
EIQ SERVER SECURITY AND PRIVACY ACCESS PROFILES AND INDEXES

REVISION 2.6

Introduction

There are four major aspects of a comprehensive security/privacy model according to Verisign (paraphrased):

- Authentication – verification that the user is who they say they are
- Security/Privacy – ensuring that data is not misused, disclosed to unauthorized people, and personal identities are protected as far as possible
- Integrity – protection against data being improperly modified or duplicated
- Accountability – enabling an irrefutable means of tracking operations

The following text discusses each of the above four aspects and assumes that there are accepted data integration/information sharing/interoperability (referred to hereafter as “information sharing”) standard data models and standard field names used to request data and for query modification terms. Data sources (and eventually applications) do not need to have or use these standard data models and standard field names; however, they are used for information sharing, and data source (and eventually application) data models and field names are mapped to the standard data models and standard field names.

In this document, a security access system is proposed for a multi-organization information sharing system is proposed that controls access to data through a combination of:

- Virtual schemas that are subsets of standard data models and standard field names
- Query constraints that are use virtual schemas
- Indexes that can impose specific security and privacy filters at the row, column and data element levels; controlling access to specific data source records, fields and data

It is further proposed that, with the exception of security and privacy indexes, the security access system could be used for almost any multi-organization information sharing system and not just for WhamTech’s.

General Comments on Secure VPNs

We know from our talks with agencies that there are at least four separate VPNs:

- Unclassified
- Classified
- Secret
- Top Secret

We know that there are security access levels within these VPNs.

We know that users on higher security VPNs can access lower security VPNs, but not vice versa, and usually that access is through separate physical networks. There are usually “air gaps” between VPNs.

General Comments on EIQ Server Operations

Without delving into detail that is available in other documents, EIQ Server communicates with other EIQ Servers usually in a hierarchical structure, using peer-to-peer communication, through sockets (which can be used with Secure Socket Layer (SSL) products), as well as RPC and Java RMI. There is no structural constraints on how EIQ Servers communicate, as they can be independently configured, which could lead to more network-like structures than hierarchical or other structures. EIQ Servers can also be accessed directly through Web Services.

Using a Web Services paradigm, any metadata associated with a query is passed along with the query, and any metadata associated with results is passed back with the results. Currently, a proprietary format is used, but WhamTech is tending towards a Web Services model where XML is used as the file format. It is envisaged that metadata associated with the query organization, user (including ID certificates), role (if separate from user), and application, along with query processing rules, result sort and merge rules, etc., will be passed to and back from EIQ Servers. All key parameters will have unique IDs to facilitate and optimize query processing, results merging, and audit.

General Comments on Security/Privacy Issues

In general, privacy issues are perhaps more difficult to deal with than security issues; where personal data and information for people on US soil (citizens, permanent residents, visa-holders, and illegal) need to be protected.

Also, in general, there is overlap between security and privacy issues. If the appropriate security is in place, privacy is less difficult to deal with. A case in point is, if genuine role-based security is in place and operates at all levels (query organization, user, (role), application, data source organization, and data source), AND audit logs are archived, audited and analyzed, there should be accountability and significantly less abuse of all kinds, including privacy, than if carte blanche data access is allowed. Role-based Access Control (RBAC) allows data and information to be used for specific purposes – no more. Data and information becomes available only on an “as needed” basis. Ideally, any data and information that is retrieved should somehow self-delete or expire after it has been used for a specific, designated purpose, but this is beyond the scope of this document.

The above said, it is important to address specific privacy issues, and this document addresses these later.

Identity Authentication

Typically, current identity authentication requires a system-level logon and maybe application-level logon. These logons usually involve a user name and password, which tends to be an accepted low-level security solution WITHIN organizations, but is not well accepted BETWEEN different organizations or parts of the same organization. There are various ways to approach this issue, most of which involve some form of digital certificate, such as Public Key Infrastructure (PKI) for SSL; however, at some point, the information sharing system middleware (in WhamTech’s case, EIQ Server), the application that executes the query (in WhamTech’s case, EIQ Server), or both, will have to authenticate the identity of the person (and perhaps the hardware, organization and application) submitting the query. ID authentication could be passed through (a) normal communications with HTTPS and/or SSL, (b) secure Web Services protocol, e.g., WS-Trust, WS-Security or Security Assertion Markup Language (SAML), (c) OASIS XRI/XDI, or (d) some other means. WhamTech believes that any of these methods would fit EIQ Server-based operations well.

It is also envisioned that there will be universally accepted ID authentication systems, security clearances and access controls in the long-term, but in the short to medium-term; any system will have to potentially recognize and mediate between multiple, different ID authentication systems, security clearances and access controls in different organizations and maybe even within the same organizations.

Security and Privacy Access Profiles (SPAPs)

EQ Server uses SPAPs, which are mainly virtual schema, to accommodate various access privileges with an RBAC approach, depending on the:

- Organization submitting the query
- Security level of the user within a particular organization submitting the query
- Role the user has either within the organization or for the application
- Application being used to submit the query
- Combination of data requested
- Ultimate decision of the data source organization
- Data source providing results

The above results in at least six, maybe seven SPAPs:

1. Query Organization Access Profile (QOAP) – specifies the virtual schema that can be used to access data by a particular organization or part of an organization – QOAP is modified by subsequent SPAPs.
2. User Access Profile (UAP) – specifies the virtual schema that can be used to access data by a particular user or specifies the Role Access Profile (RAP) a particular user has (see SPAP 4.), which is usually a more efficient way to manage user access controls.
3. Content Access Profile (CAP) – specifies query constraints for almost any SPAP, e.g., an organization, user (role) or application is only allowed to access data in their state, at recent data, or internal personal data, but not external personal data. CAPs could be part of almost any and all profiles. Multiple CAPs would normally be additive.
4. Role Access Profile (RAP) – specifies the virtual schema that can be used to access data according to user-roles within an organization. A less efficient option is to define role-based access in an UAP, as an UAP may be defined depending on a user's role – a single user with multiple roles may have multiple user logons and multiple UAPs. Another option may be to define role-based access in an Application Access Profile (AAP) (see next profile), where a user could be assigned a role at the application level avoiding the need for a RAP – a single application may have multiple role-based AAPs. In the case that the application manages user logon, another option is for the AAP to specify the RAP to be used instead of using role-based AAPs or in the UAP specifying the RAP.
5. Application Access Profile (AAP) – specifies the virtual schema that can be used to access data by a particular application
6. Data Source Organization Access Profile (DSOAP) – specifies the virtual schema that can be used to access data sources within an organization and that is available to the query organization – this relates directly to the agreement/relationship between the query organization and the data source organization. There may be different DSOAPs depending on user security levels. DSOAPs may have to be extended to accommodate

inclusion/exclusion of entire data sources depending on the query organization, user security level, and/or application; this would be accommodated

7. Data Source Access Profile (DSAP) – specifies the virtual schema that can be used to access data from a particular data source.

Of the above SPAPs, 3. CAP is different, as it is implemented through query modification terms rather than Boolean inclusion of virtual schema.

Ultimately, a Boolean addition of SPAPs is executed and only if all SPAPs allow a particular field/column/data element in the virtual schema would it be able to be queried and returned as part of a result-set.

The Send SPAP is the virtual schema that is sent from a query organization to the information sharing system:

- QOAP AND UAP (or RAP) AND AAP (or RAP) = Send SPAP

If an application server, Web-based application, or Web Service is used, then the Send SPAP would not include the AAP and this would be applied in the middleware.

In addition, the query sent with the Send SPAP could have CAP constraints added.

At each data source organization and for each data source belonging to that organization, the Execute SPAP is the virtual schema that is ultimately used to define the results that are returned to a user from a specific data source in a data source organization:

- Send SPAP AND DSOAP AND DSAP = Execute SPAP

Results from multiple data sources are merged and other rules applied before a composite result-set is returned to the user/application that submitted the query.

WhamTech created a relational database schema to accommodate the SPAPs associated with multi-organization information sharing.

System Approaches

There are various approaches for managing SPAPs, all using an independent SPAP management system within the information sharing system, including:

- **Entirely Independent System** -- Make the associated changes in organization, user, application, and data source access to and from the information sharing system. This approach does not make, or makes very little, use of existing systems and would probably result in a lot of work, as it would have to include new ID authentication, user management, application rewrite/new development, enable external data source access, etc.
- **Entirely Interdependent System** -- Map existing organization, user, application, and data source access to and from the information sharing system. The information sharing system would be responsible for mediating between organizations for this access; becoming a trusted infomediary. This approach makes use of existing systems and although more complex in establishing trust and information sharing protocols, would probably be more seamless and therefore more acceptable solution in the long-term. Emerging standards such as OASIS XRI/XDI and ISO 11179 will eventually help in this effort.

- **Hybrid Independent/Interdependent System** -- In the short to medium-term, a combination of making changes and mapping existing systems could be the best balance among practically getting an information sharing system running, establishing trust, and tending towards an entirely interdependent system.

WhamTech Development Status

WhamTech is developing the first phase: Boolean inclusion of SPAPs server based on registered query organization, registered user, registered role, registered application, registered data source organization, and registered data source. Each organization would be responsible for its own SPAPs whether that be on a central or multiple distributed servers. One thought was to create a Domain Name Server (DNS) approach to allow multiple distributed servers to be used in a diverse information sharing system.

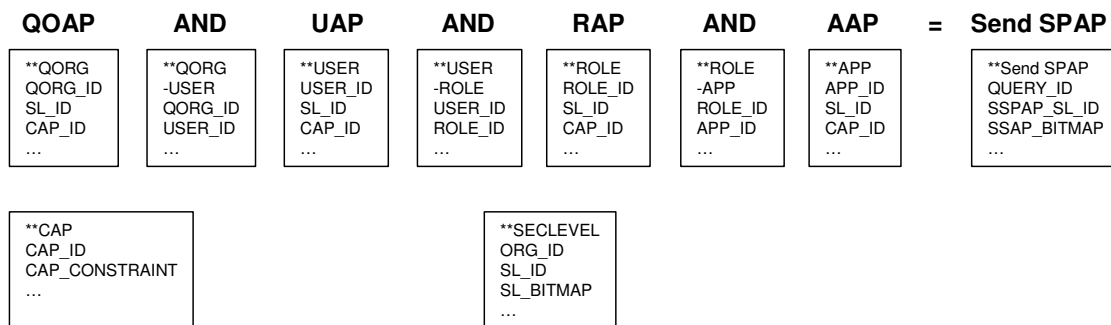
This relational database schema can also accommodate combinations of data from a query request point-of-view, by excluding virtual schema fields in the SPAPs, e.g., if someone asks to include name and SSN in the result-set, but they are not entitled to see both pieces of information together, the name or the SSN can be requested, but not both together. More complex rules on combinations of data will have to be developed and accommodated; however, the current approach is a good basis.

The relational database schema can also be extended to accommodate the next phase where query terms can be modified for security/privacy reasons, for example:

A user can search for approved metadata standard fields, but only in the state of Texas – an extra query term would have to be added to the query as follows: “where STATE = “TX””; or only on US Citizens: “where NATIONALITY = “US””.

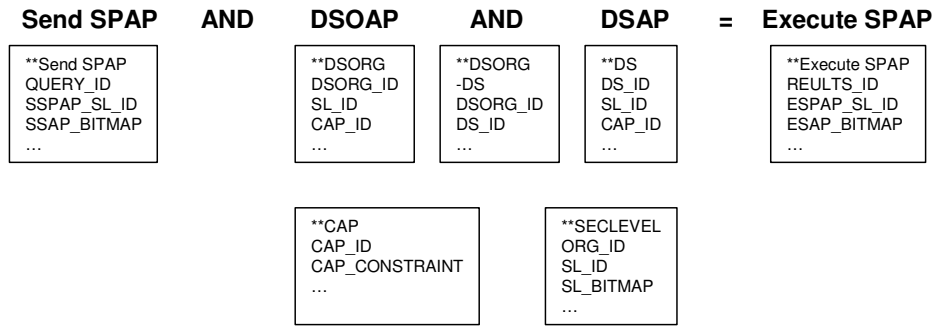
SPAPs Relational Database Schema

Please note that the following diagrams do not follow standard ER relational database notation; ER connections are excluded for clarity. These diagrams are illustrative only and are different from actual implementation. Note that query organizations could be the same as data source organizations and that the associated tables and schema can co-exist.



XML QUERY PACKAGE TO INCLUDE: Send SPAP, QORG_ID, USER_ID, ROLE_ID, APP_ID, CAP_ID1, CAP_ID2, ...

Figure 1: Relational Database Schema for SPAPs in a Query Organization



XML RESULTS PACKAGE TO INCLUDE: QUERY_ID, RESULTS_ID, DSORG_ID, CAP_ID1, CAP_ID2, DS_ID, CAP_ID3, CAP_ID4 ...

Figure 2: Relational Database Schema for SPAPs in a Data Source Organization

Privacy Issues

There are various levels related to privacy issues, including:

- As discussed in the earlier section, “General Comments on Security/Privacy Issues”, restricting data access through RBAC goes a long way to solve the situation where a user attempts to access data that is not needed for the purpose intended, i.e., data should only be accessed on an “as needed” basis
- It may be sufficient to inform a user that results are available, but not to provide the results until some higher-level permission is obtained. With EIQ Server, an interim stage is available where the user can be provided with the number of records isolated in EIQ Indexes with a query, without any direct contact with a data source and without retrieving source data
- When EIQ Indexes are built and maintained, sensitive data can be either encrypted or replaced with alias tokens. Likewise, once isolated, retrieved sensitive result-set data can be either encrypted or aliased. Currently, WhamTech uses a fast one-way 64-bit encryption system; other higher encryption levels and two-way encryption can be used.

Row, Column and Data Element Security and Privacy Indexes

The SPAPs focus on controlling data access up to the point of query execution on data sources; however, there may be a requirement to impose access controls on specific rows/records, columns/fields, and/or specific data elements within a data source. For example, a data source may contain a combination of unclassified and classified data, or specific columns have security designations, or are assigned security designations or considered private by a system owner. These types of constraints can be accommodated by EIQ Server indexes where as indexes are being built and updated, specific data could be read and using data transformation rules, used to assign security and privacy classifications to rows, columns and/or data elements. Other rules based on date could be used to assign security and privacy designations, with expiry dates. There is no limit as to how security and privacy indexes could be created and used.

Integrity Issues

Data sources themselves will/should have their own integrity assurance systems; however, there remains the question of how assurance is provided for the integrity of:

1. EIQ Server queries, including metadata
2. EIQ Indexes
3. Result-sets, including metadata

Implementing SSL or some other secure system could solve most integrity issues associated with 1 and 3; however, WhamTech can offer another option:

WhamTech owns an extremely fast one-way 64-bit CRC algorithm, which has a collision rate of 1 in 18 billion (extremely high for a CRC algorithm). Any kind of ASCII text can be “shingled” – sampled in a number of places and used as a checksum; therefore queries and results, including metadata, can be shingled and a checksum run to ensure what was sent was also received. If not, a query or result is rejected and a request is issued for a resend.

Placing EIQ Indexes in a secure environment should minimize the chance of their integrity being compromised; however, the above shingling technique can also be applied to indexes. EIQ Indexes can be routinely/constantly shingled and a checksum made against latest shingled values when a query is made to assure integrity.

Accountability Issues

WhamTech is a strong proponent of a RBAC for a federated data and information integration and sharing system such as EIQ Server. This would address many of the privacy issues, forcing an “as needed” basis for data and information requests, and avoiding some of the security issues raised by the ex-FBI spy, Robert Hanson, who had clearance to see the data and information he had access to, but there was no match between access and application, and little accountability. The other aspect of the Robert Hanson case was that he accumulated data and information over time that he had legitimate access to. As mentioned in the General Comments on Security/Privacy Issues section earlier, any data and information that is retrieved should somehow self-delete or expire after it has been used for a specific, designated purpose, but this is beyond the scope of this document.

EIQ Servers can record every single query operation, including metadata about the query and results provided, in an audit log. Actual result sets could also be stored, although this introduces other security/privacy issues.

Routine analysis could be run against, or intelligent software agents could monitor, audit logs to catch potential misuse very shortly after it occurs. Intelligent audit agents running on EIQ Server itself could also monitor and prevent misuse before it occurs.

Summary

The EIQ Server platform provides a universal and uniform means of distributing any security/privacy controls along with queries and result-sets, and is flexible enough to accommodate almost any standards, including future Web Services standards.

EIQ Server addresses the four aspects of a comprehensive security/privacy system effectively:

Authentication

EIQ Servers currently communicate with other EIQ Servers through sockets, which allows for currently widely accepted SSL secure communications, plus, identity authentication systems such as enterprise PKI can already be used with EIQ Server. Future options for identity authentication, such as SOAP/XML standards, can be easily incorporated.

Security/Privacy

Controlling access at all levels in a federated data and information integration and sharing system IS a complex matter. EIQ Server offers SPAPs that can accommodate most, if not all, of the aspects through a comprehensive relational database schema that is flexible enough to accommodate future additions, such as biometric data. It is believed that SPAPs can be used by almost any information sharing system and does not need to be tied to EIQ Server systems. Whereas SPAPs control access to the point of query execution on data sources, in addition, EIQ Server systems can offer row, column and data element indexes to control access from data source systems themselves.

Integrity

A secure system goes a long way to protecting the integrity of system traffic and files located on the system. WhamTech can also offer additional integrity protection through a very fast 64-bit CRC algorithm that can sample or “shingle” queries, results and EIQ Indexes, to assure integrity.

Accountability

Audit logs that are actively monitored and analyzed are key to assuring accountability. Also, intelligent agents running on the audit logs or EIQ Servers can potentially catch misuse shortly after it happens or prevent it in the first place.

For more information on the above, please contact:

Gavin Robertson, CTO, WhamTech, Inc.

972-380-4645 x223

gavin.robertson@whamtech.com