CTNDCI: Identifying the Challenges Towards a distributed Nano Data Center Infrastructure

MELANIE HAUSER, LMU Munich, Germany

DIANA IRMSCHER, LMU Munich, Germany

KATRIN KOLB, LMU Munich, Germany

MENGCHU LI, LMU Munich, Germany

KATHARINA RUPP, LMU Munich, Germany

ANDREAS SCHOLZ, LMU Munich, Germany

In this paper we identify the challenges currently preventing nano data centers from becoming the dominant form of content provision on the internet. With the global increase in IP traffic the question of how to provide and deliver data is becoming increasingly important. Monolithic data centers, as they are used today, pose several problems, such as high energy consumption and lack of scalability. An alternative solution mitigating the problems of monolithic data centers has been proposed in the form of a distributed nano data center infrastructure. Research has shown this to be a superior solution. However, no widespread solution based on a nano data center infrastructure has been implemented as of yet. By identifying the main challenges nano data centers are facing steps can be taken to overcome these challenges in a more focused way, leading to a more economic data distribution.

Additional Key Words and Phrases: Green IT; Nano data center; Energy consumption; Security; Availability; Scalability; Data distribution

1 INTRODUCTION

2 STATE OF THE ART

Several researchers and experts have examined the idea of Nano Data Centers. Some have even developed systems and ideas, but also stumbled upon issues and hurdles. Their papers are summarized in the following paragraphs.

2.0.1 ECHOS. ECHOS introduces a concept for Nano Data Center that can or should completely replace monolithic data centers [9]. The authors call it a radical solution for data management and provision. According to this concept, so-called "boxes" are set up at the edges of the network, eg. in home gateways (see [10]). These boxes communicate with each other via a peer-to-peer system. The peer-to-peer system as well as the bandwidth is controlled by a central unit, such as the ISP. However, the approach of networking boxes via a peer-to-peer system, and thus providing or sharing content, requires some conditions. So it is first necessary to provide a distributed hosting edge infrastructure. Furthermore, there are still some problems added. In ECHOS these are listed as follows [9]:

- "Lack of service guarantees due to uncontrolled interface between different application [...]."
- "Inefficient use of network's and other peer's resources and consequently supoptiomal performance [...]."
- "Even if sufficient status information is in place, still P2P is inherently unable to use it as it was designed around selfish user behavior and free-riding prevention mechanim [...]."
- "Absence of security and control make it impossible to guarantee the integrity and security of content."

2.0.2 . Eum et al. present the application scenario CATT, which is supposed to mainly decrease delay and disruption for video watching. In their opinion, there are three main problems, which can be solved individually, but there is no single solution for all of them. CATT is supposed to provide this solution. The first problem is that mobile and fixed-line operatiors are afflicted with increasing operational cost. The second problem is that mobile users have to accept long delays because of the distance of the providers. Furthermore, mobile users suffer from service disruption and disconnections during handover. With CATT ICN nodes can work as independent content providers, which removes the latency and multiple hops. // A prototype of the system has been developed, but the idea has not been fully developed. The authors point out that during development special attention has to be given to the protection of not only the data, but furthermore of a malicious network resource abuse. [6].

-

3 CHALLENGES TOWARDS A DISTRIBUTED NANO DATA CENTER INFRASTRUCTURE

In the following sections, the methodology is first described, and after that political and technical challenges are discussed.

3.1 Methodology

To identify the challenges towards a distributed nano data center infrastructure we used two typed of methods. First, our research consisted of studying previous work and papers. We apportioned work we found among the team members. Every team member extracted the challenges, which were stated in the papers, and analyzed them for our work.

The second method we used was conducting a qualitative interview with an expert. For this interview we chose an expert on nano data centers, who works for the Leibniz Supercomputing Center.

We purposeful excluded other techniques from our methodology. Observations and experiments simply did not fit our research topic and its current research status. Online surveys, questionnaires, and interviews with pedestrians were also excluded, because the topic demands a certain background knowledge and understanding. This knowledge and understanding cannot be expected from citizens.

Our results from the research and interview are stated in the next sections.

3.2 Political and legal challenges

Due to its distributed nature a nano data center infrastructure will face additional political and legal challenges when compared to a traditional monolithic data center infrastructure. In the following sections we will discuss important political and legal challenges. We acknowledge that there will be many more challenges to be overcome in these fields. However, most of these will arise during implementation and are therefore difficult to predict. Hence, we restricted ourselves to challenges of each field which will certainly have to be overcome.

3.2.1 Political challenges. Once the decision to transition to nano data centers has been made the question arises of who will be responsible for that infrastructure. Considering that nano data centers would likely be integrated into today's modems, with which each household connects to their internet service provider (ISP), the ISP would be an obvious choice. The ISP already has knowledge of their network and is also already distributing modems. However, this raises other challenges:

How to distribute data between ISPs?
Most internet users only have one ISP they use to connect to the internet. As not every ISP can be expected to store all of the worldwide information that is currently stored in monolithic data centers, a solution has to be found to access data which is not currently provided by the ISP a person is connected to. This might include international

partnerships which further increase the challenge to introduce political policies for such cooperations.

How to ensure net neutrality?
If ISPs are not only responsible for providing bandwidth but also for providing data, a conflict of interest might arise as to which data to provide with which priority. Especially when considering cooperations between different ISPs as mentioned above, an ISP might want to prioritize data deliverance to its own customers before servicing those of other ISPs. To ensure net neutrality, policies have to be introduced to ensure the same quality of service for each customer across different ISPs.

If these challenges prove too difficult to overcome, ISPs might not be the best choice to manage nano data centers. In this case a third party would have to be introduced for managing the nano data center infrastructure. By having an ISP independent third party the issue of net neutrality would be greatly reduced. However, the challenge on how to distribute data across different ISPs would be unaffected. Additionally the question of how to distribute the management software of the nano data center infrastructure onto ISP dependent hardware would arise.

3.2.2 Legal challenges. Legal regulations will pose challenges on the way to a distributed nano data center infrastructure. Some of these regulations are already in effect, others will yet have to be introduced to deal with the new state of the art. In this section we will introduce one challenge each as an example of what kind of legal challenges can be expected. As before, we acknowledge that there will be many other challenges to be overcome.

General Data Protection Regulation (GDPR) of the EU.
On May 25th, 2018 the GDPR EU regulation will come into affect which states, that any individual has the right to request information about data, which can personally identify them. This includes information about where this data is stored, whether it will be transferred to other entities, how long it will be stored, for which purpose it is stored and many others. It also provides an individual with the right to request access, rectification and erasure of this data as well as restricting further processing of that data [5].

[ab hier erst mal nur Notitzen] "Wenn ich jetzt da personenbezogene Daten auf so verteilten Systemen [...] speichern will, dann könnte ab 25. Mai nächsten Jahres einer der jetzt in den Daten gespeichert ist verlangen, "wo sind denn meine Daten überall?", ja? Also da seh ich schon noch Sachen wo ich sag, mit so verteilten Systemen und Rechenzentren wird des relativ schwierig, ja, was nicht heißt, dass das nicht funktioniert, was aber vielleicht heißt, dass der Aufwand das zu realisieren mehr ist als wie wenn mans bei uns macht [?]."
   - Who is liable for illegal data stored on a persons nano data center?

3.3 Technical challenges

3.3.1 Energy Consumption Model for Nano Data Centers. While nano data centers are motivated by the energy consumption problem [11], current research reveals that the advantage of the nano data centers in energy efficiency relies on certain technical foundations. To give an overview of the technical challenges towards nano data center development, we first need to figure out the energy consumers of the nano data center applications.

Considering the nano data center (NaDa) platform proposed in [11], users will host tiny managed nano servers on their end-user devices such as Triple-Play gateways and DSL/cable modems, and communicate with each other following a Peer-to-Peer (P2P) philosophy. Thus, suppose a user (client) wants to access the content stored in a nano server hosted by another user (manager), the energy consumption for this process consists of [8]:

- the energy consumed by the client for requesting the content, denoted as E_{req};
- the energy consumption of the content transportation process, denoted as E_{trans}; and

• the energy consumed by the manager for storing the content and processing the request, denoted as E_{serv}.

If we denote the total energy consumption as E_{total}, we can derive the following formula:

$$E_{\text{total}} = E_{\text{req}} + E_{\text{trans}} + E_{\text{serv}}. \tag{1}$$

To detail the formula from technical aspects, we need to understand the internet protocol (IP) network. [2] modeled the IP network as the combination of three domains: the access network, the metropolitan and edge networks and the core network. For centralized data center applications, [2] visualized the IP network model as shown in Figure 1. The access network connects each end-user to the metropolitan and edge network, which serves as the interfaces to the core network. Centralized data centers are usually directly connected with the core network, but for nano data centers, since nano servers are hosted in end-user devices, the data has to traverse the access network twice [7].
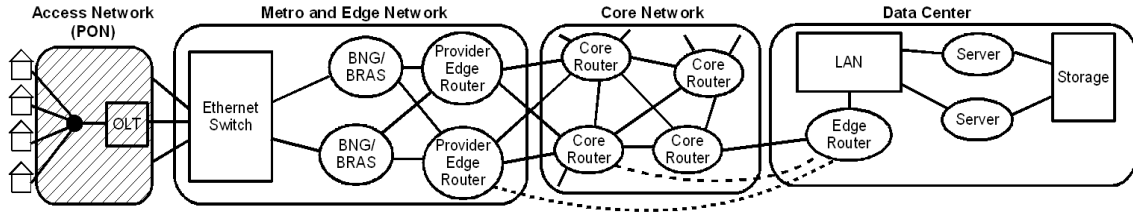

Fig. 1. IPTV network model for centralized data centers [2]

We can now detail formula (1) and adapt it to the energy consumption model proposed in [8] as the following:

$$E_{\text{req}} = E_c + E_{\text{access}}, \tag{2}$$

$$E_{\text{trans}} = E_{\text{edge}} \cdot h_{\text{edge}} + E_{\text{core}} \cdot h_{\text{core}}, \tag{3}$$

$$E_{\text{serv}} = E_{\text{access}} + E_m, \tag{4}$$

where E_c represents the energy consumed in the end-user device of the client, E_{access}, E_{edge}, and E_{core} represent the energy consumed in the access network, edge network and core network, respectively, h_{edge} and h_{core} represent the number of hops in the edge and core networks, and E_m represents the energy consumed in the end-user device of the manager.

In the following, we will introduce three technical challenges that we regard as fundamental and argument our selections with respect to the proposed energy consumption model. These challenges are: the activation of nano servers; the selection of the access network that the nano servers are attached to; and the trade-off between the distance among nano servers and the number of data replications.

3.3.2 Activation of Nano Servers. A nano data center platform is constructed with end-user devices as nano servers. We refer a nano server as *active*, when it is on and the user is accessing the nano data center service, and we refer a nano server as *idle*, when it is on but the user is not accessing the nano data center service. Whenever a nano server is on, no matters it is active or idle, it consumes energy. [11] proposed a thorough study on how the activation status of nano servers affects the total energy consumption. [11] denoted the active time of a nano server as t_{act} and the idle time of the nano server as t_{idle}, and introduced a coefficient R that represents the ratio of the active time of nano servers to the whole duration when nano servers are on:

$$R = \frac{t_{\text{act}}}{t_{\text{act}} + t_{\text{idle}}}. \tag{5}$$

As proposed in [11], R is involved in the calculation of both E_c and E_m in the energy consumption model. As a result, the energy consumption of nano data centers correlates with R as shown in Figure 2, where five different active time ratios (0.01, 0.05, 0.2, 0.5, 1) are chosen for comparison, and the energy consumption of a centralized data center is also shown as a reference. We can see that the energy consumption of nano servers increases as the active time ratio decreases, and when the ratio is large than 0.2, the energy consumption of nano servers surpasses the energy consumption of the centralized data center, i.e. the nano data center becomes less energy efficient.
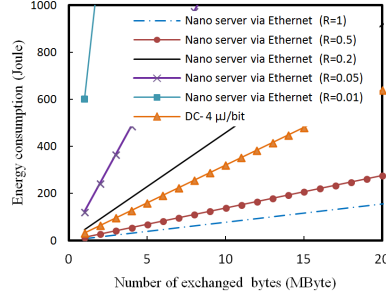

Fig. 2. Energy consumption of nano servers with different active time ratio, and of a centralized data center [11]

.

A disappointing fact is that we cannot simply assume the active time ratio R to be higher than 0.2 most of the time. Taking the widely-used video delivered by Internet protocol (IPTV) as an example: according to [11] and [4], the IPTV user activity shows large variation throughout the day, as shown in Figure 3. Even the in the peak hour, fewer than 20% of customers are active (i.e. $R < 0.2$); as for in the midnight, less than 5% of customers are active (i.e. $R < 0.05$). On average, the active ratio R is around 0.07, which means that if all nano servers are on in the whole day, it is energy **in**efficient to apply nano data centers.
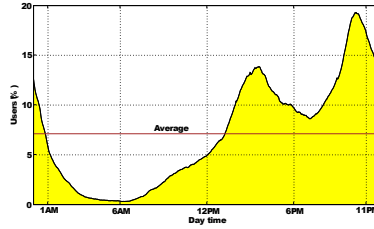

Fig. 3. User activity of the IPTV service [11]

.

Therefore, to make a nano data center platform energy-efficient, it is necessary to increase the active time ratio. From our view, a possible solution is to turn off the nano server time to time to reduce the idle time of the nano servers. But for the nano servers attached to DSL-modems or other functional devices, such as the NaDa model proposed in [11], turning off nano servers can lead to side-effects on the usage of other internet services of users, and thus the implementation of this solution requires further study. Another possible solution is to modify the applications for the nano data center, such as running multiple applications that have different peak hours, aiming to maximize the active time of the nano servers. So far we are unaware of any research tackling these

problems, and thus we list the activation of nano servers as one of the four technical challenges towards nano data center development.

3.3.3 Access Network. As discussed in Section 3.3.1, the access network connects nano servers to the edge network. According to the analysis proposed in [7] and [8], the power consumption of the access network needs to be counted twice for nano data centers, i.e. E_{access} needs to be counted both in E_{req} and E_{serv}. Thus, the energy efficiency of the access network has a large impact on the energy efficiency of the nano data center platform.

The energy consumption of the access network has been studied in [8] and [3]. Both studies refer Passive Optical Networks (PON) as energy-efficient access networks, and indicate that wireless networks (WiMAX [3], 4G, WiFi [8] are relatively energy-inefficient. Figure 4 shows a comparison of energy consumption between nano data centers using different access networks and centralized data centers with different energy consumption values, where the curves for GPON and Ethernet almost overlap, and the curves for WiFi and centralized data center with $20\mu J$/bit almost overlap. We can see that different access networks result in huge difference of energy consumption, and nano data centers attached to energy-inefficient access networks consume even more energy than centralized data centers.
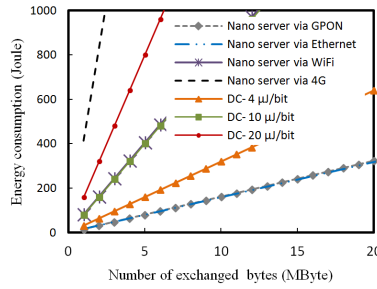

Fig. 4. A comparison of energy consumption between nano data centers using different access networks and centralized data centers with different energy consumption values [11]

.

For nano data centers, since each nano servers is possessed by an end-user, it requires further study on the access network that these end-users may or may not use, before one can determine whether the implementation of nano data centers can save energy or not.

Another concern regarding the access network of nano data centers is the trade-off between energy consumption and the access rate. Figure 5 shows their correlation from two aspects. As shown in the left figure, for a single user, the power consumption increases with the average access rate, and as shown in the right figure, the energy per bit decreases with the average access rate. For the two relatively energy-efficient access networks – Point-to-Point optical access network (PtP) and PON, their performance show different trends with the increase of the access rate. While PON is more energy efficient for a low access rate, it consumes more energy than PtP when the access rate surpasses 300Mb/s. Thus, it also requires further study on the selection of the access networks with respect to the expected access rate.

3.3.4 Trade-off: Distance and Replication. The energy consumption of the data transmission process (E_{trans}) in nano data center platforms is dependent on the number of hops in the edge (h_{edge}) and core networks (h_{core}), as mentioned in (4) in Section 3.3.1. [8] estimated the average number of hops in the edge and core networks to be 3 and 5, using *traceroute* from end-user devices to WordPress [1] servers. The number of these hops can be understood as the distance between the end-user requesting data and the end-user hosting the corresponding
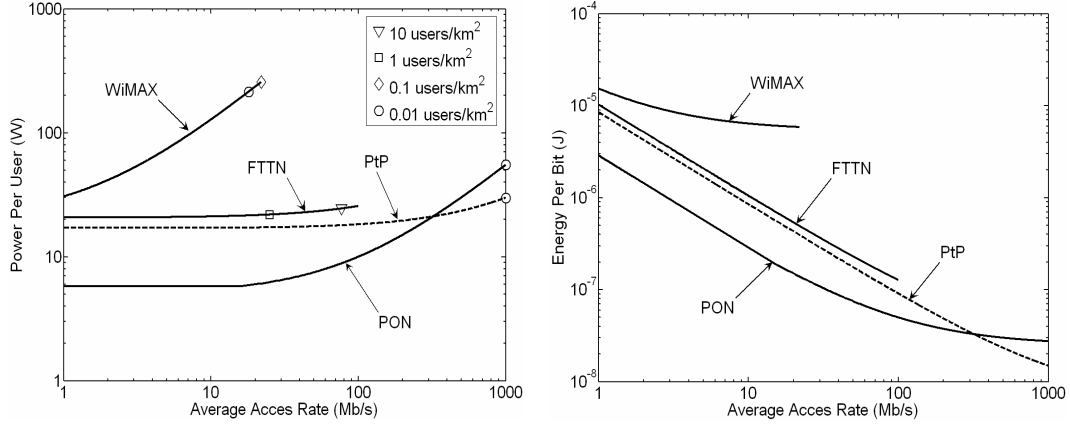
Fig. 5. Power consumption per user for different access networks and energy per bit of different access networks [3]

.

data. Regarding the location of the communicating end-users, the number of hops shows large variation: for non-local users, [8] measured h_{edge} and h_{core} as 3 and 8; and for local users, h_{edge} and h_{core} are measured as 1 and 2. Figure 6 shows the energy consumption of the the core and edge networks for the data transmission between local and non-local users, and compares them with the centralized data center. We can see that if an end-user accesses data from a non-local nano server, the energy consumption of the edge and core networks can be even higher than accessing data from a centralized data center.
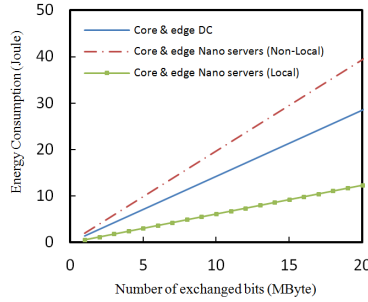

Fig. 6. Energy consumption of core and edge networks for accesing data from different locations [8]

A natural solution to reduce or balance the data transmission distance is to add data replicas [7] [2]. However, while data replication can be beneficial to the energy-efficiency in the transmission process, it also leads to the increase of the energy consumption for storing the data. [2] indicated that the data replication strategy should depend on the popularity of the data. Figure 7 shows the contribution of storage, servers and transmission to the total power consumption of IPTV services, where 20 replicas of a 2-hour SD movie are stored in 20 different data centers [2]. We can see that the storage cost dominates the total energy consumption when the movie is rarely downloaded, and the transmission cost becomes significant as the data popularity increases. Regarding this, for movies that are frequently downloaded, it is more energy-efficient to keep a relatively large number of replicas to

reduce the energy consumption in the transmission process; and for movies that are rarely downloaded, it is more energy-efficient to keep a relatively small number of replicas to reduce the energy consumption in the storage.
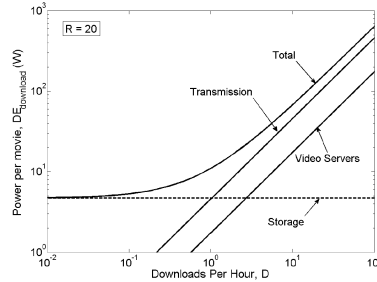

Fig. 7. The contribution of storage, servers and transmission to the total power consumption of IPTV services with respect to data popularity [2]

.

[7] compared the energy consumption for downloading movies from nano data centers and from centralized data centers with respect to the access frequency, under the scenario that a centralized data centers keeps two replicas of each data, and nano data centers keeps 2, or 10, or 100 replicas of each data. The result indicated that a centralized data center is more efficient than nano data centers for unpopular data.

Thus, we regard finding the balance between the transmission distance and the number of replicas as one the technical challenges towards the development of nano data centers. From our view, a possible solution is to detect the popularity of each individual data dynamically, and adjusting the number of replicas for each data accordingly. However, further study on the negotiation and configuration mechanism is required.

4 EVALUATION/RESULTS

5 DISCUSSION

6 CONCLUSION AND FUTURE WORK

ACKNOWLEDGMENTS

REFERENCES

[1] Automattic. 2005. WordPress.com. (2005). http://wordpress.com (Accessed on 13/01/2018).
[2] Jayant Baliga, Robert Ayre, Kerry Hinton, and Rodney S. Tucker. 2011. Architectures for Energy-Efficient IPTV Networks. *IEEE Systems Journal.* 5 (09 2011), 518–527. https://doi.org/10.1109/JSYST.2011.2165601
[3] Jayant Baliga, Robert Ayre, Wayne V. Sorin, Kerry Hinton, and Rodney S. Tucker. 2008. Energy Consumption in Access Networks. IEEE. https://doi.org/10.1109/OFC.2008.4528538
[4] Meeyoung Cha, Pablo Rodriguez, Jon Crowcroft, Sue Moon, and Xavier Amatriain. 2008. Watching television over an IP network. ACM, 71–48. https://doi.org/10.1145/1452520.1452529
[5] Council of European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). (2016). Chapter 3, Articles 12-23.
[6] S. Eum, Y. Shoji, M. Murata, and N. Nishinaga. 2015. Design and Implementation of ICN-enabled IEEE 802.11 Access Points As Nano Data Centers. *J. Netw. Comput. Appl.* 50, C (April 2015), 159–167. https://doi.org/10.1016/j.jnca.2014.07.031
[7] Anja Feldmann, Andreas Gladisch, Mario Kind, Christoph Lange, Georgios Smaragdakis, and Fritz-Joachim Westphal. 2010. Energy Trade-offs among Content Delivery Architectures. IEEE. https://doi.org/10.1109/CTTE.2010.5557700
[8] Fatemeh Jalali, Robert Ayre, Arun Vishwanath, Kerry Hinton, Tansu Alpcan, and Rodney S. Tucker. 2014. Energy Consumption of Content Distribution from Nano Data Centers versus Centralized Data Centers. *SIGMETRICS Performance Evaluation Review* 42 (2014), 49–54.

[9] P. Rodriguez N. Laoutaris and L. Massoulie. 2008. ECHOS: Edge Capacity Hosting Overlays of Nano Data Centers. *SIGCOMM Comput. Commun. Rev.* 38 (01 2008), 51–54. https://doi.org/10.1145/1341431.1341442

[10] Technicolor. 2011. Technicolor Research & Innovation - Nano Data Center. Online video. (01 2011). https://www.youtube.com/watch?v=uvkBUKqFbkI

[11] Vytautas Valancius, Nikolaos Laoutaris, Laurent Massoulié, Christophe Diot, and Pablo Rodriguez. 2009. Greening the internet with nano data centers. In *CoNEXT*. ACM, 37–48.

A   QUESTIONNAIRE

B   INTERVIEW