

Greening the Internet with Nano Data Centers

Vytautas Valancius
Georgia Tech
valas@gatech.edu

Nikolaos Laoutaris
Telefonica Research
nikos@tid.es

Laurent Massoulié
Thomson
laurent.massoulie@thomson.net

Christophe Diot
Thomson
christophe.diot@thomson.net

Pablo Rodriguez
Telefonica Research
pablorr@tid.es

ABSTRACT

Motivated by increased concern over energy consumption in modern data centers, we propose a new, distributed computing platform called Nano Data Centers (NaDa). NaDa uses ISP-controlled home gateways to provide computing and storage services and adopts a managed peer-to-peer model to form a distributed data center infrastructure. To evaluate the potential for energy savings in NaDa platform we pick Video-on-Demand (VoD) services. We develop an energy consumption model for VoD in traditional and in NaDa data centers and evaluate this model using a large set of empirical VoD access data. We find that even under the most pessimistic scenarios, NaDa saves at least 20% to 30% of the energy compared to traditional data centers. These savings stem from energy-preserving properties inherent to NaDa such as the *reuse of already committed baseline power* on underutilized gateways, the avoidance of *cooling costs*, and the reduction of *network energy consumption* as a result of demand and service co-localization in NaDa.

Categories and Subject Descriptors

C.2.4 [Computer Communication Networks]: Distributed Systems—Distributed Applications

General Terms

Design, Management, Measurement

Keywords

Energy Efficiency, Data Centers, Nano Data Centers, Video Streaming

1. INTRODUCTION

Most current Internet applications are served from a large number of collocated servers stacked together in one of multiple *data center* facilities around the world. This centralized hosting model is a classic example of the economies of scale: large numbers of similar servers yields relatively low manning requirements and eases

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CoNEXT'09, December 1–4, 2009, Rome, Italy.

Copyright 2009 ACM 978-1-60558-636-6/09/12 ...\$10.00.

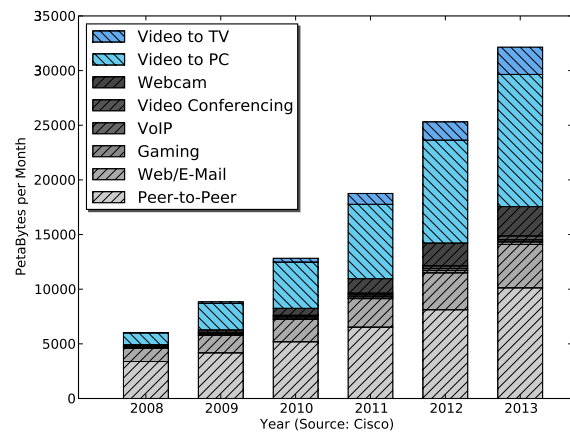


Figure 1: Growth of video content in the Internet.

procurement procedures. Homogeneous hosting environments allow for better resource optimization and management, while extensive use of virtualization technologies provide an abstraction of dedicated platforms to developers.

Centralization trend is not without its limitations. Data centers are prone to: 1) over-provisioning, 2) high cost of heat dissipation, and 3) increased distance to end-users. All of these issues, as we will see later, lead to ever increasing energy bills that the data center and network operators need to foot. The data centers are *over-provisioned* because they need to match the peak demand despite that the average load remains much lower throughout most of the day; in addition, redundancy requirements might push up the number of necessary servers significantly. Data centers also are *expensive to cool*. Despite significant efforts for improving the server power-efficiency, an average data center spends as much energy on cooling as it spends for powering its servers [23]. Even in the data centers using state-of-the-art cooling technologies heat dissipation accounts for at least 20% to 50% of the total power consumption [4]. Centralization trend also increases the data center *distance to the users*. Not every service can be hosted at well connected central interconnection points. Higher distance from end users increases bandwidth-mileage requirements and adds to the energy consumption of the networking equipment. It is not surprising therefore that data centers recently made headlines with reports that they consume 1.5% of total US electricity [3], or that their carbon-dioxide emissions are projected to surpass those of the airline industry by the year 2020 [6]. These concerns are indeed

expected to amplify in view of growth projections of data center-hosted applications like video distribution [13] (Figure 1).

While data centers are an example of centralization trend, another, opposite trend manifests through Peer-to-Peer (P2P) systems. The decentralized P2P systems free-ride on end-user computers and, being spatially distributed, incur little or no heat dissipation costs. In addition, some P2P systems can exploit locality and reduce the redundancy requirements, because each P2P user often is P2P server at the same time. Unfortunately, conventional P2P system performance typically suffers from free-riding, node churn, and lack of awareness of underlying network conditions.

Motivated by the problem of energy consumption, in this paper we propose a new way to deliver Internet services based on our *Nano Data Center* (NaDa) platform. The key idea behind NaDa is to create a distributed service platform based on tiny managed “servers” located at the edges of the network. In NaDa, both the nano servers and access bandwidth to those servers are controlled and managed by a single entity (typically an ISP) similarly to what is suggested in [20, 27]. Significant opportunities already exist for hosting such tiny servers on ISP owned devices like Triple-Play gateways and DSL/cable modems that sit behind standard broadband accesses. Such *gateways* form the core of the NaDa platform and, in theory, can host many of the Internet services currently hosted in the data centers. In this paper, however, we will focus on video streaming services, which, as shown in Figure 1, exhibited large growth in past few years. In addition, video streaming services allow for easier energy usage modeling and simulation. Figure 2 shows a high level architecture of such streaming architecture using home gateways as nano servers.

NaDa follows a P2P philosophy, but contrary to typical P2P, NaDa is *coordinated* and *managed* by an ISP that installs and runs the gateways that act as nano servers, as well as the network that interconnects them. Due to its managed nature, NaDa avoids most of the shortcomings of classic unmanaged P2P. ISPs can easily implement NaDa by providing new customers with slightly over dimensioned gateways, whose extra storage and bandwidth resources would be used by NaDa to implement services like video hosting, all of which will be totally isolated from the end-user via virtualization technologies. Thus, with a rather small investment in higher capacity devices, ISPs can go beyond their traditional role as communication carriers, and enter the potentially highly profitable service and content hosting market.

Obviously designing a system like NaDa spans a magnitude of issues that cannot all be covered here and thus in this paper we mainly focus on energy efficiency which, as discussed earlier, is probably the Achilles heel of monolithic data centers. As we will demonstrate in coming sections, NaDa can reduce the energy costs of offering Internet services, particularly content distribution. The savings come in three ways. First, through improved *heat dissipation* stemming from the distributed nature of NaDa which minimizes the need for cooling. Second, through *traffic localization* resulting from the co-location between NaDa and end users. Finally, NaDa performs *efficient energy use* by running on gateways that are already powered up and utilized for content access or other services. NaDa avoids wasting the high baseline power already paid for in an online gateway. We elaborate further on these energy saving aspects in Section 2.

Our main contributions are the following. We develop a model to evaluate the energy needed to provide services in both centralized data centers and distributed NaDa. This model relies on a large collection of empirical data from various sources, including power measurements of operational content distribution servers and end-user devices (Section 3). We then apply this model in the evalua-

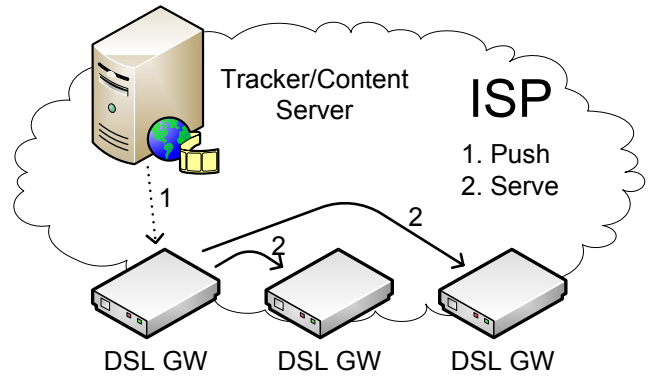


Figure 2: High level NaDa architecture. Content is served from home gateways whenever possible.

tion of energy consumption in the context of video-on-demand. We detail NaDa platform and placement mechanisms for video content (Section 4), and use trace-driven simulations to quantify energy savings under various interesting configurations. We find that even under the most pessimistic conditions NaDa, achieves at least 15% to 30% overall energy savings compared to traditional data centers, while more optimistic scenarios achieve up to 60% greater energy efficiency (Section 5). Related work and final remarks are discussed in Sections 6 and 7 respectively.

2. THE CASE FOR NADA

In this section we present a more detailed argument on why NaDa platform can be more energy efficient than conventional data centers and why it is certainly feasible to build it. We first detail the energy reduction principles of NaDa. We then quantify the availability of the end devices to be used as NaDa nano nodes.

2.1 Advantages of NaDa

Four main principles allow NaDa’s to surpass the energy efficiency of conventional data centers:

Heat Dissipation. Recent reports [4, 19] show that data centers consume large amounts of energy for energy transmission/conversion, and most importantly, cooling and heat dissipation. Such overheads are captured by the Power Usage Efficiency (PUE) metric, which is defined as the ratio between the total power consumed by a data center and the power actually delivered to its IT equipment (servers, networking equipment). The PUE factor ranges from as high as 2.0 in legacy data centers [23] to as low as 1.2 in recent state of the art facilities [4]. Operators are trying to improve their PUE by leveraging advanced scheduling [26], performing heat reuse [22], adopting free flow systems, and even exploiting colder geographical locations [12]. All these approaches, however, have limitations. Scheduling, heat reuse and advanced cooling systems save only a fraction of energy costs; placing data centers in remote locations sacrifices proximity to users and speed of service delivery. Finally, and most important, heat dissipation is hard to tackle due to the high density of co-located IT equipment, which is an inherent characteristic of monolithic data centers.

Church et al. [12] have recently proposed de-densifying IT equipment by spreading them among container sized data centers. NaDa pushes this idea to its extreme by distributing data center IT equipment to user premises where heat dissipation costs are negligible. We show that the additional energy consumed by a single gateway is negligible and thus easy to dissipate.

