POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

ISSN: 1302-0900 (PRINT), ISSN: 2147-9429 (ONLINE)
URL: http://dergipark.gov.tr/politeknik

fCDN:A novel-energy efficient content delivery architecture over next generation systems

Yazar(lar)(Author(s)): *Kasim OZTOPRAK*

ORCID: 0000-0003-2483-8070

Bu makaleye şu şekilde atıfta bulunabilirsiniz(To cite to this article): Oztoprak K, "fCDN:A novel-energy efficient content delivery architecture over next generation", ***Politeknik Dergisi***, 21(4): 999-1006, (2018).

Erişim linki (To link to this article): http://dergipark.gov.tr/politeknik/archive

DOI: 10.2339/politeknik.470675

fCDN:A Novel-Energy Efficient Content Delivery Architecture over Next Generation Systems

Araştırma Makalesi / Research Article

Kasim OZTOPRAK*
Engineering Faculty, Computer Engineering Department, KTO Karatay University, Turkey
(Geliş/Received : 15.10.2017 ; Kabul/Accepted : 23.02.2018)

ABSTRACT

Content Delivery in mobile networks are gaining popularity with an increasing demand by new mobile subscriptions. Reduction of the outgoing traffic is becoming an important issue, which is proposed to be reduced by a futuristic Content Delivery System (fCDN) for Mobile Operators. The proposed system utilizes the idea Device-to-Device Communication (D2D), with a dynamic resource allocation support of SDN-NFV based systems and the locating of a content. The approach is to use the content of a node as close as possible to a mobile user, which may reduce the cost of communication, and power consumption while increasing the Quality of Experience. While the energy consumption can be reduced by proposed model upto 8 fold according to the current state of the art, enabling the users to use 5G based systems, it reduces the energy consumption by a factor of 50 fold in the proposed solution. The simulation results demonstrate that the use of proposed method outperforms the current state of the art systems in power consumption.

Keywords: Energy consumption, mobile networks, content delivery networks, 5G.

1. INTRODUCTION

By the tremendous demand increase of multimedia services, mobile networks has an overloaded usage for years. By far, these demands could not be met with the current architecture and the capacity of the mobile networks [1]. Newly designed architectures are designed to cope with this demand, which utilizes all-in-one design for the components and separates control and data planes.

Mobile content distribution network (mobile-CDN) studies are limited when compared to CDN studies [2]. A mobile-CDN is proposed by Yousaf et al. with the assumption that future mobile networks will have multiple gateways to the Web and 70\% of the internet traffic would be short videos in 2016 and using the TCP protocol as the service layer.

In [3], it is indicated that the increasing demand for higher bandwidth cannot be satisfied using typical content delivery systems. The existing problems are proposed to be solved mostly in similar with [4]. For instance, Oztoprak et al. proposed a P2P-based communication for a scalable multimedia delivery systems. [4] was adapted in mobile communication networks by using device-to-device (D2D) instead of P2P. All available frequency spectrum were used by researchers in D2D [5,6], whereas none of these studies contains an orchestrator framework to control/coordinate the user device communication for content delivery.

In existing literature, the proposals either do not provide feasible solutions or come up with incomplete solutions for mobile devices and content management issues. Furthermore, they do not offer a way of guaranteeing the validity of the resources obtained from mobile devices. They are far from addressing the source of motivation in order to share the content they already have.

In previous study [7], a content delivery system for mobile operators (mCSDN) was proposed. The system employed D2D to achieve the maximum capacity of the physical environment, while bringing a novel solution to manage the communication infrastructure as well as mobile devices utilizing software-defined networks (SDN). It utilizes the mobile video delivery orchestrator effectively: (i) to maximize the throughput; (ii) to maximize the overlay network capacity; and (iii) to provide higher quality streams effectively. In the study, a video streaming system is designed to use multiple description coding (MDC) in video delivery. In this setting, the delivery of the video is guaranteed, by sending a single description through the content cache located in the cellular node (eNB) of the mobile network or CDN server either within the operator or CDN providers. In order to increase the quality of experience with higher throughput of the content delivery network, additional streams are collected through the D2D overlay network. Guaranteed high availability, higher quality of experience, and less energy consumption has been demonstrated in a limited scope.

In this study, a novel Future Content Delivery Network (fCDN) is proposed. The power consumption analysis given in [7] is also analyzed and elaborated. In contrast to most studies in the literature, the power usage is calculated quantitatively rather than measuring the battery usage for an interval. The method of video coding assumed is Scalable Video Coding (SVC) (rather than MDC) in accordance with the massive usage of HTTP based streaming. In addition to the improvements, a new method is proposed in which the CPU resources of the mobile devices are used for trans-coding of the video into different bitrates, which also means to have data center

*Sorumlu Yazar (Corresponding Author)
e-posta : kasim.oztoprak@karatay.edu.tr

node in the mobile devices specifically processing the video as part of the CDN.

In the following Section II, related works about mobile content delivery systems, D2D systems, software defined networks, and power consumption are briefly surveyed. In Section III, the proposed fCDN is explained. The experimentation and the results are listed in Section IV and concluded in Section V.

2. RELATED WORK

Packet data network gateway (P-GW) is the bottleneck within the mobile networks, since it is a bridge between Internet and the mobile network. According to [2], high speed mobile device radio access is provided by radio base stations (eNB). The main problem of this system is the P-GW use as a central Internet access point. In order to solve the bottleneck issue, some studies proposed to distribute the Internet access. This may provide a shorter path for the users to access the outer network. Another critical issue in mobile networks is the unavoidable traversal of the P-GW to access the resources by using the Internet Protocol (IP). mSCP was proposed by Yousaf et al. [2], which concentrates more on TCP/HTTP traffic. Mainly, two new ideas were proposed: (1) in order to solve the trust issue, CDN data content is to be controlled by the same authority. (2) The infrastructure can be designed to include CDN as an integrated part. CDN infrastructure is proposed to have its edge nodes dispersed accross every eNB together in coordination with the central CDN servers. A similar proposal was made by Wang et al. [1] as "cooperative cell caching", where the popular content is cached in base stations.

QoE-based transport optimized Mobile CDN platform was proposed by Amram et al. [8], which pushes the popular content to peering/and or sibling caches proactively. Additionally, in order to reduce the mobile network load, an exchange mechanism between caches is proposed as a peer-to-peer model together with a network aware redirection of user requests dynamically to the appropriate cache in order. Similarly, [9,10,11] proposed systems for the mobile users to share the existing content with others in the network. Additionally, [11] emphasized on the energy consumption in D2D communication for increasing the mobile device lifetimes.

D2D communication studies mostly concentrate on the device contributions in a CDN to improve throughput, to increase the energy and spectral efficiency, delays and fairness. [5,6] surveys the D2D studies under two main categories: (1) mobile operator inband sharing, (2) outband sharing. Whereas the D2D and the cellular communications in inband channel are on the same cellular channels, outband usage is to utilize all channels such as Wi-Fi, Bluetooth etc.

According to [13,14], content distribution and multimedia transmission in cellular networks is commonly performed by D2D communication. [13] proposes a scheme to keep track of locations for the users. Hence, a nearby node can be identified and data can be transferred from them directly. The weakness of the method is the higher power consumption and control overhead while tracking the location of the nodes. Similarly, [14] proposes a method to reuse the content located in the mobile devices by utilizing D2D communication. It was assumed to use extreme capacity formed by mobile infrastructure and user devices without any consideration related to energy consumption. Although, several studies [15] try to solve sustainability problem of stream by utilizing multiple peer or device, the current state of the art does not solve the reputation management problem. In other words, there is no mechanism guaranteeing the quality and correctness of the data obtained from the other devices/peers.

Caching is a proven method to reduce the mobile data traffic by two thirds [1]. [1] proposes a content caching and delivery method for 5G networks. It is mentioned that the Internet will be using a content-centric networking (CCN). By using the CCN architecture, every network node stores data, data is transported as chunks and with a receiver oriented transfer.

The literature surveyed so far indicates that mobile network nodes and devices can be used to cache the data and forward in order to minimize the energy consumption and maximize the throughput. Mainly the missing part for most studies is an orchestrator to control and coordinate the CDN. It is further commented that SDN is a new concentration to solve the problem by separating the data and control planes.

As highlighted in [3], centralized data plane functionality may create a scalability problem in LTE. It is further suggested to be using SDN to overcome the scalability issue. However, SDN adaptation might be a challenging issue because of the radio resources management, resource scarcity addressing , and the user mobility. In addition, it is highlighted that the SDN is being used to solve network configuration management problems. A method is proposed to improve the efficiency of the content management by using the SDN based solutions in a Wireless Mesh Network (WMN) in [16]. SDN is used to provide a content caching identifier, performing the location decision in the cache, as well as cache distribution. The authors pointed out that the proposal was very effective when the users are locally converged. Similar solutions are attracting both operators and network vendors around the world by the motivation of having the ability to automate resource allocation anywhere in the system dynamically in addition to flexibility and agility. The idea behind the SDN philosophy can be summarized as; i) using white label hardware to get rid of vendor lock-in as much as possible, ii) allocating necessary resources through Network Function Virtualization (NFV) whenever requested, and iii) having a centralized automation/orchestration mechanism to control and coordinate all assets and configuration of the network.In contrast to having many studies on this area, there is still a need for a framework to control, orchestrate and automate all mobile devices as a part of the network infrastructure and as a content delivery platform, especially in the coming era of 5G.

3. FUTURE CONTENT DELIVERY NETWORKS - fCDN

3.1 . Motivation Behind the Study

The energy consumed for mobile communication is one of the main interest area of the researchers among decades. The main concern in those studies are to optimize the total power consumed per bits transmitted. [17] starts from the basics of wireless communication and derives the total energy consumed for different type of communication. Similarly, [18,19] worked on power optimization in mesh of wireless sensor networks. It is clear that the total energy consumed is the sum of i) the energy consumed during the transmission, ii) energy dissipation during the reception, and iii) the energy spent for the amplification in the sender side in order to compensate the propagation distance. [18] calculates the total energy in order to transmit k-bit message below;

$$E_T x(k,d) = E_T x - elec(k) + E_T x - amp(k,d)$$
$$E_T x(k,d) = E_{elec} * k + \varepsilon_{amp} * k * d^2 \quad (1)$$

and to receive the above message, the total energy spent becomes;

$$E_R x(k) = E_R x - elec(k)$$
$$E_R x(k) = E_{elec} * k \quad (2)$$

It is obvious from Equations (1) and (2) that, the energy consumed to transfer a bit is directly related with the distance between communicating parties. Although, [17] indicates that the order of the distance is in between 2.7 and 3.5, [18,19] use it as 2. In this study, it will be used as 2 as depicted from Equations (1) and (2).

In addition to the above, Fodor et al. [12] proposed to use the D2D communication in a cellular network. The results of the experiments conducted shows that knowing the location of a mobile device can reduce the energy consumption of mobile devices by 25 fold when communicating with the ones geographically very close. The results are very similar to the studies in [18,19].

In most of the content delivery networks, DASH (Dynamic Adaptive Streaming over HTTP) is becoming the de-facto standard method for streaming media files over HTTP [20]. DASH is based on adaptive streaming in order to improve the Quality of Experience (QoE) of the users by adjusting the bit-rate of the streams. Each stream is divided into small Group of Pictures (GoP) (of 2-10 seconds each). Those group of segments were coded in different resolution and bit-rate by using Scalable Video Coding (SVC). According to the experience obtained from the network the most appropriate chunk is delivered to the viewer. Studies using the resources from a device is becoming more popular in video streaming while the rate of the streaming remains same. On the other hand, the rapid evolution in edge computing because of the increasing capacity of mobile devices enables us to utilize video trans-coding in order to serve to the receiver the video content in different speeds. This approach brings trans-coding operation from the cloud to the edge.

The features of the proposed method are defined in the following subsection.

3.2 . Features of the Proposed Solution

The proposed fCDN is composed of the following components: (i) a CDN server farm located in the data-center or the core of the mobile operator as the main source of the content seed. This is valid only when the operator also acts as a content provider. Otherwise, the origin CDN server stays outside of the operator; (ii) content caching application located in the cellular base station in order to perform network assisted caching in order to keep to content as close as possible to the mobile users; (iii) the mobile devices (users) with the capability of D2D communication preferably with multiple radio network interfaces (i.e. Multiple LTE Channels, Bluetooth and Wi-Fi); and iv) the management software of the proposed energy aware content delivery platform is designed to optimize the usage of the capacity of the system, energy consumption as well as tracking the locations of the mobile devices (users) across the system. The system promotes the use of the resources of the nearby nodes in order to get better QoE as well as the least energy consumption. The reliance to the content shared among 5G system have a critical role in the effectiveness of such systems, otherwise, the gain in energy consumption and faster access is lost with improperly advertised content.

3.3 . Proposed Method

The proposed model starts with building a discovery mechanism for the content replicated throughout the mobile devices which is identical to the served content by the CDN servers. The overlay management software responsible from the content discovery operation performs indexing all the content and registering available CPU cycles into its resource pool either in the CDN server or in the mobile devices connected to the system.

The system is assumed to allow users to acccess and share video contents encoded by using SVC [20] with four different streaming capacity as follows: 100 Kbps, 200 Kbps, 400 Kbps and 800 Kbps. All video contents are assumed to exist in a CDN server either in the data-centers of the Mobile operator or through a CDN providers outside of the operator.

In order to reduce the latency and optimize the uplink capacity, caching application is onboarded to the virtual eNB node in the mobile base stations with the functionality of caching the most frequently used content in its cell area. Although the idea of caching content for a cellular area in base stations is offered by several studies [5,6] previously, the massive usage of this approach will be used after having the caches installed in eNBs in a virtual network function. The aim of this approach is bringing the intelligence of the systems as close to the edge as possible which is consistent with the telecommunication trends for 5G era. The proposed caching application is located in the cellular base station, which is the closest point to the users in the mobile network.

The proposed method aims to minimize the power consumption as well as guaranteeing the service quality above some levels. In addition to the lower power consumption,the propagation delay is less in a communication with closer nodes.

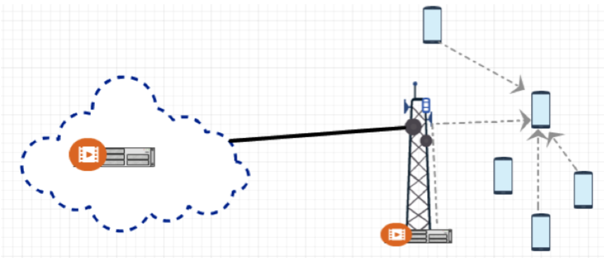

Figure 1. A Mobile system with CDN server, content cache in eNB as well as mobile devices

The proposed system depicted in (Figure 1) has two main resource pool. Although a content cache and a CDN server is located in the system, the main motivation is to utilize the content located in the neighbor devices. The system benefits from the devices three folds. Those are faster access to data, lower power consumption for the access, and more scalable streaming capacity with many devices in the system. The proposed method is summarized in Algorithm (1):

1. **Input:** ***S*** (set of nodes), ***T*** (set of content), ***N*** (set of nodes returned with searched content), *Interest* (***S*** → ***T***).
2. A node X queries for a content ***t*** ε ***T***.
3. **Output:** System returns all ***n*** ε ***N***, n_t ← (the total number of devices having the searched content and CPU resources)
4. **while** (***N*** satisfies selected quality) **do**
   a. Select node n_i ← closest to ***X*** with the searched GoP d_i in specified rate which will minimize the use of power p_i and increases the throughput
   b. If a node has higher rate for a GoP (which is unavailable in others) and having CPU resources available, utilize transcoding
5. A node ***X*** connects to Content Cache or CDN for a content t_l ε ***T*** if there is a missing GoP in D2D
6. **end while**
7. Play out the content

Algorithm 1. fCDN with caching and SLA guarantee

In the proposed algorithm (1), when a user queries the fCDN system for a content, the discovery system returns with the location of the devices having the content and other resources to share. In order to guarantee continuous streaming, the content cache (or from the CDN server in the worst case) located in the eNB nodes will be used as emergency resource.

The remaining GoPs are taken from the devices (through the stream collection part of the discovery system) based on the availability of the content and the distance between the user and the candidate device. By selecting devices that are as close as possible to the user, power consumption is reduced and throughput is increased.

If the client could only locate a GoP with higher bandwidth, it can pay the cost for transcoding the available rate into required one and then downloads into its buffer if the time permits for it.

When there is an inability to locate a GoP through D2D system, the content cache is searched for the GoP. If it's not found there either, the content is pulled from the CDN. In this case, the content in the cache servers located at the eNBs are updated accordingly.

3.4. Locating the Content

Locating the searched content is one of the critical problem to be addressed in this study. Although centralized content indexing is the fastest solution to the problem, it has scalability problems which forces us to use distributed content index in each eNB. This can easily be performed by using NFV based systems with the ability of installing series of Virtual Network Functions (VNF) at least; i) eNB; ii) index server; and iii) content cache. SDN/NFV based infrastructure on the mobile edge enable us to easily deploy cache and content servers and allocate enough resources to the system dynamically.

8. **Input:** ***S*** (set of nodes), ***T*** (set of content), ***N*** (set of nodes returned with searched content), *Interest* (***S*** → ***T***).
9. **Output:** System keeps track of all returns ***s*** ε ***S***, s_t ← (all devices and content associated to them, as well as CPU resources)
10. **while** (forever) **do**
    a. A node ***S*** arriving to the system registers its content ***t*** ε ***T***.
    b. A node ***S*** already within the system updates its status and location within every ***y*** seconds
11. **end while**

Algorithm 2. fCDN registration

In the proposed model, all participating device within the base station area registers themselves and their content to the index server located as a VNF in the same physical system of the base station. The index server acts as a gateway for the subscribers of the system not only keeping the records of the content hold by the devices within that base station area but also acts as a proxy initiating searches to foreign systems (CDN, etc.) for the content could not be located within the area of that base station.

In addition to keeping track of the content shared by the devices, it is necessary to keep the location of the devices holding resources within the system in order to optimize the use of energy. All devices joining to the system should register to the index server, and update the status and location through *hearthbeat* messages as depicted in Algorithm (2).

3.5 . Outbound Communication Alternatives

It is indicated that D2D can occur either through inband or outband channels. Although using inband allows the system to control and manage the subscribers through licensed cellular system, total communication capacity is limited to the capacity of the cellular system. On the other hand, subscribers have the flexibility to use unlicensed spectrum through cellular communication mediums as well as Wifi.

In addition to the link capacity Miettinen et al. [21] demonstrated energy usage characteristics of different systems including wifi and 3G systems. The simulation results indicate that, energy consumption of 3G for transferring data from short distance is twice more than transferring and processing same document through wifi. It becomes three times more for when performing the comparison between transferring from far node in 3G and Wifi system.

This fact leads us to use Wifi whenever available allowing us to reduce the energy consumption more than 50%.

4. EXPERIMENTAL STUDIES

In the simulation study, the maximum radius of the cell station is assumed to be 500 meters. There are a total of 100 users in the macro cell, of which 60 are uniformly distributed into 6 different clusters within 50 meters of the cluster centers, whereas the remainings are randomly positioned within the range of the macro cell. The distance between clusters are assumed to be 150 meters while individual nodes without belonging to a cluster is assumed to be 300 meters away from the clusters on average. In the case of devices outside of the clusters are assumed to be 200 meters away from each other on average while the average distance of those subscribers to the base station is assumed to be 400 meters. The total power to transmit and receive a bit is assumed to be as 100 nj and propagation of a bit for 1 $meter$ is assumed to be as 50 nj as proposed by [18], while transcoding a video part into half rate is assumed to be 48 nano joule as calculated from the proposal by [22]. The parameters for the simulation are given in Table (1).

4.1 . Simulating Constant Speed Video Delivery

In the first part of the experiments, a video with the length of 200 seconds is assumed to be distributed among the devices with several probabilities. The video is split into 100 consecutive Group of Pictures (GoP) which form a whole video. During the experiments, the availability rate (percentage) for a GoP in a mobile node is assumed to be x% out of 100 nodes. For example; if the availability rate

Table 1. Environmental parameters for the Simulation

Parameters	Value
Cellular Station radius	500 meters
Cluster Radius	50 meters
Number of Users	100
Average Number of Users in Clusters	10
ETx+ERx	50 nj/bit
εamp	100 pj/bit/m^2
Etranscode	28 nj/bit

for a GoP is 1% in the system, it means that the related GoP is only available in only one mobile node throughout the whole system. The system is assumed to have multiple different quality levels which can be trans-coded into lower quality by processing. In the experiments, the client requesting the video to play out is assumed to be in one of the clusters. Although there is a cost for locating a suitable GoP for a client, for this study, that cost is assumed to be negligible.

The client is allowed to fetch all GoPs during the experiment. The main concern of the candidate selection was on the distance to the source device, as well as the power consumption directly related to the distance. The energy consumption during a transaction is based on total energy consumed within the system without calculating consumption for transmitter and receiver nodes seperately. The simulations are performed with varying probability of existence for a GoP from 1% to 4% respectively.

Tables (2) and (3) summarizes the results of the simulation. In Table (2), the numbers 9, 45, 68 means that finding the content searched within the cluster is 9\%, within another cluster is 45\% and within other devices is 68%. As it can see from Table (2), when the existence

Table 2. Locating GoPs from the same cluster, within clusters and total D2D systems for single data rate for varying existence probabilities

Avail	100Kbps	200Kbps	400Kbps	800Kbps
1%	9,45,68	10,45,68	9,45,68	13,45,68
2%	19,74,88	19,74,88	17,73,87	15,73,87
3%	27,93,100	29,89,98	29,93,100	27,93,100
4%	37,99,100	43,100,100	37,99,100	33,95,100

probability increases, locating any GoP within the same cluster increases from 9% to 37% for 100 Kbps speed as obvious. When the existence probability goes beyond 2%, the need for connecting to content cache and/or CDN server becomes useless since all the necessary content can be obtained through the system.

In parallel to the above, it is clear from Table (3) that whenever the client locates the resource from the D2D system, the energy consumption decreases dramatically. The energy consumption saving becomes more than three fold in any speed when the probability increases from 1% to 4% respectively. It is almost 8 times less when compared with traditional CDN-cache server usage for video access in lower quality and similar benefits for higher quality.

Table 3. Energy consumption for a video request with single data rate for varying existence probabilities within D2D system

Avail	100Kbps	200Kbps	400Kbps	800Kbps
CDN/Cache	160J	320J	640J	1280J
1%	82.22J	164.05J	328.90J	651.40J
2%	49.80J	99.60J	200.00J	403.20J
3%	31.88J	61.35J	125.90J	255.00J
4%	23.73J	45.85J	94.90J	189.00J

4.2. Simulating Varying Speed Video Delivery

In the this part of the experiments, same video from the first experiment is assumed to be distributed among the devices with different quality levels which are 100Kbps, 200Kbps, 400Kbps and 800Kbps with the probabilities of 0.4, 0.3, 0.2 and 0.1 respectively. Similarly, the client requesting the video to play out is assumed to be in one of the clusters.The client is allowed to fetch all GoPs during the experiment. The simulations are performed with varying probability of existence for a GoP from 1% to 4% respectively for different clients with all video speeds consecutively.

Tables (4) and (5) summarizes the results of the second part of the simulation. Similar to the first experiments when the existence probability increases, locating any GoP within the same cluster increases for all speeds and existence probabilities as depicted in Table (4). Secondly, since the distribution differs for all speeds (0.4,0.3,0.2,and 0.1 respectively), the impact on the existence probability can obviously seen from the Table (4).

Table 4. Locating GoPs from the same cluster, within clusters and total D2D systems with variable data rate GoPs for varying client rates

Avail	100Kbps	200Kbps	400Kbps	800Kbps
1%	9,44,61	6,29,44	2,18,26	1,6,10
2%	18,71,86	10,51,72	6,31,46	1,10,19
3%	29,86,96	18,65,83	9,42,59	3,16,26
4%	37,96,100	23,80,93	11,51,71	4,20,33

Table 5. Energy consumption for a video request with variable data rate GoPs for varying client rates

Avail	100Kbps	200Kbps	400Kbps	800Kbps
CDN/Cache	160J	320J	640J	1280J
1%	99.37J	233.70J	531.87J	1203.36J
2%	66.34J	169.04J	442.49J	1129.08J
3%	49.06J	134.42J	382.13J	1052.48J
4%	42.76J	103.37J	323.93J	990.48J

In parallel to the above it is clear from Table (5) that energy consumption dramatically increases even if we have same total success rate for the same video. This is because either from utilizing transcoding to convert the speed into suitable rate which requires additional energy, or having worse quality video in the system which would not be possible to use. The latter will lead the client to connect either to a node far from him or connect directly to the content cache or CDN server. This lead us dramatic increase in energy consumption for higher quality. While the increase is in between 60%-90% for 100 Kbps client rate, it goes up to 90\%-400% for 800 Kbps streaming quality. As explained before the energy saving goes up to 4 times. In Tables (5) and (6) the first line gives the total energy consumption of the system to fetch and playout the video either from the centralized cache server located in the base station or from the CDN server.

4.3. Simulating Varying Speed Video Delivery in the Existence of Wifi

In addition to less power consumption and cheaper cost for streaming through D2D system, there is another benefit of performing streaming through the contribution of the D2D system. Since D2D communication can use out-of-band capacity other than licensed capacity of mobile network, it also overcomes the potential congestion in the mobile network because of the high bandwidth demand of the applications. In addition to overcome congestion problems incurring in the network, it also reduces use of energy as proposed in \cite{hotcloud} where the energy transmission and reception cost halves for close distance compared with cellular ones while it goes down 3 to 3.5 fold for the distant communication. In the experiments, the propagation cost change is left untouched while the transmission cost is halved. As depicted in Table (6) the results of the simulations indicate that the energy saving is around 1% less when heavily accessing outside of the cluster. The saving becomes more than 50% when using resources heavily from the same cluster.

In this experiment, an SDN-NFV based Wifi access device is installed around the center of clusters. It reduces the radius of the distance to half of previous one, as well as eliminating the transcoding effort since the cache installed in those devices can keep copies of the content in multiple quality forms. When the system is heavily populated with those hotspots even it saves energy more than 50\% when compared with Wifi based solution. as depicted in in Table (6).

Although there are some improvements in the SDN-NFV based system proposal, the improvement in the performance seen in Table (6) does not fit into expectations from the customers of telecommunication systems in the future. First of all, the experiment did not consider the increase in the cache hit rate although it should be. Secondly, although a SDN based cache server and Wifi access point to the clusters are placed, the internet link of them has not been considered for the third experiments. This conflicts with the idea of Wifi offloading, and 5G plans. Table (6) has lines identified with 5G, and you can see the amazing change when we utilize both cache hit rate increase and internet access through 5G. These numbers indicate that the future of telecommunication sector can be defined as *"the closer to the subscriber, the better"*.

Table 6. Energy consumption for a video request with variable data rate GoPs for varying client rates in the use of Wifi

Avail	100Kbps	200Kbps	400Kbps	800Kbps
CDN/Cache	160J	320J	640J	1280J
1% Cellular	85.63J	218.31J	522.20J	1210.08J
1% Cellular Wifi	85.43J	218.01J	521.91J	1209.94J
1% Cellular Wifi+SDN	85.01J	217.51J	521.55J	1209.88J
1% 5G	2.49J	5.04J	9.97J	25.40J
10%Cellular	11.77J	35.92J	134.08J	649.04J
10%Cellular Wifi	10.36J	33.93J	131.82J	647.10J
10%Cellular Wifi+SDN	9.18J	32.44J	130.21J	646.30J
10% 5G	2.34J	5.70J	12.43J	24.18J
50%Cellular	4.64J	10.84J	33.87J	158.76
50%Cellular 5 Wifi	2.38J	6.49J	26.56J	151.80J
10%Cellular Wifi+SDN	1.33J	4.32J	22.92J	148.90J
50% 5G	1.31J	2.84J	7.64J	19.11J

The results depicted in Table (6) gives some hints about achieving the promises of 5G. Although the energy saving is achieved by providing shorter distance for the devices, it also utilizes the use of Wifi to access to the internet which consumes half of what cellular communication is consuming. One of the main findings from the table is the results of 5G network which seems loosely coupled with the availability rate since the cost of transferring data from Wifi hotspot is very cheap compared with transferring it through cellular infrastructure from cache servers or CDN. Distributed Internet access also overlaps with one of the main desires of [2]. Besides, while the proposals in [2,3] propose to use base stations as caching point for a system, the proposed method goes one step further to organize the resources actively to the field next to consumers.

While the motivation behind sharing content with other nodes is not covered in this study, the nodes allowing the others to use their resources should be promoted. This mechanism shall also cover the use of CPU resources in addition to the usage of network resources. The use of CPU resources forms and edge computing system through mobile nodes.

5. CONCLUSIONS

In this study a futuristic content delivery system for mobile operators (fCDN) is proposed to control and coordinate the video streaming by utilizing both CDN assisted cache and D2D network. The proposed solution was inspired from the idea of using D2D communication, and with the framework of mobile device management and locating the content as well as benefiting from the improvements in the telecommunication world especially in SDN/NFV technologies and the effectively using them. The enabling technology of this architectural proposal is the migration of conventional systems into SDN-NFV based systems.

The main motivation of the proposed system is "using the content from the node as close as possible to the mobile user". This will enable the users to achieve minimal energy usage, while maximizing the throughput as well as minimizing the cost.

The results obtained from the study clarified two things; i) D2D contribution is necessary for scalability, and ii) architecting the mobile networks ready for 5G will allow users to access data with cheaper cost.

In the future, the studies should head towards the availability of consecutive GoPs in a mobile node or throughout the system in order to simulate the trace of a previous playout of a video as well as the monetization effect of the methods proposed and the balance between QoE and cost.

REFERENCES

[1] Wang, X., Chen, Taleb, T., Ksentini, A., and Leung, V. C. M, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems", ***IEEE Communications Magazine,*** 52(2) : 131–139, (2014).

[2] Yousaf, F.Z., Liebsch,M., Maeder, A., and Schmid, S., "Mobile CDN enhancements for QoE-improved content delivery in mobile operatör networks", ***IEEE Network***, 27(2): 14–21, (2013).

[3] Bradai, A., Singh, K., Ahmed, T. and Rasheed, T., "Celluar Software Defined Networking: A Framework", ***IEEE Communications Magazine,*** (6): 36–43, (2015).

[4] Oztoprak, K., and Akar, G. B., "Hybrid fault tolerant peer to peer video streaming architecture", ***IEICE Communications Magazine,*** 91(11): 3627–3638, (2008).

[5] Asadi,A., Wang, Q., and Mancuso, V., "A survey on device-to-device communication in cellular networks", ***IEEE Communications Surveys and Tutorials,*** 16(4): 1801–1819, (2014).

[6] Liu, J., Kato, N., Ma, J. and Kadowaki, N., "Device-to-Device Communication in LTE-Advanced Networks: A Survey," ***IEEE Communications Surveys & Tutorials,*** (c) 1–1, (2014).

[7] Oztoprak, K., "mCSDN: A software defined network based content delivery system with D2D contribution",

in: Proceedings - IEEE 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2053-2057, (2016).

[8] Amram, N., Fu, B., Kunzmann, G., Melia, T., Munaretto, D., Randriamasy, S., Sayadi, B., Widmer, J., and Zorzi, M., "QoE-based transport optimization for video delivery over next generation cellular networks", ***in: Proceedings - IEEE Symposium on Computers and Communications***, 19–24, (2011).

[9] Almashor, M., Khalil, I., Tari, Z., Zomaya, A. Y. and Sahni, S., "Enhancing Availability in Content Delivery Networks for Mobile Platforms", ***IEEE Transactions on Parallel and Distributed System,s*** 26(8): 2247–2257, (2015).

[10] Tong, S.-R., Du, S..T., Chen, L.-W. Chen, S., and Yeh, E., "A Peer-to-Peer Streaming CDN for Supporting OTT Video Broadcast Service in Mobile Networks", ***International Conference on Consumer Electronics-Taiwan (ICCE-TW)*** , 260–261, (2015).

[11] Sung, J. Lee, D., Bang, Y., Lee, J., and Rhee, J.-K.K., "Energy-aware algorithms for network-assisted device-to-device content delivery networks", ***Information and Communication Technology Convergence (ICTC)***, 469–471, (2015).

[12] Fodor, G., Dahlman, E., Mildh, G., Parkvall, S. Reider, N., Miklos, G., and Turanyi, Z., "Design aspects of network assisted device-to-device communications", ***IEEE Communications Magazine,*** 50(3): 170–177,(2012).

[13] Bao, X., Lee, U., Rimac, I. and Choudhury, R.R., "Data Spotting: offloading cellular traffic via managed device-to-device data transfer at data spots", ***ACM SIGMOBILE Mobile Computing and Communications Review,*** 14(3) :37, (2010).

[14] Ji, M., Caire, G. and Molisch, A., " Wireless device-to-device caching networks: Basic principles and system performance", ***http://arxiv.org/abs/1305.5216***.

[15] Yazici, M. A., Oztoprak, K., Akar, N., and Akar, G. B., "Markov zinciri tabanlı Güvenilirlik modellemesi", ***in: Proceedings-IEEE 16th Signal Processing, Communication and Applications Conference***, Aydin, pp. 1-6. doi: 10.1109/SIU.2008.4632689, (2008).

[16] Kim, W.S., Chung, S.H., and Moon, J.W., "Improved content management for information-centric networking in SDN-based wireless mesh network", ***Computer Networks,*** 92: 316–329, (2015).

[17] Rappaport,T.S., "Wireless Communications: Principles and Practice", ***IEEE Press, Prentice Hall PTR***, ISBN: 0-7803-1167-1, (1996).

[18] Heinzelman,W.R., Chandrakasan, A., Balakrishnan, H., "Energy-efficient communication protocol for wireless microsensor networks", ***Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, Maui, HI, USA***, (2):10. doi: 10.1109/HICSS.2000.926982, (2000).

[19] Alsultan, M., Oztoprak, K., Hassanpour, R., "Power Aware Routing Protocols in Wireless Sensor Network", ***IEICE TRANSACTIONS on Communications***, Vol.E99-B No.7: 1481-1491 . DOI: 10.1587/transcom.2015EBP3453, (2016).

[20] ISO/IEC, ISO/IEC. Information technology Dynamic adaptive streaming over HTTP (DASH), in: online from ***https://www.iso.org/standard/65274.html***, 2053-2057, (2016).

[21] Miettinen, A.P., Nurminen, J.K., " Energy efficiency of mobile clients in cloud computing", ***2nd USENIX Workshop on Hot Topics in Cloud Computing (HotCloud '10),*** Boston,MA, June, (2010).

[22] Ou, Z., Pang, B., Deng, Y., Nurminen,J.K., Yla-Jaaski, A., and Hui,P., Energy and Cost-Efficiency Analysis of ARM-Based Clusters, ***12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012)***, Ottawa, ON, pp. 115-123. doi: 10.1109/CCGrid.2012.84, (2012).