ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AN D ENGINEERING TECHNOLOGY (IJRASET)

Issues of Privacy and Confidentiality in Cloud Computing

Mohd. Junedul Haque
College of Computers and Information Technology
Taif University
Taif, Saudi Arabia

Abstract—Cloud computing means entrusting data to information systems that are managed by external parties on remote servers "in the cloud."Webmail and online documents (such as Google Docs) are well-known examples. Cloud computing raises privacy and confidentiality concerns because the service provider necessarily has access to all the data, and could accidentally or deliberately disclose it or use it for unauthorized purposes. Conference management systems based on cloud computing represent an example of these problems within the academic research community. It is an interesting example, because it is small and specific, making it easier to explore the exact nature of the privacy problem and to think about solutions.

Keywords- Software as a Service (SaaS), PaaS (Platform as a Service), OPEX[operating expenditures], Elastic Compute Cloud (EC2), EDAS

INTRODUCTION

Cloud computing [1][2] refers to both the applications delivered as services over the Internet and the hardware and systems software in the data enters that provide those services. The services themselves have long been referred to as Software as a Service (SaaS). Some vendors use terms such as IaaS (Infrastructure as a Service) and PaaS (Platform as a Service) to describe their products, but we eschew these because accepted definitions for them still vary widely. The line between "low-level" infrastructure and a higher-level "platform" is not crisp. We believe the two are more alike than different, and we consider them together. Similarly, the related term "grid computing," from the high-performance computing community, suggests protocols to offer shared computation and storage over long distances, but those protocols did not lead to a software environment that grew beyond its community. In this paper a case study of conference management system is taken and issues, data privacy concerns, ways forward are discussed in different sections.

Everyone has an opinion on what is cloud computing. It can be the ability to rent a server or a thousand servers and run a geophysical modelling application on the most powerful systems available anywhere. It can be the ability to rent a virtual server,load software on it, turn it on and off at will, or clone it ten times to meet a sudden workload demand. It can be storing and securing immense amounts of data that is accessible only by authorized applications and users. It can be supported by a cloud provider that sets up a platform that includes the OS, Apache, a MySQL™ database, Perl, Python, and PHP with the ability to scale automatically in response to changing workloads.

Cloud computing can be the ability to use applications on the Internet that store and protect data while providing a service anything including email, sales force automation and tax preparation. It can be using a storage cloud to hold application, business, and personal data. And it can be the ability to use a handful of Web services to integrate photos, maps, and GPS information to create a mash up in customer Web browsers [8].

According to Gartner's Hype Cycle Special Report for 2009, "technologies at the 'Peak of Inflated Expectations' during 2009 include cloud computing, e-books… and Internet TV, while social software and micro blogging sites…have tipped over the peak and will soon experience disillusionment among enterprise users". The Internet is often represented as a cloud and the term "cloud computing" arises from that analogy.

ISSN: 2321-9653

McKinsey says that clouds are hardware based services offering compute, network and storage capacity where: hardware management is highly abstracted from the buyer; buyers incur infrastructure costs as variable OPEX [operating expenditures]; and infrastructure capacity is highly elastic (up or down) [9].

Large companies can afford to build and expand their own data centres but small- to medium-sized enterprises often choose to house their IT infrastructure in someone else's facility. A colocation centre is a type of data centre where multiple customers locate network, server and storage assets, and interconnect to a variety of telecommunications and other network service providers with a minimum of cost and complexity.

Cloud Architectures are designs of software applications that use Internet-accessible on-demand services. Applications built on Cloud Architectures are such that the underlying computing infrastructure is used only when it is needed (for example to process a user request), draw the necessary resources on-demand (like compute servers or storage), perform a specific job, then relinquish the unneeded resources and often dispose themselves after the job is done.

While in operation the application scales up or down elastically based on resource needs.

RELATED WORK

In IaaS, cpu, grids or clusters, virtualized servers, memory, networks, storage and systems software are delivered as a service. Perhaps the best known example is Amazon's Elastic Compute Cloud (EC2) and Simple Storage Service (S3), but traditional IT vendors such as IBM, and telecoms providers such as AT&T and Verizon are also offering solutions. Services are typically charged by usage and can be scaled dynamically, i.e. capacity can be increased or decreased more or less on demand. PaaS provides virtualized servers on which users can run applications, or develop new ones, without having to worry about maintaining the operating systems, server hardware, load balancing or computing capacity. Well known examples include Microsoft's Azure and Salesforce's Force.com. Microsoft Azure provides database and platform services starting at $0.12 per hour for compute infrastructure; $0.15 per gigabyte for storage; and $0.10 per 10,000 transactions. For SQL Azure, a cloud database, Microsoft is charging $9.99 for a Web Edition, which comprises up to a 1 gigabyte relational database; and $99.99 for a Business Edition, which holds up to a 10 gigabyte relational database. For .NET Services, a set of Web-based developer tools for building cloud-based applications, Microsoft is charging $0.15 per 100,000 message operations.

SaaS is software that is developed and hosted by the SaaS vendor and which the end user accesses over the Internet. Unlike traditional applications that users install on their computers or servers, SaaS software is owned by the vendor and runs on computers in the vendor's data centre (or a collocation facility). Broadly speaking, all customers of a SaaS vendor use the same software: these are one-size-fits-all solutions. Well known examples are Salesforce.com, Google's Gmail and Apps, instant messaging from AOL, Yahoo and Google, and Voiceover Internet Protocol (VoIP) from Vonage and Skype.

CASE STUDY: CONFERENCE MANAGEMENT SYSTEMS

Most academic conferences are managed using software that allows the program committee (PC) members to browse papers and contribute reviews and discussion via the Web. In one arrangement, the conference chair downloads and hosts the appropriate server software, say HotCRP or iChair.

The benefits of using such software are familiar:

- Distribution of papers to PC members is automated, and can take into account their preferences and conflicts of interest.
- The system organizes the collection and distribution of reviews and discussion, can rank papers according to scores, and send out reminder email, as well as email notifications of acceptance or rejection; and
- It can also produce a range of other reports, such as lists of sub-reviewers, acceptance statistics, and the conference program.

HotCRP and iChair require the conference chair to download and install software, and to host the Web server. Other systems such as EasyChair and EDAS work according to the cloud computing model: instead of installing and hosting the server, the conference chair simply creates the conference account "in the cloud." In addition to the benefits described previously, this model has extra conveniences:

- The whole business of managing the server (including backups and security) is done by someone else, and gains economy of scale;

- Accounts for authors and PC members exist already, and don't have to be managed on a per-conference basis;
- Data is stored indefinitely, and reviewers are spared the necessity of keeping copies of their own reviews;

The system can help complete forms such as the PC member invitation form and the paper submission form by suggesting likely colleagues based on past collaboration history. For these reasons, EasyChair and EDAS are an immense contribution to the academic community. According to its Web page, EasyChair hosted over 3,300 conferences in 2010.

Because of its optimizations for multi conferences and multitrack conferences, it is mandated for conferences and workshops that participate in the Federated Logic Conference (FLoC), a huge multi conference that attracts approximately 1,000 paper submissions.

DATA PRIVACY CONCERNS

i. Accidental or deliberate disclosure:

A privacy concern with cloud-computing- based conference management systems such as EDAS and EasyChair arises because the system administrators are custodians of a huge quantity of data about the submission and reviewing behavior of thousands of researchers, aggregated across multiple conferences. This data could be deliberately or accidentally disclosed, with unwelcome consequences [5].

- Reviewer anonymity could be compromised, as well as the confidentiality of PC discussions.
- The acceptance success records could be identified, for individual researchers and groups, over a period of years; and
- The aggregated reviewing profile (fair/unfair, thorough/scant, harsh/undiscerning, prompt/late, and so forth) of researchers could be disclosed. The data could be abused by hiring or promotions committees, funding and award committees, and more generally by researchers choosing collaborators and associates. The mere existence of the data makes the system administrators vulnerable to bribery, coercion, and/or cracking attempts. If the administrators are also researchers, the data potentially puts them in situations of conflict of interest. The problem of data privacy in general is of course well known, but cloud computing magnifies it. Conference data is an example in our backyard.

When conference organizers had to install the software from scratch, there was still a risk of breach of con rather confidentiality, but the data was just about one conference. Cloud computing solutions allow data to be aggregated across thousands of conferences over decades, presenting tremendous opportunities for abuse if the data gets into the wrong hands.

ii. Beneficial data mining:

In addition to the abuses of conference review data described here, there are some uses that might be considered beneficial. The data could be used to help detect or prevent fraud or other kinds of unwanted behavior, for example, by identifying:

- Researchers who systematically unfairly accept each other's papers, or rivals who systematically reject each other's papers, or reviewers who reject a paper and later submit to another conference a paper with similar ideas; and
- Undesirable submission patterns and behaviors by individual researchers (such as parallel or serial submissions of the same paper; repeated paper withdrawals after acceptance; and recurring content changes between submitted version and final version).

The data could also be used to understand and improve the way conferences are administered. ACM, for example, could use the data to construct quality metrics for its conferences, enabling it to profile the kinds of authors who submit, how much "new blood" is entering the community, and how that changes over different editions of the conference. This could help identify conferences that are emerging as dominant, or others that have outlived their usefulness. The decisions about who is allowed to mine the data, and for what purposes, are difficult. Policies should be decided transparently and by consensus, con rather than being left solely to the de facto data custodians.

ISSN: 2321-9653

WAYS FORWARD

i. Policies and legislation:

An obvious first step is to articulate clear policies that circumscribe the ways in which the data is used. [3][4] For example, a simple policy might be that the data gathered during the administration of a conference should be used only for the management of that particular conference. Adherence to this policy would imply that the data is deleted after the conference, which is not done in the case of Easychair (I don't know if it is done for EDAS). Other policies might allow wider uses of the data. There is no privacy policy linked from its main page, and a search for "privacy policy" (or similar terms) restricted to the domain "easychair.org" does not yield any results. In most countries, legislation exists to govern the protection of personal data. In the U.K., the Data Protection Act is based on eight principles, including the principle that personal data is obtained only for specified purposes and is not processed in a manner incompatible with the purposes; and the principle that the data is not kept longer than is necessary for the purposes. EasyChair is hosted in the U.K., but the lack of an accessible purpose statement or evidence of registration under the Act mean I was unable to determine whether it complies with the legislation. The Data Protection Directive of the European Union embodies similar principles; personal data can only be processed for specified purposes and may not be processed further in a way incompatible with those purposes.

ii. Processing encrypted data in the cloud:

Policies are a first step, but alone they are insufficient to prevent cloud service providers from abusing the data entrusted to them. Current research aims to develop technologies that can give users guarantees that the agreed policies are adhered to. Hardwarebased security initiatives such as the Trusted Platform Module and Intel's Trusted Execution Technology are designed to allow a remote user to have confidence that data submitted to a platform is processed according to an agreed policy. These technologies could be leveraged to give privacy guarantees in cloud computing in general, and conference management software in particular. However, significant research will be needed before a usable system could be developed. Certain cloud computing applications may be primarily storage applications, and might not require a great deal of processing to be performed on the server side. In that case, encrypting the data before sending it to the cloud may be realistic. It would require keys to be managed and shared among users in a practical and efficient way, and the necessary computations to be done in a browser plug-in. It is worthwhile to investigate whether this arrangement could work for conference management software.

CONCLUSION

It has been argued that the professional honor of data custodians (and PC chairs and PC members) is sufficient to guard against the threats described. Indeed, adherence by professionals to ethical behavior is essential to ensure all kinds of confidentiality. In practice, system administrators are able to read all the organization's email, and medical staff can browse celebrity health records; we trust our colleagues' sense of honor to ensure these bad things don't happen. But our standpoint is that we should still try to minimize the extent to which we rely on people's sense of good behavior. We are just at the beginning of the digital era, and many of the solutions we currently accept won't be considered adequate in the long term. The issues [6][7] raised about cloud computing- based conference management systems are replicated in numerous other domains, across all sectors of industry and academia. The problem of accumulations of data on servers is very difficult to solve in any generality.

The particular instance considered here is interesting because it may be small enough to be solvable, and it is also within the control of the academic community that will directly benefit or suffer according to the solution we adopt.

REFERENCES

[1] K. Kumar and L. Yung-Hsiang, "Cloud Computing for Mobile Users: Can Offloading Computation Save Energy?," IEEE Computer , vol.43, no.4, pp.51-56, April 2010. doi: 10.1109/MC.2010.98.

[2] Brian Hayes. Cloud computing. Commun. ACM, 51(7):9–11, 2008.

[3] Armbrust M et al, "Above the Clouds: A Berkeley View of Cloud Computing", UC Berkeley Reliable Adaptive Distributed Systems Laboratory Technical Report, February 2009.

[4] Ken Birman, Gregory Chockler, and Robbert van Renesse. Toward a cloud computing research agenda. SIGACT News, 40(2):68–80, 2009.

[5] Alexander Lenk, Markus Klems, Jens Nimis, Stefan Tai, and Thomas Sandholm. What's inside the cloud? an architectural map of the cloud landscape. In Proc. Of the

ICSE Workshop on Software Engineering Challenges of Cloud Computing, pages 23–31, 2009.

[6] M. A. Vouk, "Cloud computing - Issues, research and implementations," presented at the ITI 2008 30th International Conference on Information Technology Interfaces, 2008.

[7] J. Voas and J. Zhang, "Cloud Computing: New Wine or Just a New Bottle?," IT Professional, vol. 11, p. 3, Mar./Apr. 2009.

[8] Jason Carolan and Steve Gaede, Introduction to Cloud Computing architecture, SUN Microsystems Inc., pp. – 1-40, June 2009.

[9] McKinsey & Co. Report presented at Uptime Institute Symposium, “Clearing the Air on Cloud Computing”,April 18, 2009