Enabling Utility Computing in Clouds

Bhuvan Urgaonkar
bhuvan@cse.psu.edu
Department of Computer Science and Engineering
The Pennsylvania State University, University Park 16802, PA

Topics: Cloud Architectures and Systems; Cloud Self-Monitoring and Autonomic Control.

Summary of Current Cloud Research

The cloud computing model has opened up new possibilities for realization of the long-cherished goal of *utility computing*. Utility computing represents the desire to have IT acquired, delivered, used, paid for, and managed in a manner similar to how we use other commoditized utilities such as electricity, telephone service, cable television, etc. The principal appeal of utility computing lies in the systematized framework it could create for the interaction between providers and consumers of IT resources. In the Computer Systems Lab at Penn State, my students and I are pursuing research on developing mechanisms - for both providers and consumers of cloud platforms - that would enable such interaction. Towards this end, we plan to draw upon lessons offered by existing well-established utilities whenever appropriate and develop a variety of enabling mechanisms for cloud computing.

Provider-end Resource Usage Accounting: One of the foundational features of a utility is *accounting*—the ability of both the provider and the consumer to accurately measure the usage of the commodity being exchanged. Accurate resource accounting is desirable for a variety of important reasons, ranging from billing and auditing to effective resource allocation, anomaly detection and resolution. In our first research thread, we are developing provider-end accounting mechanisms. The mere fact that the provider has access to all the information regarding a consumer's usage does not make provider-end accounting trivial by any means. Key difficulties in such accounting arise from the distributed nature of consumers' applications as well as that of the IT systems that the providers manage. It requires the provider be able to monitor activities going on within various sub-systems of this complex environment, identify ones that are related, and aggregate their resource usage. We have developed *vPath*, a low-overhead technique that identifies causal relationships between components of a distributed application that cooperate together in the execution of a chargeable activity (e.g., a Web session) [2]. vPath is implemented completely within a hypervisor and doesnot require application modification, features that make it suitable for monitoring within a virtualized environment typical of a cloud data center. In ongoing research, (partially funded by my NSF CAREER award) we are employing the relationships identified by vPath to aggregate resources used within a provider's infrastructure on behalf of chargeable entities.

Future Research Problems

- Utility computing should allow consumers to participate in *active and informed* ways in making resource procurement decisions in a transparent "market" of competing providers. Consumers of current cloud-based offerings have a limited view of and control over resource procurement and control, a significant hindrance in the realization of a utility. We will develop mechanisms and techniques that would reduce this gap, thereby helping turn cloud-based offerings of the near future into mature utilities. Whereas emergent providers, by offering a variety of information about various options available to a consumer, bear a striking resemblance to those in other mature utilities, the other half of the utility, namely the consumers, remains largely unaddressed with regard to how best they could procure resources. We will explore techniques based on *auditing* that would allow a consumer to build probabilistic estimates of the actual allocations of various resources offered to it, independently of the provider.

- We will work on the following quintessential problem that a cloud consumer must answer: what and how many resources should I procure from the cloud to best meet my needs? This basic decision-making may help a consumer determine cost-effective ways of procuring resources from a cloud provider - in fact, from among multiple providers that might be offering comparable services. It will be important for such decision-making to be able to infer dynamically evolving mappings between the virtual resources exposed to the consumer and the physical resources underlying them. Two key themes we plan to pursue for building such functionality at a consumer are: (i) *malleable application clones* that can mimic the resource needs of an actual consumer application and can be run on a cloud to infer resource needs before moving an actual application, and (ii) *delta modeling*, whereby the complex task of modeling a consumer application's performance on a cloud would be broken into the simpler task of modeling it on a local set of hardware (accessible to the consumer) followed by applying a *delta* to this model that would be inferred by carefully conducted measurements that bring out key differences and similarities between the local resources and those at the cloud.

- Finally, we envision a cloud-based utility where a consumer would desire being able to reason about the implications of its decision-making on *multiple* features capturing its *cost versus revenue*. We will devise a hierarchy of consumer-end resource procurement and control mechanisms, spanning multiple spatio-temporal scales—from obtaining resources from (possibly multiple) providers at coarse time-scales to operating individual virtualized resources based on analytical models that are capable of capturing important trade-offs. Our recent work on modeling trade-offs that certain systems offer between performance and other metrics such as security or power consumption will offer useful guidelines and insights in our pursuit [1].

References

[1] Shiva Chaitanya, Bhuvan Urgaonkar, and Anand Sivasubramaniam. Qdsl: A queueing model for systems with differential service levels. In *In proceedings of the ACM Conference on Measurement and Modeling of Computer Systems (SIGMETRICS 2008)*, Jun 2008.

[2] Byung Chul Tak, Chunqiang Tang, Chun Zhang, Sriram Govindan, Bhuvan Urgaonkar, and Rong N. Chang. vpath: Precise discovery of request processing paths from black-box observations of thread and network activities. In *In proceedings of 2009 USENIX Annual Technical Conference (USENIX'09)*, Jun 2009.