ARMY RESEARCH LABORATORY

Fusion of Asynchronous, Parallel, Unreliable Data Streams

by Ann Bornstein, John Brand, Michelle McVey, and Andrew Neiderer

ARL-TR-5344 **September 2010**

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5067

ARL-TR-5344 **September 2010**

Fusion of Asynchronous, Parallel, Unreliable Data Streams

Ann Bornstein, John Brand, and Andrew Neiderer
Computational and Information Sciences Directorate, ARL

Michelle McVey
George Washington University College Qualified Leader (CQL)

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
September 2010	Final	June 2008–June 2010

4. TITLE AND SUBTITLE
Fusion of Asynchronous, Parallel, Unreliable Data Streams

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Ann Bornstein, John Brand, Michelle McVey,* and Andrew Neiderer

5d. PROJECT NUMBER
0TEDTC

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: RDRL-CII-C
Aberdeen Proving Ground, MD 21005-5067

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-5344

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

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

13. SUPPLEMENTARY NOTES
* George Washington University College Qualified Leader.

14. ABSTRACT

This report documents progress in an investigation of the fusion of parallel data streams composed of dissimilar data types of varying reliability, arriving at different times. The technique will serve to determine resemblance between objects, in this case, people, which can be described by a stream of data expressed as state vectors. Objects described by similar state vectors will group together. The data fusion methodology is based on the use of multidimensional scaling to evaluate data composed of attribute vectors of high-spatial dimensionality. A focus problem has been used for development of the technique, identification of high-value individuals (HVIs). The technique applies to other problems as well; the identification of HVIs is only one topical application. Data available to field operators and analysts includes traditional identification information, biometric identity files, situational and observational data, archival data, and intelligence information. These data may be used to classify individuals in terms of resemblance to key groups and cue an analyst to whether they are HVIs. A proof of concept of the methodology has been conducted with promising results.

15. SUBJECT TERMS
multidimensional scaling, groups, unreliable data

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Ann Bornstein
Unclassified	Unclassified	Unclassified	UU	70	19b. TELEPHONE NUMBER (*Include area code*) 410-278-8947

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Acknowledgments

The authors wish to thank Richard Kaste for his invaluable advice and suggestions. The project would not be the same without his contribution.

INTENTIONALLY LEFT BLANK.

1. Introduction

A serious problem in stability operations is the determination of whether a person is of interest. The importance of the person to operations is expressed as the person's value: a person assigned a high value (high-value individual, or HVI) is given more attention, whether it be surveillance, covert monitoring, or even detention. A method of determining value by estimating the likelihood of membership in a key group, such as a terrorist cell, that can be implemented either in real or near real-time in support of a field operation or as a tool that continually parses a database as a background process would be invaluable. A method for fusing data from asynchronous, parallel, unreliable data streams is under investigation.

Much current analysis of groups and evaluation of an individual's value given that group role is based on knowledge of functional links—A controls B and C, B provides a service or materiel to D, and C is in touch with D and E by cell phone. Functional link analysis requires knowledge of a subject's hidden actions and social milieu. That is, with the human intelligence (HUMINT) data needed for functional analysis, such as intercepts and informer reports, the analyst already knows the subject is a person of interest.

A great deal of overt information may be available on an individual, ranging from observable and documentary information obtainable by examination at a checkpoint to information available through documentary records of some kind. Information available from records may include intelligence reports, civil data (birth records, residence permits, etc.), criminal data, national identification data, and so on. Identification (ID) of the individual forms the link between someone standing in front of a field operative and the documentary information.

There are several field personnel identification devices presently in use. The availability of different identification devices allows focus of multiple differing data streams concerning an individual on a single problem set, identification of individuals, and evaluation of the importance of those individuals as possible members of groups of interest. The fielded identification devices use traditional fingerprints, facial imagery ("mug shots"), and iris scans to allow the Soldier in the field to check an individual's identity records, if any, against the situational context and the identity documents in the individual's possession.

These identity data enable integration of different data streams that may consist of quantitative, qualitative, dichotomous, or even rank-order data. Examples might include the fit of several facial identification characteristics to an image captured from different aspects than the reference image, fit of identification features of a captured fingerprint to a set of reference prints, voice recognition features, and physiognomic measures, such as weight, height, and bodily feature proportions. Observer judgment may also be included, perhaps via a subjective, albeit

numerically expressed, confidence rating scale. In addition, the technology for sensing physiological conditions, such as stress, widely used in applications such as home exercise equipment, may allow estimation of identity issues as well as the potential for immediate, personal violence.

The basis of the analysis is thus a set of data elements reflecting personal characteristics, observables, intelligence data, and civil records. These data elements may be of different types: nominal, ordinal, interval, or ratio. The information in the component characteristics of the descriptive data set must, at present, be expressed numerically for analysis.

2. Information in the Personal Data Vector

The personal data vector may have data elements that are themselves the result of data fusion. In general, the personal data vector will contain the following:

- Situational information, including claimed identity.
- Observational information, including human recognition (observer ID) and observational evaluation of behavior (nervousness, etc.).
- Biometric information, including biometric identity determination (e.g., fingerprints).
- Documentary information, including personal identity documents and archival identity records.
- HUMINT information.

This analysis required generation of a population database with notional personal characteristics and the resulting attribute vectors. The rough distribution of types of information is illustrated in figure 1. The situational information includes, for this study, ground truth information used for development of the database, such as the information development sequence, a miniscenario, which would not be present in an operational database. For instance, a person is detained at a checkpoint based on human observation of the person's nervousness, visual descriptive information is noted, a biometric identity kit is used to take biometric identity data and measure biometric stress indicators, identity documents are checked against the biometric identity database, and any HUMINT information is obtained.

The data describing these kinds of information may be characterized as nominal, ordinal, interval, or ratio. A methodology to analyze attribute vectors must accept or compensate for these differing measurement scales.

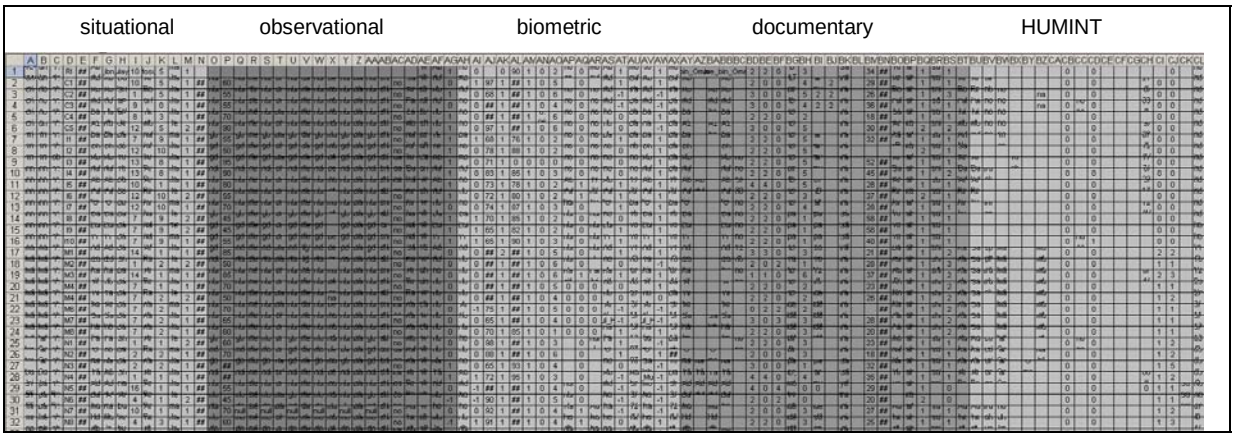


Figure 1. Illustration of the different types of information in the notional persons database.

2.1 Identity

There are several kinds of information that can be used to evaluate identity (i.e., a person can be identified several ways). For instance, a detained person may identify him or herself. This claimed identity may be true, or an accepted variant in that culture's naming conventions, but may also be false or misleading. Due consideration must also be given to nicknames, ambiguous naming conventions, and difficulties in transliterating local script. A person may have several "identities," such as an intrinsic or birth identity, a claimed identity, a local nickname identity, a social role-based identity, false or deceptive identities, and possibly others.

For instance, consider the case where there are several encounters with the same person, under different circumstances and with, perhaps, several names offered and incorporated into the record of the encounter. An example might be a notional Iraqi man, born Umar Zukeed. He registers for conscription as Umar ibn Zukeed, is drafted and released from service, is arrested by local police and detained as Umar al-Tikriti, and is stopped at a checkpoint and registered with biometric characterization but with false identity papers under the name Abdul Karim. A fingerprint left at a bomb scene and attributed by informers to Fath al-Din creates an intelligence record under the name Fath al-Din, which is later linked by the fingerprint left at the bomb site to the original Umar ibn Zukeed's military conscription record, to civil police arrest records under the name Umar al-Tikriti, and to the biometric registration at the checkpoint as Abdul Karim. The collation of records by fingerprint gathers all these records together.

These records may be considered "personas" of the same "underlying or intrinsic identity or personality." The personas and the underlying or intrinsic ID the person was born with may be linked in a number of ways. *

*The term "ID" is commonly used for "identification document" and "identity." The two are not the same thing, though in common usage they are assumed to be at least consistent. In practice, they are not always. In this document the specific meaning of ID is determined by context.

However, “person” is also the term applied to an individual standing in front of a field operative or described by a record of some event, even though the actual identity or even the claimed identification of that individual is uncertain or unknown. That “person” will claim an identity that may or may not be the intrinsic ID.

Similarly, when Umar is arrested carrying false ID documents as Abdul Karim, the fictitious name is also an “alias.”

For the purposes of this investigation, the underlying or intrinsic ID is considered to be the oldest documented identity of a person. Ideally, the intrinsic ID is the one registered at birth through a birth record, if one is available. In other cases where birth records are not available or traceable, the intrinsic ID may be based on a record such as some other official identification document or, if background documentation cannot be found, based on registration in a U.S.-controlled identity database. The oldest documented identity may contradict the identity or persona presented to the field operative.

Other data may be necessary to disambiguate different encounters. A persona may be described by associated data elements, including observation of physical appearance and behavior, biometric data, documents, identity or historical data in intelligence or criminal databases associated with that persona through physical or documentary ID. This disambiguation may be facilitated by a separate entry, subordinate to the relevant persona, linked to the appropriate intrinsic ID, for each encounter, whether physical or documentary.

One possible taxonomy of the identification hierarchy is shown in figure 2. Four encounters with a notional person are shown diagrammatically. The person has several personas, including a false identity. Encounters with the different personas are shown—in two cases, biometric ID (bio-ID) leads to the underlying identity. In one encounter, a bio-ID kit is unavailable; in another, the information available does not yet lead to the person.

In this study, the issue of multiple personas is not addressed, with one exception. In that case, an encounter with a person leads to an identification of the intrinsic identity, but the record is also notionally linked to a persona described in intelligence material whose identification is not supported by archival material. In this way, two personas are linked to one person.

One way of depicting the persona/intrinsic ID issue in the database might be as shown in figure 3. This is a modification of the notional persons database used in this phase of the investigation. Columns have been added for the intrinsic or birth identity common to three personas of a petty criminal. Each persona was generated by an encounter with occupation troops, and each, in turn, was linked to the same identity trail in Iraqi civil and military records.

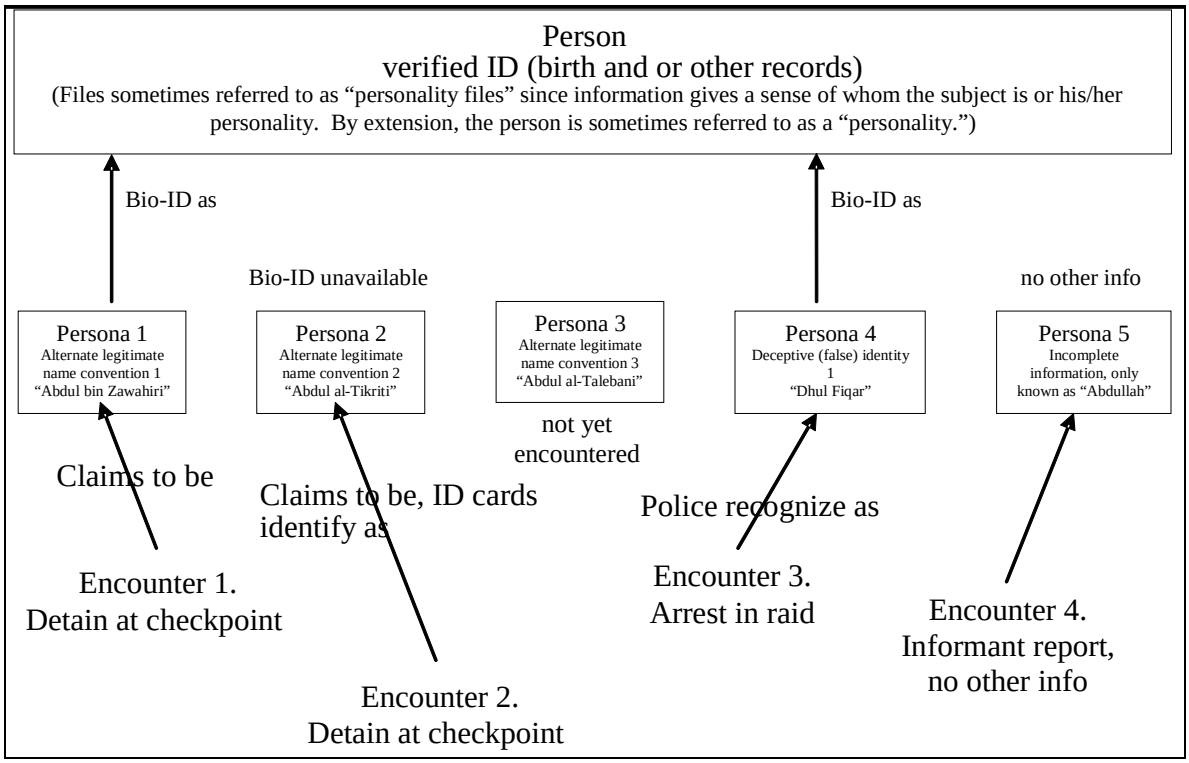
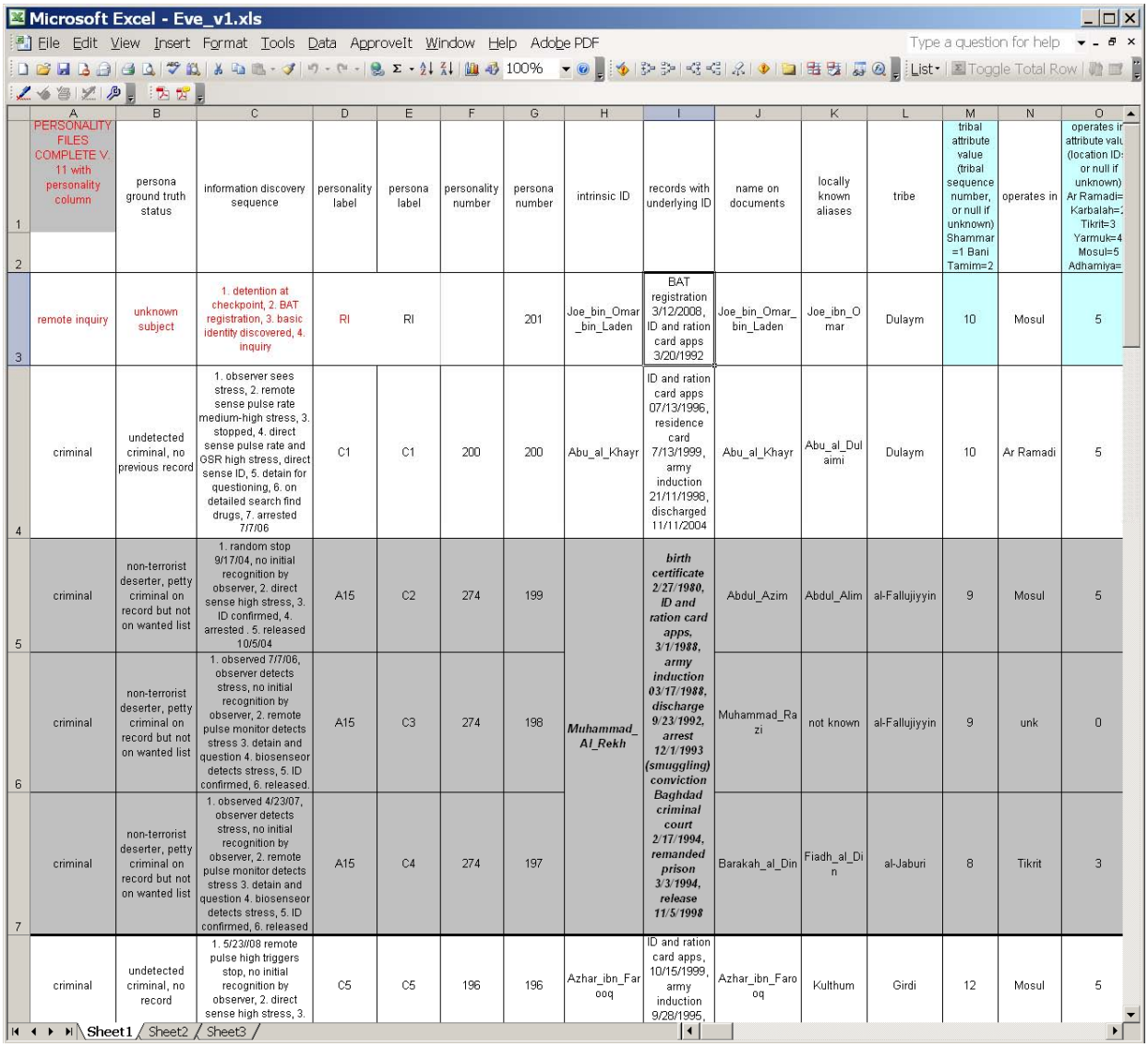


Figure 2. Possible taxonomy of the relation schema for the identification hierarchy.

A relational database is ideal for this sort of linkage. It also allows linkage of individuals to multiple entities, such as locations, persons, events, or documents and flagging of links with different indicators of confidence. A preliminary design indicates that use of a relational database for the next version may be the best option.

The analyst may have a different level of confidence in the link than in the identity documentation. For instance, if a high-confidence identification method, such as fingerprints, links an individual to a trail of high-confidence identity documents culminating in a birth record, the analyst may hold a great degree of confidence in both the link and the identity documentation and, hence, in the person's ID. If other, less accurate methods of identification, such as identification by facial imagery and testimony of people in the neighborhood, link an individual to a set of high-confidence identity documents, the analyst may not have as much confidence in the person's ID. For instance, a person in the neighborhood, Abdul Jalil, may be said by informants to be Abdul Fattah**,** a former high-ranking member of the Ba'ath Party, who disappeared after the occupation. The existence of the individual Abdul Fattah is not in doubt, but the identification of Abdul Jalil as Abdul Fattah may be considered unproven.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	PERSONALITY FILES COMPLETE V. 11 with personality column	persona ground truth status	information discovery sequence	personality label	persona label	personality number	persona number	intrinsic ID	records with underlying ID	name on documents	locally known aliases	tribe	tribal attribute value (tribal sequence number, or null if unknown) Shammar =1 Bani Tamim=2	operates in	operates in attribute valu (location ID= or null if unknown) Ar Ramadi= Karbalah=2 Tikrit=3 Yarmuk=4 Mosul=5 Adhamiya=
2															
3	remote inquiry	unknown subject	1. detention at checkpoint, 2. BAT registration, 3. basic identity discovered, 4. inquiry	RI	RI		201	Joe_bin_Omar_bin_Laden	BAT registration 3/12/2008, ID and ration card apps 3/20/1992	Joe_bin_Omar_bin_Laden	Joe_ibn_Omar	Dulaym	10	Mosul	5
4	criminal	undetected criminal, no previous record	1. observer sees stress, 2. remote sense pulse rate medium-high stress, 3. stopped, 4. direct sense pulse rate and GSR high stress, direct sense ID, 5. detain for questioning, 6. on detailed search find drugs, 7. arrested 7/7/06	C1	C1	200	200	Abu_al_Khayr	ID and ration card apps 07/13/1996, residence card 7/13/1999, army induction 21/11/1998, discharged 11/11/2004	Abu_al_Khayr	Abu_al_Dulaimi	Dulaym	10	Ar Ramadi	5
5	criminal	non-terrorist deserter, petty criminal on record but not on wanted list	1. random stop 9/17/04, no initial recognition by observer, 2. direct sense high stress, 3. ID confirmed, 4. arrested . 5. released 10/5/04	A15	C2	274	199	***Muhammad_Al_Rekh***	***birth certificate 2/27/1980, ID and ration card apps, 3/1/1988, army induction 03/17/1988, discharge 9/23/1992, arrest 12/1/1993 (smuggling) conviction Baghdad criminal court 2/17/1994, remanded prison 3/3/1994, release 11/5/1998***	Abdul_Azim	Abdul_Alim	al-Fallujiyyin	9	Mosul	5
6	criminal	non-terrorist deserter, petty criminal on record but not on wanted list	1. observed 7/7/06, observer detects stress, no initial recognition by observer, 2. remote pulse monitor detects stress 3. detain and question 4. biosenseor detects stress, 5. ID confirmed, 6. released.	A15	C3	274	198			Muhammad_Razi	not known	al-Fallujiyyin	9	unk	0
7	criminal	non-terrorist deserter, petty criminal on record but not on wanted list	1. observed 4/23/07, observer detects stress, no initial recognition by observer, 2. remote pulse monitor detects stress 3. detain and question 4. biosenseor detects stress, 5. ID confirmed, 6. released	A15	C4	274	197			Barakah_al_Din	Fiadh_al_Din	al-Jaburi	8	Tikrit	3
	criminal	undetected criminal, no record	1. 5/23/08 remote pulse high triggers stop, no initial recognition by observer, 2. direct sense high stress, 3.	C5	C5	196	196	Azhar_ibn_Farooq	ID and ration card apps, 10/15/1999, army induction 9/28/1995,	Azhar_ibn_Farooq	Kulthum	Girdi	12	Mosul	5

Figure 3. One way to display information linking separate identities or personas to a common, intrinsic identity.

This set of possibilities can culminate in a situation where there is no link to a document-supported, traceable ID, in which case, the analyst must consider the underlying ID unknown. If circumstances do not permit linkage of an individual to archival identity documentation—the encounter with the individual may not involve any identity biometrics—the lack of a provable underlying ID may not be significant. On the other hand, if the person is available for physical identification but has no traceable documentary history or background, the lack of an archival, documentary background may be very significant.

2.1.1 Types of Personal Information Available

Initially, for this methodology development, a set of personal ID information streams has been proposed based on current technology and current operations. The intrinsic identities of the

notional persons for this study are those identities linked to the persons under investigation by the birth record or other civil records. In the event no birth record is available, the oldest identity document is considered to reflect the intrinsic ID.

Regardless of the taxonomy of identities and personas, interpretation and fusion of the identity information data streams in real operations is likely to require expertise beyond that available in a tactical setting.* Field identification systems presently in use in Iraq include the Biometric Automated Toolset (BAT), which is used for registry and ID confirmation, and the Biometric Identification System for Access, which is used for control of base access (*1*). These systems record and use fingerprint, facial imagery, and iris pattern data. As of 2007, the field teams did not have direct uplink of information to central data management facilities (*2*). Connection with the Iraqi government automated fingerprint system does not appear to be automatic as of this writing. The information environment assumed for this study is described in appendix A.

This information flow is illustrated in figure 4. The link between the field and support elements is depicted as a satellite link, but the link may be any link of capacity sufficient to deal with the data load. The field element is shown using personal observation and biosensory tools to gather information on an individual. The individual may be detained, perhaps at a checkpoint, or under observation and may or may not be aware of the observation. Biometric identity tools may produce on-site positive ID or, as shown in this case, the user of the biometric tool may require help from an analyst with access to records gathered in a central location. In this scheme, the fusion would be performed using multidimensional scaling (MDS), and the analysts' conclusions communicated to the field element.

An example of the information support function is the assessment of the importance of an individual with whom the friendly unit comes in contact. The information at the base area could include intelligence reports, dossiers and identification characteristics from other, possibly civil law enforcement, databases. The information provided by the subject in the field could include description, video imagery, fingerprints, or possibly voice clips. Information may be gleaned from several other sources: (1) objects such as documents produced by or discovered by search of the subject's person, (2) contextual information such as presence next to a partially completed bomb, or (3) other individuals that could include informants, neighbors, relatives, or other kinds of witnesses.

Identity data link the subject with the relevant situational and record data. Biometric ID provides an objective assessment of identity by reference to established identity databases and also allows confirmation of the identification. Some biometric data may allow estimation of subject veracity and, hence, estimation of the veracity of self-identification independent of identity databases.

*In fact, there is presently a data fusion center operating in support of identity data for operations in Iraq, the Biometrics Fusion Center. See http://www.wvbiometrics.org/department-of-defense-biometrics-fusion-center.html, accessed 25 January 2007. This center, located in West Virginia, responds to inquiries from the field.

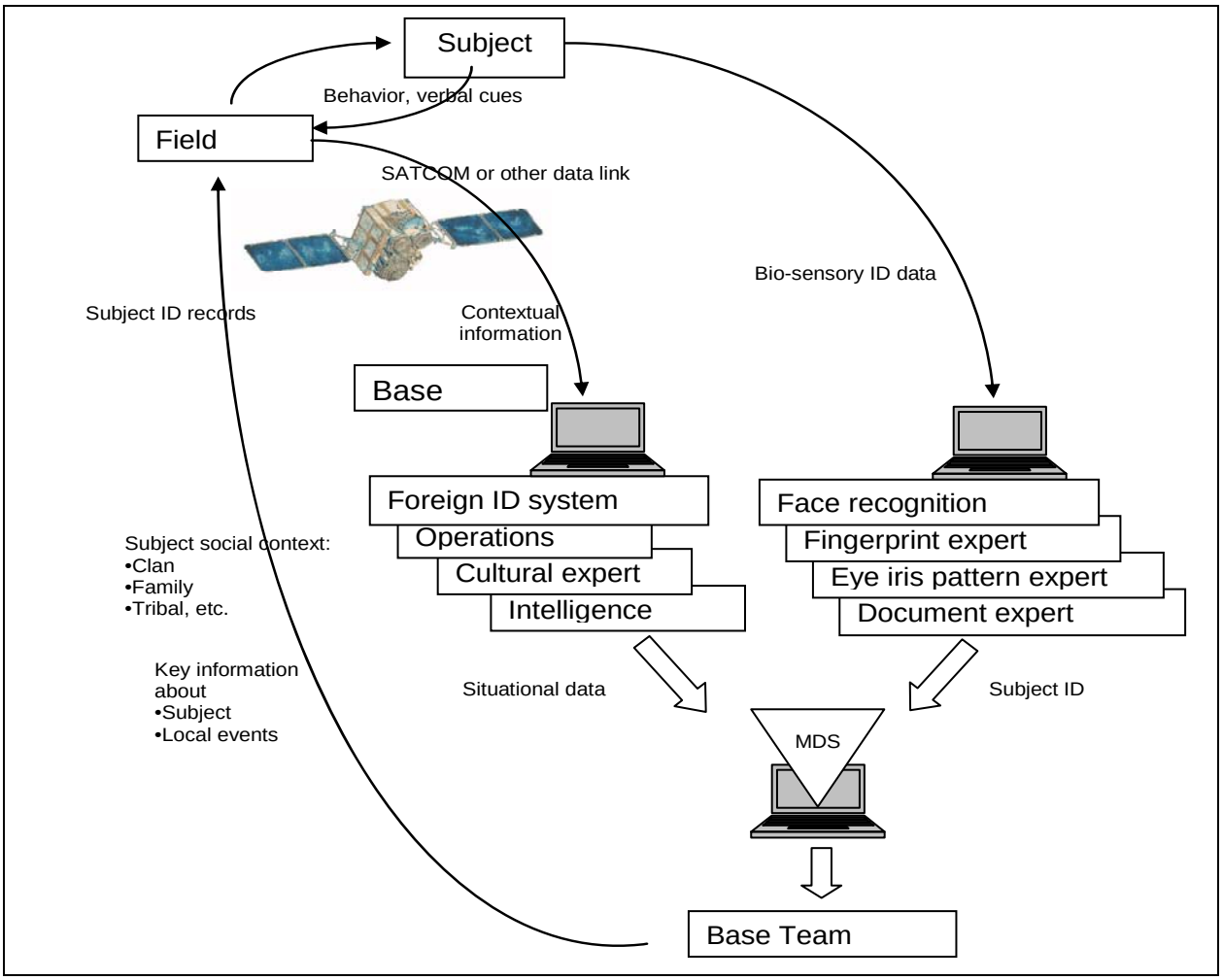


Figure 4. Information flow concept in support of a field unit inquiry.

The latter is especially important where identity databases are rudimentary, of questionable reliability, or entirely lacking. The latter will also constitute a new branch of biometric identification.*

*Presently, there are two branches of biometric ID systems: verification of identity and determination of identity. "A biometric [ID] system can be designed to test one of only two possible hypotheses: (1) that the submitted samples are from an individual known to the system; of (2) that the submitted samples are from an individual not known to the system. Applications to test the first hypothesis are called "positive identification" systems [verifying a positive claim of enrollment], while applications testing the latter are "negative identification" systems [verifying a claim of no enrollment]" (*3*). [Interpolations ours.]) Verification of identity is performed by reference to a set of characteristics from a set of cooperative enrollees. Forensic identification is performed by comparison of subject characteristics to a central repository of personal signature data. If reliable self-identification can be realized, it will constitute a third branch of bio-ID. This third branch may be of great importance in the present theater of operations. This is potentially enormously valuable information, and is summarized by the question, *"Are you who you say you are?"* It should be possible to phrase the question, probably with additional questions, so as to lend credibility to an individual's declaration of identity. The question might also be accompanied by some variation of *"What are you doing here?"*

2.1.2 Human Observational Input

A number of personal characteristics included in this investigation are based on human input. These include observer recognition, identification, and stress evaluation. A trained observer may be able to recognize individuals from imagery files of persons of importance—"mug shots." An observer may also use cues, such as gait or posture, perhaps subconsciously, to determine identity. Observer estimate of stress is also an important factor both for cuing the observer to the possible importance of the subject and for estimation of the truthfulness of any documentary evidence or claimed personal information by the subject. An observer may be cued to falsity of identity documents by observation of subject cues such as eye movement, appearance of superficial blood vessels, posture, "fidgetiness," and a host of other cues (*4*). Some of the techniques are discussed in appendix B.

2.1.3 Machine Input

Biometric technology to evaluate stress both remotely and directly exists. Likewise, biometric identification technology is readily available. An analyst need not be an expert in biometric technologies to use the data produced by them, but an analyst must have at least a cursory knowledge of the technologies to take into consideration factors such as reliability, applicability in different situations, usability in the field, the types of information produced, and so on. A meaningful scenario cannot be developed without this knowledge, and the mathematical methodology for analyzing and interpreting the data produced cannot be chosen. A short discussion is, therefore, in order.

2.1.3.1 Stress Indicators. Biosensory stress indication is a fairly mature technology. Inclusion of these data is, in principle, quite easy and so will be considered as part of the personal attribute vector.

There are several possible channels that might be used. The two channels chosen for this study, galvanic skin response (GSR) and pulse rate, are convenient and reasonably well understood. Pulse rate lends itself to remote sensing—a good microphone can pick up heartbeat. GSR and pulse rate can be picked up directly by clipping a small sensor on a finger. Neither is a perfect predictor of stress, but they serve as convenient illustrative representatives of the general technology for this study.

Subject stress is extremely important as stress may serve as an indicator of impending violence. In an era of suicide vests, evaluation of subject stress may save many lives.

2.1.3.2 Biometric ID Inputs. Biometric ID is a reasonably mature technology. There are several biometric ID sets available. The presently fielded biometric ID sets used in Iraq employ fingerprints, measurement of iris characteristics, and facial imagery. Voice recognition is used in identity confirmation where conditions are ideal—noise level, previous registration, control of the phrases used, willing cooperation by the person wishing identity confirmation—but forensic

ID is more difficult. It may not be feasible in a field setting and a near real-time environment but is discussed for the sake of completeness. Biometric ID is discussed at greater length in appendix C.

2.1.3.2.1 Fingerprints. Fingerprints are the best standard for personal identification. They require a comparison fingerprint database. If the fingerprint matches one of the prints on file in a database, an identification of the individual with the identity registered at the time the print was taken may be assumed with high confidence dependent on the number of points of similarity. Matching of the print in identity confirmation systems is fast because the number of prints to compare is small. Matching of a fingerprint in a forensic role can be time consuming, may require access to several disparate databases, and is best left to an expert.

Presently, there are fingerprint records available through the Iraqi Automatic Fingerprint Identification System (AFIS), which "has ~750,000 records in its database, including ~280,000 criminal records which were captured prior to coalition forces taking control of the country, and these records were fingerprint cards—ink fingerprint cards that were scanned into [the] system" (*1*).

2.1.3.2.2 Iris Patterns. The patterns in the human iris are being increasingly used as an identification method. They obviously cannot be left as latent evidence at a crime scene, but their value in checking a person against a central registry appears to be substantial. The descriptive statistical properties of the pattern are compared with the statistical properties of the iris pattern information contained in the registry (*5*, *6*).

2.1.3.2.3 Facial Metrics. Facial metrics allow trained observer recognition and use of face recognition algorithms. These algorithms are presently of moderate reliability, at best, but are based on facial proportions not easily altered. Facial recognition obviously depends on, in the case of expert human assessment, a set of "mug shots" or surveillance videos of high enough quality to be useful and, in the case of automatic recognition algorithms, digitized facial imagery. Such imagery can be gathered using visible and thermal infrared; the thermal IR video may allow detection and penetration of disguises. Use of such artifices as disguises would, of course, also flag a person for detention and investigation.

2.1.3.2.4 Voice Records. Voice records may be used in an identity confirmation role where the subject is cooperative and has previously registered a high-quality excerpt that will be compared with an identification sequence based on an identical script. Voiceprints are generally difficult to use in a forensic role. They depend on the existence of a voice record of high enough quality to make the voice characteristics recognizable and the willingness of the subject to speak a similar phrase in a natural way. These circumstances may not occur frequently enough to be worthwhile but may be worth considering as a subject of very high value may be known only by voice intercepts and possibly fictitious names. Certainly if a person stopped at a checkpoint while trying to enter an important area refuses to speak to the observer, that is important enough to flag the individual for detention and further investigation.

2.1.4 Automated Identity Sets

2.1.4.1 Biometric Automated Toolset. The BAT is presently in use in Iraq. The BAT is a system composed of a digital camera, a fingerprint scanner, and an iris scanner, linked to a Toughbook* computer. The system is used to register or identify individuals in Iraq and Afghanistan. The individuals' estimated or declared weights and heights are recorded for the personal file. The biometric characteristics of individuals can be compared to a large database resident on the computer or downloaded to a centrally managed file for comparison with other databases (*7*). Presently, over 2000 BAT systems are deployed, and the identity database contains over 560,000 enrollments. A BAT is shown in figure 5.

Figure 5. BAT principal components are as follows: digital camera, Toughbook computer, fingerprint reader, and iris scanner.

The BAT database contains pictures, iris scans, fingerprints, and physical measurement data. The ability to do on-site, real-time biometric comparison is impressive, as is the ability to exchange information with a central repository (*8*).

2.1.4.2 Handheld Interagency Identity Detection Equipment. The Handheld Interagency Identity Detection Equipment (HIIDE) is a very portable device with fingerprint scanner, camera, iris scanner, and "biographical contextual data." The HIIDE is "interoperable with BAT for biometrics data exchange back to DoD biometrics [sic.] Data Repository." Plans are to field 6664 devices. The device is held in two hands and is 5 in high × 8 in wide. It is shown in figure 6 (*8*).

*Toughbook is a registered trademark of Panasonic Corporation of North America, One Panasonic Way, Secaucus, NJ 07094.

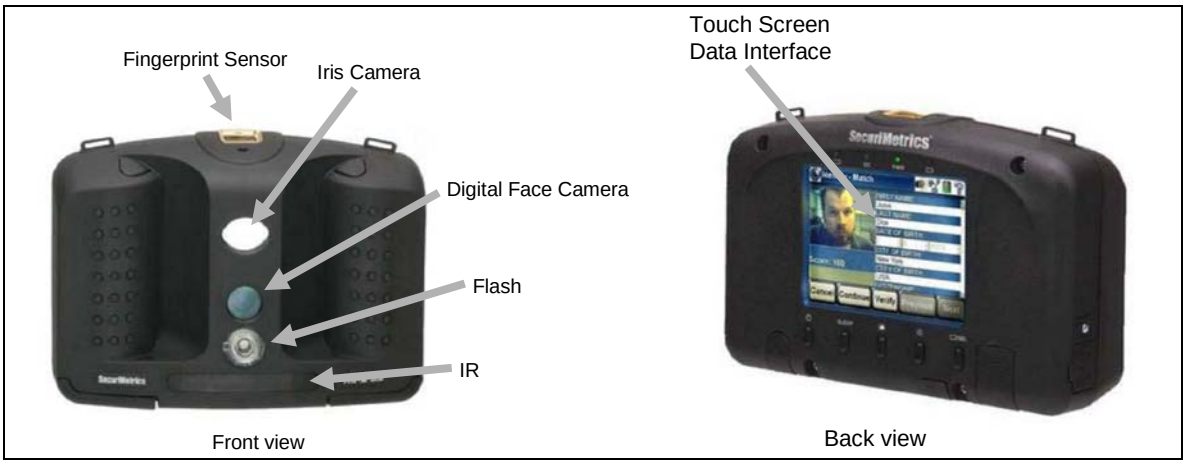


Figure 6. The Handheld Interagency Identity Detection Equipment (HIIDE), front and back views.

2.2 Identity Documents

Identity documentation can include several kinds of identification papers. In the case of Iraq, identity papers may include the following:

- Iraqi Nationality Certificate
- Iraqi Civil Status ID
- Residence Address Card
- Food Ration Card
- Birth Certificate

The Food Ration Card "is a very important form of ID and can be checked relatively quickly by the Iraqi authorities." Some individuals may produce a passport as an identity document as well. An extract of a description of the use of Iraqi identity documents is provided in appendix D (*9*).

Advice on the ways in which these documents may normally be found in everyday use will be necessary; this may require assistance from intelligence sources. For instance, does every person stopped at a checkpoint usually have all of these identity documents? If a person might not ordinarily have all of these identity documents, which are most likely to be in their possession? Is a forgery of any of these easy to detect? Do enemy operatives usually have one but not others?

2.3 Records

Records include databases in U.S. Government possession; Allied Government possession; publicly available documents such as those published in print, broadcast, and on the Internet; and archives of captured documents. U.S. Government data include intelligence, theater civil affairs,

and law enforcement data. These possess a records aspect and are also a dynamic database; for this study, both the working files and the archival or historical data are considered "records." For this study, only U.S. intelligence data and host country material are considered. This restriction was imposed for reasons of data manageability. A convenient taxonomy of the overall database or record structure is discussed later.

2.3.1 U.S. Government Data

U.S. Government information sources available now include the following types of intelligence:

- Biographic
- Economic
- Sociological
- Transportation and telecommunications
- Military geographic
- Armed forces
- Political
- Science and technological (*10*)

Local and international law enforcement data may also be available. This is an enormous span of information to consider. A team approach is inevitable for successful exploitation and use of this range of information sources.

2.3.2 Host Country Data

For this study, host country data is assumed to be available. Host country data is assumed to include civil registration data, criminal law enforcement data, secret police data, and military administrative records data. All of these databases are assumed to be available and keyed to the persons under consideration by intrinsic ID or by fingerprints. Fingerprint data, in turn, allows access to civil and military administrative data; the earliest record concerning a person—ideally the record of birth—is assumed to be the intrinsic identity.

2.4 Deception

The presence or absence of deception in an encounter would be an important cue to the value of an individual. As of this writing, the suspected presence of deception other than false documents is not a component of the personal attribute vector, as deception is so difficult to detect. Inclusion of deception would be very useful in a later study, given practical advice from the field.

The employment of deceptive countermeasures to identification would be an extremely important input to an individual's attribute vector. These countermeasures may include disguise and forgery. The employment of identity countermeasures, if detected, may serve as a marker of the individual's importance. Successful deception of the many biosensory identification channels available seems problematic. Deception of fingerprint systems seems unlikely. If the person is dirty enough, fingerprint sensors may give inaccurate readings, but dirt can be removed. A dirty finger is unlikely to give a false identification; if the fingerprint is too badly obscured, the counter-countermeasure is a bit of soap and water. Fingerprint identification spoofing of a computer identification system used for access control has even been the subject of a TV reality show, but the deceptive method shown on TV would be obvious to a human examiner (*11*).

Nevertheless, even if any particular biometric channel might be susceptible to countermeasures for a few combatants given special training, all enemy combatants are unlikely to receive such training. In addition, there will be several channels of information in use simultaneously, each of which must be flawlessly countered in a different way and countered so that no channel actually appears under attack. Detection of countermeasures will draw attention to the individual, defeating the purpose of the countermeasures.

Forgery is an ongoing deception problem in all identity operations and is beyond the scope of this report. Identification of an individual by fingerprint or other reliable ID method may contradict a claimed identity or even the identity documents being carried by the individual at the time of the encounter. Detection of the discrepancy between documents carried and the underlying ID is reflected in several examples in this study. Discovery of such a deception may be performed by fingerprint and other ID methods that have been built into the scenario constructed for this study, but other instances of deception, such as disguises, have not.

3. The Attribute Vector

3.1 Form of the Vector

An attribute vector may be constructed representing available information. Complex information may need to be simplified or condensed to be meaningfully represented in the attribute vector that describes a single individual, and many individuals in the registered population may not share equal amounts of information in any particular category or personal attribute. An example might be information in credit reports: there may be a great deal of detailed information on individual 1 regarding loans or credit cards, as well as the credit agency's rating, while individual 2 may have little or no credit history but be represented by a single credit rating number, to all appearances just like individual 1. Individual 3, from a foreign country, may have

no such information at all, yet be extremely wealthy. Any method for assessment of an individual must accommodate cases where no data are available.

The case of no available data may be represented in the attribute vector as NA. The MDS software tool, PERMAP, uses this same abbreviation.

The impact of the lack of information may vary depending on the situation. In some cases, the lack of information concerning a given characteristic is an important datum; in other cases, it is not. For example, a person detained at a checkpoint may have good identity documents representing a given persona, but there may be no corresponding HUMINT reports filed against that persona. If intelligence coverage of the theater is reasonably comprehensive, the lack of HUMINT reports, though not conclusive, is important and may be represented as zero records or documents available. If the theater is undeveloped, so that there is very little HUMINT, lack of HUMINT data on a given subject may mean little, and the data must be regarded simply as missing data. There is an obvious gradient for the importance of missing or nonexistent data. Incorporation in the MDS methodology of the nature of the importance of missing information may well be a useful research topic.

If a person is known only by a presumably false name in a HUMINT report, with no other information, the lack of identity documents must be represented as no available data, or NA. On the other hand, if a person were stopped at an identity check and found to have no identity documents, the lack of identity documents would be a significant datum and would be entered as zero for the documentation attributes rather than as just the lack of data, NA.

Attributes include information determined by observation of the individual during a physical encounter, such as height, weight, and identity documents, and data in data repositories that can be linked by bio-ID to the individual, such as prior arrests or intelligence information and situational information. Gathering of the ensemble of available data allows systematic construction of personal attribute vectors that permit comparison of a given individual with all previously encountered individuals.

There are two other interesting factors relating to the attribute vector: (1) the weight or importance assigned to an attribute and (2) normal or expected behavior. The weight of a given attribute is probably best allocated by the analyst based on factors such as doctrine, experience, or expert opinion. For instance, is presence of a past fingerprint record or traceable birth certificate providing a positive identification more important than presence of a Ration Card? Probably, but the analyst must decide. The issue with PERMAP is that weights are applied uniformly to all individuals under analysis, when, in fact, the weight or importance of a given attribute may vary individual to individual.

Weight may be used as a surrogate for confidence in a given attribute value—reflecting the uncertain nature of the reliability of real data—but such a makeshift is unlikely to be satisfactory. An explicit analysis of the reliability factor is needed. One problem is the situational aspect of

information about individuals taken from different contexts; the value of an attribute for one person may be very reliable (high confidence that the data reflected are accurate) but not for another. This is particularly important for intelligence data and may prove to be relevant for observational or remotely sensed biometric data as well. Certainly, observational data such as an estimate of subject behavior is more reliable from some observers than from others, especially in cases where cross-cultural differences may, unconsciously, affect the estimate.

The expected presence or absence of data for a specific situation may be important. The importance of any deviation from reasonable expectations must also be a judgment call by the analyst. For instance, if the individual has an identity card or weapons permit in his possession, the absence of a fingerprint record that can be linked to the identity document may be very important. Such civil documentation can only be obtained in the normal course of events by a process that involves fingerprinting. Lack of fingerprints may be due to inefficient or corrupt administrative procedures endemic to some areas of the world or may be indicative of attempted deception. Possible deception would make that attribute more important and imply greater weight. Lacking expert advice, the authors did not attempt a weighting scheme at this time. For this study, all subject attributes carried the same weight; such uniformity is likely not representative of tactical data in general. Any weighting schema may be scenario dependant as well.

We have, then, a set of n subjects, each represented by an m-dimensional attribute vector. If the set of n subject vectors is $\{v_i\}$, where $i = 1, \ldots, n$, and the m elements in the attribute vector are represented by indices $k = 1, \ldots, m$, then the k^{th} attribute of the i^{th} subject is v_{ik}.

3.2 The Person Attribute Vector

The person attribute vector used in this study is based on a database populated with, as of this writing, 52 notional persons representing the background population plus a 53rd person as a test case. It is based on the notional HUMINT message set devised for use in the Soft Target Exploitation and Fusion (STEF) Army Technology Objective. The 19 individuals in the HUMINT message set and its underlying scenario were incorporated in an ontological database. Those persons were extracted, and the related data concerning each persona incorporated in this database.[*]

The STEF scenario was based on HUMINT reports; most of the personas concerned were not actually physically encountered. Thus, a great deal of the kind of observational data, biosensory data, information concerning personal identity documents, and information that could be derived from host civil and military record archives did not exist.

An additional 33 persons besides the 19 STEF individuals were invented plus the test case. The problem of multiple identities was not addressed in this version of the database except for one

[*]Kindly provided by Dr. Kofi Apenyo, Tactical Information Fusion Branch, Information Science Division, Computational and Information Sciences Directorate, U.S. Army Research Laboratory.

terrorist identity constructed so as to be linked to a terrorist from the STEF data set. Many of the issues discussed previously concerning persons/personas/underlying identity that surfaced from the attempt to deal with this particular set of two linked personas; otherwise, in this version of the database, the issue of multiple personas does not arise, so the term used in this discussion is, for the sake of simplicity, person.

Most of the persons added to the STEF persons were the subjects of physical encounters, such as at checkpoints. Encounter histories were invented to define the information available in the database on these persons and the order in which it became available. These encounter histories implied a great many personal characteristics from which a number of attributes defining the persons could be derived. The database presently has 52 individuals plus a test case, for a total of 53 individuals, each with 115 unique personal characteristics, generating 6095 cells (see appendix E).

There is some redundancy since several of the personal characteristic columns in the database are numerical representations of textual data, such as a numerical code for tribal affiliation. That is, the text entry for a given tribal name might be Darraji, but the numerical representation of that tribe, required for analysis, is 5. In this case, two data columns have exactly the same information content expressed differently. The 115 data columns represent 89 unique characteristic columns, yielding 4717 unique data cells plus 26 numerical attribute columns. The set of attribute vectors is derived from the 26 numerical attribute columns.

This database is tailorable according to the needs and conditions of the theater. Each theater may be culturally unique, with different amounts of data available. For instance, in Iraq there was a well-developed police state, with records on and documents for individuals. In Somalia there is next to no records infrastructure since central government is practically nonexistent. In the latter theater, the population may be "ground truthed" by the systematic use of biometric identity devices, but initially the available data may not support similarity analysis to the same extent as in a country such as Iraq. The database framework is still useful as attributes may be parked in the MDS software used in this study, or the attribute vector macro in the Excel representation of the database may be modified to omit empty data attributes. This plasticity and versatility may be useful.

Many of these cells are not populated because the persona or encounter history for an individual did not imply existence of data. For example, if a person is only known as a name overheard by an informant or by an intercept with no other data, most of the cells for that persona must, perforce, be blank. This implies a sparse data matrix.

Many of the notional persons belong to groups in the background population. Presently, the 19 STEF-derived individuals are loosely grouped into the team responsible for a hostage terrorist operation, a small part of the notional Mashhadan terrorist cell, and a variety of others. The key is variety; those individuals do not form a cohesive group in the scenario, and this lack of

"groupness" will be demonstrated later in the analysis. Some additional personal information other than that presented in the STEF-derived ontology was invented, and some was considered implicit in the scenario and applied to the database.

The database contains a number of groups of persons in addition to the STEF-derived groups. The additional 34 persons invented for this study included individuals in several groups: 10 innocent distractors (including innocent neutrals and some friendlies), one innocent test case, five common petty criminals, eight militia members, and 10 non-STEF terrorists. The latter, due to the scenario assumptions (i.e., existence of a great deal of observational, documentary, and other information) did not group with the STEF terrorists.

The database is embodied in Excel. A screen shot illustrating the general organization of the database is shown in figure 7. Characteristics columns that are highlighted serve as the numerical attributes selected for analysis. The key for determining attribute values is shown in the head of each column, in the first two rows. The 53rd individual, labeled "unknown subject" in the third row of the spreadsheet (highlighted), is the test case invented to assess the efficacy of the methodology to determine which group(s) an individual most resembles.

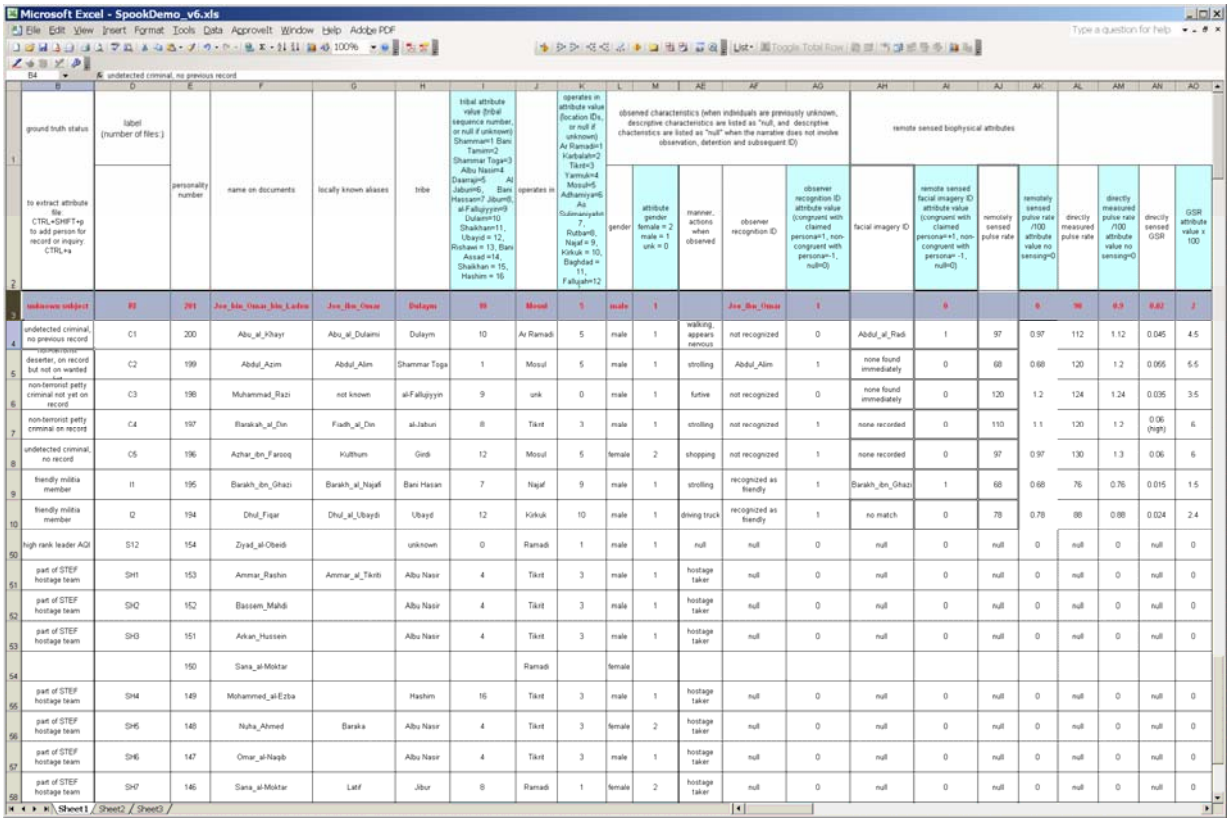


Figure 7. An excerpt of the notional persons database used in this study. The subject of remote inquiry is in the third row, labeled "unknown subject."

The 115 unique characteristics include information visible to an analyst. Other cells not included in these characteristics include ground-truth data that would not be available to an analyst during an encounter, such as information on the encounter scenario for that individual. An example of the encounter scenario might be the following sequence of events: (1) an observer might be cued to the possible importance of the person based on the person's nervous behavior, (2) stop the individual for a document check, (3) use one of the presently fielded biometric ID devices, (4) obtain confirmation of ID, (5) compare that ID to the ID claimed by the person and shown on ID documents on the person, and so on.

A list of the characteristics is provided in appendix E. Some characteristics (e.g., height, pulse rate) are numerical in nature and are suitable for analysis. Many of the characteristics are not numerical, such as stated tribal affiliation and encounter location. Selected non-numerical (categorical) characteristics must be converted into numerical attributes prior to analysis.

Each individual is represented by a vector of 26 attribute values plus an ID label. The attribute vectors for the 53 individuals are shown in appendix F. A discussion of the rules for development of the information in the database and for development of the level of information for each individual—each individual's slice of the scenario—is provided in appendix G.

4. MDS and PERMAP Overviews

A discussion of the analyses performed in this study must touch on both MDS as a technique, widely practiced in different forms over several decades, and as embodied in the software tool chosen for this work, PERMAP (*12*).*

The objective of MDS is to reduce the observed complexity of a database; the perceived set of relationships can be scanned at a glance, making the complex data more accessible to the human mind. MDS techniques can be employed to "inspect" the data and provide a visual representation of the pattern of (dis)similarities among a set of individuals (e.g., subjects of interest). The measurements of (dis)similarity between these pairs of subjects are mapped as distances between points in a low-dimensional space while "matching" the original (dis)similarities as closely as possible. For this analysis, the input space is an m-dimensional attribute vector, and the solution set is of dimension l, with $l < m$. PERMAP v.11.6e used in this study allows attribute vectors with dimensionality up to size 60 and solutions of reduced dimensionality $l = 1, 2, \ldots, 8$.

PERMAP requires the analyst to select from a variety of options for each of the input parameters. Some input parameter options may be limited as they are dependant on the type of

*The equations and definitions in this section are taken from ref (*12*), except where noted.

MDS analysis that is initially selected. Understanding how the analysis is carried out requires some knowledge of how PERMAP (and MDS in general) functions when the input parameters are varied. That is outlined in this section; the impact of those choices is outlined in section 4.2.

PERMAP input can take the form of either a dissimilarity matrix or a set of attribute vectors from which a dissimilarity matrix can be derived, using any one of the attribute-to-dissimilarity functions available in PERMAP. Dissimilarities are based on some measure of association (either nearness or distance) between pairs of entities in a set. Dissimilarity values may be assigned, calculated, or measured. The pairwise dissimilarity between entities *i* and *j* is represented as D_{ij}. For this study, PERMAP was instructed to compute the D_{ij} from the attribute vectors.

PERMAP treats every attribute as the same type of data. This is referred to as "coercing" the data. For instance, if a member of a set of persons is described by a pulse rate (say, 82 beats/min), a tribal affiliation (assigned a descriptive number of 3), and a rank in a hierarchy (for instance, 72 out of 100), the numbers 82, 3, and 72 are different data types (e.g., ratio, nominal, and ordinal, respectively).

Coercing, or converting, the attribute data up or down is done to force the attribute data to be homogeneous, or of a single consistent type. Typically, down conversions are preferred over up conversions. Up conversions can be controversial, albeit interesting, but may also lead to nonsensical results. On the other hand, down conversions generally produce rigorously correct results, even though the results do not take full advantage of the information in the data.

The dissimilarity matrix D_{ij} is used to calculate a set of distances, $d_{ij,}$ between points located in a reduced number of one to eight dimensions. There are several algorithms the analyst may select from for calculating the $d_{ij;}$ the choice of algorithm will depend on the specified relationship between the $d_{ij,}$ D_{ij} and the data types.

There are a number of measures for distance between points representing a pair of entities and, hence, (dis)similarity of pairs of entities, such as Euclidean and City Block.

The distance metric allows definition of a "badness," or stress, function, B_{ij}, which describes the degree to which the set of distances, $d_{ij,}$ "match" the input dissimilarities, D_{ij} :

$$B_{ij} = f(d_{ij} - D_{ij}) \cdot \tag{1}$$

If the d_{ij} in the reduced dimension solution space are the same as the D_{ij} in the higher dimension input space, then B_{ij} is 0.

There are several selectable stress functions in PERMAP; the ones used in this study are Stress and Stress.

This leads to the question of how good the match is, summed over the entire data set. One parameter that may be used is the objective function. An objective function is a measure of the

degree to which a mapping departs from the input. The objective function may also be referred to as an error function or merit function, depending on whether the function is maximized or minimized, respectively, as objects are moved away from a good configuration. Many mathematical formulas are possible but "most objective functions are defined to be the sum over all object pairs of the pair's weight factor times the square of the pair's badness measure B_{ij}" (*12*). This is called the weighted sum-squares form.

A better way to determine the fit of an MDS plot is with the Shepard plot. A good fit is assumed if three conditions are met relative to the Shepard plot:

1. There is clustering about the 45° line if carrying out ratio or interval MDS; otherwise, clustering about the central monotone line if carrying out an ordinal MDS. This clustering is evaluated by R^2, or coefficient of determination, which is the percent of variance that is explained by using the central line as opposed to using the average.
2. There are no outliers.
3. There is no evident pattern in the distribution of points.

PERMAP results are expressed both as a numerical output file of entity locations in the reduced dimension space and as a two-dimensional (2-D) mapping of the d_{ij} onto a plane representing the (dis)similarity of the entities (persons) described by the attribute vectors.

The MDS analysis scheme is summarized in figure 8.

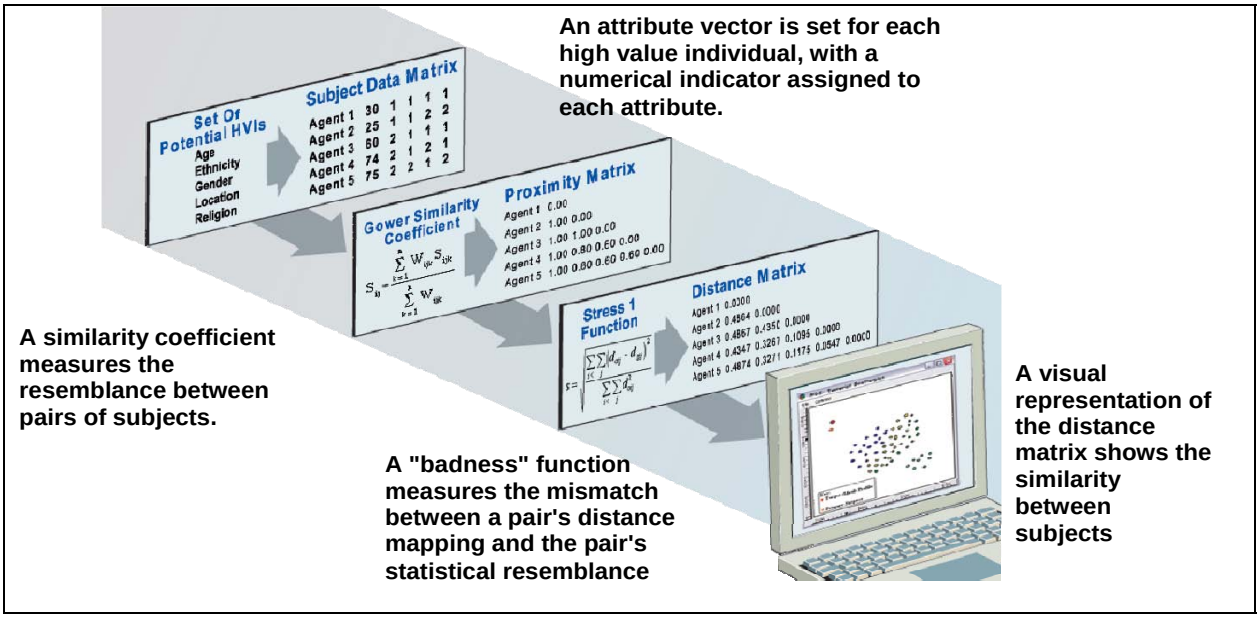


Figure 8. The MDS process.

The MDS output, then, is a visual representation of (dis)similarity between objects or entities described by attribute vectors. If one object is displayed as close to a grouping of objects, it is reasonable to assume the object is a member of the group. In fact, that may not be true. Further, the closeness of data points in the visualization may not indicate the closeness of the relationship, if one exists.

The problem of when to include an entity in a group has usually been approached from the functional description. When entities (individuals) are graphed, they may be judged as part of a group or not based on to whom and how much they communicate with other entities (individuals) (*13*). As of this writing, a quantitative treatment of an estimate of groupness based on similarity seems to be difficult to find. Lacking a numerical index of groupness, the analyst must look at the display and then look at the distance measures, d_{ij}, to judge what individuals the subject of interest most closely resembles and what groups those individuals belong to.

Outlier detection may provide methods for assessing group membership. Details of ordinary anomaly detection (outlier detection) are not given here due to the richness of the field (*14*). Most methods of outlier detection are based on simple statistical analysis of the data in a column of a table or in several closely related columns, geometrical distance calculations between points in an arbitrary *N*-space, and the relation of distances in an *N*-space to assumed probability density distributions.

Arguments based on geometrical considerations are powerful and general, but in current applications are likely to be enormously difficult to use with high-dimension, massive, sparse data sets. Measures such as whether a data point in *N*-space is at a greater or lesser distance from some object such as a centroid of a cloud of data points introduce several requirements for analyst judgment.

For example, consider the problem of assessing group membership by whether the point assessed is within the convex hull describing the array of points in *N*-space. If it is within the convex hull, the point under analysis may be reasonably assessed to be a member of that group of points. If the point is somewhat beyond the hull, the problem of judgment arises: how far is too far for membership?

Consider the problem of assessing group membership of a point by its distance from a reference object such as an arbitrary plane or the group centroid. If the point is compared to the distribution of distances from the reference object, an estimate of whether the point is outside the group may be made, but there are pitfalls. For instance, the shape of the cloud of objects may be such that the point under assessment is clearly isolated from the rest, but its distance from the centroid is still similar to that of other objects in the group. Such situations may be analyzed and detected by visualization, but for high-dimension spaces, that may be difficult.

Another problem may lie in the nature of the society providing the data. The Iraqi society has been a police state, nominally fascist, since the formation of the Ba’ath party in the 1930s.

Conformity has been dictated centrally and enforced brutally. Further conformity may be assumed to be present from the nature of Iraqi society in terms of tribal and clerical control, enforced, at least in the latter case, by capital punishment.

A member of a group is thus likely to be similar in many observable respects to other members of the group. Deviations might be expected to be relatively minor and concentrated along a few dimensions not under rigid social control.

Outlier detection methodology may well provide guides to aid the analyst in assessing membership of an individual in a group. Assessment of outlier analysis methodologies may be well worthwhile in the next phase of this investigation if time and resources permit.

---

5. Asynchronous, Parallel, Unreliable Data

5.1 Asynchronous Data Streams

Information may be expected to flow into the fusion element at different times. This is due to differing response times for different types of data. For instance, the biometric ID kit may have a previous record of an individual, but gathering the correct supporting identity records may involve a time delay. On the other hand, human observations and biosensed data, such as heart rate, are immediately available. Thus, the analyst sees an increasing amount of data and gathers an increasingly detailed set of information as time progresses.

A fusion methodology must deal with incomplete data, and analyses must be carried out iteratively as data arrives. For this reason an MDS tool that tolerates incomplete data must accommodate asynchronous data.

5.2 Parallel Data

Parallel data streams may include different measures of the same phenomenon, contrasting measures of the same phenomenon, or measurements of many phenomena or characteristics that together allow a global assessment or characterization of an individual.

Measures of the same phenomenon allow increased confidence in assessment of that phenomenon. For instance, GSR and pulse rate both indicate stress and so are examples of different estimations of the same characteristic. If the measures are in agreement, higher confidence may result. If different, an assessment of stress may be viewed with less confidence.

5.3 Reliability of a Fused Data Stream

Understanding the information channels to be fused is vital. What the data in a channel represent and how reliable any individual channel happens to be is important to properly forming the data

vector that integrates the whole data stream and estimating the reliability of the fused data. The reliability of some data streams may be so variable, for example, intelligence data, or completely unknown that an estimate of the reliability of the fused data may not be possible. *

Thus, some of the attributes in the persona vector may have associated uncertainty intervals, such as a measurement of pulse rate as 85 ± 5 beats/min, while others may not. Uncertainty intervals may not apply to other attributes, such as tribal affiliation. In any case, uncertainty intervals cannot be propagated algebraically through an analysis carried out with software such as PERMAP. Obtaining uncertainty intervals by analyses with the actual variables set at the respective extrema of the uncertainty intervals for each attribute proliferates beyond the scope of this study.†

5.4 Fusion

The fusion of the information contained in the multiplicity of different attributes should allow an estimate of the importance of an individual. The attributes may reflect different phenomena or different measures of the same phenomena. MDS allows the analysis of the individual in the light of all the data available and permits a best estimate based on incomplete data, refined as more data becomes available.

The reliability of data describing any phenomenon may have considerable variation without compromising the analysis. Use of bounds in the analysis also allows elimination of data that varies beyond preset limits. This may also screen out important indications of a change, so bounds must be approached with caution.

Use of MDS as described in this report, then, should allow an analyst to assess the characteristics of individuals using multiple data sources in parallel, which produces data that are of varying reliability and arriving at different times.

6. Concept Evaluation

6.1 Concept Evaluation and Demonstration

A statistically controlled experiment is necessary to gain a defensible estimate of the utility of the methodology. An estimate of operational or intelligence value of an individual can start with

*Intelligence reports may be rated by a combination of the estimated reliability of the source and by the estimated reliability of the information reported. These are qualitative measures.

†This view is supported by ref (*12*): "In addition to considering changes in attribute information level one should consider each attribute values imprecision and how this imprecision might affect the MDS map. This approach might involve assigning error terms to each attribute value and then using the rules of the theory of error analysis to flow these errors through to the resulting dissimilarities and then to the object positions on the map. This is almost never done. Actually, it is probable that it has never been done. To get useful results one would have to assign probability distributions (using ranges yields unreasonably pessimistic results) for the uncertainties present in each attribute value. Obviously, this approach is very difficult for problems of any significant size and we know of no MDS program that supports this kind of sensitivity analysis."

an estimate of the degree of resemblance of that individual to key groups in a background population. An experiment using a notional personal database will determine the following:

- Whether a methodology based on MDS will allow intelligence officers to determine resemblance of a person under investigation to key groups within the population faster than unassisted parsing of population records (time).
- Whether the assignments are correct more often than when unassisted parsing is used (accuracy).

The concept evaluation will consist of two phases—an exploratory/training phase and an experimental phase. The evaluation will use test participants—(ideally) intelligence officers briefed on the scenario and their individual roles—with each participant using the MDS-HVI methodology and a representation of the population such as an Excel spreadsheet. The test participants would be presented with a training population database and a test population database comprising individuals with varying degrees of resemblance to key population groups. Both databases will be "ground truthed."

Two sets of role scripts, one set for the training/exploratory phase and one set for the experimental phase, and two scenarios will be necessary.

The training/exploratory phase may result in procedural/experimental changes, in which case the exploratory phase will continue until the test participants' training is saturated.

Care must be taken to follow insurgency tactics, technology, and procedures in developing the scenarios and roles.

6.2 Notional Scenario

The STEF database was used as the basis for the notional person database in this study. The STEF scenario was expanded to include encounters with random individuals. These encounters included personal confrontation during document checks and personal confrontation cued by suspicious behavior. In some cases, biometric stress indicators are included due to the easy availability of the technology. Each encounter is governed by a person-specific encounter scenario, which governs the development and extent of the information available to the friendly force.

Since the encounters were assumed to include random detention and document checks, as well as cued detentions and document checks, nonhostile elements were included in the database. These elements included friendly indigenous persons, some innocent distractors, common civil criminals, and additional insurgents, including militia members.

Consider notional checkpoint and surveillance scenarios to illustrate how information fusion might support a requirement for determining whether a person resembles a profile of an HVI.

6.2.1 Surveillance

A surveillance team is deployed to provide additional security to an important event to be held in an arena. The surveillance team is provided with remote sensing devices and the BAT. The sensing devices provide the ability to remotely sample pulse rate as a stress indicator and check facial video against the ID database. The BAT is presumed supplemented by a heart function sensor and a GSR device, other means by which to measure possible stress. Individuals are chosen either at random or by cues (e.g., apparent nervousness, anomalous appearance, or behavior) to the trained observers. The remote sensing devices cue to stress. Other physical information, such as estimated height and weight, is also gathered. An individual, once stopped, is required to produce ID documents and provide GSR and fingerprint data; better facial video and iris patterns are also obtained. The different bio-ID channel data plus descriptive information are used to populate an attribute vector, which is then compared to the attribute vectors associated with the different groups.

In addition to personal attributes, situational information is gathered. The Civil Affairs database is queried by the surveillance team for data on reconstruction efforts in that area and the efforts of the enemy to disrupt them. The local law enforcement and local intelligence databases are queried for information concerning known or suspected enemy agents operating in that area. The Allied intelligence database is queried for information concerning enemy activities in that area, including leadership command links, cellular organization, and logistic elements known or suspected to be active in that area. The tactical database is queried for enemy or allied actions in that area.

6.2.2 Checkpoint

The checkpoint scenario is essentially the same as the surveillance scenario. The checkpoint is assumed to be set up to check unknowns at a location rather than confirm identity against a cooperative register of acceptable persons. The individuals are assumed to be queued and the remote sensor scheme outlined in the surveillance scenario is assumed to be applied to the individuals standing in line.

---

7. Preliminary Results

7.1 Analysis Parameters

PERMAP was used to explore the analysis of the notional persons attribute vectors. Different input options for MDS Type, Distance Metric, Badness Function, solution dimensionality, and Attribute-to-Dissimilarity Function were evaluated.

PERMAP was run iteratively, recording the best solution based on the objective function value for a given set of input parameters. A series of runs were conducted with different input parameter options. The results are presented in table 1.

Table 1. Comparison of results using different sets of input parameters.

MDS Type	Distance Measure	Data Type	Badness Function	Dim.	Attribute Function	Objective Function	Shepard R^2
Ordinal	Euclidean	Unknown	Stress	3	Spearman	0.0281	0.06
Ordinal	Euclidean	Unknown	Stress	3	Euclidean (raw)	0.0083	0.934
Ordinal	Euclidean	Unknown	Stress	2	Euclidean (raw)	0.0208	0.874
Ordinal	Euclidean	Unknown	Stress	4	Euclidean (raw)	0.0141	0.921
Ordinal	Euclidean	Unknown	Stress	3	City Block (norm.)	0.0154	0.826
Ordinal	Euclidean	Unknown	Stress	3	Spearman	0.0159	0.16
Ordinal	Euclidean	Unknown	Stress	3	Guttman	0.0171	0.07
Ordinal	Euclidean	Unknown	Stress	3	Euclidean (norm.)	0.0153	0.809
Ordinal	Euclidean	Unknown	Stress	2	Euclidean (norm.)	0.0387	0.671
Ordinal	Euclidean	Unknown	Stress	3	Nominal SMC	0.0267	0.743
Ratio	Euclidean	Unknown	SStress	2	Euclidean (norm.)	0.1524	0.355
Ratio	Euclidean	Unknown	SStress	3	Euclidean (norm.)	0.0739	0.630
Ratio	Euclidean	Unknown	Stress	2	Euclidean (norm.)	0.0781	0.523
Ratio	Euclidean	Unknown	Stress	3	Euclidean (norm.)	0.0333	0.712

Note: Dim. = dimensionality and norm. = normalized.

A good analysis of the Shepard plot provides a better way to determine the fit of an MDS map than does comparing the objective function value to some arbitrary critical value. The Shepard plot shows the d_{ij} plotted against D_{ij}. For this study, combinations of analysis parameters that yield R^2 values of about 0.7 or better and attribute functions of about 0.2 or lower were good starting points. The overall objective was to achieve a fit, or R^2, of 0.9 or better and an objective function value of >0.1. The input parameters and resultant R^2 and objective function values highlighted in table 1 represent the solution set chosen as the best representation of the MDS-HVI dataset.

To determine whether the three-dimensional (3-D) map was stable, the 3-D restriction was removed and the solution "relaxed" into a higher dimension, in this case, a four-dimensional (4-D) solution. The 4-D solution generated a map similar to the final 3-D solution map, indicating that three dimensions were adequate for describing the MDS-HVI data set. The objective function value was 0.0141 and the R^2 was 0.921, both values very close to those of the 3-D model, indicating that the 3-D model was stable.

Another question was, for reasonable input parameters, what sort of variation could be expected from multiple iterations of the program utilizing the same input parameters? A set of 20 runs was carried out using the final model input parameters of Euclidean Distance Measure, Euclidean (Raw) Attribute-to-Dissimilarity Function, Stress Badness Function, and 3-D solution

set. The values of the Objective Function and R^2 showed no variation between iterations, remaining constant at 0.0083 and 0.934, respectively.

7.2 General Observations

The final mapping for this proof of concept study was produced using the following input parameters: Ordinal MDS Analysis, 3-D solution, Stress Badness Function, Euclidean (Raw) Attribute-to-Dissimilarity Function, and Euclidean Distance Measure. The mapping of the points represented by the attribute vectors is shown in figure 9. This run yielded an objective function of 0.00831 and an R^2 value of 0.934.

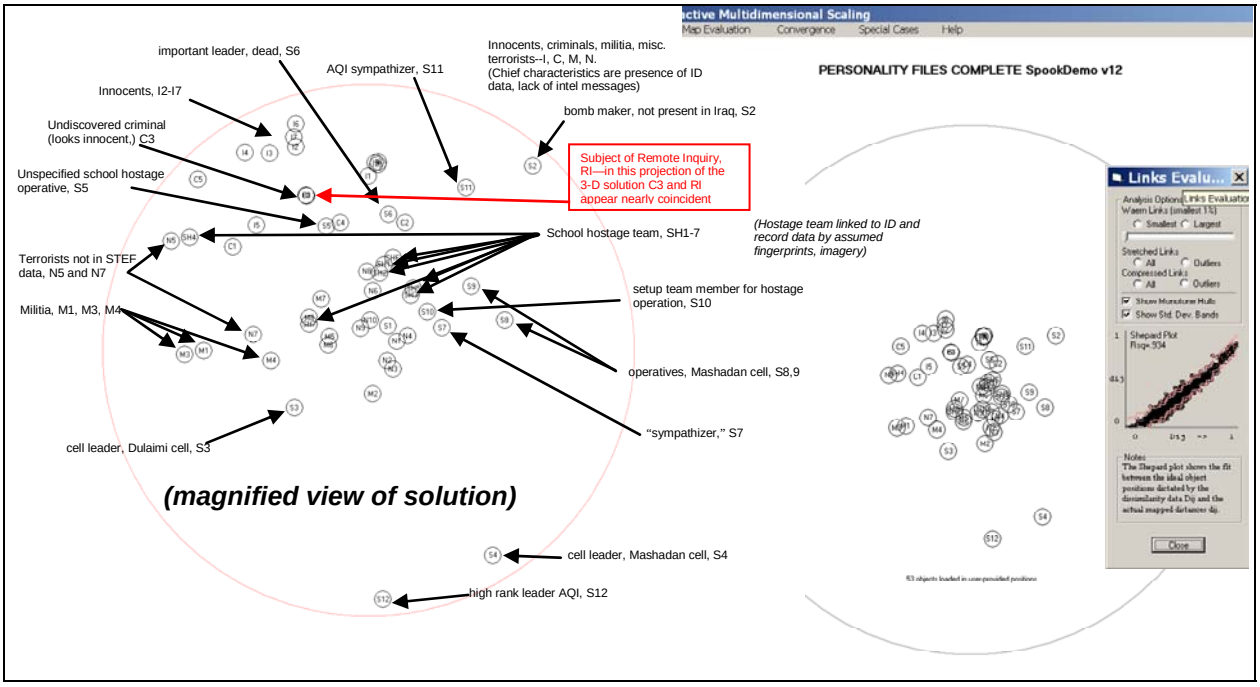


Figure 9. The mapping of the MDS-HVI solution set. The map is magnified for convenience in labeling.

The points representing individuals are labeled. A notional individual representing an innocent person undergoing an identity check is shown as the Subject of Remote Inquiry (RI). Several observations are in order.

The notional innocent, Subject of Remote Inquiry, RI, is grouped with the innocents. The criminals and militia are grouped together, respectively, but separately from the other groups.

The STEF individuals are, for the most part, scattered. This reflects the fact that most of them represent "eaches," or a variety of individuals: organizer, courier, and so on. Identification for these individuals was limited by the scenario. They are whispers on a phone, glimpses by an informer, and so on, but the personas are not bound to an underlying ID. The exception is the group of STEF persons labeled "School Hostage Team."

The School Hostage Team individuals were, in the scenario, members of a team involved in a hostage event. Their files were very similar. In a hostage scenario, the terrorists could reasonably be identified by fingerprints and possibly imagery. This sets them apart from the rest of the STEF personas, who were identified mainly by informer reports, surveillance imagery, and intercepts.

The specific resemblance of the RI to other members of the database is discussed in more detail in the next section.

7.3 Specific Similarities

The immediate neighborhood of the RI yields substantial information. For an analysis yielding good values of the objective function and R^2, inspection of the Waern links indicates similarity of an entity represented in the plot to its sequentially closest neighbors (*15*).

The resemblance of the RI to other individuals may be shown by examination of the Waern links. The smallest 4% of the links,* or greatest resemblances, of the individuals nearest the RI is shown in the expanded view of figure 10b. The persons of greatest resemblance to the RI are an innocent, I5, and an undiscovered criminal, C3, who, since he is undiscovered, closely resembles an innocent in terms of observables. Relaxing the link magnitude to the 5% level adds four more innocents (I1, I8, I9, and I10). At the 6% level of link magnitude (not shown), two more innocents (I6 and I7) are added; at 7% (not shown), a criminal, C4, is added.

The progressive relationships are shown in table 2. The RI thus appears most similar to the group of innocents, which is in accord with the ground truth.

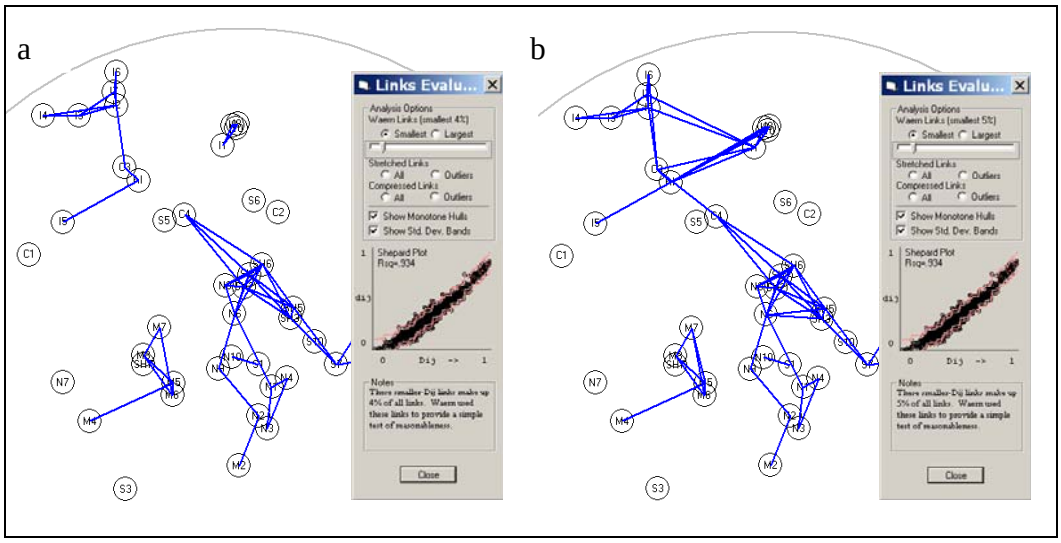


Figure 10. Expanded views showing the smallest 4% and 5% of Waern links.

*There were no links from RI to any other entity, of magnitude less than the least 4% of links.

Table 2. Waern link analysis showing nearest neighbors to the RI.

Smallest Percentage of Waern Links	Number of Neighbors Less Than That Distance	Neighbors
4	2	I5, C3
5	6	Add I1, I8, I9, I10
6	8	Add I6, I7
7	9	Add C4

Clearly, the RI looks more like an innocent than anything else, with some resemblance to the profiles of petty criminals. Again, this class of analysis shows what a subject looks like, not what the subject is.

The list of distances in the 3-D solution set is provided in appendix H.

7.4 Output Visualization

A solution for arbitrary viewing in 3-D space of the MDS-HVI analysis output represented as a graph structure of nodes and links is briefly described. Such dimensionality reduction meant that data was formatted as three-component position vectors, which could then be transformed and projected onto an arbitrary plane before mapping to a 2-D computer monitor. These affine transformations and projection were done using an Extensible 3D (X3D) application programming interface and then viewed in an Xj3D Version 2_M1_DEV_2009-05-18 browser from Yumetech, Inc. This is an open source, standalone browser that uses some 170 primitive X3D objects and includes an unlimited number of prototype definitions for individually defined scene descriptions.

Affine transformations of data are followed by projection onto any plane using the X3D standard application programming interface for viewing. X3D is an International Standards Organization specification for an extensible markup language description of scene content across the Web that supports layering. In this case, text is displayed in a layer of an Xj3D viewer both statically and dynamically: (1) a legend for quick identification within the scene and (2) a console that displays a resultant X3D event chain for text animation when a network node in the scene is touched. A directed acyclic graph of X3D objects defines the scene. Both Java and ECMAScript bindings are used to access and manipulate the scene graph at run time.

An X3D computer program was written in-house to support the visualization of MDS output. The program has been labeled d^2NetVis—a dynamic generation of nodes and links defining a dynamic network structure—and is written for the immersive profile of X3D to include user interactivity (e.g., navigation) within a scene. Affine transformation of data is followed by projection onto an arbitrary plane, which is sometimes referred to as a 2 1/2-D view.

Figure 11 shows d^{2}NetVis applied to the MDS-HVI output. This example includes 53 nodes, as summarized in the legend layer at the left of the window: the subject of remote inquiry (RI), 5 criminals (C), 10 innocents (I), 8 militia (M), 10 non-STEF (NS), 12 STEF (S), and 7 STEF hostages (SH). A tooltip capability allows one to pass the mouse pointer over any network node for quick identification of any network node. When a node is touched (i.e., a mouse button is pressed when over the node), a more detailed text message can be displayed in the console layer. This capability is, however, not shown in figure 11.

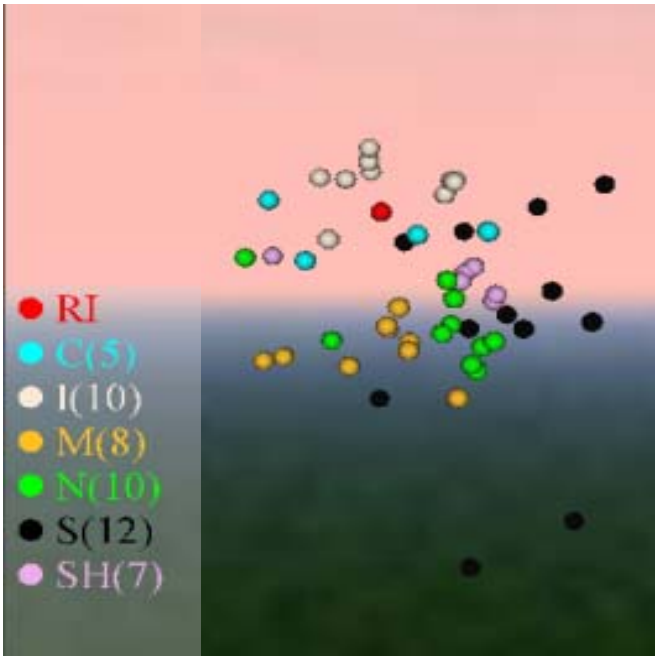


Figure 11. d^{2}NetVis of the MDS-HVI solution set.

Note that all spheres in an X3D figure have the same radius, and any apparent size difference is due to the distance from the viewer before projection. The actual position of a sphere is an attribute of the data and graphic components.

8. Summary

The opportunity exists for the creation of an information support tool to facilitate and enhance the process of grouping subjects of interest into categories based on observables, HUMINT data where available, and previous information on the social and intelligence milieu. This scheme uses existing biometric tools and procedures and complements functional link analysis based on HUMINT information. The present approach employs MDS.

A proof of concept based on a notional personal information database, including reasonable and plausible data, indicates that the methodology will group individuals in accordance with ground truth based on a notional scenario, including assumed and plausible observational, documentary, archival, and intelligence information.

Real personal data will be sought, but such data may not be made available for research purposes. Assuming resources are available, the analysis method will be evaluated by an experimental program.

9. References

1. Brook, D., LCDR. Department of Defense Bloggers Roundtable with Lieutenant Colonel John W. Velliquette JR., USA, Iraqi Biometrics Manager, Coalition Police Assistance Training Team Mission, via Teleconference from Iraq Topic: The Role of Biometrics in the Counterinsurgency Moderator: Lieutenant Commander Brook Dewalt, USN, Office of the Secretary of Defense for Public Affairs Time: 9:01 A.M. EDT Date: Wednesday, August 15, 2007, [capital letters edited]; also at http://www.defenselink.mil/home/blog/docs /20070815BloggersRoundtable%20wLTC%20Velliquette_transcript.pdf (accessed 25 January 2008).

2. Skillings, J. Biometrics and Security in Iraq, dated 17 August 2007 11:15 AM PDT; also at http://www.news.com/8301-10784_3-9761654-7.html?tag=more (accessed 25 January 2007).

3. Wayman, J.; Jain, A.; Maltoni, D.; Maio D., Eds. *Biometric Systems, Technology, Design, and Performance Evaluation*; Springer: New York, 2005.

4. Inbau, F. E.; Reid, J. E.; Buckley, J. P.; Jayne, B. C. *Criminal Interrogation and Confessions*; 4th ed.; Aspen, 2001.

5. Zhu, Y.; Tan, T.; Wang, Y. Biometric Personal Identification Based on Iris Pattern. *ICPR2000: the 15th International Conference on Pattern Recognition*, pp 801–804.

6. Daugman, G. High Confidence Visual Recognition of Persons by a Test of Statistical Independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence Archive* **November 1993**, *15* (11), 1148–1161.

7. Hall, B., CPL. New Face of Recruitment. *Marine Corps News*; 19 November 2007; also at http://www.usmc.mil/marinelink/mcn2000.nsf/main5/E62EE05210FBA4D4852573980037DA0?opendocument (accessed 1 February 2008).

8. Isaacson, A.; Vemury, A. Biometric Automated Toolset (BAT) and Handheld Interagency Identity Detection Equipment (HIIDE), Biometrics Task Force, 19 September 2007; also at http://fingerprint.nist.gov/standard/archived_workshops/xmlandmobileid/Presentations /Vermury-BAT-HIIDE.pdf (accessed 1 February 2008).

9. UK Home Office, Border and Immigration Agency. Iraq Country Policy Bulletin February 2007, issued 17 December 2007; also at http://www.ind.homeoffice.gov.uk/sitecontent /documents/policyandlaw/countryspecificpolicybulletins/Iraq_Country_Policy_Bulleti1.pdf?view=Binary (accessed 22 January 2008).

10. FM 2-22.3. *Human Intelligence Collector Operations*, HQDA, Washington, DC, September 2006.

11. Mythbusters, Episode 59: Crimes and Myth-Demeanors 2; 23 August 2006; also at http://dsc.discovery.com/fansites/mythbusters/episode/episode-06.html (accessed 13 December 2007).

12. Heady, R. B.; Lucas J. L. *PERMAP 11.6 Operation Manual*; 23 March 2007; also at http://www.ucs.louisiana.edu/~rbh8900/PermapManual.pdf (accessed 22 April 2008).

13. Freeman, L. C. Visualizing Social Groups. *American Statistical Association 1999, Proceedings of the Section on Statistical Graphics*, 2000, pp 47–54; also at http://moreno.ss.uci.edu/80.pdf (accessed 28 March 2008).

14. Hodge, V.; Austin, J. A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review* **October 2004**, *22* (2), pp 85–126.

15. Waern, Y. Structure in Similarity Matrices, A Graphic Approach. *The Scandinavian Journal of Psychology* **1972**, *13* (1), pp 5–16.

Appendix A. Data Map

This study assumes linkage of a great deal of information, ranging from biometric identity data to archival host country data. The relationships and information flows change with shifts in tactical requirements and even national diplomacy. The basic ingredients for the fusion described in this study are summarized next.

Biometrics is widespread. There are pieces of biometric data in many databases that have been obtained to address specific needs, nearly always related to personal identity. Identity confirmation and forensic identity determination rely on similar databases, although their formats may vary, depending on the need and use for the data.

Data fusion applied to tactical problems must rely on the intelligence backbone, embodied in the Distributed Common Ground Station–Army. The biometric information itself is the key to identity, to which other information is associated. A person may have several identities, ranging from alternate naming conventions to aliases, perhaps including mistaken identities. The method for creating the associations may be a relational database. Import of the data for the tables to be associated depends on the data, its location, and its format.

In the case of the data to be fused for support of an operator in the field, consider the plethora of biometric data gathered and held by entities operating in the field and in support of those operational entities. This data may include information from biometric identification kits or local civil records. The U.S.-gathered biometric information will conform to the Electronic Biometric Transmission Specification* host or allied country records may not. An additional electronic specification is the National Information Exchange Model (NIEM), which has an xml schema for intelligence records.[2]

The basic information flow for an analyst attempting to fuse information from several sources is similar to that shown in figure A-1. Control or access to host country records and means of transmission is indicated as unknown.

*The Integrated Automated Fingerprint Identification System (IAFIS). http://www.fbi.gov/hq/cjisd/iafis.htm (accessed 2 September 2010).

[2]NIEM 2.1. http://www.niem.gov/niem/NIEM-2.1-schema-index.html (accessed 7 September 2010).

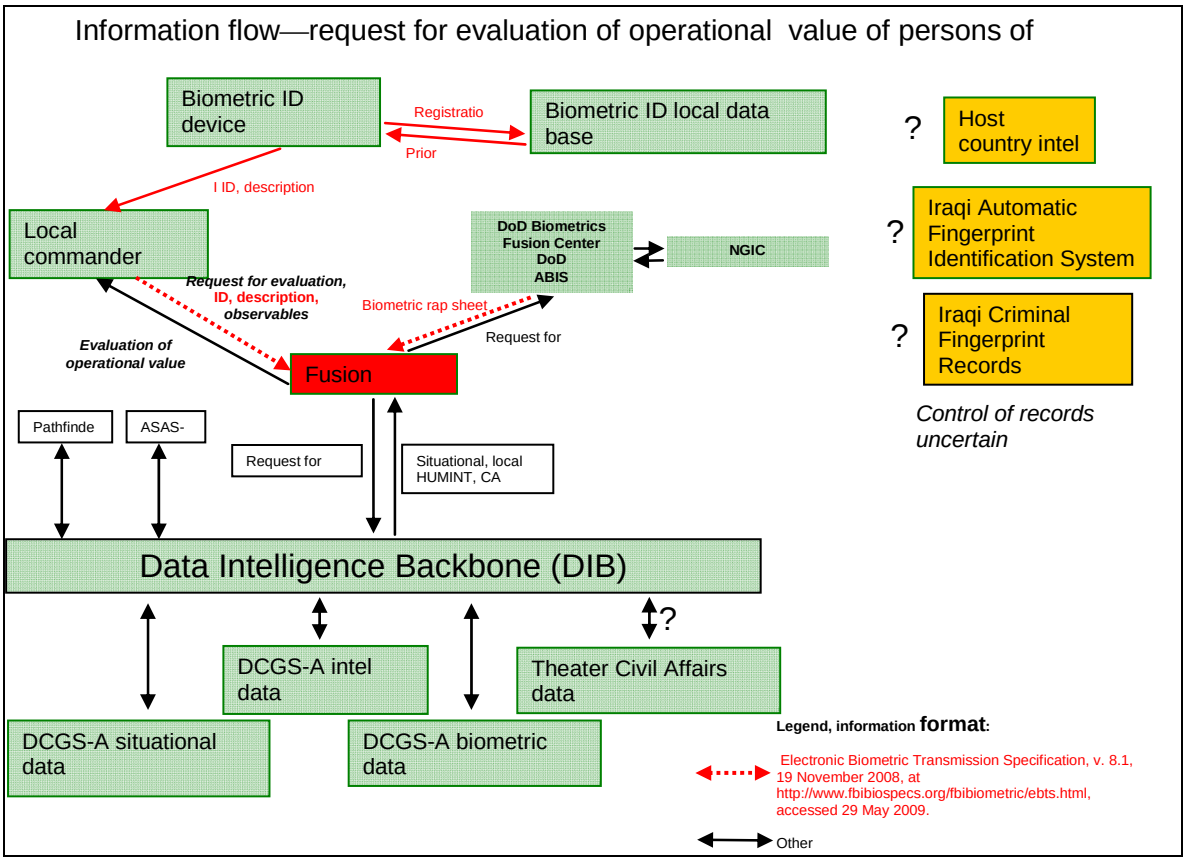


Figure A-1. Basic information flow for fusion of biometric, situational, and intelligence data.

The key players are the field elements using the biometric identification (ID) devices and the local biometric databases. Another key player is the Department of Defense Biometrics Fusion Center operated by the Biometric Task Force. Information from originators of biometric information is centralized and linked based on the biometric identity. The biometric identity is based on the biometric signature. A person may have one or more names associated with the identity. The biometric identity is characterized by a biometric file primary key, which may be associated with several other identity or identity inquiry records identified by other keys, such as the Transmission Control Number (TCN). Linkage of the available data produces a consolidated list of all information relating to an individual. The list is referred to as a "rap sheet." A sample of an xml biometric rap sheet is shown in figure A-2.

Typically, biometric registration in the field is done by fingerprint, facial imagery, and iris pattern. The record is then identified by both the registration name and a biometric record number. In this way, additional records comprising any biometric identity record (e.g., fingerprint) linked to a name or a record inquiry number with no associated name can be associated with the original record, which is identified by its record number.

```
<?xml version="1.0" encoding="UTF-8" ?>
- <Result repository="IAFIS" type="1" xmlns="http://www.bah.com/biometrics/match/2.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.bah.com/biometrics/match/2.0 file:///C:/projects/aims/aims-xml/match-xml/src/jaxb/matchML.xsd">
  <DateTime>2008-08-29T10:08:41</DateTime>
  <Remark>Biometrics Match For TCN: DD001I001-20080829120258-BATS-400V-12457 Biometric Matched TCN 623367AD4 Biometric Matched
TCN DD001I00120080829120258BATS400V12457</Remark>
- <Sensors>
- <Sensor>
  <Eftid id="0.1" description="BFC EFT File Primary Key">1002761597</Eftid>
- <FileName id="0.2" description="Full File Path and Name"> <![CDATA[ 623367AD4  ]]>
  </FileName>
  <DateProcessed id="0.3" description="Record Creation Date in EFTDB">2008-08-29T00:00:00</DateProcessed>
  <Source id="0.4" description="File Data Source">FBI</Source>
  <FileCorrupted id="0.5" description="File Corrupted Flag (0 - Good 1 - Bad)">0</FileCorrupted>
- <File id="0.7" description="Filename"> <![CDATA[ (getFileName())  ]]>
  </File>
  <DateReceived id="0.8" description="Date Received by BFC">2008-08-29T10:01:00</DateReceived>
  <TOT id="1.4" description="Type of Transaction">-1</TOT>
  <DAT id="1.5" description="Date">2008-08-29T00:00:00</DAT>
- <TCN id="1.9" description="Transaction Control Number"> <![CDATA[ 623367AD4  ]]>
  </TCN>
  <IDC id="2.2" description="Image Designation Character">0</IDC>
  <RET id="2.5" description="Retention Code">Y</RET>
  <FBI id="2.14" description="FBI Number">623367AD4</FBI>
  <NAM id="2.18" description="Name">UMAR, AL-TIKRITI IBN NASIB</NAM>
  <POB id="2.20" description="Place of Birth">XX</POB>
  <CTZ id="2.21" description="Country of Citizenship">XX</CTZ>
  <RAP id="2.70" description="Request for Electronic Rapsheet">Y</RAP>
```

Figure A-2. Excerpt from a notional biometric rap sheet.

As seen in figure A-2, the rap sheet is listed as a match for other records identified by TCNs (<Remark>, lines 5–6).

The biometric data may then be linked via association table with combat incidents associated with locations (towns), military grid references (map coordinates), units or hostile entities involved (organizations where known or suspected individuals were seen, killed, or captured), modus operandi (tactical pattern), forensic information, surveillance imagery, or intelligence reports.

INTENTIONALLY LEFT BLANK.

Appendix B. Human Observation

The observer has the ability to assess the behavior of the subject either remotely or in person during an interview. In principle, the mechanical biosensory mechanisms of this proposed method are extensions of the interrogator's sensorium into the biosensory field: subject pulse and blood pressure, electrical assessment of electrodermal skin response (also known as galvanic skin response), and changing electrical fields outside the skull, representing neural activity within the brain. In terms of human observation, breathing and sweat are visual indicators of physiological indices. Inferences may be drawn based on observation of pulse in the great arteries of the neck and the blood flow in the face—blushing, pallor, and visible pulse in superficial vessels.

Behavioral cues may be used to assess the state of subjects according to the principles of the Reid Technique.[1, 2] The Reid Technique codifies empirical rules for observations of the following:

- Verbal channel
- Paralinguistic channel
- Nonverbal channel

B.1 Verbal Channel

This includes word choice and arrangement of words used in responses. There is an extensive array of stress or deception indications from use of unnecessary qualifiers, extra detail, nonuniform levels of detail, inclusion of emotional qualifiers, and the like.

B.2 Paralinguistic Channel

Paralinguistics deals with the characteristics of speech "falling outside the spoken word," including rhythm, pauses, inflections, tone, and other similar properties.

B.3 Nonverbal Channel

This includes posture; individuals may cross their legs, lean forward, look away, look down, fiddle with an ear, and so on.

[1] Inbau, F. E.; Reid, J. E.; Buckley, J. P.; Jayne, B. C. *Criminal Interrogation and Confessions*; 4th ed.; Aspen, 2001.

[2] Inbau, F. E. *Essentials of the Reid Technique: Criminal Interrogation and Confessions*; Jones and Bartlett, Inc., 2005.

A test using expert observers and videotaped interviews of criminal suspects was conducted by the originators of the Reid Technique. Test data indicated accurate assessments of truthfulness 86% of the time and deceptiveness 83% of the time. Inconclusive assessments were excluded. Tests conducted with paid subjects who had little to gain or lose indicated accuracy no better or a little better than chance. This is not surprising; if the subject doesn't care about being caught in a lie, signs of deception will probably not be evident.

Appendix C. Biometric ID

Physical measurements, such as physiognomy and fingerprints, will be taken to attempt to identify the subject (via biometric identification [ID] or bio-ID) in order to link the subject to the situational information. If the subject is unknown to the interrogator or the bio-ID data does not correspond to a previous record or registration, the link to the situational data must be geographic and situational rather than personal: where was the subject at the time (s)he was encountered, and what was (s)he doing? Bio-ID data can include fingerprint and facial video data. Other more exotic identification technologies may be used but will not be described here.

Presently there are two branches of bio-ID systems:

> A biometric [ID] system can be designed to test one of only two possible hypotheses: (1) that the submitted samples are from an individual known to the system; (2) that the submitted samples are from an individual not known to the system. Applications to test the first hypothesis are called "positive identification" systems (verifying a positive claim of enrollment), while applications testing the latter are "negative identification" systems (verifying a claim of no enrollment).*

Confirmation of membership in a list of registered individuals for access to either a site or to information is extremely important and is also fundamentally different from identification of an individual from a general population. In the one case, the individual is previously registered in a small and well-defined database, and the interest is in successful, timely, and accurate identification. In a random encounter, such as at a checkpoint, the reverse may be true.

The possibility of real-time or near real-time determination of deception leads to an interesting possibility: verification of self-identification. This is summarized by the question "Are you whom you say you are?" The question may lend credibility to an individual's declaration of identity. The question might also be accompanied by some variation of "What are you doing here?" If this possibility can be realized, it will constitute a third branch of bio-ID.

A number of commercial biometric ID systems have been developed. These use physiological measurements such as iris patterns, fingerprints, and facial measurements. Commercial ID systems are developed in support of identity verification rather than identity determination. That is, the subject is compared to an existing register of identifying characteristics to verify that the individual is in that register. The commercial system may use a different set of characteristics than those used in large government identification databases. Unless a commercial ID system is specifically developed to interface with the local government identification records, the system

*Wayman, J.; Jain, A.; Maltoni, D.; Maio D., Eds. *Biometric Systems, Technology, Design, and Performance Evaluation*; Springer: New York, 2005.

may not permit easy access to or use of existing identification databases. In general, the issues will be physical access to the information in the identification database, permission to use the data, and possibly translation of any government identification data format to a format that will allow the commercial system to search for and compare files with the government identification records.

Fingerprint matching would be the workhorse biometric ID technique. There are several fingerprint readers on the market. If the criminal or intelligence fingerprint archive were digitized, candidate matches in near real-time would be possible, with the final match done by a specialist with appropriate software, a role very well suited to an expert at the base element.

Other allied nations may not have digital fingerprint records, so the automatic extraction of features for possible matches with their records may not be possible. The U.S. Federal Bureau of Investigation maintains the Integrated Automatic Fingerprint Identification System (IAFIS), which will select several candidate matching records for submission to qualified requestors for on-site investigation.[2] The candidate prints from the automatic system may then be examined by a qualified expert with the proper software.

Facial video may be matched with surveillance imagery, "mug shots," or imagery from other sources. The expert consensus is that automatic facial ID software is not very accurate, perhaps 70%, but may serve as a selection tool for records that may be further utilized by experts. A cursory examination of the literature indicates that facial recognition is done basically by abstracting key feature measurements such as pupillary separation, height of ears with respect to the brow, length of nose, distance from base of nose to lip, and so on. Images can also be analyzed by wavelets and similar techniques.[3, 4] The probability of correct visual ID has been estimated as 70%.[3]

Automatic identification using imagery from files does not have a very good record, about 50%–75%, according to one expert.[5] The metrics illustrated in papers on recognition algorithms are reminiscent of the Bertillon system, relying heavily on facial and feature proportion and arrangement.[6] As such, agreement of personal physiological metrics can be indicative, and useful within its limits, but probably not definitive. Nevertheless, surveillance imagery may, in many cases, be the only ID record available for many subjects of interest.

---

[2]The Integrated Automated Fingerprint Identification System (IAFIS). http://www.fbi.gov/hq/cjisd/iafis.htm (accessed 2 September 2010).

[3]Neti, C. V.; Senior, A. Audio-Visual Speaker Recognition for Video Broadcast News. http://www.clsp.jhu.edu/ws2000/groups/av_speech/papers/hub4_avspeaker.pdf (accessed 13 February 2007).

[4]Neti, C. V. *Visual Feature Extraction, Audio-Visual Speech Recognition Workshop 2000 Final Report*; Technical Report WS00AVSR; Johns Hopkins University, CLSP, 12 October 2000; also at http://www.cs.cmu.edu/~iainm/papers/ws00avsr.pdf (accessed 13 February 2007).

[5] Schneider, J. K. Ultrascan Corporation. Private communication, 10 August 2006.

[6]The Bertillion System. http://criminaljustice.state.ny.us/ojis/history/bert_sys.htm (accessed 20 October 2006).

Voice analysis, or “voiceprinting,” may be possible with considerable accuracy for cooperative identity confirmation, with comparison of a specific phrase with a prerecorded registration sequence. It is not clear that an automatic identification system for noncooperative use is readily available, nor that such a process can be done in real-time. Matching whatever voice data that could be enticed from a subject with an intercept might be very difficult, perhaps impossible, for applications such as interrogation support in the field.

A real problem with using bio-ID in a foreign country is the lack of a database of enrolled subjects’ biometric characteristics. Local police may have “mug shots” and fingerprint files, but ID files on a significant number of individuals may be lacking, or secret police files may be unavailable to friendly forces. The possibility of confirmation of self-identification may reduce dependence on prior registration.

Other bio-ID modalities may be adopted as they become available.

INTENTIONALLY LEFT BLANK.

Appendix D. Iraqi Identification Documents

This section is an excerpt from Iraq Country Policy Bulletin 2/2007 Issued 17 December 2007, UK Home Office, Border and Immigration Agency, section 3.*

D.1 Documents

Documents which are widely available are the Iraqi Nationality Certificate as well as the Iraqi Civil Status Identification (ID), both of which are issued by the Directorate of Travel and Nationality/Ministry of Interior. In the Kurdistan Regional Government (KRG) area, these documents are issued by the Directorate of Nationality and Civil Status/Ministry of Interior (Governorate of Sulaymaniyah) and the Directorate of Nationality and Civil Identification (Governorates of Erbil and Dohuk). These documents are obtained by applying in person as there is no reliable postal service. These documents are the main identification documents within Iraq and are requested for any kind of interaction with the authorities, such as an application for a Food Ration Card, school registration, and the issuance of death and birth certificates. Although the Iraqi Civil Status ID card is more commonly used, both can be used for general purposes and at roadblocks. Iraqis will also often present an ID from their place of work. Iraqi Civil Status ID or Iraqi Citizen Papers are needed to cross from KRG into Government-Controlled Iraq.

D.1.1 Residence Address Card

Another document used at times is the Residence Address Card, which certifies the holder's address and is needed to buy real estate, a car or mobile phone, or to submit a job application. Instead of the Residence Address Card, one can also obtain a one-time document certifying a person's residence from the local mayor (Mukhtar). In the KRG area, only one-time documents certifying a person's residence are available.

D.1.2 Food Ration Card

The Food Ration Card, which allows its holder to obtain the monthly food ration through the Public Distribution System (PDS), is issued by the Ministry of Trade and is also widely accepted as an identification document. In the KRG area, the Food Ration Card is issued by the Directorate of Food/Ministry of Trade (Governorate of Sulaymaniyah) and the General Company for the Trade of Food Items/Ministry of Finance and Economy (Governorates of Erbil and Dohuk).

*UK Home Office, Border and Immigration Agency. Iraq Country Policy Bulletin February 2007, issued 17 December 2007; also at http://www.ind.homeoffice.gov.uk/sitecontent /documents/policyandlaw/countryspecificpolicybulletins /Iraq_Country_Policy_Bulleti1.pdf?view=Binary (accessed 22 January 2008).

D.1.3 Birth Certificates

Birth certificates are usually obtained in public hospitals or health centres. A copy of the birth certificate has to be sent to the PDS center to include the newborn on the family's Food Ration Card. Death certificates are issued by public hospitals, indicating the time, date, and reasons of the death. Deaths occurring outside a hospital need to be approved by the Civil Status Court that issues a certificate proving the death. A copy of the death certificate is to be sent to the PDS center to exclude the deceased from the family's Food Ration Card. In the KRG area, birth/death certificates need to be sent to the Directorate of Food (Governorate of Sulaymaniyah) and the General Company for the Trade of Food (Governorates of Erbil and Dohuk) for (de)registration of a person in the PDS.

In order to relocate within Iraq, an individual must be in possession of all the following documents:

- Personal identification number which is issued by the General Directorate of Citizenship in accordance with Iraqi civil law number 65 (1972).
- The Iraqi Nationality Certificate which shows that the holder is Iraqi.
- A letter of confirmation from the place of work in the intended relocation town and/or the approval of the Mukhtar of the intended relocation town.
- A declaration from the security services that the person is not involved in criminal activities.

Without all this documentation supporting official relocation, individuals would be unable to access rationed food and would be denied access to work. A person wishing to remain in, or move to, an area of Iraq other than his hometown does not have to visit his hometown to obtain the requisite documentation for a relocation application.

The Ministry of Migration and Displacement branch offices issue documents to internally displaced persons (IDPs) that certify they are IDPs. These documents are used by the Ministry of Trade to honor the ration cards that IDPs have with them (from their place of origin) at the place of displacement. Ration cards are used for basic provisions, such as rice, beans, cooking oil, detergent, tea, salt, flour, petrol, and other fuel products. They can only be used in a specific shop or store in the person's neighborhood. But local Ministry of Trade offices can exchange the ration cards if persons move to another district. The majority of people comply with the arrangements for relocation and rationing. The ration card is a very important form of ID and can be checked relatively quickly by the Iraqi authorities.

Appendix E. List of Characteristics in the Notional Persons Database

The characteristics in the database include the data elements in the following list. The list summarizes the columns in the Excel spreadsheet, some of which are characteristics of the notional individual and some of which express ground truth. Entries marked with an asterisk (*) are included in the attribute vector. This list applies to the file SpookDemo_v12.xls of the notional persons database used in this study.

Note that entries corresponding to columns 1, 2, and 3 in the database are scenario related and describe ground truth and are not counted as personal characteristics. Columns 4 and 5 are a label and a tracking number, respectively. This leaves the 111 personal characteristics noted in the body of the report. The scenario data, such as the information discovery sequence (column 3) for the encounter helps select the detailed information to be included in the data that is presumed available to investigators and corresponds to the ground truth characteristics of the underlying identity.

Columns:

1. Group (innocents, STEF persona, criminal, etc.)
2. Ground truth status concerning the persona (scenario information)
3. Information discovery sequence
4. *Label
5. Persona number
6. Claimed name
7. Alias/naming variants in encounter (obtained from neighbors, police, etc.)
8. Tribe
9. *Tribal attribute value
10. Home/ residence location
11. *Home location attribute value
12. Gender
13. *Gender attribute value
14. Recorded height
15. Observed height agreement with file description
16. Recorded weight
17. Observed weight agreement with file description
18. Recorded build
19. Observed build agreement with file description
20. Recorded eye color
21. Observed eye color agreement with file description
22. Recorded complexion
23. Observed complexion agreement with file description
24. Recorded facial hair
25. Observed facial hair agreement with file description

26. Recorded color hair
27. Observed hair color agreement with file description
28. Observed clothing
29. Armed
30. Location observed/encountered
31. Manner, actions when observed
32. Observer recognition ID
33. *Observer recognition ID attribute value
34. Facial imagery ID
35. *Remote sensed facial imagery ID attribute value
36. Remotely sensed pulse rate
37. * Remotely sensed pulse rate /100 attribute value
38. Directly measured pulse rate
39. * Directly measured pulse rate /100 attribute value
40. Directly sensed Galvanic Skin Response (GSR)
41. * Direct GSR attribute value x 100
42. Iris pattern ID
43. *Iris pattern ID attribute value
44. Location and date iris pattern recorded
45. Facial imagery ID
46. * Facial imagery ID attribute value
47. Location and date facial imagery recorded
48. ID according to fingerprint
49. * ID according to fingerprint attribute value
50. Location and date reference earliest fingerprint registered
51. Iraqi Civil Status ID
52. Iraqi Nationality Certificate
53. Residence Address Card
54. Food Ration Card
55. Weapon permits
56. * Number of ID docs congruent with claimed name given to observer during processing attribute value
57. *Number ID docs congruent with true identity attribute value
58. *Number ID docs not congruent with true identity attribute value
59. Other civil, Iraqi military and police, U.S. BAT docs, known IDs on file, dates
60. *Number of official civil records for relating to true identity attribute value
61. Recorded events other than ID applications and birth records, includes passports
62. Number of recorded events in records other than ID applications attribute value (not used)
63. Recorded nationality
64. *Recorded nationality attribute value
65. Birthplace
66. Recorded age
67. DOB (Date of Birth)
68. Ethnicity
69. *Recorded ethnicity attribute value
70. Recorded sect

71. *Recorded sect attribute value
72. Human_log
73. Languages
74. Profile
75. Skills
76. *Recorded skills attribute value
77. Status
78. CommunicatedWith
79. *Number enemies communicatedWith attribute value
80. ContactWith
81. *Number known enemies contactWith attribute value
82. HasCoconspirator
83. HasLinksTo
84. HasSource
85. Heads (hostile organization)
86. *Head of hostile organization attribute value
87. IsAffiliateOf
88. IsMemberOf
89. *IsMemberOf attribute value
90. IsParticipantIn
91. IsRelativeOf
92. IsSourceFor
93. IsSupervised by
94. *IsSupervisedBy attribute value
95. IsSuspectedBy
96. IsSuspiciousOf
97. Made
98. MaybeMemberOf
99. OperatesIn
100. Owns
101. Perpetrates
102. Reports
103. SameAffiliation
104. SameNationality
105. Seeks
106. Supervises
107. *Supervises number of hostile individuals attribute value
108. Supplies
109. Targets
110. UseSameFacility
111. Local locations
112. Dates member of Iraqi armed forces
113. Assessment number of recorded operations
114. *Number of entries in HUMINT message file attribute value
74. Profile (duplicate of item 74, for convenience)
115. Notes (HUMINT messages)

INTENTIONALLY LEFT BLANK.

Appendix F. Attribute Vectors

The attribute vector of the subject of remote inquiry, RI, is the first item. Attributes are listed in the comment lines above the ATTRIBUTE LIST.

personality_files_v12.txt - Notepad

File Edit Format View Help

TITLE PERSONALITY FILES COMPLETE SpookDemo v12
NOBJECTS=53
NATTRIBUTES=26

Clabel	tribal	home	gender	observer ID	remote facial ID	remote pulse	direct pulse	direct GSR	iris ID	direct facial ID	direct fingerprint	# docs ~claimed ID	# docs ~bio-ID	#docs not ~ claimed ID	# civil records	nation ality	ethn	sect	skills	comms with	contact with	headof	member of	superv by	# super-vises	# HUMINT msgs
ATTRIBUTELIST																										
RI	10	5	1	3	1	0.7	0.9	2	1	3	3	3	3	0	3	1	1	1	1	0	0	7	1	1	0	0
C1	10	1	1	1	2	0.97	1.12	4.5	1	2	2	2	0	2	4	1	1	1	9	0	0	7	1	1	0	0
C2	1	5	1	3	1	0.68	1.2	5.5	1	2	2	3	0	3	5	1	1	4	1	0	0	7	1	1	0	0
C3	9	NA	1	1	1	1.2	1.24	3.5	1	2	2	3	0	3	4	1	1	1	1	0	0	7	1	1	0	0
C4	6	3	1	1	1	1.1	1.2	6	1	1	3	2	2	0	2	1	1	1	1	0	0	7	1	1	0	0
C5	15	5	2	1	1	0.97	1.3	6	1	1	2	3	0	3	5	1	2	3	1	0	0	7	1	1	0	0
I1	7	9	1	3	3	0.68	0.76	1.5	1	3	3	2	2	0	4	1	1	1	3	0	0	7	2	1	0	0
I2	12	10	1	3	1	0.78	0.88	2.4	1	1	3	2	2	0	4	1	2	1	5	0	0	7	2	1	0	0
I3	13	8	1	3	3	0.71	NA	NA	NA	NA	3	2	2	0	5	1	1	1	7	0	0	7	1	1	0	0
I4	13	8	1	2	1	0.83	0.85	2.5	1	1	3	2	2	0	5	1	1	3	9	0	0	7	1	1	0	0
I5	10	1	1	2	1	0.73	0.78	2	3	3	1	4	4	0	5	1	1	1	3	0	0	7	1	1	0	0
I6	12	10	2	1	1	0.72	0.8	2.1	3	3	3	2	2	0	3	1	2	1	1	0	0	7	1	1	0	0
I7	12	10	1	1	1	0.74	0.87	2.5	1	3	3	2	2	0	4	1	1	1	3	0	0	7	1	1	0	0
I8	7	9	2	3	3	0.7	0.85	2.3	1	3	3	2	2	0	1	1	1	1	1	0	0	7	1	1	0	0
I9	7	9	2	3	3	0.65	0.82	1.8	1	3	3	2	2	0	1	1	1	1	1	0	0	7	1	1	0	0
I10	7	9	1	3	3	0.65	0.9	2.5	1	3	3	2	2	0	1	1	1	1	1	0	1	7	1	1	0	0
M1	14	1	1	2	1	1.65	1.25	5	3	3	3	3	3	0	3	1	1	3	8	0	0	7	8	6	0	3
M2	1	2	2	1	1	1.05	1.3	5.5	2	2	2	2	0	2	1	1	1	4	8	0	0	7	8	6	0	2
M3	14	1	1	1	1	1.3	1.35	6	3	3	3	1	1	0	6	1	1	3	9	0	0	7	8	6	0	3
M4	7	1	1	1	1	1.15	1.28	4.5	3	3	3	2	2	0	3	1	1	3	9	0	0	7	8	6	0	1
M5	7	2	2	1	1	1	1.3	4	3	1	3	2	2	0	2	1	1	3	5	0	0	7	8	6	0	2
M6	7	2	1	1	2	0.75	1.25	4.5	2	2	2	0	2	0	2	1	1	3	1	0	0	7	8	6	0	3
M7	7	2	1	1	1	0.65	1.1	3.5	2	2	2	3	0	3	3	1	1	3	1	0	0	7	8	6	0	1
M8	7	2	1	1	3	0.7	0.85	1.2	3	3	3	2	2	0	3	1	1	3	1	0	0	7	8	6	0	1
N1	1	1	2	1	1	0.98	1.05	3	2	2	2	2	0	2	2	1	1	1	5	0	0	7	7	2	0	2
N2	2	2	1	1	1	0.88	1.3	6	1	1	1	2	0	0	0	1	1	2	5	0	0	7	6	7	0	2
N3	2	2	1	3	1	0.65	0.93	4	2	1	2	3	0	3	1	2	1	2	6	1	1	6	7	3	1	5
N4	1	1	1	3	2	0.72	0.95	2.5	2	2	2	4	0	4	4	1	1	1	6	0	0	7	6	2	0	3
N5	16	1	1	1	NA	NA	1.1	4	1	1	2	4	0	4	1	2	1	1	3	0	0	7	3	2	0	1
N6	4	1	2	2	NA	NA	1.3	4.5	1	2	2	2	0	2	2	1	2	1	3	0	0	7	4	2	0	2
N7	10	1	1	1	1	0.92	1.25	4.3	2	2	2	2	0	2	3	1	1	1	9	0	0	7	8	2	0	2
N8	4	3	1	1	3	0.91	1.05	4	3	3	3	2	2	0	5	1	1	1	3	0	0	7	3	2	0	3
N9	4	1	1	2	1	0.8	0.9	5	1	1	2	2	0	2	0	1	1	3	3	0	0	7	3	5	0	4
N10	5	6	1	NA	1	NA	NA	NA	NA	3	2	2	0	2	2	1	1	1	7	0	1	7	7	4	0	1
S1	5	6	1	NA	2	NA	NA	NA	2	2	2	2	2	0	1	1	1	1	7	0	1	7	7	4	0	4
S2	4	13	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	1	1	1	4	0	5	7	6	2	0	1
S3	10	1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1	1	7	1	5	3	4	4	4	4
S4	5	4	1	3	3	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	1	1	1	7	2	6	2	5	4	3	14
S5	11	7	1	NA	1	NA	NA	NA	NA	1	3	NA	NA	NA	2	1	1	1	3	0	4	7	4	3	0	1
S6	NA	11	1	3	3	NA	NA	NA	NA	3	3	NA	NA	NA	0	1	1	1	7	0	0	4	7	2	0	1
S7	5	4	1	3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	1	1	1	2	0	5	7	5	5	0	3
S8	1	4	1	3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	1	1	1	2	0	5	7	5	5	0	1
S9	5	6	1	3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	1	1	1	2	1	5	7	5	5	0	3
S10	4	4	1	3	3	NA	NA	NA	NA	3	3	NA	NA	NA	4	1	1	1	7	2	6	7	3	4	0	1
S11	5	12	1	NA	NA	NA	NA	NA	NA	NA	3	NA	NA	NA	2	1	1	1	9	1	1	7	3	4	0	2
S12	NA	1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	1	1	1	7	3	6	4	7	3	3	13
SH1	4	3	1	1	3	NA	NA	NA	NA	3	3	NA	NA	NA	2	1	1	1	3	0	0	7	3	2	0	1
SH2	4	3	1	1	1	NA	NA	NA	NA	3	3	NA	NA	NA	2	1	1	1	3	0	0	7	3	2	0	1
SH3	4	3	1	1	3	NA	NA	NA	NA	3	3	NA	NA	NA	2	1	1	1	3	0	4	7	4	3	0	1
SH4	16	3	1	1	3	NA	NA	NA	NA	3	3	NA	NA	NA	1	6	1	1	3	0	0	7	4	3	0	1
SH5	4	3	2	1	1	NA	NA	NA	NA	3	3	NA	NA	NA	2	1	1	1	3	0	4	7	4	3	0	1
SH6	4	3	1	1	3	NA	NA	NA	NA	3	3	NA	NA	NA	1	2	1	1	3	0	0	7	3	2	0	1
SH7	8	1	2	1	3	NA	NA	NA	NA	3	3	NA	NA	NA	2	1	1	1	3	0	0	7	7	2	0	3

Appendix G. Descriptive Information, Encounter Rules, and Information Development

G.1 Naming Conventions

The naming conventions used in Arabic are complex by European standards and are only intended in this database to add some degree of realism. Names such as 1, 2, or 3 would serve the requirements of the mathematics used in this study as well as more realistic names but make for a poor demonstration.

Naming conventions include the person's "given" name (such as Achmed), a series of patronymic names spanning several generations (Achmed ibn Abdul, or Achmed son of Abdul), or what in English would be considered nicknames based on a characteristic or location (Achmed al-Tikriti, or Achmed associated with Tikrit). In this version of the database, the patronymic modifiers ibn/bin (son of, as in Osama bin Laden), or bint (daughter of) are used, although in the lingnet.org outline of current Iraqi use they do not appear. The term "al-" (roughly, "of the") does appear in the lingnet.org outline of the Iraqi naming conventions and is used here.

Another naming variant is qualification of an individual as "father of" or "mother of" a son. Hence, Abdul Abu Husayn, Abu father of Husayn, and Alia Umm Muhammed, Alia mother of Muhammad. The term Abd, or "servant/slave of," is mentioned in the outline (Muhammed Abd Allah) and is used in this database.

A person may be known and recognized under all of these conventions simultaneously.

Not all Arabic speaking countries use the above naming conventions.

G.2 Tribal Affiliations

Tribal membership is very important in determining a person's conduct and associations. Tribal membership is based on what is gleaned from the Internet on tribal associations with areas and sects. The lingnet.org data on association of certain tribes with religious sects has not been cross-checked. Tribal associations or membership labels such as Tribe A, B, etc., would satisfy the mathematical requirements but, again, make for a very poor demonstration.

Tribal membership is important for at least four reasons:

1. Tribal identity links to area and, hence, defines context.
2. Tribal identity defines religious sect membership. That is, a Shi'a tribesman (refers to both sexes, for convenience) is unlikely to embrace a Sunni agenda.
3. Tribal leadership determines to a large extent the loyalties and actions of a tribesman; the tribesman fights whomever his sheik tells him to fight.
4. A tribesman is likely to associate with members of his own tribe. If a leader of a cell is a member of a tribe, he is likely to recruit relatives and fellow tribesmen.

G.3 Encounter Rules

There are several groups in the surrogate database. The members of the groups are assumed to encounter security personnel in an open setting, such as a roadblock or control point, or passing by in the street.

The primary group is the set of individuals in the Soft Target Exploitation and Fusion (STEF) file. Descriptions and biometric identification (ID) data are considered present only if the narrative supports the data having been gathered forensically. Thus, a highly visible operation such as a hostage operation is considered to have created imagery records and fingerprints. If the subject has been observed, the subject is considered to have facial imagery on file and may or may not have fingerprints on file. If the subject is described as having been "seen," a verbal description is considered to be on file, but no biometric files are considered to be available. This is, of course, a simplification.

A second group is a set of known and previously unknown terrorists modeled after the STEF subjects. A small group of individuals are STEF subjects who are previously unidentified and assumed to be detained based on various cues. Upon detention, the individuals are identified as members of the STEF group. Other groups are assumed to be affiliated with known militias such as the Mahdi Army, unidentified terrorist groups or militias, or branches of al Qaeda. Determining indicators for these groups is difficult. Previous records are tautological, and other indicators are only detectable on detention: tribal membership, locality, and biometric stress indicators, and, of course, ID through description and biometric files.

A third group is a set of criminals. These are wanted and sought by police and security service personnel, although not actively sought because the case is too insignificant or too old, or, as yet, undetected and so not yet sought. The rule for observer visual facial ID or recognition is that if the case is too cold or the criminal as yet undetected, the observer cannot be cued by recognition. In these cases, the observer judgment of agreement of the subject's appearance with a description, if any, buried in the files, is rated as null. Cueing for stopping or detaining the subject must be by observer visual cues of stress or remote sensing of biometric stress indicators.

A fourth group is a set of friendly distractors. These subjects are assumed to be identifiable as members of friendly tribal or militia organizations. A part of this group is a set of persons assumed to be innocent distractors. These are not sought by security personnel and are assumed to be stopped randomly or on evidence of entirely reasonable stress, given the environment.

An example of an encounter scenario might be as follows:

A security element observes a market. An observer, with a link to the telefusion element, observes a passer by. The passer by demonstrates cueing behavior of apparent nervousness, the remote biometric sensor detects very rapid pulse, or the observer recognizes the subject from a wanted list of published descriptions or published imagery. This might be termed "remote information."

The observer stops the subject for examination of documents and direct biometric sensing, including GSR, pulse rate, fingerprints, and facial video. Identity documents are examined. Information is datalinked to the telefusion element. This might be termed "direct information."

The biometric data and self-declared (including identity documents) name is used to interrogate the database to link the individual with intelligence and situational data. This may be termed "linked data."

The data elements in the remote, direct (if any), and linked data are consolidated into an attribute vector and processed to determine where the subject's attributes cluster with respect to the groups. This may be termed "MDS data."

G.4 Development of Information With Time

The status of the information available for grouping the subjects is very different depending on the application. The MDS-HVI methodology may be applied in two very different, albeit equally important ways. The first and simplest is as a background process that parses records of persons known to the security services and police. These are grouped, highlighted, and prioritized for specific investigation. The second is real or near real-time (RNRT) support of field operations by an information fusion team operating at a distant site, with high-capacity information access and processing capability.

G.4.1 Background Operation

The information files are basically dossiers, with no observer input, no remotely sensed biometric information, and no directly sensed biometric information. This case is not being addressed at this time.

G.4.2 Real/Near Real-Time Remote Support

The basic scenario is a human encounter with information support. The human encounter is either random or cued. A random encounter is supported by biometric stress indicators, but a human or bio-ID of the subject is considered to trigger a cued encounter. Thus, the encounter

begins with human and biometric assessment of stress. Personal and document searches may add to the information and trigger fingerprint ID. Fingerprint ID is done by means of an automatic fingerprint reader. The fingerprint files may be digitized or on paper; in Third World environments paper files are most likely, though political identity files may be significant enough to be on a parallel, higher technology digitized system.

Digital fingerprint matching is assumed for this scenario. Matching of fingerprints using paper files requires manual processing—probably unlikely to result in an answer during the assumed time scale of these encounters due to time constraints and skilled labor requirements.

Facial imagery ID is assumed to be digital in nature. Identification using digital imagery is not very reliable but, if the files exist at all, a working assumption is that the ID is made within the time scale of this scenario. It may, in practice, be impossible.

Observer recognition based on "mug shots" is assumed applicable only to the highest priority subjects. A petty criminal or a criminal in a case that is not recent is unlikely to trigger observer recognition. Rank and file terrorists or militia are assumed to not trigger human recognition.

Agreement of observer evaluation of appearance with prior records is assumed to happen only when some event triggers delivery of the description to the observer or the subject is important enough for an observer to remember the description.

Success of digital facial imagery ID is assumed when the scenario justifies existence of prior imagery, such as military or criminal records, or photographic surveillance. Whether the infrastructure in realistic theaters of operation will permit this is uncertain. Digital facial imagery in previous notional U.S. biometric registration encounters would permit digital facial imagery matching.

Appendix H. Distances Between Selected Individuals, in the Three-Dimensional (3-D) Solution Set, SpookDemo_v12.

The final model is based on the following input parameters: Ordinal MDS type, Stress Badness Function, 3-D solution dimensionality, and unnormalized (raw) Euclidean Attribute-to-Dissimilarity Function.

Table H-1. Distances between the subject of remote inquiry and the persons in the solution set.

Persons	Distance
C3	0.227
C4	0.0961
I5	0.1151
I1	0.1327
I6	0.1376
I7	0.1387
I10	0.1395
I9	0.1411
I8	0.1456
C1	0.1581
I2	0.1591
I4	0.1756
I3	0.1891
C5	0.1933
C2	0.2353

NO. OF COPIES	ORGANIZATION
1 (PDF only)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA 8725 JOHN J KINGMAN RD STE 0944 FORT BELVOIR VA 22060-6218
1	DIRECTOR US ARMY RESEARCH LAB IMNE ALC HRR 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB RDRL CIM L 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB RDRL CIM P 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB RDRL D 2800 POWDER MILL RD ADELPHI MD 20783-1197
	ABERDEEN PROVING GROUND
1	DIR USARL RDRL CIM G (BLDG 4600)

NO. OF
COPIES ORGANIZATION

1 DIRECTOR USARL
RDRL CII
B BROOME
2800 POWDER MILL RD
ADELPHI MD 20783-1197

1 DEFENSE ACADEMY FOR
CREDIBILITY ASSESSMENT
T BROWN
7540 PICKENS AVE
FT JACKSON SC 29207

1 DEFENSE ACADEMY FOR
CREDIBILITY ASSESSMENT
J DENVER
7540 PICKENS AVE
FT JACKSON SC 29207

1 PM DOD BIOMETRICS
200 STOVALL ST
STE 10N07 HOFFMAN II
A LAZAREVICH
ALEXANDRIA VA 22332

1 PREDICTIVE ANALYTICS GROUP
UNIT 111
3510 SOUTH MICHIGAN AVE
CHICAGO IL 60653

1 DIRECTOR
AMSRD AAR AEF R
G HERC
BLDG 95 RM 253
PICATINNY NJ 07806

1 SAIC
L GIBSON
801 MAIN ST STE 300
LOUISVILLE CO 80027

ABERDEEN PROVING GROUND

18 DIR USARL
RDRL CII C
A BORNSTEIN (10 CPS)
J BRAND (2 CPS)
E HEILMAN
T HANRATTY
A NEIDERER
J RICHARDSON
M THOMAS
D WELSH

INTENTIONALLY LEFT BLANK.