAs). Content providers normally share these CDN surrogate servers with other content providers. Leading large scale CDN companies¹, such as Akamai and Limelight are providing services to Facebook, Twitter, BBC, CNN, and YouTube.

Theoretical CDN models have been developed to effectively solve pertinent problems, especially in resource management and re-direction policies [16,17]. Utility is one of the metrics that can be used to quantify traffic activities in a network and can be used to represent the usefulness of servers in terms of data replication within the network. Similarly, the CDN Utility is used to identify the performance of surrogate servers in a CDN [18] and can also be called as CDN utility.

Stamos et al. [18] appraised the CDN surrogate servers' utility and pinpointed certain parameters that influence the surrogate server's utility. The authors analyzed the CDN utility, hit ratio, mean response time, and byte hit ratio. Moreover, the study also considered several aspects, such as the utility of CDN against network topology, generation of requests, and policies for re-direction. The study also revealed that there was no change in peaks in CDN utility with discrete network topologies, when cache size is increased.

In case of frequent transmission of files of larger sizes, the CDN utility increases that has also proven effective in terms of pricing to content providers. While considering CDN utility, authors identified a performance peak invariant of network topology, the network traffic model, and the website model. Surrogate servers were evaluated, it was found that the closest surrogate server with cooperation performed significantly better as compared to other policies for request re-direction.

Moreover, a performance peak in the CDN utility is also observed due to the cooperation among the servers. On the other hand, such a performance peak is not exhibited by the closest surrogate server without cooperation. The performance of individual server is a key determinant about the amount of uploaded content.

In case of random surrogate server with cooperation, random distribution of requests creates a plateau in CDN utility. The authors also observed that in terms of mean response time, the closest surrogate server with cooperation outperformed the other two approaches, whereas in due to high network traffic, the response

¹ <https://www.datanyze.com/market-share/cdn/>

time for random surrogate server with cooperation was observed very poor. Random surrogate server does not consider the proximity criteria and consequently the response time is increased. However, the closest surrogate server without cooperation outperformed the other two approaches in terms of mean response time.

The CDN utility and CDN utilization seem to be similar concepts, but they do not address the same metrics. The CDN utility is more about upload to download ratio while the CDN utilization considers surrogate servers serving IoT requests over a defined period. In most of researchers concentrate on improving the CDN utility [25]. However, CDN resource (surrogate servers) utilization is ignored [19].

Few works consider CDN utilization based models to evaluate energy consumption in CDN surrogate servers [20,21]. The changing trend of large scale cloud and fog enabled distributed systems is leading us towards considering new metrics when analyzing the performance of such systems [22]. Several works have been done towards consideration of utility as a cost and pricing estimation model [23]. Keeping in view current and future importance of CDNs in Internet traffic distribution, efficient utilization of resources is one the most important parameters to consider [24].

2.1. CDN utility

Stamos et al. [18] analyzed the utility of the surrogate servers in the CDN and presented a metric called CDN utility. Essentially, CDN utility considers traffic activity to demonstrate performance of surrogate servers for data circulation in the CDN. Several methods have been proposed for the improvement of this utility. Several parameters, such as cache size, network topology, pattern distribution request, and the redirection policies have significant impact on the CDN utility [25]. It is also important to understand the relation between "the number of bytes of contents served (by surrogate servers) against the number of bytes of contents pulled (from other surrogate servers or from the origin server)". For a CDN surrogate server s , net utility γ_s is quantified as:

$$\gamma_s = \frac{2}{\pi} \times \arctan(\varepsilon) \quad (1)$$

If a surrogate server serves more content than is replicated by it, it is said to have high utility and vice versa. Therefore, the parameters ε can be visualized as follows:

$$\varepsilon = \frac{\text{bytesServed}}{\text{bytesReplicated}} \quad (2)$$

The metric considers a surrogate server practical if the server has high net utility. In other words, more content is uploaded by the surrogate server than is downloaded by it. The parameter ε is the ratio of the uploaded content (bytes) to the downloaded contents (bytes) and γ_s has value range [0 . . 1] from Eq. (1). A similar type of net utility metric is used by Mortazavi et al. [26] for a peer to peer system. Considering the N surrogate servers of a CDN, the CDN utility is defined as follows:

$$\gamma = \frac{\sum_{j=1}^N \gamma_s}{N} \quad (3)$$

2.2. CDN Utilization

In a CDN scenario, the IoT clients send requests for contents that are subsequently forwarded according to the redirection policy to the surrogate servers. In case a request regarding an object is received from the IoT client c_1 by the surrogate server s_1 , the resource is locked by the s_1 . The surrogate server s_1 checks the

cache for the demanded object and if the requested object is present in the cache, the contents are forwarded to c_1 and the resource is unlocked. Consequently, the request is accomplished in a faster way. In another scenario the contents required by c_1 are not present in the cache of s_1 , the object can be pulled from another surrogate server s_2 or from the origin server depending upon redirection policy of the CDN.

The activity (of a surrogate server) of pulling an object from another surrogate server is called cooperation. Cooperation among the surrogate servers causes in the locking and unlocking of a certain resource in another surrogate server (from which object is pulled). Once the requested object is received, it is stored by the surrogate server s_1 in its cache and is sent to the c_1 . When a surrogate server locks a resource, we say it is to be utilized. Surrogate server is considered in utilization until it releases (unlock) a resource (connection).

The amount of IoT requests a surrogate can serve at most, represents its overall capacity. Utilization of a surrogate server can, thus be modeled as ratio of current number of clients requests a surrogate server is serving to its capacity [27]. Utilization ratio (UtR) of a surrogate server s between time intervals $[t_1, t_2]$ is computed as:

$$UtR_{s[t_1, t_2]} = \frac{ConCurr_{s[t_1, t_2]}}{ConMax_s} \quad (4)$$

$ConCurr$ shows current number of client requests surrogate server s is serving. A capacity of surrogate server is presented as $ConMax$. So, utilization (Ut) of a surrogate server s between time intervals t_i and t_j ($t_j > t_i$) is computed as:

$$Ut_{s[t_i, t_j]} = \frac{\sum_{k=i}^{j-1} UtR_{s[t_k, t_{k+1}]}(t_{k+1} - t_k)}{t_j - t_i} \quad (5)$$

Hence, utilization Ut of a surrogate server s during a life time T of an experiment will be:

$$Ut_s = Ut_{s[0, T]} \quad (6)$$

3. Elements influencing the CDN utilization

The elements of privacy and security are very crucial in affecting the performance and utilization of the CDN based Internet of Things (IoT). Besides, there are certain other parameters that significantly affect the surrogate server utilization. These parameters are described in the following subsections.

3.1. CDN Security

The matters of CDN security, for instance, CloudBleed, demonstrate that edge systems make transactions among execution, privacy, confidentiality, and security among others. A CDN resembles a metal sphere for the Internet. Not considering caches at the edges of networks to accelerate the heap spells of images, movies, and other data, most of the processes may end without execution.

3.2. CDN Privacy

Uploading content to CDNs is a basic segment for any user having Internet connectivity. The failure of devices in hours where the Internet traffic is high, due to several reasons, e.g., malicious attacks or maximum cache hit, can possibly bring down servers and stop subscribers to access a webpage or any other application. Thus, a CDN should have good features for avoiding attacks and privacy issues.

3.3. Network link speed

The connection duration and mean response time are sufficiently dependent on the speed of the network link. With a network link of higher speed, request and contents are sent and received at a faster pace that eventually leads to the reduced connection duration, provided the other parameters are the same.

3.4. Content request frequency

The high frequency of requesting content also increases the server utilization because it leads the servers to establish and manage more connections in limited time. Moreover, a higher number of requests in small time interval also leads to an increase in congestion in the network that eventually results in the increased response time.

3.5. Object size

The size of an object is in direct relation i.e., directly proportional to the size of object. The request completion rate can be high for a website with smaller objects; however, connection duration is smaller in that case. On the other hand, for the websites with larger object sizes, the connection duration is longer that eventually results in increased time to complete a request.

3.6. Cache size

The cache size also plays a vital role in the utilization. The surrogate server stores the objects in the cache. Therefore, in case of a smaller cache, a small number of objects can be served and consequently the possibility of completion of requests decreases. If the cache of the surrogate server does not possess requested objects, the contents must be pulled either from the other surrogate servers or origin servers that ultimately augments the response time and the connection duration. However, with the larger cache size, the probability to serve more requests by the surrogate servers is increased.

3.7. Content popularity

The duration of the connection is also affected by the popularity of contents because the frequently requested contents are considered as the popular. Therefore, the surrogate server attempts to hold such contents in the cache. Moreover, as opposed to the unpopular contents, the popular contents are served in a smaller amount of time.

3.8. Client and server location

The locations of both the IoT clients and the surrogate servers are also important. The requests made by the clients that are close to the location of surrogate servers are responded more efficiently depending upon the availability of the contents. In such cases, the connections established to fulfill the requests are usually of short duration as compared to the ones where client and server are located far apart from each other.

4. Performance evaluation

To evaluate the proposed idea, we have opted for a test bed as well as Internet topology due to the lack of availability of a real CDN deployment for research purposes. This test bed consists of: (i) a system model that simulates the CDN infrastructure, (ii) network topology generator, (iii) a website with infotainment data for



Fig. 1. A Schematic view for Experimental Testbed: A Typical CDN enabled Smart City Infrastructure [13].

consumption by IoT devices, (iv) a client request stream generator configured to generate IoT data consumption requests.

4.1. Simulation setup

The CDNSim [28] provides discrete event simulation environment for CDN infrastructure. With the implementation in C++ programming language, the CDNSim is meant to simulate CDN infrastructure. OMNET++ is based on the OMNeT++ library² that offers a framework for performing discrete event simulations. Moreover, a detailed implementation of TCP/IP, packet switching, packet retransmission upon misses, and freshness are also provided by the CDNSim. Moreover, it is possible to add new policies in CDNSim.

4.1.1. System model

Most of CDN issues, such as selection of surrogate server, queuing, propagation, bottlenecks, and processing delays are addressed dynamically.

4.1.2. CDN infrastructure

The CDNSim provides the key CDN infrastructure and related functions of a CDN environment. We considered different amount of homogeneous surrogate servers (CDN infrastructure), i.e., 10, 20, 30, 40, 50 for this experiment. Each surrogate server is capable of handling 500 client requests simultaneously. Surrogate servers are supposed to be located in different areas of the world as shown in Fig. 1.

4.1.3. Cache size

Though literature suggests having larger cache size to make as few request re-directions towards the origin server or nearby surrogate servers as possible we have opted for a more conservative approach. Each of the surrogate servers has a cache size that can store 40% of the website contents. Surrogate server cache is updated using the standard Least Recently Used (LRU) cache replacement policy.

4.1.4. CDN topology

To evaluate the CDN infrastructure, we adopted IoT based network topology consisting of the following elements: Origin Server where the traffic is generated, homogeneous surrogate servers acting as IoT gateway nodes, routing components and IoT based web

² OMNeT++ is a modular, component-based C++ simulation library and framework. <https://omnetpp.org/>.

Table 1
Summary of Simulation Parameters.

Parameter	Value
Website size	1GB
Website number of objects	50000
Website z for size	1
Content Popularity vs. Size correlation	0
Client requests	200k, 400k ... 1000k
Mean interval time of requests	0.0033 s
Distribution of interval time	exponential
Requests stream z	1
Link speed	16 Mbps
Network topology type	AS
No. of Routers	3037
Surrogate servers	10, 20, 30, 40, 50
Incoming connections per surrogate server	500
Outgoing connections per surrogate server	500
Client groups	100
Content providers (origin server)	1
Cache size percentage (of the website's size)	40%
Policy parameter z value	0, 1

clients. We also considered a tracking server responsible for management of IoT requests and redirection towards origin or surrogate servers that have cached requested content based upon the request and cache status. Routing elements are responsible for connecting all the surrogate servers through backbone internet with clients. Different network organization schemes result in affecting the redirection policy. AS level real Internet topology [18] is used that consists of routing information gathered from 7 BGP peers dispersed at different locations as shown in Fig. 1.

The origin server, surrogate servers, and clients are randomly attached to a set of routers. The routers use TCP/IP protocol for packet forwarding and the link speed is set to 16Mbps. As reported by Akamai International the global average connection speed, in second quarter of 2014, for broadband Internet remained 4.6Mbps. For High Definition (HD) content streaming, required bandwidth is between 10 and 20 Mbps [29].

4.2. Web server content

A synthetic yet realistic website of 50 k objects of 1 GB size is created. Zipfian distribution is used to consider size of objects

[30]. Table 1 summarizes parameters and Fig. 1 shows schematic views of simulation testbed. The distribution scope is affected by the values of z such that when z increases, the distribution slope becomes steeper. The value z = 0 represents that the sizes of all the website objects are identical. We consider z = 1, where the size of object fades exponentially.

4.3. Web traffic requests

To generate requests, a custom modified IoT Web traffic generator is used that demonstrates access patterns very close to a realistic wireless sensor network. It takes a sensor readings range and web transactions are generated as the sequences of page traversals on the graph. Following the standard way to forward the requests based on proximity, the requests are forwarded to the closest surrogate server. We evaluated our results with a varying number of client requests (thousands), i.e., 200, 400, 600, 800, 1000. Moreover, we use different mean inter arrival time of client requests (i.e. 0.001, 0.005, 0.0033, 0.0025, 0.002, 0.00125) in order to change the intensity of overall load as shown in Fig. 2.

5. Results and discussion

In this section, we present results comparing the CDN utility and CDN utilization. Simulation parameters used for these experiments are described in Table 1. We used 10 different seeds to check the validity of our evaluation results. Our main objective is to compare the CDN utility and the CDN utilization under different CDN scenarios. The intent is to demonstrate the ultimate utilization of CDN resources, if we consider CDN utility as a sole metric to consider CDN performance.

All the results are presented with average values and standard deviation. We considered load balance and load unbalance CDN redirection policies. Load balance policies balance load of client requests among all the available surrogate servers. Load unbalance policy will redirect the majority of client requests to a set of surrogate servers. Fig. 3 shows utilization behavior of 50 surrogate servers under load balance and load unbalance policies. It shows how load balance and load unbalance policies work. We have performed experiments for a varying number of surrogate servers serving a different amount of client requests. Here we have shown some selected graphs.

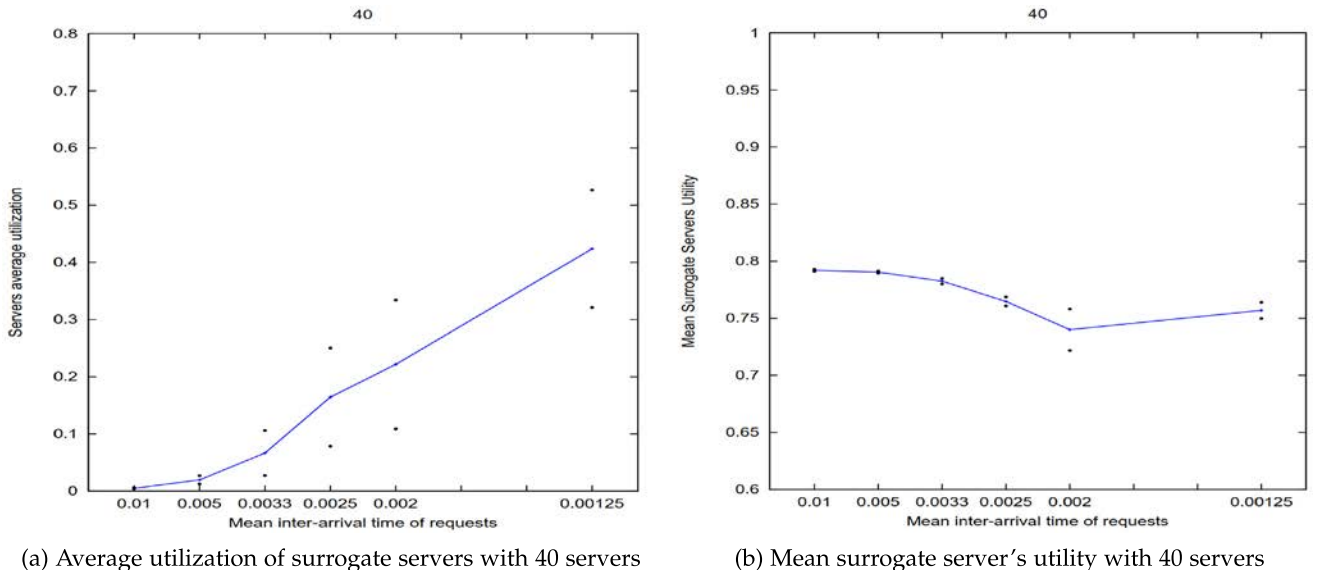
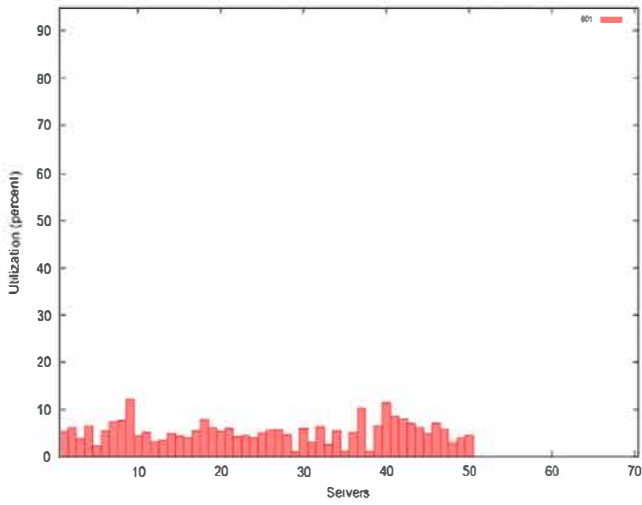
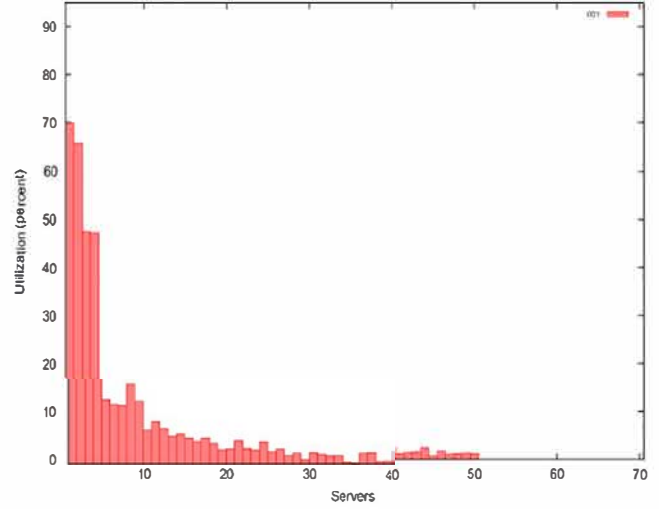


Fig. 2. Mean surrogate server's utility vs. Average utilization of surrogate servers, using Load-balance CDN redirection policy. With different mean inter-arrival time of IoT requests.

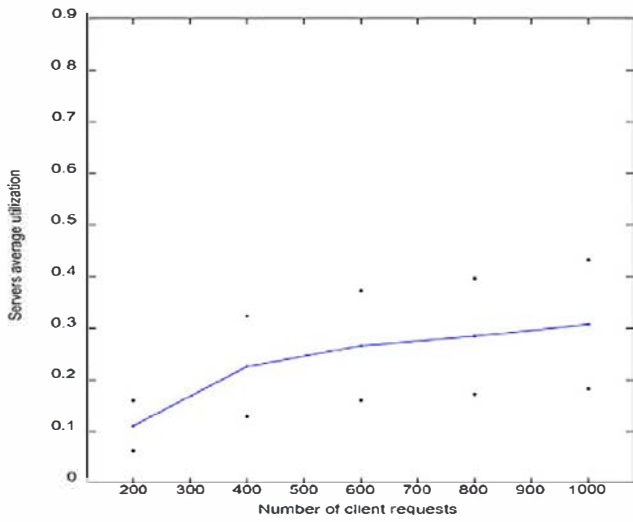


(a) Load-balance

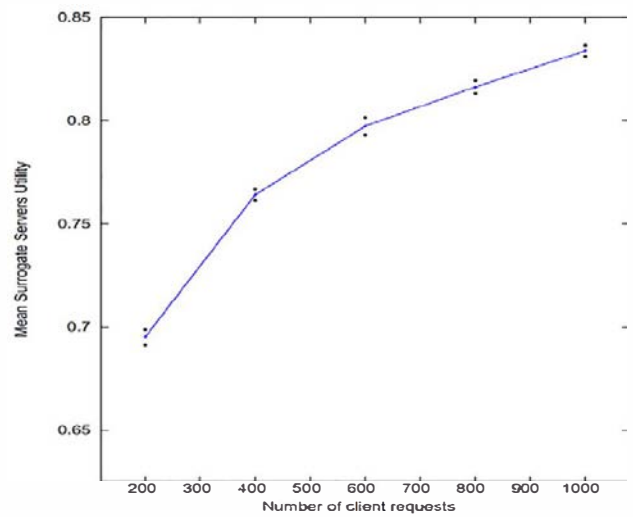


(b) Load-Unbalance

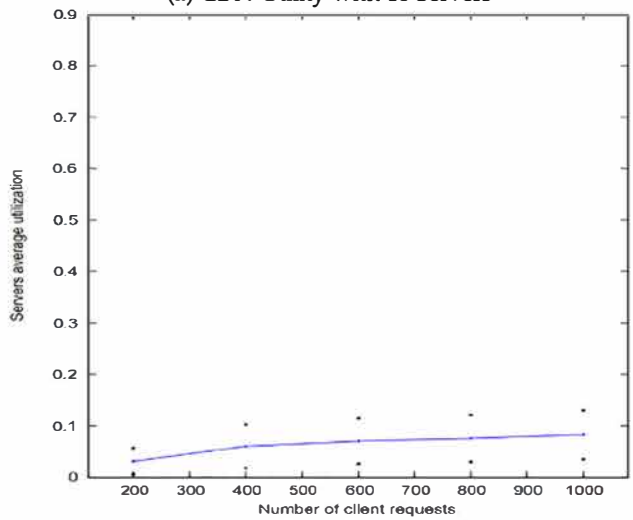
Fig. 3. Load-balance vs. Load-unbalance CDN redirection policies. 50 surrogate servers serving 1000 k requests.



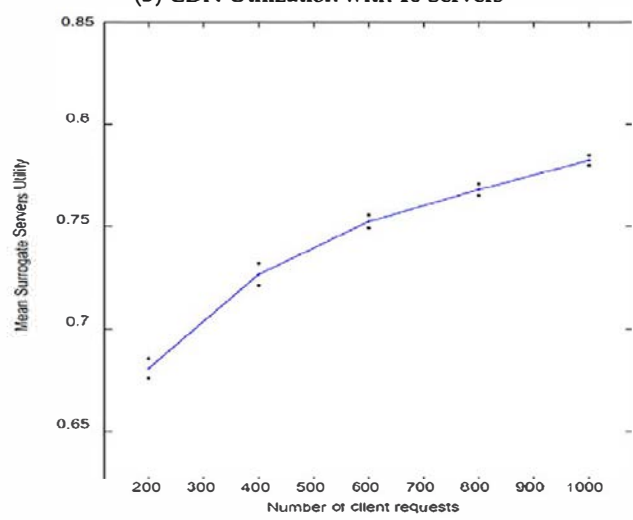
(a) CDN Utility with 10 Servers



(b) CDN Utilization with 10 Servers



(c) CDN Utility with 40 Servers



(d) CDN Utilization with 40 Servers

Fig. 4. Mean surrogate server's utility vs. Average utilization of surrogate servers, using Load-balance CDN redirection policy. Representing 10 and 40 surrogate servers serving varying number of client requests.

We evaluated our results with higher and lower loads to observe the behavior of mean surrogate server utility and utilization. Fig. 4 compares the CDN utility and CDN utilization with 10 and 40 surrogate servers taking various numbers of requests from clients. It shows that the CDN utility and the CDN utilization both increase with the increasing number of requests. With more CDN requests, surrogate server caches become more intelligent and keeps the popular contents. We can say that the surrogate server caches have become smarter and consequently impacts the performance of surrogate servers because it uploads more contents than it downloads. Therefore, we observed an increase in the CDN utility with the increase in the number of serving client requests. Similarly, when the number of requests increases, there is a subsequent increase in the number of connections at surrogate servers.

Augmentation in the number of connections at surrogate servers increases the utilization as shown in Eq. (4). However, our purpose is to compare the CDN utility and the CDN utilization. Fig. 4 shows higher values of CDN utility (69% to 83%), which is an indicator of decent performance of the CDN. However, it shows that the surrogate servers have lower utilization (10% to 30% average utilization) for different number of client requests. If we only

consider the CDN utility to evaluate the CDN performance, then we have inefficient CDN resources utilization. Same is the case in load unbalance policy as well, as shown in Fig. 5. Though load unbalance policy presents an improvement in the CDN utility (67% to 81%) and the CDN utilization (20% to 60%), the global behavior while comparing both of the metrics (CDN utility and CDN utilization) remains same as in load balance.

Results also exhibit that even surrogate servers have higher loads (up to 60%), there is still a considerable difference between the values of CDN utility and CDN utilization. It shows that the CDN utility does not exhibit CDN resource utilization. To demonstrate the behavior of the CDN utility and the CDN utilization with a smaller number of requests, we performed experiments with different number of surrogate servers serving 200 k requests as shown in Fig. 5. It shows that both policies exhibit same kind of behavior while comparing CDN utility and CDN utilization.

The CDN utility increases. Each surrogate server serves a smaller number of requests as there is increase in the number of surrogate servers. Serving lesser requests does not permit caches to be more intelligent to keep popular contents. Therefore, the surrogate server is compelled to extract the contents from the neighboring surrogate

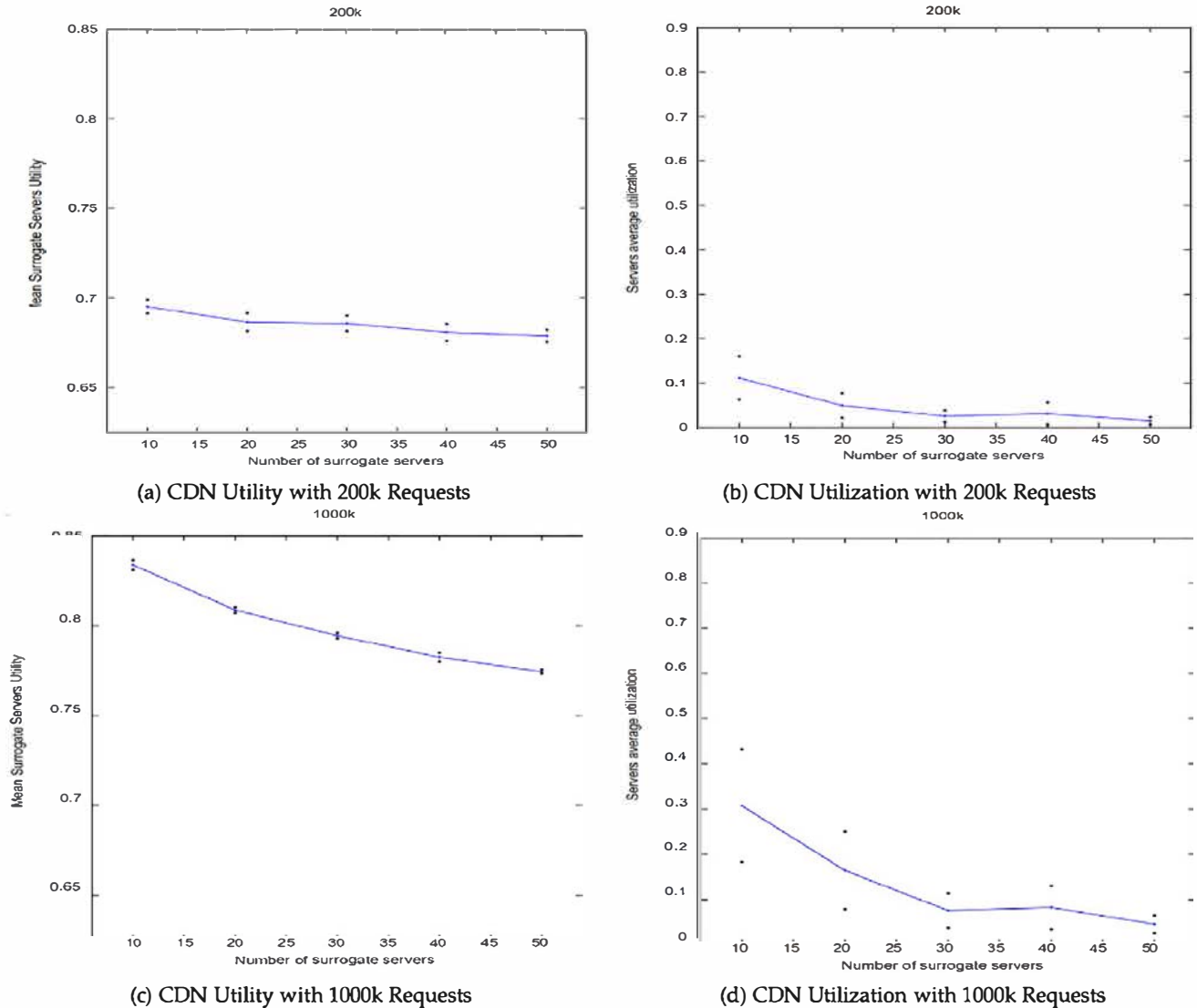


Fig. 5. Mean surrogate server's utility vs. Average utilization of surrogate servers, using Load-Unbalance CDN redirection policy. Representing 50 surrogate servers serving varying number of client requests.

servers. As a result, surrogate server pulls more contents than it serves. Therefore, its CDN utility is decreased.

In case of CDN utilization, same number of client requests for more servers' results in less connections. Convincingly, there is a decrease in CDN utilization with an increase in number of surrogate servers. The higher values of the CDN utility are indication of the higher performance. While, Considering CDN utilization, CDN resources are critically utilized. Surrogate servers exhibit very low utilization values e.g. in case of load balance CDN utilization is from 11% to 1%. If we consider only CDN utilization for CDN performance (as did in [26,18] and in [25]) then CDN is performing quite nicely. However, when we consider CDN utilization, it shows the opposite picture of story. Hence, CDN utility is not enough to evaluate CDN performance. We must consider how CDN resources are efficiently utilized.

6. Conclusion

Energy management and efficient resource utilization are key challenges in large scale implementations of Internet of Things (IoT) based deployments systems. Since the conception of IoT, demand for data is increasing with the increase in IoT devices being connected to the internet every day. The CDNs play a vital role for content delivery worldwide, especially when we consider millions of devices being connected to the web. Efficient resource utilization in CDN enabled IoT is crucial to minimize energy costs and reducing CO₂ emission. Energy aware client redirection policies may help to increase energy savings in the CDNs. Though the CDN utility provides useful information regarding performance but it does not reflect surrogate server's utilization. The CDN utilization is a metric that exactly addresses how the CDN resources are being utilized.

To evaluate the performance of CDN, it is important to consider surrogate servers' utilization. Experimental results of the paper demonstrate that there are opportunities to reduce energy consumption by changing the CDN redirection behavior based upon IoT requests. Moreover, we analyzed the CDN utility and CDN utilization under different scenarios and demonstrated that the CDN utility does not actually address the utilization of CDN surrogate servers. Moreover, the results from the study also dictate that the CDN utilization should be considered as significant metric to evaluate the overall performance of the CDN. Furthermore, infrastructure size and volume of the IoT request traffic have also been considered. Even if the surrogate servers are poorly utilized, the CDN utility exhibits higher values. Conclusively, evaluating the CDN performance only based on CDN utility without considering utilization of surrogate servers is an inefficient approach. Therefore, the authors propose to consider the utilization of CDN resources as an important metric while analyzing the performance of a CDN.

In the future, we intend to implement utilization aware net utility model for IoT based CDN infrastructures that will enable us to analyze the CDN utility in a more realistic fashion.

Acknowledgment

The authors are grateful to the Deanship of Scientific Research, King Saud University for funding through Vice Deanship of Scientific Research Chairs.

References

- [1] K. Mokhtarian, H.-A. Jacobsen, Flexible caching algorithms for video content distribution networks, *IEEE/ACM Trans. Networking (TON)* 25 (2) (2017) 1062–1075.
- [2] A.H. Alavi, P. Jiao, W.G. Buttler, N. Lajnef, Internet of things-enabled smart cities: State-of-the-art and future trends, *Measurement* 129 (2018) 589–606, Available: <http://www.sciencedirect.com/science/article/pii/S0263224118306912>.
- [3] T. Addabbo, A. Fort, M. Mugnaini, E. Panzardi, A. Pozzebon, V. Vignoli, A city-scale iot architecture for monumental structures monitoring, *Measurement* 131 (2019) 349–357.
- [4] H.A. Khattak, H. Farman, B. Jan, I.U. Din, Toward integrating vehicular clouds with iot for smart city services, *IEEE Network* 33 (2) (2019) 65–71.
- [5] A.J. Kadhim, S.A.H. Seno, Maximizing the utilization of fog computing in internet of vehicle using sdn, *IEEE Commun. Lett.* 23 (1) (2019) 140–143.
- [6] M.S. Mekala, P. Viswanathan, Clay-mist: Iot-cloud enabled cmm index for smart agriculture monitoring system, *Measurement* 134 (2019) 236–244.
- [7] J. Iqbal, M. Khan, M. Talha, H. Farman, B. Jan, A. Muhammad, H.A. Khattak, A generic internet of things architecture for controlling electrical energy consumption in smart homes, *Sustainable Cities Soc.* (2018).
- [8] E. Siow, T. Tiropanis, W. Hall, Analytics for the internet of things: a survey, *ACM Comput. Surveys (CSUR)* 51 (4) (2018) 74.
- [9] J. Liu, Q. Yang, G. Simon, Congestion avoidance and load balancing in content placement and request redirection for mobile cdn, *IEEE/ACM Trans. Networking* 26 (2) (2018) 851–863.
- [10] Y. Wang, K. Wang, H. Huang, T. Miyazaki, S. Guo, Traffic and computation offloading with reinforcement learning in fog computing for industrial applications, *IEEE Trans. Industr. Inf.* 15 (2) (2019) 976–986.
- [11] C.V.N. Index, Global mobile data traffic forecast update, 2015–2020 white paper, 2016.
- [12] S. Wistow, Assessing content delivery network target content consumption, Mar. 30 2017, uS Patent App. 15/278,796.
- [13] H.A. Khattak, Z. Ameer, U.I. Din, M.K. Khan, Cross-layer design and optimization techniques in wireless multimedia sensor networks for smart cities, *Comput. Sci. Inform. Syst.* 16 (1) (2019) 84–101.
- [14] G. Neglia, D. Carra, P. Michiardi, Cache policies for linear utility maximization, *IEEE/ACM Trans. Networking (TON)* 26 (1) (2018) 302–313.
- [15] M. Gupta, A. Garg, A perusal of replication in content delivery network, *Next-Generation Networks*, Springer, 2018, pp. 341–349.
- [16] T. Bektaş, I. Ouveysi, Mathematical models for resource management and allocation in cdns, *Content Delivery Networks*, Springer, 2008, pp. 225–250.
- [17] T. Bektaş, J.-F. Cordeau, E. Erkut, G. Laporte, Exact algorithms for the joint object placement and request routing problem in content distribution networks, *Comput. Operations Res.* 35 (12) (2008) 3860–3884.
- [18] K. Stamos, G. Pallis, A. Vakali, and M.D. Dikaiakos, "Evaluating the utility of content delivery networks, in: Proceedings of the 4th edition of the UPGRADE-CN workshop on Use of P2P, GRID and agents for the development of content networks, ser. UPGRADE-CN '09. New York, NY, USA: ACM, 2009, pp. 11–20. [Online]. Available: <http://doi.acm.org/10.1145/1552486.1552509>.
- [19] S. ul Islam, K. Stamos, J.-M. Pierson, A. Vakali, Utilization-aware redirection policy in cdn: a case for energy conservation, in: *International Conference on Information and Communication on Technology*, Springer, 2011, pp. 180–187.
- [20] S. ul Islam, J.-M. Pierson, Evaluating energy consumption in cdn servers, *ICT-GLOW*, ser. ICT-GLOW'12, Springer-Verlag, Springer, Berlin, Heidelberg, 2012, pp. 64–78, https://doi.org/10.1007/978-3-642-32606-6_6.
- [21] V. Mathew, R.K. Sitaraman, P. Shenoy, Energy-efficient content delivery networks using cluster shutdown, *Sustainable Comput.: Inform. Syst.* 6 (2015) 58–68.
- [22] S.U. Islam, N. Javaid, J.-M. Pierson, A novel utilisation-aware energy consumption model for content distribution networks, *Int. J. Web Grid Serv.* 13 (3) (2017) 290–314.
- [23] H. Arshad, H.A. Khattak, Z. Ameer, A. Abbas, S.U. Khan, Estimation of fog utility pricing: a bio-inspired optimisation techniques' perspective, *Int. J. Parallel, Emergent Distributed Syst.* (2019) 1–14.
- [24] J. Bruneau-Queyreix, Multi-criteria optimization of content delivery within the future media internet, Ph.D. dissertation, Bordeaux, 2017.
- [25] G. Pallis, Improving Content Delivery by Exploiting the Utility of CDN Servers, in: A. Hameurlain, F. Hussain, F. Morvan, A. Tjoa (Eds.), *Data Management in Cloud, Grid and P2P Systems*, ser. Lecture Notes in Computer Science, 7450, Springer, Berlin, Heidelberg, 2012, pp. 88–99.
- [26] B. Mortazavi, G. Kesidis, Model and simulation study of a peer-to-peer game with a reputation-based incentive mechanism, *Information Theory and Applications Workshop*, 2006.
- [27] S. Ul Islam, J.M. Pierson, Evaluating energy consumption in CDN servers, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, ser. ICT-GLOW'12, vol. 7453, LNCS, Berlin, Heidelberg: Springer-Verlag, 2012, pp. 64–78, https://doi.org/10.1007/978-3-642-32606-6_6.
- [28] K. Stamos, G. Pallis, A. Vakali, D. Katsaros, A. Sidiropoulos, Y. Manolopoulos, Cdnsim: A simulation tool for content distribution networks, *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 20 (2) (2010) 10, <https://doi.org/10.1145/1734222.1734226>.
- [29] Akamai, "Akamai's state of the internet", Akamai's Creative & Brand Development, Tech. Rep., Mar 2017. [Online]. Available: <http://www.akamai.com/stateoftheinternet/>.
- [30] V.N. Padmanabhan, L. Qiu, The content and access dynamics of a busy web site: findings and implications, *SIGCOMM Comput. Commun. Rev.* 30 (4) (2000), <https://doi.org/10.1145/347057.347413>.

Saif ul Islam received his Ph.D. in Computer Science at the University Toulouse III Paul Sabatier, France in 2015. He is an Assistant Professor at the Department of Computer Science, Dr. A. Q. Khan Institute of Computer Science and Information Technology, Rawalpindi, Pakistan. Previously, he served as Assistant Professor for three years at the COMSATS University, Islamabad, Pakistan. He has been part of the European Union-funded research projects during his Ph.D. His research interests include resource and energy management in large-scale distributed systems (Edge/Fog, Cloud, Content Distribution Network (CDN)) and the Internet of Things (IoT).

Hasan Ali Khattak received his PhD in Electrical and Computer Engineering degree from Politecnico di Bari, Bari, Italy in April 2015, Master's degree in Information Engineering from Politecnico di Torino, Torino, Italy, in 2011, and B.CS. degree in Computer Science from University of Peshawar, Peshawar, Pakistan in 2006. He is currently serving as Assistant Professor of Computer Science at COMSATS University Islamabad since January 2016. His current research interests focus on Web of Things, Data Sciences, Social Engineering for Future Smart Cities. His perspective research areas are application of Machine Learning and Data Sciences for improving and enhancing Quality of life in Smart Urban Spaces through predictive analysis and visualization. He is a Senior Member IEEE, Professional Member ACM and an active member of IEEE ComSoc, IEEE VTS and Internet Society.

Jean-Marc Pierson received the Ph.D. degree in computer science from the LIP Laboratory, École Normale Supérieure de Lyon. He has been a Full Professor with University of Toulouse, since 2006. His research interesting includes distributed systems and high-performance computing, distributed data management, security grid, cloud computing, and energy-aware distributed computing.

Ikram Ud Din received his PhD degree in Computer Science from the School of Computing, Universiti Utara Malaysia (UUM) in 2016, MSc in Computer Science and MS in Computer Networking from the Department of Computer Science, University of Peshawar, Pakistan. He also served as IEEE UUM Student Branch Professional Chair. Ikram has 10 years of teaching and research experience in different universities/organizations. His current research interests include traffic measurement and analysis for monitoring quality of service, mobility and cache management in Information-Centric Networking, and Internet of Things.

Ahmad Almogren holds Ph.D. degree in computer science from Southern Methodist University, Dallas, TX, USA, in 2002. Previously, he was an assistant professor of computer science and a member of the scientific council, Riyadh college of technology. He also served as the Dean of the college of computer and information sciences and the head for the council of academic, Al Yamamah university. He is currently a Professor and the vice dean for the development and quality with the college of computer and information sciences, King Saud university. His research areas of interest include mobile and pervasive computing, cyber security and computer networks. He has served as a guest editor at several computer journals.

Mohsen Guizani received the B.S. (with distinction) and M.S. degrees in electrical engineering, the M.S. and Ph.D. degrees in computer engineering from Syracuse University, Syracuse, NY, USA, in 1984, 1986, 1987, and 1990, respectively. He is currently a Professor at the CSE Department in Qatar University, Qatar. Previously, he served as the Associate Vice President of Graduate Studies, Qatar University, University of Idaho, Western Michigan University, and University of West Florida. His research interests include wireless communications and mobile computing, computer networks, mobile cloud computing, security, and smart grid. He is currently the Editor-in-Chief of the IEEE Network Magazine, serves on the editorial boards of several international technical journals and the Founder and the Editor-in-Chief of Wireless Communications and Mobile Computing journal (Wiley). He is the author of nine books and more than 500 publications in refereed journals and conferences. He guest edited a number of special issues in IEEE journals and magazines. He received the 2017 IEEE Communications Society Recognition Award for his contribution to outstanding research in Wireless Communications. He served as the IEEE Computer Society Distinguished Speaker from 2003 to 2005.

Mansour Zuair is currently an Associate Professor in the Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He received his M.S. & Ph.D. degree in Computer Engineering from Syracuse University, his B.S. degree in Computer Engineering from King Saud University. He served as CEN chairman from 2003 to 2006, vice dean 2009–2015 and dean 2016–now. His research interest is in the areas of computer architecture, Computer Networks and Signal Processing.