Comparison between security solutions in Cloud and Grid Computing

David Munoz Sanchez
Helsinki University of Technology
dmunozsa@tkk.fi

Abstract

Purpose: This paper helps Grid and Cloud system administrators and researchers to have a clear definition about Grid and Cloud computing and also wants to give clear differences between Cloud and Grid computing, define which are the main security issues to be considered, and how some of these security issues are solved. It also shows how existing security methods used in one of these technologies can be used in the other.
Design/methodology/approach: A study about 11 different papers has been done in order to find common security solutions.
Research limitations/implications: This is research done for the subject T-110.5290 Seminar in network security in the Master's Programme in Mobile Computing - Services and Security. No practical demonstrations were done while writting this paper.
Practical implications: The findings of this study can be used by researchers of Grid and Cloud applications to develop new security solutions. This paper demonstrate that technologies such as TPM[9] and trust relationship management could be used in both technologies.
Keywords: Cloud Computing, Grid Computing, Security, Mobile services, research, reusability.
Paper type: Research paper

1 Introduction

Cloud computing is currently a buzzword in the world of computing, and a few moments of browsing the Web will soon reveal an extensive range of Cloud services and products. The reason for this great interest is that it offers what some earlier technologies have wanted to do, namely, access to processing and storage power on demand. Yet, despite its high profile, many people do not really know what it is.

In essence, Cloud computing is the outsourcing of what has been described as an unlimited range of copmuter resources, for instance processing power and storage capacity, by means of virtualization, i.e. It lets a private user to do the simulation of artificial intelligence tests that consume lots of resources in little time.

Cloud computing appeas to all sorts of users, from businesses to private individuals, for various reasons. Chief among the attractions is the fact that these users have the opportunity to pay as they go: pay-per-use(PPU) or pay-on-demand(POD). This method of paying is not new. Indeed, we have been using this business model so many years; for instance, electricity suppliers employ this method to charge for electricity used at home, where households pay for as much as they use, instead of paying monthly. Another example would be for pay-as-you-go for mobile phones. What the PPU system offers, however, is flexibility: we only pay for what we use and when we use it.

What is also not as novel as we may think is that Cloud computing shares a number of similar issues with older technologies, such as Grid Computing, in which the resources of many computers in a network are utilized simultaneously, making it possible for users(companies, researchers, governments, etc. to store, exploit and retrieve masive amounts of data.

Cloud computing may seem to offer great benefits, and both, individuals and organisations are adopting it rapidly. However, there are problems associated with this system/environment. The main problem according to Ian Foster, Yong Zhao, Ioan Raicu and Shiyong Lu[3] is that, everytime a company wants to provide a service, there arises a common need to use and manage large facilities. This kind of managing of systems is used for using and requesting resources that are provided by central facilities like datacentres. Highly paralel computations are implemented in these environments and executed through thes distributed resources.

What has happened is that we have created a new environment in which vast amounts of information,etc. are being distributed through a large number of computers. One consequence of this is the issue of security. The large and extensive scale of Cloud computing activities require effective security.

One source of solutions is found in existing technologies, therefore, if we take into consideration that Cloud Computing and Grid Computing have a lot of architectural and design issues in common, we should be able to, based on one study of the security in both technologies, reuse some of the existing security solutions and thus avoid reinventing the wheel.

The purpose of this paper is to explore the main security problems in Cloud and Grid Computing and compare them: define which common security solutions, used only in one of the technologies, would be valid for both systems.

2 Background

This section summarizes the basic concepts of Cloud and Grid Computing before we go deeper into security details in section 3.

2.1 What is Cloud Computing and what are its benefits?

One good definition of Cloud Computing is given by Y. Zhao, M. Wilde, I. Foster, J. Voeckler, J. Dobson, E. Gilbert, T. Jordan and E. Quigg in 2005: "A Large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scallable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet"[10].

Cloud Computing offers its users services and applications, which are provided through the Internet, and thus, a large number of computers will be in the path of the data when the data is sent to the Cloud for analysis, and, also, when the data is coming back from the Cloud, once the analysis has been already carried out. For instance, a Mobile phone or PDA with low processing power and/or low storage capacity could take advantage of the Cloud and store or process data inside it, which would allow the mobile phone or the PDA to use applications which require more processing power than the processing power available in these kinds of electronic devices and, which will also require a permanent broadband connection in order to send and receive data at any moment from any place, thus allowing the user to access the services wherever you are.

As another example of possible users that would be interested in Cloud Computing, there are lots of small and medium-sized companies that want to provide users with a new service which requires a lot of processing power, storage power and even networking capacity. To establish this infrastructure, the company should invest a lot of money in a large number of computers to process all the incoming requests, even if the peak load appears only once in a while and for these companies it makes it impossible to deploy the service due to the lack of resources. To solve this, one very small company which wants to provide a new service to a large number of customers could either use its own private Cloud (which can be either one or two computers, for example) and then, every time a peak demand appears, use the public Cloud to process the data (one example is Amazon EC2) or directly put its server into the Cloud and pay for the amount of processing power or storage capacity used at the end of the month: instead of contracting two servers for the full month, when required, they could contract one small server and increase the processing capacity through the Cloud processing and/or storage services.

As stated by David Linthicum [3], Cloud Computing can be given in some different ways:

- **Storage-as-a-service:** It offers disk space on demand.
- **Database-as-a-service:** Gives the possibility of using remote services which other users might be using at the same time.
- **Information-as-a-service:** Allows access to data stored in a remote server.
- **Process-as-a-service:** To be able to have processes running in the Cloud.
- **Application-as-a-service:** It means to have one application provided by the Cloud provider available on demand through the Cloud, for instance, a very heavy application that requires a lot of storage or processing power.
- **Platform-as-a-service:** SDK provided by the Cloud provider that allows the programmer to develope new applications that will run in the Cloud.
- **Integration-as-a-service:** Allows to have an integrated system which acts like a intermediary between the data given by the company and the Software allocated into the Cloud(This software can make some XML parsing or even format conversion).
- **Security-as-a-service:** Grants the possibility to have security applications running directly in the Cloud: for instance, an antivirus system.
- **Management-as-a-service:** Service that allows the controlling and managing of multiple Clouds from one place.
- **Testing-as-a-service:** Allows us to make testing through specific testing software running in the Cloud, for example, testing Artificial intelligence software that requires lots of processing power.
- **Infrastructure-as-a-service:** Provides datacenter-as-a-service, which means to have access remotely to computing resources.

About the computing model in Cloud Computing, according to Ian Foster, Yong Zhao, Ioan Raicu and Shiyong Lu[4]," the resources in the Cloud are shared by all users at the same time, which will allow latency sensitive applications to operate inside these Clouds, but the QoS assurance is not an easy issue". In addition, Cloud computing uses virtualization in order to share resources. The reasons for Cloud Computing to rely in virtualizations are[4]: 1)It allows to execute multiple applications on the same server. 2)The configuration of the resources for each application can be easily managed. 3)If more resources are needed, the application can be moved to one or more servers with more resources. 4)Resource, provisioning, monitoring and maintenance can be automated, and common resources can be cached and reused. 5)Processor manufacturers introduced hardware support for virtualization, which is why the performance of this method has developed further.

2.2 What is Grid Computing and what are its benefits?

Alternatively we have Grid Computing, which is well defined by Mary Humphrey and Mary R. Thomson: "A Computational Grid is a collection of heterogeneous computers and resources spread across multiple administrative domains with the intend of providing users easy access to these resources."[6]

Ian Foster[5] made a three point checklist to define what a Grid is. In this list he said:

- **"coordinates resources that are not subject to centralized control" :** Which means that not only one entity manages all the system but some different system administrators could be managing different parts of the same Grid at the same time.
- **"using standards, open, general-purpose protocols and interfaces" :** This will allow to all the companies involved in the Grid to use and access these standards.
- **"to deliver nontrivial qualities of service" :** In Grid computing not a fixed rate of load is going to be managed and this can be given small or big rates. This phenomenon causes the given quality of service to vary and not always stay constant.

Grids allow the use of idle resources. Through this, companies create a Grid in order to share those idle resources and, if necessary, they can access more computational resources(shared by other companies) than they usually can, and share their own resources while they are not carrying out any computationally demanding tasks.

About the architecture, a lot of heterogeneous hardware is used in order to create the Grid and, in addition, these devices are not managed by only one person but by different system administrators in each of the companies. This situation causes the security, administration policies and network managing to become heterogeneous too, thus more difficult to manage.

However, not only companies can make use of these Grids; users can do it too. One successful example would be the Seti@home project, a program that attempts to find extraterrestial life, where users install a program and, while not using the resources, idle resources are used to process small packets or tasks, and return the processed information to the main server.

Finally, if we take a look at what Ian Foster, Yong Zhao, Ioan Raicu and Shiyong Lu[4] have stated: "most Grids use a batch-scheduled compute mode, in which a local resource manager (LRM), such as PBS Condor, SGE manages the compute resources for a Grid site and users submit batch jobs (via GRAM) to request some resources for some time.", this necessitates having a fair scheduling system that avoids that one process from using all the resources and leaving none for the rest.

2.3 What are the common issues between Cloud and Grid Computing?

As stated earlier, Cloud and Grid Computing are similar technologies that share lots of issues. The most important things that both technologies have in common:

1. To achieve good scalability, data must be distributed over many computers[4].
2. People can be afraid of sending sensitive data through a large number of computers.[4].
3. Data must be moved repeatedly to distant computers, which generates the bottleneck of the process, since the data is not always available everywhere and sometimes it is necessary to make this data available[4]
4. Data can be requested regardless of its location[4].
5. "Cloud and Grid computing provide service-level agreements (SLAs) for guaranteed uptime availability of, say, 99 percent. If the service slides below the level of the guaranteed uptime service, the consumer will get service credit for receiving data late. " [6]
6. Both systems must be able to determine the amount of unused resources[6].

2.4 What are the differences between Cloud and Grid Computing?

Since our main objective is to know the difference between Cloud and Grid computing and afterwards compare them from the point of view of security, here we have a list of differences between Grid and Cloud in order to clarify what are the main differences between Cloud and Grid computing:

1. Cloud computing normally runs in a set of homogeneous computers, but Grid, on the other hand, runs on heterogeneous computers[4].
2. Grid computing is normally focused on an intensive calculus, while Cloud Computing offers two types of calculuses: standard and intensive.[6].
3. Grid computing is open-source while Cloud Computing is not.[7]. This creates interoperatibility problems between today's Clouds[4].
4. Most Grids use a batch-scheduled compute model, while in Cloud Computing all users share all the resources at the same time. And this is why latency sensitive applications, which could run in Grids, could not be supported in Cloud Computing[4].
5. Grids do not rely on virtualization as much as Cloud does[4].
6. Cloud Computing does not support as much provenance as Grid does[4], which is a "derivation history of a data product, including all the data sources, intermediate data products,and the procedures that were applied to produce the data product"[4].
7. High Performance computing is better supported in Grid computing than in Cloud Computing[4].
8. Grid is distributed, has multiple research user communities (including users accessing resources from varied administration domains) and is grouped in Virtual Organisations[7]. Cloud Computing, on the other hand, usually has only one research community and a common group of system administrators that take care of the entire Cloud.
9. Grid is mostly publicly funded at local, national and international levels[7], while Cloud Computing is funded mainly by its users.

10. Grid computing, must share user and resource interfaces to allow providers to connect their resources, while Cloud Computing attempts to share only the user interface while the resources interfaces are hidden and given in an abstract way.

11. In Grid computing, the trust model is based on identity delegation, where users can access and browse resources,which are not highly abstracted and virtualized at different Grid sites. In Cloud computing resources are abstracted and virtualized and this trust model does not exist since everything is inside the same Cloud.

12. While in Grid computing, storing very small files (e.g. 1-byte files) is not economically suitable, in Cloud Computin it is.[6]

3 Security issues in Cloud computing

As with every distributed system, Cloud Computing has lots of problems. We have to take care of the network infrastructure, which is not always in our control, and be very careful with the data in order to avoid third parties from capturing it.

Some security solutions and problems have been proposed by ArmMichael Halton[2]:

- Web application vulnerabilities: Cross scripting, SQL injections. **Solution:** Develop a security oriented framework that teaches the best programming practices.

- Vulnerabilities inherent to the TCP/IP stack and/or the operating systems: DoS, and DDos(Distributed denial of service). **Solution:** Deactivate unused services, update applications and control rights.

- Authentication problems: IP spoofing, RIP attacks, ARP poisoning. **Solution:** Use encrypted protocols if possible prevents IP spoofing, controlling rights to access ARP tables,etc.

- The verification, tampering and loss of data.**Solution:** encrypted data would be a solution, but, 'since the unencrypted data must reside in the memory of the host running the computation', this must be encrypted in order to avoid memory copies[9].

- Physical access.**Solution:** Control rights and log actions when accessing the hardware.

- Privacy control of data.**Solution:** Use Service-level agreements.

Another solution, given by the Trusted Computing Group(TGC)[10] is the Trusted platform module(TPM): Chip with Private key and cryptographic algorithms(not rewritable). The public key is signed by the manufacturer to guarantee the correctness of the key. It protects the booting process through hashes of the SW involved in the boot sequence[9].

As stated by Nuno Santos,Krishna P., Gummadi Rodrigo Rodriguez[9], "customers cannot protect their VMs on their own and "anyone with privileged access to the host can read or manipulate customers' data". These insiders create a need to have a "technical solution that guarantees the confidentiality and integrity of computation, in a way that is verifiable by the customers of a service". So, they[9] proposed a new system called TCCP(Trusted Cloud Computing Platform), which is mainly focused in IaaS(Infrastructure as a Service) and based on two components:

- To install a TVMM, each node uses a TPM[10] to ensure the boot process.

- Trusted virtual machine monitor(TVMM): Prevents privileged users from inspecting or modifying VMs and protects its own integrity.

- Trusted Coordinator(TC): Manages which nodes are trusted and which are not through a record. A node, therefore, only cooperates with other trusted nodes.

Some other issues were considered by Ian Foster, Yong Zhao, Ioan Raicu and Shiyong Lu[4]:

- Monitoring can be less important because Cloud Computing has an abstract layer that could respond to failures and qualities of service automatically.

- "Cloud Computer security infrastructure typically relies on Web forms(over SSL) to create and manage account information for end users, and allows users to reset their passwords and receive new passwords via email through unsafe and unencrypted communication"[4]

In addition, taking into account that Cloud Compuging relies on web page management where the user can register himself into a web page and have almost instant access to the resources through a credit card, a credit card validation system is needed[6] and some other issues like web-based vulnerabilities prevented.

4 Security issues in Grid computing

As with Cloud Computing, Grid Computing has a lot of security issues that should be considered since it is a distributed system where a heterogeneous set of computers share their idle resources.

Marty Humphrey,Mary R. Thompson[8] defined various scenarios and gave various clues on how Grid security should be managed and what are its problems. Here is a list of the main challenges and solutions:

- A machine is sharing its resources and the user is running applications. Then it is needed to assure that the machine has not been compromised. **Solution:** A specialized scheduler that allows users with enough rights to run applications.

- Local user ID and Grid user ID must be mapped. **Solution:** It can be done through the use of centralized domain controllers, for instance, OpenLdap, that provide user authentication and authorization methods. Others[8] give another solution: a single Grid sign-in mechanism.

- Access to logs that are controlled by various users. **Solution:** To do this special security libraries can be used to access the data and control who did what.

- Determine access policies to services either locally or remotely. **Solution:** The authorization policy must locally be digitally signed by the owner and kept securely. Remotely, the owner must be able to have a secure connection and authenticate himself.

- Data integrity and confidentiality should be achieved. **Solution:** Integrity is achieved through MAC algorithms. Confidentiality is achieved through encryption methods and keys with a limited life time.

- Proper key management. **Solution:** One possibility is to use smart cards.

- Trust relationships between users and domains/hosts become imperative. **Solution:**Authentication is achieved by SSL credentials or secure DNS and IPSec.

- Delegation of rights to one or multiple persons is a problem with no clear solution yet.

- Information must be available and can be requested from everywhere. **Solution:** So in order to have availability, information services are used. LDAP can be used to these purposes since it gives user/password access control and can map the user's id to his service's directory.

- Firewalls and VPN's between Grid's domains became a challenge. **Solution:** Infrastructure servers can be configured to run on known ports which can be allowed by the firewall. In the case of VPN, certificates like x509 identity certificates would be a good solution to allow access to other Grid domains.

- Physical access security has to be considered as well.

As with Cloud Computing, some other issues about Grid Computing have been considered by Ian Foster, Yong Zhao, Ioan Raicu and Shiyong Lu[4] as well:

- "Public-key based GSI (Grid Security Infrastructure) protocols are used for authentication, communication protection and authorization"[4].

- Grid computing is more heterogeneous and has dynamic resources, which is why it should address some issues: A single sign-in method to access multiple Grid sites, privacy, integrity and segregation should be taken into account so that resources owned by one user cannot be accessed by unauthorized users and/or tampered with during transfer[4]

- Community Authorization Service(CAS) is used for advanced resource authorization within and across domains[4]

5 A comparison between Cloud and Grid Computing security solutions

Nowadays, and knowing that Grid Computing is a much more mature technology than Cloud computing is, we would agree that the security is better in Grid than in Cloud Computing. This thought is shared by Ian Foster, Yong Zhao, Ioan Raicu and Shiyong Lu[4] who said that "[The] Cloud Computing security model seems to be relatively simpler and less secure than the security model adopted by Grids. ".

Talking about how a software is normally considered secure or not, we could think about the development process, where testing is done through it and bugs are taken out of the code; and compare it to Cloud Computing, which has been less tested than Grid computing, and so it is very probable that in the future more bugs will be discovered, compared to Grid technologies that have already been tested.

To compare Cloud and Grid security, we will use security issues from the previous sections and show what are the main differences in section 5.1 and in section 5.2 we will describe some methods that can be used for both technologies.

5.1 The main differences between Grid and Cloud Computing

- As long as Cloud computing has more homogeneous platforms and Grid computing more heterogeneous ones and dynamic resources, Grid has to consider some issues that Cloud Computing does not have: Single sign-in to access multiple Grid sites, privacy, integrity and segregation should be taken into account so that resources owned by one user cannot be accessed by unauthorized users and/or tampered with during transfer[4].

- Cloud users can use Cloud easily and almost instantly through a credit card, while Grid security is stricter and does not give this feature[4].

- The strict security approach given by Grid computing adds security, helping to prevent unauthorized access[4], while Cloud computing does not.

- Defining a framework which has to be used to program for the Cloud gives an additional posibility to manage security, while the more open system given by Grid does not[6].

- Grid computing, must share user and resource interface to allow providers to connect their resources, while Cloud Computing tries to share only the user interface while the resource interfaces are hidden.

- Since Cloud computing relies on web applications, it has the vulnerabilities of web applications, which Grid does not have.

- In Grid computing all the resources are shared with other users so the machine's security should not be compromised. In Cloud computing, virtualization is normally done directly with the support of processors' virtualization methods, so the resources are accessed in an abstract and more secure way.

- Grid computing needs to have multiple IDs [8] while Cloud computing only needs one.

- In Grid computing, data integrity and confidentiality must be considered.Integrity is not normally mandatory in Cloud computing,since your information is already inside a virtualized environment, but confidentiality is necessary to prevent man-in-the-middle attacks.

5.2 Methods that can be used for both technologies

Some methods that nowadays are used by one of the technologies could be used in the future to make the other more secure. This is a very interesting issue due to the lack of testing that has been already done with Cloud Computing due to its youngness. Some possibilities will be given in this section:

- As previously mentioned, the Trusted Computing Group(TGC)[9] proposed the Trusted platform module(TPM), which is a chip with Private key and cryptographic algorithms, which are not rewritable. This method has been proposed to Cloud Computing but, since the same chip could be used in different domains of the Grid, using public key cryptography to secure the communication.

- Another thing that could be considered is the trust relationship management methods used in Grid computing. Nowadays, since the APIs used by different Cloud providers are not common and since is not very feasible to combine multiple Cloud providers at the same time because of the lack of standard APIs, the trust relationship management is not necessary but if we think about the future, where Cloud providers agree to use common APIs and the possibility of sharing data and combining Clouds become possible, trust relationship management methods used in Grid could be useful for Cloud Computing as well.

6 Conclusions

Lots of technologies are nowadays on the market and some of them share their features. In this case, as it has been demonstrated, already existing security solutions can be used for other technologies. In addition, in some specific issues, new technologies can be more secure than older ones due to that the design of new solutions can be more suitable to avoid security problems and will make this task much easier.

Relating to this matter, in this paper we demonstrated that technologies like the Trusted platform module(TPM) proposed by the Trusted Computing Group(TGC) [9] can be used for Cloud computing as proposed but it can be used in different domains of Grid computing as well in order to secure communication through public key cryptography.

On the other hand, looking into the future, we considered using trust relationship management methods, like the ones used in Grid computing, to be used in Cloud Computing when building common APIs that allow the of sharing data between Cloud providers. Trust must, therefore, be considered between the Cloud providers.

[4, 1, 6, 8, 2, 9, 11, 3, 5, 7, 10].

References

[1] Armbrust, Michael; Fox, Armando; Griffith, Rean; Joseph, Anthony D.; Katz, Randy H.; Konwinski, Andrew; Lee, Gunho; Patterson, David A.; Rabkin, Ariel; Stoica, Ion; Zaharia, Matei. *Above the Clouds: A Berkeley View of Cloud Computing*. February 2009. `http://www.ibm.com/developerworks/web/library/wa-cloudgrid/?S_TACT=105AGX01&S_CMP=HP`.

[2] ArmMichael Halton. *Security Issues and Solutions in Cloud Computing*. June 2010. `http://wolfhalton.info/2010/06/25/security-issues-and-solutions-in-cloud-computing/`.

[3] David Linthicum. *Cloud Computing, Deep Dive.* 2009. MPI-SWS `http://www.scribd.com/doc/26495704/Cloud-Computing-Deep-Dive-Report`.

[4] Foster, I.; Yong Zhao; Raicu, I.; Lu, S. *Cloud Computing and Grid Computing 360-Degree Compared*. Grid Computing Environments Workshop, 2008. GCE '08 , vol., no., pp.1-10, 12-16 Nov. 2008. `http://people.cs.uchicago.edu/~iraicu/publications/2008_GCE08_Clouds_Grids.pdf`.

[5] Ian Foster. *What is the Grid? A Three Point Checklist.* July 2002. Argonne National Laboratory & University of Chicago `http://dlib.cs.odu.edu/WhatIsTheGrid.pdf`.

[6] Judith M. Myerson . *Cloud computing versus grid computing*. March 2009. EECS Department. University of California, Berkeley.Technical Report No. UCB/EECS-2009-28. `http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html`.

[7] Marc-Elian Begin. *Comparative Study: Grids and Clouds, Evolution or revolution.* May 2005. CERN `https://edms.cern.ch/file/925013/4/EGEE-Grid-Cloud-v1_2.pdf`.

[8] Marty Humphrey,Mary R. Thompson. *Security Implications of Typical Grid Computing Usage Scenarios*. February 2009. Computer Science Department,University of Virginia. Distributed Security Research Group,Lawrence Berkeley National Laboratory. `http://dx.doi.org/10.1023/A:1015621120332`.

[9] Nuno Santos,Krishna P., Gummadi Rodrigo Rodrigue. *Towards Trusted Cloud Computing.* May 2009. MPI-SWS `http://www.usenix.org/events/`

hotcloud09/tech/full_papers/santos.pdf.

[10] Trusted Computing Group. TCG https://www.trustedcomputinggroup.org.

[11] Y. Zhao, M. Wilde, I. Foster, J. Voeckler, J. Dobson, E. Gilbert, T. Jordan, E. Quigg. . *Virtual Data Grid Middleware Services for Data-Intensive Science*. 2005. http://dx.doi.org/10.1002/cpe.968.