Grid Security: Lessons for Peer-to-Peer Systems

Mike Surridge and Colin Upstill,
IT Innovation Centre, 2 Venture Road, Chilworth Science Park, Southampton SO16 7NP, UK

Overview

The vision of the Grid is to provide a computational infrastructure supporting flexible, secure, co-ordinated resource sharing among dynamic collections of individuals, institutions, and resources [1]. Interest in the Grid has increased as major science programmes look to Grid technology to provide for their computing needs. This has led to substantial investment in the Grid by vendors and governments, notably through the UK e-Science programme and similar programmes in other nations, and more recently at European level. As a result, far more people are joining the effort to develop Grid infrastructure and applications.

The Grid by its nature involves access to computer systems and data outside one's own company or institution. Security is therefore a major element in any Grid infrastructure, as it is necessary to ensure that only authorised access is permitted. However, early developments of the Grid were strongly motivated by the performance benefits of sharing resources, and Grid security models were designed not to interfere with this. We show by comparison with mainstream e-Commerce experience that early Grid security models exhibit several weaknesses [2].

The early development of the Grid also largely failed to take account of operational realities such as network administrator responsibilities and network devices such as firewalls. Early Grid systems were simply not operable outside academic institutions and closed research networks, and we contend that the most common strategy for making them work "in the real world" represents a short-term fix that is likely to produce conflict between users and application developers on the one hand, and those responsible for network administration and security on the other. We believe that the peer-to-peer community is also likely to face similar conflicts between its decentralised management approach and the day-to-day concerns of those entrusted to maintain our security.

IT Innovation is playing a leading role in the UK E-Science Programme and the exploitation of Grids for industrial and commercial purposes in the European Framework programmes. We have found it necessary to propose and begin development of radical solutions to some of these problems, including "proxy-free" delegation models and semantically-aware firewalls.

Grid Security Myths

Computer security is normally designed to protect against five basic types of threat:

- an intruder reading your confidential data;
- an intruder changing or removing your data, confidential or otherwise;
- denial of service to legitimate users;
- an intruder using your system to launch attacks against other people's systems; and
- an intruder using your system for other nefarious (and possibly illegal) purposes.

The challenge for computer system developers and operators is to allow legitimate users to go about their business, while preventing unauthorised users from perpetrating these various types of abuse. Unfortunately legitimate users often require very extensive privileges, if not for themselves then for some of the software they use. Systems are therefore vulnerable if anybody can pass themselves off as a legitimate user, or if the software contains bugs that allow users to exceed their

normal privileges. These threats are especially serious for systems connected to public networks like the Internet, including e-Commerce systems and Grid systems.

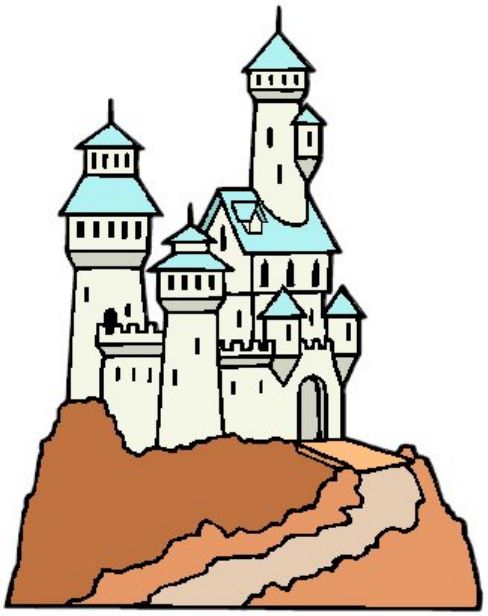

Figure 1. Security in Depth

In the world of e-Commerce, experience has shown that computer systems and their operators are not 100% perfect: authentication systems can be compromised, and software typically does contain bugs that experts can exploit to gain control of the system. Today, e-Commerce systems are designed to have "security in depth" [3], so they can contain the consequences of a breakdown in security and give system administrators time to eject the intruder and effect repairs.

This approach requires the system to be designed like a medieval castle with concentric defences: a moat, drawbridge and portcullis, a high perimeter wall, an even higher inner wall, and a keep. As a last resort, we should also have a few hidey-holes where we can store data we will need to recover, or to produce in court if Interpol ever catches up with the perpetrator. Users' privileges grow only slowly as they move deeper into the castle, so it isn't possible to do serious damage or cover one's tracks without first breaking through several layers of defences. Furthermore, e-Commerce systems have to be subject to constant review, so that any software bugs are detected and patched very quickly, any new threats are recognised, and the overall operation (including human as well as technical aspects) enhanced to deal with them.

The Grid has always placed emphasis on security, but few Grid systems have been designed like e-Commerce systems with "security in depth". For example, Globus [4] was designed from the outset to use encrypted, authenticated communications between participating systems and organisations. Users must authenticate themselves using a PKI certificate when they log onto a Globus grid.

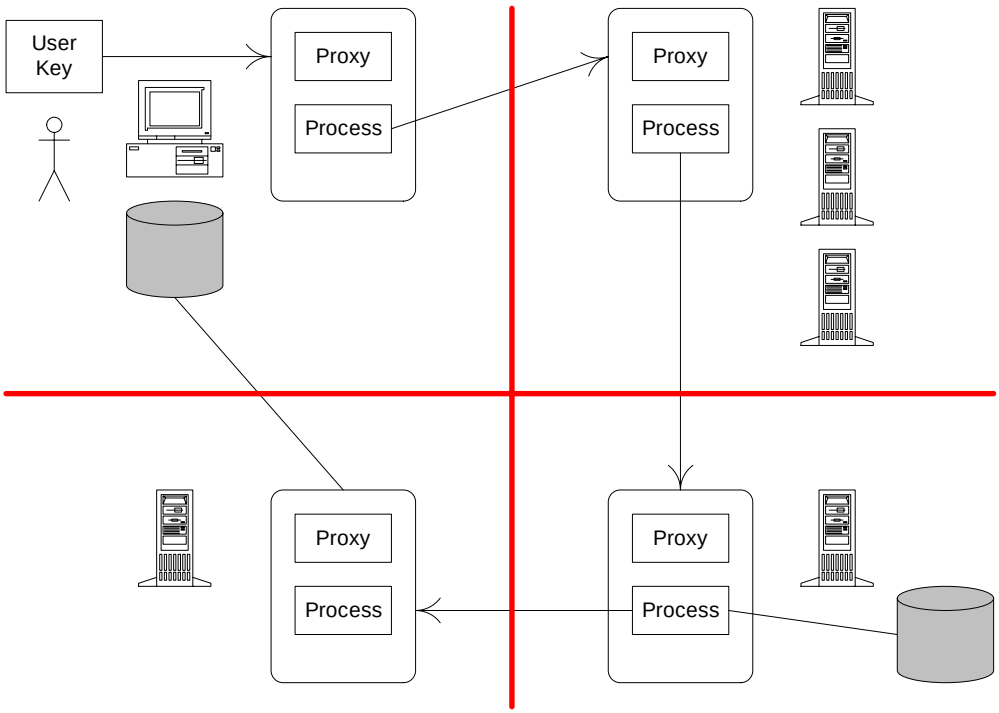

Figure 2. A Grid Certificate Proxy Chain

However, in the interests of transparency and especially performance, Globus applications are allowed to spawn new processes without an authenticated communication direct from the user. Instead, each process is provided with a so-called "proxy certificate" allowing it to initiate and sign its own communications. This proxy certificate is signed by the parent process, and the first (login) process is signed by the user, so the chain of proxies can be traced back to the user and is considered to be under the user's control [4]. The process bearing the proxy certificate is granted some or all of the privileges of the original user, which typically includes uploading and running the user's own code on machines to which they have access.

The performance and scalability benefits of proxy certificates are well proven, and the ability to spawn arbitrary processes running one's own code under a proxy certificate is very attractive to computational scientists with large-scale computing needs. However, the high degree of trust placed in these users and their proxies makes all participants and users of such a grid vulnerable if any site is compromised. This approach provides very few rings of security: if the user's own key pair, or the Grid infrastructure software, or any (user-defined) application process, or any of the sites on which the user's application executes is compromised, the resulting intrusion is immediately serious and yet may be almost impossible to detect. Concerns such as these led IETF to reject certificate proxies from the X.509 operating standards, and to date there is no X.509 compatible proxy mechanism.

Firewalls and the Grid Security Arms Race

Another serious security problem afflicting many Grid systems is their inability to work well in the presence of firewalls. Most sites now use firewalls to prevent outsiders making connections except to specific, low-risk or tightly managed services. Although this is reasonably effective, it also prevents legitimate distributed applications from working. This problem is particularly evident for Grid applications, many of which:

i) use dynamic collections of resources that require peer-to-peer connectivity and bi-directional event notification;

ii) use high-performance UDP-based protocols, whose point of origin is easy to spoof, which leads most firewall operators to block them by default;

iii) involve users sending their own code to run on a grid resource [5], undermining the use of firewall to partition networks into trusted and untrusted domains.

A current solution to these problems is to open various tunnels through the firewall to allow the Grid traffic to pass through. However, since many Grid systems allow users to run their own codes or even access the shell, this simply bypasses the firewall, and leaves one's internal network open to abuse by a malicious imposter or a misbehaving application. Not surprisingly, few system administrators are willing to open the tunnels needed by many Grid systems.

Recently, Grid developers have started using Web Services [6,7] as the basis of a new Open Grid Services Architecture [8]. These exploit "low-risk" protocols such as HTTP [9] that were allowed through by many firewall administrators. The problem is that malicious users can exploit the same mechanisms, so network administrators will soon have to block even these protocols. This "arms race" between network administrators and distributed application developers (including crackers) makes security more complex and expensive and decreases its overall effectiveness. Worse still, it is likely that a Grid infrastructure or application that works today may stop working at some future time if it depends on a particular firewall administration policy.

The Semantic Firewall

To address several of these issues, a group of semantic web researchers from the University of Southampton led by IT Innovation are embarking on a new project known as "The Semantic Firewall". This project will investigate a completely different approach to authorisation and network administration, in which Grid processes are obliged to negotiate with a firewall in order to secure access through it. The Semantic Firewall is so-called because it will exploit semantic web technology, and use machine reasoning about the messages and policies presented to it at a semantic level.

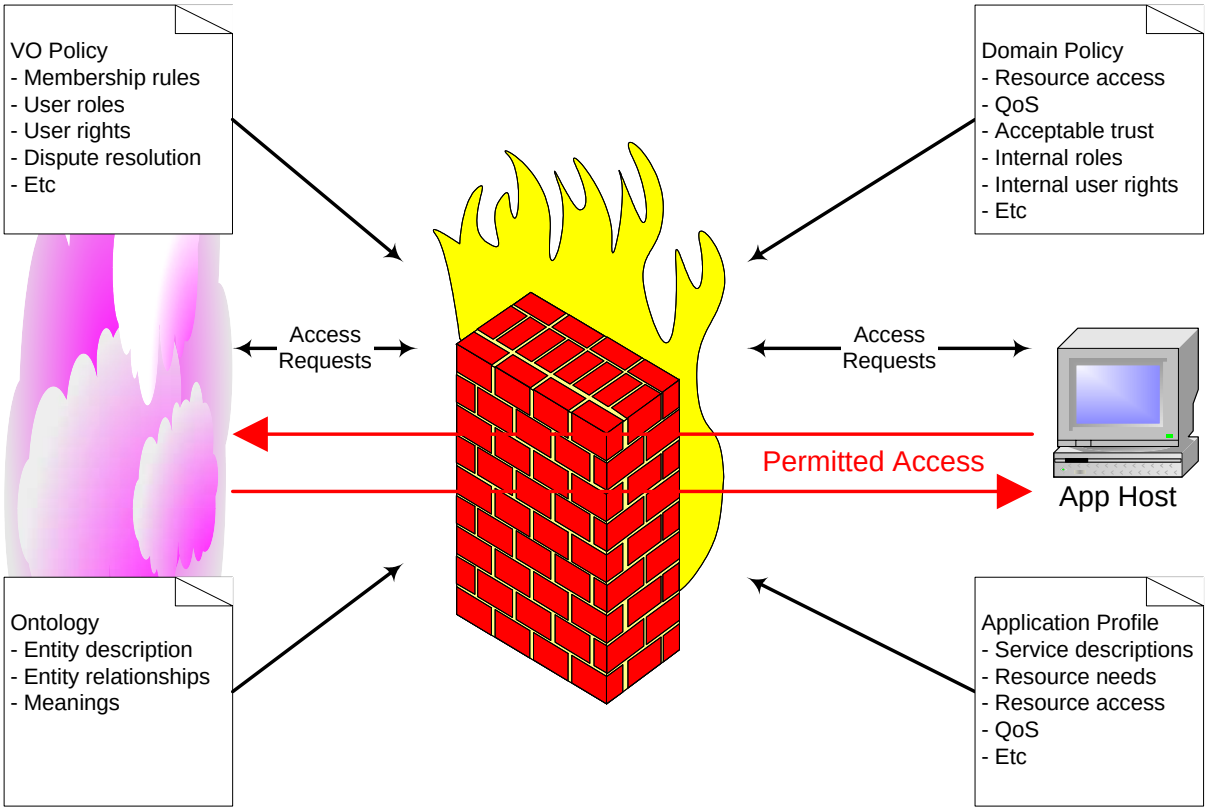

Figure 3. A Semantic Firewall

The idea here is that applications will no longer be responsible for managing Grid security and containing intruders. Instead, applications will present a profile describing their needs to the semantic firewall (at either local and remote sites). The firewall itself will also have access to the policies of the Grid community that the application instance belongs to, and of the domain whose resources it protects. Only if the application requests (from both local and remote processes) are consistent with the policies will the firewall allow traffic to pass through, regardless of which ports or protocols are to be used.

The current focus of our work is to understand the semantics that one might wish to express at each point, covering resource requirements, user rights and (proxy-free) delegation mechanisms. Results are being fed back constantly into our industrial Grid development work, notably in the European 5th Framework IST projects GRIA [10] and GEMSS [11].

Conclusions

There are numerous other problems with the operational security of Grid systems, such as the scalability and appropriateness of Grid authorisation management systems, etc. The Grid community is seeking to address these through a range of developments to support the virtual

organisation as depicted by Foster et al in [1]. Most of these challenges will be familiar to the peer-to-peer community, and experience from the Grid suggests there are no easy solutions.

Our main concern today is that even more basic problems like proxy-free delegation and firewall-friendly operation appear to require quite radical technical solutions. The difficulties experienced by the early Grid pioneers are not technical in nature, but stem from human, organisational and in some cases legal aspects of operating a decentralised infrastructure. The Grid Security Arms Race is a direct consequence of adopting solutions that don't take these non-technical and operational issues fully into account.

Technical people can always find technical solutions to their current problems, but unless the operational needs of organisations are taken into account, such solutions will not provide us with a sustainable future in either Grid or peer-to-peer infrastructure.

References

[1] Ian Foster, Carl Kesselman, and Steven Tuecke. The Anatomy of the Grid, Enabling Scalable Virtual Organisations. International Journal on Supercomputer Applications, 2001.

[2] Mike Surridge, "A Rough Guide to Grid Security", IT Innovation, 2002. Also published by the UK E-Science Programme as UKeS-2002-05. See http://www.nesc.ac.uk/technical_papers/.

[3] Numerous books have been published on computer security. See e.g. Bruce Schneider, "Secrets and Lies: Digital Security in a Networked World", John Wiley & Sons, 2000 (ISBN 0-471-25311-1), or Bob Toxen, "Real-World Linux Security: Intrusion Prevention, Detection and Recovery", Prentice Hall PTR Open Source Technology Series, 2001 (ISBN 0-13-028187-5).

[4] The Globus project is described in: I Foster, C Kesselman, "Globus: A Metacomputing Infrastructure Toolkit", Intl J Supercomputer Applications, 11(2):115-128, 1997. For up to the minute information, see http://www.globus.org.

[5] Volker Roth, "Empowering Mobile Software Agents", Proceedings of the 6th IEEE Mobile Agents Conference, IEEE CS Press, Los Alamitos, Calif., pp. 238-244, 2002.

[6] Gudgin, M., Hadley, M., Mendelsohn, N., Moreau, J. and H. Nielsen, "SOAP Version 1.2 Part 1: Messaging Framework", W3C Working Draft, June 2002.

[7] E. Christensen, E., Curbera, F., Meredith, G. and Weerawarana, S. "Web Services Description Language (WSDL) 1.1" http://www.w3.org/TR/wsdl. 2001.

[8] The OGSA standardisation proposal has been published as a series of working papers at http://www.globus.org/ogsa. The principle documents are "The Physiology of the Grid" by I Foster, C Kesselman, J M Nick and S Tuecke, and "Grid Service Specification" by S Tuecke, K Czajkowski, I Foster, J Frey, S Graham and C Kesselman. The OGSA specifications are now being refined by the Global Grid Forum OGSA Working Group. See http://www.ggf.org/ogsa-wg.

[9] Fielding, R., Gettys, J., Mogul, J., Nielsen, H., Masinter, L., Leach, P. and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1", RFC 2616, June 1999.

[10] See http://www.gria.org.

[11] See http://www.ccrl-nece.de/gemss.