Letter from the Special Issue Editor

Fuelled by the advances virtualization and high-speed network technologies, cloud computing is emerging as a dominant computing paradigm for the future. Almost all technology assessment groups such as Forrester and Gartner project very aggressive growth in cloud computing over the next decade or two. Cloud computing can roughly be summarized as "X as a service" where X could be a virtualized infrastructure (e.g., computing and/or storage), a platform (e.g., OS, programming language execution environment, databases, web servers), software applications (e.g., Google apps), a service, or a test environment, etc. A distinguishing aspect of cloud computing is the utility computing model (aka pay-as-you-go model) where users get billed for the computers, storage, or any resources based on their usage with no up-front costs of purchasing the hardware/software or of managing the IT infrastructure. The cloud provides an illusion of limitless resources which one can tap into in times of need, limited only by the amount one wishes to spend on renting the resources. Cloud computing is by no means a novel/new thought. In a recent tutorial on cloud computing at EDBT 2012, Amr Abbadi & Divy Aggrawal pointed out one of the early references to cloud computing in a speech by John McCarthy at the MIT centennial in 1961 where he states "If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility Ĕ The computer utility could become the basis of a new and important industry". While McCarthy was possibly 40-50 years ahead in his projection, there is no doubt that the spirit of his statement is finally coming to fruition.

A sales pitch for cloud computing emphasizing its key characteristics could look something as follows: use as much as **your** needs dictate; pay **only** for what you use; don't **worry** about hiring staff to manage any system administration issues such as **loss of** data due to failures; and have better **control** over your IT investment (no up-front costs, cheaper due to economies of scale). Hidden in the sales pitch is what is, perhaps, the largest challenge facing cloud computing - **"your only worry is loss of control"**. Indeed, perception of loss of control over ones resources, whether it be infrastructure, software, or data, has been identified by many pundits as the important and immediate challenge facing cloud computing. The key operative issue here is the notion of trust. Loss of control, in itself, is not as much of an issue if clients/users could fully trust the service provider. In a world where service providers could be located anywhere, under varying legal jurisdictions; where privacy and confidentiality of ones data is subject to policies and laws that are at best (or under some circumstances) ambiguous; where policy compliance is virtually impossible to check, and the threat of "insider attacks" is very real - trust is a difficult property to achieve. Loss of control over resources (by migrating to the cloud) coupled with lack of trust (in the service provider) poses numerous concerns about data integrity, availability, security, privacy and confidentiality to name a few. Whether or not one migrates to the cloud depends upon how one balances the perceived risks (due to loss of control) with the benefits (the cloud offers) and this, in turn, depends upon the end users and their needs. It is perhaps fair to state that, given the widespread adoption of services such as email (e.g., Gmail), document management (e.g., Google Drive) on the internet, the end-users have already decided that the benefits outweigh the risks. The major question facing the cloud computing market is the extent to which small/medium/large organizations (including the government) - i.e. the "real" paying customers - will adopt cloud computing solutions. The answer to this question depends upon the perceived risks of migrating to the cloud and the organizations' risk-tolerance. The research community can significantly facilitate such a migration by developing technological solutions that help alleviate these risks.

This issue of the Data Engineering Bulletin focuses on the privacy and security aspects of outsourcing data to the cloud. The bulletin presents 11 papers by leading researchers who are exploring issues relevant to security, privacy, and confidentiality in cloud computing from different perspectives.

The bulletin starts with a paper by Montjoye, Wang, and Pentland that lays out a bold vision of a privacy preserving architecture for personal data sharing in the cloud based on trusted intermediaries. The second paper by Chen & Sion evaluates the economic viability of implementing secure outsourced data management

in untrusted clouds. Their thesis that today's cryptography and security solutions are simply not expressive enough to support outsourcing in a cost-effective way can be viewed as a call for innovative approaches that indeed offer practical solutions.

The next three papers focus specifically on cloud as a vehicle to offloading the complex task of information sharing. Nabeel & Bertino address an important problem of fine-grained access control based on a broadcast based group key management scheme that provides a scalable solution to selectively share data with others in an untrusted cloud. Anh and Datta define a design space for fine-grained access control solutions based on the level of access, the trust one has in the cloud, and how the work is split between the client and the cloud - this provides an elegant approach to viewing current solutions and exploring future challenges. Hamlen, Kagal, and Kantarcioglu identify an approach to policy enforcement wherein application code self-censor their resource accesses to implement efficient access control.

The following three papers explore mechanisms for secure data processing in the cloud. Khadilkar, Oktay, Kantarcioglu, and Mehrotra explores the design space for data and workload partitioning in hybrid clouds wherein in-house computing resources are integrated with public cloud services to support a secure and economical data processing solution. Mukundan, Madria, and Linderman tackle the challenge of data integrity, specifically, provable data possession in the presence of multiple replicas that enables owners to verify that cloud providers maintain the requisite number of replicas for data availability based on the service level agreement. Arasu, Blanas, Eguro, et. al. describe the Cipherbase relational database technology they are building at Microsoft that leverages novel customized hardware to store and process encrypted data. The last three papers in the series explore cloud in the role of applications as a service. Feldman, Blankstein, Freedman, and Felton introduce two cloud deployable application frameworks – SPORC (for collaborative applications such as text editor and shared calendars) and Frientegrity (that extends SPORC to online social networking) that do not require users to trust cloud providers with either confidentiality or integrity of data. Sanamrad, Wider, et. al. introduce a middleware approach that enables users to encrypt and store calendar entries and emails in Google Calendar and Gmail. Finally, the paper by De Cristofaro, Soriente, Tsudik and Williams describes a system they call hummingbird that implements a functionality similar to twitter using which users can tweet, follow and search while simultaneously protecting the tweet content, hashtags, and follower content from being exposed to the hummingbird service provider.

As we have become accustomed to in reading the Data Engineering Bulletins, the range of articles differ significantly in the level of their depth and treatment of the subject - while some lay out a vision for the future, other offer technically mature approaches based on significant prior work by the authors. Irrespective of the nature of the papers, collectively they provide a good summary of the state-of-the-art research and also a wealth of interesting new ideas/thoughts. This bulletin makes a wonderful reading for anyone interested in either learning about or intending to do research on what is possibly one of the most important challenges for computer science in the next decade.

Finally, I would like to acknowledge the generous help by Kerim Yasin Oktay in meticulously following up with the authors, collecting, formatting, and compiling the papers into the bulletin.

Sharad Mehrotra
University of California, Irvine