Security, usability and the new types of grid users

A paper for the 'Designing for e-Science: Interrogating new scientific practice for usability, in the lab and beyond' workshop at NeSC, January 2006

Contents

1. Abstract

Who will be the grid users of tomorrow? Past experience looking at users of the Internet or World Wide Web tells us that from a dedicated core of highly technically gifted (in IT) individuals, other interested parties begin to take over, at least in terms of absolute numbers. Current trends also indicate that, in the not so distant future, the majority of grid users will be researchers in a variety of fields. These are likely to vary from scientists actively interested in computation to researchers needing computer power or distributed data sets, but who are disinterested in computing *per se*. We propose a categorisation of 'future grid' users into the following categories: Service End-User, Power User (with three distinct sub-types), Service Provider and Infrastructure Sysadmin. A further basic type could be argued as Third Party Beneficiary. This paper outlines the possible characteristics of these 'types' of users. We discuss briefly the levels of security, trust and responsibility that are associated with each type of user outlined above. For users that have layers of applications or, for example, a portal between them and the grid resource, it is almost certain that heavyweight security solutions, as we have with client digital certificates, are too onerous and unnecessary. It is likely that some users will, however, need client digital certificates, due to the level of control that they may exert on individual grid resources. We also outline a Customer-Service model of grid use. The Service End-Users (SEUs) may become the most numerous and most demanding users. It is therefore imperative that we consider their possible profile, even though these users may not yet exist in large numbers. This paper explores this briefly and examines whether access management mediated via Shibboleth would be more appropriate for these users. It may be that authentication and authorisation for the SEU 'customers' should be the responsibility of the Service Providers (SPs). This would hint at a more legal framework for delegating authority to enable grid use, but one which could be more secure and easier to administer. Such a model could also simplify the challenges of accounting on grids, leaving much of this onerous task to the Service Providers.

2. Introduction

The Pew Internet & American Life Project produces reports that explore the impact of the Internet on families, communities, work and home, daily life, education, health care, and civic and political life in the USA. The Project wrote in 2005 that 'The Web has become the "new normal" in the American way of life' (Pew Internet, 2005). It is used by two thirds of Americans for a variety of purposes from checking email to participating in auctions (Fox, 2005). On any given day in 2004, it is estimated that 70 million American adults did something on the Internet. It was not always thus. Before Netscape's browser, Mosaic, was given away free in 1994, the Internet was the domain of the educated and technically knowledgeable. Even within that educated elite, the use of the Internet was dominated by a few research subject areas, possibly arenas in which the development of computing itself had been highly relevant for many years.

What changed? The introduction of a graphical interface that was easier to use and more intuitive did appear to increase the rate of uptake of home computing, but that rate was still not as high as with other expensive consumer items, such as the television and the VCR in previous decades (see Dutton, 1999, figure 9.1). Why did the use of the Internet and home computing not grow faster? There were clearly a variety of reasons. However, one major factor was that developers and providers failed to understand their users. Dutton points out that ICT designers often hold great misconceptions of their users. In 1993 it was estimated that 47 per cent of US Adults were below the level of literacy required to read and interpret a bus schedule (NCES 1993). Nevertheless, many commercial developers assume that their software is written for consumers with at least a high-school education, knowledge of the English language and a background within the mainstream culture (Dutton, 1999). Up to the current day, and presumably beyond, one of the major factors shaping patterns of Internet use is skill of the user (Di Gennaro & Dutton, forthcoming). Many activities that are exclusive to the skilled elite will almost certainly enter the mainstream just as soon as usable and intelligible interfaces are created for such activities. If commercial software and hardware developers are prone to misunderstanding their users, with the enormous market pressures acting directly upon them, then largely academic-based grid developers are almost certain to make the same mistakes.

Dutton (1999) warns of the ICT industry assuming a high-tech, ICT-centric user, as computer system developers frequently overestimate the technical agility of such users. Those writing applications and interfaces for the research/academic community are working with people who are at least as well educated as themselves and are often of a bafflingly high technical competence in fields not well understood by the systems developers. It is thus difficult to imagine how a systems developer could *not* overestimate the ICT abilities of the researchers. The misconception is surely forgivable, but it needs to be identified as a misconception, nevertheless.

Many interested groups must hope that grid technology must be approaching the metaphoric 'release of the browser' stage some time soon. Whether there will be a surge in take-up, as seen with Internet technologies after 1994, or whether it will be a more steady increase remains to be seen. However, it is the availability and ease of use to the greater community that will make the breakthrough. This paper is focussed mainly upon the educational and research use of grid technology. The engagement of the average citizen with grid technology will take much longer. However, we believe that the experience of take-up of the Internet is relevant to the divide between researchers experienced in programming or scripting and the rest of the research community.

Studies into the take-up of grid technology by researchers in the latter category are difficult to find and are clearly difficult to perform. Most surveys have to obtain data from current users of grids – clearly people who have fully overcome, or are somehow coping with, any usability issues – and therefore collectively present a skewed picture. Users without much grid experience find it difficult to engage with questions involving grid technology, even though these users are very important to the findings (Beckles *et al.* 2004). Even requirements analyses, where the users are clearly identified, may suffer from difficulties in eliciting information from those users. Such individuals may feel either inexperienced in the technology or very junior in role (Gavaghan *et al.*, 2004). In 2004, Beckles *et al*. called for a detailed requirements analysis of the UK academic/scientific community's needs with respect to computational grids. A formal approach such as this is clearly to be encouraged. It is very difficult to question potential grid users when they have either never heard of a grid or are too intimidated by the technology to consider using it, although Beckles (2004a) highlights some approaches to do just that. This approach would be expensive, potentially time-consuming, but – given the scale of the development – surely worth it.

Anecdotal evidence of researchers refusing to engage and benefit from grid technology suggests that when an application-interface is presented that is easy to use, the uptake is strong (e.g. Sinnott, 2006). As the Market for Computational Services Project notes, the inability to use a simple 'service' such as a resource broker in itself leads to a lack of ease of use and little motivation for the end user leading to little or no take-up for real use (Grid Markets 2003).

Researchers in many of the sciences during the 1960s and 70s, who were reliant upon large sets of numbers and statistics, often found it more productive to program their own spreadsheet applications if they could gain access to, or build, a computer. Later, those same researchers were able to use an 'off the shelf' spreadsheet application, and became disinterested in the technology behind the application: they could focus fully on the findings of their science. At that time, presumably, more researchers were 'enabled' to work with arrays of figures and statistics, whereas before they may not have been able to enter the field due to a lack of knowledge, or confidence in, programming. Clearly, the more that the tools (or 'vehicles') for research – and other activities – are developed and improved, the greater the uptake is of the technology. But something else happens as well. Not only do the numbers of users increase, but the proportions of the *types* of users may change drastically. Taking our reference point from the developments of the Internet and the early World Wide Web, we predict that grids will change from the enigmatic domains of highly technical computer experts to areas of greater access to all researchers. This presents a challenge to the development of grid middleware and user interfaces. Before we try to improve the experiences for the grid users of today by building up their new and existing requirements, should we not consider who are the grid users of tomorrow?

Authors have previously noted that the current grid middleware is too intimidating for many users, and have often focussed on the security aspects (e.g. Beckles, 2004b). These aspects are important as they are often the most onerous for the non-computer specialist. In temporary lieu of the work, noted above, to collect requirements from current non-users (Beckles, 2004a), we believe that we should examine the types of users that are emerging within grid computing and consider their generic security profiles as well as their likely access management requirements. This may assist in identifying such users in order to carry out a real world requirements analysis. However, until such an analysis is made, our work is merely a guide to the likely categories of users.

The following sections of this paper present our view of these users of tomorrow. This is a personal view, based partly on experience and partly on predictions arising from the use of the Internet and the Web.

3. Types of grid users

3.1. Categories of users and their relative abundance

Table 1 presents a summary of a projected set of tomorrow's grid users. In most cases, these users currently exist to a certain extent and therefore their relative abundance can be compared between today and the future. There is likely to be an almost perfect inversion in terms of numbers of users in some categories. This is important as much effort is expended in serving the needs or gathering the requirements of the most common type of user at the time the system is being developed. If the most common type of user, at present, is the 'power user' or the system administrator, this can be highly distracting in terms of freeing-up the power of grids to the rightful constituency of users: the true end users. Until this is accepted, a significant barrier to entry for the main beneficiaries of the technology will remain.

Table 1 Grid users of the future

Type of user	Typical characteristic	Main role	Current proportion of grid users in this category	Future proportion of grid users in this category
SEUD	Service End-User (data). Little or no computing expertise.	User of applications served by SPs. Uploads data or runs queries.	Low	High
SEUX	Service End User (executables). Some understanding of code creation.	As SEUD, but runs either executable code or scripts via SPs	Low	Medium
PUA	Power User Agnostic of grid resource node. High degree of computing expertise.	Develops programs and data but does not care where processing takes place.	Medium	Low/medium
PUS	Power User requiring Specific grid resource nodes. High degree of computing expertise.	As PUA but may have more platform etc. dependent expertise and some sysadmin expertise.	High	Low
PUDS	Power user Developing a Service. High degree of computing expertise.	As PUA/PUS but developing expertise like SP.	Low	Low
SP	Service Provider. High degree of computing expertise.	As PUA/PUS but has expertise in authorisation and possibly identity management.	Low/medium	Low
Grid-Sys	Infrastructure sysadmin. High degree of computing expertise.	System administration of grid nodes, possibly with infrastructure delivery and security expertise.	High	Low

Note that there are clearly omissions from Table 1. Two notable actors are the Third Party Beneficiary (TPB) and Resource Owner. A TBP could be a person or organisation who/which does not interact directly with the grid but whose personal data are being handled on the grid. Resource Owners clearly have important functions, but they do not necessarily interact with the grid, unless playing one of the seven main roles shown in Table 1 at a particular moment in time. In designing future grids, the requirements of both of these actors would have to be given much thought and would impact upon the likely architecture and security mechanisms of those grids. However, for the purposes of this paper, the general requirements (or expectations) of only the seven main roles are considered in relation to access management and other security needs.

Throughout this paper, the abbreviation SEU is taken to represent SEUD and SEUX where a statement could apply equally to either category.

3.2. Example illustrations of the seven major actors

Table 2 shows some examples of the activities and needs of these seven major grid actors. Where it is possible to give current real-world examples of each actor's activity, this is displayed in italics. The remaining text comprises purely imaginative illustrative examples.

The majority of today's users come into the PUS and Grid-Sys categories and it would therefore seem irrelevant to pick out real-world examples of these actors. Many future PUA users probably represent a subset of the current PUS users that will become apparent when and where resource brokers are widely used.

Table 2 Example activities for the seven main categories of grid users

User/actor	Examples of activities
SEUD	Little or no computing expertise. User of applications served by SPs. Example 1: *PhD Biologist submitting large data sets for processing to a service. A current example includes an application such as BASIS. BASIS (Biology of Ageing e-Science Integration and Simulation system, http://www.basis.ncl.ac.uk/) is a UK e-Science pilot project which delivers a grid-enabled system that serves the biology of ageing research community by helping to integrate data and hypotheses from diverse biological sources. From the user's point of view the service is presented through a web portal.* Example 2: User or organisation receiving regular output (without necessarily sending input) e.g. the BBC or Meteorological Office receiving bulletins from a 'Weather' SP. Example 3: *Social scientist submitting various problems or scenarios to a social modelling and simulation service (possibly with full graphical interface as suggested by the MoSeS project (Modelling and Simulation for e-Social Science - http://www.ncess.ac.uk/research/nodes/index.shtml#moses).*
SEUX	User of applications served by SPs, but for executable code or scripts. User may have the means to produce executable code or scripts, but with little computing expertise. Example: Scientist wishing to run Matlab code in a grid environment.

PUA	Power User Agnostic of grid resource node. High degree of computing expertise. Technical expert programmer supporting end-user. Submits the programs and data to a resource broker or primary node, which, in turn, submits jobs to (other) grid resource nodes. The PUA does not care which resource takes on the job. Example: Takes data from PhD Biologists as there is no service available for their needs. Packages data and algorithms and submits these to the grid for processing.
PUS	Power User requiring Specific grid resource nodes. High degree of computing expertise. Example 1: (The example of an expert serving the needs of PhD Biologists or Humanities researchers fits equally well here). Example 2: PUS has an open-ended project that calls a grid-connected telescope studying sunspot activity. PUS has to be specific about the telescope and may or may not have to be specific about other computing resources used.
PUDS	Power User Developing a Service. High degree of computing expertise. As PUA/PUS but developing expertise like SP. *Examples would include the developers on the BASIS project, the BRIDGES project (http://www.brc.dcs.gla.ac.uk/projects/bridges/), the NeuroGrid project (http://www.neurogrid.ac.uk/) and many more.*
SP	Service Provider. High degree of computing expertise. May have expertise in identity management and authorisation. Many of the developers, administrators and owners of projects already mentioned will play the role of SP when the applications mature. A popular method of providing this service is to build a portal, possibly using web services to give an easy interface to the SEU.
Grid-Sys	Infrastructure system administrator. High degree of computing expertise. A Grid-Sys may manage dedicated grid resource nodes (including clusters) and any grid system objects such as resource brokers, authentication, authorisation or accounting points. As well as possibly managing a resource, a Grid-Sys is likely to be responsible for (and may be expert in) security and access management. A Grid-Sys may be the resource manager of a node that accepts jobs (from PUAs) from a resource broker, or of a node that may authenticate and/or authorise PUS users directly where they wish to be specific and use the Grid-Sys' resource without any involvement of the resource broker. A special type of Grid-Sys is someone who hosts a grid resource node for a particular SP, or a set of SPs.

Note that we have not divided the users into the kinds of grid jobs that result from their activity. This may, however, be another valuable approach. For example, one type of user may run a job that executes (or interacts with) only one grid node (a 'single point' job) whereas another may run a job that is divided, or subsequently splits, into many sub-jobs that interact with many grid nodes. These are useful definitions but are probably applicable to nearly all of the actors in Table 2.

3.3. Access management characteristics of these actors

Table 3 describes the access management or security characteristics of the seven user types. The final column of 'Security risk to grid node' tries to capture both the concepts of the threat to the grid resource and the risks (or costs) associated with managing these kinds of users. For example, the threat from an individual user of this type may be fairly low, but the difficulties of managing many users of this type give rise to an associated threat of attackers posing as those users. These are separate concepts, but they have been combined in this case, as it would appear to be appropriate.

The SEUD only ever uses a service, probably presented through some sort of gateway to the grid beyond. Therefore, the security risk to the grid resources from this user should be much lower than the other users. It is assumed that this user cannot interact directly with any grid nodes. Whatever threats that may exist from this type of user, there is the added defence of a restrictive application layer between the user and the grid node.

A similar profile could be expressed for the SEUX. However, a greater risk exists from those users due to executable code being run. Note also that there are similarities between the SEUX and PUA, using a resource broker. The main differences are in computing expertise, the use of a SP and that one is a true end user.

The PUA does not interact directly with any grid node (apart from the resource broker) and therefore should pose a lower security risk than the other users, apart from the SEU. Nevertheless, code written by or submitted by the PUA will be run on a grid node somewhere and therefore the security risk may be seen as being moderate.

The PUS interacts directly with grid nodes, running code on those nodes. The security risk, from the viewpoint of the resource owners, is therefore much higher from this type of user and, if her identity is not known it would be a requirement that it could be traced easily and very quickly, should any ‘breach in security’ occur. Therefore, despite the lack of necessity of the identity of the PUS being known to the grid node, it has been acknowledged in Table 3 that some grid node owners may feel more reassured if this is the case (but see the later section on *Emotional security* on page 10). Nevertheless, the benefits of a system whereby the user can be traced accurately, when problems occur, should outweigh the benefits of logging identities at each grid node.

The PUDS has a similar security profile to the PUS, but is beginning to take on some of the aspects of a SP and therefore could pose a threat to both the grid and to the test SEUs involved in the development. When interacting with the grid, there may therefore be a requirement for the PUDS to be explicitly identified.

As Table 3 indicates, the SP has a complex security profile. The SP (machine) is likely to be trusted by and/or to be explicitly identified to both SEUs and to grid nodes. The SP (user) is also likely to have a profile similar to a PUDS when developing and testing and connecting to grid nodes directly. The SP machines may or may not be considered as part of the grid: these machines may simply be gateways to the grid and not contribute directly to grid computation.

The security profile of the Grid-Sys has been expressed as ‘Moderate’. This is due to two opposing influences. Firstly, for each grid node, there will be very few system administrators, almost certainly in single figures. This means that the task of managing these users’ authorisation information – and possibly authentication mechanisms – is relatively simple. Secondly and conversely, if an impostor were to be able to bypass the access management system, the risks are very high to the grid node.

Table 3 Access management/security characteristics of the seven user types

User/actor	Access management/security characteristic	Security risk to grid node
SEUD	SEUD does not need to be 'known' by a grid access management service (should one exist) as the grid trusts and accounts the SP not the user. SP may need to authenticate, authorise and account for the user as well as possibly taking on 'metering' responsibilities.	Low (shielded by gateway/ application).
SEUX	The SEUX may have a similar access management characteristic to the SEUD due to the possible greater absolute numbers. The presence of SEUX will probably mandate the automated trace/isolate functionality discussed in section 4.3 on page 13.	Moderate.
PUA	The PUA's identity need not be managed by a grid access management service (should one exist) but some sort of mapping to a billing account may be necessary. It could be possible for the identity of the PUA to be concealed behind another entity, as occurs with the SEU. This entity could be a SP providing grid brokering services. Either the SP or the grid access management service is likely to require status (and other) information from an identity manager/provider for authorisation purposes.	Moderate (shielded by resource broker).
PUS	As for PUA, above in some scenarios. However, in addition, grid node owners may wish to have a direct authentication, authorisation (and accounting) relationship with the PUS. Alternatively, authentication elsewhere may be acceptable if a more transparent assertion of identity is given in order to satisfy the security instincts of grid node owners.	Moderate/high.
PUDS	As for PUS but moving into arrangements like SP (see below). May need to begin interacting with and accounting for SEUs in an experimental manner.	High (as for SP, see below).
SP	A SP may be trusted to provide services only to those authorised to use the grid or the SP may offer services to any end user, and be simply billed by the grid, or by the nodes that it uses. The SP may wish to manage identities and to authenticate SEUs or the SP may be willing to devolve these tasks. The SP probably needs to manage or recognise status (authorisation-related attributes). The SP needs strong/secure assertions of identity/authentication between it and the grid resource nodes. Accounting may be required between the grid resource nodes (or access management service) and the SP and between the SP and the SEU, although this latter requirement may not need to be met using grid middleware. As an individual, the SP could use any method (including that of devolved authentication) of access management to his/her machines (to which the SEUs connect or utilise in some way). Moreover, those machines may or may not be considered to be part of the grid.	High (impacts security both of grid nodes and of SEUs).
Grid-Sys	A Grid-Sys is likely to need to authenticate directly to particular grid resource nodes. However, in theory, it is possible that he may authenticate elsewhere and the node computer may trust that external authentication point (or identity provider).	Moderate (High risk but more easily managed).

3.4. An access management scenario

3.4.1. Two major routes of entry to the grid

Figure 1 outlines a likely scenario illustrating the access management 'behaviour' of these different types of grid users. As we have already established in Table 3, there are a variety of ways in which the access management requirements of each set of users, and of each resource protecting itself from each user, may be fulfilled. The scenario presented in Figure 1 is merely one of many that are possible. Nevertheless, we have depicted the PUA acting in two different ways, as these two ways are likely to be significant.

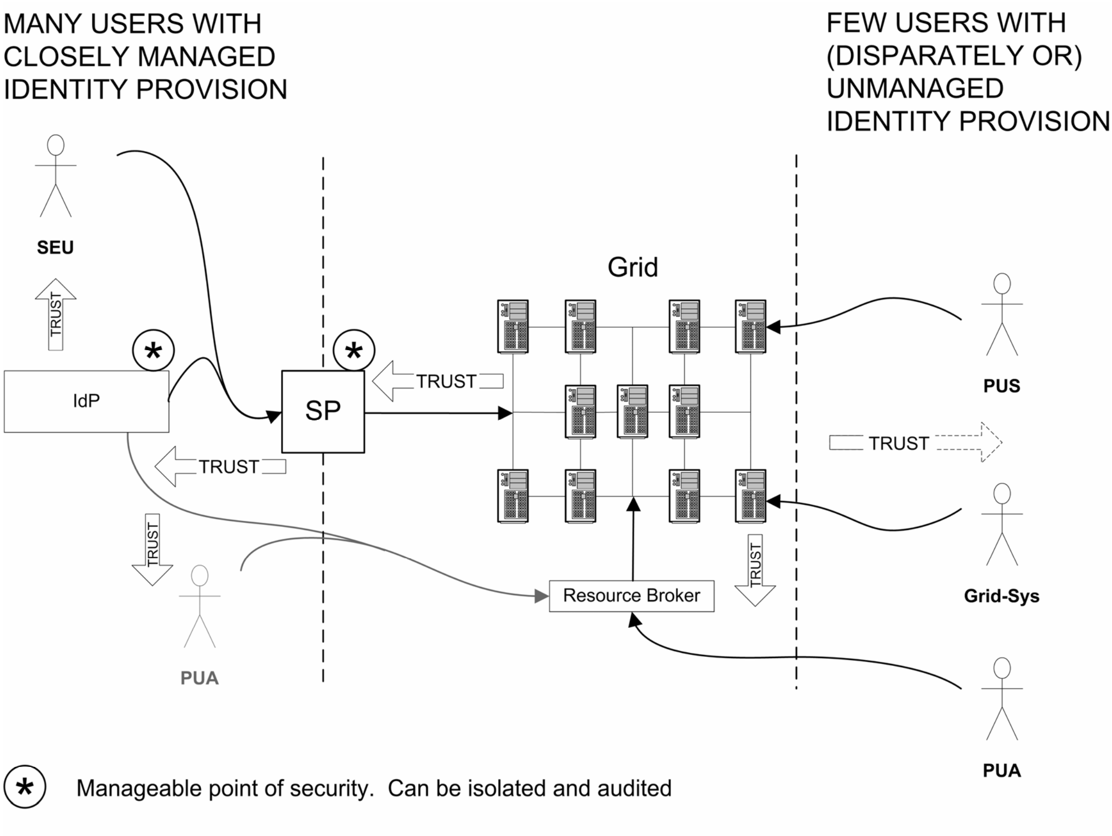
MANY USERS WITH CLOSELY MANAGED IDENTITY PROVISION
FEW USERS WITH (DISPARATELY OR) UNMANAGED IDENTITY PROVISION
SEU
TRUST
*
IdP
*
SP
TRUST
TRUST
TRUST
Grid
PUS
TRUST
Grid-Sys
TRUST
Resource Broker
PUA
PUA
*
Manageable point of security. Can be isolated and audited

Figure 1 Possible access management behaviour scenario of the different types of grid users. (The PUDS has been omitted as it should contain elements of the PUS and SP).

3.4.2. A note about identity management

At this point, it is worth discussing briefly the concepts of 'identity' and 'identity provision' and the management of identity. In a perfect on-line world, identity 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!

3.4.3. Identity and attribute management

Currently, it is easier to combine identity, authentication and authorisation to some degree. Identity tokens (accounts and passwords, 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. In Figure 1, possibly the greatest problem on the right side of the diagram is the difficulty of managing the identities. With grids, this is often not performed well, due to the over-centralised nature of the identity provider, usually the certification authority (CA): when the user changes organisation, the CA may not be informed. Nevertheless, if this identity management is handled better, the difficulty then becomes managing the authorisation attributes. These issues are not of great relevance if the numbers of users are relatively low. On the left side of the diagram, the identities are managed at the same place as the authorisation attributes. This is far more scalable to high numbers of users.

3.4.4. Trust and auditable security

In Figure 1, broad arrows have been used to depict trust 'routes' worthy of discussion. The trust relationships would clearly be far more complex (for instance, the SEU would trust the IdP with his private information, and both would trust the SP). The grid (or grid nodes) needs to trust most of the other actors to a certain extent and this is formalised through security and use policies. The most difficult trust relationship is between the grid and the less well-managed users (PUA, PUS, PUDS and Grid-Sys). This is mostly due to the absence of a point that can be easily audited. In contemporary grids, the CA could be audited, but the CA does not often have a close relationship with the users. Furthermore, there is often no clear point that could be audited with respect to authorisation attributes. Again, this could be tolerable from the viewpoint of the grid node owners if the numbers of such users remains low.

When considering the SP, IdP and SEU (and possibly PUA) combination, a trust route is clearly visible and easily understood. There are also obvious points for a security audit: the SP itself and the IdP. This appears to be more robust and should scale to many more users than the unmanaged or disparately managed identity mechanisms.

3.4.5. An appropriate scenario for power users?

In the above sections we have built up the idea that power users may rely on centralised identity and attribute management, rather as most grid users do today. Alternatively, non-technical end users (SEUs) may rely on devolved authentication and attributes managed at organisations local to them (see section 5 for more discussion of this concept).

One category of power users – the PUAs – may, however, have access management requirements or capabilities in similarity with the SEUs. The grid community may be willing to allow devolved authentication for these users, rather like they may with the SEUs. This is why the PUAs appear twice in Figure 1. It is relatively easy to envisage them behaving as the other power users (the PUSs and Grid-Sys's, for example) but there is also a case for them to have devolved authentication. This is because a trusted resource broker plays a role between the PUA and the grid. The two possibilities have been outlined in Figure 1 as it is unclear as to which mechanism – or if both – would be most acceptable.

3.4.6. 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 she 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.

3.4.7. Trace and isolate

Bringing the focus back to grid security, 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 without knowing identity 'up front' or by logging permanent identities. Therefore, the main requirement should be the detection of misuse or security breeches and the quick tracing of the identity of the user, rather than constant logging of identity.

The thinking behind Figure 1 anticipates that the power users of the future will be required to transmit their identities and authenticate at every transaction. This gives emotional security which is – on its own – quite unsatisfactory. The current "requirement" for identities to be used throughout grid middleware is the current *solution*, not the requirement. The real requirements are for:

- good authorisation procedures;
- quick identification of wrong-doers;
- rapid revocation of rights, possibly throughout the gird;
- (in most cases) the rapid revocation of ongoing jobs throughout the grid.

4. The Customer-Service grid relationship

4.1. SEUs dominate

It is clear, from our earlier assumptions, that the vast majority of 'users' of the grid, in future, will probably be Service End Users and, individually, these SEUs pose the lowest security threat as their activities are highly controlled by the SP and the service application. If we accept these as basic assumptions, we can see some advantages for the simplicity of a multi-tiered security architecture. As Figure 1 shows, there is trust between the grid and the SP and between the SP and the entity or organisation managing the user's credentials. Furthermore, the SP and the IdP are clear auditable points. We can thus envisage the SP as the true grid user. It is the SP entity that runs jobs on the grid and, if it were a commercial grid, the owners of the grid nodes could charge the SP for the use of their resource. Thus a clear relationship between the grid and the SP begins to emerge. Where particular authorisation requirements exist – such as "only members of organisation *A* can use this grid at this time" – the grid could mandate that the SP honour that requirement, and the SP could be audited for this. The SP is thus required to take responsibility for authorising users.

The SP is therefore behaving rather like a traditional regulated service provider. For example, a business in the UK can apply for a franchise to sell electricity to consumers and to attach them to the national (electricity) grid. The organisation that runs the national grid does not need to know who the end users are, unless anything goes wrong or laws are being broken. The grid simply charges the electricity SP for the number of units drawn and the SP charges all of its customers. As a fictitious extension to this, there could be a requirement to only supply electricity to British citizens. It would therefore be clearly the responsibility of the SP to check on the eligibility of the end user, not of the engineers and power station organisations who run the grid.

This customer-service concept does not need to rely upon any financial requirements. Even in an academic world – but one in which access is restricted to only certain communities – it would be appropriate to run application-based services to high numbers of users in this way.

For ease of use, the vast majority of users will access the power of grids via portals, portlets or similar server-based applications. If we accept that this is true, then we can take the opportunity to tighten up security for all of these users. The portal/server represents a point to which we can – technically or legally – devolve the responsibility for authentication and authorisation. This is a truly synergistic opportunity by:

- improving usability to users who would never benefit from the grid if it meant that they had to perform technical computing operations to reach that point;
- introducing an 'auditable' point of security to which authentication and authorisation may be securely devolved.

The BRIDGES project built both a data and a compute Grid infrastructure accessible by a portal which allowed biomedical researchers to authenticate (using a simple username/password mechanism). Scientists were then able to upload nucleotide (or protein) sequences and compare them against a variety of local and remote genomic databases. Explorations in rolling out X.509 *user* certificates to the BRIDGES scientists, for identity/authentication purposes, were largely unsuccessful. Instead, solutions utilising X.509 *server* certificates were adopted. Scientists were more comfortable with username/password solutions and to encourage uptake, these requirements were directly catered-for. Numerous other challenges were addressed in BRIDGES such as re-engineering of client side tools for simplicity and user friendliness, e.g. to make them "google-like". In short, the scientists wanted a familiar environment in which to work, which shielded them as far as possible from the underlying Grid infrastructure.

Similarly, the Market for Computational Services project (Grid Markets, 2003) asserts that the evolution of the grid is constrained by the fact that users can only use machines where they have accounts. This approach is largely – but not entirely – aimed at power users in that the user has to engage at a much more technical level with each grid node. The user experience is greatly simplified in the Grid Markets project by interacting via a central broker. A logical extension to the findings of the BRIDGES and Grid Markets projects means that a lack of usability can mean an absolute lack of take-up, which in turn makes it difficult to survey users regarding their usability comments.

4.2. The main threat with the Customer-Service model

The main threat with the Customer-Service model, if implemented efficiently, is likely to be from denial of service (DoS) attacks on SPs. From the grid's or grid node's point of view, the user is the SP. Should any breach in security occur, the normal reaction would be to revoke the SP's privileges, temporarily or permanently. This seems reasonable. However, this means that all users benefiting from the service provided by the SP and the grid will by stymied.

A balance would need to be struck between the risk of this threat and the ability of the SPs to build reasonably safe applications. With such an application as the BLAST technology provided-for by the BRIDGES project, for example, it is difficult to see many threats to the SP other than:

- users submitting jobs incessantly, and thus tying up the databases and the compute cycles or
- submitting a cleverly formulated nucleotide sequence that never resolves and stays busy (as an extreme example).

Clearly, problems will occur, as they do with any multi-user application, but they should be able to be either mitigated-for in advance, or dealt with as they arise.

If an SP provides an application with very poor security then that SP clearly deserves to be suspended until such problems are fixed.

4.3. Automated suspension

Rogue, clever, end users may exist and these will need to be quickly identified. Where the service holds the authentication responsibility itself, then this should be very simple. Where the service relies upon third parties to perform the authentication (as outlined in Figure 1, above), policies should be in place to trace the users very rapidly with the co-operation of those third parties. Reasonably, an automated process should be in place whereby the SP can immediately cause the IdP to suspend the activity of a particular user. This would then be followed by human attention within the IdP to identify the user and investigate which actions should be taken.

5. Shibboleth

Building on the previous sections, we have established how the majority of users of a grid may be 'funnelled' via a server-based application so that requests and jobs may be run on the grid for them. 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. There are arguments to say that this is far more secure than managing these issues centrally (Norman 2005).

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 possible that Shibboleth could – one day – be extended to accept authentication within one organisation and the retrieval of attributes from another (virtual) organisation. This is mere speculation at this stage and depends as much upon general attitudes within the community as it does on technical possibilities.

The BRIDGES and ESP-GRID projects are producing a Shibboleth-enabled portal with which to authenticate and authorise people to access the BRIDGES/BLAST application. 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 numbers of power users grow greatly, as we argue in this paper.

6. Conclusions

The main conclusions of this hypothetical thinking regarding the likely users of future grids are:

- Like the mature web, we predict that most users will require simple, secure, ring-fenced applications to obtain the great benefits of grid technology.
- If such applications are placed in portals (probably using web technology), the security threat profile of this vast majority of users is relatively low (being controlled by the application). Thus, heavyweight security solutions will not be needed for the majority of users.
- In such a scenario, power users will exist as a small proportion of users. Those users probably merit heavyweight security solutions to be applied to them.
- Where applications are based for the benefit of most users, these provide convenient ‘funnels’ for such users. Such funnels are suitable for security auditing and therefore are a substantial aid to scalability.
- The SP ‘funnels’ could rely on third parties for authentication and some authorisation purposes. These third parties could be classed as Identity Providers (IdPs) and could, similarly, be audited. This, we believe, adds to the scalability and security of the grid.
- Shibboleth could provide such scalability and security but only if implemented wisely. However, there are some drawbacks to Shibboleth that warrant attention.
- We have categorised the majority of users as Service End Users (SEUs) who interact directly with Service Providers (SPs). The SPs are the users who interact with the grid directly.
- Grids will be used by many power users. We have tentatively named these as PUAs, PUSs, PUDSs, (SPs) and Grid-Sys’s (see Table 3 on page 8 for full definitions). There are more ‘actors’ in such a system, but we believe that these capture most of the users who interact with the grid directly.

- We described the concept of the SEU-SP interaction as the 'Customer-Service Model'. This has obvious benefits where grid services require accounting and financial recompense. However, it also has benefits for authorisation issues that will exist in nearly any kind of grid.

7. References

Beckles, B. (2004a) User requirements for UK e-Science grid environments. UK e-Science All Hands Meeting (2004) http://www.allhands.org.uk/2004/proceedings/papers/251.pdf.

Beckles, B. (2004b) Removing digital certificates from the end-user's experience of grid environments. UK eScience All Hands Meeting (2004) http://www.allhands.org.uk/2004/proceedings/papers/250.pdf.

Beckles, B., Brostoff, S., and Ballard, B. (2004) A first attempt: initial steps toward determining scientific users' requirements and appropriate security paradigms for computational grids (2004). Proceedings of the Workshop on Requirements Capture for Collaboration in e- Science, Edinburgh, 14-15 January 2004, 17-43.

Di Gennaro, C., and Dutton, W. H., (Forthcoming) 'Youth, Proximity to the Internet and Political Participation: The Case of Britain', Parliamentary Affairs.

Dutton, W.H. (1999), Society on the Line (Oxford University Press): 227-56, especially table 9.1, page 228.

Fox, S (2005) Digital Divisions, Pew Internet & American Life Project http://www.pewinternet.org/pdfs/PIP_Digital_Divisions_Oct_5_2005.pdf.

Gavaghan, D., Whiteley, J., Pitt-Francis, J., Slaymaker, M., Lloyd, S., Boyd, D., Mac Randal, D., Kleese van Dam, K., and Sastry, L. (2004) Gathering Requirements for an Integrative Biology Project. UK e-Science All Hands Meeting (2004) http://www.allhands.org.uk/2004/proceedings/papers/77.pdf.

Grid Markets (2003). A Market for Computational Services: A Proposal to the e-Science Core Technology Programme. http://www.lesc.ic.ac.uk/markets/Resources/Tag.pdf. Also http://www.sve.man.ac.uk/Research/AtoZ/MCS/RUS/.

NCES (1993). National Center for Education Statistics, *Adult Literacy in America* (Washington: NCES).

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.

Pew Internet (2005) Trends 2005: Internet: The Mainstreaming of Online Life. http://www.pewinternet.org/pdfs/Internet_Status_2005.pdf.

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.

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.

8. 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.

The authors are especially grateful for the critical comments made by Bruce Beckles during the preparation of this paper.