CSR:Small:Storage Architecture for the Next Generation of Smart Mobile Platforms

NSF Program 11-555

Activity Category: CSR
Award number: CNS-1218520

Umakishore Ramachandran
Moinuddin K. Qureshi

College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280
Phone: (404) 894-5136
FAX: (404) 385-2295
e-mail: rama@cc.gatech.edu
WWW URL: http://www.cc.gatech.edu/~rama

Annual Report via Fastlane August 12, 2014

1 Goals of the Project

(Note: Goals of the project are Unchanged from 2013 Annual Report. This section is reproduced here for completeness of this report.)

Mobile computing is sweeping the world. According to recent statistics, there are up to 4 billion mobile users (more than half of the world's population), with accelerated penetration in regions such as India and Africa. This trend is even more significant in the light of predictions about mobile devices dominating most personal computing landscape in the near future. Recent studies are pointing to the permanent storage in such devices being a weak link. Due to considerations of power and physical footprint, flash has been the storage technology of choice hitherto in mobile platforms. Newer storage technologies such as PCM, and STT-RAM are on the horizon as a competitor to flash. Each of these technologies have their own idiosyncracies both from the point of view of architecting them (from a hardware perspective) and integrating them in the operating system (OS) stack (from a system software perspective). It is time to take a critical look at the OS software stack and architect it to plan for future evolutions of the storage technologies. In this proposal, we focus on the system software and hardware issues, in an integrated fashion, in the construction of high-performance storage architecture for mobile platforms.

The goals of the project are summarized below:

- We will collect traces of applications (both current and futuristic) for mobile platforms and conduct detailed storage-centric measurements of their performance on current and emerging mobile storage technologies.
- We will analyze the source of storage-related performance bottlenecks on mobile platforms at different levels of the software stack (application, system software, and hardware) using a combination of measurement, simulation, and real implementation on existing and emerging mobile platforms.

- We will propose solutions both at the system software and architecture levels, in an integrated manner, for advancing the state-of-the-art in storage for mobile platforms.
- We will conduct detailed quantitative evaluation of our solutions both via simulation and implementation of research prototypes. In addition to validating our research ideas, this effort in of itself may lead to new insights on how to perform quantitative studies that are closer to real implementation.

2 Activities and Results

2.1 Research and Education

We document below the research accomplishments in the second year of the project (June 1, 2013 to July 31, 2014).

2.1.1 NSF I-Corps Investigation - Multi-tiered Storage

Storage architecture of high-end servers are interesting to study to understand the role played by Flash memory in the storage hierarchy. It will also give us insight into architecting the storage of Smartphones. Video is the dominant workload on the Internet. Smartphones have become the *de facto* client devices for viewing video. One of the research questions we have been studying is the exploration the architecture of multi-tiered storage that serve the video to the clients. A paper entitled, "*FlashStream: A multi-tiered storage architecture for HTTP video streaming incorporating MLC Flash Memory*", appeared in **ACM Multimedia 2013** [1].

Following the academic success of *Flashstream*, we embarked on an NSF I-Corps expedition [2] to identify the market potential for multi-tiered storage in data centers and content distribution networks that are increasingly catering to the demands of video streaming from clients. The intuition behind this expedition can be summarized as follows.

Web content is experiencing an explosive growth. Further, due to the increasing video traffic on the web, backend storage servers not only need increased storage capacity but also increased storage bandwidth to serve the web clients in a timely manner. Over time, hard disk drives (HDDs) have advanced in capacity per dollar while main memory (in the form of DRAM) has advanced in speed per dollar. This trend has been ongoing for a long time and given the nature of the two technologies it will likely continue to perpetuity. Flash memory technology is getting attention from industry due to its significant cost decrease in the past few years. Flash-based Solid-State Drives (SSDs) are now at the price point where it can bridge the capacity-speed gap between HDDs and DRAM. In addition, its energy consumption is far less than the two storage technologies. Therefore, multi-tiered storage architecture, is an ideal vehicle to optimize capacity, speed, and energy consumption of storage by exploiting the advantages of the three technologies. With the explosive consumption growth of static digital content (e.g., text, image, sound, video, etc.) on the Internet, service providers (e.g., Facebook, Flicker, YouTube, Hulu, Netflix, etc.) are facing huge I/O demand for serving their contents. Our business model is plug-in software for popular web servers such as Apache and Nginx so that CDN providers can adopt it without modifying their infrastructure. Through the I-Corps program, we planned to discover the needs and potential for commercialization of low-energy high-performance multi-tiered storage technology for web content.

Web storage is an ideal application for flash memory because the web content access is mostly read-only. Especially, with the rapid growth of video content on the web, this fact is becoming even more amplified. Our multi-tiered web storage technology exploits the read-only characteristic of web storage, delays write operations to flash memory for caching hot web contents, and transforms the random writes to sequential writes with a very small amount of in-device DRAM buffer. The upshot is a significant energy reduction that can be achieved by incorporating low-end flash memory in our proposed storage architecture. We expect such a solution to be attractive for architecting the storage infrastructure of large data centers. We explored the potential for commercialization of this technology by interviewing CDN service providers, web service providers, and other stake holders. The NSF I-Corps team consisted of the PI Ramachandran, Mr. MK Ryu (student whose research through this NSF grant, CNS-1218520, led to the I-Corps investigation), and Mr. Gareth Genner (a serial entrepreneur in the Atlanta area) who served as the business mentor.

The upshot of the investigation is the formation of an LLC *flashboost-io* [3] and we are currently in the process of validating the performance advantage of our technology with field trials at customer sites including

CNN and StorTrends. MK Ryu has since graduated [4] and has accepted a position with Google. He will continue to direct the validation studies for this technology together with the PI Ramachandran.

2.1.2 Integrating Mobile Storage with Cloud: Cloud Cache

The motivation behind this research is to provide an illusion of infinite storage space for a user on a mobile device by seamlessly extending the limited NAND Flash storage on the mobile to the cloud storage. To this end, we have investigated the storage usage patterns in mobile devices (such as memory footprint and data characteristics; access frequency, data types, data size, etc.). We have collected the traces of file I/O operations for several applications running on mobile devices. Additionally, we have created a temporal history of user directories and files to understand locality patterns. Using a pattern-driven design, we seamlessly migrate (in both directions) data on the mobile with the cloud storage. With this seamless integration of the local storage and the cloud storage, user generated data (images, audio, video, and emails) is automatically synchronized between multiple local devices that a given user may access and the cloud [5].

2.1.3 Trading Reliability for Performance

In our 2013 annual report, we reported on our studies conducted on mobile platforms exploiting the user context to perform "informed" optimizations on storage activities and hence boost application performance. A paper documenting this approach entitled, "Fjord: Informed Storage Management for Smartphones," appeared in MSST 2013 [6].

Continuing the promising results of Fjord, we are undertaking a detailed study of several applications to understand their storage activities from the point of view of determining opportunities for relaxing reliability to gain performance. The intent is to explore the tradeoff between data resilience and performance for cloud-backed mobile applications. Due to the use of frameworks such as Android in the development of mobile apps, storage performance is critical to the application performance. Our studies have shown that frequent flush operations between kernel buffers and the storage results in negatively impacting application performance. Such flush operations are required to guarantee data fidelity in case of failures (power and software). However, depending on the criticality of the data, it may be possible to relax the tight consistency between the kernel buffers and the storage, and thus boost application performance. Therefore, the first order of business is to explore I/O access patterns of mobile apps, and the sensitivity of the applications to I/O failures.

We have instrumented Android/Linux system stack on Nexus-7 platforms to intercept the storage system calls so that we can carry out such a comprehensive study. In order to generate workload typical of a human user, and make the workload repeatable, we have set up well-known tools such as Monkeytalk or Monkeyrunner. Additionally, we have built software fault-injection features inside the Android device kernel. Using this instrumentation, we study the sensitivity of apps running on the mobile device to I/O failures. We log information system-wide ranging from application-level to the block device level using this infrastructure. This will lead to designing the appropriate knobs in the system to trade reliability for performance of cloud-backed apps running on mobile devices [7].

This study also aids the research we identified in Section 2.1.2, since the criticality of writes to data objects can inform the automatic migration policies of cloud cache. Eventually, this comprehensive study on performance-resiliency tradeoff will help us to extend the research to optimal data placement policies when the system architecture includes heterogeneous memories and storage types (e.g., DRAM, flash, and PCM).

2.1.4 Memory Virtualization for Smart Mobile Platforms

This research concerns memory management in virtualized environments, with the primary focus being on improving the latency for inter-VM memory management. This aspect is important from the point of view of evolution of computing and storage capacities of mobile hand-held devices, their increasing usage for a variety of applications (some of which may be security sensitive) and the gradual introduction of virtualization into the software stack of these mobile devices (for increasing the versatility and security of such devices). Although memory management in virtualized environments has been well-researched in the context of servers and cloud computing environments, the focus of our research is on reducing the latency for dynamic memory allocation in a mobile computing environment.

With a goal of measuring and analyzing the end-to-end latency involved in the ballooning principle, we

designed and implemented directed ballooning mechanism based on the split driver principle followed in the other drivers built for the Xen based para-virtualized setup. We also analyzed the latency involved in this mechanism and compared it with another mechanism called Transcendent Memory designed for a similar purpose. Our study showed that directed ballooning is a better approach to reducing the latency for memory allocation in a virtualized environment [8].

Building on this initial study, we have been pursuing the following research activities [9]:

1. **Scheduler enhancement**: This involves detailed analysis of the interference caused by default scheduling policies in the Xen hypervisor and the Linux Kernel in the ballooning process and the resulting effect on latency. Based on the analysis, we modified the scheduler in the Linux kernel to recognize ballooning related work items and temporarily enhance the priority of the kernel threads dealing with those work items. Similarly, a mechanism was introduced to provide higher priority to the vcpus of the VMs which are actively involved in the ballooning process. Analysis shows significant advantage of using the scheduler enhancement principle in the presence of competing workloads.
2. **Analysis of the memory requirements of applications**: This task involves analyzing the memory requirement pattern of some commonly used applications, specifically focusing on browser (used for social media tasks) and multi-media play back applications like YouTube and DASH streaming clients. Our analysis shows that all these applications show bursty behavior in their memory requirements and given their prevalence and significant usage in modern hand-held devices, there is a need for low latency memory management mechanisms in virtualized setups designed for mobile and hand-held devices.
3. **Analysis of self ballooning**: This task analyzes the behavior and latency involved in memory management using self ballooning mechanism where the VMs periodically increase or decrease their memory allocation based on their respective requirements. In order to analyze and appropriately compare this approach with that of directed ballooning, there is a need for introducing appropriate control mechanisms inside the hypervisor (Xen) and the VM kernel (Linux). First and foremost, we have introduced a control mechanism to start and stop the self ballooning mechanism so that it allows us to accurately measure the latency involved in mitigating a memory demand using the self-ballooning approach. In addition, we have built a mechanism inside Xen to control the process of memory release to a VM, so as to provide it only the memory that has been released by other VMs through the self ballooning process.
4. **Design and implementation of the directed ballooning mechanism with much of the logic inside Xen**: In our first implementation using the split driver approach, the Xen hypervisor's role is confined to memory pages reclamation and release. Much of the decision making and logic rests with the Dom-0 (control domain) and the data transfer for decision making happens with the help of XenBus- and XenStore-based communication mechanism. Our preliminary analysis has shown that XenStore operations significantly contribute to the overall latency. So in order to reduce the latency, this activity aims at transferring all the logic inside the Xen hypervisor and using Xen upcalls and hypercalls for the necessary data transfer.

2.1.5 Advanced OS Course on Udacity MOOC Platform

In Summer 2014, the college of Computing announced offering an online MS program in Computer Science [10] using the Udacity [11] MOOC platform. The PI Ramachandran led the faculty task force for the College of Computing that resulted in this program offering [12].

PI Ramachandran also developed one of the first courses to be offered in this program, namely, *Advanced OS* [13]. This course was taught to the first batch of Online MS CS students in Spring 2014.

2.1.6 Advanced Topics in Memory Systems - A new Graduate Course

Professor Moin Qureshi has developed a graduate course on Memory Systems [14]. This is a research oriented course where students learn about the new advances in memory technologies, memory architectures, and storage systems. Students also do a research project addressing open research problems.

3 Training and Development

3.1 Graduate Students

We had six students associated with the project for the reporting period though only 2 of them were supported by the funding. *Wonhee Cho* and *Yeonju Jeong* are the students funded from this grant. In addition to them, four more students contributed to the project. Here is a complete list and their work assignments:

1. Wonhee Cho and Yeonju Jeong are the lead students on the project, and are working on software support for increasing the efficiency and lifetime of emerging storage class memories, which is the technology of choice especially for Smartphones.
2. Dushmanta is working on virtualization technologies for personal mobile platforms that has implications for the storage architecture of Smartphones.
3. Moonkyung Ryu worked on P2P video streaming algorithms and multi-tiered storage architecture for CDN servers that help characterize the workloads on Smartphones and their implications on the system software stack for Smartphones. He graduated in May 2014 [4] and is currently working for Google.
4. Lateef Yusuf worked on constructing transient social networks using Smartphones. Given that social networks drive the use of Smarphones, his research played a vital role in informing the kinds of applications and their storage access patterns on Smartphones, which are essential for the proposed research. He graduated in May 2014 [15] and is currently working for Amazon.
5. Kirak Hong worked on system support in the form of distributed programming models and runtime systems for marrying mobile computing platforms with cloud computing in the context of situation awareness applications using camera networks. Given the data intensive nature of camera networks, his research was influential in thinking about the storage performance needs of mobile platforms. He graduated in August 2014 [16] and is currently working for Google.

3.2 Undergraduates

In Summer 2013, two students, Andrew Wilder and Daniel Whatley were associated with the project and investigated various “hooking" techniques for trapping system calls with a view to profiling the I/O access patterns of mobile apps on Nexus-7 platform. In Fall 2013, Andrew Wilder continued his investigation and his explorations contributed to the project positively. In Spring 2014, we had one undergraduate, Kyle Kelly, working as an UG RA with the help of the NSF REU supplement, under the direction of my PhD student, Lateef Yusuf. Kyle’s work was in developing a DASH streaming client application on Nexus-7 mobile platform for use as a workload in the storage research.

3.3 Full-time Employment and Internships

We have successfully placed many of our PhD students on full time employment and/or internships.

Lateef Yusuf has graduated in May 2014 and taken a full-time position with *Amazon* in Seattle, Washington.

Moonkyung Ryu has graduated in May 2014 and taken a full-time position with *Google* in Mountain View, California.

Kirak Hong has graduated in August 2014 and will be taking up a full-time position with *Google* in Seattle, Washington.

Wonhee Cho is currently (Summer 2014) working as an intern for *Microsoft Research*, Redmond, Washington. His internship is closely allied with the NSF-funded research; specifically, he is looking at in-SSD processing of large data sets to circumvent the I/O bandwidth problem with the host.

4 Outreach Activities

4.1 Industrial Collaboration

VMware is very interested in aspects of this research and has given an unrestricted gift jointly to Professors Ramachandran and Qureshi in support of their research entitled, “Rethinking Operating System Structure with Heterogeneous Main Memory." This collaboration will enhance the research output from this project beyond the NSF funding.

Dr. Qureshi spent several years at IBM Watson Research before joining the faculty at Georgia Tech. He has strong connections with them on PCM technology. These connections will help further the quality of the research results we will be able to produce from this project.

4.2 International Collaboration

4.2.1 Germany

Dr. Ramachandran successfully completed a collaborative project with Professor Kurt Rothermel of Institute for Parallel and Distributed Systems (IPVS, `www.ipvs.uni-stuttgart.de`) entitled, "Complex Event Processing in the large." This project is funded by the state of Baden Württemberg, Germany for four years from 2010-2013. The project enabled Dr. Ramachandran to spend a month each year (2010-2013) at IPVS. Additionally, it enabled extended student visit from IPVS to Dr. Ramachandran's lab resulting in several joint publications in premier outlets such as ACM DEBS and ACM SIGCOMM MCC.

Through the German Academic Exchange Service (DAAD), Professors Rothermel and Ramachandran have secured funding for three graduate students (every year) from the University of Stuttgart to visit Georgia Tech for a year-long stay involving course and project work, which can be used towards their MasterŠs degree from the University of Stuttgart. The first batch of students (Steffen Maas, Fabian Mueller, and Patrick Alt) completed their studies in Summer 2014. During the course of their stay, they participated in the research projects of Dr. Ramachandran. The next batch of three students just arrived to Georgia Tech. Involvement of such students in this project will enhance the intellectual output beyond what is possible with the NSF funding.

4.2.2 South Korea

During August 3-9, 2014, Dr. Ramachandran along with the Dean Zvi Galil, visited several leading S. Korean companies (Samsung, LG, SKhynix) with a a view to foster research collaboration between the College of Computing and these companies. Given the attention paid to storage class memories and smart phones paid by these companies, it is certain that this NSF-funded project stands to gain through the establishment of such collaborative ties.

4.3 CERCS

Georgia Tech has an NSF I/UCRC grant administered by the the Center for Experimental Research in Computer Systems (CERCS). We have an advisory board meeting for CERCS that comprises researchers from leading industries. We have a board meeting once every 6 months as part of this I/UCRC award. One of the fixed items on the agenda during these meetings is a poster and demo session by the students. This serves as a great opportunity for students on this project to meet and discuss their research with leading researchers from industries.

4.4 Open Source Software

Through the efforts of one of the students who graduated recently from our group (Dr. Hyojun Kim, currently with IBM Almaden Research), we have been successful in open-sourcing our *FlashFire* [17] software, which to date has a user community of over 100,000.

4.5 Harman Innovation Council

PI Ramachandran is a member of the **innovation council** of Harman International (`www.harman.com`). He attended their innovation council meeting in October 2013, and presented a talk entitled, "System Support for Internet of Things."

5 Plans for the Third Year

We plan to continue research as proposed. The Third year plans includes:

Measurements and Analysis of Storage Access Traces of Apps on Android. We will complete this study in Fall 2014. The results from this task will serve as a guide in identifying opportunities for sacrificing storage reliability to gain performance without sacrificing correctness of apps running on Smartphones. This study will also help to prune and select the design space exploration for both software and hardware solutions in the context of emerging storage class memories.

System Software Investigations and Architectural Innovations. System software investigation will continue this year; in addition we will explore architectural innovations in support of our findings from system software investigation. We will work on these two aspects in an integrated manner so that the solutions reinforce each other. The evaluation methodology (next task) will be heavily used in quantifying the efficacy of the solutions explored.

Performance Evaluation Studies. This task will be ongoing through the lifetime of the project. The methodologies developed through this task will be used as a general framework in the entire project.

6 Publications and Products

6.1 Publications

See the *highlighted* references at the end of this document for publications/artifacts that appeared or in preparation as a consequence of this grant.

7 Contributions

The activities we are currently undertaking are starting to bear fruits and there are papers in the publication pipeline. These are listed in the references (highlighted) at the end of this report.

7.1 Human Resources

For the reporting period, several students participated either centrally or peripherally on the project:

1. Four students (Moonkyung Ryu, Wonhee Cho, Yeonju Jeong, and Dushmanta Mohapatra) played central role with their respective research work.
2. Two other PhD students (Lateef Yusuf and Kirak Hong) played a peripheral role in that their respective research helped inform this funded project.
3. Three PhD students (Lateef Yusuf, Moonkyung Ryu, and Kirak Hong) completed their PhD dissertations and are gainfully employed
4. One PhD student (Wonhee Cho) took his PhD qualifying exam in Spring 2014.
5. Three MS students on exchange from Germany (Steffen Maas, Fabian Mueller, and Patrick Alt) worked on aspects of research that were peripherally connected to the primary focus of this project.
6. Three undergraduates (Andrew Wilder, Daniel Whatley, and Kyle Kelly) worked on aspects of the project.

8 Impact

The techniques developed in this project to date enhances the performance of SSDs built using NAND flash memory.

Specifically,

- Hyojun Kim's dissertation research and the MSST 2013 paper showed that mobile device performance is critically dependent on storage performance and the need to pay attention to designing the storage system software recognizing that the storage is flash.
- MK Ryu's dissertation research and the ACM Multimedia 2013 paper showed that with the increase in video traffic on the web, the multi-tiered storage architecture in data centers have to pay attention to read performance more than write performance.
- MK Ryu's dissertation research also paved the way for exploring commercial opportunities for a novel multi-tiered storage architecture that caters to streaming data requests from a large number of clients.

9 Special Requirements

None.

References

[1] Moonkyung Ryu and Umakishore Ramachandran. FlashStream: A multi-tiered storage architecture for HTTP video streaming incorporating MLC flash memory. In *ACM Multimedia 2013: The 21st ACM International Conference on Multimedia*, October 2013.

[2] Moonkyung Ryu and Gareth Genner and Umakishore Ramachandran. Flash in Action for High-Performance and Low-Energy Web Content Storage. NSF I-Corps January-February 2014 Bay Area Cohort.

[3] Flashboost.IO. `http://flashboost.io/`. An LLC formed as a result of the NSF I-Corps Bay Area Cohort, Jan-Feb 2014.

[4] Moonkyung Ryu. *TOWARDS A SCALABLE DESIGN OF VIDEO CONTENT DISTRIBUTION OVER THE INTERNET*. PhD thesis, School of Computer Science, College of Computing, May 2014.

[5] Wonhee Cho and Umakishore Ramachandran. Cloud-Cache: Infinite Storage Abstraction for Mobile Devices, September 2014. Manuscript Under Preparation.

[6] Hyojun Kim and Umakishore Ramachandran. Fjord: Informed storage management for smartphones. In *MSST 2013: 29th IEEE Symposium on Massive Storage Systems and Technologies*, May 2013.

[7] Yeonju Jeong and Umakishore Ramachandran. Cloud-backed application study: Trading resilience for performance, September 2014. Manuscript Under Preparation.

[8] Dushmanta Mohapatra and Umakishore Ramachandran. Evaluation of memory management schemes in virtualized environments, September 2014. Manuscript Under Preparation.

[9] Dushmanta Mohapatra and Umakishore Ramachandran. Coordinated memory management in virtualized environments, September 2014. Manuscript Under Preparation.

[10] GT OMS CS. `http://www.omscs.gatech.edu/`, 2014. Online MS in Computer Science, College of Computing, Georgia Tech.

[11] Udacity. `https://www.udacity.com/`. MOOC Platform Used by College of Computing for OMS CS.

[12] Umakishore Ramachandran. GT OMS CS - a proposal for online MS in Computer Science presented to the faculty in the college of computing. Internal to the College of Computing, Georgia Tech., 2013.

[13] Umakishore Ramachandran. Cs 6210 - advanced operating systems. `http://www.cc.gatech.edu/~rama/CS6210-External/`, Spring 2014. MOOC OMS version available at `https://www.udacity.com/ud702`.

[14] Moinuddin K. Qureshi. ECE 8873-A: Advanced Topics in Memory Systems, Spring 2014.

[15] Lateef Yusuf. *IMPROVING QUALITY OF EXPERIENCE FOR MOBILE VIDEO STREAMING*. PhD thesis, School of Computer Science, College of Computing, May 2014.

[16] Kirak Hong. *A DISTRIBUTED FRAMEWORK FOR SITUATION AWARENESS ON CAMERA NETWORKS*. PhD thesis, School of Computer Science, College of Computing, August 2014.

[17] Hyojun Kim and Umakishore Ramachandran. Flashfire: Overcoming the performance bottleneck of flash storage technology. Technical Report GT-CS-10-20, `http://smartech.gatech.edu/handle/1853/36335`, School of Computer Science, College of Computing, Georgia Institute of Technology, 2010.