University of Glasgow | Experimental Particle Physics

GLAS-PPE/2008-35
26th November 2008

Department of Physics and Astronomy
Experimental Particle Physics Group
Kelvin Building, University of Glasgow
Glasgow, G12 8QQ, Scotland
Telephone: +44 (0)141 330 2000 Fax: +44 (0)141 330 5881

Integrating Security Solutions to Support nanoCMOS Electronics Research

R. Sinnott, C. Bayliss, T. Doherty, D. Martin, C. Millar, G. Stewart, J. Watt
National e-Science Centre, University of Glasgow

A. Asenov, G. Roy, S. Roy
Dept of Electronics and Electrical Engineering, University of Glasgow

C. Davenhall
National e-Science Centre, University of Edinburgh,

B. Harbulot, M. Jones
e-Science North West,University of Manchester

Email: r.sinnott@nesc.gla.ac.uk or t.doherty@physics.gla.ac.uk

Abstract

The UK Engineering and Physical Sciences Research Council (EPSRC) funded *Meeting the Design Challenges of nanoCMOS Electronics* (nanoCMOS) is developing a research infrastructure for collaborative electronics research across multiple institutions in the UK with especially strong industrial and commercial involvement. Unlike other domains, the electronics industry is driven by the necessity of protecting the intellectual property of the data, designs and software associated with next generation electronics devices and therefore requires fine-grained security. Similarly, the project also demands seamless access to large scale high performance compute resources for atomic scale device simulations and the capability to manage the hundreds of thousands of files and the metadata associated with these simulations. Within this context, the project has explored a wide range of authentication and authorization infrastructures facilitating compute resource access and providing fine-grained security over numerous distributed file stores and files. We conclude that no single security solution meets the needs of the project. This paper describes the experiences of applying X.509-based certificates and public key infrastructures, VOMS, PERMIS, Kerberos and the Internet2 Shibboleth technologies for nanoCMOS security. We outline how we are integrating these solutions to provide a complete end-end security framework meeting the demands of the nanoCMOS electronics domain.

Integrating Security Solutions to Support nanoCMOS Electronics Research

R. Sinnott, C. Bayliss, T. Doherty, D. Martin, C. Millar, G. Stewart, J. Watt
National e-Science Centre, University of Glasgow

B. Asenov, G. Roy, S. Roy
Dept of Electronics and Electrical Engineering, University of Glasgow

C. Davenhall
National e-Science Centre, University of Edinburgh,

B. Harbulot, M. Jones
e-Science North West,University of Manchester

r.sinnott@nesc.gla.ac.uk

Abstract: The UK Engineering and Physical Sciences Research Council (EPSRC) funded *Meeting the Design Challenges of nanoCMOS Electronics* (nanoCMOS) is developing a research infrastructure for collaborative electronics research across multiple institutions in the UK with especially strong industrial and commercial involvement. Unlike other domains, the electronics industry is driven by the necessity of protecting the intellectual property of the data, designs and software associated with next generation electronics devices and therefore requires fine-grained security. Similarly, the project also demands seamless access to large scale high performance compute resources for atomic scale device simulations and the capability to manage the hundreds of thousands of files and the metadata associated with these simulations. Within this context, the project has explored a wide range of authentication and authorization infrastructures facilitating compute resource access and providing fine-grained security over numerous distributed file stores and files. We conclude that no single security solution meets the needs of the project. This paper describes the experiences of applying X.509-based certificates and public key infrastructures, VOMS, PERMIS, Kerberos and the Internet2 Shibboleth technologies for nanoCMOS security. We outline how we are integrating these solutions to provide a complete end-end security framework meeting the demands of the nanoCMOS electronics domain.

Keywords: authentication, authorization, virtual organizations, PERMIS, VOMS, Kerberos, X509, Shibboleth

1 Introduction

The constantly decreasing scaling of transistors in complementary metal oxide semiconductor (CMOS) integrated circuits has fuelled the phenomenal growth and success of the global semiconductor industry. This has been well captured over the past 40 years by Moore's law [1]. However, the International Roadmap for Semiconductors (ITRS) [2] is nearing the limits of physical scaling with sub-10nm transistor dimensions scheduled for mass production in 2016. With the dimensions of a typical transistor in current processor technologies at 65nm, the race is on between major semiconductor manufacturers to demonstrate that they are capable of manufacturing devices with such nano-scale dimensions.

Taking on the scaling challenge demands that the fundamental discreteness of charge and matter, at these scales, must be considered and integrated into the entire electronics design process. There are at present no integrated methodologies that can capture the full complexity of this problem and allow the accurate prediction of both the characteristics and scale of these intrinsic fluctuations in transistor performance and power consumption, and their subsequent impact on the performance of circuits and systems. Rather, the current approach is to assume that all transistor devices behave in a similar manner. This assumption, whilst greatly simplifying the task of circuit designers, is no longer valid on these nanometer scales, since the variation in the number and distribution of dopant atoms within each macroscopically similar transistor makes each one microscopically different, which introduces significant differences from device to device [3-5]. In the presence of such intrinsic parameter fluctuations, the emphasis has shifted from predicting the characteristics of a single transistor to predicting the statistical behaviour of ensembles of macroscopically identical but microscopically different devices. This requires very large number of *ab-initio* simulation of ensembles of devices capturing both the atomic and electronic structure. These simulations can then be used to understand and predict the behaviour of circuits and systems comprising billions of devices where small variations in the behaviour of individual transistors can have a huge impact upon circuits they comprise. Fault

tolerance, power consumption and yield are just some of the aspects of circuit design that are impacted by variability of transistors. Understanding the trade-offs in design and the impact of transistor device variability is key to the future of the semiconductor industry, especially in understanding the constraints and tolerances imposed by device technologies on the entire design process.

The EPSRC Pilot Project *Meeting the Design Challenges of nanoCMOS Electronics* (www.nanocmos.ac.uk) began in October 2006 and has been funded to explore and develop Grid based solutions addressing the research challenges inherent in this space. However, this is not solely a high performance computing project. Given the potentially huge commercial impact of this work (the semiconductor industry is a multi-trillion dollar enterprise), the need for protection of intellectual property (IP) is essential, since numerous major commercial semiconductor partners are directly involved in the project. This IP protection must apply to the commercial software tools used, themselves often costing hundreds of thousands of dollars per year per license; the designs of next generation transistor device architectures; the design and development of circuits and systems; the associated data sets generated and indeed the whole process of design and analysis. Within this context, it is therefore paramount that fine grained security is supported across the entire collaboration. There are many authentication and authorisation based approaches that the e-Science and security communities more generally have developed. In this paper, we present a selection of such technologies that are being explored and integrated for nanoCMOS researchers including Kerberos [6], X509-based PKI solutions [7], VOMS [8], PERMIS [9] and the Internet2 Shibboleth technology [10]. We describe how we have applied these solutions and show how they can interoperate to support user-oriented single sign-on and seamless, secure access to the services and data sets demanded by the commercially sensitive nanoCMOS research domain.

The rest of the paper is structured as follows. Section 2 gives a brief overview of the processes, services and associated data sets that typify the electronic design process, and introduces the overall architecture of the nanoCMOS infrastructure focusing especially on those aspects that require security. Section 3 introduces the various authentication and authorisation technologies used and outlines their primary features. Section 4 outlines key scenarios demonstrating how we have integrated implementations of these security solutions exploiting novel Grid security interoperability standards. Finally in section 5 we conclude on our experiences in applying these security solutions and provide a summary of our plans for the future.

2. NanoCMOS Security Requirements

The overall sets of services and data associated with nanoCMOS research can be classified into several core business-oriented areas – all of which have their own explicit demands for security. These broadly break down into process simulation; device simulation; compact model extraction and circuit and system simulation and timing and power extraction as depicted in Figure 1.

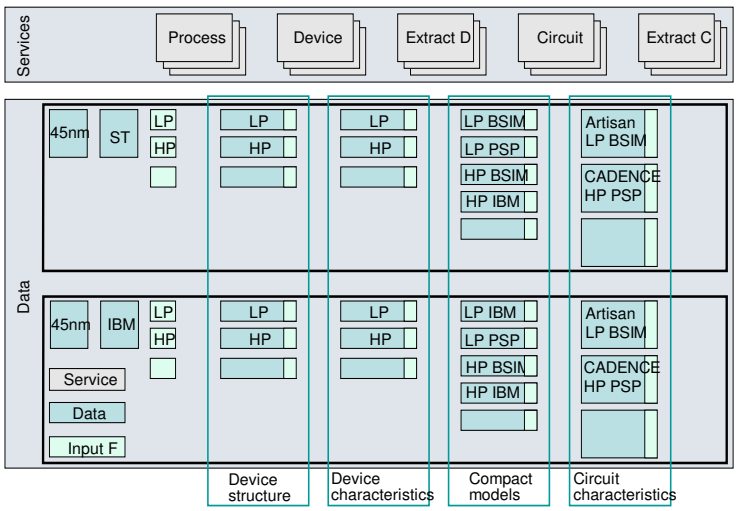

Figure 1: Conceptualisation of the nanoCMOS Services, Data and Design Processes

2.1 Process Simulation

Process simulation is concerned with the physical steps necessary to turn a bare piece of silicon into a working semiconductor device. This includes dopant implantation information; oxide growth; etching, deposition of metals etc. This information is typically supplied by a semiconductor foundry such as IBM or TSMC in a *technology interchange (.tif)* file. This information is extremely commercially sensitive and contains enormous IP value.

2.2 Device Simulation

Device simulation involves the solution of sets of coupled quantum mechanical/drift diffusion equations describing the distribution and flow of electrons in a given device structure. The inputs to a device simulation include information from the process simulation, i.e. the provided *.tif* file containing the doping profile for this device and further input files including the simulation mesh used as the basis for the device simulation. Statistically relevant

ensembles of device simulations need to be run to characterise the behaviour of a particular device. Each of these will have different distribution of individual dopant atoms caused by the stochastic nature of implantation mechanisms.

The output of a device simulation is typically a current/voltage (I/V) curve describing the characteristics of a device with a particular dopant profile. The ensemble device simulation process generates hundreds of thousands of such I/V curves as well as other output. Knowing the current/voltage characteristics of a given device architecture is extremely sensitive information for a semiconductor manufacturer since this information is crucial to understanding the power consumption and performance of the next generation devices. We note that the device simulations themselves are extremely computationally intensive and generate copious amounts of file-based data, all of which can potentially contain valuable IP related results. Device modellers thus need to be able to maintain fine grained access control in terms of where jobs are executed and who is allowed access to the resultant data sets. It is highly unlikely that certain IP sensitive simulations and data sets will be allowed to run on publicly accessible resources such as the UK e-Science National Grid Service (www.ngs.ac.uk) or other similar public HPC resources.

2.3 Compact Model Extraction

Having generated the set of I/V curves for a particular device, it is necessary to abstract this information to a higher level so that multi-device circuit/system simulations can be performed. Compact models are semi-empirical analytical descriptions of the response of a device. A compact model is generated through identification of an extraction strategy (identifying the subset of device model parameters which most influence the curves) and exploitation of commercial tools. This phase typically requires domain knowledge and expertise in identifying the particular parameters that most influence the generated I/V curves. The compact models resulting from this process are used by circuit and system designers in the design of chips, circuits and associated subcomponents such as registers, arithmetic logic units etc. Such compact models have considerable IP associated with them.

2.4 Circuit and System Simulation

Once compact models have been generated they can be used by circuit simulators to predict the behaviour of circuits and systems built from multiple combinations of these compact models. Typical examples of the kinds of behaviour analysed at the circuit/system level with these compact models are to identify how the set of connected components respond to a stepped input voltages or to explore particular tolerances of the integrated circuit. Of course, circuits and systems are themselves extremely commercially sensitive with strong IP-protection demands placed upon their access, usage and interpretation more generally. Similarly, commercial (licensed) applications are often used for circuit simulation. Feedback from circuit simulation may require modifications to the generated compact models which in turn may require device simulations to be redone.

Seamless linkage of each of these processes is essential to the understanding of how atomistic variation of devices impacts upon system level design and profitability. Given that each of these steps also has commercially sensitive IP/license protected processes which need to be addressed, the overall framework that has been adopted in nanoCMOS is based upon the definition of and support for fine-grained security protocols. It is important to note that it is unlikely that a single person, institution or industrial partner will have access to all of the information associated with all of the steps. Rather, each stakeholder needs to have their own access and usage policies in the overall security framework. We note also that this security is driven by non-disclosure agreements, which partners must agree to be bound by before dealing with IP protected designs and data sets.

3 Authentication and Authorisation Technologies

There are many authentication and authorization infrastructures existing today. Username/password challenge responses are perhaps the simplest and most widely adopted authentication solution. One key characteristic of Grid based infrastructures that is essential for nanoCMOS research, is the support for single sign-on. That is, once authenticated, the user is allowed access to a range of resources across many sites without further re-authentication challenge/responses. To support this, the most commonly adopted model for Grid based authentication is based upon public key infrastructures (PKI) [11].

3.1 X509-based PKIs

In PKIs, certificates are used to bind the identity of a user to their public key. These certificates are typically based on X.509 standard[12] and are issued by trusted third parties known as certification authorities (CAs). Through trusting a CA and the procedures they adopt for issuing and revoking an individual's certificates, sites can validate the identity of the individual in possession of the corresponding private key. This process is typically supported through delegation of the identification verification process to a local registration authority at the user's institution. This PKI based approach has been adopted in the UK (www.grid-support.ac.uk/ca) with a single centralised CA and a direct trust chain to users, i.e. the CA does not use subordinate CAs to issue certificates to users.

Once allocated, X509 certificates are often used in the Grid environment through the creation and use of proxy certificates which have a shorter life time, typically of the order of 12 hours. Once a proxy certificate is created, a user is able to access remote resources across a range of sites which recognise both the CA that issued the X509 certificate, and where that user has a local account registered. This account registration is commonly achieved with middleware such as Globus [13], through grid mapfiles which map the distinguished name (DN) of the individual with a certificate to a local user account. Tools such as *grid-proxy-init* allow for creation of the proxy certificates themselves. As

discussed in [14-16], there are many limitations associated with this model of authentication and access control, not least is the scalability of knowing all end user DNs and the fact that this model, from a security perspective, does little more than provide access to a local account to run arbitrary applications/code. Furthermore, the geographical distribution between the identity management process and the issuing authority also leads to problems. Thus a user that moves institutions (or is expelled from a given institution) will still have access to an X509 certificate issued by the CA which can be used to access resources at other sites – in principle without those resource providers being aware that the user is no longer at the same institution.

This model of security has immediate issues with regards to its suitability for nanoCMOS security requirements. since it provides little more than user certificate authentication and there is no direct way to determine whether a user certificate has been compromised and is being used by an imposter. Given that these certificates (or proxy certificates) can be used for arbitrary purposes on distributed resources, the dangers are immediately apparent. Ultimately the problem stems from these certificates being used for authentication whilst nanoCMOS requires finer grained security (authorisation) to be supported. Nevertheless the X509 based authentication approach is widely adopted by the Grid community, hence for non-commercially-oriented simulations, the ability to access and use computing resources such as the UK National Grid Service and ScotGrid (www.scotgrid.ac.uk) is essential.

3.2 Internet2 Shibboleth

To overcome the issues with the identity management of X509 based certificates, UK academia has been exploring the roll-out of the Internet2 Shibboleth technologies for federated access management. The UK Federation (http://www.ukfederation.org.uk) was established in November 2006. Numerous other international access management federations have now been established. The basic model of access control is based upon trust relationships between identity providers (IdPs), who authenticate their own users and service providers (SPs) who provide resources that can be accessed by users from trusted IdPs. In the simplest model, a user attempting to access a Shibboleth protected SP is redirected to a "Where Are You From Service" hosted by the federation and asked to select their home institution. Once selected they are redirected and asked to log in at their home institution. Each institution is free to have whatever local authentication system they wish. Once a user has successfully authenticated, a signed Security Assertion Markup Language (SAML) [17] assertion is returned to the SP which uses this information to make its own autonomous access control decision.

This model depends upon two key components. Firstly, the underlying trust relationship that exists between IdPs and SPs in the federation. This trust relationship has an underpinning PKI where IdPs and SPs are issued with certificates by the federation, i.e. to verify that the signed assertions are actually from an IdP in the federation. Secondly, this model is based upon standardisation of the attributes that are being exchanged. The UK federation is based on a small core set of attributes based around the *eduPerson* schema [18]. These attributes allow the description of which institution a user is from and their role at that institution, e.g. lecturer at Glasgow University.

In the context of the nanoCMOS project, the SP that has been established is based upon a Grid portal. This portal allows access to a range of services and data sets associated with device modelling, compact model generation and circuit simulation. More information on these services is available in [19-21]. In the context of nanoCMOS, Shibboleth offers a distinct advantage in overcoming user identity management and authentication issues caused by X509-based PKIs using a centralised CA. Thus, individuals no longer at institutions within the nanoCMOS project will no longer be able to authenticate at known, trusted IdPs. We note that software support for scoping of the trust relationships associated with IdPs in the UK Federation exists and is described in [22]. The linkage between signed SAML assertions and Grid-based X509 certificates is being addressed by numerous projects including ShibGrid, SHEBANGS and their follow on project SARoNGS [23]. Within the nanoCMOS portal, this linkage is supported through support of proxy credential repositories accessible through the project portal. Thus end users authenticate via Shibboleth to the portal and are then able to create proxy credentials through a portal interface to a MyProxy service [24]. We should point out that an issue with Shibboleth based access control is the scope of the UK Federation itself, this is currently aligned to UK higher/further education academic institutions and is not yet available to industry. Scoping work is on-going in the UK to explore similar federations for, amongst others, industrial collaborations and healthcare access management federations.

Whilst Shibboleth overcomes identity management issues and simplifies access to resources (through not demanding that end users acquire and manage their own X509 certificates), it does not in itself address the finer-grained security needs of nanoCMOS. Thus the core *eduPerson* attributes that are defined in the UK Federation must be extended to support the specific needs of nanoCMOS collaborators, for nanoCMOS-specific access control *authorisation* decisions. This might be through information on the possession of particular commercial licenses and/or specific roles that are meaningful within the context of the nanoCMOS project. With such additional information, access to services can be enforced both at the portal level, i.e. a user not in possession of a given license should not have access to the client interface that allows to invoke that licensed service, but also at the potentially remote service and/or data provider level. Thus the portal is a mechanism through which services and data can be accessed and these can and do exist remotely across multiple institutions which are autonomous. To support such scenarios demands that nanoCMOS-specific authorisation decisions are made. As described in [25] there are several ways in which attributes required for authorisation can be defined and used for authorisation decisions. The decentralised model (as is the case with Shibboleth and multiple independent IdPs) and the centralised model as reflected with the Virtual Organisation Membership Service (VOMS) [8].

3.3 VOMS

VOMS has gained widespread acceptance across the Grid community due to the relatively simple model for defining the roles specific to a particular virtual organisation and how they can be used/enforced by sites, and through being designed specifically to be compliant with existing X509 based Grid authentication approaches. VOMS is based upon a centralised server where roles are agreed across a given collaboration. Once defined, sites are responsible for configuring their resources to use these roles where the typical resources protected with VOMS are pooled accounts on HPC-oriented compute clusters. This is often implemented by local administrators through tools such as Local Centre Authorization Service (LCAS) and the Local Credential Mapping Service (LCMAPS) [26]. These tools map the user role information into group identities (*gid*), user identities (*uid*) on associated pool accounts established on the local cluster for that particular virtual organisation. Typically these accounts are set up and configured with software specific to the particular end users of that virtual organisation. This overcomes one of the main limitations of grid mapfiles which are established on a *per user* basis.

The VOMS model assumes a centralised administrator responsible for assigning roles to individuals. Once assigned, VOMS allows for creation of proxy certificates which embed the roles a particular user has. For VOMS-enabled resources, these attributes are used to make local access control decisions and if successful will map end users to appropriate pooled accounts with associated uid/gid's. For sites that are not VOMS-enabled, the proxy certificates are used as normal X509 proxy certificates, i.e. the embedded VOMS credentials are ignored.

It is important to note that collaborating sites may use VOMS information at their own discretion. This might be providing a mapping to a local pooled account, not allowing access to individuals with certain VOMS roles but allowing access to others etc. This local enforcement is not defined within the virtual organisation policy given by the definition of the roles in the VOMS server, but is left up to local administrators to interpret and enforce.

3.4 PERMIS

The Privilege and Role Management Infrastructure Standards Validation (PERMIS) project (www.permis.org) was an EU funded project that built an authorisation infrastructure to realise a scalable, X.509 attribute certificate (AC) based privilege management infrastructure. Through PERMIS, an alternative approach to centrally allocated X.509 public key certificates can be achieved through the issuance of locally allocated X.509 ACs.

The PERMIS software realises a Role Based Access Control (RBAC) authorisation infrastructure and offers a standards-based application interface that allows developers of resource gateways (gatekeepers) to enquire if access to a particular resource should be allowed. The PERMIS RBAC system supports the definition of policies comprised of rules specifying which access control decisions are to be made for given resources. These rules can include definitions of: subjects that can be assigned roles; sources of authority (SoA), e.g. local managers trusted to assign roles to subjects; roles and their hierarchical relationships; what roles can be assigned to which subjects by which SoAs; target resources, and the actions that can be applied to them; which roles are allowed to perform certain actions on certain targets, and the conditions under which access can be granted to roles. A typical PERMIS scenario is to define rules of the form "Can this person with this role access this resource and perform the following action".

Roles are assigned to subjects by issuing them with X.509 ACs. Policies can then be defined, digitally signed by a manager and stored in one or more LDAP repositories that used these ACs to make access control decisions. The PERMIS infrastructure offers potentially extremely fine grained authorisation capabilities, both in terms of policy expression and enforcement. Policy editing tools allow for easy development of the policies with recent enhancements to support the natural language expression of policies.

One of the primary benefits that PERMIS provides for the nanoCMOS project is its alignment with the Grid middleware and standards community more generally. PERMIS is currently integrated with Globus and the Open Middleware Infrastructure Institute (OMII-UK www.omii.ac.uk) software stack. Typical scenarios using PERMIS and middleware such as Globus include support for SAML callouts that are automatically raised by a policy enforcement point (PEP) and passed to a PERMIS based policy decision point (PDP) when a Grid service is to be invoked as discussed in [27]. More recently, standards for pushing and pulling attribute certificates between various attribute authorities now exist and are now supported by PERMIS as discussed in section 4.

PERMIS is well suited for the expression and enforcement of policies on, access to and usage of, given Grid services. Where PERMIS is less applicable is in dealing with data sets, especially in the case of the highly distributed and dynamic data sets generated from device simulations. In this situation, distributed file systems with fine grained security are required. One such solution, that has been evaluated and now rolled out in the nanoCMOS project is the Andrew File System (AFS) [28], specifically we have adopted the openAFS (www.openafs.org) implementation of AFS. Underpinning AFS is the Kerberos security infrastructure.

3.5 Kerberos

AFS is a distributed networked file system which uses a set of trusted servers to present a homogeneous, location-transparent file namespace to clients. AFS was originally developed as part of the Andrew Project at Carnegie-Mellon University, a distributed computing project which started 1983. (The name comes from the names of the founding benefactors: Andrew Carnegie and Andrew Mellon). Through AFS, a user can log on and securely access a virtual file space crossing multiple heterogeneous resources. To support both structuring and security of this file space, AFS uses organizational units called cells. A cell can be considered as the collection of all the files belonging to an

organisational unit. A cell may correspond to an actual organisation or, as is the case in nanoCMOS a virtual organization. A cell comprises one or more servers and one or more clients. Each server, as might be expected, hosts a collection of files and makes them accessible throughout the cell. Each client allows access to the files hosted by the various servers. Underpinning the security of cells and hence AFS is the Kerberos security infrastructure [29].

Kerberos was initially developed as part of MIT's Athena project over 20 years ago with Kerberos V5 the most current release. Kerberos uses a central ticket issuing authority - a Kerberos Domain Controller (KDC). Like public key based systems, principals are identified by a name and realm in the form *name@REALM*. Realms are approximately synonymous with the domain the entities are hosted in.

Normally users access Kerberos in a single sign-on manner. When a client wishes to communicate with a service or access a protected file store, they first send the KDC a secure message requesting a connection to the service. The KDC, upon authentication of the request, generates a session key and creates a credential and ticket which are returned to the client. The ticket, intended for the server, identifies the client and contains a copy of the session key encrypted with the service's key. The credential identifies the service requested and contains the session key encrypted using the client's key. The client then sends the ticket to the service along with its details encrypted with the session key. The server first decrypts the ticket to get the session key then uses it to decrypt the second half of the message. To complete the handshake the server takes the time stamp from the second half of the message, increments its value by one, encrypts it with the session key and returns it to the client. If the client can decrypt the response from the server and the value is valid then the connection is authenticated.

Kerberos's use of symmetric keys makes it less computationally expensive than public key systems. It also has the advantage of offering greater security per key bit than public key systems. Unlike public key systems trust in Kerberos is established between entities at specific hosts. This helps prevent stolen credentials being used on another host as the session keys are keyed to the host they were generated on.

Kerberos includes a federation system allowing transitive trust relationships to be established between realms. Furthermore, in PKI-based systems certificates become invalid when they expire or when they are added to the certificate revocation list (CRL). Given that a CRL needs to be updated regularly it is possible for compromised certificates to be used after they are technically revoked. Kerberos, through the enforced use of the KDC, is less susceptible to this since any session keys will usually expire within hours and only allow communication with a specific host.

Through use of AFS clients, nanoCMOS researchers are able to securely access and share data sets using Kerberos tokens. Each AFS directory has an associated access control list (ACL). Individual files do not have their own ACLs. Subdirectories inherit the ACL of their parent directory unless explicitly given an ACL of their own. ACLs are typically administered by a directory's owner and/or local system administrators. An ACL comprises a list of entries prescribing who can access the directory and its contents and with what permissions. Each entry comprises a user or group and the permissions granted to that user or group. Groups are a collection of one or more named users. Users are able to create new groups, remove groups and to add individual users to groups, as well as the permissions they have, e.g. to create subdirectories, read and/or edit files etc.

AFS and its support for Kerberos offers many features which make it well aligned with the requirements of the nanoCMOS project. Firstly, it offers a mechanism to manage large quantities of highly distributed files as generated from device modeling simulations, secondly the fine-grained security-oriented access permissions on these directories and files allow collaborators to share their data sets on a *per user/group* basis as required. However, it is essential that the multiple different security solutions identified here are aligned and interoperate to ensure the complete end-end security of the nanoCMOS infrastructure. In the next section we show through a variety of scenarios how this interoperability is currently being supported.

4. nanoCMOS Security Case Studies

The ideal scenario for nanoCMOS research is to have fine-grained end-end security across all resources for all partners involved in the nanoCMOS project. This should protect access to services, data, and designs on an as needs basis. At the time of writing we are working on the complete integration of all of these solutions, however we have implemented a variety of interoperability scenarios between security solutions which we describe.

4.1 Integration of Shibboleth and PERMIS for Portal Configuration

To improve the usability and uptake of Shibboleth technology in Grid environments it is necessary to securely extend the attributes supplied by the UK Federation with attributes applicable to the requirements of specific virtual organizations. The SPAM-GP project (www.nesc.ac.uk/hub/projects/spam-gp) was proposed to provide a set of portlets to support the process of establishing and enforcing, fine grained Grid security in a Shibboleth environment. Specifically the project is developing a family of JSR-168 compliant portlets, with which a Grid portal administrator can tailor access to the resources available behind the portal, i.e. the Grid services which themselves have authorization requirements that need to be met. These include portlets for the scoping of attribute acceptance policies (the Scoped Attribute Management Portlet - SCAMP); security-oriented content configuration of portals (CCP) and for the creation and use of attribute certificates (ACP) as required for potentially remote service authorization.

The SCAMP portlet allows restricted and syntactically correct manipulation of the attribute acceptance policy of a Shibboleth SP to streamline the subset of IdPs from whom a portal will accept user attributes. Thus rather than accepting all IdPs in the UK Federation, nanoCMOS only wishes to accept attributes from those involved in the actual

collaboration itself. To support this, the SCAMP portlet parses the federation metadata associated with the SP for the list of all the IdPs within the federation, and stores the values of the 'scope' entry for each IdP. When the SP is provided with a scoped attribute, the suffix will by definition be one of these scoped values. The list of IdP scopes in the federation is provided to the user/portal administrator in the form of a drop down list, one per user attribute, where the institutions from whom attributes are to be recognized/accepted from may be selected. Once the required sites are selected, these changes can then be added to the attribute acceptance file. This policy information will then subsequently be available for the next browser session referencing that resource, i.e. only allowing access to the resources from known and trusted sites with expected attributes. Figure 1 illustrates the application of the SCAMP portlet. The front end access to the portal is depicted at the bottom of Figure1 below. We note that this portal displays the various attributes (roles) that have been released by the identity provider and attribute authority at the University of Glasgow. We note that in this case, the only attributes that are recognised by the portal are those prefixed with *NanoCMOS* from the trusted (scoped) University of Glasgow IdP. For illustrative purposes, the top part of Figure 1 shows another Shibboleth protected portal but this time without scoping of attributes.

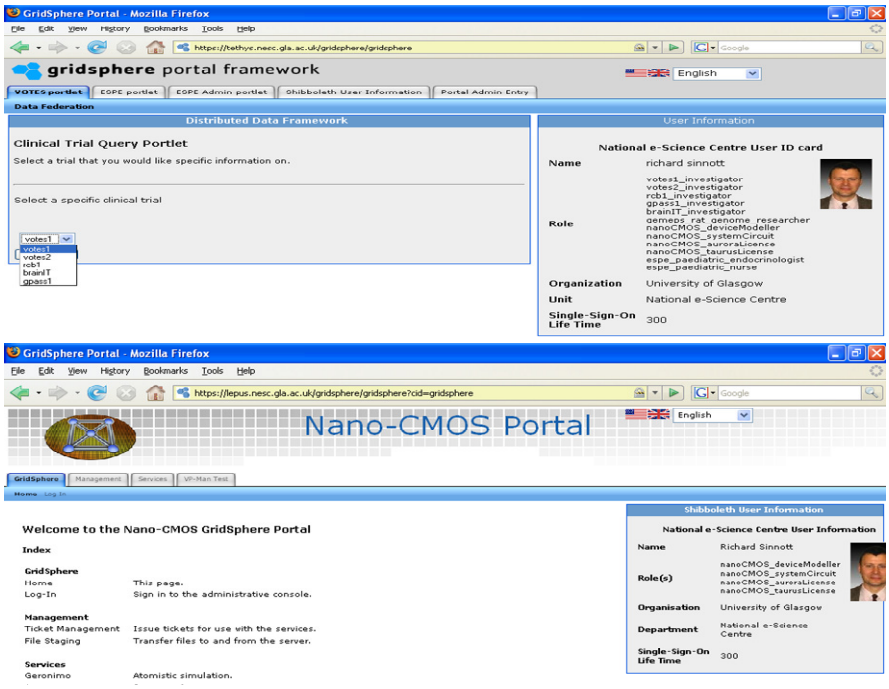

Figure 1. NanoCMOS Portal with Attribute Scoping (below) and Other Portal without Attribute Scoping (above)

This scoping allows the portal to be restricted to only accept attributes from known and trusted sources, e.g. the nanoCMOS partner sites or only from specific individuals at those sites. The attributes themselves are then used to restrict access to the associated services available within the portal. To support this, the CCP portlet is used by the portal/nanoCMOS administrator to enforce access control decisions on the interfaces to the services and data sets themselves (portlets). Thus an individual with a given set of roles or licenses should only be able to see the interfaces to those services that their privileges allow. To support this, the GridSphere portal framework has been extended with capabilities for user /virtual organization defined roles. These roles are then used to render the appropriate interfaces (portlets) to the user.

The ACP portlet provides a mechanism which addresses the need for site autonomy. It is highly unlikely that access control decisions to nanoCMOS resources will be left up to a potentially remote portal managed at Glasgow University. Instead, sites must make their own decisions on access to, and usage of, their own resources. To facilitate this, the ACP creates attribute certificates using the role information provided by Shibboleth and signs and stores them in an LDAP server associated with the portal. Remote services that wish to make their own local authorisation decisions can then be configured to use pull these attribute certificates in order to ensure that the user is authorized to access and use that remote service. The attributes themselves are signed by the virtual organization and agreed *a priori*, i.e. a remote service provider must provide the necessary information (roles required etc) for these attributes to be created on the fly by the ACP. Once defined, the use of these attributes in making authorization decisions is entirely transparent to the end user.

4.2 Integration of VOMS and PERMIS for Security-oriented Job Submission

In the development of the required Grid infrastructure the project has utilised many of the technologies provided by OMII-UK. The early phase of work has focused upon development of a family of OMII-UK services, which support the device modelling and compact model generation phases of electronics design. These services have been developed to exploit the OMII-UK GridSAM job submission system.

The aim of GridSAM is to provide a web service for submitting and monitoring jobs managed by a variety of Distributed Resource Managers (DRM). This web service interface allows jobs to be submitted from a client in a Job Submission Description Language (JSDL) document and supports their status retrieval as a chronological list of events detailing the state of the job. GridSAM translates the submission instruction into a set of resource-specific actions: file staging, launching and monitoring using DRM connectors for each stage. A variety of resource specific DRM connectors are available including connectors for Condor, Sun Grid Engine and Globus. The work is currently focused on supporting the Globus DRM connector for the *GRAMSubmissionStage* part of the DRM connector sequence. Here, authorisation is decided before the JSDL document is submitted to the GridSAM instance and converted to a Globus specific Resource Specification Language document and submitted to a GRAM manager. This is achieved through extraction of the VOMS attributes from the GridSAM invocation (themselves embedded in the JSDL document) and using these to authorise access to specific connectors.

In order to support VOMS integration within the nanoCMOS domain, a *nanoCMOS* virtual organisation was established in a VOMS server at the National e-Science Centre in Glasgow. In this, the key roles of *deviceModeller* and *circuitSimulator* were established. These roles have been used within vanilla VOMS scenarios to map end users within the nanoCMOS domain to appropriate pooled accounts and gids/uids for the nanoCMOS project on the ScotGrid (www.scotgrid.ac.uk) resource at Glasgow. Additionally, in order to demonstrate how these roles could be used to limit the resources to which a given individuals jobs could be submitted the architecture shown in Figure 2 has been adopted.

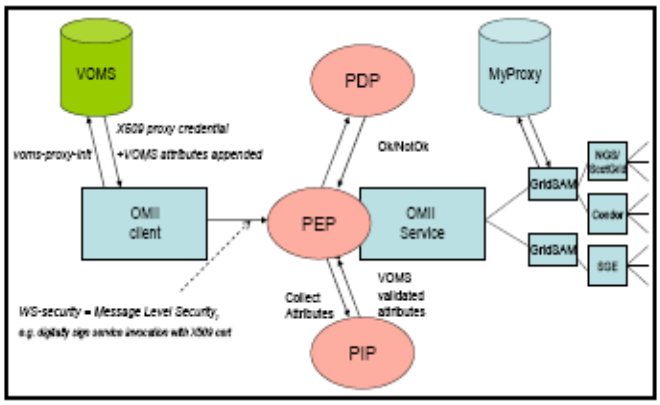

Figure 2: VOMS and PERMIS Integration Architecture

Here, the end user client creates an X509 proxy certificate with the appropriate VOMS ACs embedded and this information is then used to enforce the access control decision. The Acacia tool was used for this purpose. The PERMIS PEP then validates these attributes and passes them to the decision engine to ensure that the person with that role is allowed to submit that job to that particular resource. The system supports job submission to major clusters such as the UK e-Science National Grid Service, as well as local Sun Grid Engine clusters and Condor pools in Glasgow. In this scenario, it may well be the case that a particular role is not allowed to submit jobs to shared resources outside of Glasgow, e.g. if the data sets or simulations have IP-restrictions. In this situation the decision will be enforced that only local resources can be used for job submission. That is, the authorisation step at the GridSAM::DRM connector level will ensure that individuals can submit jobs to local resources and not to non-local resources such as the National Grid Service.

It is worth noting that the above scenario can be equally well supported using non-VOMS based attribute certificates from one or more attribute authorities, e.g. LDAP servers in Glasgow. We recognize however that a common set of attributes simplifies the roll out of the infrastructure and services. Where necessary the adoption of hybrid approaches to attribute authorities, i.e. use VOMS but if further information is required by a PEP/PDP then pull the other information from one or more trusted authorities. It should be noted that with recent Grid authorization standardization efforts, the above architecture can function equally well in a pull mode. For example, rather than the end user embedding the VOMS AC in their proxy certificate, it is equally possible for them to simply use an X509 proxy certificate and have the service pull the needed (VOMS) attributes required to make authorization decision. Within the nanoCMOS project we expect to work with both of these scenarios as required.

4.3 Integration of X509 Certificates and Kerberos

Access to AFS is through AFS tokens, typically obtained via Kerberos. However, many Grid based infrastructures such as ScotGrid and the UK National Grid Service exploit Globus middleware and its use of X509 based PKIs. As a consequence, such systems and their associated gatekeepers do not directly support such tokens. We note that most cluster based systems do not propagate tokens from cluster head nodes to worker nodes anyway. To work around this, we have adopted the *gssklog* application. *gssklog* is able to take an X509 based proxy certificate and authenticate with a gssklogd server. The *gssklogd* server is then able to return appropriate AFS tokens. To support this scenario on Globus-enabled resources requires that *gssklog* is called before the globus-gatekeeper invokes any jobmanager – otherwise it will not be possible to write to the appropriate AFS directories. Of course, this scenario also mandates that appropriate AFS cells are established and importantly for cluster based usage, that AFS clients are installed on

worker nodes. The use of gssklog compared to more typical Kerberos usage for creation of AFS tokens is shown in Figure 3.

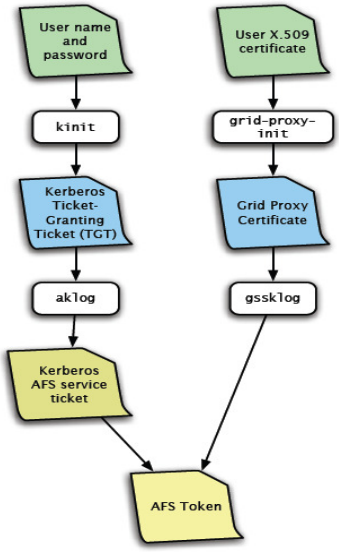

Figure 3: Usage of gssklog for AFS Token Creation

The typical scenario explored in nanoCMOS for integrating X509 certificates and Kerberos is as follows. The user creates an X509 proxy credential either directly on their client machine or through the Shibboleth enabled portal and the MyProxy service that is offered there. Using this proxy credential, the user is able to select input files and simulations that they wish to run on compute clusters such as the AFS-enabled ScotGrid resource. When these jobs are submitted, the associated gatekeeper invokes the *gssklog* application to obtain the appropriate AFS tokens and the job is submitted and data staged to the appropriate AFS directories upon job completion.

One of the benefits of this model of job submission is that since we are working with a global file system, the notion of file staging to and from the cluster is moot. That is, the virtual directories can be considered as local to the AFS enabled cluster even when potentially remote.

5. Conclusions

The challenges facing nanoCMOS electronics research demand that fine grained security is supported, in order to address the myriad IP constraints associated with the commercial domain. We recognize that no single security solution meets all nanoCMOS needs: X509 certificates have known limitations but are widely accepted as the way to authenticate for access and use Grid resources; VOMS attribute authorities are well recognized for defining the roles applicable across virtual organizations but have their own limitations; PERMIS provides fine grained service-level security but is not well matched to distributed file systems; Shibboleth provides simple user-oriented access to resources, whilst Kerberos underpins technologies such as AFS to provide secure access to federated file based data. This work is still very much in progress however many of the issues in integrating these multi-security solutions are being addressed both within the standards community and through early adopters such as the nanoCMOS project.

That said, we are acutely aware that the best security solution is often the simplest. The complexity of many of these solutions is a continuous source for concern both for us and importantly for establishing trust with our commercial partners. To address this we continue to embrace simple solutions for access control and management. Thus rather than attempting to deal with secure enclaves on shared public resources such as the National Grid Service, we are often taking more pragmatic solutions such as only allowing Intellectual Property-oriented jobs to execute on local resources. This has the advantage of overcoming security concerns but does of course mean that less compute facilities are often available. Such pragmatic considerations incorporating any associated risks are continuously being monitored and assessed in delivering the nanoCMOS infrastructure.

5.1 Acknowledgements

The authors would like to thank their collaborators in nanoCMOS. The nanoCMOS project is funded by a grant from EPSRC. The authors also acknowledge funding from Joint Information Systems Committee for VPman project which has helped deliver the VOMS-PERMIS implementation and from OMII-UK for the SPAM-GP project.

6. References

[1] G.E. Moore, *Cramming more components onto integrated circuits,* Electronics, Vol. 38, No. 8, April 19, 1965.

[2] International Technology Roadmap for Semiconductors (ITRS), http://www.itrs.net

[3] D. J. Frank and Y. Taur, *Design considerations for CMOS near the limits of scaling*, Solid-State Electronics, vol. 46, pp 315-320 (2002).

[4] K. Takeuchi, R. Koh and T. Mogami, *A study of the threshold voltage variation for ultra-small bulk and SOI CMOS*, IEEE Trans. Electron Dev, vol. 48, p. 1995 (2001).

[5] A. R. Brown, A. Asenov, J. R. Watling, *Intrinsic Fluctuations in Sub-10 nm Double-Gate MOSFETs Introduced by Discreteness of Charge and Matter,* IEEE Transaction on Nanotechnology, Vol. 1 pp. 195-200 (2002).

[6] C. Neuman, T. Yu, S. Hartman, K. Raeburn, *The Kerberos Network Authentication Service (V5).* RFC 4120 (Proposed Standard), July 2005. (Updated by RFCs 4537, 5021).

[7] R. Housley, W. Polk, W. Ford, D. Solo, *The Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile*. RFC 3280 (Proposed Standard), April 2002. (Updated by RFCs 4325, 4630).

[8] R. Alfieri, et al., *VOMS: an authorization system for virtual organizations*, 1st European Across Grids Conference, Santiago de Compostela, Spain, February 2003.

[9] D.W. Chadwick, A. Otenko, E.Ball, *Role-based Access Control with X.509 Attribute Certificates*, IEEE Internet Computing, March-April 2003.

[10] Internet2 Shibboleth Architecture and Protocols, http://shibboleth.internet2.edu

[11] ITU-T Recommendation X.509 (2001) | ISO/IEC 9594-8: 2001, Information technology – Open Systems Interconnection – Public-Key and Attribute Certificate Frameworks.

[12] R. Housley, T. Polk, Planning for PKI: Best Practices Guide for Deploying Public Key Infrastructures, Wiley Publishing, 2001.

[13] Globus toolkit, http://www.globus.org/toolkit

[14] R.O. Sinnott, J. Watt, D.W. Chadwick, J. Koetsier, O. Otenko, T.A. Nguyen, *Supporting Decentralized, Security focused Dynamic Virtual Organizations across the Grid*, 2nd IEEE International Conference on e-Science and Grid Computing, Amsterdam, December 2006.

[15] A.J. Stell, R.O. Sinnott, J. Watt, *Comparison of Advanced Authorisation Infrastructures for Grid Computing*, Proceedings of International Conference on High Performance Computing Systems and Applications, May 2005, Guelph, Canada.

[16] R.O. Sinnott, O. Ajayi, A.J. Stell, J. Watt, J. Jiang, *Single-Sign on and Authorization for Dynamic Virtual Organizations*, International Conference on Virtual Enterprises, (PRO-VE'06), Helsinki, June 2006.

[17] Assertions and Protocol for the OASIS Security Assertion Markup Language (SAML) v1.1, 2 September 2003, http://www.oasis-open.org/committees/security/.

[18] eduPerson Specification, http://www.educause.edu/eduperson/

[19] L. Han, A. Asenov, A. Berry, C. Millar, G. Roy, S. Roy, R. Sinnott, G. Stewart, *Towards a Grid-Enabled Simulation Framework for Nano-CMOS Electronics*, in Proceedings of the IEEE e-Science 2007 Conference, Bangalore, India, December 2007.

[20] R.O. Sinnott, T. Doherty, D. Martin, C. Millar, G. Stewart, J. Watt, *Supporting Security-oriented Collaborative nanoCMOS Electronics e-Research*, International Conference on Computational Science, Krakow, Poland, June 2008.

[21] R.O. Sinnott, A. Asenov, A. Brown, C. Millar, G. Roy, S. Roy, G. Stewart, *Grid Infrastructures for the Electronics Domain: Requirements and Early Prototypes from an EPSRC Pilot Project*, UK e-Science All Hands Meeting, Nottingham, UK, September 2007.

[22] J. Watt, R.O. Sinnott, J. Jiang, G. Stewart, A. Stell, D. Martin, T. Doherty, *Federated Authentication and Authorisation for e-Science*, in Proceedings of APAC 2007 conference, Perth, Australia, September 2007.

[23] JISC funded Shibboleth Access to Resources on the NGS (SARoNGS), www.jisc.ac.uk/whatwedo/programmes/programme_einfrastructure/**sarongs**.aspx

[24] J. Basney, M. Humphrey, V. Welch, *The MyProxy Online Credential Repository*, Software Practice and Experience, Volume 35, Issue 9, July 2005, pages 801-816.

[25] R.O. Sinnott, D. Chadwick, T. Doherty, D. Martin, A. Stell, G. Stewart, L. Su, J. Watt, *Advanced Security for Virtual Organizations: Exploring the Pros and Cons of Centralized vs Decentralized Security Models*, 8th IEEE International Symposium on Cluster Computing and the Grid (CCGrid 2008), May 2008, Lyon, France.

[26] L. Cornwall, J. Jensen, D. Kelsey, Á. Frohner, D. Kouril, F. Bonnassieux, S. Nicoud, K. Lörentey, J. Hahkala, M. Silander, R. Cecchini, V. Ciaschini, L. dell'Agnello, F. Spataro, D. O'Callaghan, O. Mulmo, G. L. Volpato, D. L. Groep, M. Steenbakkers, A. McNab, *Authentication and Authorization Mechanisms for Multi-Domain Grid Environments.* Journal of Grid Computing 2(4): 301-311 (2004).

[27] R.O. Sinnott, A.J. Stell, D.W. Chadwick, O. Otenko, *Experiences of Applying Advanced Grid Authorisation Infrastructures*, Proceedings of European Grid Conference (EGC), LNCS 3470, pages 265-275, Volume editors: P.M.A. Sloot, A.G. Hoekstra, T. Priol, A. Reinefeld, M. Bubak, June 2005, Amsterdam, Holland.

[28] R. Edward, R. Zayas. Andrew File System (AFSv3) Programmer's Reference: Architectural Overview, 1991.

[29] J.T. Kohl, B.C. Neuman, T.Y. T'so, *The Evolution of the Kerberos Authentication System, Distributed Open Systems*, pp78–94, IEEE Computer Society Press, 1994.