Abstract
From an enterprise perspective, one key motivation to transform the traditional IT management into Cloud is the cost reduction of the hosted services. In an Infrastructure-as-a-Service (IaaS) Cloud, virtual machine (VM) instances share the physical machines (PMs) in the provider’s data center. Increasing the number of PMs leads to lower downtime cost at the expense of higher infrastructure and other operational costs (e.g., power consumption and cooling costs). Hence, determining the optimal PM capacity that minimizes the overall cost is of interest. In this paper, we show how a cost analysis and optimization framework can be developed using stochastic availability and performance models of an IaaS Cloud. Specifically, we develop and solve a cost minimization problem to address the capacity planning in an IaaS Cloud: what is the optimal number of PMs that minimizes the total cost of ownership for a given downtime and performance requirement set by service level agreements? We use simulated annealing, a well-known stochastic search algorithm, to solve the optimization model. For each point in the search space, we need to determine the performance, availability and power consumption requirements. Hence we develop scalable analytic models for the performance, availability and power consumption analysis of an IaaS Cloud. The essence of our approach is in reducing the complexity of analysis by dividing the overall model into multiple interacting stochastic process sub-models and then obtaining the overall solution by (fixed-point) iteration over individual sub-model solutions.

Short Biography
Kishor S. Trivedi holds the Hudson Chair in the Department of Electrical and Computer Engineering at Duke University, Durham, NC. He has a B.Tech (EE, 1968) from IIT Mumbai, M.S. (CS, 1972) and PhD (CS, 1974) from the University of Illinois, Urbana-Champaign. He has been on the Duke faculty since 1975. He is currently a visiting professor at IIT Gandhinagar. He is the author of a well-known text entitled, Probability and Statistics with Reliability, Queuing and Computer Science Applications, first published by Prentice-Hall; a thoroughly revised second edition (including its Indian edition) of this book has been published by John Wiley. He is a Fellow of the Institute of Electrical and Electronics Engineers. He is a Golden Core Member of IEEE Computer Society. He has published over 500 articles and has supervised 45 Ph.D. dissertations. He is the recipient of IEEE Computer Society Technical Achievement Award for his research on Software Aging and Rejuvenation. His research interests in are in reliability, availability, performance, performability and survivability modeling of computer and communication systems. He works closely with industry in carrying our reliability/availability analysis, providing short courses on reliability, availability, performability modeling and in the development and dissemination of software packages such as SHARPE and SPNP.