Functional Specification of the OSS Electronic File Room

James A. Rome
ORNL

May 4, 1996

(Revised July 14, 1996)

Table of Contents

Functional Specification of the OSS Electronic File Room

James A. Rome
ORNL
May 4, 1996

(Revised July 14, 1996)

STATEMENT OF PURPOSE

The Electronic File Room (EFR) shall provide a secure, searchable, and user-friendly electronic repository for the storage and retrieval of documents at many different classification levels.

PHASE I FUNCTIONAL DESCRIPTION

1. Overview

The Department of Energy Office of Safeguards and Security (OSS) File Room (IMC) currently has about 700 linear feet of documents in storage, with an additional 25% in long-term storage. These documents range from Unclassified to Top Secret, and are subject to handling caveats and need-to-know restrictions on access. The Department is legally required to preserve these documents for varying periods of time.

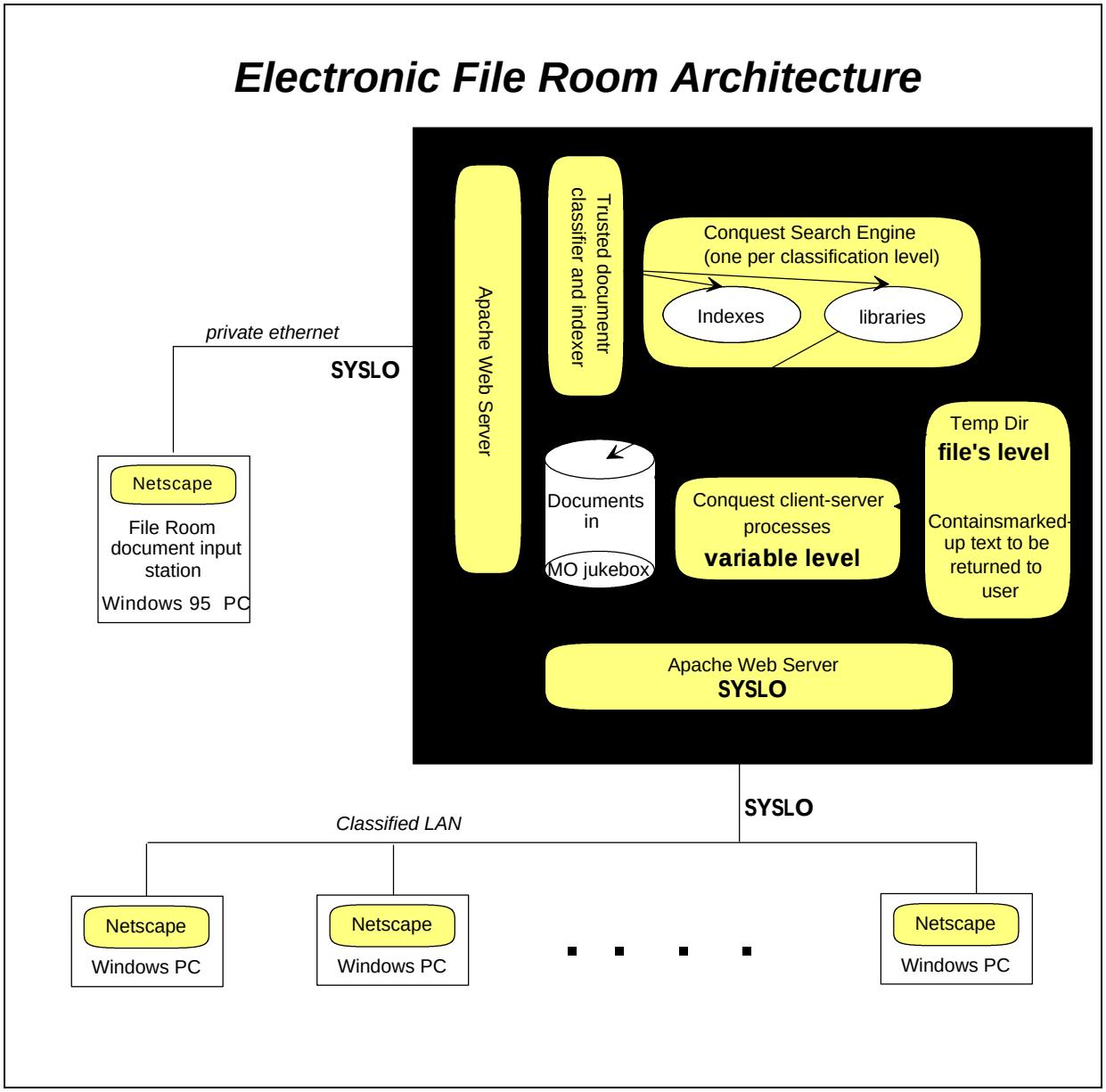
There is a great need to be able to search these documents in an efficient and timely manner which implies that they need to be converted into an electronic format. Legally, this format must preserve the “look” of the original document. In addition to searches over the document text, it is also desirable to search over keywords such as author, subject, date, etc.

Initially the EFR shall be connected to a classified LAN on which all users have appropriate clearances, but not necessarily all need-to-know categories. Eventually, it would also be desirable to allow access to this database from the unclassified LAN.

Therefore, the challenge is to provide a secure but user-friendly method of accessing, searching, and retrieving documents from the EFR. For implementing these requirements, the architecture chosen employs the ubiquitous and familiar World Wide Web technology in an “intranet” environment coupled with a Compartmented Mode Workstation (CMW) to enforce the multilevel security. Commercial off-the-shelf software (COTS) is used whenever possible to reduce the amount of custom programming required.

The name given to the EFR is “Ask-OSS” (pronounced “ask us”).

At the end of Phase I, this system shall be functional and installed in the OSS IMC ready for real-life use testing.



2. Architecture

The EFR consists of a text-retrieval database, Conquest (hence abbreviated as CQ), that is fronted by a Web server on a CMW computer system. Users connect to this system using the Netscape Web Browser over a private “intranet,” namely the classified LAN.

3. Platform

Because this system is the first of its kind, it is difficult to predict its performance. The addition of security always results in a performance degradation. Accordingly, we chose one of the most powerful workstations available as the CMW operating system platform —

the Hewlett-Packard (HP) J-200. This workstation has the capability of being upgraded to a two-processor configuration if more processing power is ever required. HP's implementation of the CMW operating system was developed by SecureWare, which has now been purchased by HP. The CMW system enforces mandatory access controls (MAC) to apply classification dominance tests to all system transactions.

The J-200 houses the Web server and CQ text search servers. A separate and distinct Web server process is launched for each Web transaction, a transaction occurring each time a new Web page is retrieved or generated for the browser on a user's machine. Thus, each user is connected to a different server process each time he/she selects a new item on a Web page. This paradigm avoids any contamination of the authentication process that can occur when multiple Web server processes are running.

3.1 Hardware specification

3.1.1 DOE-OSS system

The J-200 components consist of

- The CPU with 128 Mb of memory, an internal 2 GB hard disk, and two ethernet cards.
- high resolution color monitor
- CD-ROM reader
- Floppy disk
- Uninterruptable power supply
- DAT tape drive for multilevel backups
- 8-bay hard disk bay containing 2 4-GB hard disks for index storage
- HP-UX 10.09 CMW operating system
- Pinnacle Micro Apex Magneto-Optical (MO) jukebox with 2.3 GB platters (upgradeable to 4.6 Gb/disk).

3.1.2 Oak Ridge system

A similar but less powerful system has been procured to use in Oak Ridge in order to provide remote support and development once the J-200 system has been transferred to Germantown. This system consists of an HP C-100 which uses the same processor as the J-200. Similar peripherals, including a jukebox, have been procured.

4. Document processing

Documents shall be input into the EFR via one or more PCs located in the OSS File Room and connected to the J-200 via a private ethernet. These PCs shall be running Windows 95 for enhanced performance.

4.1 Scanning hard copy documents

Documents shall be scanned in using the existing Fujitsu scanners in the OSS File Room. 400 dpi resolution shall be used to reduce OCR conversion errors.

4.2 Conversion software

Adobe Capture 1.01 shall be used to convert the scanned TIFF images to Adobe Portable Document Format (PDF) format. The resulting PDF files shall be placed in a directory on the CMW workstation that is NFS-mounted by the PC. The IMC personnel shall enter the document information and interact with the J-200 using the Netscape Web Browser to interact with a private Web server running on the J-200. This interface shall be very similar to that used for CDOCS unless changes are desired. In order to ensure that each document is properly associated with its keyword data (security level, title, owner,....) a cover page containing this information shall be generated and inserted at the start of each document before it is scanned. The fonts of this cover page shall be selected to assure correct OCR conversion. Thus, the first page of each PDF file will be a cover page containing all the document information.

4.3 Electronic document storage

Once IMC personnel have entered the document information and created the resulting PDF file, a trusted program on the J-200 shall apply the correct classification to this file and move it to a disk on the MO jukebox. Finally, the file shall be indexed into the library assigned to the document's security level, and placed into the CQ text retrieval database. Depending upon performance, these updates can be performed as new documents are submitted, or at night in batch mode.

Unclassified files shall be kept on separate MO disks to facilitate an eventual move to the dual-LAN system proposed for Phase II.

4.4 Electronic document submission

The EFR can accept electronic documents either as PostScript files or as PDF files. PostScript files will be converted to PDF files using the Adobe Distiller program that will be installed on the IMC PC. A scheme must be devised for securely transferring these files to the EFR, and for integrating these documents with the usual document entry process so that the proper document information is associated with the electronic file. In particular, an "electronic cover sheet" shall be inserted into each PDF file. Therefore, electronically-submitted PDF files shall have the identical format as those made from scanned documents.

4.5 CDOCS migration

A scheme must be devised for migrating existing CDOCS documents to the EFR. This migration should be as automated as possible. The EFR can handle the TIFF

images and ASCII text used by CDOCS. However, because CDOCS stores each page as two separate files, the only way to access all the files for a given document is to use CDOCS to retrieve them. Whether this process can be automated depends upon the capabilities of CDOCS.

In preparation for the transition from CDOCS to Ask-OSS, and before the deployment of Ask-OSS, a procedure shall be devised for creating and storing either newly-scanned TIFF files in a format suitable for eventual use by Capture, or for converting them into PDF files (with cover pages) using Capture.

4.6 Reclassification

The read/write MO media and file labeling allow any document to be reclassified. This process involves several steps. These steps will be performed by File Room personnel from the scanning PC, using a Web-based interface.

- The security field in the PDF file must be changed to reflect the new classification.
- The file must be removed from its old CQ library and index.
- The file must be inserted into the new CQ library and index.

5. Text retrieval database system

The CQ (CQ) text retrieval database (also known as Retrievalware) is the COTS program that is used for indexing, querying, and retrieval of the EFR documents. CQ will be used in a client-server mode that is connected to a Web-based user interface. CQ uses the concept of a “semantic network” to allow users to make very sophisticated queries without the need to set up any predefined query aids.

6. User Web interface

Users of the EFR shall perform all interactions with the EFR database using a Netscape Web Browser (NWB) that runs on their PC. In addition, a copy of the Adobe Acrobat Reader (free) shall be required to view the PDF files. PDF files can also be viewed with a copy of Acrobat Exchange (< $100) which allows users to rearrange pages and add comments to facilitate group interactions. Acrobat Exchange also supports “plug-ins” which will eventually allow the user to view “hits” (words or phrases in the document matching query parameters) as highlighted or marked up text in the PDF document.

Users may make private copies of any retrieved files on their PCs. Users will be unable to change anything on the database itself.

It is assumed that all users are cleared for access to the LAN.

On the J-200, the Apache Web Server (AWS) is launched from the *inetd.conf* file. This means that a new copy of the AWS is launched for each user interaction, which avoids the contamination of the authentication process that can occur when multiple servers are

running, and the user is connected to a different (random) server each time he selects a new item on a Web page.

All users must have an account on the system, and their security permissions are specified when this account is set up. However, ***no user logins are allowed on the CMW system.***

6.1 User authentication

It is assumed that there are no sniffers on the net so that clear-text user IDs (uid) and passwords may be used. In a later phase of this project, strong authentication and encryption will be implemented, probably using Fortezza technology.

When the user contacts the EFR web page the first time, a uid and password screen is displayed by the user's NWB. These are remembered by the user's copy of Netscape and retransmitted as a uuencoded string with every future transaction.

6.2 Querying for documents

The Apache Web server has been enhanced such that all user authentication occurs against the CMW's Trusted Computing Base (TCB). Each time a user access a Web server (a Web transaction), that user is authenticated. The NWB caches the user's name and password after the initial transaction, so the user must enter that information only once until the Web browser is terminated.

The first step in the chain or pipeline of processes implementing the Web server– CQ interface is to consult the CMW TCB to determine the user's clearance.

6.2.1 CQ data structure

CQ stores a private data structure for each user of the system. This structure is destroyed after a predetermined period of inactivity — currently 15 minutes. The user is identified by his user id and *ip* address (not password). Because the user's uid and password have already been checked against the TCB by the Web front-end, and because we do not want the user's password exposed as stored data in the program, we replaced the CQ password by the user's *ip* address. This may also help to prevent session capture by some other user. This private data structure contains the user's uid and his clearance. CQ identifies the user session using an 8 character sequence of upper-case letters that is based upon the uid, ip address, and time of day. This "CQid" is used to keep track of the state of each individual user.

The CQid is passed in the HTML command line, so it could be captured by a sniffer program. This string will be encrypted in future versions of Ask-OSS, but it is no more vulnerable than the original clear-text password string. OSS has specified that there are no sniffers on their network.

6.2.2 Security level of the CQ front-end server

The front-end server is the CQ process that talks to the Apache web server. All traffic between the two processes must run at SYSLO. However, because each user is assigned a separate web server (on a separate port), the inter-user traffic is separated.

6.3 Retrieving documents

Documents stored in the EFR database will be in PDF format with the possible exception of migrated CDOCS documents. Eventually, the relevancy-ranked database hits can be displayed directly in the PDF document if the user has a copy of Acrobat Exchange. Initially this capability will not be available, so the relevancy-ranked hit list can be displayed in the NWB window as ASCII text. Thus, the user has a choice of viewing the PDF file which displays the document's original format, or the marked-up text that has been extracted from the PDF file by CQ.

7. Web server operations

Two separate Apache Web Servers will be utilized. One will communicate with the database users on the classified LAN. The other will communicate with the IMC PCs on a private ethernet.

7.1 User authentication

The AWS is set up to require authentication each time a user makes a new Web transaction (every time he clicks on a button to initiate display of a new Web page). Only those users listed in the *.htaccess* file are allowed on the system. However, the usual authentication that is performed using the *.htpassword* file is replaced by authentication using the uid and password that is stored in the CMW trusted computing base (TCB).

Authentication will also be applied on the private AWS to assure that only authorized IMC personnel can access the J-200.

7.2 CQ Web facilities

On the server, the CQ Web interface product is used for providing the user interface for the EFR. HTML Web pages are generated dynamically based on user selections and text search results. In addition, a set of online help pages offers guidance on EFR usage and operations.

8. Document server operations

There will be three copies of the CQ document server running, one at each of the security levels, *Unclassified, Confidential*, and *Secret*. Each level will include all compartments.

It has been agreed by OSS that provided a user has a clearance that dominates a library, it would not be an unpardonable breach of security if the user saw the title of a document that was for a compartment not included in the user's allowed set. However, to prevent this, all queries will be accompanied by an automatic prequery that will require that the user's clearance contain all of the compartments in any document's security keyword field. This additional security will be provided by the CQ database engine, and not by the MAC of the CMW system.

8.1 CQ library management

The EFR documents shall be organized into three separate libraries: *Unclassified*, *Confidential*, and *Secret*. Currently, the *Top Secret* category will not be used for security reasons. Each document in a library will be stored as a separate file that will be properly labeled (by the CMW system) with this clearance as well as the need to know categories and handling caveats (i.e., compartments). All documents will be stored on the MO disks.

The CQ indexes will be stored on the external hard drives for speed. Because the index for each library will contain documents with all of the categories and handling caveats, the indexes will be classified at the level of the library plus all of the compartments.

8.1.1 Library access restrictions

The clearance (security level + compartments) is used to ensure that the user is only allowed to see CQ libraries up to and including the user's clearance level. (However, these libraries may contain information for compartments that the user is not allowed to access.)

The library access is enforced at two places.

- The dynamically-configured library selection html page has been written to only display libraries that can be accessed by the user.
- The user could conceivably modify his html request to include libraries above his clearance level. To thwart this, we check the user's selection in *cqfeserv* to eliminate any such attempts.

8.2 CQ query mechanisms

Interaction between the displayed web page and the CQ database is handled by the program *cqcgi*. A proprietary set of macro commands embedded in the html page allows us to get and set CQ variables and to perform simple program constructs (if, then, else, loops, ...).

The user's query is transmitted by the program *cqcgi* to the CQ front-end server, *cqfeserv*. When *cqfeserv* gets the user's query, it is supplemented by a Boolean query over the compartments to prevent the user from seeing hits outside his need to know categories. For example, suppose the system has compartments A, B, C,...F and the

user has access to C and D only. The prequery will be

KEYWORD.COMPARTMENT NOT (A OR B OR E OR F).

Thus far, this mechanism has been 100% reliable in all of our tests, and no hits have been returned to documents that the user should not see.

8.3 CQ document retrieval

Whenever a database document is opened for retrieval, the classification level of the process that opens the document will be changed to the user's classification level so that CMW MAC can be applied to deny access to unauthorized documents. If access to the file is denied, an error "Access denied" will be returned to the user. This situation should not occur if the above prequery mechanism is successful.

Once the document has been successfully opened (because MAC was satisfied), MAC is no longer applied; the document can be read at any level using its file handle. Therefore, the CQ server level will be set to SYSLO so that the document content can be sent out the port to the unlabeled PC on the classified LAN.

In fact, what actually happens is that CQ opens the PDF file, extracts the text, marks it up to show the ranked hit list, and stores the result in a temporary file. This temporary file is what is actually sent back to the user. There is one temporary directory used for all servers, and each file is labeled with its correct security label. In any event, once the file is opened, the server must return to SYSLO in order to send the data back to the front end.

8.4 Document retrieval extensions

CQ has many powerful features. For example, a retrieved document can be used as input to a new query. Most of these extensions will not be supported on the first release of the EFR.

9. Security

The role of security must be considered in each part of the EFR system.

9.1 LAN interface

The users of the EFR system all have PCs connected to the classified LAN. These PCs run at a single security level, and do not understand the multilevel ip security packets that are used by the CMW system. In CMW parlance, these PCs are "single-level unlabeled hosts."

One of the CMW security features is that all computers that are allowed to communicate with the CMW host must be entered into a database called *M6RHDB*. The Network Security Officer (one of the CMW roles) must register each external host and specify its ip address, name, and security protocols. Because PCs are unlabeled hosts, they must come into the CMW system at a single level. It makes no

sense to pick a level other than SYSLO or SYSHI. We believe that allowing them to come in at SYSHI would be a mistake because they would dominate any MAC test in the system. Accordingly, all the OSS PCs are entered as coming in at SYSLO, the lowest possible security level in the system.

In addition, the OSS Novell network assigns the PCs to dynamic ip addresses. Accordingly, all of the OSS PCs will be entered into the *M6RHDB* file using the wildcard *ip* address entry 0.0.0.0. This entry allows any *ip* address to access a Web server, but prevents any PC from logging into the system or running any other services such as FTP. This is exactly the security policy that should be enforced.

A further complication is that the ports used by the Web server and clients also enforce MAC. Therefore, if the PCs are specified as being at SYSLO, either the port MAC must be overridden, or else information must be downgraded in level in order that it can be sent out the port. We feel that overriding the port MAC is a bad idea because the port provides resistance to external attack. Therefore, we will write trusted code (code that enforces the principle of least privilege) that will raise and lower the clearance of the CQ Server as required.

9.2 CMW server operations

The Apache Web Server has been modified to check user access using the facilities of the CMW TCB. It is launched from the *inetd* demon so that a new copy of the server is launched for each user. Because OSS users sometimes share PCs, and because ip addresses are dynamic, we also associate the user's uid with the connection's ip address to prevent spoofing.

The server starts out running as *m6nobody*, the user assigned to any untrusted port. Currently, the server switches to user *root* and drops all privileges except *chsubjident* which allows the server to switch to a different uid. We plan on trying to get the server to run as user *www* rather than root. However, this may not be much of an issue because we drop almost all potential privileges from the process each time it starts.

A log file is maintained of all connections and errors.

9.3 User authentication and clearances

A user has a single clearance and access to any number of compartments. If a user is a member of a compartment, he can access it up to the level of his clearance. Thus, a user has the same clearance in all compartments.

For unclassified information, the concept of hierarchical compartments is used to allow a large number of custom need to know categories. In other words, it is possible to create a l*abel encodings* file that allows one compartment to dominate another.

The user's clearance and compartments are assigned by the ISSO.

9.3.1 User database

It is important that the owners of “need to know” categories can ascertain who has access to which compartments. A trusted program will be written to provide a list of users who have access to any compartment.

9.4 Compartment and label enforcement

Compartments and labels are enforced by the CMW system whenever an object is accessed by a process.

9.5 Auditing

On a CMW system, almost every system event can be audited. However, to ease the burden of looking at these audit events, it is suggested that auditing be minimized. We suggest auditing each actual document retrieval by keeping the user name, document security level, and title.

9.6 Testing plan

The security of the system must be tested.

9.6.1 Architectural review

The overall architecture of the system should be reviewed several times by an IV&V team. A preliminary review shall be conducted during the summer of 1996, with a final review to occur in the first quarter of FY ‘97, after the system is installed at DOE Headquarters..

9.6.2 Test database

Once the system is operational, a test database must be constructed so that the security and usability of the system can be tested.

9.6.3 Penetration testing

The system shall be tested as harshly as possible by the developers to try and determine any security flaws.

INSTALLATION AND OPERATIONS

10. Installation at DOE Headquarters

The J-200 system shall be assembled and thoroughly tested in Oak Ridge. After this testing phase, the entire system shall be moved to the OSS File Room (IMC). The system must be connected to the Classified LAN, and the private LAN to the File Room PCs must be set up.

Software will be preinstalled on the new IMC PC in Oak Ridge.

User accounts (without login privileges) must be set up on the J-200. At this time, user clearances will have to be assigned.

11. Training

Training manuals shall be written when online-help is not available.

11.1 Administrators

While the J-200 is being installed in the IMC, the future system administrator(s) should be prepared to work with Oak Ridge personnel to learn how to operate and administer the CMW operating system.

11.2 Users

Users have access to PCs running Microsoft Windows, and are connected to the OSS classified LAN. The PCs must have a TCP/IP stack and the ability to run the Netscape Web Browser. Users must also have a copy of Adobe Acrobat Reader or Acrobat Exchange. Very little training is required to learn to use these tools, and they all have online help facilities.

11.3 Document processors

The IMC personnel require powerful PCs to be able to process documents in a timely fashion. When the system is installed and functional. Oak Ridge personnel will work with the IMC personnel to teach them how to enter documents into the EFR system.

12. System administrators

The following CMW roles may be occupied by different individuals, or they may be all assumed by the same person.

12.1 System Administrator

The Administrator creates new accounts and backs up the system. He also assigns privileges and clearances to anyone who can assume the role of ISSO.

The administrator will perform system backups. The MO disks have a lifetime of 100 years. Once the disks are full, they should be write protected and mounted as read only. At this time, a DAT tape backup of the disk should be made, but further backups are unnecessary unless flies are reclassified. The system files and indexes should be backed up also using the CMW's trusted backup programs which retain all file labels. Because the database indexes can be recreated, and because the system software should be relatively static, monthly backups will probably be sufficient.

12.2 Information System Security Officer (ISSO)

The ISSO can access to any file in the system, so the ISSO must be cleared for all information in the system and trusted completely. In principle, the ISSO cannot initiate a new account, and cannot change his own privileges (the Administrator performs these functions), but because he can access the files that contain these items, with some effort he could override these controls.

The ISSO assigns each user his privileges on the system. This includes his clearance and compartments. The ISSO also performs audits and checks the integrity of the TCB.

12.3 Network Security Officer (NSO)

The NSO is responsible for entering hosts into the M6RHDB database and specifying their correct security levels. Because no logins are allowed, and a wildcard entry is used for Web access, this role will only be active at system installation time. Later, it can be filled by the ISSO.

PHASE II FUNCTIONAL DESCRIPTION

In Phase II of the EFR project, the system, as installed in the OSS File Room, must be subjected to a full IV&V inspection and be approved before it can become fully operational.

Phase II adds functionality in two areas: enhanced security as well as inter-LAN access.

13. Final IV&V inspection and approval

Before the EFR system can be fully trusted, it must be subjected to a full IV&V review and approved by the review team and by DOE. Preparation for this review requires that a full set of detailed system specifications be prepared.

14. Enhanced authentication

One weakness of the Phase I system lies in the area of authentication and encryption. The technology to provide these enhancements is just becoming available. A primary goal is to eliminate clear-text user ids and passwords. There are several options:

14.1 More secure Web server

There are two options available for enhancing the Web server security. There exists a version of the Apache Web server that supports encryption and enhanced authentication (of the server only) using the secure socket layer (SSL) protocol. However, the license agreement for this version specifically excludes anything having to do with nuclear weapons work! Some negotiations are called for.

The other option is to use the Netscape Commerce Server, provided that it has hooks for the CMW authentication that is required for this project. Netscape has announced that this server will support the Fortezza card, so this option may be particularly attractive.

14.2 Fortezza Card support

The Fortezza card is used to provide strong user authentication and encryption. The user-to-Web server connection will be secured using the Fortezza card. It is still difficult to obtain these cards, and every PC will have to install a PCMCIA card reader in order to use this technology. To implement the Fortezza card support, the Web server will have to be modified, and it may be necessary to write a Windows-based program to run on the user's PCs.

14.3 Security options for Conquest

Conquest version 5.2 has a security option package available that utilizes Kerberos to encrypt all sessions and to provide user authentication. This option costs about $5000 additional. Because of the extensive code customization that we have performed, we are currently using Conquest version 5.0.3 and may upgrade to 5.0.5 if time permits.

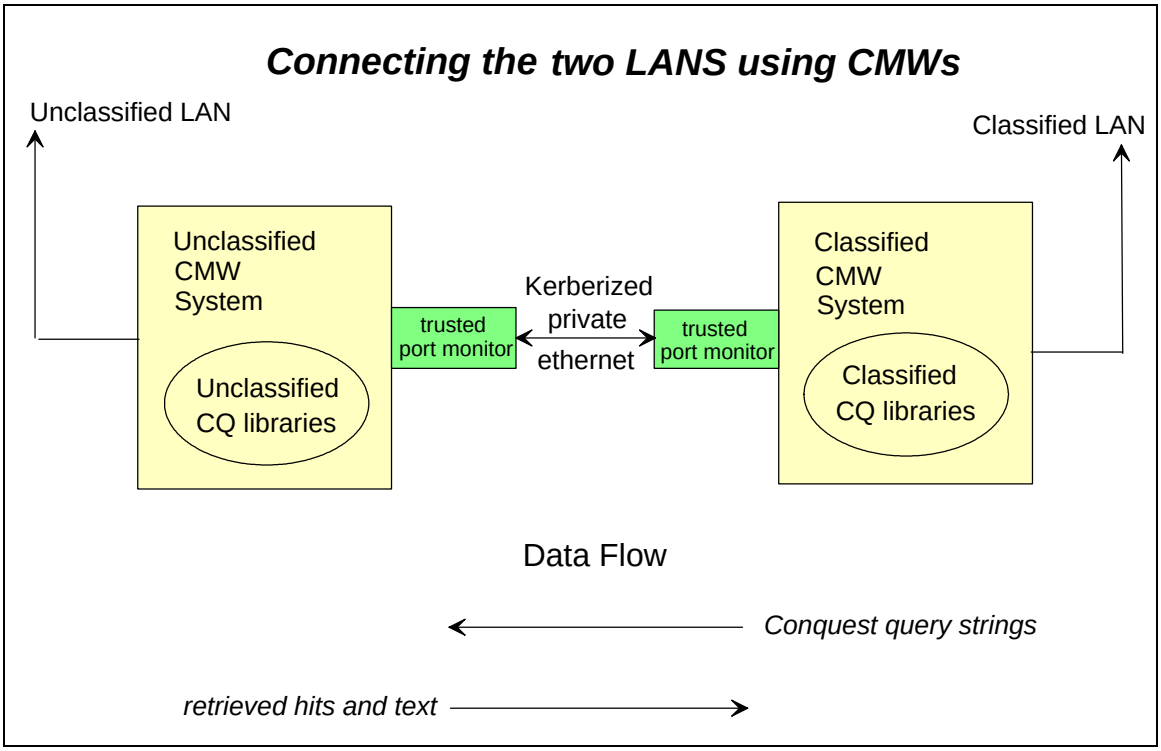
15. Connection between LANS

One vision of the EFR project is to be able to allow both cleared and uncleared users access to appropriate parts of the EFR database. The technology to do this securely will be available when the Fortezza-card support has been implemented.

Two CMW workstations will be used, one on each LAN, and configured as in Phase I. The unclassified library will only be mounted on the unclassified J-200. The two workstations will be connected via a separate, private ethernet circuit using additional ethernet cards.

The CQ text retrieval system supports queries to remote databases. Version 5.2 of CQ enhances this option by providing Kerberos authentication and encryption of the traffic between the *cqfeserv* front end and the search engine and its document libraries and indexes. The copy of CQ on the unclassified system will not allow any remote databases to be mounted, thus queries will be restricted to the unclassified database. CQ on the classified system will mount its own classified libraries plus the remote unclassified library.

Trusted programs on each machine will monitor the inter-LAN ethernet ports to ensure that the only traffic comes from the CQ queries (going from the high LAN to the low LAN), and the CQ retrievals (going from the low LAN to the high LAN). The attractive feature of this scheme is that most of the data flow is from the low LAN to the high LAN (write-up), and except for the possibility of classified words being used in a query, no classified information can flow from the high LAN to the low LAN.


Connecting the two LANS using CMWs
Unclassified LAN
Classified LAN
Unclassified
CMW
System
Unclassified
CQ libraries
trusted
port monitor
Kerberized
private
ethernet
trusted
port monitor
Classified
CMW
System
Classified
CQ libraries
Data Flow
Conquest query strings
retrieved hits and text

Acronyms Used in This Document

AWS	Apache Web Server
CQ	Conquest text retrieval engine
CMW	Compartmented mode workstation
EFR	Electronic file room
HP	Hewlett-Packard, Inc.
HTML	Hypertext markup language
IMC	The OSS Information Management Center, i.e., the File Room.
ip	Internet protocol — the ip address is used to specify a host on the Internet
ISSO	Information system security officer
LAN	Local area network
MAC	Mandatory access controls
NWB	Netscape Web Browser
PC	Personal computer
PDF	Adobe portable document format
SSL	Secure socket layer
SYSHI	The highest security level on the CMW system
SYSLO	The lowest security level on the CMW system
TCB	Trusted computing base — the core of the CMW operating system

Task Breakdown — Phase I

Item	Section	Time to Complete (person weeks)	Status
System architecture		0	Complete
Obtaining hardware		1 2—3 month delivery	Jukeboxes, hard disk bay and Oak Ridge equivalent workstation (an HP C-100) all ordered. IMC PC must be ordered.
Programming Jukebox		2	Awaiting equipment arrival
Document preprocessing web interface	4.2	0	Working
PDF file modification to add information fields	4.3	0	Completed
Procedure to classify file and add it to database	4.3	2	Just starting
User Web interface	6, 7.2	1	Functional, ready for fine-tuning
CQ client-server security programming	6.3, 8.1–8.3	3	Working. Undergoing testing
Reinstallation of CMW with correct security model	9.3	2	Awaiting new CMW version
Apache Web Server programming	7, 7.1, 9.2	0	Complete
Installation of second LAN for data input from PCs	4.2	1	Awaiting ethernet card
CQ indexing and library management	8.1	3	Working. Procedures for reclassification underway
User database	9.3.1	1	Future
Preparation for initial IV&V	9.6.1	2	Underway
Training manuals	11	8	Just started
Creation of a test database	9.6.2	0	Completed
Penetration testing	9.6.3	2	Future
Training DOE personnel	11	2	Future
Setting up auditing	9.5	2	Future
Installation in IMC	10	2	Future
Scheme for direct electronic input of documents	4.4	4	Future

Scheme for reclassification of documents	4.6	3	Future
Devising a CDOCS migration scheme	4.5	8	Just starting

Task Breakdown — Phase II

Item	Section	Time to complete (person weeks)
Final IV&V of Phase I system	13	8
Trouble shooting, bug fixing, and maintenance of Phase I system		12
Implementation of Fortezza technology	14.1	20
Development of multi-host CMW to connect classified and unclassified LANS	14.2	52