Operational Requirements Document (ORD)
for the
Cross Compartment Intelligence Alert System (CCIA)

Abstract

This paper discusses a new network-based technology that will help organizations better utilize sensitive data by addressing fundamental conflicts such as:

- privacy-rights versus public-interest,
- collaboration versus self-interest, and
- the detection-avoidance of specific threats to the nation versus the general need to protect data sources.

The paper focuses on the latter conflict. The current intelligence data establishment is deeply divided by security compartments that impede the timely and the effective fusion of intelligence data that is essential for the detection and elimination of threats such as terrorist attacks and Ponzi schemes. There is an urgent need to improve data fusion but an equally compelling need to prevent the intelligence data from being abused or exploited to betray vital secrets and damage intelligence information sources.

The key technical mechanisms are explained in the paper and reduced to a set of vendor-independent requirements.

Paul Baker
pbaker@wwnsoftware.com

Contents

1.0 General Description of Operational Capabilities

1.1 Capability Gap

Homeland security relies on knowledge of our enemies and early warning of threats. Quality information is never handed over in a package, however. It emerges by carefully assembling the evidence regarding an enemy's plans and preparation. The national intelligence enterprise is divided into several components and all are in a position to collect intelligence information, but there is a danger that no single component has enough evidence to characterize a threat. Without corrective action, the division of the intelligence enterprise introduces a capability gap.

The capability gap is particularly acute with respect to international cooperation as highlighted by the failure to act on information about Umar Farouk Abdulmutallab because the fragments of information were divided between UK, US, Yemen and Nigeria. The picture became clear only a potentially disastrous incident occurred and the various governments exchanged information on that specific suspect.

In brief, there is a gap between the ability to cooperate and understand an incident after it occurs and our current inability to recognize, predict, and thwart a forthcoming incident.

This capability gap is bridged, in principle, by teamwork between agencies. Analysts representing the interests of their agency and enjoying access to their agencies' data can sit down together in secure setting and exchange related information. Clearly however, the throughput and response time for such manual analysis is unable to deliver a timely warning of all impending threats.

Timely intelligence alerts of threats to the nation can be assured only when every component of the intelligence enterprise cooperates fully with the information integration effort. A capability gap will remain so long as there are organizational issues that impede cooperation between agencies and governments. Several issues discourage cooperation:

- Much of the information that is critical for early threat detection is also very sensitive in regard to how it was gathered. Sharing the information widely, even for a valid purpose like intelligence analysis, creates a risk for sources and the operations that rely on them.
- Intuitions and suspicions that should guide integrated information analysis often start with an insight from an agent working at a low level, near the problem. That agent may have difficulty influencing the intelligence analysis effort to perform the necessary analysis.
- When information is simply made available for intelligence analysis, traceability and accountability suffer. The components that supply information for intelligence analysis want to know where and how their work is used so that they

can receive credit for successes and protest data misuse. When traceability is lacking, there is no incentive to comply with the mandate to share data.

There is additional discussion of the organizational issues in Section 3.2, below.

The *Cross Compartment Intelligence Alert System* or CCIA will fill the capability gap by enabling intelligence analysis across compartments of the intelligence data enterprise while maintaining the protection afforded by the secure data compartments. The terms *compartment* and *intelligence analysis* are used in the following senses. A *compartment* is a selection of intelligence data protected by security barriers. A compartment is established and managed by one of the components of the intelligence enterprise. Conventionally, access to a compartment is based on an individual's level of clearance and the individual's need to know. For this reason, each component of the enterprise maintains a number of compartments distinguished by their access rights. The compartments for intelligence data are essential components for security, but they are also barriers to intelligence analysis.

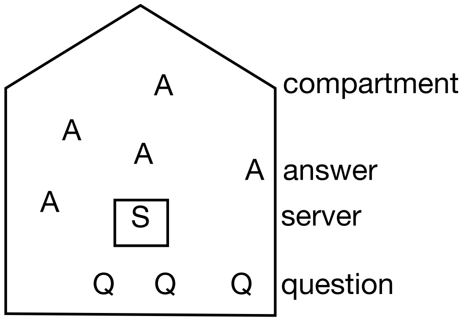
Intelligence analysis is the process of integrating the intelligence information and assigning significance to a product. The second step is important because many apparent associations of facts are chance associations. Therefore, a full intelligence analysis algorithm is both a method for integrating and associating intelligence data records and also a methodology for evaluating the significance of the results.

In summary, existing systems provide disincentives for the cooperation necessary for a successful intelligence analysis program. To remedy this situation, the CCIA introduces enabling technology that performs integration of information across security compartments without revealing the content of the data behind a compartment's barrier.

1.2 Description of the Proposed System

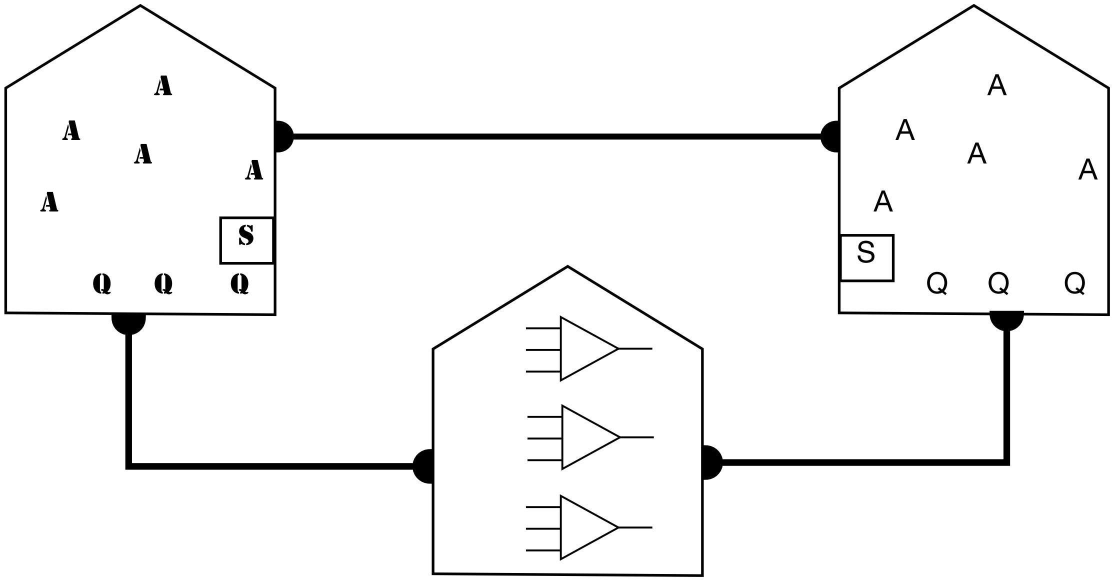
1.2.1 Intelligence analysis with Encrypted Data

The proposed CCIA system will perform intelligence analysis operations across the intelligence enterprise without removing the essential security barriers around data compartments. The explanation of the method begins with an overview of the intelligence analysis process.



The figure to the left illustrates a compartment as a five sided figure reminiscent of a building. The compartment maintains a collection of classified information represented by the letters "A" for answer. Usually, the compartment has a type of server that answers questions. We represent a server with the letter "S" in a process box. During routine operation, an analyst presents a question to the server and may get an answer. There is no reason to store the question. However, good analysts will ask deeply probing questions that have no answer within the data collection in this data compartment. These questions will remain unanswered. We represent the unanswered questions with the letter “Q” for question. The compartment protects both the answers and the questions.

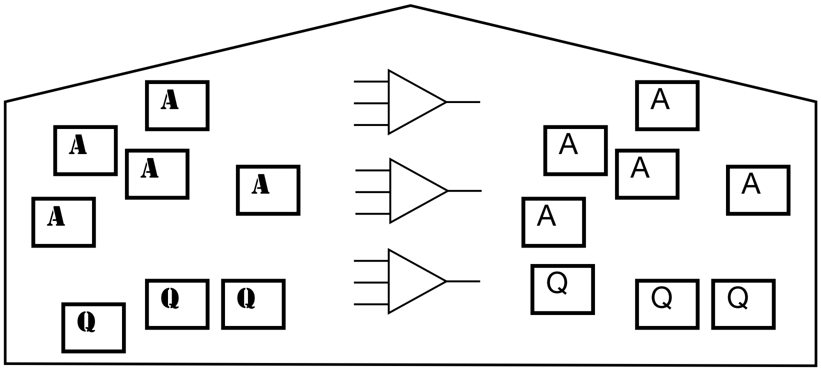
Successful intelligence analysis must work across data compartments to find associations that define threats to national security. Currently, that means opening the compartments to allow access. Two obvious approaches are illustrated in the figure below. The horizontal line connecting left and right data compartments in the figure below represents a cryptographically secured communication link between the two compartments. The compartment on the right allows its server to respond to questions submitted by an analyst in the compartment on the left through the encrypted connection.

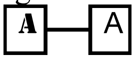
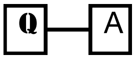


If the query succeeds, the analyst receives back a query-answer pair e.g. Q — A. There are two drawbacks to this approach. An analyst can ask a query against the entire database. Thus, the analyst can abuse the access privilege and learn facts beyond any reasonable need to know. Second, the server can record the queries for study. That allows one compartment to learn about ongoing investigations in another compartment. For sensitive investigations, that feature is undesirable.

The second obvious approach illustrated in the figure above relies on an automated process in another compartment to run intelligence analysis algorithms against data stored in other compartments. The third compartment, shown in the lower middle of the figure above, illustrates this method. The algorithms are represented by triangular process symbols. The processes look for associations of facts found in the different compartments, e.g., A — A. The disadvantage of this method is that data from the compartments is transmitted to the common location bypassing the protection of the original compartments.

The CCIA builds on the architecture of the second method by locating the intelligence analysis processes in a separate secure compartment which functions in the system as a middleman or matchmaker. The matchmaker stores no data but performs an intelligence analysis service for the other compartments. In the CCIA method, the secure compartments encrypt data before providing it to the middleman service. The middleman is unable to decrypt the data. The figure below illustrates this distinctive step in the CCIA method. In this figure, the encryption applied to the data is illustrated by drawing solid rectangular boundaries around the letters representing data answers and questions.



At the stage illustrated above, the two data compartments have encrypted each intelligence data record, “A”, and each analyst query, “Q”, using a common encryption key and then sent the encrypted information to the matchmaker’s compartment for intelligence analysis. The matchmaker does not possess the encryption key; therefore, the security of each data compartment is respected during intelligence analysis process. Because of this level of protection, the matching algorithms provide only encrypted output, e.g. associations of related encrypted data  and pairs of encrypted queries and answers . The output of intelligence analysis is useless until the results are returned to the compartments where the information originates. The encrypted results are fully useable in the source compartments where the required encryption key is available to render the results in clear text.

The advantages of intelligence analysis with encrypted data include the following:

- No compartment receives access to the full complement of information in another compartment. Only the information in the matched data will be transferred between compartments.

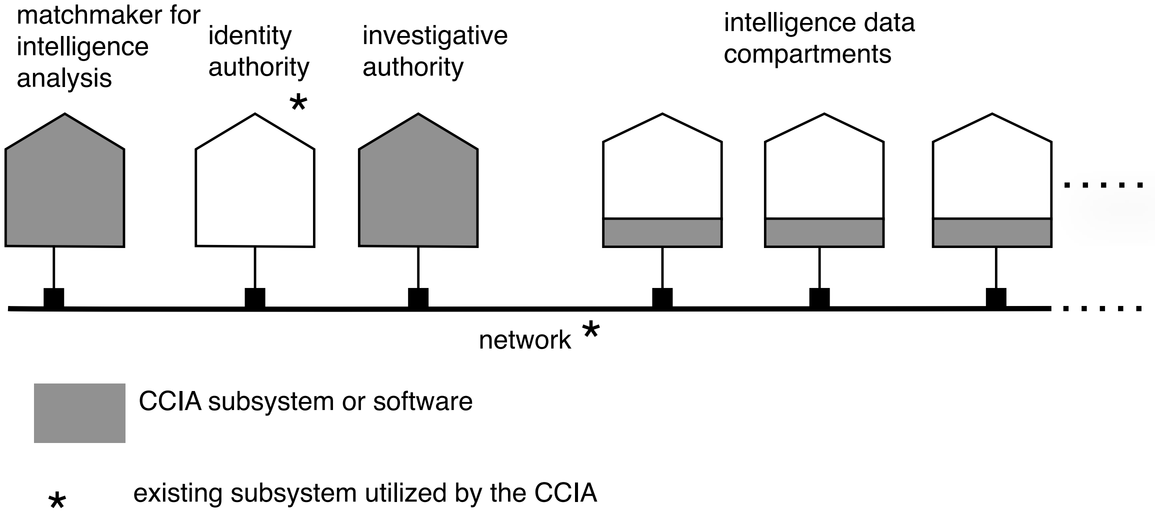
- The intelligence analysis center never works with clear text; therefore, adding the center to the system does not reduce security.

- One compartment knows when another compartment receives some information because both compartments receive copies of the fused data.

- The risk of sharing data for intelligence analysis is drastically reduced; therefore, we expect better compliance with data sharing directives.

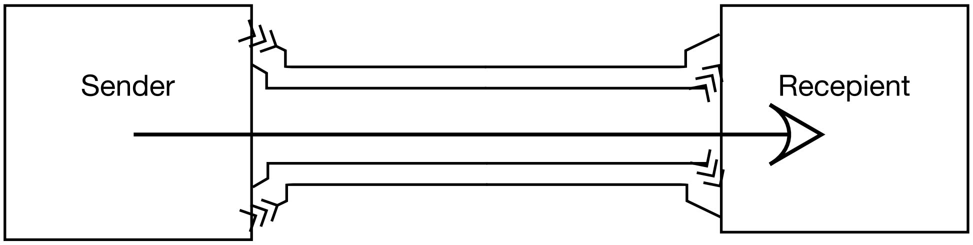
1.2.2 CCIA System Architecture

In operation, the CCIA uses new subsystems and software combined with existing subsystems to leverage existing investments and reduce development time and risk. The high-level system architecture is illustrated in the following figure:


matchmaker for intelligence analysis
identity authority *
investigative authority
intelligence data compartments
network *
CCIA subsystem or software
* existing subsystem utilized by the CCIA

The CCIA connects to an existing secure communication network but provides its own encryption service that may supplement that available in the existing network. Each compartment installs a component of the CCIA software to join the larger CCIA system. The CCIA system adds a major new component – the intelligence analysis matchmaker, which is tasked with matching or analyzing encrypted data from the existing intelligence data compartments. The remaining two subsystems shown above support the organizational goal of the CCIA – the increased cooperation of the components in the intelligence enterprise.

The goal of cooperation is reached by two means: trust and accountability. Before trust can develop, it is imperative to know whom you are dealing with. That identification problem occurs on every data transfer; fortunately, there is an existing cryptographic solution that we illustrate in the following figure:



Every communication in the CCIA has a specific sender and recipient. Two layers of encryption disguise the communication's content while positively identifying both the sender and recipient. The right pointing arrow in the above figure represents the transfer of the content over the secure connection on the network. The layer of encryption around the message can only be removed by the intended recipient. Another layer around that can only be removed using the public encryption key associated with the actual sender. Thus, the communication link successfully transfers the message only if the message arrives at the intended recipient and only if it was generated by the claimed sender. This standard methodology relies on an identity authority that supports public key encryption by associating agents on the network with unique certificates. The identity authority is incorporated as an existing subsystem that must be present in the CCIA.

The CCIA encourages trust by positively identifying the participants and preventing casual access to a participant's data. Obviously, some data must be shared to achieve the goal of integrating intelligence information. However, only the data items that integrate successfully are shared and then they are shared only with the compartments that provided a component of the matched data records. The CCIA system limits the amount of data placed at risk by reducing data sharing to the minimum possible amount – specifically to those items selected by analysis of encrypted information because they convey useful, related information between the security compartments.

It is true, however, that there remains some risk. Consequently, it is important to have a method to investigate data abuse. The primary CCIA intelligence analysis method is so secure that it is not easy to investigate suspicions of misuse. The matchmaker has the results of intelligence analysis, but the results are encrypted, they cannot be read, and cannot be used as evidence. To correct this shortfall, the CCIA includes an investigative subsystem. That subsystem is invoked when a problem is suspected. It can retrieve encrypted intelligence analysis results from the matchmaker. Then, it can decrypt the results with the cooperation of any one of the original parties in the intelligence analysis operation. Consequently, the investigative subsystem has the capability to investigate any incident when it has the cooperation of the matchmaker and one or more of the intelligence data components. This restriction limits the scope of the investigation to serious matters and prevents the casual search and possible abuse of the data.

In summary, the proposed system affords a level of protection for the operating components of the intelligence enterprise that will encourage a greater degree of cooperation between the components. Ideally, more widespread cooperation on intelligence analysis will lead to threat identifications that arrive earlier and are more specific than those produced with the current system.

1.3 Operations Concept

All current operations must continue to perform intelligence analysis while the CCIA is deployed and brought into operation. In terms of the mechanics of intelligence analysis, the CCIA performs the same intelligence analysis operations as the current automated operation but differs from the current system by offering a higher level of security and superior tracking of data movement. Initially, the two can operate side-by-side. However, the information available to the CCIA will grow quickly because the security advantage will allow the use of information too sensitive for inclusion in the current analysis system.

The selection of an organizational entity to operate the secure, encrypted intelligence analysis process is a major decision that will help the CCIA win the trust and cooperation of the existing components of the intelligence enterprise. The assignments of responsibility for the matchmaker and investigative subsystems of the CCIA are considerations that are non-technical and will be decided during the testing and deployment of the new system.

2.0 Existing System Shortfalls

2.1 Overview

The existing system is built with hardware, software, and human assets. Current software may meet the specified functions of the system but it fails to address serious legal and organizational. The consequence is a failure to achieve the desired performance in intelligence data analysis. The following table lists the main areas of shortfall:

Table 2-1 General Areas of Shortfall

Information that is available for "connect the dots" intelligence analysis is likewise available for espionage, blackmail, or other forms of theft.
Risky tradeoffs are made between utilizing intelligence data and protecting the original sources of the data.
Information may be used without proper credit to the analyst discouraging future cooperation and increasing intramural rivalries.
The mission to investigate all possible threats impinges on the privacy rights of citizens.
Individuals may withhold data from the system to protect the organization against perceived dangers. The organization may then fail to "connect the dots".

Section 2.2 provides a longer description of these shortfalls. The shortfalls can only be remedied with the proper combination of new software and new organization of the human assets. However, a number of approaches are being considered or tried. Each has its own shortfalls. We list some major approaches in the following table.

In Table 2-2, each approach is referenced to the subsection below that discusses the approach. The third column in the table summarizes the shortfalls and the last column gives the deployment status.

Table 2-2 Shortfalls of Enhancements to the Current System

Approach	Section	Shortfalls	Deployed
Manual	2.3	Slow and error prone. Productivity insufficient.	yes
Open Source Spying (Intellipedia)	2.4	No protection at all from espionage, blackmail and other data theft. Data are certain to be withheld from the system to protect sources and prevent abuses.	yes
Anonymous Data	2.5	Only a partial solution. Substitution of a name token for a name is not secure against attack.	available
Encrypted databases	2.6	Encryption key is widely distributed and therefore adds little to the existing security barriers.	proposed

2.2 Organizational Causes for the Shortfalls

Intelligence analysis methods assume that data is freely available for analysis. Yet, it is clear that each component of the intelligence apparatus must guard its data to protect its own sources from reprisal and to prevent an enemy from assessing the capabilities of intelligence gathering and potentially even exploiting gaps in those capabilities. An outside enemy may intend harm, but insiders pose a risk too. Each component of the U.S. intelligence enterprise regards the other as an ally but perhaps an unreliable one. In every large organization, there will be staff members of uncertain quality and loyalty. Consequently, access to sensitive data must be restricted. Moreover, allied components of the intelligence enterprise often adopt goals and policies that are not fully acceptable to their fellow components. Conflicting policies provide another excuse for a component to restrict the use of its intelligence information.

In addition, intramural rivalry complicates the relations of allied components. One component will be naturally reluctant to share preliminary data with others in order to preserve a leadership advantage in counterterrorism operations. The counterterrorism operation might be more effective if based on integrated data; but that consideration will be weighed against the realization that only one organizational component will claim a successful operation.

Clearly, when an opportunity for cooperation is missed, the event will not be documented unless there is a serious event that leads to an investigation. As a result, on a day-to-day basis, no one can be blamed for the routine failure to cooperate on intelligence analysis. In fact, each component sees an advantage to working alone. Without a new element – such as a radically novel system capability like the CCIA – no organizational improvement is likely.

From a legal standpoint, another issue limits the application of intelligence analysis. In the course of successful data gathering activities, information may be obtained that compromises the constitutional guarantees of U.S. citizens and runs counter to legislation regarding privacy protection. It is neither possible nor desirable to avoid gathering information in such a grey area. Any comprehensive intelligence gathering operation should find it or there is something wrong with the information acquisition method. Since it is inevitable that private data are collected, such data must be handled in a way to minimize any conflicts with statute.

If private data is revealed publicly, the specific content of private data tells more about how it was gathered than the route by which it was leaked. It follows that the intelligence component that gathers such information will be held accountable if the information is misused while the component that inadvertently or intentionally exposes the data can easily escape any blame. In this situation we see another organizational conflict of interest – the wrong people get blamed for data abuses. Finally it should be noted that private data could be leaked and used for private criminal activities such as blackmail and extortion. Any data system must guard against such criminal activity.

In summary, the intelligence enterprise is characterized by a structural problem: a conflict of interest between minimizing the risk of data misuse and maximizing the intelligence analysis output. Furthermore, different components of the enterprise may adopt different goals and policies that impede the cooperation on the intelligence analysis effort. As a consequence, organizational support for intelligence analysis falls short of the optimal level.

The next two sections will discuss methods that are currently deployed to address intelligence analysis issues but which fail to achieve optimal results.

2.3 Liaison Staff

Without an innovative technical solution, the shortfalls in intelligence analysis across security compartments must be managed by tasking staff to fill the gap. That is the role of cross-agency and cross-department analysis centers staffed by representatives of the individual components. Each representative is provided with a secure link back to their home organization and a computer analysis terminal. The representatives can work together on a problem, each can query their respective databases, and all can discuss and contribute to an integrated report. Each representative is responsible for protecting the interests of her home organization while contributing to the general mission. This approach succeeds by engaging skilled people to handle the organizational issues. Its limitation is bandwidth – the amount of intelligence information that can be reviewed in a day is limited. This method gives up the speed and comprehensiveness of automatic analysis in order to preserve maximum safety for the data. It is a safe but limited alternative.

2.4 Web Methods

A recent trend attempts to bridge the capability gap by implementing "open source spying" wherein intelligence information is entered into a universal system based on Web 2.0 concepts. In such open source inspired systems, the users – the intelligence information analysts – create integrated intelligence by contributing, reviewing, and cross-referencing information. The Intellipedia is the most prominent example of such a system and like all such systems it cites the successful adoption of Internet such as Wikipedia, Google and on-line issue-oriented forums. A good early discussion of this trend was reported in "Open Source Spying" by Clive Thompson, in the New York Times December 3, 2006. Director of National Intelligence Mike McConnell announced recently the deployment of the concept as reported by Siobhan Gorman in the Wall Street Journal Jan. 22, 2008. This deployment elevates data integration to the highest mission priority while downgrading the protection of original sources and privacy concerns. It may be too drastic a step. One possible outcome of the deployment is that sensitive data will be held back instead of included in the intelligence analysis effort.

An additional risk is that the Intellipedia succeeds too well and falls prey to the public outcries that doomed the Total Information Analysis effort in the early 1980's.

The next two sections discuss common technical proposals that partially address the intelligence analysis shortfalls.

2.5 Shortfalls in Anonymous Data

A common proposal - particularly in response to privacy laws - advocates replacing any reference to personal identity with a number. The result is called anonymous data. The number that replaces the identity is a key into a special, highly secure database table that is not distributed. According to these proposals, it should then be possible to share depersonalized data without concern. The NIH uses such an approach for clinical research studies. The proposal is sufficient if the purpose is compliance with the letter of privacy laws. When data protection is essential to the mission, it is insufficient.

Anonymous data is vulnerable to the exposure of personal identities when it is combined with other data. The other available data can be correlated with the anonymous but unencrypted data. Matches with the anonymous records can frequently reveal the identity. For a published example see “A Face Is Exposed for AOL Searcher No. 4417749”, by Michael Barbaro and Tom Zeller Jr, NYT, August 9, 2006.

Moreover, important forms of data abuse do not require identity. For example, an enemy agent with access to anonymous data can study the information to see relationships that indicate the source of the data. For example, in the aftermath of the current war in Iraq, U.S. Intelligence information was made available to a trusted Iraqi who eventually shared it with Iran. Iranian intelligence analysts were immediately able to identify the American information as derived from a poorly encrypted Iranian communication system. The Iranian system was immediately upgraded depriving the U.S. of valuable intelligence intercepts.

2.6 Shortfalls in Encrypted Databases

Other proposals involve the encryption of the data fields of a relational database using a selected encryption key. Each compartment in the intelligence enterprise would know and use the same selected key. The encrypted database located in a compartment could then be made available for searching with the confidence that information records are protected by encryption. Any of the compartments can now search any other compartment’s database by an encrypted query procedure.

Proposals for encrypted databases are not responsive to the shortfalls described in this ORD. They allow broad search opportunities that lead to data abuse. In their architecture, they allow a wide distribution of the secret key; therefore, the key is not sufficiently secure. Furthermore, if an outsider finds a method to access the encrypted database, the outsider can apply code-breaking methods to the entire database potentially allowing the encryption to be broken. Any secure system must limit the amount of encrypted text available to a code-breaker, but the encrypted database proposals place no limits.

Finally, existing proposals for encrypted databases ignore the special problems of biometric data. Biometric data is obtained from living, changing people using measurement instruments under variable field conditions. The results are not exactly reproducible from one measurement to the next. All of the existing encrypted database proposals depend on the fact that two values will be exactly equal after encryption if and only if they were exactly equal before encryption. For biometric data, however, we seek a near or close match of values. Two values that are close before encryption become far apart after encryption and will not match. Biometric data are assuming more and more importance in intelligence analysis and law enforcement and biometric data must be supported in future systems.

3.0 Capabilities Required

The following table provides an overview of the required capabilities. The capabilities are listed and explained in greater detail in the sections below. Each overview capability cites references to the more detailed list in the text below.

Table 3-1 High Level Capabilities for the CCIA

High-level Capabilities	**Detail**
The system provides an intelligence analysis service operating on information contained in two or more secure compartments without requiring any release of unencrypted information.	GC-1
The system will be capable of secure operation when the operational components of the system are housed in geographically distant, secure facilities.	GC-4
The intelligence analysis service is housed with and operated by a matchmaker who cannot decrypt the input data and who, moreover, cannot decrypt the results of matching.	GC-7 QM-1 QM-2
Secure data compartments encrypt all information before providing it to the intelligence analysis service.	GC-2 KM-1 KM-2
The intelligence analysis processes shall link related encrypted information records and provide a copy of each linked composite of records to any of the secure data compartments that contributed a portion of the composite record.	GC-3 GC-10
The system shall manage the keys used to encrypt data before intelligence analysis and all system process maintain a separation of key and encrypted data to prevent decryption outside of the secure compartments.	KM-2
The system will utilize asymmetric key encryption or an equivalent to restrict delivery of all documents to the intended recipient, to verify the identity of the sender, and to verify the integrity of each document's content.	GC-6 GC-5
The system will support the operation of a special investigative unit to audit suspicious activities or reported issues in the system.	GC-9 GC-8 KM-3
The system shall be able to perform useful searches and comparisons on encrypted biometric measurements.	EC-3

3.1 General Capabilities

GC-1: The system shall be capable of intelligence analysis on specially encrypted information records without having available a key to decrypt the records. This requirement protects the providers of the encrypted records from unauthorized disclosure of information by individuals operating the intelligence analysis process. The encryption key required to implement this requirement shall be known henceforth as the "innermost encryption key" to distinguish it from other keys which may be implied by the capabilities. The intelligence analysis operation will operate on fixed sets of encrypted information provided for a specific intelligence analysis session. The system will perform frequent intelligence analysis sessions to reduce latency in responding to new information. Between the discrete intelligence analysis sessions, the encrypted information may be replace or updated according to the needs and policies of the information-owners.

GC-2: At each facility that maintains a secure database of intelligence data, the system shall perform a special encryption process that applies the innermost encryption key to information records so that specific terms and values in the records are encrypted but general terms, metadata, and/or keywords are not encrypted. The special encryption can be regarded as partial encryption because information about the syntactic or grammatical structure of the record is not encrypted. It will be possible to perform meaningful intelligence analysis on these specially encrypted records because their syntactic structure is recognizable.

GC-3: The intelligence analysis process will recognize related information records from two or more sources and fuse them by providing links between records and annotation on the links explaining the relation. The fused records contain encrypted information that cannot be interpreted by the operators of the intelligence analysis process.

GC-4: The system will be capable of secure operation when the operational components of the system are housed in geographically distant, secure facilities.

GC-5: The system will incorporate and use the services of an outside identity authority that will positively and securely identify the components of the system by associating each component with the public key of an asymmetric encryption key pair.

GC-6: The system will provide a means to transfer information encrypted by the innermost encryption key to the operational entity that operates the intelligence analysis process in a manner that positively identifies the source of the data and that prevents all parties except the intended party from receiving and using the information. At minimum, a conforming system will apply two layers of asymmetrical key encryption to the information records that have already been specially encrypted with the innermost encryption key (see GC-1 and GC-2). These two additional layers of encryption use the private asymmetric key of the data source and the public asymmetric key of the intelligence analysis operator.

GC-7 The intelligence analysis process is operated by an organizational entity acting as a matchmaker or broker on behalf of two or more organizational entities that are sources of intelligence data. The system will enable source entities to control the information presented for intelligence analysis at a level of detail that includes the following. First, one source, call it A, may control the information presented for intelligence analysis with information from another source, call it B, where A and B may be any two distinct sources for the system. Second, each source may change the information presented for each session to include new information and to reflect any changes in the source's policy on data inclusion.

GC-8: The intelligence analysis process shall store any results derived by intelligence analysis. The results are encrypted but are annotated with the time and other production details.

GC-9: There shall be an option for a special purpose investigative entity that does not participate in routine intelligence analysis operations. The optional investigative entity may review the results of an intelligence analysis session if it obtains the cooperation of the intelligence analysis operator and one of the members in a "comparison set". The term "comparison set" is explained in KM-1 below.

GC-10: The fused information records (see GC-3) shall be provided to all members of the comparison set. Each member of the comparison set is in possession of the innermost encryption key and has the capability to decrypt the fused information records and act on information contained in it.

3.2 Encryption Key Management Capabilities

KM-1 The organizational entities that are sources of information for the intelligence analysis process shall agree mutually on comparison sets. A comparison set consists of two or more distinct organizational entities that have agreed to submit their results together for intelligence analysis. If the comparison sets have the minimum size of two entities, then the number of sets may be large. For example, if A, B and C are organization entities supplying intelligence data for intelligence analysis then comparison of all data sources requires three comparison sets: [(A,B), (B,C), (A,C)]. If the comparison set is expanded to three sources, then obviously one set is sufficient for A, B, and C. Because intelligence analysis is based on mutual agreement, some potential comparison sets may be missing. For example, the pair (B,C) will be missing if either B or C refuses to work with the opposite entity.

KM-2: There is a mechanism by which members of a comparison set can generate and agree upon the innermost encryption key to be used for each session of intelligence analysis. The mutually agreed innermost encryption key is never revealed outside the comparison set except under the investigative procedure referenced in GC-9 and described further in KM-3.

KM-3: An organization entity may be asked to cooperate with an investigative entity by providing that entity with the innermost encryption key that was used for a particular comparison set and a particular intelligence analysis session. To support this cooperation, each entity will retain a record of the keys used with its data; furthermore, the system will provide a secure method to transfer an encryption key to the investigative unit.

3.3 Query Management Capabilities

QM-1: The system will be capable of executing general intelligence analysis algorithms. When these algorithms are applied to the specially encrypted information records (see GC-2), the fused data records will contains a link (see GC-3) that is annotated with a reference to the algorithm that produced the link.

QM-2: The system will accept queries from any source in a comparison set (see KM-1). Such queries will be written in a standard query language such as SQL. Each query will be encrypted at the source as provided in GC-2 and the intelligence analysis process will execute the query against encrypted data provided from the set of sources, that is, the comparison set. If the query yields results, the encrypted, fused record contains both the query and the result. Note that members of the comparison set do not receive copies of such queries unless the query produces a result. The purpose of this capability is the initiation of sensitive lines of inquiry without unnecessary disclosure of the query.

3.4 Encrypted Comparison Capabilities

EC-1: The intelligence analysis software shall have a capability to compare encrypted values for equality. This capability is trivial and is mentioned only to point out a basic principle of encrypted information analysis.

EC-2: The intelligence analysis process will have the capability to test an encrypted value for membership in a predefined set of values. This capability might be used, for example, to test whether a name is one of several known aliases.

EC-3: The intelligence analysis process will have the capability to test a concealed value belonging to the set of real numbers and determine whether the concealed value lies in the bounds of a concealed range. This capability might be used, for example, to compare biometric measurements obtained under difficult conditions for the purpose of finding match between biometric identity records. For such values, an equality test (EC-1) would fail because the measurements are imprecise. The term “real numbers” should be interpreted to mean the representation of numbers by floating point computer representation. The method of concealment shall apply the innermost encryption key to the floating-point value. The result of this concealment operation shall be considered encryption for the purposes of GC-2. Note that concealment is not equivalent to encryption under some formal definitions of encryption.

EC-4: The intelligence analysis process shall have a capability to perform intelligence analysis operations on partially encrypted XML statements. An XML statement is partially encrypted if the values are encrypted but the XML tags names are not.

EC-4: The intelligence analysis process should be capable of searching encrypted relational database tables where the table values are encrypted but the table schema is not encrypted. The search shall execute a partially encrypted SQL statement where all values are encrypted but the SQL keywords are not encrypted.