Dan Wnuk

Developing Secure Software

12/14/2011

Securing Database as a Service Review & Summary

This article starts out talking about cloud computing and its benefits. Costs to maintain are much lower, because companies no longer need to have their own datacenter and hire people to maintain both the datacenter its servers. From here, it goes on to describing cloud computing and how currently it is relatively undefined. It then goes on to define three separate sub-types of cloud computing – Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). IaaS is described as providing the building, internet connection and necessary power, but customers have to provide their own hardware. PaaS is defined as providing the servers, but giving customers the control to put whatever software on the boxes that they desire. SaaS is providing access to specific applications, such as Pandora providing their software as a service, allowing customers to pay for better features. It then goes on to describe that Database as a Service (DaaS) falls under the last category because a database is simply a piece of software running on a server.

Once everything is defined, the article goes on to list some of the problems with current database implementations. Most people want to encrypt their data to keep it confidential, but if someone were to encrypt an entire database, the amount of time to access the database and perform operations on it goes up drastically because you are introducing the encrypting algorithms into finding the data. The article goes on to state that this makes the database unusable. Another pitfall listed is that most companies offering Database as a Service are hosting many databases, and so instead of directly dealing with each of the databases, they usually have a relatively simple abstraction on top, a type of management system. There are new types of databases that are attempting to address modern problems with databases as a service, such as Microsoft SQL Azure.

Cloud computing still has not been widely accepted, however. The problem with security and wondering whether the data that is being uploaded to a third party company is truly secure is only one of the problems. Another problem is the possibility of downtime. The article mentions the Amazon outages that occurred not too long ago and took down a lot of popular sites, such as Reddit. Another issue is whether or not the company hosting their data will last as long as needed. Additionally, people worry about other people that are using the same shared resources. What if they are breaking the law, and you are associated with them somehow? On top of that, sharing resources with other people in it can be a deterrent. Another instance is the privacy of the data – from others using the service, the provider itself as well as potential hackers of the system, as a recent debacle with Sony has brought to light.

The article also mentions Service-Level Agreements, which are the contract between the provider and customer, and says that customers should leverage these to get the most security out of the provider, because it is legally binding. After this, they go into Database-specifics. The first hurdle they continue to explore is encryption, where they mention that while research is being done on querying encrypted databases, the performance and reliability is not yet there. There are several techniques being used – binning techniques, privacy homomorphism encryption, early stopping comparisons and separate encrypted indexes. The article doesn't explain what these are exactly, but say

that the homomorphism encryption is good for summing data, but offers poor performance for joining. Another interesting method is to split up encrypted data among multiple hosts that cannot communicate with one another, and only the owner can access all three and combine them. This is a very interesting concept, and seems like something that could be extremely useful. While faster, this does require multiple providers, which could be a blow to the wallet.

The next area discussed is key management. Keys likely should not be stored with the database because it can then be decrypted, so it must be stored locally, which sort of defeats the purpose of moving to the cloud. There have been some new techniques which allow for hosting in the cloud, but it has yet to be applied to database encryption. Other issues included are authenticity – how do you figure out whether the data obtained from the database are authentic? There are some techniques being researched in this area, but nothing concrete.

Data pollution is another serious issue. Since most providers put each customer's services in a virtual machine, and these virtual machines are running on the same box, a malicious VM could attack other VM's. There currently aren't any known attacks out there like this, but it has been proven that they are possible. Other ways of data pollution include information for one process being accidentally injected into another process. Memory is the final issue mentioned. When something is decrypted, even temporarily, that exists in the memory of the system. Another malicious process could be running on the computer reading the memory, which would give them access to the decrypted data.

This document taught me about the various types of cloud computing services, and the issues related to them. A lot of the material was stuff I already had concerns about – such as, with your data being hosted somewhere else, does it really belong to you? Alternatively, using third party services for potential illegal activities would mean that they could get a court order to force the provider to give your information away, which is another potential issue. Several new topics in regards to databases were brought up – such as issues with encryption and the current slowness behind the algorithms for reading or searching on encrypted data. This was never mentioned in our databases course, but it makes perfect sense that attempting to encrypt every query and then decrypt results would be very time consuming, especially as the length of encryption keys becomes larger and larger.

This learning can be applied to my future by taking into account the types of issues related to cloud computing and weighing the decisions heavily on reported issues. Since encryption is a topic of interest of mine, these newly discovered issues (in my mind, at least) could even bring about a new area of study for me. I was particularly fascinated by the idea to split up encrypted data among multiple providers, and only when pulling the data and combining can it be decrypted. This seems like a great idea and potentially where the industry could go if hosting becomes even cheaper. Overall, the reading simply helps me better understand cloud computing and database issues, where they are heading and will help me make better, more informed decisions relating to these topics.

Article Information

Title: Securing Database as a Service: Issues and Compromises

Authors: Joel Weis & Jim Alves-Foss

Date Accessed: 12/14/2011

Website URL: http://www.computer.org/portal/web/csdl/abs/mags/sp/2011/06/msp201106toc.htm

PDF URL: http://origin-www.computer.org/plugins/dl/pdf/mags/sp/2011/06/msp2011060049.pdf?template=1&loginState=2&userData=Milwaukee%2BSchool%2Bof%2BEngineering%253AMilwaukee%2BSchool%2Bof%2BEngineering%253AAddress%253A%2B155.92.43.234%252C%2B%255B140.98.196.191%252C%2B155.92.43.234%252C%2B206.169.246.148%255D