Grids and Private Networks are Antithetical

Andrew Martin* and Carl Cook†

Technial Report TR-04/04, June 2004

Abstract

Achievement of suitable security in Grids is a hard and multi-faceted problem. The deployment of Grids over Virtual Private Networks (VPNs) has been proposed as a (partial) solution. We draw on the experience of the UK e-Science programme, the PlanetLab project, and the New Zealand Grid experience to show that there is a significant mismatch between the capabilities offered by VPNs and the properties required by Grids. Moreover, we argue that by giving attention to Grids implemented in private networks, certain security issues may be overlooked for too long.

*Oxford University Software Engineering Centre
†Department of Computer Science and Software Engineering, University of Canterbury

1 Introduction

The notion of a Grid (Foster and Kesselman, 1999) has gained popularity as a metaphor and guiding principle for system architectures designed to permit large-scale resource sharing across widespread heterogeneous collections of systems. An important feature is the notion of a dynamic *Virtual Organisation* (VO) in which a collection of individuals or organisations share resources in an *ad hoc* way for a period of time, with minimal effort required to set up or finalize the organisation.

It has been clear that careful consideration of security issues is central to the successful deployment of Grids. Potential resource providers will be reluctant to participate if the possibility of misuse of their resource is too great; potential customers will not use Grid services if they cannot achieve an adequate guarantee of quality of service—including confidentiality, integrity, and availability.

Grid projects such as the Teragrid (NSF, 2001–; Bunn and Newman, 2003) are constructed using continental-scale private networks. In the UK, some have proposed a comparable level of isolation for the core e-Science Grid using VPNs (Fox and Walker, 2003) (the 'Virtual Private Grid'). In this paper we will concentrate on the issues raised by this idea, considering in passing the Grid-ness of the former. Our purpose in writing is to capture some discussions which have largely been informal—though many seem to speculate that VPNs may be useful for Grid work, little has been written. Our more specific purpose is to discourage further suggestions that private networks may be useful.

We begin with a consideration of the relevant key features of a Grid—far from being a uniform notion. Section 3 explains the notion of a private network, and how VPNs can be implemented using strong cryptography and the Internet. Section 4 describes several reasons why this technology seems poorly matched to the Grid needs. A final section draws some conclusions, asking whether this notion of a Virtual Private Grid (VPG) can inform the design of new technologies and solutions.

2 Grids and their challenges

From the outset it has been clear that security would be a central challenge in achieving the Grid vision. Foster, Kesselman and Tuecke (2001) have as one of their three defining features of a Grid that it is

> *necessarily highly controlled with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs.*

Achieving these goals has proven one of the significant challenges of the eScience programme, as other papers in this workshop will record. A very non-exhaustive list of issues is:

- cross-domain working with multiple administrators raises innumerable issues related to local responsibility and accountability: how to harmonise and police acceptable usage policies (ETF, 2003);
- use of the Globus toolkit has entailed many difficulties in interactions with firewalls, and, in the way it has been used to form the UK eScience Grid, questions of how a security breach at one site may propagate to others;

- the European DataGrid exhibited an exploit whereby a simple batch job submitted via its resource broker could compromise the security of the chosen compute element, unexpectedly permitting an interactive login on that host;
- questions have been raised about the ability of end users to manage X.509 certificates and corresponding private keys in a suitable way to achieve reliable authentication;
- tools for authorisation are under development, but capabilities for their exploitation—to achieve fine-grained access control, and so forth—have been explored in only a few domains;
- the creation and destruction of dynamic VOs has proved an elusive goal, since it depends at least on being able to achieve authentication and authorisation.

Significant challenges also face the PlanetLab (*PlanetLab Home Page*, 2004) Grid architecture. Planetlab is a global overlay network—a testbed Grid of nodes running on top of the Internet. Presently, PlanetLab consists of 385 nodes distributed amongst 161 sites worldwide, and exists to support the development of new types of Grid computing schemes, as well as new routing protocols, distributed computing technologies, mass storage architectures, and network management and monitoring tools (Chun, Culler, Roscoe, Bavier, Peterson, Wawrzoniak and Bowman, 2003).

While the concept and ambitions of PlanetLab is positive for the advance of Grids and networking research in general, many architectural issues are beginning to surface as the architecture is opened up to the greater research community. Such issues include:

- the creation of new users within the lab
- the allocation of users to each planetlab site
- the funding model of each node within the network
- the allocation of resources—should it be per slice, per user, per site, or another metric?
- the authentication of users. Currently, a partially automated public key scheme is used, but is this scalable?
- should the administration be centralised to one authority—as it is with ISOC/IANA within the Internet, semi-distributed—each site of nodes administers a set of users, or totally distributed?
- what services and system calls to the host node should the virtual machines have access to?
- how should the architecture respond to security breaches?

The above questions are difficult and non-trivial. The issues of security, privacy, and data integrity are yet to be seriously addressed, and may only be resolved for PlanetLab once a general Grid solution has been found and standardised.

3 Private networks and their capabilities

Although Internet connections are now almost ubiquitous, and of very low cost, various applications and organisations find good reasons to employ leased private networks. They may use the Internet Protocol for their implementation, though many other implementations are possible. By avoiding the public network, whole classes of threat are removed. Much improved confidentiality and privacy are possible.

Private networks can also be used to link separate sites, by use of private leased lines, so that network traffic may travel over a long distance as if it were within a single site. Multinational corporations implement internal networks in this way; and projects such as the Teragrid (Bunn and Newman, 2003) have implemented Grid solutions on this basis. A recent incident (Frederick, 2004) demonstrated that such isolation cannot be total in the modern world, and is therefore no security panacea.

In New Zealand, a research initiative is currently underway which will see the provision of a dedicated nationwide private network. The role of this network is to provide a reliable, low latency, and high throughput medium for Grid computing to increase the degree of collaboration between the leading national research universities (Roxburgh, Pawlikowski and McNickle, 2004). For this research, there are several reasons why a private network is preferred:

- the by-volume cost of traffic is too expensive over the Internet, especially since there are no government subsidies for Internet research expenses
- there are no performance guarantees in terms of what the Internet can provide. This poses a significant problem, because the types of research—including Grid processing of medical imaging data and other resource-hungry tasks—have no tolerance for unpredictable network performance
- despite appropriate funding for capital expenditure, there are no devices in existence that can provide end-to-end performance guarantees over the Internet
- the administrative issues related to running a secured network over the Internet are also a dis-incentive. For example, we expect that all researchers can connect at the physical and logical layers once some basic authentication has been executed, and all other control will be the responsibility of the application layer. In the Internet, this approach will be too vulnerable to security attacks

The question of what constitutes a 'private' or a 'public' network is probably more one of degree than a simple binary choice—see Section 4.4.

3.1 Virtual Private Networks

VPNs implement the same facilities using the Internet, by means of encryption to ensure confidentiality and integrity.

To the end user, VPNs are something of a marvel. They allow a roaming device—a laptop—using any Internet connection to behave as if it were part of their home corporate network, apparently on the 'inside' of any firewall protection, and with (potentially) full access to sensitive network data and resources.

Moreover, this solution is entirely sanctioned and even encouraged by their system and network administrators.

Those administrators are also able to use a VPN to connect remote sites using the Internet. Although the traffic travels over the public network, the encryption prevents clear-text eavesdropping or tampering.

Hardware support

Elsewhere, we demonstrated that network bottlenecks are often a function of processor limitations, not memory or network access speeds (Cook, Pawlikowski and Sirisena, 2002). This performance bottleneck is particularly pertinent to VPNs, due to the encryption of all data payloads. In order to address these computational overheads incurred by the processing of VPN traffic, a new type of gateway device has been introduced in the form of VPN routers, or *concentrators*. Such devices, as described in Held (2003), have specifically-enhanced encryption modules to cut down processing overheads and handle large numbers of simultaneous connections.

Applied appropriately, concentrators can be an effective solution to improving the end-to-end throughput of VPN data flows. Unfortunately, there are several shortcomings. Concentrators are expensive in terms of capital expenditure and ongoing configuration, and must be in-place at every route point in the VPN. Whilst they can handle virtually any number of concurrent connections—which is ideal for organisations that support large numbers of remote connections—the setup costs are too high for situations where only a low number of users per route point exist.

Levels of adaptation

VPNs are not as common within the Internet as many first predicted. They have instead been largely restricted to applications where centralised access dominates, due to the flexibility and control limitations. In our own industrial experience, we note that VPNs are restricted to providing remote access for teleworkers, and LAN-to-LAN links between permanent and trusted cooperating organisations.

To explain this apparent lack of global adaptation, our experiences suggest that most organisations do not require the services of VPNs because the majority of traffic is within the LAN. To interconnect LANs the use of VPNs is a possible option, but due to lack of performance guarantees, organisations typically subscribe to carrier-provided leased lines that offer Service Level Agreements (SLAs) such as ATM and Frame Relay. Whilst VPNs certainly offer a low cost compared to leased private lines, most organisations appear to be comfortable in paying for the performance and availability guarantees.

Finally, VPN standards are not well defined, hence it is difficult for organisations to tunnel to each other (the security policies and specific protocols must be tuned, and each parties public key must be shared). Therefore, inter-organisational VPNs typically exist only where organisations have an existing, close relationship.

4 Mismatches

4.1 Point-to-point versus fully-connected Grid

As we have seen, in most Grid models the intention is that the consumer should have no particular care over where a Grid resource resides, or which instance is to be used for a particular task. The resources available are intended to function as commoditized products. It follows that a large number of resource nodes are needed (in order to create a genuine market, cope with failure, provide predictable turn-around times and so on), communicating with a possibly much larger number of user nodes (perhaps in a hierarchal structure, rather than a simple connectivity style).

In order to use a VPN over each connection between a user and a resource node, a potentially enormous number of VPNs will be needed, with associated key management challenges for each. This will almost certainly render the enterprise unmanageable. Even an model in which only the Grid nodes (compute, data, broker, other resource) participate, VPNs will exhibit an exponentially-rising set-up cost for adding new nodes.

4.2 Throughput expectations

Some Grid applications are characterized by their very high 'bandwidth' requirements: the rapid transfer of terabytes or even petabytes is envisaged. In order to implement a VPN between Grid nodes it is clear that dedicated hardware will be required, otherwise valuable compute resources will spend their cycles undertaking encryption. Such dedicated hardware will not only be expensive, it will find it hard to keep up with developments in network capability, and will introduce a 'pinch point' in the data transmission.

The management of resources amongst dynamically changing numbers of VPN endpoints is another difficult challenge. Schemes such as 'Hose' presented by Duffield, Goyal, Greenberg, Mishra, Ramakrishnan and van der Merive (1999) have been proposed to reduce the levels of complexity in managing resources across dynamically subscribed VPNs, whilst simultaneously increasing the levels of throughput. Unfortunately, these schemes have the negative effect of reducing the ability to provide guarantees on Quality of Service, which is surely pivotal to the success of most Grid computing applications.

4.3 Rapid, dynamic set-up

As noted by Schopf (2003), the lack of standard interfaces for Grid link establishment is a significant unresolved issue impeding the progression of Grid infrastructures. As VPN nodes require manual pre-configuration (which is not conducive to the dynamic setup and tear-down of temporary virtual circuits), we fail to understand why VPNs have been suggested as a solution to the establishment of ad hoc, ubiquitous Grid architectures.

A significant difficulty with VPNs, and in particular their relevance to Grid infrastructures, is that they must necessarily depend upon shared secrets—shared keys for symmetric encryption algorithms. Two parties that have neither pre-existing trust between them nor a trusted third party cannot establish a key

over an open medium (like the Internet) without the danger of a man-in-the-middle attack.

Fortunately, the man-in-the-middle attack is not a significant problem because the typical deployment scenarios for VPNs are as a dial-in substitute (so users can agree keys in person at the relevant offices) or between distant LANs connected via the Internet (where the connection will be long-lived; keys can be set up out-of-band using the postal service, for example). A PKI may provide the necessary third-party trust between the two parties, but it is important to note that each connection still requires a negotiated encryption key. While the shared secret used in the VPN may be relatively straightforward to set up using a suitable cryptographic protocol, there is no substitute for a true out-of-band set-up which is used in high integrity situations.

If the VPN is unable to guarantee a high degree of trust (because new nodes are provisioned rapidly, and new relationships established with previously unknown parties) then its security is illusionary.

4.4 Poor security perimeter

Where a VPN links disparate parts of the same organisation, those parts may be subject to the same security regime. The connection between those offices does not materially alter the likelihood of an internal security incident.

Grids are intended to link different organisations, which will almost certainly have differing security policies and enforcement cultures. If a user in organisation A abuses their secure connection to a system in organisation B, sanctions will be harder to apply. The VPN creates an illusion of a single safe, secure zone where in fact none exists.

Perhaps the distinction between a private and a public network is ultimately a subtle social one: the UK Academic Network—SuperJANET—is a private network among the universities, with specific rules, conventions, and services which differentiate it from the public/commercial Internet, but its user community is so wide (and subject to so many different acceptable use policies) that for all practical (and security) purposes, it must be regarded as a public network.

Moreover, surveys of security failures and their cost consistently point to 'insider' abuse being the overwhelming source of problems: although it is reasonable to invest in keeping the bad people away from the network, it is inadvisable to imagine that the private network is therefore safe.

4.5 Vulnerabilities of VPNs

A common misunderstanding of VPNs is that they provide *absolute* security. Whilst VPNs support strong encryption of data over a shared medium, they do not provide complete undetectability of activity. It must be noted that VPNs do not attempt to shield the presence of users within the network, or hide identity of endpoints and hosts, the types of data, and the frequency of data exchanges. If the primary uptake of VPNs within Grid infrastructures is to provide security, the designers of such Grids must be aware of the above vulnerabilities.

To address the problem of traffic snooping within VPNs, an accepted form of defence is the use of a single IPSec channel between VOs, as described in (Herscovitz, 1999). This provides a degree of secrecy and a degree of immunity

to traffic analysis, but as illustrated in (Cohen, 2003), single IPSec channels between endpoints are not always possible.

4.6 Only the network layer

The best a private network can offer is network layer confidentiality. This may be an efficient way to achieve suitable privacy and secrecy, but most requirements of this nature are best expressed as end-to-end arguments (Saltzer, Reed and Clark, 1984), thus *at the application layer*.

The end-to-end argument can be paraphrased, in part, by stating that redundancies are introduced when application-specific routines are implemented at the network layer. As an example given in the end-to-end argument, implementing encryption at the network layer is not only redundant for all clear-text applications, but it exposes the security mechanisms outside of the applications that request the security. In this regard, a private network—actual or virtual—gives an illusionary sense of confidentiality.

5 Discussion

The authors have encountered often enough the informal suggestion that VPNs may help to secure Grids, though mercifully few papers seem to advocate this approach. Our intention has been to observe that VPNs offer a very poor mechanism for the implementation of Grids. More broadly, although the use of private networks to implement Grid solutions is desirable today, they do not offer a long-term solution.

Leased private networks will be valuable in some contexts for providing high-bandwidth linkages between certain Grid sites, for the reasons we have discussed. VPNs offer no such benefits—and are more likely to place an unnecessary limit on capacity.

It is clear that the challenges of Grid security are sufficiently complex that they must be handled in the application layer through careful design and suitably fine-grained policies, and private networks can offer no help in this matter. Indeed, they may prove to be a distraction since they will cause us to make assumptions about the network which will not necessarily hold.

For example, Globus expects to use multiple TCP ports in creative ways, because all its early deployment scenarios involved services not mediated by firewalls. Trying to deploy the software across UK Universities where each campus deploys its own border security has been very problematic (ETF, 2003). We must avoid building non-scalable solutions of this kind.

We are reassured to revisit 'The Anatomy of the Grid' (Foster et al., 2001), and find

> *VPNs and static configurations make many VO sharing modalities hard to achieve. ... A basic problem is that a VPN is not a VO: it cannot extend dynamically to encompass other resources and does not provide the remote resource provider with any control of when and whether to share its resources."*

In this paper, we have illustrated the significant differences between the capabilities afforded by VPNs and the demands required by Grid infrastructures.

There is a possibility that the overwhelming majority of the community has already decided that VPNs and Grids are a mismatch, in which case our fears are, to a degree, unfounded. To the best of our knowledge, however, this argument has not been previously documented in any formal manner.

References

Bunn, J. J. and Newman, H. B. (2003). Data-intensive grids for high-engery physics, *in* F. Berman, G. Fox and T. Hey (eds), *Grid Computing: Making the Global Infrastructure a Reality*, Wiley.
`http://www.grid2002.org/`

Chun, B., Culler, D., Roscoe, T., Bavier, A., Peterson, L., Wawrzoniak, M. and Bowman, M. (2003). Planetlab: An overlay testbed for broad-coverage services, *SIGCOMM Comput. Commun. Rev.* **33**(3): 3–12.

Cohen, R. (2003). On the establishment of an access vpn in broadband access networks, *IEEE Commmunications Magazine*, Vol. 41 of *2*, IEEE Press, pp. 156–163.

Cook, C., Pawlikowski, K. and Sirisena, H. (2002). ComAN: A multiple-language active network architecture enabled via middleware, *5th International Conference on Open Architectures and Network Programming (OPENARCH)*, IEEE, New York, USA.

Duffield, N. G., Goyal, P., Greenberg, A., Mishra, P., Ramakrishnan, K. K. and van der Merive, J. E. (1999). A flexible model for resource management in virtual private networks, *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, ACM Press, pp. 95–108.

ETF (2003). Building an e-science grid for the uk, *Technical report*, UK e-Science Core Programme.
`http://tyne.dl.ac.uk/ETF/public/l2g_final_report.pdf`

Foster, I. and Kesselman, C. (eds) (1999). *The Grid*, Morgan Kaufman.

Foster, I., Kesselman, C. and Tuecke, S. (2001). The anatomy of the grid: Enabling scalable virtual organizations, *International J. Supercomputer Applications* **15**(3).
`http://www.globus.org/research/papers/anatomy.pdf`

Fox, G. and Walker, D. (2003). e-Science gap analysis, *Technical Report UKeS-2003-01*, UK e-Science Core Programme.
`http://www.nesc.ac.uk/technical_papers/UKeS-2003-01/index.html`

Frederick, D. (2004). Security incident notice.
`http://news.teragrid.org/announcements/archive/20040407_02.php`

Held, G. (2003). Focus on the Asant FriendlyNet VR2004 Series VPN Security Router, *International Journal of Network Management* **13**(6): 427–432.

Herscovitz, E. (1999). Secure Virtual Private Networks: The Future of Data Communications, *International Journal of Network Management* **9**(4): 213–220.

NSF (2001–). Teragrid project.
`http://www.teragrid.org`

PlanetLab Home Page (2004). Online document. The PlanetLab Consortium.
`http://www.planet-lab.org`

Roxburgh, A., Pawlikowski, K. and McNickle, D. (2004). Grid computing: the current state and future trends, *Technical Report TR-COSC 01/04*, Department of Computer Science and Software Engineering, University of Canterbury, Christchurch, New Zealand.

Saltzer, J. H., Reed, D. P. and Clark, D. D. (1984). End-to-End Arguments in System Design, *ACM Transactions in Computer Systems* **2**(4): 277–288.

Schopf, J. M. (2003). Grids: The top ten questions, *International Symposium on Grid Computing*, Taipei, Taiwan.