# Lazy Shadowing: A scalable, energy-aware resilience framework for extreme-scale systems

Xiaolong Cui, Rami Melhem, and Taieb Znati

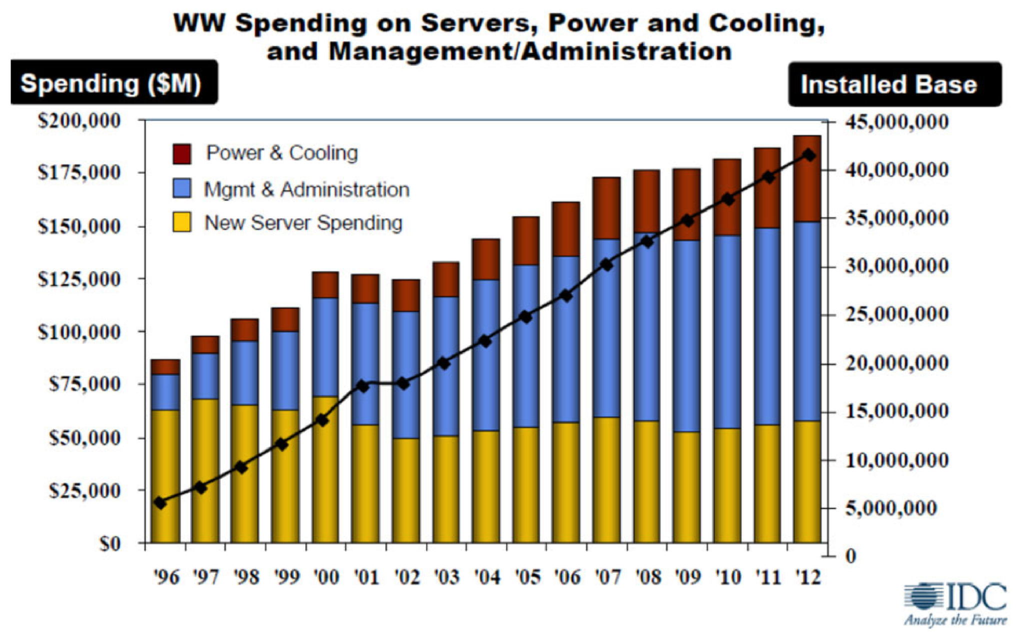
## Problem

As the demand for computing capability continues to increase, there will be a multifold increase in the number of computing, storage and communication components in large scale systems, such as HPC supercomputers and cloud data centers. This increase has two direct implications:

1. Increased failure rate
2. Increased energy consumption

| Number of nodes | Projected System MTBF |
|---|---|
| 1 | 5 years |
| 10 | 6 months |
| 100 | 20 days |
| 1000 | 2 days |
| 10000 | 4 hours |
| 100000 | 25 minutes |

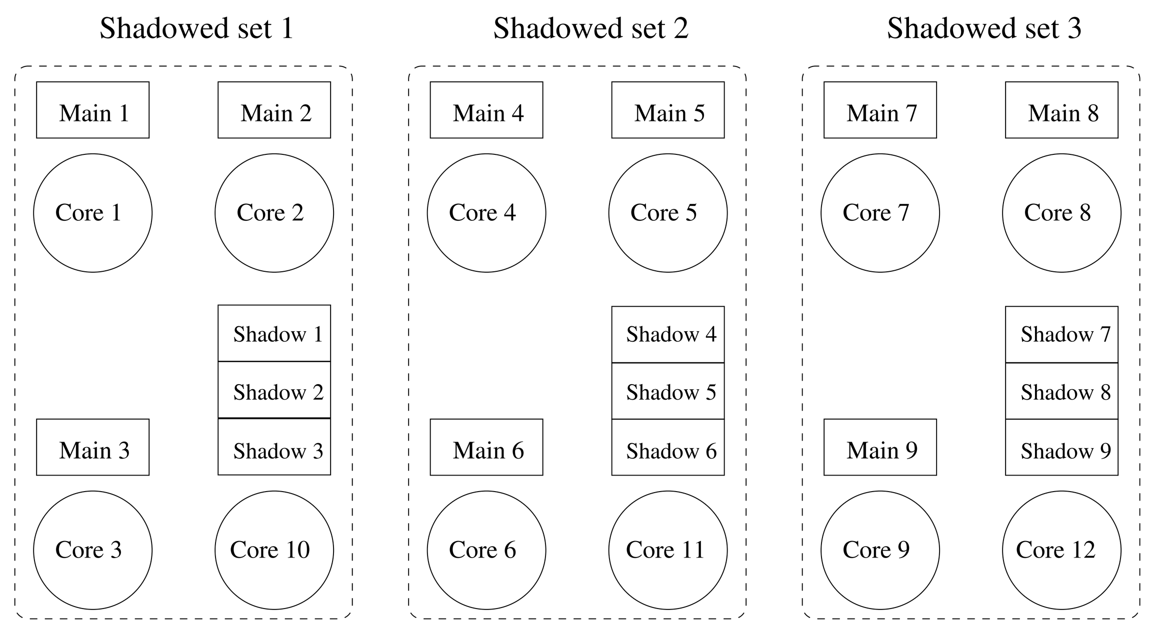
Failure




Energy

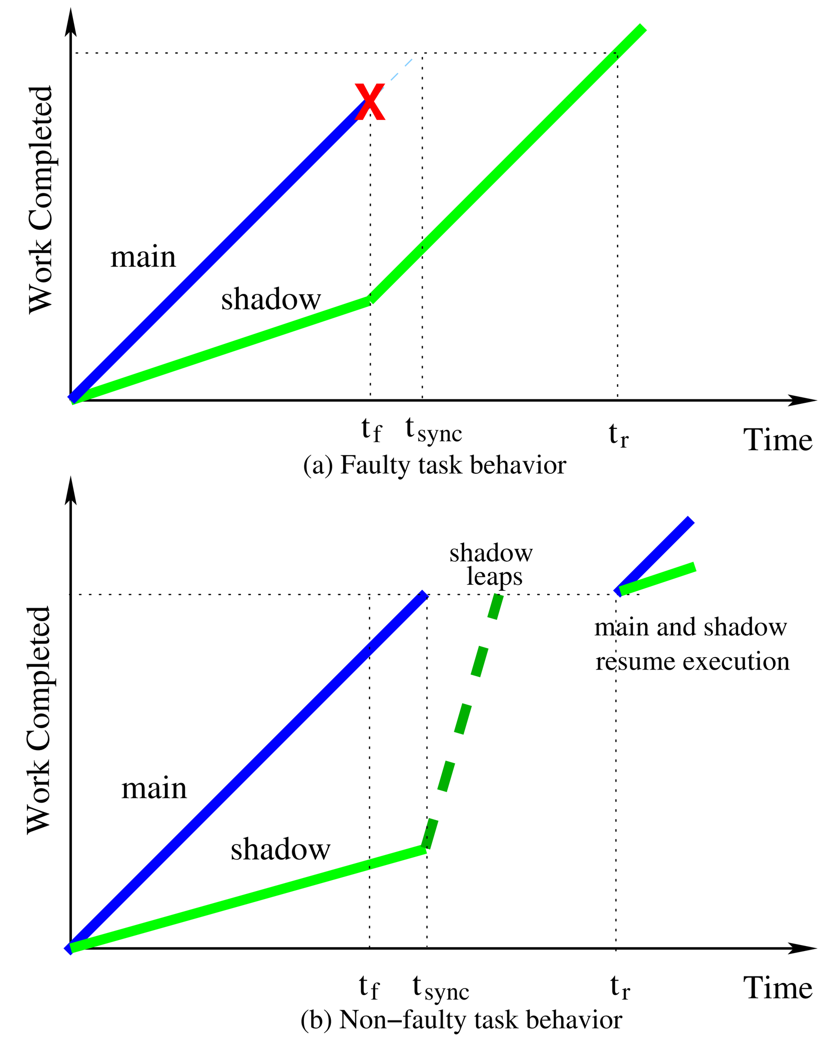
## Implementation

**Collocation** is used to achieve the desired execution rates of the shadow processes.




An example of collocation.

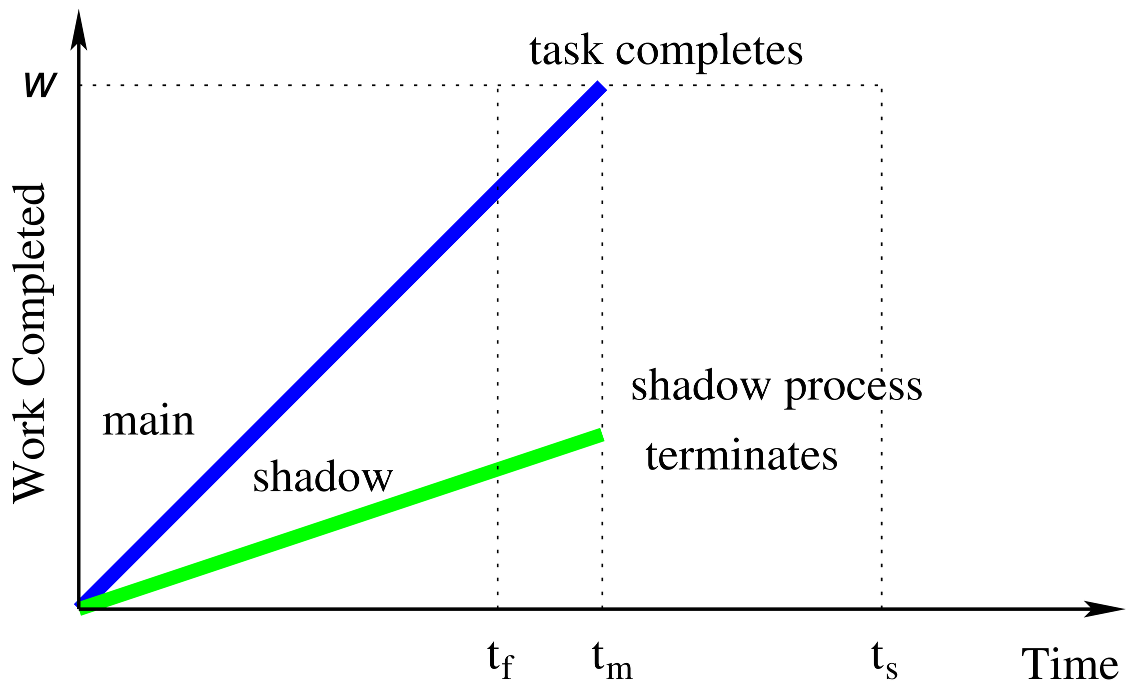
**Leaping Shadows** is an optimization for tightly-coupled parallel applications.



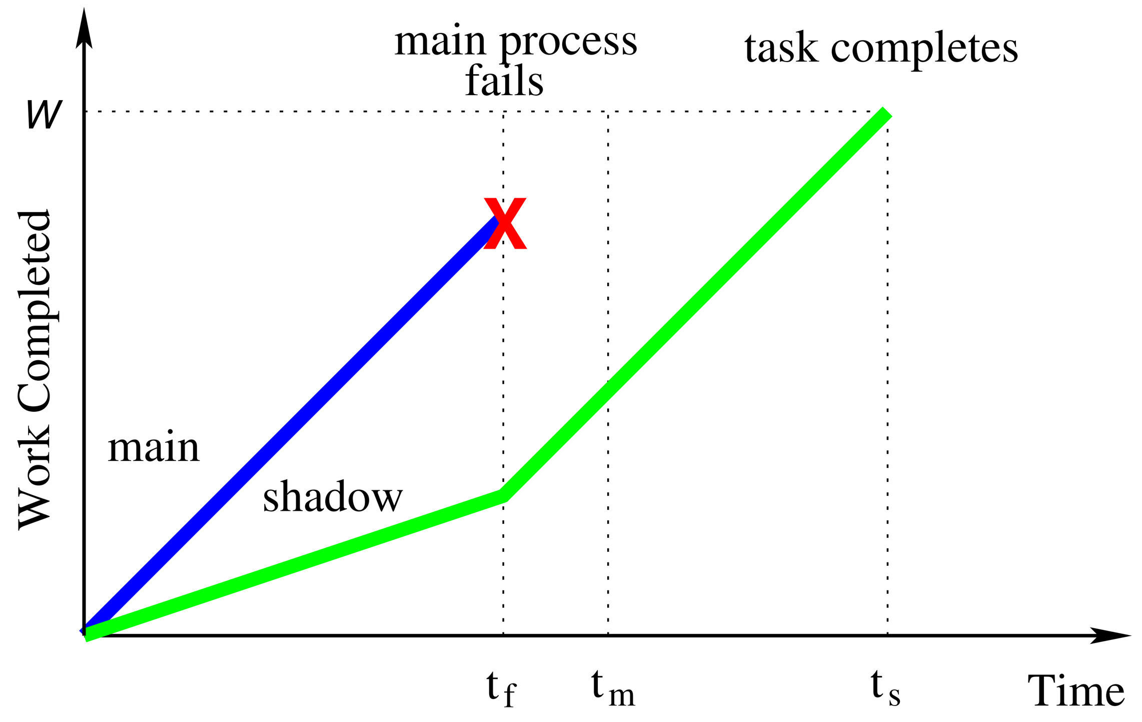

Leaping shadows.

## Method

The basic tenet of Lazy Shadowing is to associate with each main process a suite of shadow processes, whose size depends on the criticality of the application and the reliability of the underlying system. The shadows execute simultaneously with the mains, but at a slower speed to save energy.



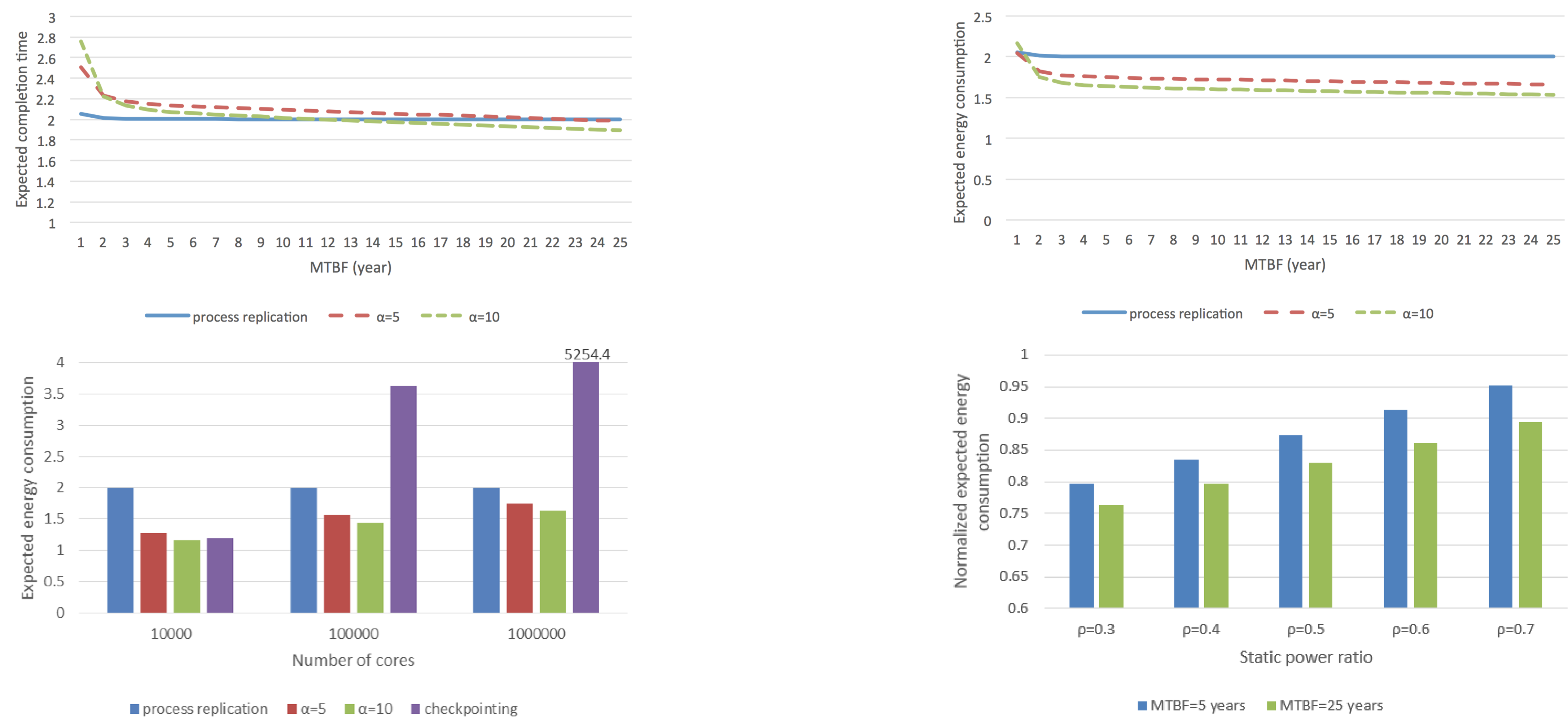

Main process successful completion




Main process failure

## Evaluation

Compared to existing fault tolerance methods, Lazy Shadowing can achieve 20% energy saving with reduced solution time at scale.




## Future Work

There are two directions:

1. Build a simulator to validate the model
2. Implement a prototype for Lazy Shadowing and evaluate the performance

## Conclusion

Understanding the interplay between fault-tolerance and energy consumption is critical for the viability of future large scale systems. To this end, we propose Lazy Shadowing as a scalable, energy-aware fault-tolerance approach. The beauty of Lazy Shadowing is its ability to explore a parameterized tradeoff between hardware and time redundancy to tolerate failures and balance between response time and energy consumption.

## Acknowledgment

This research is based in part upon work supported by the National Science Foundation under Grant Number CNS12-53218 and CNS12-52306.