Thoughts on Grid Security standards and practices in the Cloud Era

David L. Groep (OGF security area, Nikhef, ed.)
with contributions from Jens Jensen (RAL/STFC) and Ralph Niederberger (FZ Jülich).

Security in Clouds is likely the one aspect that needs to be firmly grounded. There are many different angles and issues to security and privacy that we can think of, and all of them affect the way you may interact with a cloud. Some of the issues will depend on the kind of cloud you're taking to, since issues that affect machine-image-clouds may not be present in application-hosting clouds, and vice versa. In quite a few cases, though, issues around securing access to services (and thus access to clouds) have already been dealt with in our recommendations and community practices, whereas others are subject of current research- and working groups, or are today in the realm of 'best practice' in some of large grid infrastructures. It covers the issues of access to cloud services and their use, data security and privacy, network and firewall issues, incident handling and policy coordination.

To the customer, the cloud is as remote as any type of grid: at least conceptually, the user of a grid resource or cloud is not supposed to know where a job, workload or file is sent, which means that building end-to-end trust and securing communications are obvious characteristics regardless of whom you're talking to. With this observation it becomes clear that quite a few aspects we dealt with previously in the Open Grid Forum and in current production grids are applicable to clouds: distributed logging, bookkeeping and accounting for customers to keep track of what was done where and in which cloud; securing data on cloud storage against prying eyes; and the coordination of end-to-end trust – not only in the technical sense, but especially the policies and practices that surround a true trust fabric. Production-oriented grid infrastructures today have also realised that you need simple, coordinated policies such that the user having agreed to one acceptable use policy can be reasonably sure it will comply with the policy of the other infrastructures as well – essential if you want to transparently access resources from multiple providers (one could for example image combining application-hosting and virtual machine clouds in a transparent way). Establishing the interoperation best practices and standards in OGF has been quintessential to get current infrastructures to work as they do today.

It is revealing to see that in a few cases even issues we faced in very complex multi-domain grids re-appear in unexpected places when you want to use clouds to process sensitive information, where you may want to have assurance and assertions by trusted third parties about a specific cloud. In this paper we want to take you on a tour of considerations on cloud security from a grid perspective, and highlight aspects of the OGF suite of documents and activities.

Lets now look at a few of the issues: how to gain access to a cloud and use it; securing your data whilst it is in there; interconnecting a cloud to your own networks; and how to coordinate of trust, policies, and how to handle security incident response.

Access to cloud services and the usage of clouds

This is conceptually one of the easier things to do, and does not differ significantly from access to other kinds of resources. For those clouds that have a programmatic interface to manipulate it, the work on security communications is just as relevant, and GFD.132 ("Secure Communications 2.0") also addresses the mechanisms for slightly more complex scenarios involving intermediaries. Username-password access is (still) common, and OGSA SC2.0 supports policies like username/password-digest, besides other, more secure mechanisms.

In machine-images clouds, you can essentially build your own private cluster or single-domain grid, thus making the authorisation and trust issues easier than in the multi-domain grid case. Within it, you will not have to deal with multiple authorities and cross-domain trust, and you can – provided your favourite cloud provider does firewalling around your set of machines, and you trust this firewall – even create the illusion of a local area network. Of course for any service you provide yourself based on the resources allocated to you in the cloud, all conventional issues (and existing documents) will again apply.

However, don't discard the multi-domain issues yet – they will reappear later if you want to use resources in the cloud for sensitive or otherwise security- and privacy-classified work.

Also trust and mutual authentication, above what is offers by simple 'let's trust DNS naming, that's enough'-style PKI certificate authorities that offer you SSL certificates for a few Euros remains important here. The 'cheap' certificates will merely assert that you're talking to the domain name you entered, but provide no hint about who is running this domain, or whether the name is just a variation of a well-known legitimate one. It is equally relevant to both grid and clouds. Are you sending your (private) application and data to the 'right' cloud provider? Or are you the victim of a phishing attack? Even the Extended Validation certificates that are today promoted for internet banking are already a lot better, but will still only assert organisation name, and nothing more. Are you prepared to send your applications or machine images there? The International Grid Trust Federation (IGTF) and the CAOPS Working Group have put down a model for dealing with mutual trust through 'authentication profiles', that define common minimum baselines, based on policies jointly set by customers, relying parties and the 'trusted' third parties. But since this topic is akin to the data security issues will discuss below, lets defer the discussion on this topic just a bit ...

Especially services that use re-usable authentication, like cookies not linked to the action you're about to execute, are vulnerable to cross-site request forgery attacks: you can get implicitly redirected to 'another' site without the user ever getting even a conscious choice and perform the action the attacker want you to take. And since the user/customer is already 'logged in' to the remote site, or has the appropriate cookies, the malicious actions is just carried out, no questions asked. Browser-based management systems are obviously most vulnerable, but are not the only possible attack vector.

Data security and privacy

Storage in the cloud, when you are doing your own computing in a trusted environment, is relatively easy to protect from prying eyes. Not much standards work as such has been done in this area, but grid middleware exists in several large deployment projects that provides strong data encryption whilst making sure that no single resource provider has enough information to crack your data. Systems like Hydra (from the gLite middleware stack of EGEE) are integrated with data management solutions that support our SRM standards, and it uses a secret sharing scheme (SSSS) underneath so that no single site has enough information to gain knowledge of the encryption key. Access to the key servers is then protected with traditional GSI and PKI technology. There is no associated standard yet, but it is used in production infrastructure for storing medical patient data. However, this is the easy bit.

Protecting your data in a compute environment is highly non-trivial, and always requires a large amount of trust in whoever is operating your infrastructure. During processing, your data must be decrypted at some point, accessed in clear-text, and then re-encrypted. There is no way to protect this data during processing. Partial solutions have been thought of before (the long-extinct Trusted Computing RG tried part of it, and research groups keep coming up with other bits and pieces), but the fundamental problem remains: you have to trust the resource provider.

Of course one might try and split the work amongst many providers (that are hopefully independent), such that no-one can collect enough information together to make use of it and thus infer private information. Frankly, both splitting the work and trying to be inference-proof are un-doable propositions for any sort of realistic problem.

Another option would be administrative controls, for example by using (micro-)contract with providers where you negotiate the level of privacy before allocating the resources, and thus – by policy – trust them to honour the agreed (micro-)contract. Such a contract might be “process this data, but securely wipe any scratch areas used, as well as any machine images submitted”. Or, “ensure that my resources are never shipped out of the country I designate, even if this means terminating my application or machine images”. The latter agreements will be of course relevant when processing sensitive personal data or even medical patient data, preventing the shipping of personal data to countries that don’t provide a safe harbour for such data.

Trusting cloud and providers for specific purposes is also a place where, suddenly, some interesting ‘multi-domain’ issues re-surface. For example, thing of the processing of sensitive data in a cloud. Processing of personal data is normally subject to specific protections, and it cannot legally be handled by organisations that are not subject such restrictions. So, if a cloud resource provider wants to provide service to customers that handle personal data, it should comply. But, how does the customer know a provider meets the criteria of a safe harbour? It may needs assertions by a trusted third party about such compliance, in order to fulfil its own policies and for regulatory compliance. The same holds for medical and patient data, etc. And there are all different trusted third parties for various aspects of the

service. Combining assertions from different sources and making informed decisions based on them is essentially what multi-domains grids have been working on for those past years.

In short: who is authoritative for which assertions about a cloud? How many assertions from how many different sources do I need before I can use a cloud within my policy and regulatory constraints?

Of course, solutions will always be partial, but the ground work on standards and best practices has been laid. A tandem group in OGF like the CAOPS-WG/IGTF (document templates in the working group and its policy instantiation in the IGTF) has put together a framework for agreeing on policies and international trust based on 'minimum baseline' assurance levels coordinated by both providers and customers of the trust infrastructure; a model that now gets copied into other non-grid infrastructures as well.

Security is a domain where technical standards have to complemented by community best practice in trust building and policy, like we see in the authentication trust fabric.

Network and firewall issues

Integrating cloud infrastructures (be they machine-image and application/service types) into the local infrastructure also throws up the firewall issue very prominently. If you have 'outsourced' part of your capacity in a cloud, of course your systems in the cloud have a special relationship with your 'local' infrastructure. In order to do their work, the systems and apps in the cloud will likely need access to parts of your local infrastructure – say, access to data bases, storage of results, backup, and the like.

Will 'your' part of the cloud have access to your local network? Unlikely, unless you start punching holes in your local firewall. But then, will the IP addresses used from within the cloud be uniquely associated with 'your' application or machine images? Can you guarantee that? The very same questions that have plagues multi-sites and multi-domains grids over the years. The Firewall Interest RG gap analysis and 'current solutions' overview present exactly where the problems are, and why current solutions are too labour intensive and lack the management capabilities to be usable in a production data-centre environment.

When integrating clouds with local infrastructure, or when trying to work with resources from two different clouds as one system, these issues surface like they do in all other grids. The interfaces that are currently being standardised in the Firewall Virtualisation WG (INFRA), with the proof-of-concept implementation, are quite relevant in the cloud space as well.

Policy and incident handling

Although usually outside the scope of a standard organisation, harmonisation of incident handling procedures is critical and ought to be part of the security strategy. Running in the cloud makes you and the cloud provider jointly responsible for acting on security incidents. So, does your cloud provider inform you of incidents on its infrastructure that affected your machine images or application? Do you know if systems are kept up to date according to your policy (or regulatory requirements)? If you are

running in a machine-image cloud, can you yourself escape through the hypervisor, i.e. if your systems get exploited, do you pose a risk to the cloud as a whole? Does the provider have an active introduction detection to check for such intra-cloud attacks? How can you coordinate with your provider in case of an incident? Do you have a trust relationship, such as exists within the CSIRT community or FIRST, and like the one that has been forged in the (scientific) grid infrastructures?

Conversely: how can a cloud provider know that you are impacted by a particular vulnerability? Should the cloud warn you beforehand? And, is the providers incident response policy compatible with your business model, or will you get shut out pre-emptively and be left out in the cold while the provider figures out whether or not you are the source of an incident? Liability clauses in contracts typically tend to protect the provider unless agreed otherwise – does that match your business policy?

Open security-related questions with no simple answers, as there is no 'right' and 'wrong' in these cases.

Policy issues, including acceptable use and auditing, are usually left open or implicitly agreed to, but major grid infrastructures have realised that common policies ease interworking significantly. It does not matter too much if you stay with a single provider, but if you use two providers, it is quite inconvenient to have different AUPs apply to both. Acceptable behaviour in one may not be acceptable in the other, and resource allocation is complicated by having to take this into account. Interoperable science grids, such as OSG, EGEE, DEISA, NAREGI and others have (almost) identical base policies that are then 'instantiated' for a particular management body. It makes for smooth interworking, and a users' consent given to one grid can be accepted as sufficient by the grid peers. When common policy agreements break and individual providers withdraw from the common policy, we now know (unfortunately by painful experience) that this causes havoc in operations. Time agreeing on common policies, even if you find it agonizingly slow, still is time well spent.