2013 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing

# Extending the Capabilities of Mobile Devices for Online Social Applications through Cloud Offloading

Alexandru-Corneliu Olteanu
University Politehnica of Bucharest,
Romania
Email: alexandru.olteanu@cs.pub.ro

Nicolae Ţăpuş
University Politehnica of Bucharest,
Romania
(*Supervisor*)

Alexandru Iosup
Delft University of Technology,
The Netherlands
(*Supervisor*)

***Abstract*—Handheld devices are becoming an attractive option for users to interact with their social network, through online social applications. We are witnessing a rapid adoption of smarter devices all around us, which brings with it orders of magnitude in heterogeneity. Thus, researchers in the field of distributed systems are faced with new challenges: How to optimize performance for devices that are so diverse in terms of energy consumption, processing power and communication capabilities? My PhD research focuses on this challenge, adopting techniques for offloading operations from mobile to more powerful cloud-based infrastructure, and brings a three-fold contribution. *First*, we have characterized and modeled workloads of online social applications, and empirically validated them using traces of hundreds of real applications. *Second*, we are currently investigating offloading mechanisms, including: communication offloading, lossy performance offloading, and lossless performance offloading. We have been testing and evaluating these mechanisms with several mobile applications, measuring performance and energy consumption. *Third*, we will create an integrated cloud-based offloading system that aims to improve the performance of online social applications. We will empirically evaluate this system using both simulations and open-source real-world applications.**

## I. Problem Statement

Modern handheld devices, such as smartphones and tablets, offer portability, increased computational power, and communication capabilities. Thus, they are becoming an attractive option for users to interact with each other, through social applications, and with their environment, through home automation. Facebook, who has anounced recently their increase to over 1 billion monthly active users, reports that more than a half of their users reach their social network using a mobile device [1]. Although mobile devices are growing in functionality and computing power, we believe the role of more powerful infrastructure, to augment the capabilities of mobiles, will increase. Thus, my PhD research focuses in general on techniques for offloading operations from mobile to more powerful cloud-based infrastructure.

The convergence of mobile and grid computing has been studied for a number of years, with results in job scheduling [2][3] and resource discovery [4][5]. However, the idea of applying this research specifically to online social applications is *new*. Online social applications are applications dedicated to socially interconnected users, developed for various purposes, such as gaming, multimedia streaming, travel, communication, etc. The way the users are socially related influences the workload, as the users can alter the way their friends interact with the application. Thus, the *first challenge* we are facing is to analyse user behavior and workload in the context of user social interactions.

Offloading has been attracting some research effort in the past few years, also in the CCGrid community [6][7]. Software

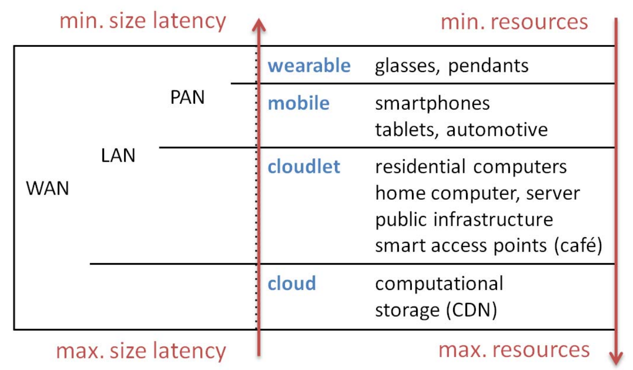


Fig. 1. Hierarchical view of computer systems that people are using today, based on the way they are interconnected: using *Personal Area Networks* (PAN), *Local Area Networks* (LAN),*Wide Area Networks* (WAN). Trade-offs are highlighted.

that uses the interaction of mobile devices with the cloud is already on the market. However, recent research materials have identified the cloudlet as an offloading target [8], emphasizing the trade-off between communication and computation costs. Inspired by this research, we see mobile devices as part of a hierarchy of computing systems people are using today, which is essentially comprised of wearable devices, handheld devices, cloudlets, and clouds (see Figure 1). Given the rate at which smart devices become smaller and closer to the user, we find it foreseable that in the near future mobile devices will serve not as sources, but as targets for offloading, from smaller, wearable devices. The *second challenge* we are facing is to decide how to conduct offloading in online social applications.

The recent popularity of smart mobile devices is motivating many manufacturers to produce devices for many types of consumers, thus leading to orders of magnitude in heterogeneity. The processing unit may vary from single-core CPU, to quad-core CPU, and even to hybrid architectures that include a dual-core CPU and a GPU. The battery lifetime may vary from tens to hundreds of hours in standby, and is greatly influenced by user behavior, as intense device usage can reduce its battery life to barely a few hours. The *third challenge* we are facing is to assess the benefits of offloading under such heterogeneity.

My PhD research focuses on ways to extend the capabilities of mobile devices, to improve the way they interact with other devices, with cloudlets, and the cloud. We empirically measure how well this improvement matches the requirements of several types of mobile applications, in terms of performance and energy consumption. This paper describes the status of my PhD research and a plan for the next year.

978-0-7695-4996-5/13 $26.00 © 2013 IEEE
DOI 10.1109/CCGrid.2013.52

TABLE I. TIMELINE FOR MY PHD RESEARCH OVER THE NEXT YEAR.

| | Month 1-3 | Month 4-6 | Month 7-9 | Month 10-12 |
|---|---|---|---|---|
| workload characterization and modeling | - characterization and modeling of user behavior<br>- more validation on extended data set<br>- collect data for mobile specific workloads | - characterization and modeling using mobile specific workloads | - model validation using tune-up application | - wrap-up thesis |
| offloading mechanisms | - finish offloading survey | - experimental evaluation for lossless performance offloading | - | - wrap-up thesis |
| integrated offloading-based system | - finish system design<br>- tests with open-source app (OpenTTD) | - implement performance benchmark<br>- experimental evaluation | - test different scheduling and allocation policies<br>- experimental evaluation | - wrap-up thesis |

## II. PROPOSED RESEARCH

Given the challenges identified in Section I, we investigate several research questions:

1) How does the number of active users evolve over time for online social applications?
2) How to decide, online and efficiently, what, when, where, and how to offload?
3) How to measure the improvement that a mix of offloading mechanisms can bring for various devices?

In the remainder of this section, we describe the current status of our work, and the future development as summarized in Table I. Our research has a strong experimental focus, which we present in Section III.

### A. Research Overview

To answer the first research question, we have been conducting *workload characterization and modeling* for online social applications. Several studies about the workloads of online social applications exist [9]. However, in our work we use traces for a larger number of applications, over longer periods of time. Moreover, we try to extend our research over more types of online social applications [10], which are specifically designed for mobile devices. This work is detailed in Section II-B.

To understand offloading, we have been investigating *offloading mechanisms* in the context of several mobile applications. Computation and communication offloading have been investigated in several works in the past few years [11][12][13]. We have proposed new offloading techniques, such as offloading communication to custom hardware extensions, which provide communication protocols that are not embedded into off-the-shelf smartphones. We have analysed several types of applications, such as video and image processing applications [14][15], which raise new opportunities for offloading, that are detailed in Section II-C.

The offloading investigation helps answering the last two research questions. However, to fully answer them, we will develop an *integrated system* to test offloading, scheduling and allocation, in the context of a real-world application. We give an overview of this system in Section II-D.

### B. Workload Characterization and Modeling

We characterize and model the workload of online social applications using an empirical approach. This analysis helps to understand the user behavior and to propose a system design that maximises performance and minimises costs. We will also use the model to generate workloads (Section III-D).

Due to business considerations, most of the private companies that produce or host applications with a large number of users, do not want to share them with researchers, and even less to make them public. Thus, obtaining device specific traces can be difficult.

We have started our characterization efforts with a data-driven analysis of traces from hundreds of applications hosted by Facebook. Such traces are publicly available on third-party websites, from which we have collected data over a period of almost three years, since 2010. We have characterized and modeled two components: the popularity distribution, which describes how applications attract users, and the evolution model, which describes how users evolve over time for any given application. We find that the most popular applications gather many more users even than the ones ranked immediatly below them. Also, the evolution of the number of users follows a pattern: it increases steadily up to a peak, and then steadily decays. The traces we have collected for this analysis are currently being imported in the Game Trace Archive [16], for public access. We are also extending the workload model with a user behavior graph, with a focus on social activities that might impact performance, such as visiting a neighbour in online social games.

We have conducted a use-case based analysis on a custom collaborative learning application. We have used our computer cluster [17] to test the performance in use-cases addressed to a group of 10-15 users [10]. We plan to further develop our application and to integrate it with Moodle, the course management system that University Politehnica of Bucharest uses, to better understand the user behavior and to collect more workload traces.

Recently, we have reached an agreement with a company that develops a mobile tune-up application, which has tens of thousands of users. The application currently makes simple statistics over the system usage and makes recommendations to the user. We plan to aggregate the information in the cloud and to use statistical methods to make better recommendations. These findings could also prove useful in developing a method to dynamically decide what and when to offload.

### C. Offloading Mechanisms

Based on the findings in Section II-B, we are investigating several offloading mechanisms, which we divide into: *communication offloading*, *lossy-performance offloading*, and *lossless-performance offloading*.

Communication offloading has been described as the offload of communication intensive polling to a cloud resource, which assumes the polling function and pushes information to the device only when needed [13]. We have proposed an extension: offloading communication to custom hardware extensions, which provide communication protocols that are not embedded into off-the-shelf smartphones. We have implemented such methods in two mobile applications, for home automation [18] and for pollution tracking [19]. For example, the home automation application is designed to control a home automation system. Usually networks of devices for home automation use low energy, low bit-rate communication protocols, such as ZigBee. Off-the-shelf mobile devices can interact with a device in the network by connecting through WiFi or 3G at a gateway, which mediates communication between the network and the Internet. We have used an USB dongle, capable of ZigBee communication, to enable a second communication channel, through which the mobile device can connect directly to a network device, if the user is at home. We have compared the two communication channels and we have proposed that the ZigBee communication channel may be preferred for small packages, due to the energy communication patterns. This work is currently under publication [18]. The pollution tracking application polls a Bluetooth dongle with air quality sensors. We have analysed the energy consumption for polling the sensor dongle and we have proposed an adaptive query algorithm for location oriented applications [19].

By lossy-performance offloading we denote a mechanism that compares if it is better to offload processing some data, or simply to skip processing that data. This type of mechanism can be applied in any application where response time is more important than precision. We have analysed frame-skipping in two video-processing applications, for tracker-based augmented reality [14], and for surveillance using smartphones [15]. The augmented reality application overlays media, such as images and videos, on top of bar-codes that it detects in the camera capture stream. The processing is done in a pipeline, using separate threads. Buffers between the stages of the pipeline enable frame-skipping, which prevents lags in the response on the screen. The same mechanism is implemented in the surveillance application, in which frame-skipping is implemented to yield better detection rates.

We plan to investigate lossless-performance offloading, as a complement to lossy-performance offloading. We plan to further analyse applications that do processing in a pipeline, if input data cannot be skipped. This type of applications can provide an added value to current research efforts, because they can be partitioned in more ways. At run-time, depending on dynamic conditions, our algorithms will decide which way the processing should be split among the mobile device and the offloading target.

### D. Integrated System

We plan to develop an integrated system to test offloading, scheduling and allocation, in the context of a real-world application. In its implementation, we will use the findings regarding offloading mechanisms (see Section II-C) and regarding workloads (see Section II-B). We will investigate the tradeoffs in tuning the system for improvements in performance, energy consumption, or costs, and we will investigate the impact that cloud specificities, such as multi-tenancy, might have.

As a test application with many users, we are considering *OpenTTD*, a popular open-source simulation game. In multiplayer mode, OpenTTD currently supports up to 255 simultaneous users on the same map, one of which must host the server. We plan to investigate several use-cases, in which the game is either hosted on a mobile device, or in the cloud, and in which all the clients are in the same Local Area Network, or dispersed in multiple networks.

We will also develop a system-wide scheduler, that will take run-time decisions about what, where, and when to offload, based on the findigs from studying the tune-up application (Section II-B).

## III. Empirical Evaluation

In line with the CCGrid community, we believe experimental research has the potential to close the gap between theory and practice. In this section, we describe the empirical evaluation of the three components of the thesis, which provides validation and mathematical verification of the conceptual contributions.

### A. Overview

We are conducting empirical evaluation, that is, testing in real environments, for all the solutions we propose. We have *validated the workload model* presented in Section II-B (see Section III-B). We are conducting *energy and performance measurements* for the offloading mechanisms presented in Section II-C (see Section III-C). We will conduct *evaluation of the integrated system* both with simulated and real-world workloads (see Section III-D).

We are continuously trying to find ways to address the threats to validity that our empirical findings might have, in terms of repeatability and user base coverage. For example, we try to collect even more data regarding the applications we used for our workload model and to use real-world applications.

### B. Validation of the workload model

We have validated the results of our characterization and modeling of online social apps (see Section II-B), with a dataset that describes 630 apps over thirty one months, that we have collected from third-party websites. We find that the Log-Normal distribution fits best the popularity distribution and that its shape does not change much in the thirty-one months. For evaluating the evolution model, we could use only 53 apps from our dataset, the ones for which we have full lifetimes, and we find [20] distributions that characterize each of the evolution model's components.

In the next three months, after we analyse the traces collected from the tune-up application, we plan to improve our model adding device specific model components. The amount of users suggests there will be enough data for modeling and evaluation.

### C. Energy and performance measurements

Various methods have been proposed for energy and performance measurements on mobile devices [21] [22]. We have used the Android API to read some of the data directly, such as current consumption, number of bytes transmitted and received through WiFi, etc. We have approximated other data, such as the voltage of the WiFi module, to estimate indicators such as the energy consumption.

We have conducted experimental evaluation of the communication offloading mechanism on two Android applications described in Section II-C. Using the home automation application, we have investigated energy consumption for communication, expressed as energy per byte, and indicated [18] that communication using WiFi can be orders of magnitude more inefficient for small commands, like switching the lights, than for larger commands, like transferring photos or synchronizing the local database. In the pollution tracking application, we show [19] that using the adaptive query algorithm to gather data from the Bluetooth dongle is up to 20 times more energy efficient than traditional polling techniques.

We have also conducted an experimental evaluation of the lossy-performance offloading mechanism, using two video-processing applications (described in Section II-C). For the augmented reality application, we have indicated [14] that the frames that the application processes are around 25-50% of the total frames obtained from camera capture and that the loss increases when the complexity of the overlayed image gets higher. For the surveillence application we have shown [15] that the framerate is proportional with the CPU frequency and is not affected by the number of cores.

For the lossless-performance offloading mechanism, we plan to use the same energy estimation method. Furthermore, we will need to find a way to express the performance and the quality expected by the user.

### D. Evaluation of the integrated system

The integrated system presented in Section II-D, adapted for OpenTTD, will be tested using both simulated and real-workloads. For simulating workloads, based on the findings in Section II-B, we will use a private cloud infrastructure, obtained by deploying OpenNebula on top of two grid based systems: DAS-4 in the Netherlands [23], and NCIT in Romania [17]. We also plan to test our system using a public cloud infrastructure, such as Amazon Web Services.

Ideally we need evaluate our implementation on a large number of real users. However, if we fail to attract a sufficient number of OpenTTD users, we also plan to test a mobile learning application dedicated to the students of the UPB university in Romania. The mobile learning application, as described in Section II-B, could be built around Moodle, the course management system that UPB uses, and could address up to 2,000 potential users.

## IV. Summary

The main contributions of my PhD are:

1) We have collected traces of online social applications;
2) We have proposed a workload model for online social applications with two components: popularity distribution and evolution model;
3) We have proposed a new way to do communication offloading, using custom hardware extensions, which provide communication protocols that are not embedded into off-the-shelf smartphones;
4) We have investigated lossy-performance offloading in two applications based on video processing;
5) We have conducted empirical evaluation on all the components of the thesis developed until now;

We plan the following contributions during the next year:

1) We will extend our workload model to include user behavior in online social applications;
2) We will collect device-specific traces from a tune-up application, and we will use them to improve the workload model and investigate a benchmark for dynamic decisions on offloading;
3) We will investigate lossless-offloading, with a focus on apps with pipelined processing;
4) We will implement our integrated offloading system, which also leverages scheduling and allocation to improve performance, and we will conduct experimental evaluation using OpenTTD;
5) We will address threats to validity that our empirical findings might have, in terms of repeatability and user base coverage.

## Acknowledgments

I would like to thank Nicolae Ţăpuş and Alexandru Iosup. I would also like to thank my colleagues and collaborators from University Politehnica of Bucharest, from Delft University of Technology, especially Siqi Shen and Boxun Zhang, and from the University of Applied Sciences in Dresden. This work is funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family, and Social Protection through the Financial Agreement POSDRU/107/1.5/S/76909 and by the STW/NWO Veni grant 11881.

## References

[1] Facebook reaches one billion users. [Online]. Available: http://money.cnn.com/2012/10/04/technology/facebook-billion-users

[2] P. Ghosh, N. Roy, and S. Das, "Mobility-aware efficient job scheduling in mobile grids," in *CCGrid*. IEEE, 2007.

[3] K. Hummel and G. Jelleschitz, "A robust decentralized job scheduling approach for mobile peers in ad-hoc grids," in *CCGrid*. IEEE, 2007.

[4] D. Bruneo, M. Scarpa, A. Zaia, and A. Puliafito, "Communication paradigms for mobile grid users," in *CCGrid*. IEEE, 2003.

[5] A. Gomes, A. Ziviani, L. Lima, and M. Endler, "Dichotomy: A resource discovery and scheduling protocol for multihop ad hoc mobile grids," in *CCGrid*. IEEE, 2003.

[6] H. Chen, Y. Lin, and C. Cheng, "Coca: Computation offload to clouds using aop," in *CCGrid*. IEEE, 2012.

[7] K. Sinha and M. Kulkarni, "Techniques for fine-grained, multi-site computation offloading," in *CCGrid*. IEEE, 2011.

[8] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The case for vm-based cloudlets in mobile computing," *IEEE Perv. Comp.*, 2009.

[9] A. Nazir, S. Raza, and C. Chuah, "Unveiling facebook: a measurement study of social network based applications," in *IMC*, 2008.

[10] V. Zamfirache, A. C. Olteanu, and N. Ţăpuş, "Collaborative learning assistant for android," in *RoEduNet*, 2013, in publication. Available: http://bit.ly/11cQ9u3.

[11] G. Chen, B. Kang, M. Kandemir, N. Vijaykrishnan, M. Irwin, and R. Chandramouli, "Studying energy trade offs in offloading computation/compilation in java-enabled mobile devices," *IEEE TPDS*, 2004.

[12] H. Eom, P. St Juste, R. Figueiredo, O. Tickoo, R. Illikkal, and R. Iyer, "Snarf: a social networking-inspired accelerator remoting framework," in *MCC*. ACM, 2012.

[13] R. Kemp, N. Palmer, T. Kielmann, and H. Bal, "Energy efficient information monitoring applications on smartphones through communication offloading," *Mobicase*, 2012.

[14] A. Gherghina, A. C. Olteanu, and N. Ţăpuş, "A marker-based augmented reality system for mobile devices," in *RoEduNet*, 2013, in publication. Available: http://bit.ly/XXPLYV.

[15] A. Făsui, A. C. Olteanu, and N. Ţăpuş, "Fault tolerant surveillance system based on a network of mobile devices," in *RoEduNet*, 2013, in publication. Available: http://bit.ly/WCOr0K.

[16] Y. Guo and A. Iosup, "The game trace archive," in *NetGames*, 2012.

[17] NCIT-Cluster. [Online]. Available: https://cluster.grid.pub.ro/

[18] A. C. Olteanu, G. D. Oprina, N. Ţăpuş, and S. Zeisberg, "Enabling mobile devices for home automation using zigbee," 2013, in publication. Available: http://bit.ly/YJfGJZ.

[19] D. O. Rizea, D. Ş. Tudose, A. C. Olteanu, and N. Ţăpuş, "Adaptive query algorithm for location oriented applications," in *RoEduNet*, 2013, in publication. Available: http://bit.ly/WGXRbg.

[20] A. Olteanu, A. Iosup, and N. Ţăpuş, "Towards a workload model for online social applications: Extended report," TU Delft, Tech.Rep. PDS-2013-003, January 2013. [Online]. Available: http://bit.ly/VzTei2

[21] A. Carroll and G. Heiser, "An analysis of power consumption in a smartphone," in *USENIX*, 2010.

[22] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy consumption in mobile phones: a measurement study and implications for network applications," in *IMC*. ACM, 2009.

[23] DAS-4. [Online]. Available: http://www.cs.vu.nl/das4/