KEN BIRMAN,
An exclusive interview, conducted in Athens on March 27, 2010

"We're still in the early days… Year 1 for cloud computing may be closer to 2015"

— A wave of innovation will enable organizations to add capabilities and save money by leveraging this powerful new technology. But before the cloud can be trusted for mission-critical uses, tough challenges must be overcome by researchers and service providers.

The world was plunged into deep crisis by the financial sector. Could the information technology industry spark an upturn?

Everything I've read suggests that the technology sector is leading a recovery from the recent economic crisis. The evidence ranges from the strong sales of products like iPads and Kindles to the impressive profits of companies like Intel, Cisco and Microsoft. But what excites me most is the potential I see as we look ahead. Cloud computing is enabling companies to virtualize and externalize their data, giving them efficiencies and flexibility in running their IT systems, and making it possible to deploy new services cheaply. This is a tremendous opportunity and I think it could ultimately dwarf the web revolution. However, sophistication is needed because, in some ways, cloud technology is just not ready yet. The balancing act is to move aggressively but prudently as this exciting technology evolves.

How do you see this wave of new technologies developing in the near future?

I anticipate all sorts of advances in areas associated with the cloud, and the Internet as a whole—a basic transformation in the ways that many corporations operate. Today we think of TCP/IP as the bricks and mortar of the Internet, but the protocol of the future will be HTTP, the web services protocol used in cloud computing applications. The Internet and the cloud computing systems that talk to client platforms are evolving at stunning speed, and this is reshaping everything: hardware, software, even the way we think about IT applications. I don't think we've seen anything so sweeping since the client-server revolution in the 1980s!

So do you think that corporations should already be shifting as much of their IT infrastructure as possible to the cloud?

Technology revolutions take time to mature. Today, one focus for innovation involves the capabilities of the Internet itself, both in terms of speed but also reliability, and in the ways that the network functions with the cloud. In some sense the network is becoming a projection of the cloud, hosting content, playing roles in routing requests but also voice traffic and other media, and even performing some cloud functions directly, for example to support applications on small mobile devices. Simultaneously, we're seeing a second wave of innovation within the cloud itself, as providers discover ways to leverage cloud computing in support of all sorts of IT functionality.

Is cost the main incentive?

Even at this early stage, cloud computing already offers 10- to 20-fold reductions in the cost of administration, computing and storage, and the advantage could grow. The challenge is to seize this opportunity. If we think about the big, traditional corporate IT systems that do resource management, planning and fulfillment, handle customer relations systems—these are too complex to port quickly to the cloud model. One possible response is to build private clouds—to "virtualize" existing applications and host them on an in-house cloud. But to get the full benefit one needs to learn new ways of computing that involve renting resources from Amazon or another cloud provider, or rebuilding entire systems on a cloud platform like Microsoft Azure or Salesforce.com. That path offers much greater cost savings, but can require new ways of designing systems. So we've entered a period that challenges corporate planners with a wide range of choices and some compelling economic incentives, but the decisions aren't simple ones.

Can we be sure these trends will be successful?

The problem is that the area is changing so quickly. So it's easy to say that the Internet and cloud will evolve, but until the solutions are really in hand, how can we be sure that we'll solve all these hard problems? Nonetheless, I'm very optimistic.

What are the risks involved with cloud computing?

To me, the question centers on trust. Corporations manage all sorts of sensitive data, and we've learned to do this well. But as we displace sensitive, mission-critical applications into new settings, we need to trust new kinds of technologies that may not have the properties required. I'm not talking just about applications that keep credit card data, or that control power plants or weapons, run electric power grids, or operate hospitals and banks. This issue of sensitive data and critical computing also includes mundane things like e-mail and telephone services. As we look to the future, the concerns become even stronger. For example, to improve the quality of care (and of life) for an elderly person, we might develop computer systems that enable that person to live at home instead of in an assisted care facility. But suppose that the computing system of a small hospital with 1,000 of these home-care patients malfunctions, or the Internet fails. You could see very fragile individuals cut off from medical care. Meanwhile, cost is such a crucial parameter for many industries that it becomes almost impossible for governments not to think about shifting expensive activities such as healthcare into the cloud. So creating trustworthy cloud computing is, for me, the highest priority—and the lack of trust is the biggest risk today.

So trust is the key to success? By "trust," do you mean security, or something more?

Here at Cornell, we do view security and privacy as central needs, but we also think about properties like high availability, data consistency, fault-tolerance, robustness against attack and rapid response. Sometimes these are in tension: today's cloud systems often relax consistency to guarantee quick response. To this we must add a further element: if data or applications run on third-party platforms, what happens if the cloud hosting company goes out of business, or changes pricing in an abrupt way, or suffers from reliability issues? For cloud computing to really succeed, we need to learn to create trusted applications that run on untrusted platforms.

But for Amazon, Microsoft, Google, Akamai and Salesforce, cloud computing is already a business reality!

Absolutely, and their solutions have been quite reliable—despite some very visible glitches. But keep in mind that all these companies have

their own reasons for offering cloud resources cheaply, and their corporate strategies could shift. Still, with this caveat, cloud computing is certainly here to stay, if for no other reason than that cloud computing is the right model for search, and search has been a tremendous success. Just the same, I think we're yet at Year 1 for cloud computing for large industries or banks, on account of their numerous obligations and the need to ensure the infrastructure and guarantee customer privacy and trading 24/7, if required. A bank cannot casually become dependent upon third-party systems. My guess is that some industries will end up building their own clouds, but it may be five years before that becomes common. Moreover, it takes some time to build, validate, gain confidence and gradually migrate applications. So we're still in the early days… Year 1 for cloud computing may be closer to 2015.

The increasingly connected world affects companies and their IT systems in particular. How will their processes utilize and make better use of the information collected than they do today?

One thing that excites me is the ability to replace inexact models with queries running directly on the underlying data. To give an example, consider the resource planning models used in ERP systems, or the pricing models used by banks. These are really approximations—mathematical equations we use to represent reality, because reality is too complicated to work with. But if the data itself can be captured, we can often dispense with the models and answer those questions with the full accuracy the data permits.

And what are the benefits of this kind of shift?

Earlier we talked about the financial crisis, so let me use that for an example. We all know that the root cause of the crisis was that banks were unable to compute the actual value of the complex mortgage-based derivatives they were trading. Those mortgages depend, ultimately, on the value of real-estate, and until recently, there was no easy way to pull up prices for properties on a house-by-house basis. So banks depended on models that estimated pricing for entire regions, and those models in some sense caused the financial meltdown. But today, we can obtain those house-by-house prices, as well as information about the credit worthiness of the owner. One could take one of those mortgage-backed securities and arrive at an accurate measure of their value and the risk, based on the real facts. If we had had that kind of data five years ago, we might have avoided the worldwide economic crisis. And this is just one example among many.

Globalization has rendered the economy increasingly complex, and the downturn has increased the need for regulation. What are the impacts on information systems?

Our governments are demanding transparency in the corporate decision process, and that implies much better record keeping. For instance, banks are now required to do quite a bit of record management and archiving to comply with the oversight regulations in Europe and in the United States. This vast array of records contains all sorts of information, e-mails, transactions, telephone call logs, and so on. It must be possible to access them during the correct period but impossible to access them subsequently, and only the right people should be authorized to do so. And they must be kept in a safe place. Such situations pose a mixture of legal, regulatory and technical challenges, and this is new for those of us who do IT research. We've always looked at problems through purely technical lenses; suddenly, we're facing problems that demand much more sophistication.

Surely such vast issues require a certain amount of cross-disciplinary action?

In the United States, one answer has been the NSF-funded TRUST science and technology consortium, which Cornell helped create in 2004, along with Berkeley, Stanford, Carnegie Mellon and Vanderbilt University. TRUST includes economists, lawyers and researchers from many disciplines in addition to those at the top of the field in security and reliability. We work as a team, and this lets us tackle the sorts of complicated challenges we see in settings like finance and health care.

What will be the role of the IT service provider in the next wave of innovations?

I think the opportunity is enormous, but also that it demands a special kind of sophistication. Cloud computing and the evolution of the Internet are changing the IT landscape and no corporation can afford to be left behind. Yet concerns about trust and privacy and worries about dependency on third-party providers are very real. We've seen a rush to cloud computing over the past decade, but more recently there is a growing awareness of the dangers. Under these conditions, corporations will need to work with IT partners who have the sophistication to understand what works today, what will be possible tomorrow and how best to leverage these new capabilities with minimal risk. I'm certain that providers who justify the trust that their clients place in them will thrive in this new environment.

Ken Birman is the Rao Professor of Computer Science at Cornell University (Ithaca, New York) and was the winner of the 2009 IEEE Kanai Award for his research in distributed systems. A leader in the field of cloud computing for more than two decades, software developed by Professor Birman has been used by the New York Stock Exchange and the French air traffic control system.

His publications include *Reliable Distributed Systems: Technologies, Web Services and Applications* (Springer Verlag, 2005).

Low tide