The Cloud is the future Internet: how can we engineer a cloud?

James Roberts

The Cloud is the future Internet: how can we engineer a cloud?

Infocom 2013 Keynote

Jim Roberts, Inria RAP

Abstract:

The future Internet will be a network of data centers. Of course, this is already true today in the sense that most content and services delivered to end users come from a data center of some form or other. We postulate a future where this cloud of compute and storage resources is rationalized to yield a much more economically efficient structure. Users will be attached to the Internet via a data center that, among other functions, will act as router for the small proportion of demand that cannot be handled locally. Independently of this prediction, the performance of data center networks and the cloud is among the hottest topics in networking today. More specifically, researchers are seeking ways to fulfill service level agreements, typically expressed in terms of minimum bandwidth and maximum latency guarantees, in the multi-tenanted cloud infrastructure. This is clearly reminiscent of the huge amount of effort expended over the last decades on trying to solve the issues of Internet QoS. Unfortunately, it seems lessons from the failure of much of that research have still not been widely learnt. We would argue that the main reason for the inapplicability of QoS service models is the failure to take proper account of the random nature of demand. While modelling demand in terms of stationary stochastic processes is a tried and tested approach in engineering many networks and systems, it is hardly ever applied in evaluating schemes for sharing Internet (or cloud) resources. This is unfortunate since, in our experience, it is an approach that brings essential understanding and enables the design of simple and effective traffic control mechanisms. We must gain similar understanding to engineer the cloud. The challenge is to succinctly characterize cloud traffic and then to elaborate models to accurately predict performance for given demand and network capacity.