A case for Shibboleth and grid security: are we paranoid about identity?

A paper for the UK e-Science All Hands Meeting, September 2006

Mark Norman,
University of Oxford

1. Abstract

The findings in this paper represent some of the output of the ESP-GRID project following the consultation of current grid users regarding the future nature of grid computing. The project found that there was a clear purpose for Shibboleth in a future grid and that, for the majority of users, this would be secure and improve their experience of grid computing. Client-based PKI remains suitable and desirable for Power Users and we must be careful of the means by which we mix these two access management technologies. PKI is currently used to define grid identities but these are problematically conflated with authorisation. The grid community should work harder to separate identity/authentication and authorisation. This paper also questions whether we need identity to be asserted throughout grid transactions in every use case. Currently, this is a solution to a security requirement: it should not be a requirement in itself. We propose that the grid community should examine methods for suspension of a rogue user's activities, even without identity being explicitly stated to all parties. The project introduced the concept of a Customer-Service Provider model of grid use and has produced demonstrators at the University of Glasgow.

2. Introduction

2.1. The ESP-GRID project

This paper represents some of the output of the Evaluation of Shibboleth and PKI for Grids (ESP-GRID) project (URL in References). The project also has thoughts and findings on the types of users who may populate a future grid and on the idea of a Customer-Service Provider model of grid use. These are found in a separate All Hands paper (Norman, 2006).

The ESP-GRID project has evaluated the access management requirements of grids both from the existing literature and the projected future set of users. It has also investigated the technologies available for policy management and looked at the concept of virtual organisations. Much of the technical output of the project has been in the form of demonstrators developed at the National e-Science Centre Hub at the University of Glasgow,UK (URL in References).

2.2. Grid security: what are we trying to secure?

Grid computing tends to be thought of as displaying a quite different threat model to that of other network environments (e.g. the world wide web). With grid computing the concept is that the user has some degree of control of the remote grid machine that she is accessing: instead of – for example – merely returning a document, she is able to take up the processor of the machine for an extended amount of time and she is usually able to modify the environment on that machine as well. A rogue user in such a situation clearly could pose a far greater threat than in more traditional 'Internet' situations. Alternatively, we have argued (Norman, 2006) that in the near future – if grid computing is truly successful – most users may access grid services in a very controlled manner that has many similarities with the world wide web. Such users are not likely to be able to modify the environment on the grid machine and may be limited to very predictable actions.

Therefore, most activities on a grid may pose a much lower threat to grid machines than the activities that dominate today. For the sake of this paper, let us assume that we have a mixed economy of users: some exerting relatively deep control over distant grid machines and many users with little scope or interest in modifying the computing environment or how the jobs run on the grid.

2.3. Sections of this paper

Within this paper, we examine identity management and who is best to take on this task. This is followed by an examination of the perceived requirement for constant identity assertion throughout the grid and the 'case for Shibboleth'.

3. On the grid is it appropriate to devolve identity management?

Traditionally, user 'identities' have been managed in the higher education community on a per-institution (organisation) basis. There has been little drive to be very rigorous about checking real-world identity accurately when issuing identity credentials at such organisations for the first time, although it is likely that these procedures have been better than many believe. Those working in a stricter (usually PKI) culture may consider these procedures to be inferior. However, there are strengths and weaknesses to both the (usual) PKI approach and to the per-institution approaches.

3.1. In the UK, we are already trusting the old ID-establishment processes

At present, an applicant for a digital certificate only needs to present *some form* of photo ID (undefined in the UK e-Science Grid CA - Certificate Policy and Certification Practices Statement). Usually this is taken as a person's university card. This means that we are trusting the procedures for issuing the university card in the first place. It follows that the original choice for choosing client-based PKI for grid security is somewhat flawed, as the strongest part – the greatest benefit – of PKI: the establishment of a long-term, highly trustworthy, ID is compromised. This is an argument for another place, however.

A further difficulty with mixing old university procedures and newer, very centralised, PKI procedures can be summed up in this scenario:

> *Post-grad A. Newman begins work at Cotswolds University. He finds he needs a digital certificate for some of his grid-based research. He talks to his local registration personnel about this who know nothing of the "grid" and then finds he has to travel to his local Registration Authority at Oxford University, after applying on-line. He travels to Oxford and presents his 'Cotswolds Card'and the RA grants his certificate request.*[1] *A little later, it turns out that Newman is a thief and a fraudster and Cotswolds University revokes all of his university accounts, swipe cards etc. etc. Unfortunately, the good registration folks at Cotswolds don't have anything to do with e-Science (they haven't been on the RA training course) and therefore Newman is allowed to keep his digital certificate for the rest of the year.*

It is clearly better that the registration or personnel people closest to the user should look after the identity of that user. PKI is usually over-centralised and managed at a very remote, often national, level, as in the UK. This is highly problematic. This case

[1] Assume that the "Cotswolds Card" is his university ID. However, a lovely twist to this story would be that he (theoretically) could have used a "Cotswolds Card" that was issued by his local swimming pool (with inadequate ID checks), but that still contained his photograph.

has been made at greater length elsewhere (Norman, 2005).

3.2. Where PKI should work in managing identities

We should address the concepts of 'identity' and 'identity provision' and the management of identity. In a perfect on-line world, identity management would be completely separated from authorisation. However, at present, this is rarely the case. Grids using client digital certificates, for example, tend to have the *Organisation*, to which the person belongs, included on the certificate. The certificates are issued typically for a year and users are able to obtain a certificate only if they are a member of a particular research or grid community. All of these factors are attributes associated with authorisation decisions. If such authorisation decisions were handled quite separately from the identity token (e.g. digital certificate) then users would be able to keep the token for life. It would not need to be managed, except for the instances where it was issued mistakenly or wrongly or if it had been 'stolen' by another entity. The person will still be the same entity in ten years' time, even if she had undergone a sex change, been convicted of defrauding other grid users etc. etc. Her identity would not have changed, but her authorisation attributes certainly would!

3.3. Identity and attribute management

Currently, it is easier to combine identity, authentication and authorisation to some degree. Identity tokens (accounts, user names, digital certificates etc.) are issued by organisations such as education establishments and it is these same organisations that help the resource providers (e.g. grid nodes) to make authorisation decisions about users. This need not be the case, but if this 'identity problem' were to be solved then the problem would just transform to a problem of managing authorisation-associated attributes.

4. The grid requirement for identity

4.1. Emotional security

When discussing and planning security mechanisms it is always surprising how often one's emotions can cloud the issues. We tend to assume that a system is more secure if the users and other entities therein are always explicitly and fully identified (i.e. there are logs of identities associated with most actions). This is only true if those identities may be checked accurately, the data is current and the authorisation is similarly accurate. Without those caveats, explicit identities can give a thoroughly false sense of security.

Emotionally, we always want to know "who" the user is, in case they do something wrong. Actually, as the "who" is really quite difficult to check and the authorisation credentials even more difficult, it should be the "can I trace this user easily if he does something wrong" that should be far more important, as should the concept of, "actually, I don't mind who this is right now, just as long as I'm fairly sure that they are authorised". But those don't give us a warm feeling of security. They are, nevertheless, far more secure than relying on poorly maintained identity (mixed with authorisation) information.

Bruce Schneier writes about the great insecurity of relying too much on ID (Schneier 2004a) and also gives examples of where this can lead to surprisingly (and possibly unexpected) reduced levels of security (Schneier 2004b). These examples include airline traveller programs whereby travellers can register beforehand, go through an identity check and thereafter reduce the chance of having their baggage searched at airports: clearly a first-time terrorist gains an advantage by such a situation. Schneier cites excellent examples of terrible security which makes people feel

better and points out how good security may seem counter-intuitive until looked at in depth. With regard to the use of ID, he rightly suggests that if you make something easier (i.e. lower security) if ID is used, then the bad guys will just get ID. And the rest of us are left not paying enough attention to security because the ID has given us a false sense of security.

4.2. Do we need identity throughout, for every service?

Currently, grids' supposed requirement for 'up front' identity assertion throughout may be exaggerated. Some services certainly need to know the identities of users. However, many do not: the hard requirement for identity has probably come about as it is a *solution* to the requirement to suspend the activity of wrong-doers or for when certain users' credentials have been stolen.

4.3. Rapid suspension and slower identification/revocation?

Explicit identity may be useful at times, but it is clearly secondary in importance to a guaranteed method of quickly detecting wrong-doers and of removing their privileges. This may be achieved with or without knowing identity 'up front' or by logging permanent identities. Therefore, the main requirement should be the detection of misuse or security breaches and the quick tracing of the identity of the user, rather than constant logging of identity.

The real requirements are probably for:

- good authorisation procedures;
- quick detection of wrong-doers;
- rapid suspension of rights, possibly throughout the grid;
- (in most cases) the rapid revocation or suspension of ongoing jobs throughout the grid;
- an investigation into the activities of the individual.

These requirements are expanded upon and tied to the Customer-Service Provider model (see 5.1 below).

5. The case for Shibboleth

5.1. The C-SP model will dominate

As outlined elsewhere (Norman, 2006), it is very likely – on many mature production grids – that the majority of users will benefit from the power of grid computing through an application-interface on a server: for example, via a web portal. We have called this the *Customer-Service Provider model* (C-SP model). With such a restrictive point in terms of the possible range of actions that a user can undertake, the use of Shibboleth to enable authentication and authorisation is highly appropriate. The use of the grid via the C-SP model is summarised in Figure 1. The abbreviations SEU (Service End User), IdP (Identity Provider), and SP (Service Provider) are described at greater length in Norman (2006). The SP uses a set of host certificates to interact with the grid. The grid machines could cause the revocation of one or more of these host certificates if an attack were suspected and/or the SP and IdP could be made to suspend a user's activities automatically in such a case. This could then, typically, be followed by human attentions within the IdP and SP to identify the user and investigate which actions should be taken.

It is a widely-held principle that the organisations interacting most frequently with the end user are the most appropriate to manage their identities and/or their most common authorisation attributes. Conversely, exceptions to this exist in two main areas:

- If it were possible to truly separate authentication from authorisation on the grid, there is little reason

why long term identity tokens could not be issued. This would then mean that authentication could take place in a variety of places.

- Authorisation attributes may also be held with virtual organisations and a secondary query may be necessary.

Nevertheless, in the first exception cited above, it may still be most convenient for SEUs to be authenticated at their home organisation (IdP) for single sign-on reasons. Similarly, it may be convenient for the virtual organisation to allow authentication at the IdP or the SP before releasing the attribute information.

If we put the above two exceptions aside, then Shibboleth (http://shibboleth.internet2.edu/) is a good fit for devolving authentication and much of the management of authorisation attributes. Shibboleth would provide a useful single sign-on (like) experience for the user: he would only need to authenticate at his home organisation. This would benefit him in terms of having to learn only one sign-on interface, and would place the task of managing identities and attributes with the most appropriate organisation.

Shibboleth may not be appropriate if identities are established long-term, although the authentication of these identities may sit well with the home organisation, nevertheless. It is also likely that a forthcoming release of the Shibboleth software will be extended to accept authentication within one organisation and the retrieval of attributes from another (virtual) organisation.

5.2. Demonstration of the C-SP model with Shibboleth

The BRIDGES, DYVOSE, VOTES (URLs in References) and ESP-GRID projects have produced a Shibboleth-enabled portal with which to authenticate and authorise people to access a variety of applications. This activity proves that Shibboleth and the grid can interoperate, but it avoids the issues of supporting Power Users. These issues may be unimportant unless the number of Power Users grows greatly. The need for something very easy to use for good uptake by researchers was discovered early on in the BRIDGES project, in particular, by the developers at Glasgow, both in terms of access management (and the need to avoid client digital certificates) and in a clean, easy, "Google-like" interface (Sinnott, 2006).

5.3. Most users are not Power Users

Missing from Figure 1 are the other types of grid user (described and discussed in Norman, 2006). These include the most common type of user that exists today. Such users, who typically work at the command line, write and/or compile code and often wish to modify the environment at a remote grid node, we have termed 'Power Users' in the ESP-GRID project. Power Users are likely to be able to tolerate the difficulties of working with PKI and any security advantage derived from the use of PKI is of benefit to the grid as such users pose a greater threat to an individual grid node.

5.4. Can Power Users benefit from Shibboleth?

There are several initiatives under way that are attempting, in different (and similar) ways to bring together the security of PKI and the ease of use of Shibboleth (GridShib, SHEBANGS, ShibGrid, MAMS, SWITCHaai, among others).

Some approaches need a mapping between an individual's Distinguished Name (DN) on his digital certificate and an 'attribute' that the user's home enterprise directory (or Attribute Authority – AA – in Shibboleth terms) can manage and supply, when requested. This would allow the identity to be the same, however the authentication were performed (e.g. via username/password and Shibboleth or via presentation of a digital certificate). Other

approaches (e.g. GridShib) mandate the use of a certificate for authentication but then use a Shibboleth AA for authorisation purposes. Some of these projects are also examining using Shibboleth and institutional single sign-on mechanisms to release digital certificates for use on the grid, but see *6.3-Mixing trust models*, below. There is much effort being applied to the Shibboleth-enabling of MyProxy servers which may prove very fruitful.

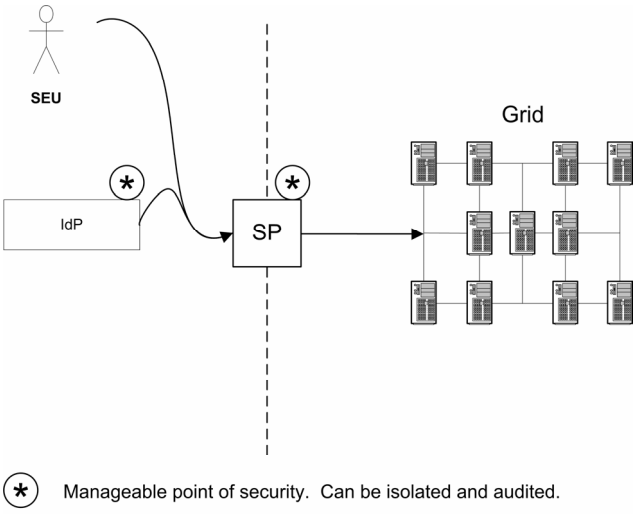

Figure 1 The C-SP model of access to the grid: the SEU is authenticated by the IdP (trusted by the SP) and the SP accesses the grid via a host certificate.

6. The cases against Shibboleth

6.1. Power Users

Power Users are probably an inappropriate group to benefit from the Shibboleth model of accessing the grid. If we accept that the use of PKI is beneficial to the grid as a whole, then it is this set of users who should be using client digital certificates, as today. There may be some isolated benefits to these users through using Shibboleth, such as 'away from home' access to pre-prepared proxy certificates or access to basic level assurance certificates for some tasks. If, however, our prediction proves to be incorrect and grid use does not grow via the C-SP model, and Power Users remain the dominant group, Shibboleth may be of limited benefit.

6.2. Delegation

An extension (RFC 3820) to the original standard (from RFC 2459 profiling X.509 version 3 certificates) allows for delegation via proxy certificates. This work has arisen largely due to the use of PKI on grids and Globus-based grids in particular. There are some 'philosophical' difficulties with such an approach; notably where the (proxy) private key does not remain in the sole control of the original user. However, this activity has proved a way forward for delegation on grids and is clearly a mechanism for *constraining* delegation. Shibboleth in itself does not provide a mechanism for delegation. (Shibboleth is based upon machine-to-machine trust and is, to some extent, incompatible with this concept, but see next section for an analysis of trust). Some would say that the use of proxy certificates gives rise to the situation of machine-to-machine trust and therefore, the need for this kind of delegation may be inappropriate, but this is an argument outside the scope of this paper.

6.3. Mixing trust models

The routes of trust for Shibboleth and for client-certificate PKI are a little different. In PKI, a user has a certificate and invokes it directly when interacting with a grid machine. If we ignore the checking of signatures and CAs, the trust is from human to machine (the human with the certificate does not need to rely upon another entity or machine for her certificate to be believed and trusted for authentication). This PKI represents 'human to machine trust' for the authentication/identification step.

Shibboleth requires the user to log in at his home organisation's identity provider (usually a web single sign-on interface). The assertion that "this user has been authenticated" therefore comes from a machine. Therefore Shibboleth represents 'machine-to-machine trust' for the authentication/identification step.

We need to take extra care how we combine these two methods. A mixing of the two trust models is problematic and – at the very least – brings down the overall security (or assurance) level to that of the least secure component. If we were to use Shibboleth to make the user experience with PKI less onerous, we are certainly reducing the assurance level of the assertion. (However, if the Shibboleth home organisation identification and authentication procedures were very robust for that user, it may not reduce the assurance level that much). This issue is in addition to the security challenge posed by the complexity of mixing two access management 'systems': the resulting system may be very complex and the more complex something becomes, the more likely it is to develop security problems.

It could be argued that, if we wish to use Shibboleth, then we should avoid the use of client-based PKI completely: the user could employ Shibboleth to mediate authentication and authorisation and then a 'gateway' machine could be trusted by the PKI-based grid. This is, effectively, the C-SP model.

7. Summary

Some of the outcomes of the ESP-GRID project include that PKI is used at the moment to manage identities, but that these identities are problematically conflated with authorisation. In the UK and elsewhere, our current implementation of client-based PKI is very good at establishing identities, but is very poor at managing authorisation. The grid community should work harder to separate these two.

We should also question whether we need identity to be asserted throughout grid transactions in every use case. Identity being asserted and logged 'up front' before every transaction gives us a *feeling* of security. The real need is for rapid suspension of the rogue user-initiated activity and the later revocation of credentials and/or rights: people have confused this requirement with the current *solution* of identity assertion throughout the grid.

Shibboleth is a great opportunity to allow the appropriate people to manage identities and authorisation-enabling attributes. It is certainly worth pursuing in the grid world: it could have the benefit of increasing the level of security on the grid as well as the ease of use for non-computer technical users.

The ESP-GRID project postulates that, in order for the grid to scale, some sort of Customer-Service Provider arrangement is necessary to enable the new users who are not expert computer scientists. This C-SP model lends itself to Shibboleth very well but, equally, the authentication point could be at the service provider portal instead.

The project has worked with the National e-Science Centre at Glasgow University to produce some demonstrators which are good examples of both the use of Shibboleth and grid and of the C-SP model.

Mixing Shibboleth and client-side PKI for grid users is difficult and potentially insecure, although there will be cases where

it is useful and appropriate. Indications from the ESP-GRID project are that client-based PKI is appropriate for grid Power Users (the current majority of grid users), but that Shibboleth, combined with their local institutions' single sign-on technologies would benefit the vast majority of the future End Users.

8. References

BRIDGES (Biomedical Research Informatics Delivered by Grid Enabled Services) project web site http://www.brc.dcs.gla.ac.uk/projects/bridges/

Certificate Policy and Certification Practices Statement for the UK e-Science Grid CA http://www.grid-support.ac.uk/ca/

DYVOSE (Dynamic Virtual Organisations in e-Science Education) project web site http://labserv.nesc.gla.ac.uk/projects/dyvose/

ESP-GRID (Evaluation of Shibboleth and PKI for Grids) project web site http://www.oesc.ox.ac.uk/activities/projects/eprojects/esp-grid/index.xml

GridShib project web site http://gridshib.globus.org/

MAMS (Meta Access Management System) project web site https://mams.melcoe.mq.edu.au/

Norman, M.D.P. (2005) The case for devolved authentication: over-centralised security doesn't work. JISC Core Middleware: developments within Security and Access Management, 20 October 2005. http://www.dcoce.ox.ac.uk/docs/JiscNeSCMiddwareBriefingOct05.pdf.

Norman, M.D.P. (2006) Types of grid users and the Customer-Service Provider relationship: a future picture of grid use. Proceedings of the 2006 UK e-Science All Hands Meeting

Schneier, B. (2004a) San Francisco Chronicle, February 3, 2004 http://www.schneier.com/essay-008.html.

Schneier, B. (2004b) Boston Globe August 24, 2004 http://www.schneier.com/essay-051.html.

SHEBANGS (Shibboleth Enabled Bridge to Access the National Grid Service) project web site http://www.sve.man.ac.uk/Research/AtoZ/SHEBANGS

Sinnott, R (2006) Development of Usable Grid Services for the Biomedical Community. Proceedings of *Designing for e-Science: Interrogating new scientific practice for usability, in the lab and beyond*' workshop at the UK National e-Science Centre, January 25-26, 2006.

SWITCHaai web site http://www.switch.ch/aai/

VOTES (Virtual Organisations for Trials and Epidemiological Studies) project web site http://labserv.nesc.gla.ac.uk/projects/votes/

9. Acknowledgements

The ESP-GRID project is funded from the Joint Information Systems Committee (JISC) and the authors are grateful for the support of the Core Middleware: Technology Development Programme.