Research Statement

Xiaolong Cui
Department of Computer Science
University of Pittsburgh
xic51@pitt.edu

I. Introduction

The main motivation of my proposed research stems from the observation that, as Cloud Computing continues to grow, both the failure rate and power consumption are highly-likely to increase significantly, driving the system to extremely lower efficiency while consuming unprecedented amount of energy. Hence, understanding the interplay between fault tolerance, energy consumption and profit maximization is critical for the viability of Cloud Computing. To this end, my research aims at the design and implementation of novel, scalable, and energy-aware fault tolerance frameworks for green and sustainable Cloud Computing.

Cloud Computing has emerged as an attractive platform for compute- and data-intensive applications, as it allows for low-entry costs, on demand resource provisioning and reduced cost of IT infrastructure maintenance. More and more of our daily apps, on both mobile and desktop computers, are moving their back-ends into the "cloud". Recent studies predict an annual growth rate of 17.7% by 2016, making Cloud Computing the fastest growing segment in the IT industry. As the demand for Cloud Computing accelerates, cloud service providers will need to expand their infrastructure to ensure the expected levels of performance, reliability and cost-effectiveness, resulting in a multifold increase in the number of computing, storage and communication components in their datacenters.

Service level agreement (SLA) is a critical aspect for the Cloud Computing business. Basically, SLA is a contract between the cloud service provider and customer that specifies the terms under which service is to be provided. Failure to deliver the service as specified not only subjects the cloud service provider to penalties, but also threatens customer' confidence in Cloud Computing. Unfortunately, the direct consequence of expanded datacenters is its increased propensity to failure. While the likelihood of a server failure is small, the sheer number of computing, storage and communication components that can fail is daunting. In addition, datacenters are fast becoming a major source of global energy consumption, exacerbating the impact of CO_2 emission on the environment. It is reported that energy costs alone account for 23-50% of the Cloud Computing expenses and this mounts up to $30 billion worldwide. Altogether it raises the question of how fault tolerance might impact energy consumption, the profit of Cloud Computing business, and the environment we live in.

Currently there are two fault tolerance approaches, both of which fail to answer the question above. Checkpoint/restart, which uses time redundancy, requires full or partial re-execution when failure occurs. Such an approach can incur a significant delay subjecting cloud service providers to SLA violations, and high energy costs due to extended execution time. On the other hand, Process Replication exploits hardware redundancy and executes multiple instances of the same task in parallel to guarantee that at least one instance completes without delay. This solution, however, requires additional hardware resources and increases the energy consumption proportionally, which in turn might outweigh the profit gained by providing the service. The trade-off between profit and fault tolerance calls for new frameworks to take into account both SLA requirements and energy consumption in dealing with failures.

To tackle the above challenge, I propose an energy-aware and scalable fault tolerance framework, referred to as Shadow Replication, for profit maximization and energy reduction in Cloud Computing. Similar to Process Replication, Shadow Replication ensures successful task completion by simultaneously running multiple instances. However, Shadow Replication is distinctive in that it differentiates the execution rates of the instances, enabling a parameterized trade-off between response time and energy consumption. This allows cloud service providers to maximize the expected profit by accounting for income, potential penalties, and energy cost, ultimately promoting green and sustainable Cloud Computing.

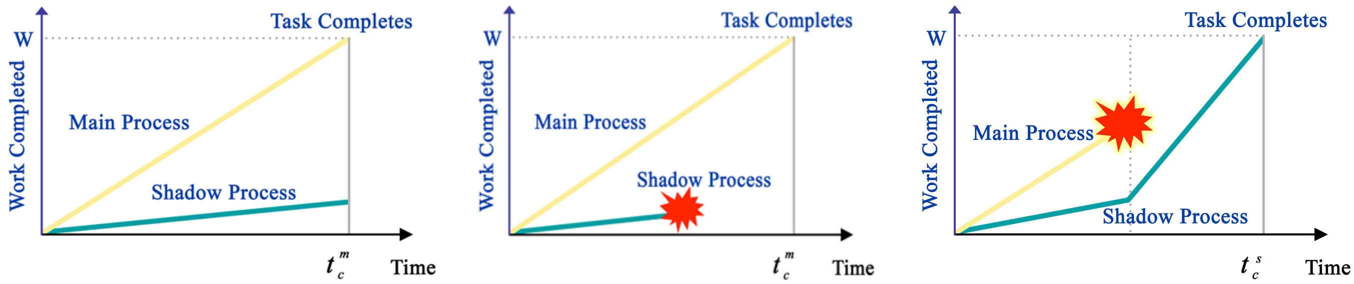
The rest of the statement is organized as follows. Section II introduces current progress in Shadow Replication and presents our preliminary results. Section III points out directions for future exploration. Section IV concludes this statement.

II. Current progress

I have been working on this project with my advisors, Dr. Taieb Znati and Dr. Rami Melhem, for one and a half years. So far, we have accomplished the following goals:

- A design of the Shadow Replication framework;
- Development of a profit-based analytical model to explore the feasibility of Shadow Replication, and to determine the optimal execution rates of all task instances to maximize profit and minimize energy;
- Development of an event-driven simulator to verify the above analytical model;
- A comprehensive evaluation using both the analytical model and simulator to analyze the profit and energy savings achievable by Shadow Replication, compared to existing fault tolerance approaches.

I have submitted three papers on this work. [1] is published in Energies 7, no. 8 (2014) and [2] is accepted to CLOSER 2014. Our third paper is currently under review.



(a) No Failure (b) Shadow Process Failure (c) Main Process Failure

Fig. 1: Shadow Replication with a single shadow process.

A. Shadow Replication

The basic tenet of Shadow Replication is to associate with each main instance a suite of "shadows" whose size depends on the criticality of the application and its performance requirements. Each instance is executed using a process. To mitigate correlated failures, the main and shadow processes execute on separate computing nodes.

Formally, the Shadow Replication fault tolerance framework is defined as follows:

- A main process, $P_m(W, \sigma_m)$, that executes at the rate of σ_m to complete a task of size W;
- A suite of shadow processes, $P_s(W, \sigma_s^b, \sigma_s^a)$ $(1 \leq s \leq S)$, where S is the size of the suite. The shadow processes start execution simultaneously with the main process at rate σ_s^b. Upon failure of the main process, all shadows switch their rates to σ_s^a, with one shadow process designated as the new main process. This continues until the completion of the task.

To illustrate the behavior of Shadow Replication, I use only one shadow process and consider the scenarios depicted in Figure 1, assuming at most one process failure. Figure 1(a) represents the case when neither the main nor the shadow fails. The main process, executing at a higher rate, completes the task at time t_c^m. At this time, the shadow process, progressing at a lower rate, stops execution immediately. Figure 1(b) represents the case when the shadow fails. This failure, however, has no impact on the progress of the main process, which can still complete the task at t_c^m. Figure 1(c) depicts the case when the main process fails while the shadow is in progress. After detecting the failure of the main process, the shadow begins executing at a higher rate, completing the task at time t_c^s. Given that the failure rate of an individual node is much lower than the aggregate system failure, it is very likely that the main process will always complete its execution successfully, thereby achieving fault tolerance at a significantly reduced cost of energy consumed by the shadow.

The novelty of Shadow Replication lies in its differentiation of the execution rates. Specifically, it executes the main process at the rate required for response time constraint, while slowing down the shadows for energy saving, thereby enabling a parameterized trade-off between response time and energy consumption. A closer look at the model reveals that Shadow Replication is a generalization of existing fault tolerance approaches, namely Checkpoint/restart and Process Replication. If the SLA allows for flexible completion time, Shadow Replication would take advantage of the delay laxity to trade time redundancy for energy savings. If the target response time is stringent, however, the shadows would execute simultaneously with the main at high rates. The flexibility of the Shadow Replication provides the basis for the design of a fault tolerance strategy that strikes a balance between task completion time and energy saving, thereby maximizing profit.

B. Analytical model and simulator

One challenge of Shadow Replication resides in determining jointly the execution rates of all processes, with the objective to minimize energy and maximize profit. To achieve this, I propose an analytical model, from which an optimization problem is formulated to derive the optimal execution rates. The model considers the time and energy needed for a cloud job, under different system specifics and failure distributions. The profit is modeled as the difference between the payment from customers, which depends on the completion time, and expenses for running the cloud job, which are mainly energy costs. For more details please refer to [1].

To verify the correctness of the analytical model, I build an event-driven simulator that simulates the behaviors of Shadow Replication under various configurations. It can report all necessary statistics, such as number of failures encountered, time to completion, and energy consumption.

C. Preliminary results

Several important parameters are identified that impact the energy consumption and profit gains of Shadow Replication. Correspondingly, I conduct a series of sensitivity studies where Shadow Replication is compared to state-of-the-art approaches. The results from both the analytical model and simulator show that Shadow Replication can achieve significant energy savings and profit gains, without violating the SLA constraints. Specifically, Shadow Replication can achieve 15%-30% energy savings and 20%-30% profit increase under reasonable configurations [2]. Furthermore, Shadow Replication would converge to Process Replication, when target response time is stringent, and to Checkpoint/restart when target response time is relaxed or when failure is unlikely.

III. FUTURE DIRECTIONS

Current results reveal that Shadow Computing is promising for significant energy saving within SLA constraints. The direct benefits include profit gains for cloud service providers and reduced CO_2 emission, making Cloud Computing more environment-friendly and more sustainable. Inspired by that, I will be fully committed to solve some challenging questions in the next academic year.

The first plan is to further improve the efficiency of Shadow Replication for tightly-coupled jobs. Our current design works well for loosely-coupled jobs, such as MapReduce jobs, where synchronization among tasks is minimized. In a tightly-coupled job, however, even a very short recovery time $\big((t_c^m - t_c^s)$ in Figure 1$\big)$ may be amplified by the frequent synchronizations, resulting in a delay in the job completion time. In order to minimize this effect and further improve performance, I plan to explore the potential benefits of a new technique, referred to as "Leaping Shadows". The idea is to take advantage of the recovery time and align the execution state of the slow shadow processes with their faster main processes to achieve forward progress. Remote Direct Memory Access (RDMA) is a possible way to implement Leaping Shadows. This step will take approximately 3 months to complete.

The next research direction is to evaluate the feasibility and performance of using process collocation in Shadow Replication. Our current work assumes Dynamic Voltage and Frequency Scaling (DVFS) in controlling the execution rates. The effectiveness of DVFS, however, may be markedly reduced in computational platforms that exhibit saturation of the processor clock frequencies or large static power consumption. An alternative is to collocate multiple processes on a single computing node, while keeping the node running at maximum rate. Time sharing can then be used to achieve the desired execution rates. The two alternatives are equivalent in terms of completion time, since they have the same effect on the execution rate control. In terms of energy, however, each of them has its advantage. Process collocation requires less hardware resources and this reduces the energy linearly, while DVFS uses more hardwares but can reduce energy superlinearly. It needs further analysis to determine which alternative consumes less overall energy. Furthermore, more efforts are needed to study the potential issues with process collocation, such as correlated failures and collocation overhead. This will take approximately 3 months to complete.

The last and most challenging step is to build a prototype, in order to experimentally evaluate the performance of Shadow Replication using real life applications. This effort includes the design and implementation of a software library that supports the main components of Shadow Replication, including process collocation, required consistency protocols, message logging and message forwarding protocols, and execution state transfer in support of Leaping Shadows. For Shadow Replication to be scalable and efficient, it is necessary to minimize its overhead to the normal execution of the running processes as well as to the operating system. This step will take approximately 6 months.

IV. CONCLUSION

The main motivation of this work stems from the observation that, as Cloud Computing continues to grow, both the failure rate and energy consumption are highly-likely to increase significantly. Hence, understanding the interplay between fault-tolerance, energy consumption and profit maximization is critical for the viability of Cloud Computing.

To this end, I propose Shadow Replication as a novel, scalable, and energy-aware fault tolerance framework to minimize energy and maximize profit. Our preliminary results predict that Shadow Replication is able to achieve significant energy saving and profit gains while satisfying SLA requirements. I will continue to explore and optimize Shadow Replication in support of green and sustainable Cloud Computing.

REFERENCES

[1] X. Cui, B. Mills, T. Znati, and R. Melhem, "Shadow replication: An energy-aware, fault-tolerant computational model for green cloud computing," *Energies*, vol. 7, no. 8, pp. 5151–5176, 2014. [Online]. Available: http://www.mdpi.com/1996-1073/7/8/5151

[2] X. Cui, B. Mills, R. Melhem, and T. Znati, "Shadows on the cloud: An energy-aware, profit maximizing resilience framework for cloud computing." in *4th International Conference on Cloud Computing and Services Science*, 2014.