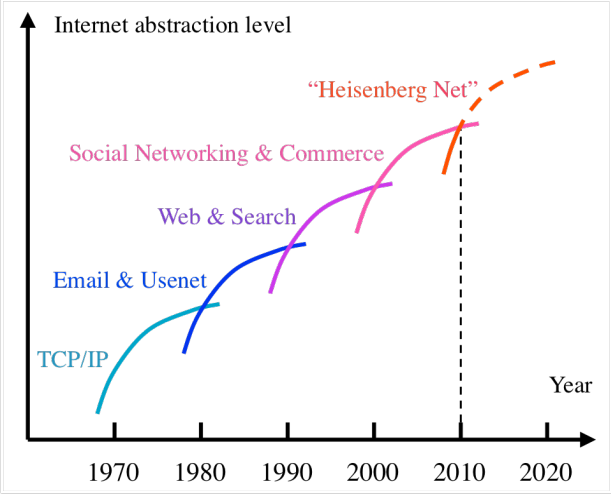
The Clouds of Revolution

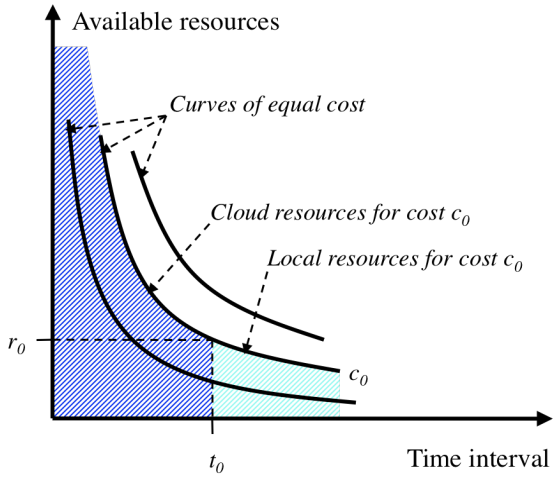
Peter Van Roy, Seif Haridi, Alexander Reinefeld, Jean-Bernard Stefani

Extended version of an editorial to appear in
Public Service Review, European Union, Issue 19

January 2010



The Internet has been revolutionizing society for four decades. Old timers have seen four Internet revolutions in their lifetimes: TCP/IP (1970s), Email and Usenet (1980s), the Web and its search engines (1990s), and social networking and Web commerce (2000s). We are now on the brink of the next revolution, which will bring the Internet to a new level of functionality [1,2]. The new revolution is based on cloud computing and the concept of elasticity. Any Internet application will be able to access immense computational and storage resources for short time periods. Cloud resources are typically sold according to a "pay per use" principle. This means that the product of the amount of resources and the time they are used is approximately proportional to the cost. For a given cost, large amounts of resources can be made available for short times. This is analogous to Heisenberg's uncertainty principle in physics, which makes available arbitrarily high energies if the time period is short enough. We call it the computational Heisenberg principle. There already exist common applications that use this principle. For example, Google Search has revolutionized the way people use the Web. It is based on the PageRank algorithm, which needs vast computational and storage resources. Applications that use vast resources in short bursts will become ubiquitous and lead to qualitative enhancement of many Internet operations, making them much more intelligent.



The figure to the left illustrates the computational Heisenberg principle for cloud computing. The figure shows the available resources (a combination of computing power and storage capacity) as a function of the time interval they are used. For a computing cloud, the cost is approximately proportional to the product of the time t_0 and the amount of computing resources r_0. The figure shows a family of hyperbolas of equal cost $c=rt$. The light blue shading shows the local resources available for a maximum cost c_0. The dark blue

shading shows the additional resources available in the cloud, for the same maximum cost. The figure clearly shows the computational Heisenberg principle in action: for a given cost, as the time interval becomes shorter more resources are available. The cloud can absorb very large computing needs for short intervals. The dark blue area will be the home of new kinds of applications that did not exist before.

One simple operation that will be enhanced by the computational Heisenberg principle is voice communication. We predict that in the next ten years, voice communication will do real-time language translation (we suspect that Google is already hard at work implementing it!). You will be able to phone a correspondent in China and speak and hear English, and the correspondent will hear you in Chinese and reply in Chinese. Today's conflicts between world languages will become irrelevant. The pieces of this system already exist (for example, the IRCAM research institute has implemented many of them); all they need is resources. Computational and storage intensive algorithms, based on techniques of concatenative audio synthesis, will analyze your English speech, look up the appropriate fragments in an enormous database, and find and assemble the corresponding Chinese fragments in real time. The general approach depends on a large and growing corpus of words and phrases as used by real speakers, and it will rely on bilingual speakers to improve it. The approach has enormous computational and storage requirements, but clouds can provide them using the computational Heisenberg principle. If an average person spends 15 minutes per day on the phone (about 1/100 of their time), then each person will get an average of 100 times the resources that would be available to one person.

The Heisenberg Net will drive research in distributed systems for the next ten years. It will bring to the fore new concepts such as elasticity (the ability to scale resource usage up and down rapidly on demand) and fractal self management (the ability of a large system to manage itself at all scales). It will drive the invention of new programming abstractions to handle the new algorithms and their "scale-free" use of resources. As researchers, we look forward to help create the Heisenberg Net, and as users, we look forward to enjoy its fruits.

References

[1] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Andy Konwinski, Gunho Lee, David A. Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. "Above the Clouds: A Berkeley View of Cloud Computing". UC Berkeley, Technical Report UCB/EECS-2009-28, Feb. 10, 2009.
[2] European Commission Expert Group Report. *The Future of Cloud Computing: Opportunities for European Cloud Computing Beyond 2010*. Version 1.0. Editors: Keith Jeffery, Burkhard Neidecker-Lutz. Rapporteur: Lutz Schubert. Jan. 2010.