Secure Data Systems and Performance: Friends or Foe?

Manos Athanassoulis
Harvard University
manos@seas.harvard.com

The increasing trend in data collection and data processing has lead to the aggressive development of public clouds which store, process, and query data collections from completely heterogenous sources, with different requirements not only in performance, but also in confidentiality and security. The performance considerations have been leading to bigger public clouds where economy at scale allows for more efficient solutions. Today data management is being commoditized and offered as *database-as-a-service* over private and public cloud infrastructure.

At the same time the increasing number of data breaches over the past 15 years have raised the awareness of privacy and security. For a company, its data need to be private in order not to lose its business advantage, for a government it can be strategic economical and national security reasons, and for a scientist it is to protect work in progress before it is officially published. For a hospital, it may contain sensitive patient data which are protected by law. Even for an end user, data stored and processed on public clouds are sensitive because it may contain private information that the user wants to keep it so. The repetitive breaches, however, show that we are far from having strong guarantees that our data, and the processing we are doing over it, is safe-guarded in the existing private and public cloud infrastructure. We have recently seen sensitive banking information leaked (e.g., credit cards), private data leaked (e.g., photos), and government correspondence (e.g., emails). These occurrences beg the question:

What is the cost (in performance and/or functionality) to really have secure data systems?

There is a natural fight to the privacy of the data stored and the queries answered. Every encryption scheme inherently reveals a bit of information. Every query reveals a bit of information. Is there a way for these information pieces not to be combined?

Tunable Secure Data Systems. To address the variable requirements of different applications a system today needs to support varying levels of security which would allow variable performance. Data systems today have to balance between functionality, security, and performance. We envision the design of secure data systems that offer tunable security and privacy and, as a result, variable performance, while keeping the overall functionality of the system maximal. The ultimate goal is to build systems that can *support any application regardless of the desired level of security*. Tunable security and privacy – and as a result performance – would allow us to simply tune the system for any application, in a similar way we tune data systems for the underlying hardware and the access patterns of their applications.

Research Challenges. As database systems are increasingly offered as-a-service on the cloud, tunable secure data systems will also be a cloud offering. In traditional data system design, performance was the ultimate goal. However, recently, performance is balanced against other emerging goals; a first rival of performance was power, today, security is a key requirement. This creates a key research challenge, since now we have to bridge different application security requirements in the same system.

Security vs. Performance in Multi-tenancy. Security brings new aspects in the research challenges of multi-tenancy. Not only different levels of security should be supported, but also, different guarantees over possibly the same data should be provided. The goal here is to build a *tunable secure data system that offers different levels of security and privacy to different applications at the same time*. Apart from doing this by segregation of applications, another approach is to design hierarchical encryption schemes which can build one on top of the other. A key contribution is to store all data in a way to support the minimal security requirements and add additional levels of encryption to data and queries when needed in a computationally practical way. While this is a hard research challenge it is becoming increasingly important as we move towards a heavily cloud-based data processing world, where storage, processing, and eventually security are all offerred as-a-service.

Security vs. Performance in Indexing. The balance between security and performance manifests in accessing data, and in particular in indexing data. In fact, every time an index is used additional information is leaked to a possible curious listener. Hence, a key research goal is to *build secure indexes that consecutive accesses cannot be aggregated* to construct information for the dataset to the eavesdropper. Realizing this goal is an open research challenge. A key contribution is to build a continuously obfuscating index which uses, for example, user passwords, trusted hardware, and different levels of encryption in successive accesses.

Secure Data Systems. Over the last years database researchers have built data systems that aim to store, manage, and query data in a *secure* way. Two prominent examples are CryptDB and Cipherbase. *Tunable secure data systems* go beyond the state of the art; the goal is to provide tunable and modular security in a cloud environment with multi-tenancy. This would allow the same infrastructure to support multiple applications with different performance and security requirements, and this would happen through tuning which would affect both performance and security guarantees.

8th Biennial Conference on Innovative Data Systems Research (CIDR '17) January 8-11, 2017 , Chaminade, California, USA.