Cloud Architecture and Application Programming Interface

NSF-PI Meeting
Spring 2011

Justin Y. Shi

shi@temple.edu
Where is the Science?

• We have three well-established research divisions:
  – Computer architecture: fastest, reliable computing possible.
  – Network architecture: fastest, reliable networking possible.
  – Application development: easiest programming possible.

• Troubles start when the applications are put together ... for the cloud?
The Cloud Question

How to decrease applications mean time between failures (MTBF) as we increase component count?
The Cloud Question

How to increase application performance as we increase component count?
A Historical Reference Point

• The low-level data transport layer has been delivering stochastic and lossless, scalable, reliable deterministic services for many decades.

• The higher level messaging (SOA), transaction processing, storage and parallel processing suffer scalability challenges.

• Some good discipline was lost in fast technology developments?
A Possible Approach

• Honest APIs: Application APIs should be more “honest” than the current designs. The “unknown” state must be explicitly programmed to avoid non-deterministic behavior.

• Semantic multiplexing: a method known for its power to sustain extreme large scale networks.

• Holistic application architecture: a practice that integrates API design, computing and networking infrastructure designs in the same consideration.
Risk Assessment

• Challenges the current architectural establishments that focuses only on performance or reliability but not both.
• Encourages exchanges amongst existing knowledge bases (app dev, computer architecture, network architecture) at fundamental levels. This is hard.
• Enables the further developments of mission critical functions and tasks without scalability limitations. This is harder.
Potential Impact

• The realization of **API’s involvement in application architecture** will have a profound impact on existing and future applications. Current APIs assume only two states for each operation: success or failure. The third state must also be handled or the reliability of transmission cannot be guaranteed.

• Statistic multiplexing is a powerful tool in addressing scalability challenges. **Semantic multiplexing** is capable of solving some computationally impossible problems. It must be promoted beyond the protocol research community. Education materials should be updated.

• **Large efficiency gains** can be expected in existing and future large scale cloud systems.

• **Cloud sustainability** can be drastically improved for all applications.
Scientific Rubber must also meet the road someday...

• Answers to the open “replication” needs with minimal negative effects.
• Answers to ACID requirements with minimal negative effects.
• Answers the two Cloud questions.
• ...
Thank you!

Contact: Justin Y. Shi, shi@temple.edu