RL-TR-93-257
Final Technical Report
December 1993

KNOWLEDGE BASED MULTI-LEVEL SECURE NETWORK TECHNOLOGY

Fuentez Systems Concepts, Inc.

Thomas C. Housman and Raymond Fuentez

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Rome Laboratory
Air Force Materiel Command
Griffiss Air Force Base, New York

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-93-257 has been reviewed and is approved for publication.

APPROVED:

HERBERT E. MARKLE, Capt, USAF
Project Engineer

FOR THE COMMANDER

JOHN A. GRANIERO
Chief Scientist for C3

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL (C3DA) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1993	Final Dec 90 - Aug 93

4. TITLE AND SUBTITLE
KNOWLEDGE BASED MULTI-LEVEL SECURE NETWORK TECHNOLOGY

5. FUNDING NUMBERS
C - F30602-89-C-0085
PE - 62702F
PR - 4519
TA - 24
WU - 05

6. AUTHOR(S)
Thomas C. Housman, Raymond M. Fuentez

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Fuentez Systems Concepts, Inc.
11781 Lee Jackson Highway, Suite 700
Fairfax VA 22033

8. PERFORMING ORGANIZATION REPORT NUMBER
FSC F085-A009

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Rome Laboratory (C3DA)
525 Brooks Road
Griffiss AFB NY 13441-4505

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
RL-TR-93-257

11. SUPPLEMENTARY NOTES
Rome Laboratory Project Engineer: Capt. Herbert Markle/C3DA/(315) 330-2091

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The application of artificial intelligence technology to assist analysts in the field to identify Sensitive Compartmented Information (SCI) which must be sanitized and disseminated to US Military commanders was investigated and documented. The current manual methods were modeled and used in the design of an EXPERT SYSTEMS prototype to perform the functions of (a) identifying time critical SCI needed by the commanders,(b) associating the SCI with related intelligence from other sources, (c) determining the processing required to sanitize the SCI to appropriately classified collateral (non-SCI) reports, and (d) automatically generating and displaying to the operator the sanitized reports. The prototype was installed and tested at Rome Laboratory, Griffiss AFB NY 13441-4505.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Artificial Intelligence (AI), Expert System (ES), Knowledge Based Systems (KBS), Sanitization	72
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

Page

III. CONCLUSIONS

Page

II. RESULTS

III. CONCLUSIONS

LIST OF TABLES

Page

II. RESULTS

III. CONCLUSIONS

I. INTRODUCTION

1. GENERAL

This final technical report documents the results of the Knowledge Based Multi-Level Secure Network Technology (KBMLSNT) project. The document details all technical work accomplished and information gained in the performance of the contract. The document describes the demonstration of artificial intelligence techniques to perform identification and sanitization of sensitive information for dissemination to tactical battlefield commanders.

2. BACKGROUND

Intelligence information acquired through sophisticated sensors are critical inputs to the tactical commanders decision process. Tactical commanders in the field require this real-time intelligence support but do not need to know the source of that data, nor do they have the means to protect its highly classified nature. The sanitization of classified data can be employed to permit wider distribution of essential intelligence information while protecting the sensitive source of that information. Manual sanitization is a time consuming process which may delay the expeditious processing and dissemination of intelligence information. Automation of the sanitization process could significantly improve the timely dissemination of critical information to battlefield commanders thereby improving the preparedness of forward area deployed forces.

Current sanitization methods require that analysts identify sensitive intelligence information and determine through clearly established procedures how that data can be sanitized. Once sanitized, the analyst must write or reformat the releasable information into a properly formatted message for dissemination to tactical battlefield commanders.

In a dynamic, active threat environment, there can be an order of magnitude increase in the volume of critical, intelligence derived information. During these high tempo operations, human sanitization would likely result in the delay or loss of information required by tactical commanders. As the volume of intelligence data increases, there exists the potential for a corresponding increase in inadvertent security breaches, critical information gaps, and lengthy delays in the delivery of critical information to tactical commanders.

3. OBJECTIVE

The objective of this project was to define and develop a knowledge based multi-level secure network interface that will increase the flow of critical Command, Control, Communications, and Intelligence (C3I) information to the tactical battlefield commander. Specifically, the efforts were to assess the feasibility of applying artificial intelligence techniques to assist the intelligence analyst in the sanitization of classified information from one classification level to a lower classification level, eventually automating the sanitization process. The target product is an implementation of a sanitization model which utilizes commercial off-the-shelf

software packages, transportable and is complainant with government and industry standards.

4. APPROACH

The approach was to define the manual sanitization procedures with sufficient clarity to initially develop a model and then implement the model as a computer software. The first research task was to conduct an analysis of existing manual sanitization procedures at a site where those rules, procedures, patterns, and information structures could be readily obtained. This analysis would result in the definition of a sanitization model. The second research task focused on identifying available commercial off-the-shelf technology which could be applied to the implementation of the model. Both expert system tools and natural language tools were surveyed. The third task was to develop a design for a computer based architecture for implementing the model and proceed with implementation of the model. The design maximizes the use of COTS software. The final task was to demonstrate the sanitization of formatted message traffic.

II. RESULTS

1. GENERAL

The work accomplished three tasks: analysis of sanitization procedures and development of a sanitization model; survey of AI technologies including expert systems and natural language processors; and implementation of the sanitization model resulting in a feasibility demonstration. The demonstration was to show how sanitization of sensitive compartmented information can be accomplished with Expert Systems computer software.

The following sections describes the accomplishments and results in detail. Section 2 presents the results of an on-site analysis of manual sanitization procedures, conducted at USAFE facilities. Section 3 presents the results from surveys of document parsing and expert system tools. Section 4 presents the design of an expert system based sanitization prototype. Section 5 presents a summary of all preliminary and formal testing. Section 6 presents an approach to implementation with a concept of operation.

2. ANALYSIS RESULTS

A feasibility analysis was conducted at the United States Air Force Headquarters in Europe (USAFE) and Allied/NATO facilities in the German Democratic Republic in November of 1989 and May of 1990. The analysis focused on determining whether the process of sanitizing information acquired through highly sensitive intelligence sources can be specified with sufficient clarity to initially assist and ultimately replace the manual operator and if such specificity is possible to develop a multi level secure sanitization model.

The analysis of existing manual sanitization procedures is the crucial first step in developing an automated approach to initially assisting and eventually automating the intelligence analyst function of identifying and sanitizing critical intelligence information. This effort identified rules, procedures, patterns, and information structures which compose the manual sanitization process and the development of a model which reflects the manual sanitization process.

An iterative development approach was used which incorporates end-users' operational experiences through a structured review/enhancement cycle, yielding the real-world model. A team of senior engineers were tasked: to analyze manual sanitization procedures and if feasible to develop an sanitization model. FSC engineers identified, studied and documented analyst's procedures and techniques which pertained to sanitization. The following approach was taken:

- Examine National, Service, and Theater level documentation to obtain a textbook definition for a sanitization model and to develop a list of questions for site intelligence analysts to answer.

- Generate a textbook sanitization model and questioning scheme to be used during the on-site data gathering knowledge acquisition process.

- Interact with site intelligence analysts to record real-world procedures which affect the model's definition.

- Integrate analyst's thoughts into the model.

2.1 Site Survey

An analysis of the Combat Operations Intelligence Center (COIC), Ramstein AFB, West Germany and the Tactical Fusion Center (TFC), Boerfink, West Germany determined that the sanitization process can be specified with sufficient clarity to initially assist and ultimately replace the manual operator.

2.1.1 USAFE Combat Operations Intelligence Center (COIC)

The COIC is divided into several functional intelligence areas charged with developing situation summaries, briefings or other reports by assimilating intelligence data with database information (order of battle files, commander's operation plans/tasking orders, enemy doctrine). These reports are then used by the reconnaissance retaskers to direct sensor sources to the tactical areas of interest. This provides the most current intelligence picture for the near-real-time targeteer to use to prepare strike plans against mobile and fixed targets.

Although the COIC does generate classified information, it does not have a comprehensive NATO reporting requirement as part of its mission. Discussions with several analysts suggests that the sanitization function for the NATO consumers is rarely performed at the COIC. For the most part, historical and background information is maintained in hardcopy reports and messages.

2.1.2 USAFE Tactical Fusion Center (TFC)

The TFC is divided into the Warning Branch and the Force Assessment Branch. The Warning Branch consist of watch standing personnel that are divided into three Indications and Warning (I&W) teams to provide around the clock real-time support to NATO. The Force Assessment (FA) Branch is comprised of day working support personnel that provide support to the I&W teams, and perform various other functions such as; data base maintenance for the Air Order of Battle (AOB), the Surface-to-Air (SAM) Missile-Order-of-battle (MOB), and ad hoc analysis and reporting functions.

The TFC receives intelligence information from a variety of sources and sensors via record message traffic. These messages include:

- NARRATIVE SPOT REPORTS
- NARRATIVE E-GRAM REPORTS
- OTHER NARRATIVE REPORTS
- FORMATTED IMAGERY REPORTS
- FORMATTED TACTICAL REPORTS
- FORMATTED TACTICAL ELINT REPORTS
- FORMATTED OPERATIONAL ELINT REPORTS.

A number of the above message types are received with a classification line which allows automatic release to NATO consumers without additional analyst intervention. Others require analysis, evaluation, and editing prior to being considered for release to the NATO community. The TFC produces the following intelligence products to various consumers:

- SAM OOB for NATO
- USAFE AOB
- Fused Intelligence Report to Europe (FIRE)
- TFC Intelligence Input to NATO (TIIN)
- TFC Wartime Intelligence Input to NATO (TWIIN) - Wartime Only
- SAM Intelligence Report (SAMINTREP)
- Intelligence Report (INTREP)
- Target Air Nomination (TAN) - Wartime Only
  (a hardcopy document produced yearly).

The following events or activities are routinely reported by TFC analysts: forward weather conditions, selected border and/or air corridor violations, selected air defense exercises, instances of communist block active Electronic Counter Measures (ECM), selected mobility training exercises, selected deployment operations, Missile-Order-of-battle (MOB) information, intelligence collection flights and forward area penetrations by Soviet bombers, selected ground support exercises/operations, instances of live missile firings, and any other significant air or ground activity.

The TFC analyst uses a two step approach to determining if a message is releasable to NATO. The first step is to check formal DOD, USAF, and locally generated procedures that establish well defined rules and guidelines for release of information to NATO. The second step is to make a decision, or request assistance in making a decision based on the sensitivity of the information, the importance/need of the information and suitable cover for the information. To accomplish this, the TFC analyst will either sanitize or decompartment the information at the TFC and release that information to NATO. For this purpose, the following definitions apply:

- Sanitization - The process of editing or otherwise altering intelligence to protect sensitive sources, methods or analytical capabilities so as to permit greater dissemination of the information. Sanitization is not to be confused with "Declassification or Downgrading". A sanitized "SECRET" report is still classified "SECRET" after the sanitization process has taken place.

- Decompartmentation - The removal of information from a compartmentation system without altering the information to conceal sources, methods or analytical procedures. Normally, decompartmented intelligence products must contain certain control markings that restrict dissemination to US. channels only.

- Release - The physical or electronic transfer of intelligence products to an authorized recipient.

In supporting NATO customers, the TFC sanitizes information to the SECRET level and below. In general two intelligence disciplines are affected by this sanitization process; Signal Intelligence (SIGINT) and Imagery Intelligence (IMINT). Each discipline presents unique sanitization considerations for the reporting TFC analyst.

Locally generated Operating Instructions (OIs) provide the analyst a baseline from which a sanitization decision can usually be made. However, situations may arise that are not addressed by existing OIs and the analyst must then rely on two key considerations of sanitization; 1) sensitivity of the information involved and 2) the intelligence needs of the NATO recipients. Due to the nature of the information and processes involved, the rules governing sanitization are themselves classified and are not included in this report.

2.1.3 Survey Conclusion

The operational mission of the COIC as determined through our on site visit suggests that the COIC not be considered as a candidate for modeling of the sanitization process. The mission of the TFC and the established well described and tightly constrained process of preparing messages for release to NATO indicates the greater potential for modeling the manual sanitization process.

2.2 Sanitization Model Definition

The Sanitization model documents the manual sanitization process performed by intelligence analysts at the Tactical Fusion Center when providing timely intelligence information to Tactical Battlefield Commanders. The model specifically concentrates on tasks which analysts perform during sanitization of sensitive information. The model takes into account the knowledge which analysts acquire during the process and suggests an organization of that knowledge.

2.2.1 Model Summary.

The model encompasses four primary tasks; Identification, Association, Sanitization, and Dissemination. Figure 2.2.1-1 depicts a decision flow of the Sanitization model. At each task information known as sanitization knowledge is accessed, collected, and organized into a sanitization knowledge-base.

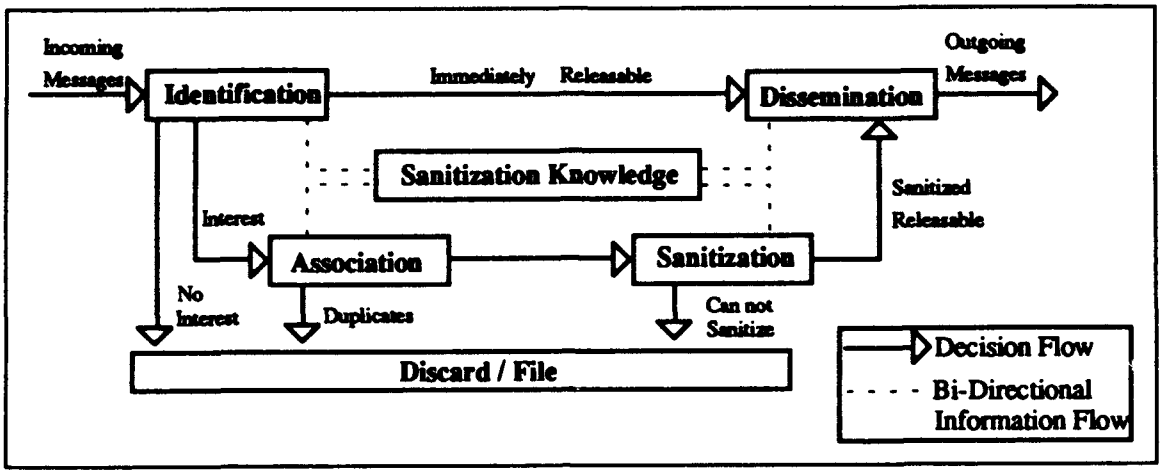

FIGURE 2.2.1-1 Sanitization Model

The Identification task identifies messages and other intelligence information into Three general categories;

1. Information of no interest to supported commanders. This information is purged from the system with no further action.

2. Information of interest to supported commanders. This information is disseminated without further processing.

3. Information of interest to the supported commanders that is classified SCI and is not releasable in its current form and needs to be considered for sanitization.

The Association task relates incoming messages and other intelligence information to existing intelligence information. Duplicate information is purged from the system.

The Sanitization task applies sanitization rules to separate messages and other intelligence information of interest to supported commanders into either information which can not be sanitized or information which can be sanitized. Information which can not be sanitized under any circumstances is stored in an SCI data base and used to support further analysis. Information which can be sanitized is processed by the analyst to produce an appropriately classified collateral (non-SCI) product for dissemination to the commanders.

Dissemination is the final task in the model and involves the review and release of collateral (non-SCI) products to supported commanders.

The Sanitization model, a sanitization knowledge base which retains the experience and education of analysts. Figure 2.2.1-2 depicts the sanitization knowledge base. Knowledge is divided into information about the message, related intelligence information, and information pertaining to sanitization. Some knowledge is temporary, while other knowledge will linger and remain within the knowledge base analogous to the human's experience and education. It is referenced again and again while processing messages and other intelligence information.

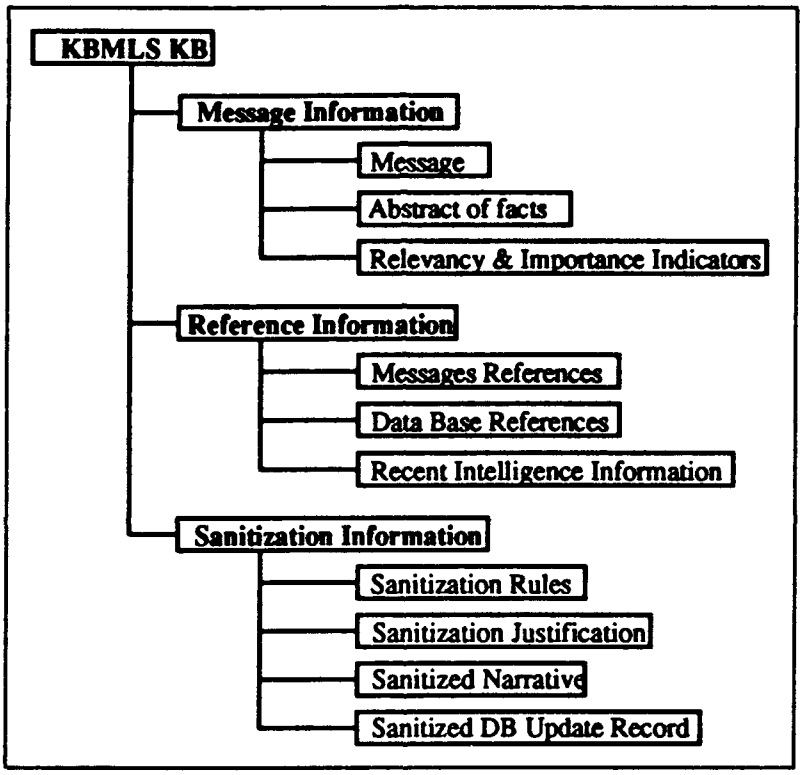

FIGURE 2.2.1-2. Sanitization Knowledge Base

2.2.2 Model Description.

This section identifies and describes the tasks, functions and procedures with the Sanitization Model. Each task is presented with a description of its purpose and processing flow. In addition, a description of the specific knowledge acquired in each task during the manual sanitization process is documented.

2.2.2.1 Identification

The purpose of the Identification task is to recognize information of interest to tactical battlefield commanders and identify candidates for sanitization. This screening procedure, performed by analysts, determines the degree of relevancy and importance of the data to tactical battlefield commanders. Figure 2.2.2-1 depicts the Identification task functions and flow.

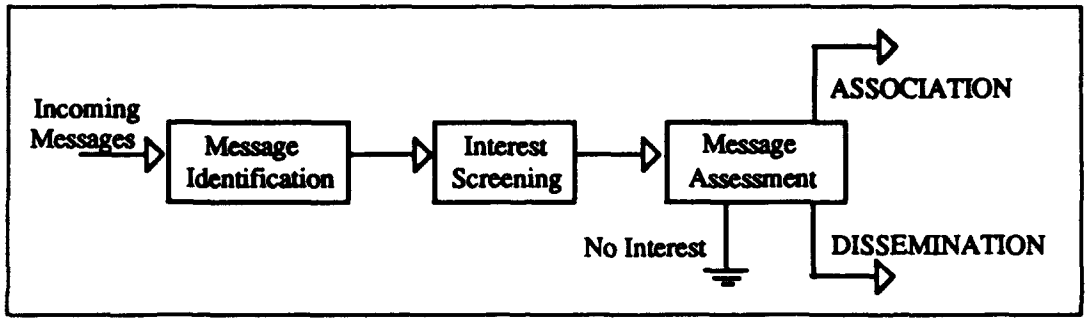

FIGURE 2.2.2-1 Identification Task Functional Flow

Incoming messages are reviewed for subject and classification by the Message Identification function. The message header contains the source, classification, and priority. The body of the message contains the answers to the questions: WHO (object), WHAT (activity), WHERE (location), WHY (history), WHEN (time), and HOW (history). The Interest Screening function identifies wing commanders which are interested in the information. The Message Assessment function determines the processing flow for the remainder of the model. Messages which contain information of little value to the analyst are set aside and referenced in the future as recent intelligence information. Messages of high interest which already satisfy releasable criteria are candidates for the Dissemination Task. High interest sensitive information which do not meet initial release criteria require sanitization and must proceed to the next step, the Association task.

2.2.2.2 Association

The purpose of the Association task is to establish links to all relevant information. Figure 2.2.2-2 depicts the Association task functions and flow.

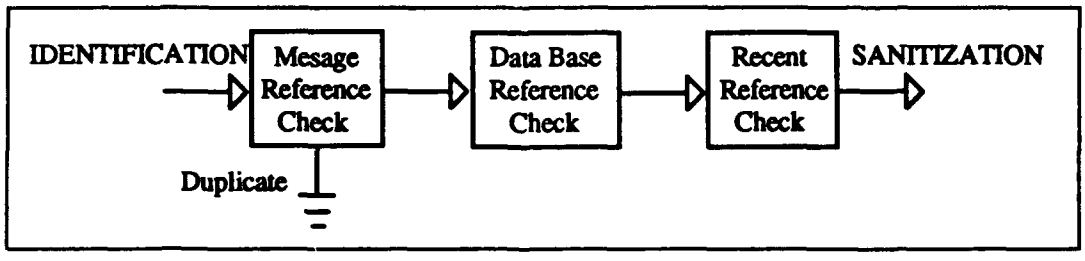

FIGURE 2.2.2-2 Association Task Functional Flow

The Message Reference Check function reviews message queues for related information. The Data Base Reference Check function reviews Order-of-Battle Data Bases. Missile-Order-of-Battle (MOB) and Air-Order-of-Battle (AOB) data bases are examined to obtain the latest information on hostile air wings and surface-to-air missiles. The Recent Reference Check function reviews information which for one reason or another was not capable of being sanitized

2.2.2.3 Sanitization

The purpose of the Sanitization task is to sanitize compartmental information by implementing specific sanitization procedures. Sanitization consists of two separate activities. The first activity is to determine if the information can be sanitized. Activities are; (1) determine why the information is SCI, (2) determine if information can be sanitized under existing criteria, (3) determine if operational situation meets sanitization criteria, (4) determine if required additional information meets criteria, and (5) select appropriate sanitization procedure. The second activity is the actual sanitization of the SCI information. Here is where the information is removed, disguised, or merged with less sensitive information to create a releasable product. Figure 2.2.2-3 depicts the Sanitization task functions and flow.

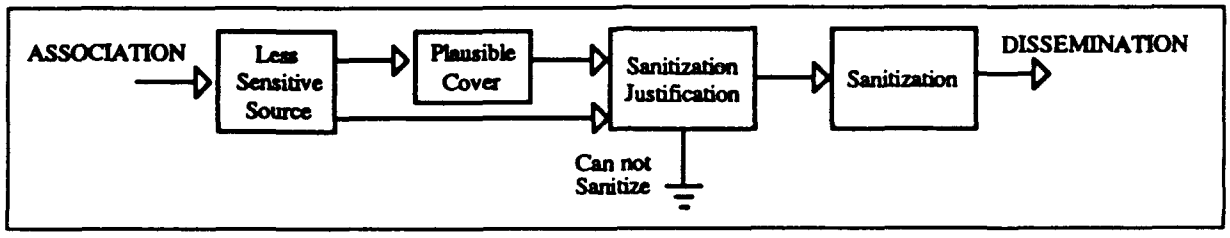

FIGURE 2.2.2-3 Sanitization Task Functional Flow

The Less Sensitive Source check attempts to find another messages which are less sensitive. The Plausible Cover function is where suitable coverage or reasonable coverage. The Sanitization Justification function contains a set of rules which determine if information can be sanitized and to what degree. The determination is based on rules derived from National, Service, Theater, and site specific manuals. The Message Sanitization function relies on sanitization instructions to minimize the risk of possible compromise of source, methods, techniques, and/or degree of success. Each message which is received by the site has a set of modifiable instructions which address methods for sanitization.

2.2.2.4 Dissemination

The purpose of the Dissemination task is to review and release the sanitized information. This task generates the proper message and reviews the message for release. Figure 2.2.2-4 depicts the Dissemination task functions and flow.

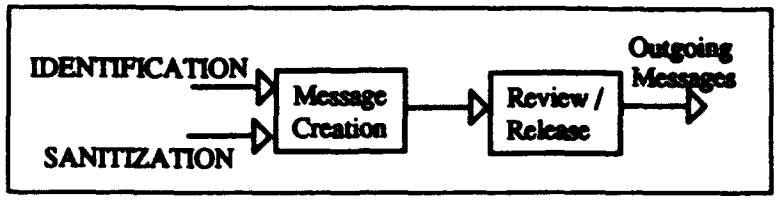

FIGURE 2.2.2-4 Dissemination Task Functional Flow

The Message Creation function is the generation and editing of sanitized messages. The Review/Release functions is a security check performed on all outgoing messages prior to release. The message is reviewed for sensitive words or phrases which might expose the source, method, or technique which was used to acquire this information. The review is performed to ensure that the sanitized information implements the approved security policy.

2.3 Conclusions

2.3.1 Current Sanitization Process verses the Sanitization Model

The current manual process for extracting information received from sensitive sources and for making changes to such information in order to make it eligible for release to users was compared to the sanitization model described in the section. The Sanitization Model was found to be accurate in its description of all major functions and in its application of guidelines and standard operating procedures of the National, Theater, and local security regulations. The model was discussed with analysts who routinely perform the sanitization. The consensus among those who reviewed the model is that it quite accurately represents the manual process and that it accounts for all the major factors involved in making the sanitization decisions.

2.3.2 Closing Summary

The feasibility analysis detailed the manual sanitization procedures used in the USAFE Tactical Fusion Center and developed an information flow model to use for further analysis. The analysis to date shows a very high potential for developing Artificial Intelligence techniques to bring significant automation to the sanitization process with no increase in security vulnerability. Using the defined model it was shown that at every major point in the process, the activities can be specified with sufficient clarity and control to develop a computer program to assist the operator in assembling and analyzing the information and the rules which govern its releasability. As a minimum, a rule oriented (Expert System) Data Base Management System seems very feasible as a means to assist the analysts in linking past messages to current sanitization issues. The DBMS could provide the basis for further automation such as key situation assessment through pattern analysis techniques. At every stage in the model, there is at least the potential to make the analyst's effort more productive and less error prone through the use of expert systems techniques. At several points, it may be possible to achieve a completely automatic function that generates a product for an operator to approve. Building on a prototype

which starts out with basic AI capabilities like the ones described above, it appears at least reasonably feasible to increase its performance and security trustworthiness to the point that it could perform the sanitization without operator intervention.

3. SURVEY RESULTS

3.1 Overview

During the architecture development phase of the KBMLS prototype, specified and derived functional, performance and supportability requirements were allocated to system software and hardware components. Based upon these requirements, and based upon the sanitization model architecture of the KBMLS prototype, criteria were derived for evaluating GOTS/COTS software tools which could satisfy the allocated requirements. The approach taken in conducting this survey included the following objectives:

- Define and establish the criteria to be used in evaluating and selecting appropriate parsing and expert system tools
- Obtain general information on as many tools of each category as possible
- Reduce the sample population to those GOTS/COTS tools which could provide the allocated document parsing and knowledge representation requirements of the KBMLS prototype architecture
- Evaluate the population subset based on a tool's ability to meet the established criteria
- Identify those tools most capable of providing the necessary parsing/expert system functionality based on the specified criteria.

General information was solicited from numerous vendors of document parsing tools and expert systems. Appendix A lists the evaluated tool population for expert systems and provides a brief description of some of the basic features of the twenty-five tools which were initially considered in the survey. Appendix B provides information on the evaluated document parsing tools.

Based on a comparison of the sample tool population features with the established evaluation criteria, KES (Software A&E) is recommended as the KBMLS expert system shell. KES provides the richest set of expert system features, is available on a larger number of host platforms, and has a software architecture which facilitates integration with other KBMLS applications.

The preliminary analysis and evaluation of document parsing tools showed a diverse range of capabilities and functions which made a direct comparison of features difficult. The survey was

unable to identify any document parsing tools which specifically meet all of the requirements of the KBMLS prototype. Most commercially available tools which satisfied a subset of the required technical and functional features could be categorized as natural language processing and text/information retrieval systems. None of these tools were specifically engineered as "document parsing" tools. As a result, each failed to satisfy many of the KBMLS prototype requirements. However, of those tools surveyed, NL Builder (Sychronetics, Inc.) provides acceptable functionality and an architecture that permits limited integration within the KBMLS prototype.

The results of the survey are based on literature reviews, independent evaluations, discussions with developers of fielded systems, vendor-supplied materials, and write-ups on fielded systems when available.

3.2 Expert System Survey Results

The following subsections summarize the evaluations, draw conclusions, and present a recommendation.

3.2.1. Evaluations

Of 25 expert systems identified, five were selected for in-depth evaluation and analysis: COSMIC CLIPS developed by the National Aeronautics and Space Administration; GURU developed by Micro Data Systems, Inc.; EXSYS developed by Exsys Corporation; KES developed by Software A&E, Inc.; and NEXPERT developed by Neuron Data Inc. All of these tools are categorized as expert system shells. Each of the five shells were extensively evaluated and reviewed against the survey evaluation criteria. Table 3.2.1-1 summarizes the evaluation of expert system characteristics.

TABLE 3.2.1-1 Expert System Evaluation Chart

	NEXPERT	CLIPS	EXSYS	GURU	KES
Power & Flexibility	9	4	6	6	9
Compatibility & Portability	9	8	7	5	10
System Interface	9	8	5	6	9
Vendor Evaluation	9	7	9	8	10
Total	36	27	27	25	38

Initial evaluations of the five expert system (ES) shells revealed that each possessed acceptable developer and user interfaces. No shell provided features significantly superior to the others in this area. In general, expert system shells, as opposed to AI languages such as LISP and PROLOG, have been specifically engineered to provide enhanced developer and user interfaces in

addition to providing extensive development and debugging capabilities. Each shell's developer interface, user interface, development capacity and debugging capabilities satisfied the established evaluation criteria for development and implementation of the KBMLS prototype. In addition, each of the five tools were evaluated as having acceptable capabilities for rapid prototype development and iterative refinement of their knowledge bases. These capabilities (i.e. enhanced user/developer interfaces and development/debugging capabilities) were the primary screening criteria used to reduce the initial twenty-five tools to five selected for further evaluation.

Technical features deemed critical to the KBMLS prototype development are: strength of rule representation and ease of rule modification (power and flexibility); availability on a wide range of conventional hardware platforms (portability); and the ability to be integrated and embedded within other conventionally developed applications (system interface). Secondary desirable features important to both the rapid development and future evolution of the system, relate to the tool's acceptance in the market place (i.e. defacto standards), and long-range vendor stability and documentation/support (vendor evaluations). Each of the shells were rated against these criteria and the resultant ratings were used in the final selection process. Table 3.2.1-2 provides the rating results. A dot indicates the expert system has the capability.

TABLE 3.2.1-2 Expert System Rating Results

	KES	NEXPERT	EXSYS	CLIPS	GURU
Development & User Interface					
Explain Facility	●	●	●	●	●
Help Facility	●	●	●	●	●
Justify Facility	●	●	●	●	
Why Facility	●	●	●	●	
Menu Driven	●	●	●	●	●
Window Interface	●	●		●	
Power & Flexibility					
Backward Chaining	●	●	●		●
Forward Chaining	●	●	●	●	
Production Rules (i.e. If/Then)	●	●	●	●	●
Class Inheritance	●	●	●		
Numerical Calculations & Function	●	●	●	●	●
Certainty Factors	●	●			
Handle Unknowns	●	●			
Compatibility & Portability					
Language Developed	C	C	C	C	C
UNIX OS	●	●	●	●	●
VAX-VMS OS	●	●	●	●	●
DOS OS	●	●	●	●	●
OS2 OS	●	●	●	●	●
System Interface Capability					
Seamless Embedding	●	●		●	
Ext. Language Calls	●	●	●	●	
Bridges to Ext. Data and Applications	●	●	●	●	●
Part of Integrated Environment					●
Vendor Evaluations					
Training Courses	●	●	●		●
Consulting	●	●	●		●
Telephone Support	●	●	●	●	●
Documentation	EXCELLENT	GOOD	GOOD	GOOD	POOR

The following paragraphs provide a brief description of the criteria used in the evaluation process and their implications for the KBMLSNI demonstration system. Following the criteria description is an evaluation of the capabilities of each of the five shells in relation to the specified criteria.

3.2.1.1 Power/Flexibility of KB Representation & Inferencing

Expert system shells offer many different ways to represent knowledge as well as many different ways to reason with the knowledge. Some types of knowledge representation are more naturally suited one set of problem domains than another. In evaluating the shells, consideration was given to the type and complexity of knowledge representation and reasoning the sanitizing problem requires. Each of the shells were evaluated with respect to their ability to encode that knowledge and adequately reason (inference) using it. The automated sanitization process requires that the selected expert system shell possess a very powerful knowledge representation scheme and inferencing mechanism. Simple rule-based tools and tools representing knowledge as context trees were deemed to lack sufficient power to develop the sanitization knowledge base. The more complete methods of representing knowledge and inferencing found in hybrid tools was determined to be appropriate for this application.

The EXSYS shell is classified as a simple rule-based tool. It represents knowledge as rules and uses a backward chaining inference engine to process the rules. The system is written in the C language and is noteworthy for its speed and compactness of code. EXSYS represents facts as attribute-value pairs and uses if-then rules to represent relationships between facts. The system can handle approximately 700 rules per 64K of memory and can process an unlimited number of rules in a UNIX environment.

NEXPERT OBJECT is also written in the C programming language. Its inference engine uses both forward and backward chaining, requires 4 Mbytes of memory and runs acceptably fast on Intel 80386 based systems. NEXPERT is advertised as a hybrid tool, providing a number of knowledge representation paradigms as well as a number of complex relationships between the knowledge. An independent review of NEXPERT OBJECT indicated that this shell provides only a limited implementation of object-oriented programming and it lacks the message passing features needed to process complex information within its slots. The review indicated that it is better to view this shell as a sophisticated, structured, rule-based tool that allows inheritance in an object-oriented environment versus a pure hybrid tool.

KES, also written in the C language, has been designed with a highly modular structure using the "information hiding" technique. This structure has allowed KES designers to break down the system into many independent, logical units or modules. The shell currently supports an extensive set of knowledge base features including: Backward and forward chaining, object oriented data representation, inheritance, consistency (truth) maintenance and certainty factors. KES also provides three types of inference engines which increases both its flexibility and power. The three inference engines are the Production Rules (PS), Probability and Inference (BAYES) and the Hypothesize and Test (HT) engine. KES PS employs production rules to perform deductive reasoning, KES BAYES performs statistical pattern classification based on Bayes' theorem, and KES HT provides a higher level inference engine which uses deductive reasoning.

COSMIC CLIPS is another shell written in C. The developers of CLIPS designed the shell with the goal of portability, efficiency and functionality. CLIPS is a forward chaining rule-based tool which is based on the Rete algorithm. The collection of conditions and actions to be taken if rule set conditions are met is constructed into a rule network. While the tool only utilizes one type of inferencing, forward chaining, the developers of CLIPS feel that this limitation is offset by the complexity of use and training associated with hybrid and more complicated type of inferencing tools. Powerful systems have been built using single paradigm tools, but this fact limits, to a degree, the tool's power and flexibility of inferencing with the knowledge base.

GURU is an integrated expert system development environment, containing several different types of computer processing capabilities. When evaluating the GURU expert system shell in isolation from the rest of GURU's development capabilities, it is rated as a simple rule-based system which represents facts and rules via attribute-value pairs. GURU relies primarily on backward chaining to control its reasoning. Its structure centers on the role of variables in the rule base with the variables within each rule set forming a hierarchical network of dependencies to obtain their values. GURU is written in C. The primary strengths of the GURU package reside in its overall integrated environment for system development.

3.2.1.2 Portability/Compatibility

FSC's development strategy for the KBMLS prototype requires long term maintainability and portability of the system software. Portability ensures that future versions of the KBMLS software may be hosted on a wide range of hardware platforms. This strategy permits a relatively inexpensive implementation of the KBMLS prototype on a PC based computer while providing for migration to higher capacity platforms without significant software development.

While all five shells are written in the C programming language, only NEXPERT OBJECT and KES specifically emphasize their portability and compatibility features. The KES developer stresses that one of the major benefits of using their shell was its modular structure. The benefit of the KES modular structure is the isolation of the hardware and operating dependencies of KES into a single, easily changeable module. As a result, KES executes on more hardware platforms than any other shell surveyed.

3.2.1.3 System Interface Capabilities

A major specified KBMLS requirement for the selected expert system shell is the capability to provide external product and language "bridges". An effective and efficient design of the automated sanitization process requires that the shell also possess the capability to be embedded within "conventionally" developed applications and programs. This embedding capability will allow the COTS expert system to be invoked as a subroutine and will permit a seamless integration between the expert system shell and the surrounding layer of conventionally developed software. The survey criteria includes an evaluation of expert system features for embedded integration with conventionally developed applications software. This criteria is

essential to the rapid development and success of the KBMLS prototype.

The EXSYS expert system shell is an example of a tool that permits data to be passed back and forth for analysis but lacks an integral structure. Data is passed by writing it to a disk file which EXSYS then reads. External programs can be invoked in a using a variety of methods with data passed in both directions. The evaluation, however, does not indicate that EXSYS does not provide the explicit capability to be integrated and embedded within conventional code.

The NEXPERT OBJECT and KES shells have been specifically designed to allow easy integration into existing computing environments. NEXPERT can communicate with as well as be controlled by external programs. It can also call user-written procedures and pass parameters to these procedures. Within NEXPERT, these external procedures are called from Execute statements. NEXPERT allows applications to communicate directly and dynamically during the inference process with relational databases. NEXPERT also provides a run-time library in "C" to enable NEXPERT based applications to be seamlessly embedded within other system applications. The shell is designed for complete integration of AI applications into operational environments based on programs or processes written in C, Fortran, Ada, Cobol, Pascal, assembly language and others.

The KES shell also provides extensive flexibility for integrating knowledge-based systems with other conventionally developed software applications. KES offers three distinct methods of integrating knowledge-based applications. The "Embedded Interface" is the KES facility that allows a knowledge-based application to be run as a module of a controlling "C" or other source program. KES provides up to 150 functions that allow direct access to internal KES data types and commands. The KES "Externals facility" allows a knowledge based system to execute other software applications and to communicate with these applications via text files. The final method is the LINK facility. The LINK is a KES facility that integrates ORACLE or other relational databases with any KES knowledge base.

The COSMIC CLIPS shell also provides a mature capability to be integrated and embedded within other conventionally developed programs. The CLIPS shell allows it to be integrated with other C programs as well as software written in other languages such as FORTRAN and Ada. CLIPS includes a math library which can be accessed by other C programs provides features which allow it to be manipulated externally. The COSMIC CLIPS product also includes source code which enables developers to incorporate unique developed C routines into the shell software. An expert system developed with COSMIC CLIPS can easily be embedded within other C programs and be called as a subroutine. The manual provided with the COSMIC CLIPS package includes extensive information about embedding CLIPS within other system and the advanced programming guide describes how CLIPS can be called from a program written in virtually any language that can make external language calls.

The design approach used by the GURU expert system shell provides an expert system shell within a common development environment. GURU is an integrated tool made up of an

expert system shell, a relational database, a general purpose text processing ability, graphics capabilities and several other general computer processing functions. GURU provides a reasonably complete, self-contained development environment but does not easily provide for external bridges and cannot be easily embedded within other developed applications. GURU can interface with external databases and spreadsheets through Data Interchange Format (DIF) files. The ability to embed GURU within another conventionally developed system is not a feature provided with the package. GURU's approach is to integrate several different applications into a common development environment.

3.2.1.4 Vendor Evaluations

In addition to evaluating the technical merits of expert system shells, criteria were established to measure vendor technical support, vendor viability (long term survival), and market acceptance (defacto standards). The survey considered such things as the number of systems sold, the number of actual fielded systems, and the number of incremental releases over the product life. Incremental releases were used as a gauge to determine the type of support and enhancements the shell might receive in the future. The survey also considered training, documentation, consulting capabilities and the length of time that vendor organizations have been in business.

EXSYS, developed and marketed by EXSYS, Inc., was first released in 1983. Since that time the shell has gone through three major revisions, with the current version 3.1 having been released in July 1987. The EXSYS corporation is located in Albuquerque, New Mexico and has provides free telephone technical support. Independent product reviews indicate that the company has a solid reputation and that EXSYS is in wide use. EXSYS, Inc.'s user telephone support is very thorough and their support employees have a through understanding of the product.

Neuron Data Inc., the developer of NEXPERT OBJECT, is also a small rapidly growing company. Neuron Data was incorporated in 1985 and released the first version of NEXPERT OBJECT in the fall of that year. Since that time the company has released four major revisions to NEXPERT OBJECT and plans to release its latest version (V2.0) in the summer of 1990. The company has indicated that they have sold and distributed over 7000 systems. A distribution arrangement with Digital Equipment Corporation (DEC) permits customers to buy the tool and obtain support from DEC and Bechtel Information Systems (training support). Neuron Data provided numerous articles praising the capabilities of NEXPERT OBJECT and produced an extensive list of applications in use by noted corporations around the world.

Software A&E, the developers of KES, have identified themselves as the "oldest profitable expert system shell developer" in the expert system arena. The company introduced its first expert systems building tool, KES I, in 1983. The tool was initially developed in LISP, but was completely re-engineered in "C" in 1985 to provide more flexible integration capabilities. Since being re-engineered, KES has gone through seven major revisions. Software A&E advertises that

a new version of KES is released at least once a year and that each new release is both upward and downward compatible. KES is Software A&E's major software product and thus receives a corporate emphasis to maintain it as a leading expert system product. Because of the modular structure of the KES shell, Software A&E is able to add new and improved functionality to the shell very easily and quickly. The company offers consulting services and builds expert systems for numerous customers, including government agencies. Software A&E provides full support for KES licensees including telephone consultation, software updates and training courses whic are offered on a regular basis. Software A&E's headquarters and training/development facility located within 20 miles (in Arlington, Virginia) of FSC's development facility.

The COSMIC CLIPS shell is a product of the AI section of the National Aeronautics and Space Administration (NASA) at Johnson Space Center in Houston, Texas. The original development goal was to build a highly portable, low-cost expert system tool that could be easily integrated with external systems. The result was CLIPS (C language Integrated Production System). CLIPS is currently in use at several NASA locations and is distributed to the private industry and academic sectors through its nonprofit unit Computer Software Management and Information Center (COSMIC), located at the University of Georgia. COSMIC has served the NASA agency for twenty-five years distributing NASA-developed software. Technical support for COSMIC CLIPS is not as complete as the commercial packages, but is provided at no charge. No information was available on planned enhancements or future technical support.

GURU is a product of Micro Data Base Systems, Inc. (MDBS) located in Lafayette, Indiana. The company has been in operation for several years and is best known for its MDBS III and KnowledgeMan programming packages. Most of MDBS's products are sold to developers for specialized applications. GURU was first released in January of 1985 and since that time has undergone three major revisions. A reviews of GURU documentation indicates that while massive, it is poorly organized and difficult to use. The documentation intersperses information on the expert shell with instructions pertaining to the other components of the integrated package. GURU manuals are primarily oriented toward persons already familiar with MDBS's other products. The company does offer three-day workshops which focus on the expert system component of the package.

3.2.2 Conclusions

Based on specified and derived KBMLS requirements and on the preliminary architecture of the demonstration system, the original population of twenty-five expert system tools was reduced to eight. Two major criteria used to produce this population reduction were: selection of UNIX as the KBMLSI operating and; the contractually specified requirement to build the demonstration system with a minimal amount of new software development. Of the eight remaining expert system shells, three (OPS-2000, ESP ADVISOR, and CxPERT) were eliminated based on the absence of sufficient documentation or other literature which could substantiate their technical features. From the information available, an assessment was made that these four systems would not provide the necessary functionality required for development of the KBMLS

prototype.

Table 3.2.1-2 summarizes the ratings given to each of the five shells which were evaluated. The ratings are reflective of the varied capabilities and features of each shell. An overall score was calculated for each shell and was used in the final recommendation and selection. Scoring is based on a scale from zero to ten with ten representing full satisfaction of criteria requirements.

In evaluating the power and flexibility of knowledge representation and inferencing, both KES and NEXPERT received the highest scores. Both possess powerful knowledge representation mechanisms and both provide multiple ways of inferencing with knowledge. Based on the complexities of the sanitization problem, evaluation criteria established power and flexibility of representing and inferencing with knowledge as critical factors in the development of the demonstration system.

The ability of a shell to be embedded within other conventionally developed software was also a critical evaluation criteria. The KBMLS prototype architecture is predicated on the ability of the expert system to be embedded within conventionally developed software. The ability of the shell to provide conventional programming language interfaces was determined to be essential in building a seamless, integrated system. Three of the five shells evaluated were specifically designed with this capability. KES, NEXPERT and CLIPS all possess the characteristic of easy integration into existing computing environments. All three provide the necessary features for successful development and demonstration of the KBMLS prototype. The EXSYS and GURU expert systems do not specifically provide these capabilities and, as a result, received lower scores. While EXSYS and GURU do provide facilities for "bridges" to external programs and data, they do not provide the means to completely and seamlessly integrate an AI application with conventionally developed software.

In evaluating portability/compatibility two of the shells were found to have a clear advantage. KES and NEXPERT OBJECT specifically emphasize their portability and compatibility features. The modular structure of KES is particularly appealing to the KBMLS development strategy for migration to other hardware suites and operating systems. The KES design indicates that the company is positioned to provide frequent upgrades to their expert system, and is consistent with their commitment to release a major revision of KES at least once a year. Based on these criteria KES received the highest rating in this category.

The final criteria, vendor evaluations, was used to rate the quality, reputation, long term stability and technical support of the product. The KBMLS development requires a shell that will not be discontinued, scrapped or no longer supported in the near future. This criteria was also used as a gauge to possible future enhancements to the existing shell. The number of product updates provides an indication of that companies commitment to continuously improve their product and to stay on the leading edge of expert system technology. All five shell developers were evaluated as having quality products and solid business reputations (i.e. none were

identified as "high risk" corporations).

3.2.3 Recommendations

KES and NEXPERT received the highest overall total scores. The final recommendation and selection was ultimately a choice between these two shells. The survey results indicate that both shells possess the necessary capabilities for successful development and implementation of the KBMLS sanitization prototype.

Based on a weighted analysis, FSC recommends KES as the expert system shell for development and implementation of the KBMLS prototype. KES was selected over NEXPERT on the basis of four major items:

(1) Modular structure. One of the major benefits of KES is the isolation of the hardware and operating dependencies into a single, easily interchangeable module. As a result, KES runs on more hardware platforms than any other shell available on the market. The modular structure also allows new functionality to be added very easily.

(2) Embedded interface. KES allows the expert system to be easily embedded into other applications and programs operating as integrated modules under the control of a "C" or other source program. KES offers extensive functions (over 150 functions) that allow direct access to internal KES data types and commands.

(3) Power and flexibility. KES's power and flexibility of knowledge representation and inferencing were felt to be slightly superior to NEXPERT OBJECT's. Because of the uniqueness and complexity of the automated sanitization process the more power and flexibility in knowledge representation and inferencing, the lower the software development risk. The KES shell offers the advantages of having backward and forward chaining, and also provides three types of inference engines.

(4) Vendor proximity. Software A&E's headquarters and training facilities location was the contributed to the selection of KES. Because of the inherent complexity of generating the expert sanitization rules and in integrating the expert shell with KBMLS applications software, vendor proximity and support were established as risk reducing evaluation criteria. FSC developers should be able to receive better technical support with Software A&E being located less than 20 miles from FSC's development facilities.

3.3 Document Parsing Tools Survey

The following subsections summarize the evaluations, draw conclusions, and present a recommendation.

3.3.1 Evaluations

The KBMLS prototype requires the capability to scan narrative text messages, parse message elements, and store and disseminate information from the narrative relating to the questions: WHO, WHAT, WHERE, WHY, WHEN and HOW MANY. A major concern of the evaluators was the tool's ability to understand and parse large lists of specific words, symbols and phrases as well as its ability to handle the syntactic and semantic variations of reported items of interest.

The survey identified a number of parsing tools which provide a subset of the allocated system requirements. The evaluation criteria of the tools were:

- The ability of the COTS document parsing tools to perform keyword/phrase extraction
- Vocabulary size
- Syntactic and semantic coverage capabilities
- The ability of the COTS tool to be embedded within the system and provide a seamless integration between the document parsing tool and the other modules of the system
- Compatibility with the hardware, operating system and surrounding application software
- The ease of use, supplied interfaces and available documentation.

The majority of the tools evaluated fell into two broad categories, natural language programs and text/information retrieval tools. One additional specialty tool was also identified and evaluated. The developers of this tool categorize it as a natural language "shell"; or as a hybrid of the other two categories.

3.3.1.1 Natural Language Programs

The natural language programs were the more readily available tools reviewed. Most of the natural language programs reviewed were tools which are used to develop "friendlier" interfaces to databases and other software applications (See appendix B). For example, the BBN Parlance Natural Language Interface, developed and marketed by BBN Systems and Technologies Inc., allows users to obtain information from various relational databases by simply typing in questions in every day English. The software parses a data input string, interprets the parsed data, and translates it into SQL statements. These natural language programs supply the necessary linguistic and syntactic capabilities and allow application software developers to add words and terms to the tool dictionary. The dictionary contains all those words and phrases related to the developers specific application, making it domain-specific.

Most of these tools allow modifications to accommodate other types of applications in addition to front-end interfaces, but the majority are specifically designed for just this one purpose. Manipulating and/or modifying of these programs for use in other applications, such as the KBMLS system, is difficult to accomplish. The inability to embed these natural language programs represents the greatest roadblock. These tools, as previously described, have been specifically designed for end users and do not provide features for integration with other applications. Independent reviews and other available literature do not indicate that any of the natural language packages would be able to accomplish the task required by the design.

3.3.1.2 Text/Information Retrieval Tools

The other set of document parsing tools surveyed were categorized as text/information retrieval tools. These tools have been specifically developed to manage and manipulate large amounts of unstructured narrative text. Most users of these tools are concerned with rapidly locating word/word phrases and concepts contained in large volumes of narrative. While that application does not match a specific KBMLS requirement, the tool's ability to identity and parse word/word phrases and concepts, provides the possibility of using these tools document parsers. The three tools evaluated in this category were Elexir, Golden Retriever and Fulcrum's Text/Search (See appendix B).

Like the natural language programs, these tools are single purpose, text/document retrieval systems. None of the tools in this category demonstrate the ability to perform syntactic or semantic operations nor are they easily embedded within other applications. The tools are not flexible or versatile enough to address many of the requirements specified by the KBMLS design. While both categories of parsing tools possess a varied degree of capability to perform keyword/phrase extraction, none are able to be integrated with the KBMLS prototype.

3.3.1.3 NL Builder

Of the potential document parsing tools reviewed, one displayed the capabilities of the natural language programs and text retrieval tools, plus the flexibility to modify its inherent functions. NL Builder (Sychronetics, Inc.) is unlike the other natural language packages in that it allows the developer to define a specific natural language process without programming. In other words, a developer can build a domain specific natural language program. This task is reduced to that of building a lexical database to describe the target language or sub-language ("jargon"). NL Builder provides capabilities for syntactic and semantic analysis of sentences. Syntax and semantics are interleaved at a granularity of the developer's choice. The developer is required to build the lexical database consisting of words, word senses, features on word senses, syntactic grammar and semantic representation and the tool completes the process. Sychronetics emphasizes that NL Builder is a Natural Language shell, not a natural language program. It is written entirely in the C language and provides an extensive C language interface allowing for easy integration and manipulation in the application software environments.

3.3.2 Conclusions

After identifying and evaluating the two generic types of tools; Natural Language and Text/Document Retrieval tools, an evaluation of the tool population determined that the sampled commercially available tools for document parsing do not meet the required specifications of the KBMLS prototype. One tool that did represent potential for integration with the KBMLS prototype is NL Builder.

NL Builder is advertised as a Natural Language Shell versus a natural language program. Its primary advantage over the other surveyed tools is that it is not specifically designed as an end user product. NL Builder has the capability to be integrated and embedded into other applications and is flexible enough to allow the an application software developer to adjust the tool to handle specific application.

3.3.3 Recommendations

Based on the absence of other tools which could provide the necessary KBMLS functionality, NL Builder is identified as the only parser possessing a sufficient number of the required features needed in the development of the KBMLS prototype.

4. KBMLS SANITIZATION PROTOTYPE DESIGN

4.1 General

Implementation of the Sanitization Model resulted in the development of a single Computer Software Configuration Item (CSCI) known as the Knowledge Based Multi-level Secure (KBMLS) Sanitization Prototype. The following subsections presents the design.

4.2. KBMLS CSCI Overview

The Knowledge based Multi-level Secure (KBMLS) CSCI supports the mission of USAFE by providing an accredited interface for direct electrical exchange of information between USAFE's TS/SCI level (or "high") intelligence production centers and a SECRET level (or "low") network. This network is the Intra-Theater Intelligence Communications Network (IINCOMNET). Given tasking from the operator, the KBMLS CSCI either automatically or interactively identifies intelligence data of interest to tactical wing commanders and extracts the data from the "high" side, sanitizes the data, and distributes the data to the "low" side.

KBMLS is capable of providing automatic identification, sanitization, and product generation. In addition, KBMLS is capable of automatically categorizing sensitive information into one of the following categories for operator review.

- Immediately Releasable (requires no sanitization)

- Sanitized Releasable (has been sanitized)
- Conditionally Releasable (sanitization is authorized only if conditions are met)
- Not Sanitizable (sanitization not authorized)

KBMLS is designed to run on a UNIX System V Operating System, with OSF/MOTIF X-Windows, hosted on a DATAWATCH 386 Workstation. The target external components of the KBMLS Computer Architecture are shown in Figure 4.2 and identified as:

- USAFE IDHS - United States Air Force in Europe (USAFE) Intelligence Data Handling System (IDHS)

- KBMLS - Knowledge Based Multi-level Secure Sanitization Prototype

- USAFE GUARD - SCI to NATO Gateway system.

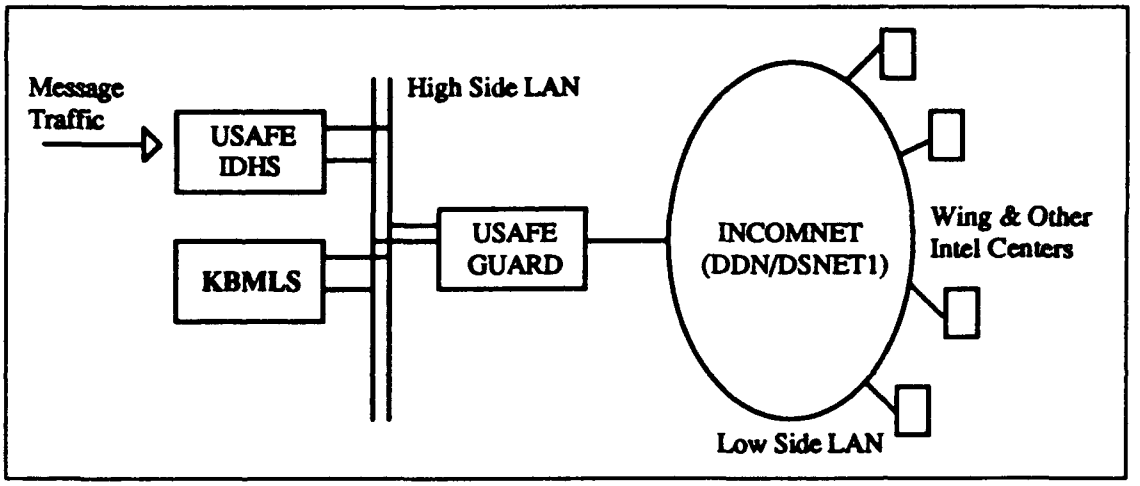

Figure 4.2 KBMLS Prototype within USAFE Computer Architecture

The KBMLS supports the mission of the KBMLS Architecture by performing high interest activity screening, associating new information to existing information, sanitization of critical intelligence information, and dissemination of sanitized intelligence information either automatically or at an operator's discretion.

The purpose of the interface to the USAFE IDHS is to receive message traffic (for example TACREPS, TACELINTS, IIRs); query and receive related messages from message queues; query and receive up to date information on routes and schedules of reconnaissance aircraft; query and receive Order of Battle (OB) data Base records for Air and Missile; and send OB update records back to the USAFE IDHS whenever update criteria has been met.

The purpose of the interface to the USAFE GUARD is to disseminate sanitized intelligence information to wing commanders. The GUARD performs a redundancy check on the classification and data to ensure the intelligence information can be disseminated to wing commanders.

4.3 KBMLS CSCI Architecture

The internal organizational structure of the KBMLS prototype is depicted in FIGURE 4.3. As shown, the KBMLS is composed of four Computer Software Components (CSC):

- Data Translation Component - Translates sensitive information into Sanitization Case Folders
- Sanitization Component - Determines need, sensitivity, cover, and method of sanitization, then generates sanitized information.
- Executive Control Component - Provides system control and system security.
- Global Utilities Component - Provides interface routines to system services, external i/o, data base, and display i/o.

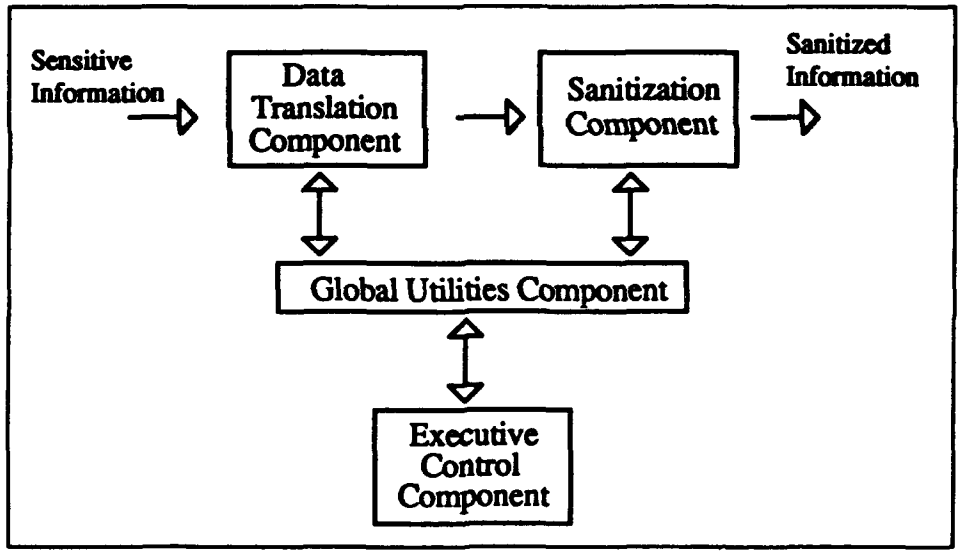

Figure 4.3 KBMLS Prototype Diagram

The purpose of the Data Translation Component is to perform automatic receipt, parsing, and storage of TACREP and TACELINT. Then route sensitive information (of high interest to tactical wing commanders) for further processing to the Sanitization Component to establish links to existing intelligence information which will be used in the sanitization process. In addition, it provides for the maintenance and tailoring of data translation rules. To identify and understand

incoming messages, a parsing technique called Translation Grammars was selected to control the parsing and translation of incoming messages into an internal data structure. The Translation Grammar Technology was furnished by the Army through FSC's involvement with software development for the Common ATCCS Support Software.

The Sanitization Component determines whether the sensitive information can be sanitized and if so, performs sanitization of sensitive information for potential dissemination to wing commanders. The decision making is implemented using Artificial Intelligence techniques. The Knowledge Base Expert Systems (KES) Commercial Off The Shelf (COTS) software is used to control the sanitization process. The sanitization component provides a set of functions to allow analysts to review, generate, update, and disseminate sanitized intelligence information or perform those function automatically. In addition, it provides for the maintenance and tailoring of sanitization tables and knowledge base.

The Executive Control Component performs system services and provide a multi-level secure environment. It controls and monitors system resources and provides maintenance and tailoring of system and security data.

The Global Utilities Component is a set of libraries which contain software common to all other components. It performs; internal/external communications, system services, and common display services. System services to manage of file and files with data structures is implemented through calls to C B TREE an A B-tree file management system. To keep inline with current trends and standards in human-computer interaction, XSIGHT an OSF/MOTIF X-Window Run-time system. was employed so the results of KBMLS automated sanitization can be reviewed and modified by authorized site personnel.

4.3.1 KBMLS Internal Interface Data Structures

KBMLS internal interface common data structures are stated below;

- Sanitization Case Folders is a complex record structure which contains a complete audit trail of events, decisions, and manipulations that occurred during the sanitization process.

- System History log packets and Security Audit Trail log packets are log entries and passes by modules to logging software. Each entry is and 80 character string with indicates either a system level event or a security relevant event.

- Message Log packet is a collection of the messages and summary record which is passed by modules to logging software for storage in KBMLS logs.

- System control messages are used to pass low level commands to software for startup and shutdown.

4.3.2 KBMLS System States and Modes

The KBMLS system has two system states and within the interactive state has three modes of operations, as defined from and end users point-of-view. The system states and modes are defined as follows:

- Automatic Mode is when the system runs automatically (without continuous monitoring by users) parsing messages, researching, and sanitizing incoming message traffic for later review by analysts.

- Interactive Mode is when some user (Analyst, System Manager, or Security Manager) is logged in. While in the interactive mode the system can be in one of three states; Analyst, System Manager, or Security Manager.

    - In the Analyst state the following activities can be performed: review case folders; release sanitized messages; update the Order-of-Battle Data base; resubmit case folders for reprocessing; and review/release daily and monthly products.

    - In the System Manager state the following activities can occur: startup and shutdown of the system; monitoring execution of the software, logging, and disk usage; and perform maintenance tasks on the case folder data base, knowledge base, criteria files, and logs.

    - In the Security Manager state the following activities can occur: update user access table; and review and archive the security audit trail log.

Only one state can be active at one time and both the automatic mode and interactive mode maybe active at the same time.

4.4 Module Descriptions

The components of the KBMLS prototype represent a collection of modules and routine libraries. Each component, with its associated modules and/or libraries is presented in Figure 4.4.

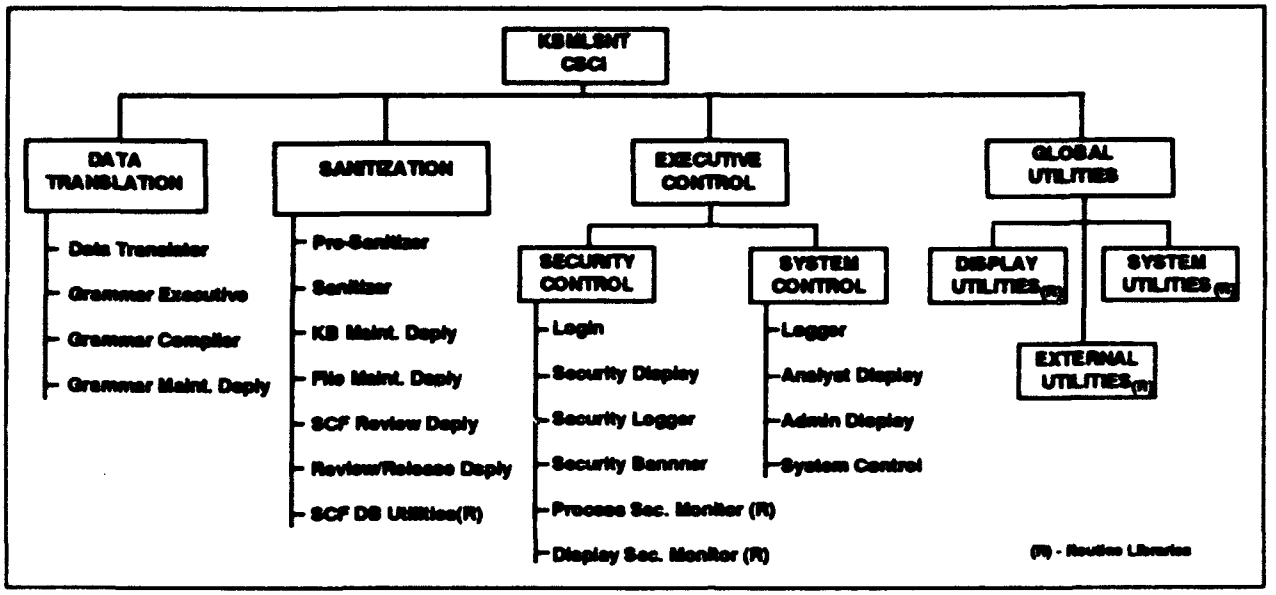

Figure 4.4 KBMLS Hierarchy Diagram

4.4.1 Data Translation Component Module Descriptions

The Data Translation component provides the functionality to parse and process incoming information. The component is comprised of four modules: the Grammar Maintenance Display to modify parsing grammars, the Grammar Compiler to generate rules sets, the Data Translator to serve as the automatic driver for processing messages, and the Grammar Executive to parse incoming messages. Figure 4.4.1 depicts the interfaces.

4.4.1.1 Data Translator Module

The Data Translator (DT) contains the functionality to receive messages from an external source, perform all the automated parsing of incoming messages and forwarding of sanitization case folders to the pre-sanitizer. In doing so, the DT module invokes the Grammar Executive to parse the messages. The DT module performs the following functions:

(1) Generation of sanitization case folders based on messages for which parsing rules are available.

(2) Creation and updating of disk-resident databases based upon analyst interaction.

(3) Duplicate message processing.

(4) Assign new security classifications to each sanitization case folder based upon message classification.

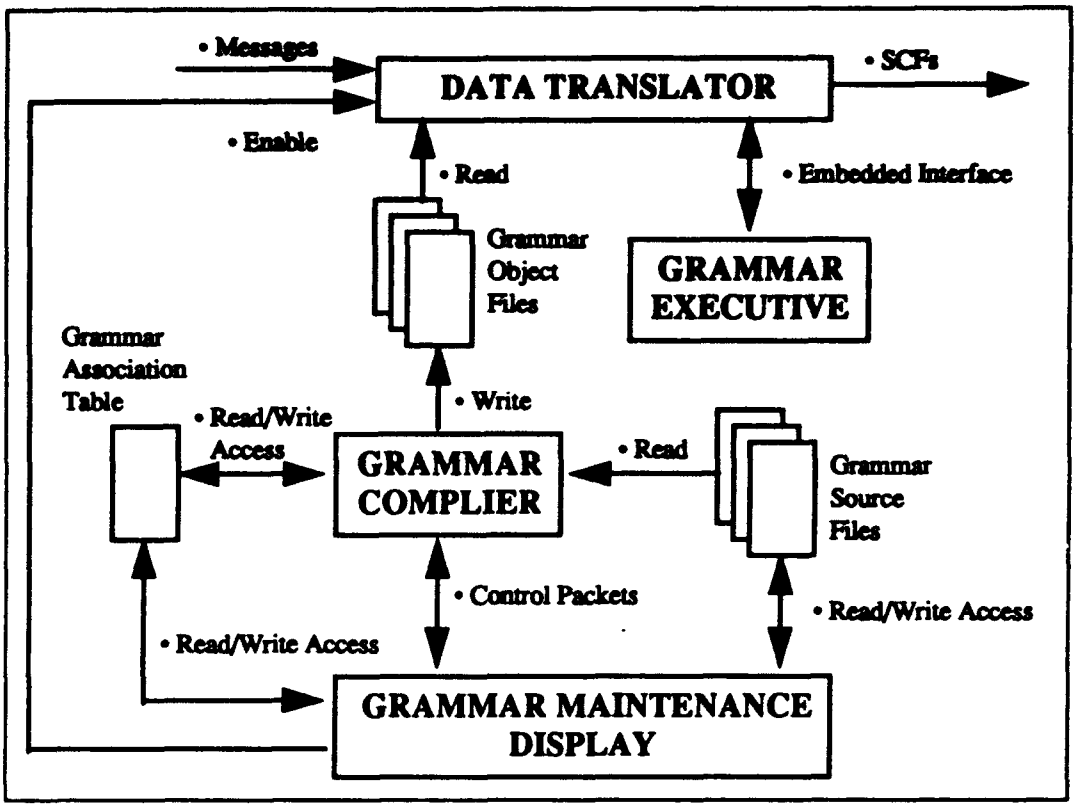

Figure 4.4.1 Data Translator Component Module Diagram

The DT process loads all of the parsing data used by the automatic parser into memory. This data includes applicable subsets of the Parsing Definition Tables. During normal operation, no disk accesses are performed by the DT software, thereby maximizing message processing throughput. After a message has been received, it is parsed via the Grammar Executive with the Message Header Rules to determine whether the message is a TACREP or a TACELINT. Message header information is extracted from the message and is stored in the sanitization folder and the message abstract. After the message header parsing, the message is checked to ensure that the message is not a duplicate of a previously received message. Each time the DT is started, certain information for the most recent 1200 messages received while the process has been active is stored in memory for as long as the process is active. If the message matches a previous message, it is not forwarded to the sanitizer. After the message type has been determined, the rules pertaining to the individual message type are utilized to parse the message. The Grammar Executive extracts information from the message and stores it in the sanitization case folder.

4.4.1.2 Grammar Executive Module

The Grammar Executive consists of a general parsing utility to extract information from a message based upon rules defined for the format of the message. The parsing tables are generated by the Grammar Compiler. The grammar executive is a utility procedure invoked by the Data Translator to parse information. The Grammar Executive utilizes parsing rules containing message format rules to parse incoming messages. Grammar rules are used for input message processing, and combine the lexical analysis and syntactic analysis tasks for parsing messages. Based on the grammar rules, the Grammar Executive extracts and formats the information that is to be put into a sanitization folder.

4.4.1.3 Grammar Compiler Module

The Grammar Compiler process generates rules for the Grammar Executive to use from english language rules created by the analyst via the user interface. Two types of tables comprise the databases: Parsing Rules and Reference tables. These tables are used to extract information from incoming messages during automatic parsing. The Grammar Maintenance Display creates and edits parsing rules. The Grammar Compiler process transforms parsing rules into the format necessary for use by the Grammar Executive. The User Interface process provides the capabilities to edit, delete, print or create rules used in parsing incoming information. These rule sets include: header parse rules, message format rules, data set rules, field level rules and common rules. This module performs the processing required to transform p-meta rule statements into parsing tables in meta object format. These tables are used by the grammar executive to perform the parsing and validation of information.

4.4.1.4 Grammar Maintenance Display Module

The Grammar Maintenance Display allows the user to manipulate the KBMLS input message parsing grammars. The user can add, edit, delete, compile, enable, and review grammars, and restore grammar file directory baselines. The compiled grammar objects are used by the data translator to parse the incoming message traffic and are on-line modifiable to allow making changes to support new message formats without requiring a rebuild of the KBMLS software. The Grammar Association Table (GAT) maintains the grammar name, the grammar dependencies, and the type of grammar.

4.4.2 Sanitization Component Module Descriptions

The Sanitization Component provides the functionality to sanitize sensitive information and format the sanitized information into an outgoing product after passing a set of rule based authorization logic which either authorizes or denies sanitization. The component is comprised of six modules and one routine library for access to the scf data base. The six modules are the Pre-Sanitizer, the Sanitizer, the SCF Review Display, the Review/Release Display, the Knowledge Base Maintenance Display, and the Sanitization File Maintenance Display. Figure 4.4.2 depicts

the modules and the interfaces among them.

4.4.2.1 Pre-Sanitizer Module

The Pre-Sanitizer determines *if* and *how* the message will be sanitized. The Pre-Sanitizer retrieves the SCF from the Data Base and searches to establish links to existing "LIKE" SCFs. A null linking results in a New Unit and continues processing. A duplicate check determines if the information in the SCF is New or duplicate. The Pre-Sanitizer contains an In_Area Algorithm. Area tests are provided for circles, rectangles and polygons.

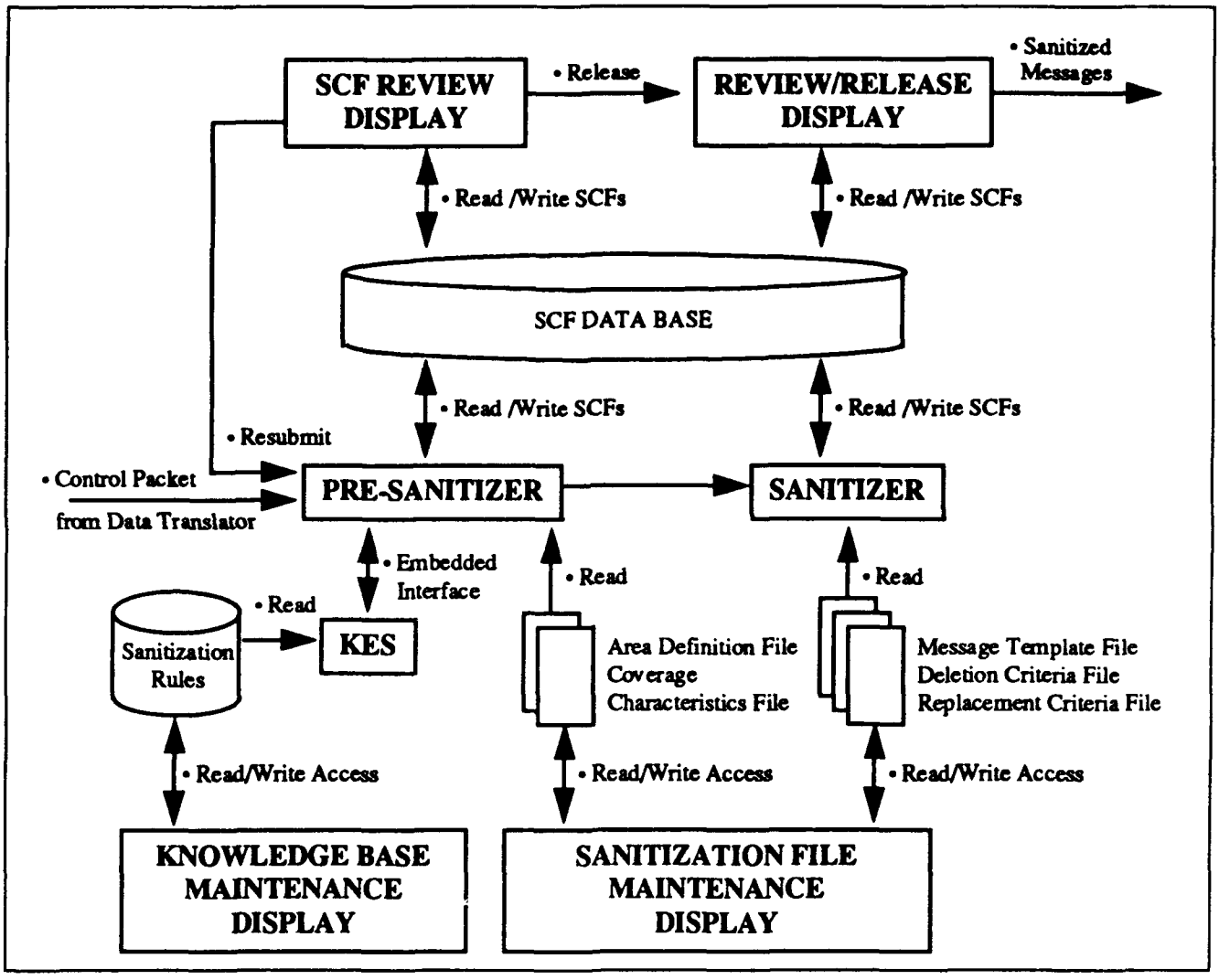

Figure 4.4.2 Sanitization Component Module Diagram

The Pre-Sanitizer interacts with the expert system by executing a script which backward chains through four sets of sanitization rules. KES is the Knowledge Based-Expert System shell.

Those rule sets are:

- Need Rules specify who has interest as it relates to the need to know. Rules are constructed based on When (time), Where (location), What (activity), and Who (unit) Information. These rules yield a list of addressees.

- Sensitivity Rules pertain to message type, classification and classification markings, source, method, and special conditions such as; confirmation needed or agreements with host nation. These rules can prevent information from being sanitized.

- Coverage Rules are divided into Allowable Source Rules and Preferred Cover Rules, which surround the Plausible Cover Algorithm. Within the Plausible Cover Algorithm a Source Equivalency Match function searches the Coverage Table for units which have similar sources and creates the Candidate Coverage List. An Area/ Time Match Function attempts to determine from the Candidate Coverage List what candidates were in the area at the Time of Intercept. The candidates which were in area at the right time are entered into a Suitable Coverage List. A Status of Forces Check Function accesses status information in the Coverage Table to determine which units qualify as either Plausible Cover or Reasonable Source. Attributes from the Sanitization Case Folder are used with the Preferred Cover Rules to select the Preferred Cover.

- Method Rules are used to determine the extent and format of sanitization. Such as; accuracy, deletion and replacement criteria, and the Format of Sanitized Information. These Rules are driven by attributes populated by the Need, Sensitivity, and Coverage Rules.

The Hold Test examines each "LIKE" SCF to determine if the new SCF satisfies any Hold Conditions such as; Needs Less Sensitive Source or Needs Confirmation.

4.4.2.2 Sanitizer Module

The Sanitizer Module performs the sanitization and reformatting of sensitive information. It receives instructions from the Pre-Sanitizer module which govern the sanitization of sensitive information. The Sanitizer builds an instruction set using Sanitization Tables. The first pass uses the instructions to create the execute instructions. The second pass makes the deletions and replacements using the execute instructions. The following search algorithms are used;

- Search / Fill from Case Folder - Replace keywords in message template with information from a specific Case Folder

- Search / Delete - Search for a word or phrase and delete the word or phrase, sentence, or paragraph where it was found

- Search / Round - Search for a real number and round to any given precision

- Search / Replace - Search for a word or phrase and replace the word or phrase, sentence, or paragraph where it was found with another specified word or phrase.

Search capabilities include; Wildcard Searching, Case Sensitive Searches. Generalized Spell Checking, Recognizing Transposed Characters, and Simple Typo Recognition.

The Sanitizer module updates the Case Folder with the sanitized message which is then written to the Sanitization Case Folder database.

4.4.2.3 SCF Review Display Module

The SCF Review Display Module is called by the Analyst Display Module when the analyst selects one of four types of SCF reviews from the SCF Review option. The options are; releasable SCFs, sanitized SCFs, conditional SCFs and browse SCFs by unit name. During SCF Review, the analyst has the following available options:

- Releasable, an additional window allows the user to review SCF messages for each specific category or by unit name.

- Summary, a window will appear containing the various fields from the case folder for the analyst to review.

- Justify, the window shows the justification for the sanitization that has taken place for this message.

- Related, a window will be displayed showing all the case folders which have the same unit name as the currently displayed case folder.

- RELEASE, the Review/Release Display Module is activated and displays the Release window.

- Resubmits SCF, the case folder will be resubmitted to the KES knowledge base and any appropriate sanitization will be done.

- Resubmit MSG, the message will be resubmitted to the data translator to be reparsed.

- Store SCF, the SCF will be put into the stored category and will no longer appear in the releasable, sanitized, or conditional list. It still can be accessed as a related unit or through the browse option which is described in a later section.

4.4.2.4 Review/Release Display Module

The Review/Release Display Module is activated when the user clicks the RELEASE option from the analyst display. The analyst may either select to send the sanitized message for output (to USAFE GUARD) or to cancel and return to the previous level menu.

4.4.2.5 Knowledge Base Maintenance Display Module

The KB Maintenance Displays allows the system administrator to review, add, modify, and delete knowledge based rule set files. The knowledge files are used by the Pre-Sanitizer's KES COTS software to execute the knowledge based rules during processing. This display allows the system administrator to change the rules that govern the criteria for sanitization. The system administrator also activates the desired rule set file for automated sanitization processing.

4.4.2.6 Sanitization File Maintenance Display Module

The Sanitization File Maintenance Display Module is an option from the system administrator's main window and allows the system administrator to tailor the KBMLS system to interpret and sanitize message data accurately. From selections the user can modify the various support tables. File Maintenance provides the support to modify information used to direct the sanitization rules. The selectable tables include the Area Definition, Cover Characteristics, Deletion Table, Replacement Table, Message Templates, and Addressee Table. Details on each of the tables is documented in the sections that follow.

- Area Definition - Area Definition Table allows the system administrator to specify a list of Areas of Interest (AOIs) for the given site missions. From this list, it is determined if an incoming message is within an AOI and sanitization processing should continue.

- Cover Characteristics - Cover Characteristics Review allows the system administrator to record a list of available covers and their associated specification. This information is used to provide plausible and reasonable cover alternatives when sanitizing data.

- Deletion Table - Deletion Table Review window allows the system administrator to specify phrases that should be deleted from outgoing message product. The deletion table phrases are grouped by classification downgrading categories. The selectable classifications are Hi to Low, Low to Ext, and Low to Gen. The user can add a new phrase or modify an existing phrase.

- Replacement Table - Replacement Table Review window allows the system administrator to specify phrases that should be replaced with a different phrase in outgoing message products. The replacement table phrases are grouped by classification downgrading categories. The selectable classifications are Hi to Low, Low to Ext, and Low to Gen. Regardless of which classification is chosen, the Replacement Table Review window is

displayed. The user can add a new phrase set or modify an existing phrase set. A phrase set consists of the Search Phrase applied to the input message and a Replacement Phrase inserted in the output message product.

- Message Templates - Message Template Review allows the system administrator to create a list of output message templates. From this list the analyst is allowed to choose a desired format for the released sanitized message output products. The user can add a new message template format or edit an existing format from the Message Template List.

- Addressee Table - Addressee Table Review window which allows the system administrator to specify a list of addressees that are acceptable for message release. This list is used by the software to determine if an input message is of interest. The user can add a new addressee or edit an existing addressee from the Addressee List.

4.4.3 Executive Control Component Module Descriptions

The Executive Control Component is divided into two subcomponents; The Security Control Subcomponent and the System Control Subcomponent. The Security Control subcomponent provides mandatory and discretionary access controls, user login authentication, security audit trail, screen sensitive labels, and maintenance and review functions. The System Control subcomponent controls the startup and shutdown of KBMLS, along with managing the message logs and system history logs.

4.4.3.1 Security Control Subcomponent

The modules which comprise the Security Control subcomponent are depicted in Figure 4.4.3.1. They are: the Login module which provides mandatory and discretionary access controls, and user login authentication; the Security Logger Module which receives security related logging event messages and maintains security audit trail; the Security Banner Module which provides screen sensitive labels; and the Security Display Module which provides the security manager with maintenance and review functions.

The Security Display module and its associated displays appear when a valid user id and password have been entered. The bar across the top of the window displays the system classification. The second bar under the classification bar is the main menu bar and defines the options available to the security manager. They are:

- USER TABLE option allows the security manager to add, delete, or modify KBMLS user accounts.

- SECURITY REVIEW option from the security manager's main window allows the review of the Security Audit Trail Log (SATL).

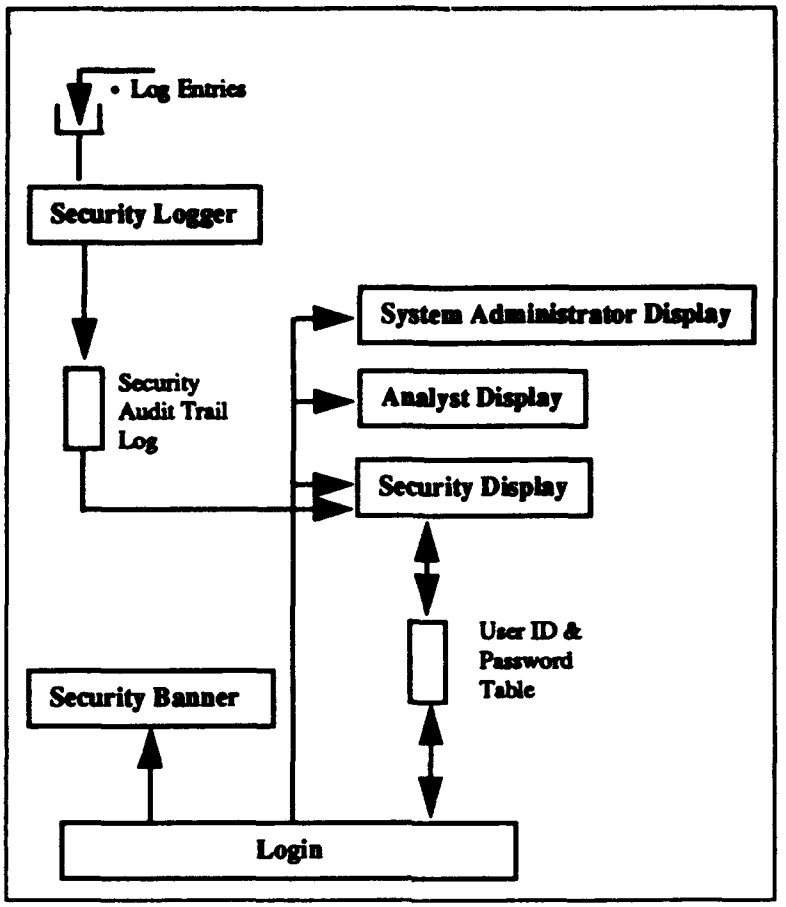

Figure 4.4.3.1 Security Control Subcomponent Modules

4.4.3.2 System Control Subcomponent

The modules which comprise the System Control Subcomponent are depicted in Figure 4.4.3.2. They are: the System Control Module which controls the startup and shutdown of KBMLS; the Logger Module which manages the message logs and system history logs; the System Administrator Display Module which provides the system administrator with access to logs, system control functions, and maintenance functions; and the Analyst Display Module which provides the analyst with access to Sanitization Case Folders.

The System Administrator Display Module is activated when a system administrator enters their user id and password. A main menu bar defines the options available to the system administrator. Those options are:

- SYSTEM CTRL option allows the system administrator to start and stop the automatic sanitization processing on incoming message data. In addition, this option permits the system administrator to monitor the KBMLS software by reviewing which software

processes are running. There are three available selections under this option: Sanitization Startup, Sanitization Shutdown, and System Monitor.

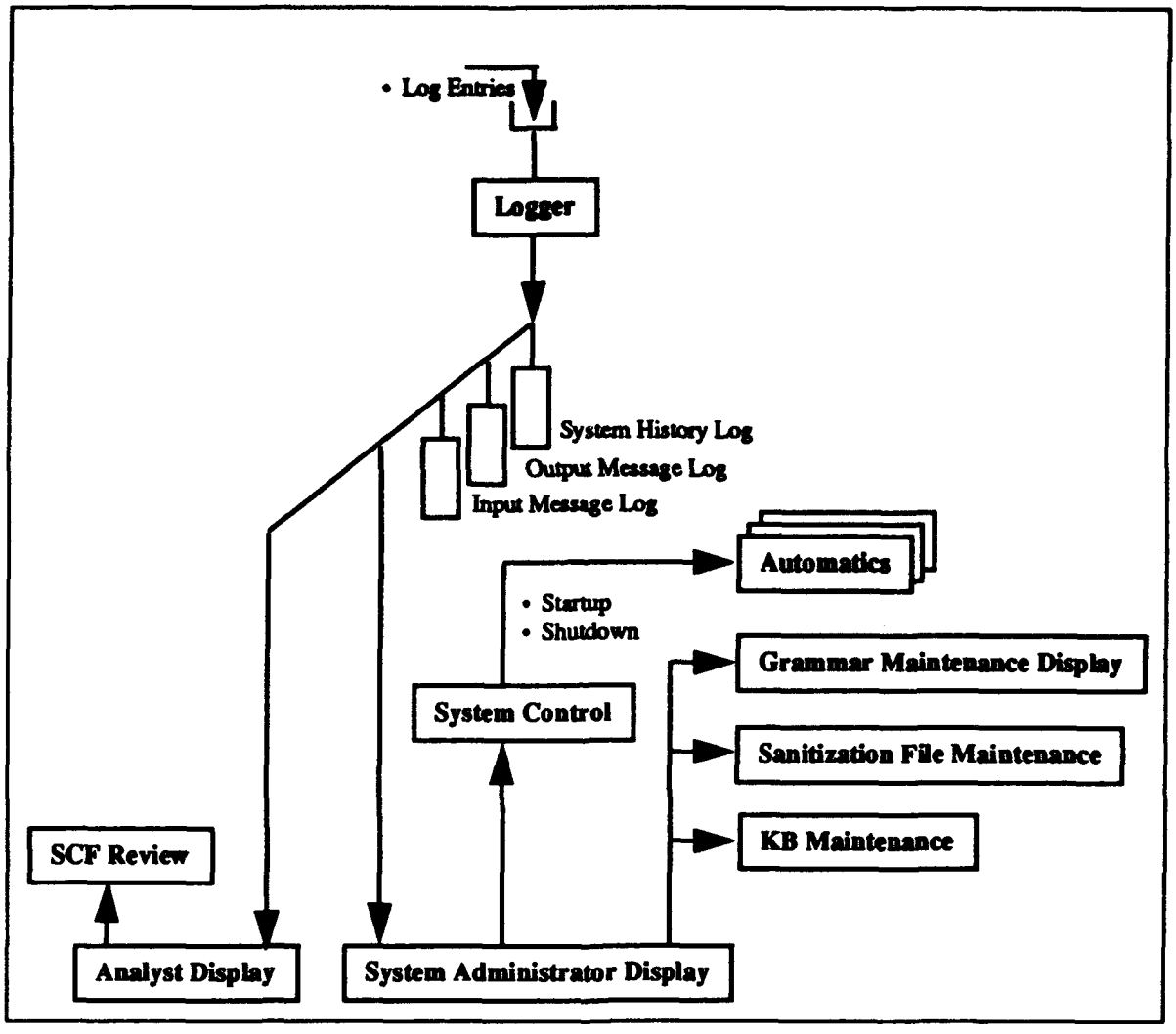

Figure 4.4.2.2 System Control Subcomponent Modules

- LOG REVIEW option allows the system administrator to review the set of KBMLS logs. The history log, input message log, and output message log are the reviewable logs.

- MAINTENANCE option from the system administrator's main window allows the system administrator to tailor the KBMLS system to interpret and sanitize message data accurately. The allowable user selections are: Table Maintenance, Grammar Maintenance, and KB Maintenance. From these selections the user can modify the various support tables, modify and enable input message grammars, and maintain knowledge base rule sets. From this option the system administrator initializes and critiques KBMLS tables and

files for the automatic sanitization processing operations.

- SYSTEM ADMINISTRATOR LOGOUT AND SHUTDOWN option on the system administrator main menu bar closes the system administrator's window or shuts down the KBMLS software.

The Analyst Display Module is activated when the role associated with the valid user name and password is an analyst. The main function provided to the analyst is the review of sanitization case folders. Access to SCFs is accomplished through the SCF Review option. The Module supports the analyst in selecting one of four types of SCF reviews: releasable SCFs, sanitized SCFs, conditional SCFs and browse SCFs by unit name. After selection, the Analyst Display Module activates the SCF Review Display Module. The Analyst Display Module also provides access to the message logs.

4.4.4 Global Utilities Component Module Descriptions

The Global Utilities Component is comprised of three separate callable libraries of routines which provide interfaces to networks, interprocess communications, and windows. They are; Display Utilities, External Utilities, and System Utilities.

4.5 Overview of KBMLS Operations.

KBMLS operations are divided into three stages of operation. They are Startup/Login, Automatic Processing, and Interactive Processing. The following describes the processing at each stage of operations.

4.5.1 KBMLS Startup and Login

Initially, the KBMLS software is booted by turning on the PC hardware. Once booted, the user types "RADC" and logs into the KBMLS system using the KBMLS account name. The KBMLS Login prompt will then be displayed. A user enters an existing KBMLS user id and password to execute the associated KBMLS role. Figure 4.5.1 depicts the process flow.

4.5.2 KBMLS Interactive Processing

There are three distinct roles in the KBMLS system. They are; the security manager, the system administrator, an the analyst. The security manager role is responsible for maintaining the KBMLS login table and reviewing the SATL. The system administrator capabilities include starting and stopping the automatic sanitization software, reviewing history logs, input logs, and output logs, maintaining the knowledge base rule files, grammar files, and all table maintenance functions. The analyst can access, modify, and review SCF's, and review the input and output message logs, and release output products.

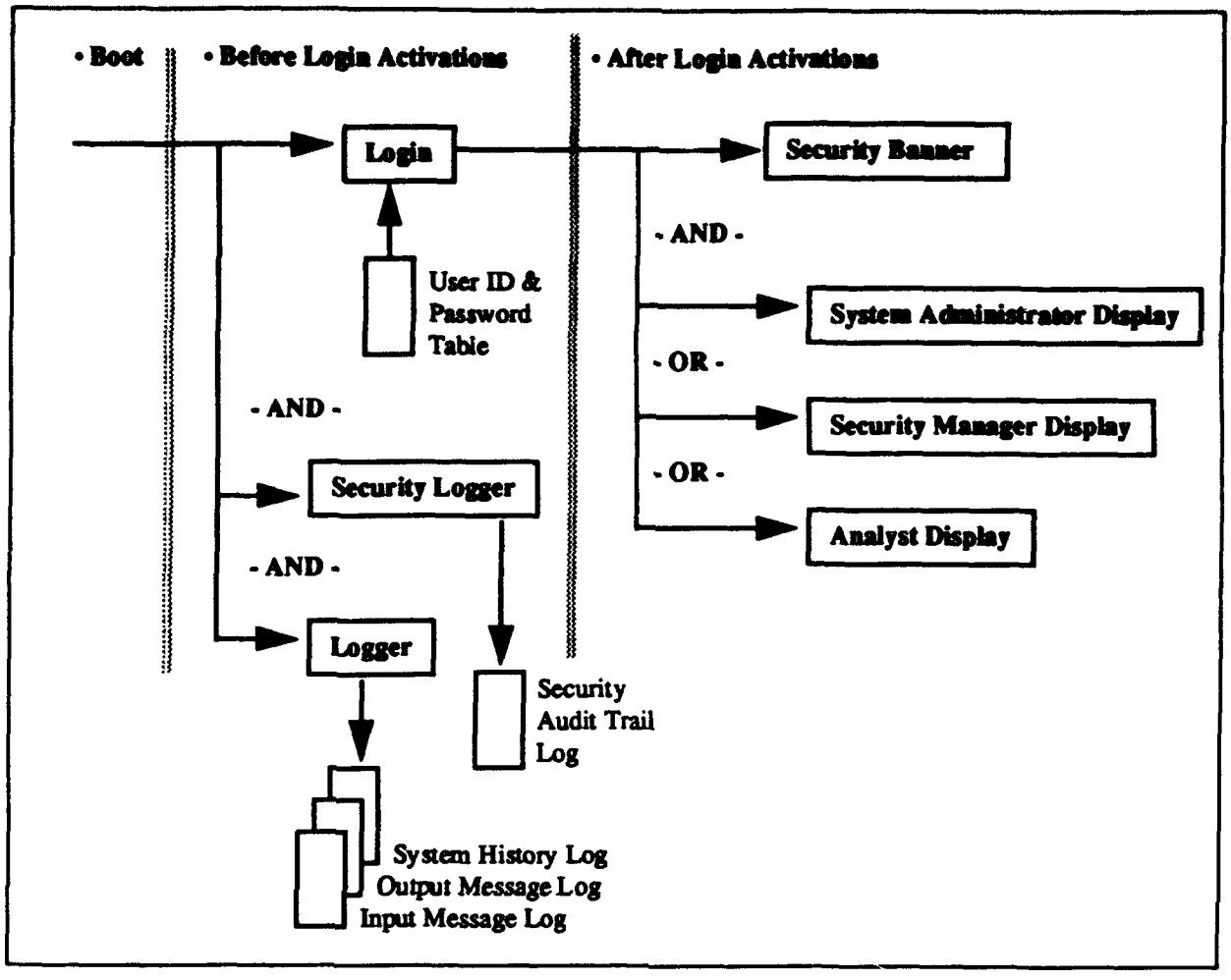

Figure 4.5.1 KBMLS Startup and Login Processing Flow

4.5.3 KBMLS Automatic Flow

The KBMLS prototype accepts formatted message input on a standard UNIX Inter Process Communication (IPC) queue, as depicted in Figure 4.5.3. Using user defined grammar definitions, the data translator process parses the incoming message and extracts the message information into a record structure. The original message and the extracted information are stored in a sanitization case folder (SCF) and written to a database. The data translator sends a message to the pre-sanitizer with the key to access the new SCF to be processed. The pre-sanitizer uses the addressee, area, and cover tables in conjunction with the KES rules file to determine if and how the incoming message can be sanitized. The pre-sanitizer marks a SCF category as sanitizable, releasable, or denied. If the SCF is of category denied, the message is not be sanitized. If the SCF is sanitizeable or releasable, a message is sent to the sanitizer with the key identifying the SCF, the sanitization instruction set to use, and the template of the output product. The sanitizer uses the message template table to get the output product template and

will fill the template keywords with information that was extracted by the parser. The sanitizer then uses the sanitization instructions, deletion and replacement tables, to perform the actual sanitization of the message.

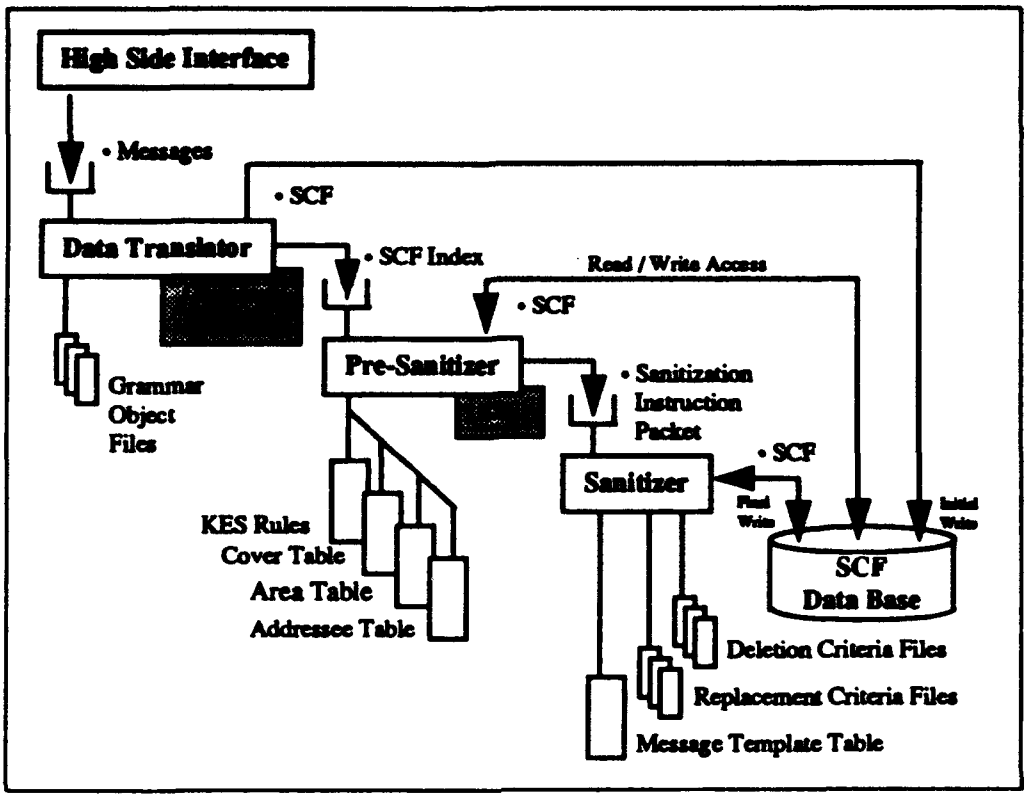

Figure 4.5.3 KBMLS Automatic Flow

5. PRELIMINARY TESTING/FEASIBILITY DEMONSTRATION RESULTS

5.1 General

The Software Testing demonstrated the use of AI techniques to implement sanitization methods and procedures. Preliminary tests on a Computer Software Configuration Item (CSCI) identified as the Knowledge Base Multi-level Secure (KBMLS) Demonstration System were conducted with Government personnel present. The following subsections summarize the testing.

5.2 KBMLS Test Descriptions

The formal tests covered were:

1. Security Manager Operation Test (SecMgr)
2. System Administrator Operation Test (SystAdm)

3. Message Input (MsgIn)
4. System Analyst (SystAnalyst)
5. System Message Processing (MsgProc)
6. System Audit Logs (AuditLogs)
7. System Shutdown (SysShut)

5.2.1 Security Manager Operational Test (SecMgr)

The security manager operational test cases verified the operations of the KBMLS security manager to create and maintain the user accounts and review security logged information. The purpose of this test was to verify:

- a user can login as security manager using an established account
- the security manager can review user accounts
- the security manager can add user accounts
- the security manager can delete user accounts
- the security manager can modify user passwords
- the security manager can review security log information
- the security manager can logout

5.2.2 System Administrator Operational Test (SystAdm)

The system administrator operational test cases verified the operations of the KBMLS systems administrator. The purpose of this test was to verify:

- the user can logon as a system administrator
- the system administrator can update the area table
- the system administrator can update the cover characteristic table
- the system administrator can specify the sanitization criteria using wildcard and keyword phases for message data deletion and replacement criteria tables
- the system administrator can update the message template table
- the system administrator can update the addressee table
- the system administrator can maintain the data translation grammars
- the system administrator can update the knowledge based rules files
- the system administrator can logout of KBMLS

5.2.3 Message Input (MsgIn)

The message input test cases verified the message input function of the KBMLS. Messages were sent from an independent machine to the KBMLS host machine to be translated and sanitized. The purpose of this test was to verify:

- formatted messages can be sent to the KBMLS

- formatted messages can originate on an independent machine.
- messages input will be processed by the data translator and sanitizer processes

5.2.4 System Analyst (SystAnalyst)

The System Analyst test cases verified the ability of the KBMLS analyst to review the sanitized messages, the sanitization case folder information, and release output products to an independent machine. The purpose of this test was to verify:

- messages received by KBMLS can be reviewed
- the sanitized output product can be reviewed and modified.
- the summary of information extracted from the message can be reviewed.
- the justification for the message sanitization can be determined.
- the sanitization output product can be released.
- the sanitization case folder can be examined in a logical ordering.

5.2.5 System Message Processing (MsgProc)

The Operational/Maintenance Mode test cases illustrated all the KBMLS functionality available to the various users during normal operations. Many of the test procedures were shown more than once and were tested in the Initialization Mode test. The purpose of this test was to verify:

- Table modifications can be made to alter the outcome of sanitization processing and SCF's can be resubmitted for sanitization.

- Rules can be modified to alter the outcome of sanitization processing and SCF's can be resubmitted for sanitization.

- The message format grammars can be altered and recompiled to change the way an input message is translated and a message already received may be retranslated.

5.2.6 System Audit Logs (AuditLogs)

The System Audit Logs test cases verified the ability of the KBMLS to record a history of user events and record account information and security events. The input and output logs were used to keep record of the messages received by and the messages released from the KBMLS system. The purpose of this test was to verify:

- messages received by KBMLS are stored in logs
- messages released by KBMLS are stored in logs
- all user account information and security events are maintained.
- all system process actions of significance are maintained.

5.2.7 System Shutdown (SysShut)

The System Shutdown test cases verified the ability of the KBMLS processes to properly terminate and the ability of the UNIX system to properly be shutdown. The purpose of this test was to verify:

- the KBMLS system can stop all processes and terminate.
- the UNIX system can properly shutdown.

5.3 Test Preparations

The KBMLS Hardware architecture for the demonstration system was a single processor architecture with external interfaces as shown in Figure 4.2 in Section 4.2. The architecture was designed to meet security requirements. In order to achieve these requirements a prepackaged tempest hardware suite was acquired from DataWatch Corporation. The following is a list of Hardware characteristics that constitute the KBMLS hardware architecture:

DATAWATCH 386/25 TEMPEST WORKSTATION

Unit Description:	80386 25 MHZ Processor
	16 MB Total System RAM
	80387 CoProcessor
	SCSI Controller
	160 MB Removable Hard Drive
	100 MB Removable Hard Drive
	1.2 MB Floppy Drive
	1.4 MB Floppy Drive
	2 Serial/1 Parallel Adapter
	VGA Adapter and VM14 VGA Color Monitor
	DataWatch Tempest Logitech Mouse
	Standard Keyboard.

TEMTEK EX1000T Dot Matrix Tempest Printer

Unit Description:	Dot Matrix Printer
	Shielded Parallel 9ft Cable.

5.4 KBMLS Test Results

KBMLS testing extended over a period of three years. The following documents each test.

5.4.1 Preliminary Testing: August 15, 1990

The KBMLS Preliminary Test was conducted during the week of August 15, 1990 at the FSC's facility. This test activity satisfied the Preliminary Acceptance Testing requirement of the KBMLS Development Contract and was witnessed by Government representatives from Rome Laboratory. All pre-test activities were conducted by the developing contractor prior to the execution of the test. Debriefing, data reduction and analysis of the test results were conducted immediately after the completion of the test. The prototype system passed all tests.

5.4.2 Preliminary Testing: February 18, 1993

The KBMLS Preliminary Test was conducted at FSC's facility during the week of February 18, 1993 and was witnessed by Government representatives from Rome Laboratory. This test activity concluded the Preliminary Acceptance Testing requirement of the KBMLS Development Contract. All pre-test activities were conducted by the developing contractor prior to the execution of the test. Debriefing, data reduction and analysis of the test results were conducted immediately after the completion of the test. Tests which were incomplete were deferred until Final Acceptance Testing.

5.4.3 Final Acceptance Testing: March 26, 1993

The KBMLS Final Acceptance Test was conducted during the week of March 26, 1993 at the Government's facility, at Rome Laboratory. The test was witnessed by Government representatives from Rome Laboratory. This test activity satisfied the Final Acceptance Testing requirement of the KBMLS Development Contract. All pre-test activities were conducted by the developing contractor prior to the execution of the test. This included interfacing to computer equipment within the Government Facility which emulated the IDHS computer (High Side) and the USAFE GUARD system. The configuration is depicted in Figure 4.2, in Section 4.2. All preliminary tests which were incomplete were rerun and were successful. KBMLS successfully demonstrated the capability to accepted multiple messages from the system high network, parse messages, invoke the sanitization rule sets, sanitize and format the information into a TUM, and release the sanitized message to the USAFE GUARD system.

6. IMPLEMENTATION PLAN

The implementation plan centers around both short term and long term operational goals. The short term goal is to demonstrate the feasibility of using AI techniques to assist intelligence analysts in performing sanitization by preparing automatically sanitized message traffic for manual review by an analyst. The long term goal is to relieve the analyst from near real time reporting responsibilities by using AI techniques to automatically sanitize and disseminate a subset of message traffic which relates to changes in the Order-of-Battle (OB) data base.

6.1 Concept of KBMLS Operations

The Primary Mission of KBMLS is to serve as a component of the USAFE Guard System. It supports the Force Assessment Branch Analyst by assisting in the sanitization of sensitive intelligence information and the generation of sanitized products for dissemination. KBMLS assists the analyst by providing:

- Automatic identification of information required by wing commanders and/or other tactical decision makers
- Automatic determination that information can or cannot be sanitized
- Automatic selection and application of appropriate sanitization procedures
- Automatic generation (or review) of sanitized product
- User friendly interface for analyst to modify system operational parameters.

The KBMLS accomplishes this by providing necessary communications interfaces, sanitization decision support functions, and a man-machine interface to permit the analyst to access and review sanitized information. The KBMLS software can be changed in the field to reflect changes in the policies and procedures mandated by cognizant authority significantly reducing life cycle maintenance costs.

The KBMLS Sanitization software is capable of residing on the same computer suite as the TS/SCI host or the USAFE GUARD. These TS/SCI host, USAFE GUARD, and KBMLS computers operate at the "system high" level. (In other words, all data in the system are protected as TS/SCI until a person determines or approves the actual classification of the sanitized data.) KBMLS will communicate with its collocated high host and GUARD via a system high local area network called the USAFE Theater Air Intelligence Network Local Area Network (UTAIN LAN). TS/SCI access will be required for persons to have physical access to the UTAIN LAN, the Guard, TS/SCI host, or KBMLS computers.

III. CONCLUSIONS

1. CLOSING SUMMARY

The Knowledge Based Multi-Level Secure Network Technology (KBMLSNT) project began with an operational requirement, from the COIC/TFC, to investigate ways of increasing the timeliness of Order-of-Battle updates to wing commanders. This operational requirement translated into the goal of applying artificial intelligence techniques to the problem of sanitizing sensitive information for dissemination to wing commanders.

FSC began by initiating a knowledge acquisition effort at the COIC and TFC sites. These sites provided the information needed to define a sanitization model. The model described the current manual process involved with sanitization and where artificial intelligence techniques such as expert systems, knowledge based systems, and natural language processors could be applied.

Hardware and software requirements followed the Air Force's standards for developing computer systems for processing classified information. The hardware was a TEMPEST 386/25Mhz Workstation with a UNIX System V3.2 operating system. The KBMLS software was written in C and Ada and contains a COTS expert system shell called KES. FSC conducted a survey of expert system shells and natural language parsers which would meet operational requirements. To support contract requirements, the software developed by FSC and all COTS packages had to be portable to other workstations.

The software developed by FSC, adheres to the open system architecture principles of portability, availability, scalability, and interoperability. The user interface was developed within X-Windows/ Motif standards. The external network interface communicates using TCP/IP and will easily transition to GOSIP network protocols in the future. The software is divided into four sections; Data Translation, Sanatization, Executive Control, and Global Utilities. Data Translation enables KBMLS to understand and process multiple message formats. Sanitizion provides all functions involved with authorization and sanitization. Executive Control provides the necessary control and security mechanism for KBMLS to solely execute on a workstation. The Global Utilities provides portability through modifications to a few routines.

Testing included interfacing KBMLS to computer equipment within the Government Facility which emulated the IDHS computer (High Side) and the USAFE GUARD system. The configuration is depicted in Figure 4.2, in Section 4.2. All preliminary tests were successful and KBMLS successfully demonstrated the capability to accepted multiple messages from the system high network, parse messages, invoke the sanitization rule sets, sanitize and format the information into a TUM, and release the sanitized message to the USAFE GUARD system. Due to time limitations, not all messages sets were developed and tested to determine if the rules were complete.

The benefits of the KBMLS system are many. It provides a mechanism to retain corporate knowledge relating to sanitization and at the same time assist in the training of new analysts. The KBMLS system has the potential to increase the flow of critical information to wing commanders and establish consistency in reporting critical events. It has the potential to reduce or eliminate future security breaches. Examples of KBMLS's overall benefit as it relates to Mission Responsiveness, Operational Support at Site, and Growth / Flexibility are;

- Mission Responsiveness — Equipped with the on-site capability to modify rules and knowledge as mission requirements change provides near real time readiness support during peacetime and wartime.
- Operational Support at Site — Integration with the existing Intelligence Community Computer Architecture and conforming to Human Factors standards reduces training and technical manual costs. Use of existing or future equipment minimizes spares, support equipment and preventive maintenance costs.
- Growth / Flexibility — Portable to future Intelligence Community Computer Architecture's. Flexible, with on-site modifications to expand rules and knowledge to accommodate other Intelligence Community domains.

Operational Flexibility is a key to mission responsiveness. Table 1 describes a situation and suggests how KBMLS would be adapted to meet the change.

Table 1 Examples of Operational Flexibility as it relates to Mission Responsiveness

As wings deploy and redeploy in and out of theater	Update sanitization rules and tables to account for the change
When changes to National, Service, and Theater level sanitization policies occur	Modify the sensitivity rules and sanitization deletion and replacement criteria tables to reflect the change in policy
As the threat changes	Activate desired rule sets, corresponding to DEFCON levels, for automated sanitization processing
Occasionally, incoming message formats change	Modify input message parsing grammars
Conversely, as outgoing sanitized message formats change	Modify the Outgoing Message Template table
As units are deployed or redeployed altering the Order of Battle	Modify the need rules to specify specific interest in a unit as it relates a list of potential addressees

2. FUTURE EVOLUTION OF KBMLS

Site acceptance and accreditation of both short term and long term goals are dependent on an evolutionary approach. Figure 1 depicts the potential evolution. Today, analysts are wholly responsible for sanitizing incoming message traffic and a senior officer is the review/release authority. It is envisioned that in the first implementation and deployment of KBMLS, the system would automatically sanitize and present to the analyst the sanitized information for verification

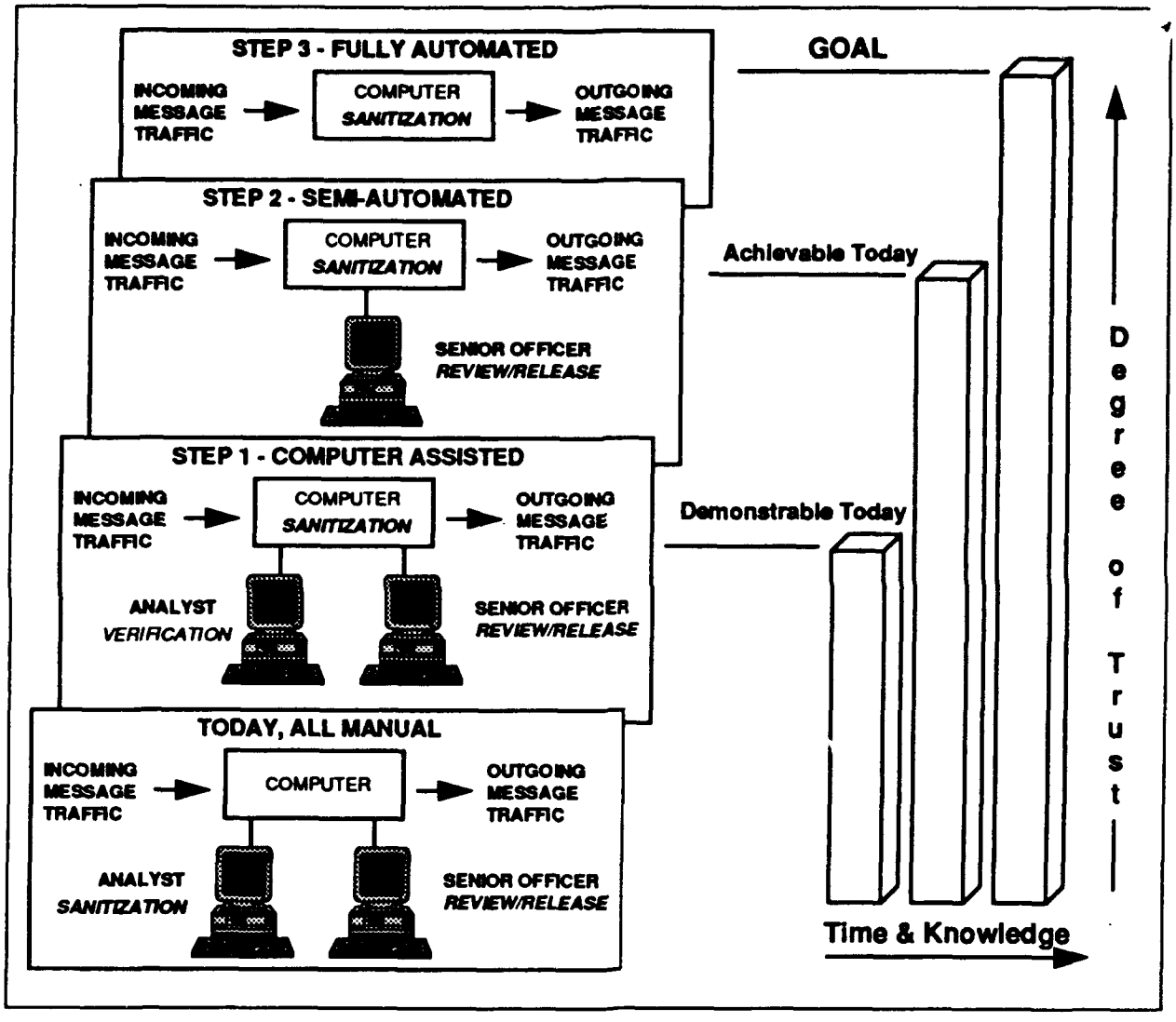

Figure 1 KBMLS Evolution

Statistics would be gathered on the performance and accuracy of the KBMLS system. The system would be enhanced to account for known deficiencies. The second implementation and deployment of the KBMLS system would relieve the analyst of the responsibility to sanitize incoming data. Only the senior officer would remain within the loop, providing review/release

authority. Lack of senior officer denial would pave the way to the final implementation. The final implementation and deployment would eliminate the senior officer from the sanitization process and the KBMLS system would automatically sanitize and disseminate information to air wing commanders.

BIBLIOGRAPHY

a. Harmon, Paul., "Expert Systems Strategies", Cutter Information Corporation, VOL 5. NO. 2

b. Vedder, R., Fortin, M., Lammermann, S. and Johnson, R., "Five PC-Based Expert Systems for Business Reference: An Evaluation", Information Technology Libraries, March 1989.

c. Harmon, P, and Maus, R., Expert Systems: Tools and Applications, John Wiley and Sons, Inc., 1988

LIST OF ACRONYMS

AFB	Air Force Base
AFMC	Air Force Material Command
AI	Artificial Intelligence
AOB	Air Order of Battle
C2I	Command, Control and Intelligence
CDRL	Contract Deliverable Requirements List
COIC	Combat Operations Intelligence Center
COMINT	Communications Intelligence
CSCI	Computer Software Configuration Item
DEFCON	Defense Condition
DoD	Department of Defense
ECM	Electronic Counter Measures
FA	Force Assessment Branch
FIRE	Fused Intelligence Report to Europe
FSC	Fuentez Systems Concepts, Inc.
FTP	File Transfer Protocol
I&W	Indications and Warning
ID	Identification
IMINT	Imagery Intelligence
INTREP	Intelligence Report
KBMLS	Knowledge Based Multi-Level Secure
KES	Knowledge Expert System
MLS	Multi-Level Secure
MOB	Missile-Order-of-battle
NATO	North American Treaty Organization
OIs	Operating Instructions
PC	Personal Computer
RL	Rome Laboratory
SAM	Surface-to-Air
SAMINTREP	SAM Intelligence Report

SATL	Security Audit Trail Log
SCF	Sanitization Case folder
SCI	Sensitive Compartmented Information
SCO	Santa Cruz Operations, Inc
SDD	Software Description Document
SIGINT	Signal Intelligence
SOW	Statement of Work
STD	Software Test Description
TAN	Target Air Nomination
TFC	Tactical Fusion Center
TIIN	TFC Intelligence Input to NATO
TWIIN	TFC Wartime Intelligence Input to NATO
U.S.	United States
USAF	United States Air Force
USAFE	United States Air Force Headquarters in Europe
USAFINTELL	United States Air Force Intelligence
USEUCOM	United States European Command
VA	Virginia
VGA	Video Graphics Adapter

APPENDIX A

EXPERT SYSTEM VENDORS AND FEATURES

ALEX 1.1: $695
Harris & Hall Associates
P.O. Box 1900
Port Angeles, Wash. 98362
(206)457-4907

Runs in Smalltalk environment. Allows programmers to develop object-oriented expert systems.

INTELLIGENCE/COMPILER: $490
IntelligenceWare, Inc.
9800 S. Sepulveda Blvd.
Los Angeles, CA. 90045
(213)417-8896

An expert system development environment that combines rules, frames, relational databases and hypertext. It supports a variety of AI techniques.

COSMIC CLIPS 4.2: $250
University of Georgia
382 E. Broad St.
Athens, GA. 30602
(404)542-3265

Machine-independent; runs under any complete C compiler. Made available for reuse under NASA's Technology Utilization Program.

CxPERT 3.0: $795
Software Plus
1652 Albermarle Dr.
Crofton, MD. 21114
(301)261-0264

Object code version is available for the IBM PC and compatibles. Source code version is written in C and can be compiled on any machine. Object code version runs under MS-DOS.

ESP ADVISOR-2: $695
ESP FRAME ENGINE: $695
Expert Systems International
1700 Walnut St.
Philadelphia, PA. 19103
(215)735-8510

ESP Frame Engine is a frame-based system that supports forward and backward-chaining rules. Advisor-2 is an enhanced rule-based system which provides direct access to external programs and languages. Runs under MS-DOS, VMS and UNIX.

EXSYS PROFESSIONAL: $2,500
EXSYS Inc.
P.O. Box 11247
Albuquerque, N.M. 87192
(505)256-8356

EXSYS is a tool for development of probabilistic rule based expert systems and provides flexible means to combine confidence factors and direct access external programs and languages. Runs under VMS, UNIX, MS-DOS and OS/2.

1st CLASS HT: $2,495
1st CLASS Expert Systems Inc.
526 Boston Post Rd.
Wayland, MASS. 01778
(508)358-7722

A combination of a forward and backward chaining expert system shell, code generator, graphics capture and display utilities and database interface. Runs under MS-DOS and VMS.

GOLDWORKS II: $8,900
Gold Hill Computers Inc.
26 Landsdowne St.
Cambridge, MASS. 02139
(617)621-3300

Hybrid shell which combines rules, frames, and object oriented programming with graphical interfaces. Runs under a variety of operating systems.

GURU: $6,500
Micro Data Base Systems Inc.
P.O. Box 248
Lafayette, IND. 47902
(317)463-2581

Expert environment for business application development with easy access to databases, spreadsheet and word processing programs. Runs under MS-DOS, OS/2 Sun, UNIX and VMS.

HUMBLE 2.0: $395
Xerox Special Info. Systems
250 N. Halstead St.
Pasadena, CA. 91107
(818)351-2351

Available for many Xerox machines Macintosh, Sun workstations, HP 9000 series, Apollo 3000 and 4000 series and some Tektronix machines. Runs in Smalltalk/80 environment.

INSTANT EXPERT PLUS: $498
Human Intellect Systems
1670 S. Amphlett Blvd. #326
San Mateo, CA. 94402
(415)571-5959

Instant Expert Plus combines rules and graphics to create expert systems applications. Uses forward and backward chaining and mixed logic. Runs under MS-DOS and Macintosh OS.

KDS 3.6: $1,495
KDS Corp.
934 Hunter Rd.
Wilmette, ILL. 60091
(312)251-2621

A large-capacity system, KDS 3.6 uses inductive reasoning to produce rules from facts. Runs under MS-DOS.

KEE: $9,000 - $30,000
IntelliCorp
1975 El Camino Real
Mountain View, CA. 94040
(415)965-5500

LISP-based development tool for creating and delivering large, complex, knowledge based applications. Runs under a variety of operating systems and on a variety of machines.

KES: $4,000
Software A&E Inc.
1600 Wilson Blvd. #500
Arlington, VA. 22209
(703)276-7910

Object oriented expert system shell designed to link with C code modules. Written in C. Runs under a large variety of operating systems and machines.

KNOWLEDGEPRO: $495
Knowledge Garden Inc.
473A Malden Bridge Rd RD2
Nassau, N.Y. 12123
(518)766-3000

KnowledgePro is a knowledge processing environment that combines hypertext, a variety of AI techniques and conventional programming productivity tools.

LEVEL5: $685
Information Builders Inc.
1250 Broadway
New York, N.Y. 10001
(212)736-4433

In addition to an expert system shell, Level5 contains a number of direct rewrite interfaces to standard databases. Runs under MS-DOS, Macintosh OS, VM/CMS, MVS/TSO and VMS.

LOGICTREE: $495
CAM Software
750 N. 200 West, Ste. 208
Provo, Utah 84601
(801)373-4080

Expert system development tool that uses decision trees to capture knowledge. Runs under MS-DOS.

M.1: $5,000
Cimflex Teknowledge Corp.
1810 Embarcadero Rd.
Palo Alto, CA. 94303
(415)424-0500

M.1 is available on the IBM PC/ST/AT and true compatibles. It is geared toward small, rule based development projects. $5000 is for site license.

MERCURY KBE: $31,000
Artificial Intelligence Tech.
40 Saw Mill Rd.
Hawthorne, N.Y. 10532
(914)347-6860

Available for VAX machines; runs under VMS. Knowledge engineering environment designed to provide high performance, stable storage and tight integration with the user's software tools and methodologies.

NEXPERT OBJECT: $5000-$8000
Neuron Data
444 High St.
Palo Alto, CA. 94301
(415)321-4488

A rule and object-oriented shell, written in C, with graphical interfaces. Fully cross-compatible with other platforms. Runs under MS-DOS, UNIX, VMS and the Macintosh OS.

PC EXPERT PROFESSIONAL: $495
Software Artistry
3500 DePaul Blvd.
Indianapolis, IND. 46286
(317)876-3042

PC Expert is developed for use with Pascal, Microsoft C, Turbo C, JPI Modula-2 and Logitech Modula-2. It contains an inference engine, a procedural language and internal development environment.

PERSONAL CONSULTANT PLUS
$2,950
Texas Instruments Corp.
P.O. Box 1444
Houston, TX 77251
(800)847-2787

Available for IBM PC/XT/AT and compatibles. Runs under MS-DOS.

SOCRATES: $295-$695
CIM Solutions
754 S. 400 East, Ste. 200
Orem, UTAH 84058
(801)374-5626

Decision tree expert system designed for knowledge-based management; Socrates is written in C and features context sensitive searching. Runs under MS-DOS and VMS.

TURBO SHELL 3.0: $119
Berkshire Software Co.
44 Madison St.
Lynbrook, N.Y. 11563
(516)593-8019

Allows development of applications in Borland's Turbo family of languages. Runs under MS-DOS.

VP-EXPERT: $249
Paperback Software Inc.
2830 Ninth St.
Berkeley, CA 94710
(415)644-2116

Rule based-expert system development tool with links to spreadsheets, database managers and ASCII files. Runs under MS-DOS.

APPENDIX B

DOCUMENT PARSING TOOL VENDORS

	VENDOR	TOOL
1.	SYCHRONETICS, INC. P.O. BOX 793 HANOVER, MD. 211076	NL BUILDER (Hybrid Tool)
2.	WESWARE, INC. 42 EPPING STREET LOWELL, MA 01852	GOLDEN RETRIEVER (Text Search Tool)
3.	THIRD EYE SOFTWARE, INC. SUITE 170 535 MIDDLEFIELD ROAD MENLO PARK, CA. 94025	ELEXIR (Text Search Tool)
4.	ENGLISH KNOWLEDGE SYSTEMS 5525 SCOTTS VALLEY DR #22 SCOTTS VALLEY, CA. 95066	JAKE (Natural Language Processor)
5.	DATAMAT VIA SIMONE NARTINI 126 00142 ROMA (EUR)	TEXT/SEARCH (Text Search Tool)
6.	BBN SYSTEMS AND TECHNOLOGIES CORP. 10 MOULTON STREET CAMBRIDGE, MA 02138	PARLANCE (Natural Language Processor)

MISSION

OF

ROME LABORATORY

Rome Laboratory plans and executes an interdisciplinary program in research, development, test, and technology transition in support of Air Force Command, Control, Communications and Intelligence (C3I) activities for all Air Force platforms. It also executes selected acquisition programs in several areas of expertise. Technical and engineering support within areas of competence is provided to ESC Program Offices (POs) and other ESC elements to perform effective acquisition of C3I systems. In addition, Rome Laboratory's technology supports other AFMC Product Divisions, the Air Force user community, and other DOD and non-DOD agencies. Rome Laboratory maintains technical competence and research programs in areas including, but not limited to, communications, command and control, battle management, intelligence information processing, computational sciences and software producibility, wide area surveillance/sensors, signal processing, solid state sciences, photonics, electromagnetic technology, superconductivity, and electronic reliability/maintainability and testability.