Multi-Institution Collaborative Computing: What Does it Really Take?

(http://www.esp.org/rjr/briite-RJR-salk-2005.pdf)

Robert J. Robbins rrobbins@fhcrc.org (206) 667 4778

Biomedical Research Institutions Information Technology Exchange

Multi-Institution Collaborative Computing: What Does it Really Take?

(http://www.esp.org/rjr/briite-RJR-salk-2005.pdf)

Robert J. Robbins rrobbins@fhcrc.org (206) 667 4778

Biomedical Research Institutions Information Technology Exchange

Multi-Institution Collaborative Computing:

What it REALLY takes is federated identity management, authentication, authorization, access-control, and auditing...

rrobbins@fhcrc.org (206) 667 4778

Biomedical research routinely spans institutional boundaries. Collaboration must be supported across these porous boundaries. Biomedical research routinely spans institutional boundaries. Collaboration must be supported across these porous boundaries.

Biomedical research requires secure computing environments. Good security requires a solid external perimeter.

Simple Fact



Challenge:

We must support IT interoperation and collaboration across multiple institutions while preserving and improving the security of our home institution.

perimeter.



Simple Fact

The computer infrastructure to support multi-site collaborative research must often implement security in a manner that cannot (and should not) depend upon the enterprise security of any one institution. Nor can it depend upon the security system of any one virtual enterprise. An ideal security system for this environment would be a totally decentralized, federated approach that provides open protocol-based components to allow the creation of and participation in numerous, independent virtual enterprises.

Topics

- Background
- Totally Decentralized Federation is Essential
- Federation is Different (& Hard)
- All Components, All the Time
- Making it work
 - Logical Simplicity
 - Social Scalability

Topics

- The Problem
- A (Possible) Solution:
 - GlAAAAS
 - GlAAAAS in Action
- Reality Check:
 - What's Really Possible
 - What Should be Done

Background

General Items

FHCRC

- Independent biomedical research organization
- 2500 employees
- Many relationships with other organizations
- Researchers collaborate outside our organization
- Much diversity within the organization
 - Four research divisions
 - Multiple research programs
 - 35 IT departments

- VP/IT at FHCRC
- PhD biologist
- Experience in community information infrastructure
 - NSF: first bio program officer for database activities
 - GDB: director, informatics core
 - DOE: genome program information infrastructure
 - caBIG: strategic & architectural planning

- BIASES
 - Database bigot
 - Even bigger TCP/IP bigot
 - Believer in decentralized components

- BIASES
 - Database bigot
 - Even bigger TCP/IP bigot
 - Believer in decentralized components
- Personal Observation
 - No big IT failure has ever occurred because of too much design and not enough execution





Bias for Action



© Scott Adams, Inc./Dist. by UFS, Inc.



• Analysis paralysis is bad

• So is rapidly heading off in the wrong direction

Background

Insights

An Assertion

- To be truly useful, an IT professional must have some knowledge of
 - Systems analysis
 - Database theory and design
 - Networking internals and design
 - Operating systems principles and design
 - Algorithms and programming

Systems Analysis Insights

- Understand your goals / Know your tradeoffs
- Understand your resources / Manage scope
- Plan for change
- Design for maintenance
- Read The Mythical Man Month

Systems Analysis Insights

- Understand your goals / Know your tradeoffs
- Understand your resources / Manage scope
- Plan for change
- Design for maintenance
- Read *The Mythical Man Month* More than once













Database Insights

- Build a good data model (schema)
 - Well normalized
 - Attributes must be properly attached
- Use single authoritative sources
 - Avoid duplicated data management
 - Data replication breeds data inconsistency
- Maintain database integrity

Networking Insights

- Efficiency ≠ Effectiveness
- No component is always guaranteed to work
- Change is inevitable
- Simultaneous upgrades are impossible
- No one is in charge
- It has to work anyway

Networking Insights II

- End-to-end protocol equivalence (interoperability) is required
- End-to-end technical equivalence is not
- End-to-end paths involve lots of negotiating and late binding of names, of technologies, and even of paths.

Algorithm Insights

- Simplicity is the goal
- Beware combinatoric explosions
- Understand algorithmic complexity
- Pay attention to scaling problems
- NP Complete ≠ impossible
 (But it's a good approximation)

Operating System Insights

- Think abstractly
- Use common abstractions
- A minimal kernel is a good kernel
- An insecure kernel \equiv an insecure system
- Modules modules modules

General Insights

- Old \neq Bad
- New ≠ Good
- Do not feel obliged to reinvent everything
- The 1972 reference monitor concept is still useful

Anderson, J. P. 1972. Computer Security Technology Planning Study. Technical Report ESDTR-73-51, Air Force Electronic Systems Division, Hanscom AFB, Bedford, MA. (Also available as Vol. I, DITCAD-758206. Vol. II DITCAD-772806). Available online at: http://seclab.cs.ucdavis.edu/projects/history/papers/ande72.pdf
Reference Monitor

- The reference monitor mechanism must be tamper proof.
- The reference monitor mechanism must always be invoked. That is, it must govern all operations and actions on the system, including the activities of the operating system itself.
- The reference monitor mechanism must be small enough to be subject to analysis and tests to assure that it is correct. That is, it must be capable of being proved to be correct.

Reference Monitor

• The reference monitor mechanism must be tamper

KEY POINT: The security mechanisms of the kernel must be small enough to function efficiently and simple enough to be UNDERSTOOD.

A "security system" that cannot be understood is not secure.

enough to be subject to analysis and tests to assure that it is correct. That is, it must be capable of being proved to be correct.

Beware Mindset Conflicts

DATABASE WORLD VIEW:

Every must be perfect, all of the time.

NETWORKING WORLD VIEW:

No component is ever guaranteed to work at any particular time.

Beware Mindset Conflicts

ENTERPRISE SECURITY:

We are designing technology to implement OUR security policies.

FEDERATED SECURITY:

We are designing technology to implement ANY security policies.

Background

Supersets and Subsets

• Downsizing a superset solution for a subset problem is usually easy.

- Downsizing a superset solution for a subset problem is usually easy.
- Upsizing a subset solution to a superset problem is often hard, sometimes impossible.

- Downsizing a superset solution for a subset problem is usually easy.
- Upsizing a subset solution to a superset problem is often hard, sometimes impossible.
- Upsizing is generally unwise; upsizing multiple times is beyond unwise.

- Downsizing a superset solution for a subset problem is usually easy.
- Upsizing a subset solution to a superset problem is often hard, sometimes impossible.
- Upsizing is generally unwise; upsizing multiple times is beyond unwise.
- Therefore, it's a good idea to know the ultimate size of your problem before going too far in the direction of a solution.

• Downsizing a superset solution for a subset problem is usually easy.

Compared with enterprise security, federated security is a superset problem.

Therefore, it's a good idea to know the ultimate size of your problem before going too far in the direction of a solution.

Federation Is Essential

Federation

Biomedical Research Occurs in a Distributed Manner

Federation

- Biomedical Research Occurs in a Distributed Manner
- Biomedical Research Demands Secure Information Infrastructure (criminal penalties apply when security is not met)

Federation

- Biomedical Research Occurs in a Distributed Manner
- Biomedical Research Demands Secure Information Infrastructure (criminal penalties apply when security is not met)
- Biomedical Research Needs a Federated Approach to Security and Access Control

Federation IS Different (& Hard)

Security and Access Control Systems are the means by which the people who are in charge enforce their decisions about about who should and who should not have access to the enterprise's computing systems.

- Security and Access Control Systems are the means by which the people who are in charge enforce their decisions about about who should and who should not have access to the enterprise's computing systems.
- In a truly federated environment, THERE IS NO CONTROLLING ENTERPRISE and NO ONE IS IN CHARGE OF EVERYTHING – there is no "privileged center" to the system.

Security and Access Control Systems are the means by which the people who are in charge enforce their decisions about about who should

A federated security model is NOT just a security model for a multi-site enterprise.

CONTROLLING ENTERPRISE and NO ONE IS IN CHARGE OF EVERYTHING – there is no "privileged center" to the system.

Q: If NO ONE IS IN CHARGE, then how do we build a security and access control system?

Q: If NO ONE IS IN CHARGE, then how do we build a security and access control system?

A: By developing a grid of components that can be used totally independently, but which can also be integrated in subsets to deliver virtual security and access control systems for virtual organizations that choose to use the components.





All Components

All the Time

All Components

In database design, one should always model the data at the finest used resolution. That is, if a use case requires that a data element be parsed into subcomponents, then the schema should decompose that data element into finer pieces.

All Components

- In database design, one should always model the data at the finest used resolution. That is, if a use case requires that a data element be parsed into subcomponents, then the schema should decompose that data element into finer pieces.
- Federated security components should be implemented at the finest used resolution. That is, if **any supported use case** requires that a service be delivered independently, then build that service as a stand-alone component.

Independent Components

- Identity Management
- Group Membership Management
- Role Definitions
- Authentication
- Authorization
- Auditing
- More...

Component Usage

In a truly federated environment, security and access control depend upon the availability of technically secure components that can be deployed in any way a user chooses (so long as the usage matches the technical specifications for the components).

Component Usage

- In a truly federated environment, security and access control depend upon the availability of technically secure components that can be deployed in any way a user chooses (so long as the usage matches the technical specifications for the components).
- Users must be free to use the components in as sophisticated (or as stupid) a manner as they choose.

Making it Work

Logical Simplicity

Logical Simplicity

- In a federated, component-based environment, the biggest challenge is managing complexity.
- This requires a commitment to simplicity.
- Components must be entirely self-contained.
- All inter-component communication occurs only through well defined protocols and interfaces.
- Systems must be designed to accommodate change.

Driving Assumption

• Many use case requirements across the federation will be inconsistent and some will be genuinely contradictory.

Driving Assumption

- Many use case requirements across the federation will be inconsistent and some will be genuinely contradictory.
- The federation must work anyway.

Making it Work

Social Scalability

Social Scalability

In a truly federated environment, long term success for a federated security model will depend upon social scalability.

Social Scalability

- In a truly federated environment, long term success for a federated security model will depend upon social scalability.
- Social scalability CANNOT be achieved through normative pronouncements.

Social Scalability

- In a truly federated environment, long term success for a federated security model will depend upon social scalability.
- Social scalability CANNOT be achieved through normative pronouncements.
- Experience suggests that social scalability is best achieved through a combination of pure laissez faire individualism and social consequences – i.e., social contracts.
Social Scalability

In a truly federated environment, long term success for a federated security model will depend

Negotiated social contracts – not mandated technical solutions – drive the emergence of standards in a federation.

achieved through a combination of pure faissez faire individualism and social consequences – i.e., social contracts.

• Every individual is free to do whatever he/she chooses.

- Every individual is free to do whatever he/she chooses.
- Every other individual is free to respond however he/she chooses.

- Every individual is free to do whatever he/she chooses.
- Every other individual is free to respond however he/she chooses.
- Interactive relationships then sort things out.

- Every individual is free to do whatever he/she chooses.
- Every other individual is free to respond however he/she chooses.
- Interactive relationships then sort things out.
- Examples:

One cuts, the other chooses.

- Every individual is free to do whatever he/she chooses.
- Every other individual is free to respond however he/she chooses.
- Interactive relationships then sort things out.
- Examples:

I am free to suppress my caller ID; if I do, you are free to refuse to answer my calls.

- Every individual is free to do whatever he/she chooses.
- Every other individual is free to respond however he/she chooses.
- Interactive relationships then sort things out.
- Examples:

You are free to run your systems in as stupid and insecure manner as you choose; if you do, I am free to refuse to have anything to do with your systems.

• Rules governing behavior can be permissions or prohibitions.

- Rules governing behavior can be permissions or prohibitions.
- The union set of contradictory permissions is a very flexible environment.

- Rules governing behavior can be permissions or prohibitions.
- The union set of contradictory permissions is a very flexible environment.
- The union set of contradictory prohibitions is the null set.

- Rules governing behavior can be permissions or prohibitions.
- The union set of contradictory permissions is a very flexible environment.
- The union set of contradictory prohibitions is the null set.
- Use case requirements across a federation will be contradictory.

If a federated security system is to deliver services greater than the null set, it must be technically implemented on the aggregation of permissions, not prohibitions.

Behavioral constraints should be achieved on a virtual organization basis, through negotiated social contracts.

ontradictory.



uautu

For example, the components of a federated security system must permit users to behave in a highly secure manner, but it should not mandate that users do so.

Negotiated social contracts will then determine who interacts with whom, with what security levels. e

Social Scalability:

Required Reading

Alexander Hamilton — James Madison — John Jay

The Federalist Papers



Social Scalability:

Required Reading

Alexander Hamilton — James Madison — John Jay

The Federalist Papers

There is no better source of ideas on how to build systems that work in a decentralized social environment.

Remember, you can't change human nature, so you must design systems that work **despite** human nature.

The Problem

Problem Components

- Systems Analysis
- Database
- Networking
- Operating System
- Algorithms

Systems Analysis Challenges

- What am I trying to do?
- What are my constraints?
 - Technical
 - Social
- How to allow for change?
 - Growth in size and complexity
 - Conceptual extensibility
 - Technological Advance and Upgrades

Database Challenges

- What information do I need to manage?
- What's my best schema?
- How should the schema be partitioned?
 - Vertically?
 - Horizontally?
 - Both?
- What can I do to ensure data quality?

Networking Challenges

- How to handle name resolution and resource discovery?
- What are the sources of dynamism?
 - System failure
 - Local changes
 - Technical advance
- How to allow for dynamism?
 - Protocol negotiation
 - Late binding

Networking Challenges

- Dealing with distributed information:
 - Authoritative sources
 - Proxy servers
 - Local caches
 - Time-to-live parameters

Operating System Challenges

- How to hook the permissions to the OS?
- How to achieve variable resolution OS support?
 - Login
 - Application access
 - Application operation
- How to achieve OS-level logs and audits?
- How to integrate with OS, but still achieve OS portability for code?

Algorithmic Challenges

- Understanding the complexity of the algorithms
- Achieving acceptable performance
 - Good algorithms
 - QoS monitoring / graceful failure

Possible Solution

Basic Issus

GIAAAS as a Solution

- A truly useful, totally decentralized federated access-control system would be a
 - Global
 - Authentication
 - Authorization
 - Access Control
 - Auditing
 - System

• That is: GlAAAAS

TCP / IP as a Model

- TCP/IP protocols are content agnostic
- TCP/IP protocols can move ANY "file"
- This is good: do not need separate networks for different data types
- This is bad: digital vermin move just as effectively as digital content
- Ultimately: local acceptable-use policies determine what is permitted

Black Box Models

- A black box model is a formal documentation of a program's function, in terms of inputs and outputs, with no concern for the program's internal technical implementation. A black box description records WHAT a program will do.
- Accompanying the black box description is a "clear box" analysis that documents the technical internal technical implementation. A clear box description records HOW a program will do what it does.

Black Box Models

- A black box model of GIAAAAS would be a formal description that documents the permission decisions to be made, the rules governing the decisions, the information necessary to make the decisions, and the source(s) of the required information.
- A sociological black box description describes WHO will make the rules, WHO will provide the relevant information, and WHO will decide if the information sources are reliable. And, WHO is accountable form security failure.

Clear Box Models

A clear box model of GIAAAAS would be a formal description that documents the technical implementation of GIAAAAS. HOW is authentication to be accomplished, HOW is information to be transmitted securely, HOW to deal with proxied requests, HOW to hook into the OS, HOW ...

Possible Solution

GIAAAAS

GIAAAAS



When users access computer resources they must be assigned specific permissions in order to carry out useful tasks.



To simplify the management of the user-by-permissions matrix we can aggregate sets of permissions (necessary to accomplish some coherent set of tasks) and bundle them as **ROLES**.



Similarly, we can also identify sets of users (with similar attributes) and bundle them as **GROUPS**.

Maximum efficiency is gained when we manage the permission matrix by assigning **ROLE** permissions to **GROUPS** of users.



Within an enterprise, however, this can sometimes appear redundant.

That is, people are often placed into GROUPS based on their job function and permission ROLES are often created to allow people in a particular job function to accomplish their tasks.

This problem is exacerbated by the linguistic problem that within an enterprise people are often placed into a GROUP based on their job function, or "role" within the enterprise.


This apparent redundancy has led some to argue that the GROUP and ROLE concept should be combined.

Such a combination violates the logical distinction between the concepts and is akin to denormalizing a database schema in order to improve performance.



Within a totally decentralized federation, however, the notions of groups and roles MUST BE KEPT DISTINCT and MUST BE IMPLEMENTED SEPARATELY.

Efforts to build a federated identity management system that does not maintain this separation will be inadequate and will ultimately result in significant disappointment.

- The ideas of "groups" and "roles" have both been used to describe how collections of permissions can be assigned to collections of users.
- The concepts have been only partially distinct and some authors have used them almost interchangeably. Others have argued that one concept is preferred and the other should be deprecated.
- In the context of a TDCF, however, the concepts are quite distinct and both are needed.



- Groups are collections of people
- Criteria for membership in a group is strictly up to the manager of that group (e.g., could be "officers of company X" or "physicians with attending privileges at hospital Z" or "people whose birthday is a prime number")
- Management of group membership can be done informally or formally (i.e., with an audit trail)

- Roles are aggregations of permitted actions that a user may take on a computer resource (e.g., the role of standard user or superuser or DBA)
- Roles are associated with computer resources (e.g., the role of standard user on computer X)
- The manager of a resource determines what roles are to be made available on the resource.





Authorization Joins Groups & Roles





NOTE: In an enterprise-free federation, it is not possible (indeed, it is inconceivable) that group membership in any particular group could always control the permission to act in a particular role on an arbitrary resource. Therefore, in a federation IT IS **ESSENTIAL THAT A CLEAR LOGICAL AND TECHNICAL DISTINCTION BE MADE BETWEEN THE CONCEPTS OF GROUPS AND ROLES.**

- There is NO central enterprise.
- Everything is (potentially) decentralized:
 - Identity Management
 - Group Membership
 - Authentication
 - Authorization
 - Auditing
- Participation is voluntary
- Solutions must scale, technically and socially.

Components of the Problem

The first step in approaching a solution is examining the fundamental questions: (a) who will manage the various components of security-relevant information, (b) what information must be available and communicated to accomplish appropriate access control, (c) when updates to the information will occur, (d) where the various components of security-relevant information will be managed, (e) why trust should be extended to security information managed by a different organization, and (f) how the necessary communications are themselves accomplished in a secure manner.

Remember the Reference Monitor Concept

The access-control system must be simple enough that it can be understood.

A permission system based on arbitrary logic over arbitrary attributes served up from (marginally controlled) vocabularies is not likely to be simple.

Remember the Reference Monitor Concept

But, simple logic and set theory says that any attribute (or set of attributes) for any object can be converted into a statement about set (or group) membership.

A simple Boolean statement over one data type (group membership) that must evaluate to true or false (or unknown) can be understandable.

Permissions based on (and only on) group membership can trivially be extended to include negative permission.

Data Requirements

Globally unique identifiers for

People

Groups

Rule Sets

Data Source Requirements

Conceptually, the data sources for all of these must be partitioned both vertically and horizontally.

Note that it is **inconceivable** that all relevant AND RELIABLE group membership information in a TDF can come from the same source that provides identity management and authentication.

Once you have the need to support more than one source of information you have a need to support n sources.

IDENTITY:

the reliable association of a particular digital identifier with a particular human being

AUTHENTICATION:

the process by which one determines that a particular use of a digital identifier is being invoked by (or on behalf) of the person it names

PERMITTED ACTION:

an individual activity that may be performed on a particular resource.

PROHIBITED ACTION:

an activity that may be explicitly prohibited on a particular resource

Definitions

ROLE:

aggregations of permitted and prohibited actions on a particular information resource

PERMITTED ACTOR:

a person (or proxy) who can be granted permission to act in one or more roles on a particular information resource.

GROUPS:

aggregations of people (actors)

GROUP MANAGEMENT SYSTEM:

a system for managing group-membership assignments. To be federation-ready, a groupmanagement system (GMS) should be capable of managing group memberships for individuals whose identities are managed elsewhere in the federation.

Definitions

AUTHORIZATION:

the granting of permission to members of a group to act in a particular role on a particular resource. Note that authorization itself should be authorized, and an audit trail of authorizations should always be available.

PROHIBITION:

the forbidding of members of a group to act in a particular role on a particular resource.

Definitions

PERMISSION MANAGEMENT SYSTEM:

a system that manages information to facilitate the unambiguous assignment of permission (or prohibition) for members of group "G" to act in role "R" on system "S". To be federation ready, a permission-management system should be capable of maintaining rules, the components of which are all defined elsewhere in the federation.

GIAAAAS In Action



Institution A maintains a database resource associated with a multisite clinical trial head-quartered elsewhere. Access to the database is tightly controlled according to rules based on groups to which individual requesting access belongs.







Institution B	
†	Institution A
	The database resource asks Institution B, WHAT INFORMATION DO I NEED TO COLLECT TO AUTHENTICATE DR JONES?.







DBA. The resource asks Dr. Jones to specify the role he wishes to use.



Jones responds: RESEARCH FACULTY. The database resource knows that the group-membership rule sets governing access to the clinical-trial resources are maintained at Institution C. The database resource queries the rule-server at C to obtain the latest rule set.













Jones is a member of the permitted group and he is not a member of the prohibited group. Therefore, he is authorized to access the database in the role of RESEARCH FACULTY. To decide whether or not to allow Jones in, the database resource used information maintained at four other, independent organizations. The decision to use these other resources was a local decision.



According to the auditing rules governing the database, Jones' request to access the database, his authorization to access the database, and all of his activities while accessing the database are logged in a logging system maintained at Institution Z. Now five other institutions have been involved in permitting and tracking Jones' use of the database resource.













Although multiple resources were involved in the access-control process, the logical was simple: (1) determine who is requesting access, (2) determine the roles and rule sets governing access, (3) determine the user's membership in the relevant groups, (4) decide to grant or prohibit permission based on a simple Boolean evaluation over a rule set, and (5) log all activities.

Reality Check

What's Really Possible What Should Be Done

Reality Check Left as an exercise for the audience... vvnat s keany USSIDIE What Should Be Done

END