Fog Computing May Help to Save Energy in Cloud Computing

Fatemeh Jalali, Kerry Hinton, Robert Ayre, Tansu Alpcan, and Rodney S. Tucker

IBM Research, Melbourne, Australia; the Centre for Energy-Efficient Telecommunications (CEET), University of Melbourne, Australia.

Group Meeting, July 8, 2016

UCDAVIS

Fatemeh Jalali, Kerry Hinton, Robert Ayre, Tansu Alpcan, and Rodney S. Tucker, “Fog Computing May Help to Save Energy in Cloud Computing,” IEEE Journal of Selected Areas in Communications, vol. 34, no.5, May 2016.

Outline

1. Introduction
2. Network Topology
3. Energy Consumption Models
4. Measurement for Energy Models
5. Energy Consumption Comparison
6. Nano Servers for Improving Energy Efficiency of Applications
7. Conclusions

Fatemeh Jalali, Kerry Hinton, Robert Ayre, Tansu Alpcan, and Rodney S. Tucker, “Fog Computing May Help to Save Energy in Cloud Computing,” IEEE Journal of Selected Areas in Communications, vol. 34, no.5, May 2016.

Fog Computing and Cloud Computing

Cloud computing is a kind of Internet-based computing that provides shared processing resources and data to computers and other devices on demand.

Fog computing is a new paradigm referring to a platform for local computing, distribution and storage in end-user devices rather than centralized data centers (DCs).

nDCs: nano data centers (nano servers), which are located in end-user premises for hosting and distributing content and applications in a peer-to-peer (P2P) fashion.

Motivation: there has been little analysis, in the literature, of the energy consumption of Fog computing.

Main Contribution of This Paper

In this work, the authors aim to identify scenarios for which running applications from nano servers are more energy-efficient than running the same applications from centralized DCs. Measurement-based models are used for network energy consumption that are more accurate than used in previous work.

- End-to-end network topology
- Energy consumption models
- Measurement for energy models
- Energy consumption comparison
- Nano servers for improving energy efficiency of applications

End-to-End Network Model for Centralized Data Centers

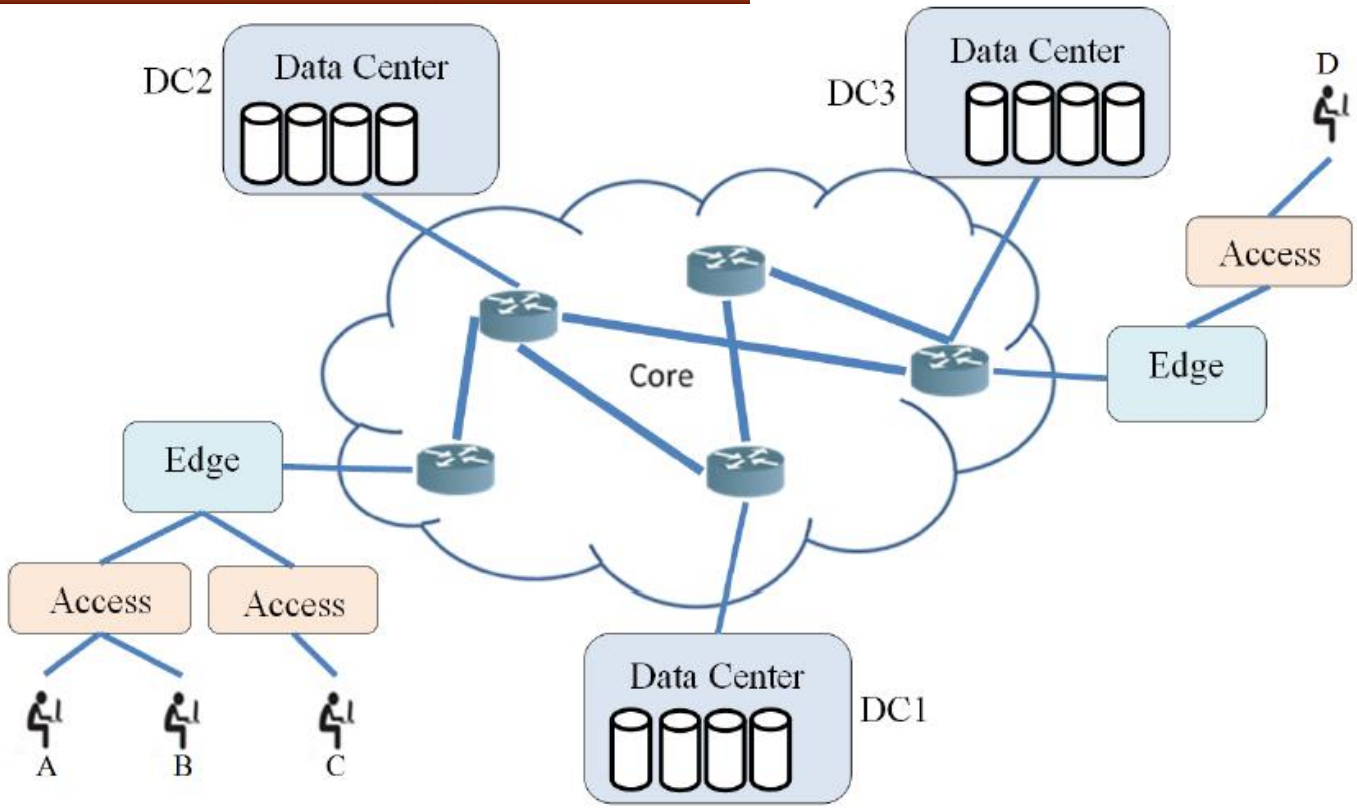


Fig. 1. Network model of centralized data centers.

One or a few centralized DCs are attached to the core of the network. Data center content is transported through large core routers and optical links to the edge network. The content passes through an access network which might be an Ethernet, WiFi, PON, 3G or 4G connection, or a combination of these to reach the end-user terminal.

End-to-End Network Model for Nano Data Centers

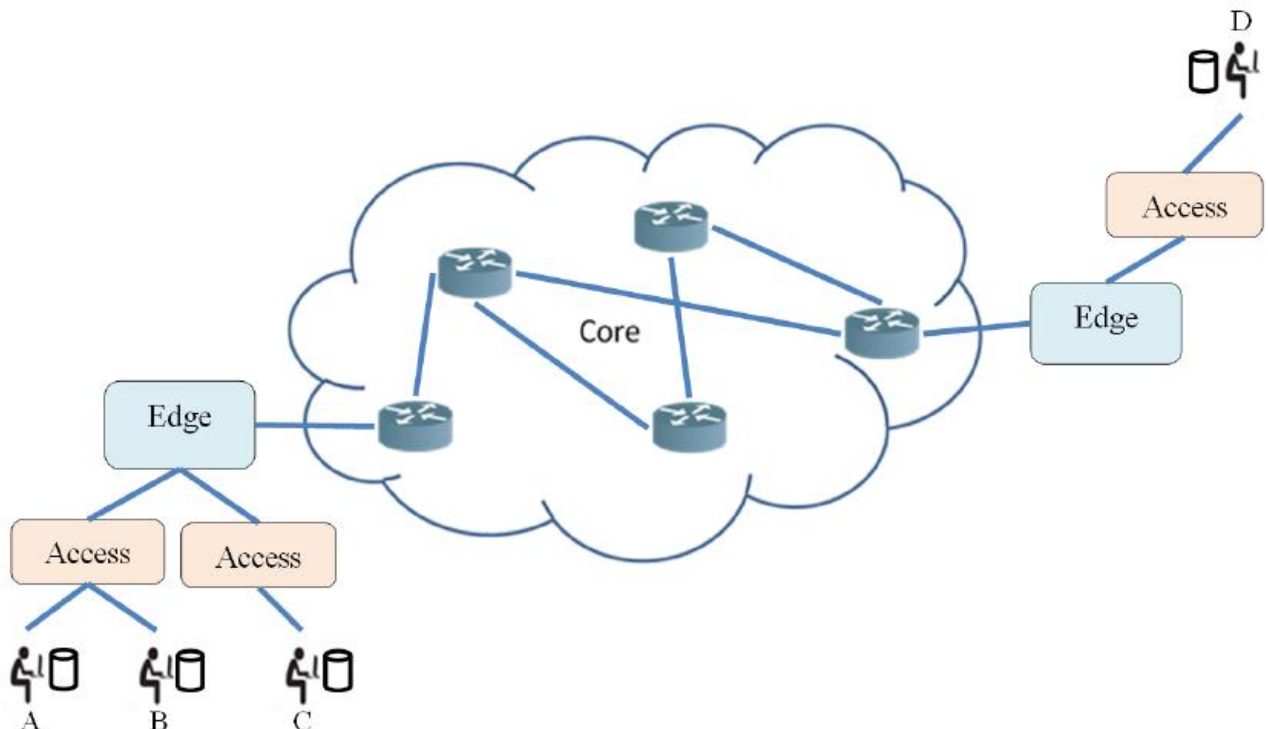


Fig. 2. Network model of distributed nano servers.

The requests are either sent from (i) “home peers” who are users located in the premises of the nano server (such as user A and user B), (ii) “local peers” who are users located in the same ISP of the nano server (such as user A and user C), or (iii) “non-local peers” who are users located in a different geographical region away from the nano server (such as user A and user D).

Energy Consumption Model

The network equipment are categorized into two types:

1) **Equipment that are shared by many users**
2) **Customer premises equipment (CPE) dedicated to a single user (or few users).**

For the highly **shared equipment** which deal with a large amount of traffic, a "**flow-based**" energy model is presented that proportionally allocates the equipment's power consumption over all the flows through the equipment.

For the equipment in **end-user premises** which are not shared by many users and services, a "**time-based**" energy consumption model is constructed based upon the amount of time that equipment spends dealing with a cloud service.

Flow-Based Energy Consumption Model

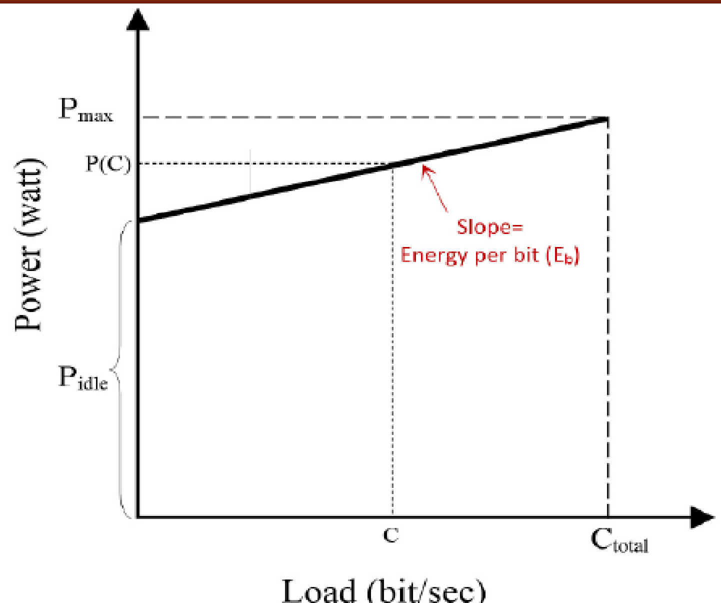


Fig. 3. Power consumption trend versus load for a network equipment (i.e. one router)

The measure of the energy consumption of a cloud service is based upon proportional allocation of the equipment's power consumption over all the flows through the equipment.

$$P(C) = P_{\text{idle}} + C\frac{P_{\max} - P_{\text{idle}}}{C_{\max}} = P_{\text{idle}} + CE_{\text{b}} \qquad (1)$$

The idle power (P_{idle}) can be a significant proportion of P_{max} (up to more than 90%), therefore we cannot ignore P_{idle} when calculating the energy consumption of a service.

Flow-Based Energy Consumption Model

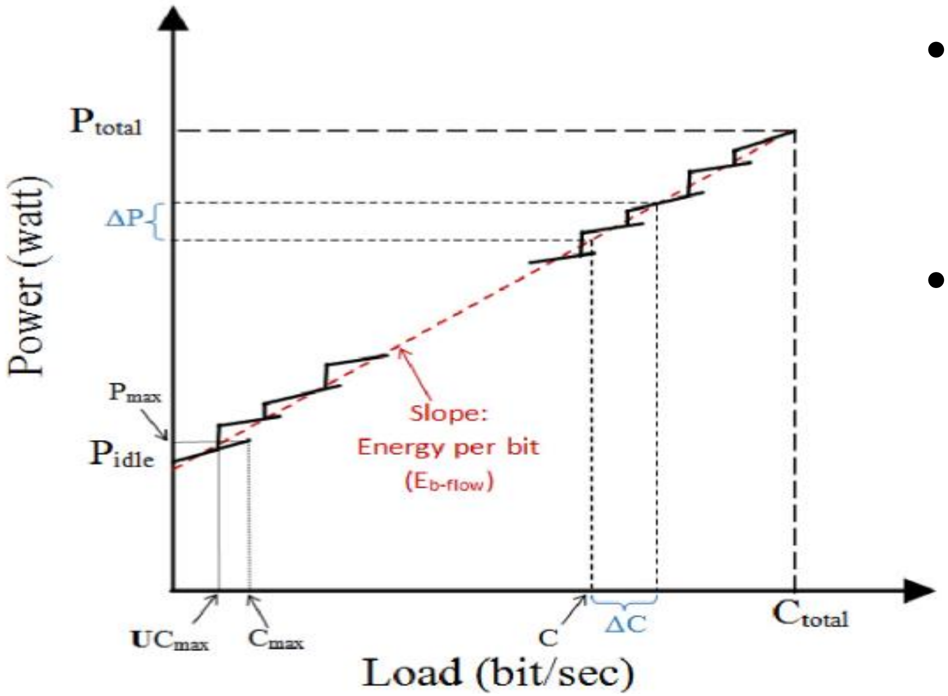


Fig. 4. Power consumption of a set of shared network equipment (i.e. routers) in one node located in a single location.

- Same model for all the equipment in the network which are shared by multiple users and services.
- Each step corresponds to the deployment of additional network equipment once the capacity per network equipment reaches the pre-set maximum operating load utilization, U.

The incremental energy per bit

$$E_{\text{b-flow}} = \frac{\Delta P}{\Delta C} \approx m\left(\frac{< P_{\text{idle}} >}{U < C_{\max} >} + < E_{\text{b}} >\right) \qquad (2)$$

The additional energy consumption of a service

$$E_{\text{k-flow}} \approx E_{\text{b-flow}} N_{\text{bit,k}} \qquad (3)$$

$N_{bit,k}$ is the number of exchanged bits of service *k* through the node by the service under consideration and *m* is the average number of network nodes in the service path.

Time-Based Energy Consumption Model

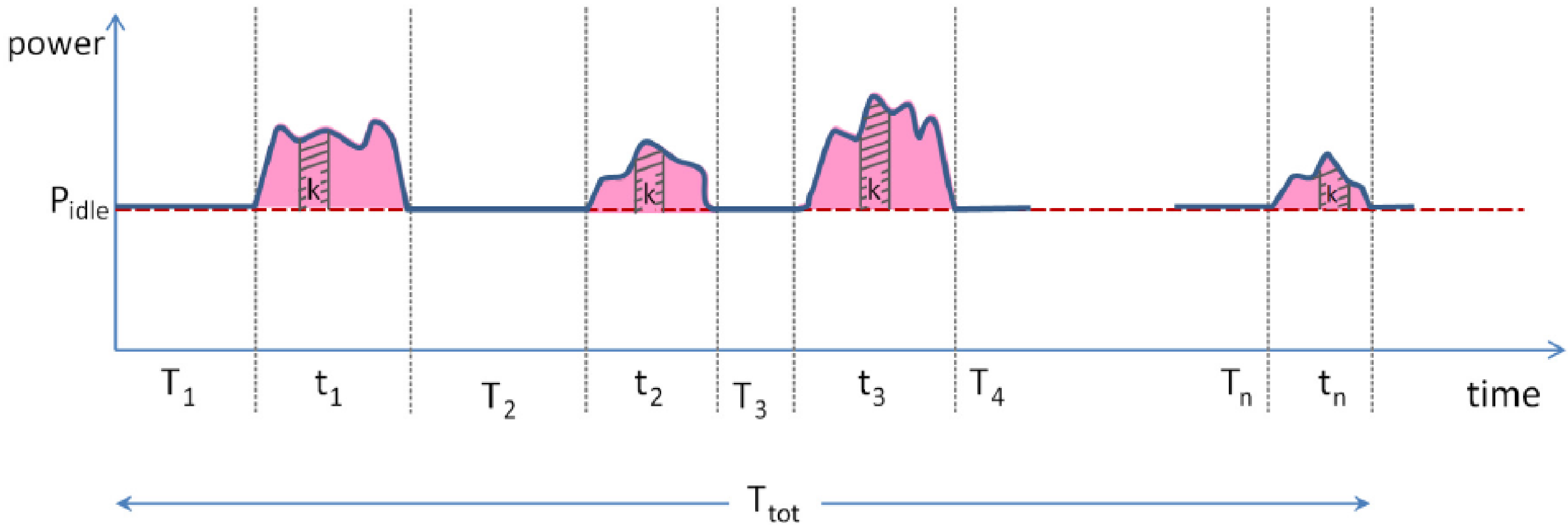


Fig. 5. Power consumption of a home equipment unit for serving/accessing services.

The energy consumption of the customer premises equipment(E_{cpe}) including the nano servers for serving multiple services is given by:

$$E_{\mathrm{cpe}} = P_{\mathrm{idle}} T_{\mathrm{tot}} + \int_{t_{act}} (P(t) - P_{\mathrm{idle}})\mathrm{d}t \qquad (4)$$

P_{idle} is power consumption of the device in the idle mode.

Time-Based Energy Consumption Model

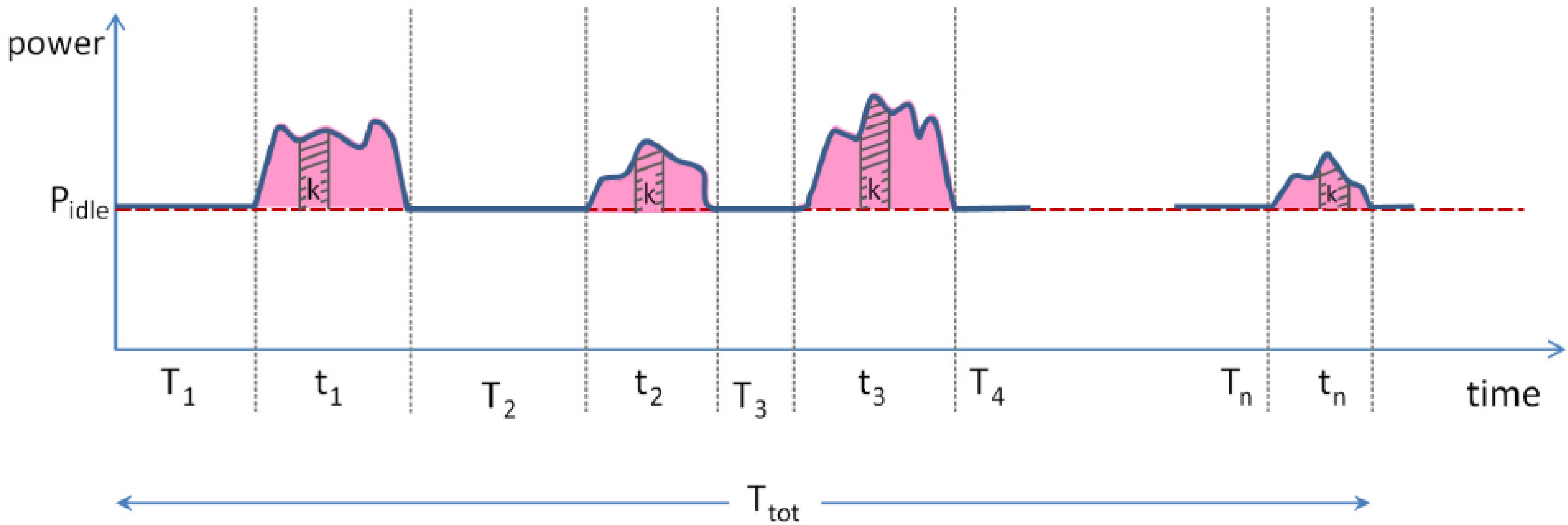


Fig. 5. Power consumption of a home equipment unit for serving/accessing services.

To determine energy consumption of one specific service running on the device such as service *k* (the hatched area in Figure 5, two parts are considered: 1) incremental energy consumption due to running this specific service ($E_{inc,k}$); 2) idle power allocated to running the service ($E_{idle,k}$).

Time-Based Energy Consumption Model

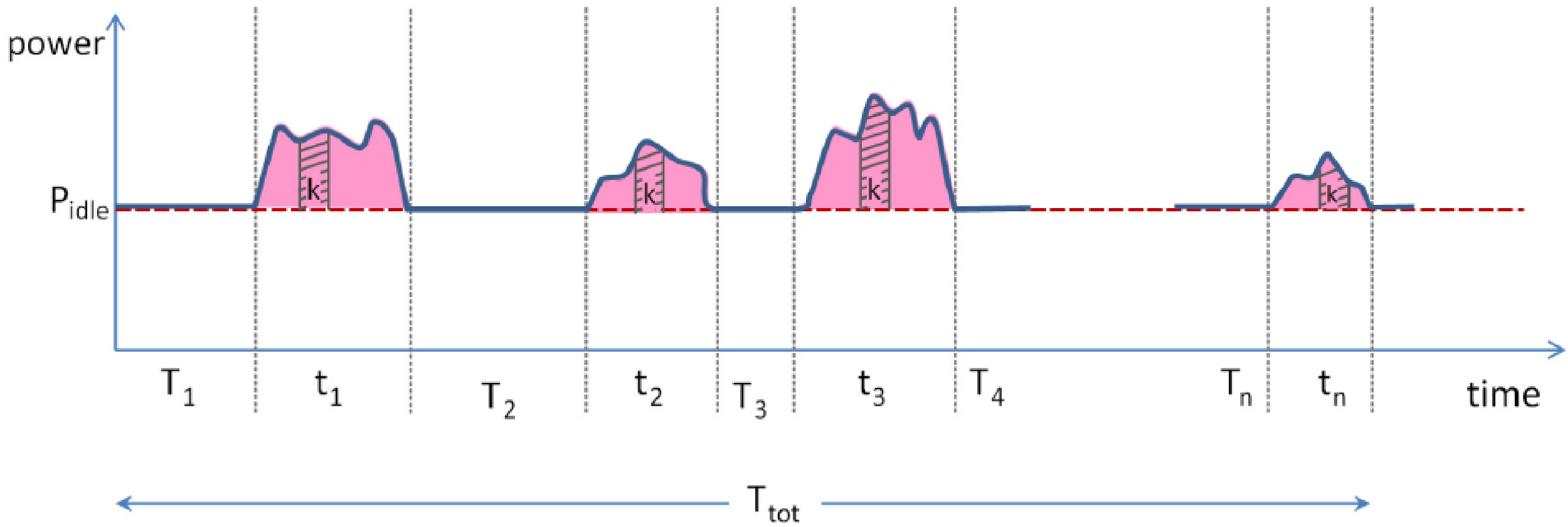


Fig. 5. Power consumption of a home equipment unit for serving/accessing services.

$$E_{\text{k-time}} = P_{\text{idle}}(\alpha + 1)t_{\text{act,k}} + \int_{t_{\text{act,k}}} (P(t) - P_{\text{idle}})\mathrm{d}t \qquad (8)$$

Coefficient (α) is the ratio of the idle time of the device to the active time.

Centralized Data Centers and Nano Data Centers

The total energy consumed by service *k* provided from a centralized DC ($E_{k\text{-dc}}$) can be expressed as:

$$E_{\text{k-dc}} = E_{\text{k-cpe}} + E_{\text{k-access}} + E_{\text{k-edge}} h_{\text{e}} + E_{\text{k-core}} h_{\text{c}} + E_{\text{k-cent}} \quad (9)$$

The total energy consumed by service *k* provided from nDCs can be expressed as:

$$\begin{aligned} E_{\text{k-ndc}} = & E_{\text{k-cpe}} + E_{\text{k-access}} + E_{\text{k-edge}} h_{\text{e}} \\ & + E_{\text{k-core}} h_{\text{c}} + E_{\text{k-access2}} + E_{\text{k-nano}} \end{aligned} \quad (10)$$

The differences between energy consumption of a service provided from a centralized DC compared to nDCs is primarily determined by the following:

– The number of bits exchanged between the user and DC (N_{bit});

– The number of hops for the two cases (h_e, h_c);

– The value of $E_{k\text{-cent}}$ compared to $E_{k\text{-access2}} + E_{k\text{-nano}}$.

Measurement of Energy Consumption Models

To quantify the models for $E_{k\text{-}dc}$ and $E_{k\text{-}ndc}$, power and traffic measurements are used undertaken using the Wordpress IV application which is an open source website and blogging tool. There are two options for Wordpress users:

1) Sign up for an account from the Wordpress website and connect to the Wordpress centralized DCs; (centralized DCs)

2) Install Wordpress software locally and create a web-server and host the content locally on a nano server. (Nano DCs)

Traffic Measurements (N_{bit})

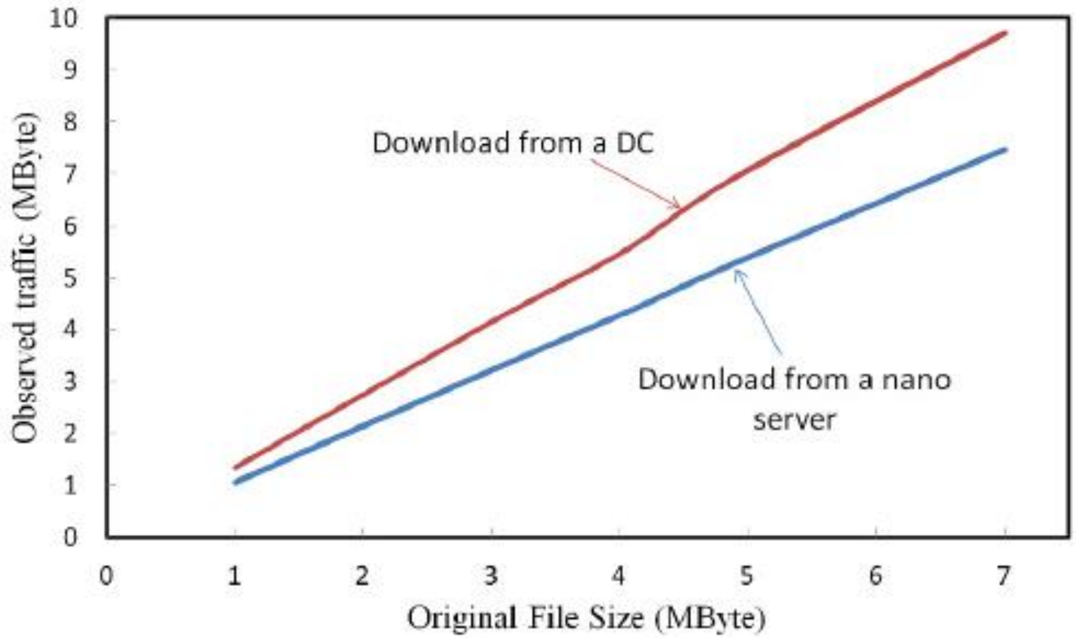
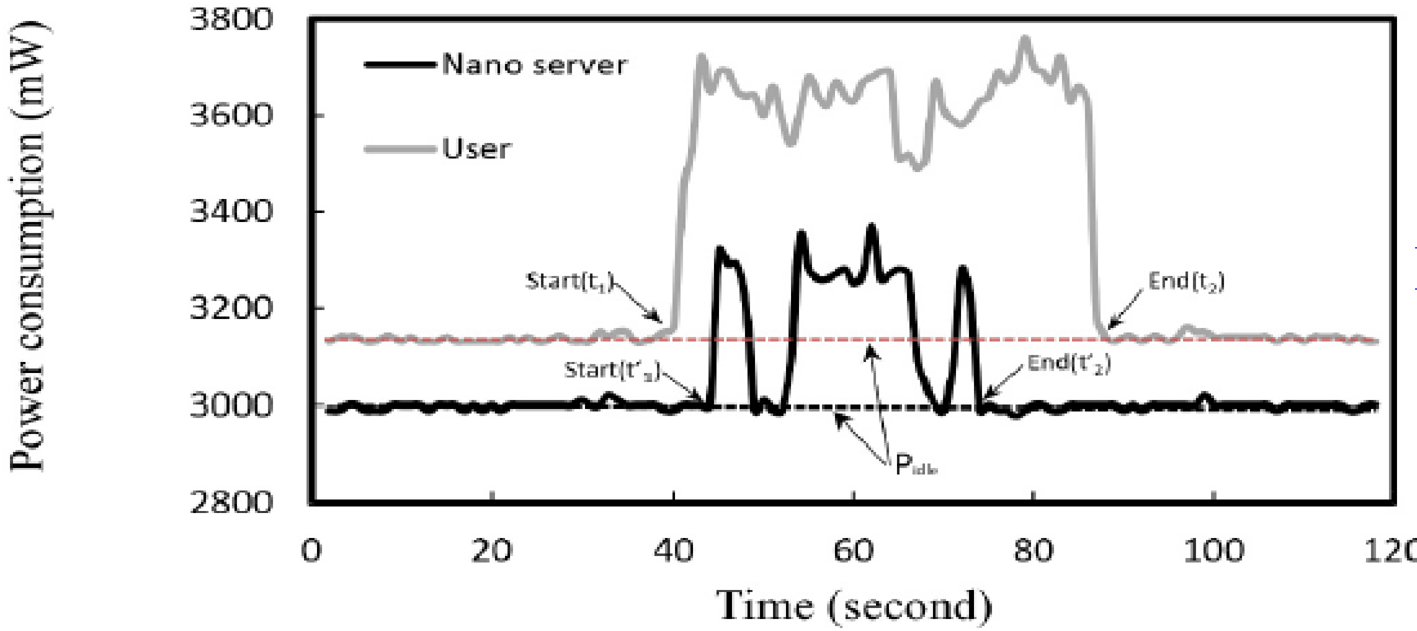


Fig. 6. Exchanged bytes during downloading files varying in size from Wordpress website versus the original sizes of files.

Wireshark

The download curve for the nano server indicates the traffic exchanged is very similar to the original size of files. However, the traffic for downloading from the DC is higher than the original file size. Post-processing the Wireshark logs reveals that the download traffic from centralized DCs is higher than the original file size due to the existence of third party applications and advertisement traffic.

Power Measurements (P_{cpe})



Power meter

Fig. 7. Power consumption of an end-user device and a nano server while uploading a file to Wordpress.

First open the web browser in the end-user device and then upload a file (t_1 in the user curve). After that, the nano server starts to process and store the file (t'_1 on the nano server curve). After storing, the local server status switches to idle mode (t'_2 in the nano server curve). Then the end-user device completes the final processing after which it also switches to idle mode (t_2 on the user curve).

User and Access Network Equipment ($E_{k\text{-}cpe} + E_{k\text{-}access}$)

TABLE IV

ENERGY PER BIT OF NETWORK EQUIPMENT IN ACCESS, EDGE AND CORE NETWORKS

	Power(Watt)		Traffic(Gbps)		Energy(nJ/bit)	
	Idle	Max	Downlink	Uplink	Downlink	Uplink
Fast Ethernet gateway (CPE)	2.8	4.6	0.1	0.1	N/A	N/A
ADSL2+ gateway (CPE)	4.1	6.7	0.024	0.003	N/A	N/A
4G gateway (CPE)	0.5	1.75	0.024	0.012	N/A	N/A
GPON gateway (CPE)	5.2	8.3	2.4	1.2	N/A	N/A
Ethernet switch	1589	1766	256	256	31.7	31.7
LTE Base-station	333	528	0.072	0.012	82820	12400
OLT	43	48	2.4	2.4	88	179
BNG	1701	1890	320	320	27	27
Edge Router	4095	4550	560	560	37	37
Core Router	11070	12300	4480	4480	12.6	12.6

As one would expect, the measurement results indicate, for a given connection technology, the energy consumption of end-user device for uploading and downloading data to the centralized Wordpress DC is approximately equal to uploading and downloading data to the nano server.

Edge and Core Network Equipment ($E_{k\text{-}edge}he + E_{k\text{-}core}hc$)

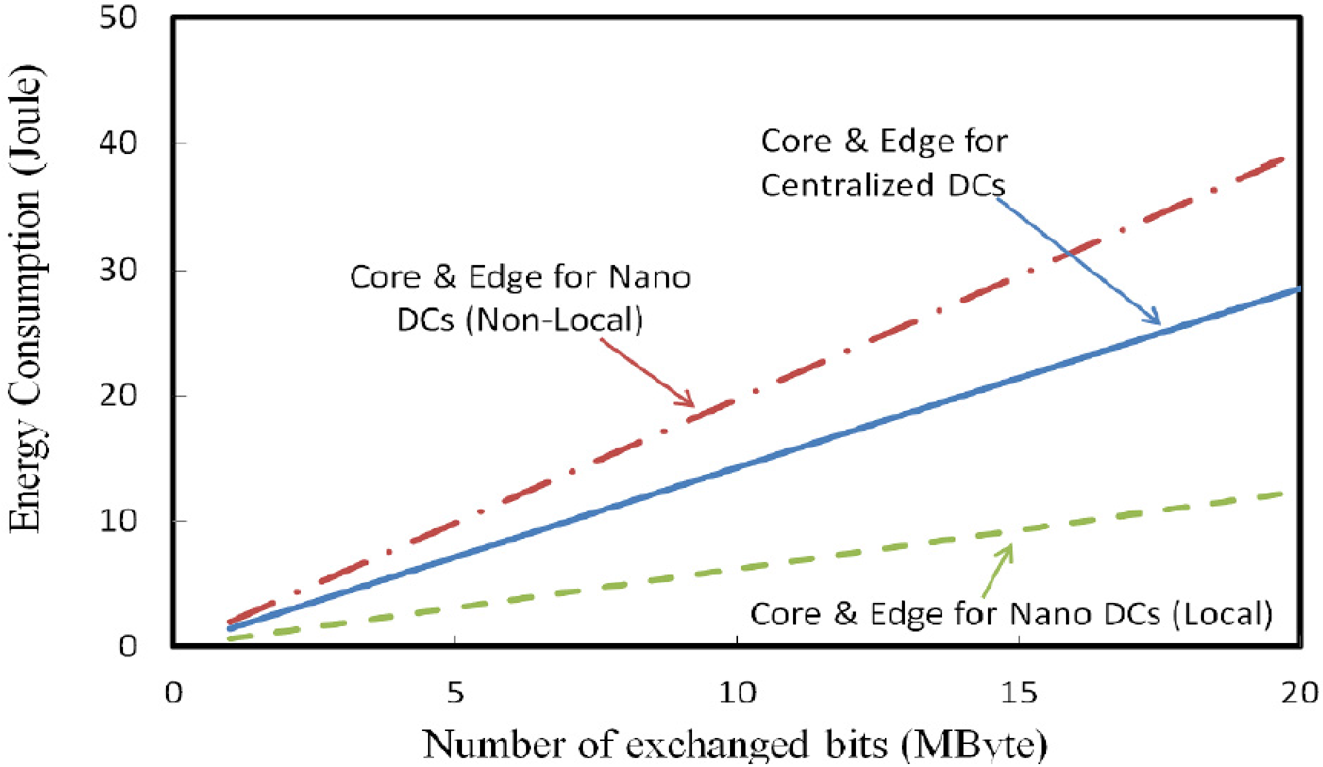


Fig. 8. Consumed energy in the core and edge equipment for accessing data from different locations.

Based on equation (9) and (10), and table IV.

The energy consumption resulting from requesting data from nano servers can be higher or lower than the energy consumed for accessing the content in centralized DCs depending on distance between the users and the stored content.

Nano Servers ($E_{k\text{-access2}}$ + $E_{k\text{-nano}}$) and Centralized Servers ($E_{k\text{-cent}}$)

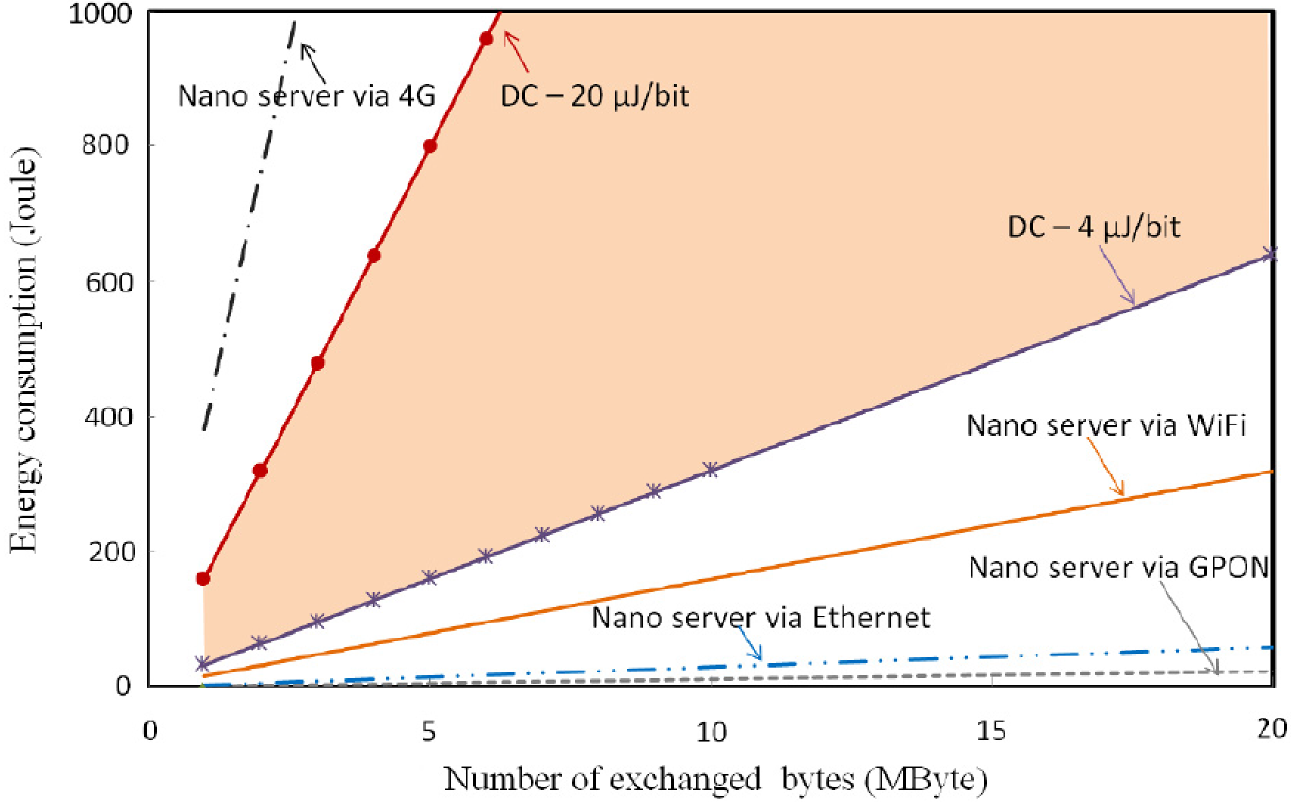


Fig. 9. Energy consumed by service *k* in various nano servers and DCs as a function of the volume of data exchanged.

Fig.9 shows how the energy consumption of the access network can affect the energy consumption of a service provided by nano servers.

Nano Servers ($E_{k\text{-access2}}$ + $E_{k\text{-nano}}$) and Centralized Servers ($E_{k\text{-cent}}$)

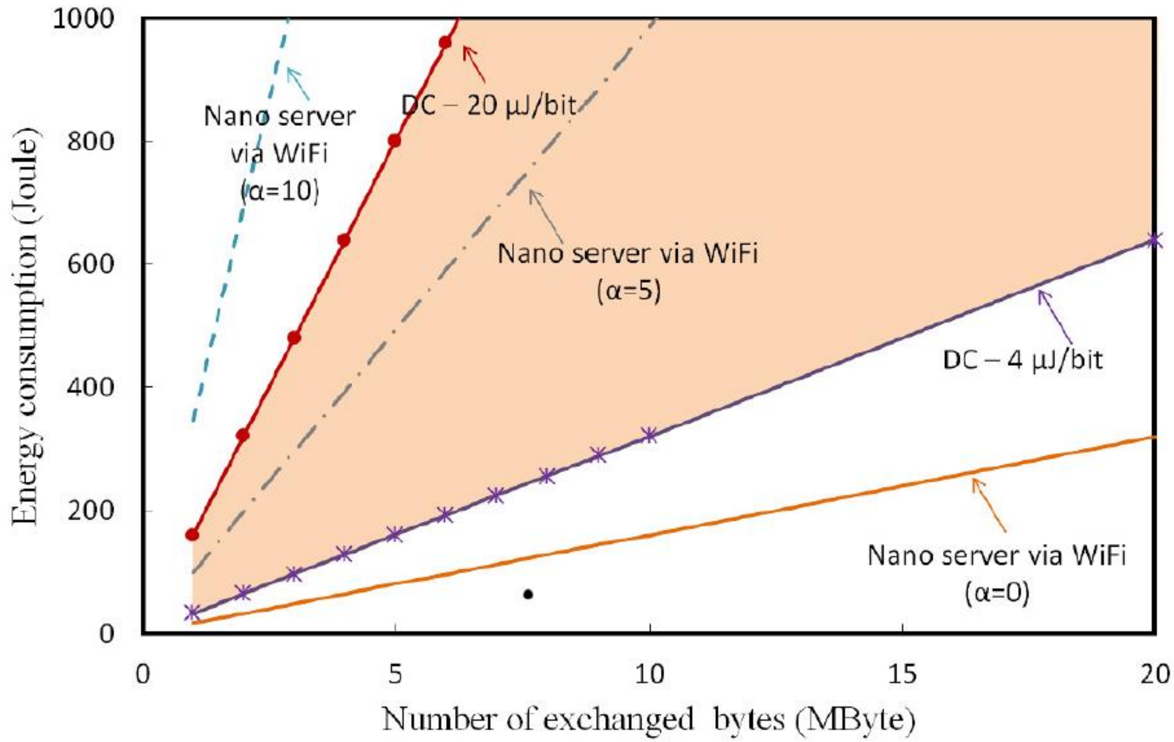


Fig. 10. Energy consumed by service *k* provided by WiFi nano servers with different ratios of idle time to active time (α) as a function of the volume of data exchanged.

Without sharing the idle time of nano server with other services and with assigning more idle time to the service *k* (increasing α), the energy consumption of the service running on the nano server increases and dominates the energy consumption of running the same service from the DCs.

Nano Servers for Improving Energy Efficiency of Applications

The total energy consumed by service *k* provided from a centralized DC ($E_{k\text{-}dc}$) can be expressed as:

$$E_{\text{k-dc}} = E_{\text{k-cpe}} + E_{\text{k-access}} + E_{\text{k-edge}}h_{\text{e}} + E_{\text{k-core}}h_{\text{c}} + E_{\text{k-cent}} \quad (9)$$

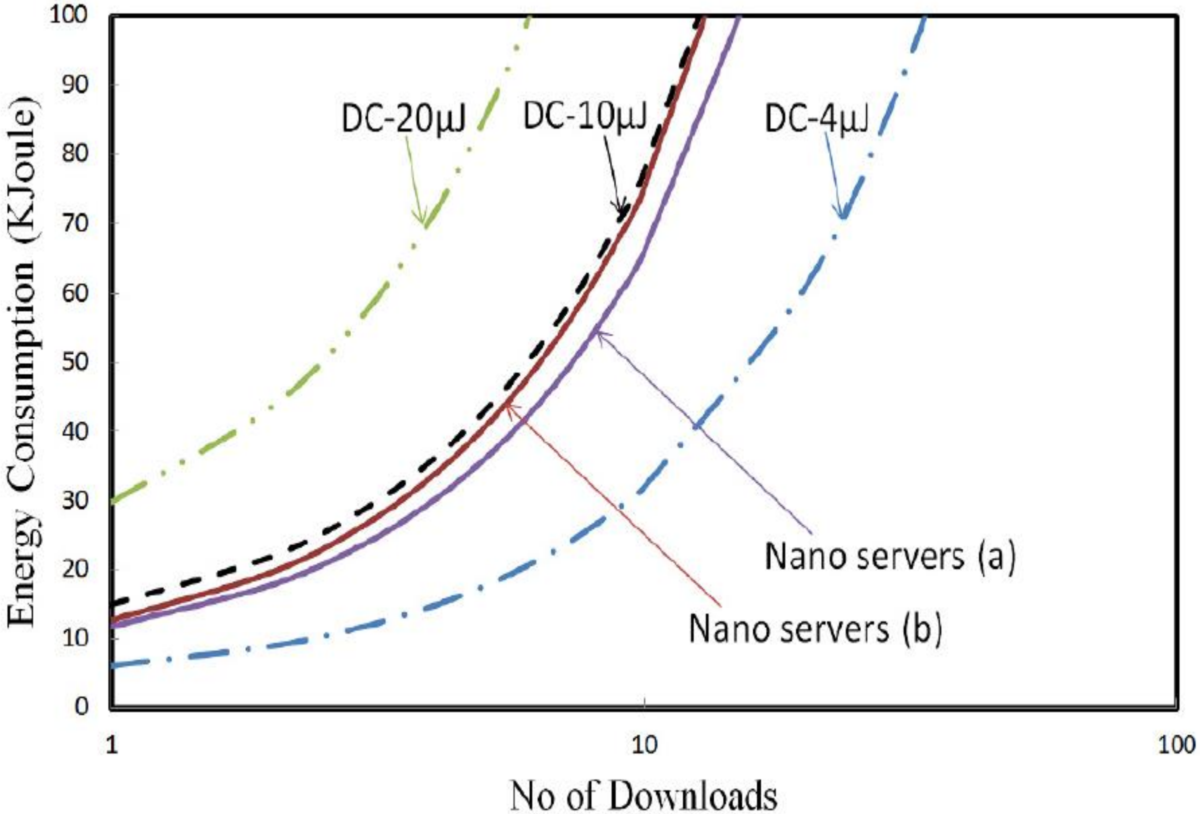
The total energy consumed by service *k* provided from nDCs can be expressed as:

$$\begin{aligned} E_{\text{k-ndc}} = E_{\text{k-cpe}} + E_{\text{k-access}} + E_{\text{k-edge}}h_{\text{e}} \\ + E_{\text{k-core}}h_{\text{c}} + E_{\text{k-access2}} + E_{\text{k-nano}} \end{aligned} \quad (10)$$

UCDAVIS
UNIVERSITY OF CALIFORNIA

Nano Servers for Improving Energy Efficiency of Applications

(1) Applications With Static Content for Which the Source of Data is Primarily in End-User Premises (static website)



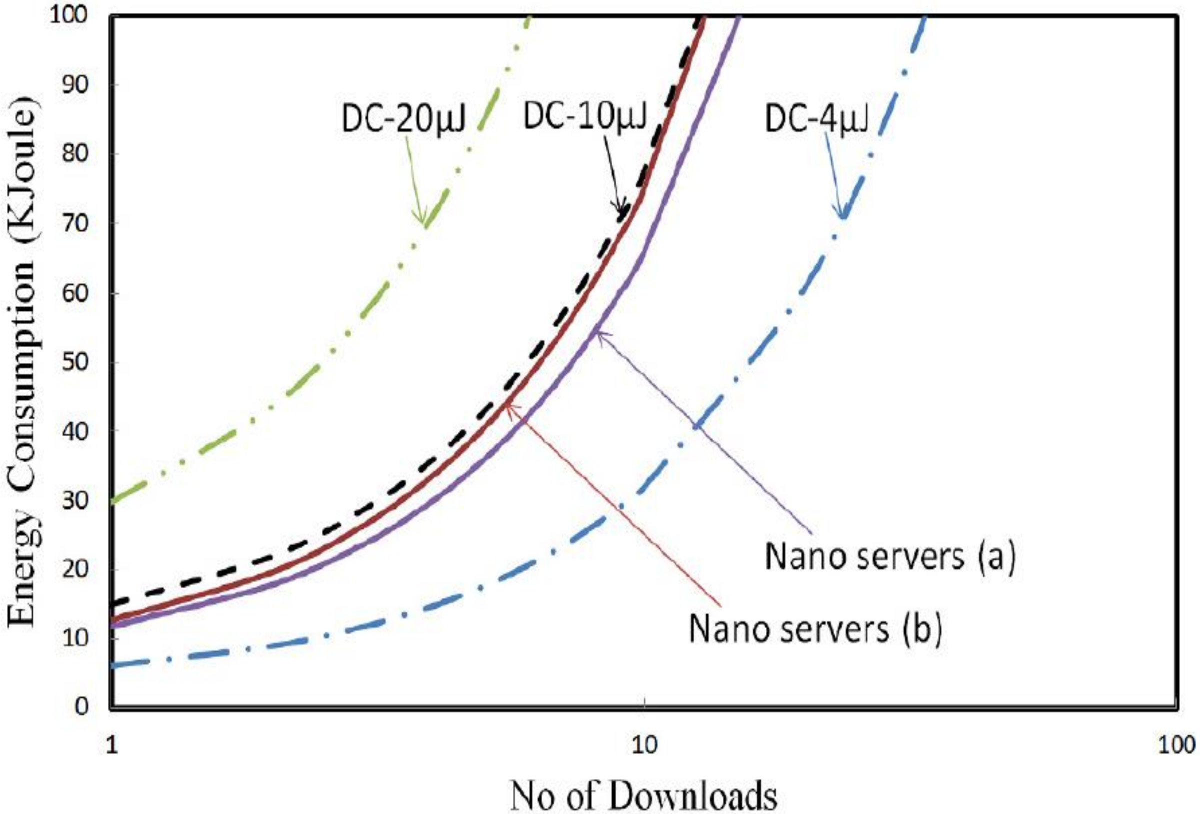
(a) 20% of access events from non-local peers;

(b) 80% of access events from non-local peers.

Fig. 11. Energy consumption of an application running form nano and centralized DCs vs number of downloads to users.

The ratio of local to non-local requests has little impact in the total energy consumption, because the energy consumption of the application is dominated by the **access network and data centers** (nano or centralized).

(1) Applications With Static Content for Which the Source of Data is Primarily in End-User Premises (static website)



(a) 20% of access events from non-local peers;

(b) 80% of access events from non-local peers.

Fig. 11. Energy consumption of an application running form nano and centralized DCs vs number of downloads to users.

However, if the initial values of energy consumption of a DC and a nano server for hosting an application are close, the energy consumption due to the use of local or non-local peers can be a determining factor for which of centralize and nano DCs are more energy consuming.

(2) Applications With Dynamic Content for Which the Source of Data is Primarily in End-User Premises (video surveillance)

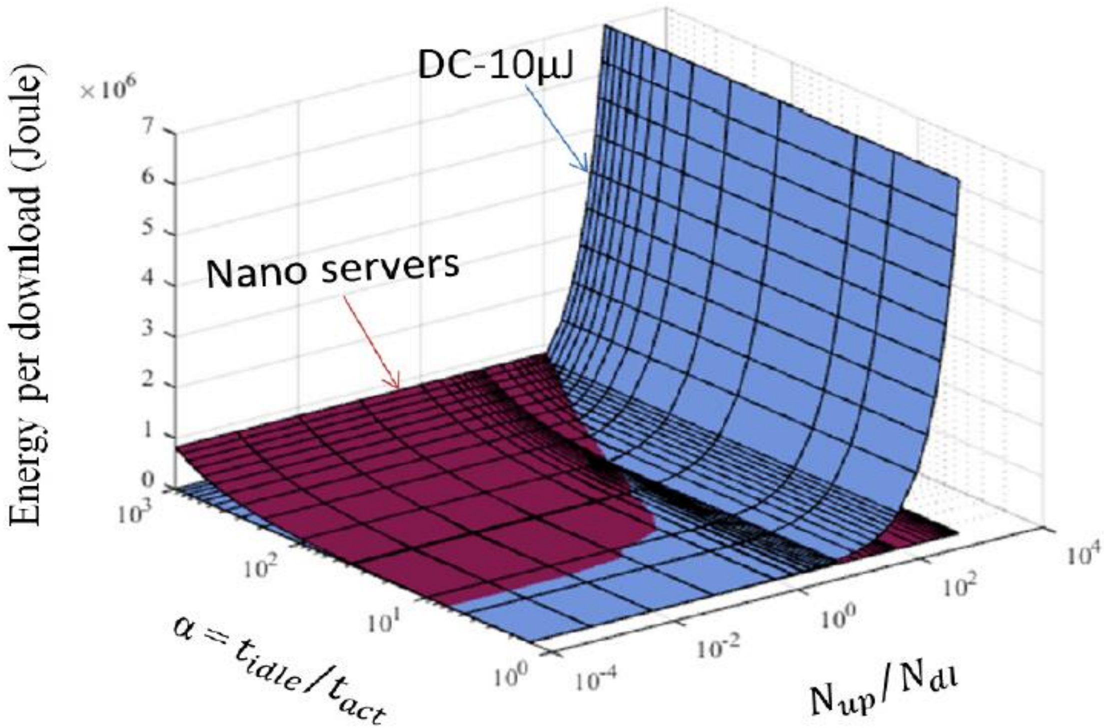


Fig. 12. Energy consumption an application running form a nDC and DC considering number of downloads and updates.

The ratio of updates to downloads of an application plays an important role in the relative energy consumption of providing an service from a centralized DC compared to a nDC. **Applications with a higher upload rate and low download rate are more energy-efficient when provided via on the nano servers architecture.**

(3) Applications Requiring Data Pre-loading (VoD)

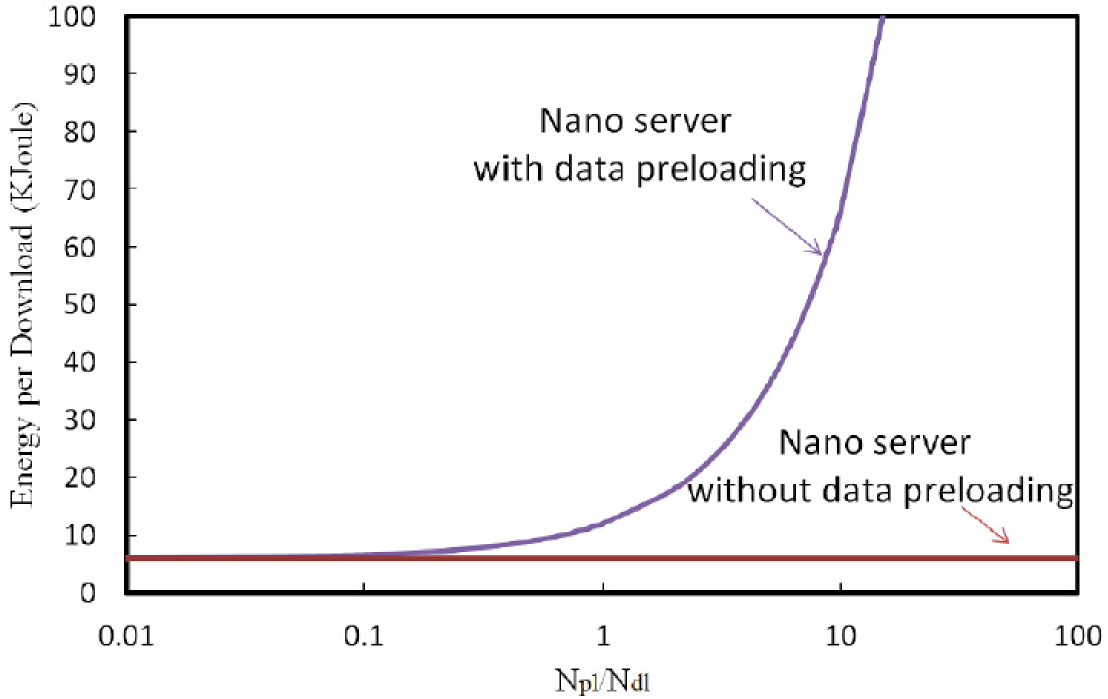


Fig. 13. Energy consumption versus number of data pre-loading to number of downloads (Npl/Ndl).

The number of preloaded data should be consistent to the number of downloads($N_{pl}/N_{dl} \leq 1$) to execute an energy-efficient application on nano servers. It means that **popular contents with more number of downloads for each data pre-loading are more energy-efficient to be run by nDCs compared to unpopular contents.**

Conclusions

nDCs might lead to energy savings depending on system design factors such as:

(a) Type of access network attached to nano servers
(b) The ratio of active time to idle time of nano servers
(c) Type of applications which includes factors like number of downloads from other users, number of updates from the origin(s) and number of data pre-loading. It was also shown that number of hops between users and content has a little impact compared to the above-mentioned factors.

The results of this work show that the best energy savings using nDCs is for applications that **generate and distribute a large amount of data in end-user premises which is not frequently accessed such as video surveillance in end-users homes.**

Thank you for your attention!

Yongli Zhao

State Key Laboratory of Information Photonics and Optical Communications (IPOC),
Beijing University of Posts and Telecommunications (BUPT), Beijing 100876, China
Tel : +86-10-61198108; Email: yonglizhao@bupt.edu.cn; Web: ipoc.bupt.edu.cn