
COPLINK

MANAGING LAW ENFORCEMENT DATA AND KNOWLEDGE

BY HSINCHUN CHEN, DANIEL ZENG, HOMA ATABAKHSH,
WOJCIECH WYZGA, AND JENNY SCHROEDER

In response to the September 11 terrorist attacks, major government efforts to modernize federal law enforcement authorities' intelligence collection and processing capabilities have been initiated. At the state and local levels, crime and police report data is rapidly migrating from paper records to automated records management systems in recent years, making them increasingly accessible.

HI

owever, despite the increasing availability of data, many challenges continue to hinder effective use of law enforcement data and knowledge, in turn limiting crime-fighting capabilities of related government agencies. For instance, most local police have database systems used by their own personnel, but lack an efficient manner in which to share information with other agencies [7–9].

More importantly, the tools necessary to retrieve, filter, integrate, and intelligently present relevant information have not yet been sufficiently refined. According to senior Justice Department officials quoted on MSNBC, Sept. 26, 2001, there is “justifiable skepticism

about the FBI’s ability to handle massive amounts of information,” and recent anti-terrorism initiatives will create more data overload problems.

As part of nationwide, ongoing digital government initiatives, COPLINK is an integrated information and knowledge management environment aimed at meeting some of these challenges. Funded by the National Institute of Justice and the NSF, COPLINK has been developed at the University of Arizona’s Artificial Intelligence Lab in collaboration with the Tucson Police Department (TPD) and the Phoenix Police Department (PPD). The main goal of

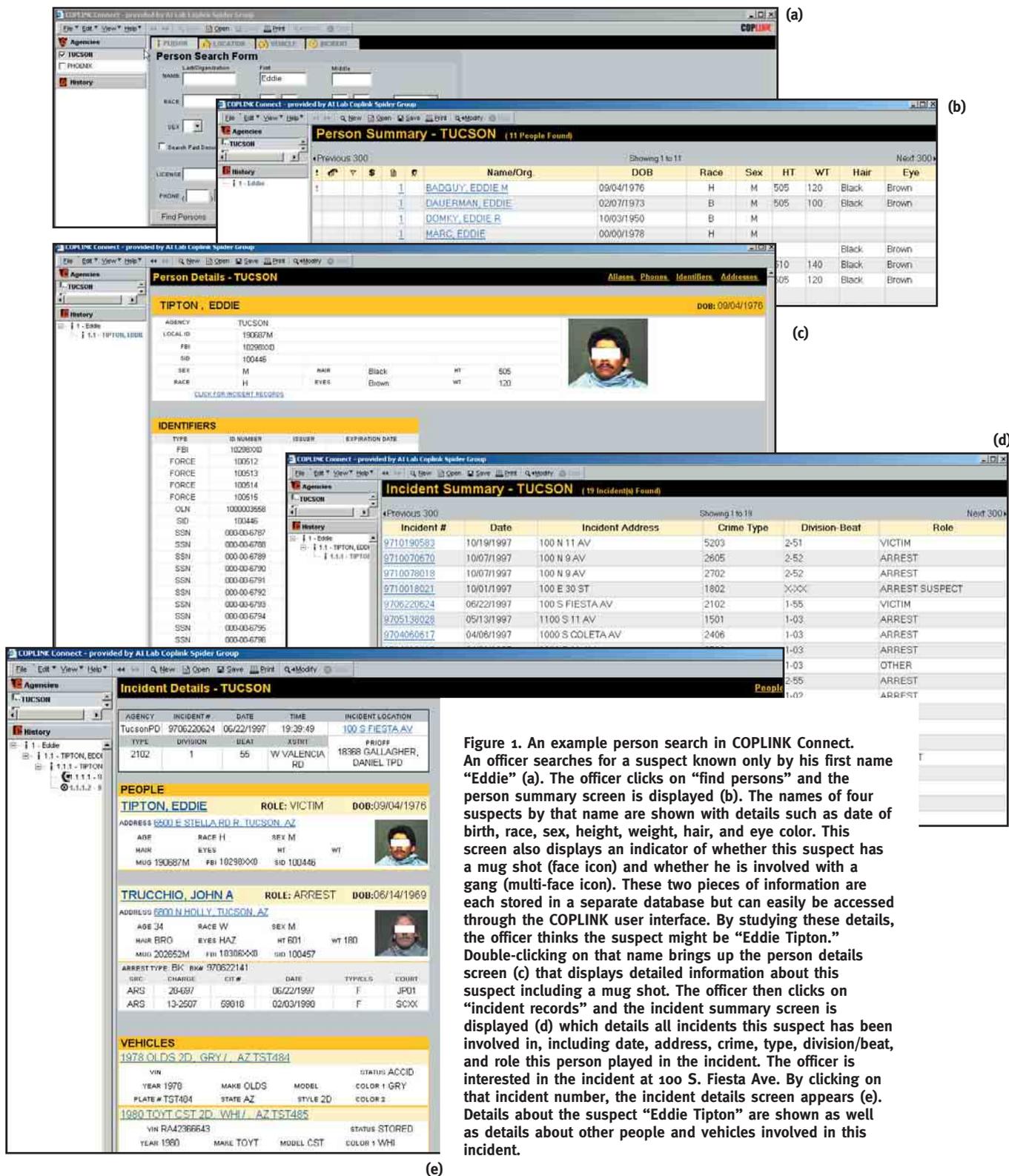


Figure 1. An example person search in COPLINK Connect. An officer searches for a suspect known only by his first name “Eddie” (a). The officer clicks on “find persons” and the person summary screen is displayed (b). The names of four suspects by that name are shown with details such as date of birth, race, sex, height, weight, hair, and eye color. This screen also displays an indicator of whether this suspect has a mug shot (face icon) and whether he is involved with a gang (multi-face icon). These two pieces of information are each stored in a separate database but can easily be accessed through the COPLINK user interface. By studying these details, the officer thinks the suspect might be “Eddie Tipton.” Double-clicking on that name brings up the person details screen (c) that displays detailed information about this suspect including a mug shot. The officer then clicks on “incident records” and the incident summary screen is displayed (d) which details all incidents this suspect has been involved in, including date, address, crime, type, division/beat, and role this person played in the incident. The officer is interested in the incident at 100 S. Fiesta Ave. By clicking on that incident number, the incident details screen appears (e). Details about the suspect “Eddie Tipton” are shown as well as details about other people and vehicles involved in this incident.

COPLINK is to develop information and knowledge management systems technologies and methodology appropriate for capturing, accessing, analyzing, visualizing, and sharing law enforcement-related information in social and organizational contexts [1].

The COPLINK system consists of two main components: COPLINK Connect is designed to allow diverse police departments to share data seamlessly through an easy-to-use interface that integrates different data sources including legacy record management systems; COPLINK Detect uncovers various types of criminal associations that exist in police databases.

COPLINK Connect

The intended users of COPLINK Connect are police officers who are typically not experienced IT users and have pressing, oftentimes mission-critical, information needs. The design of this system was closely guided by user requirements acquired through multiphase brainstorming sessions, storyboards, mock system demonstrations, focus groups, and more formally structured questionnaires and interviews. We illustrate the functionality of COPLINK Connect in Figure 1.

Most police data is scattered over distributed information sources. To find relevant information, a police officer needs to know which data sources offer what sets of data and how to access them, as well as understand each individual source's query language and user interface. He or she then must manually integrate retrieved data.

ONE OF the key functions of COPLINK Connect is to provide a one-stop access point for data to alleviate police officers' information and cognitive overload. In its current version, COPLINK Connect supports consolidated access to all major databases in the TPD. The mug shots illustrated in Figures 1a and 1e are incorporated from a separate statewide mug shot database as is the gang database illustrated in Figure 1b. Incorporating other data sources, including remote ones managed by other organizations, can be

easily accomplished. We are currently expanding the COPLINK Connect data sources to include the Arizona State Motor Vehicle database, the Computer-Aided Police Dispatching database, Tucson City Court databases, and other sources considered important by TPD officers.

There are four types of searches available to the user: person, vehicle, incident, and location. Figure 1 illustrates a person search. These search forms are chosen because they cover the most search tasks

that police officers normally perform. Based on the user requirement for simplicity (especially from field officers), we decided these four types of search can only be performed independently of each other (that is, the user is not allowed to perform a combined search such as one involving both person and vehicle search terms). A fol-

low-up user study justified such a simplified design.

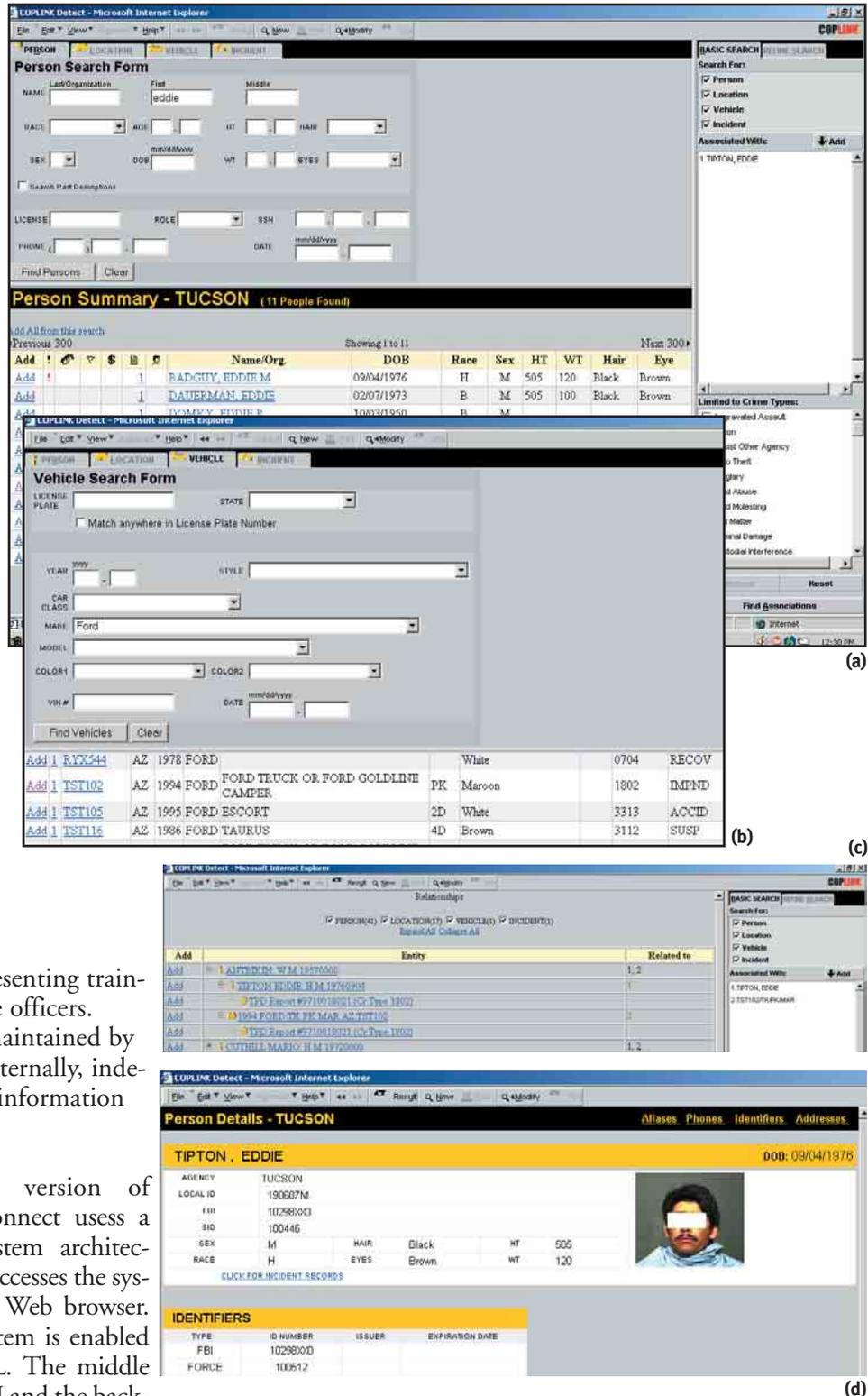
Many user interface design tradeoffs have also been made in developing COPLINK Connect. These include:

- *Partial and phonetic-based matching.* Police officers often need to conduct searches based on sketchy and incomplete information. Extensive support for partial and phonetic-based matches is built into COPLINK Connect to facilitate such searches. For instance, in Arizona, police and other law enforcement officers (for example, border patrol agents) often have to deal with misspelled Spanish names (for example, "Gweesty"). In COPLINK Connect, such names would match plausible Spanish names (for example, "Guisti").
- *Search history.* As illustrated in Figure 1d, the end user has access to his or her own search history, making data entry a bit easier and keeping the user apprised of all outstanding tasks. More importantly, the search history mechanism provides important documentation to justify and corroborate related information inquiries and subsequent actions in legal proceedings. Furthermore, search logs become important organiza-

COPLINK

Connect uses a three-tiered system architecture. The user accesses the system through a Web browser. The GUI part of the system is enabled through standard HTML. The middle tier connects the user GUI and the back-end databases

Figure 2. A sample search session in COPLINK Detect. An officer searches for a suspect known only by his first name "Eddie" and his associates for a pending investigation. A push of the "Find Persons" button displays four suspects with the name Eddie including information details such as date of birth, race, sex, height (a). Studying these details, the officer thinks the suspect is Eddie Tipton. In addition, a witness has seen the suspect drive a Ford vehicle. The vehicle search form (b) displays detailed information such as license plate, state, year, make, model, style, involvement in a gang or crime type. The officer thinks the suspect vehicle is the second one on the list and adds it to the "Associated with" list (c). All entities related to the suspect Eddie Tipton and/or the suspect vehicle Ford 1994 TST102 are displayed. The officer expands the first entity: Antriki and the ones below it. Eddie Tipton and the Ford appear as being related to it. Selecting the incident details screen (d) provides a list of incidents and find the one Eddie Tipton is involved in. All Eddie's associates are also displayed.



tional memory by representing training cases for new police officers. These search logs are maintained by COPLINK Connect internally, independent of underlying information sources.

THE CURRENT version of COPLINK Connect uses a three-tiered system architecture. The user accesses the system through a Web browser. The GUI part of the system is enabled through standard HTML. The middle tier connects the user GUI and the back-end databases using Java Servlet and JDBC, and implements the business logic using Java. This system architecture decision was based on careful consideration of the law enforcement domain:

- *Ease of installation and maintenance.* The IT

- departments of law enforcement agencies are typically understaffed. The current system architecture eliminates the needs to install or maintain software on end users' local machines.
- *Cost effectiveness.* Earlier versions of COPLINK Connect adopted a proprietary software architec-

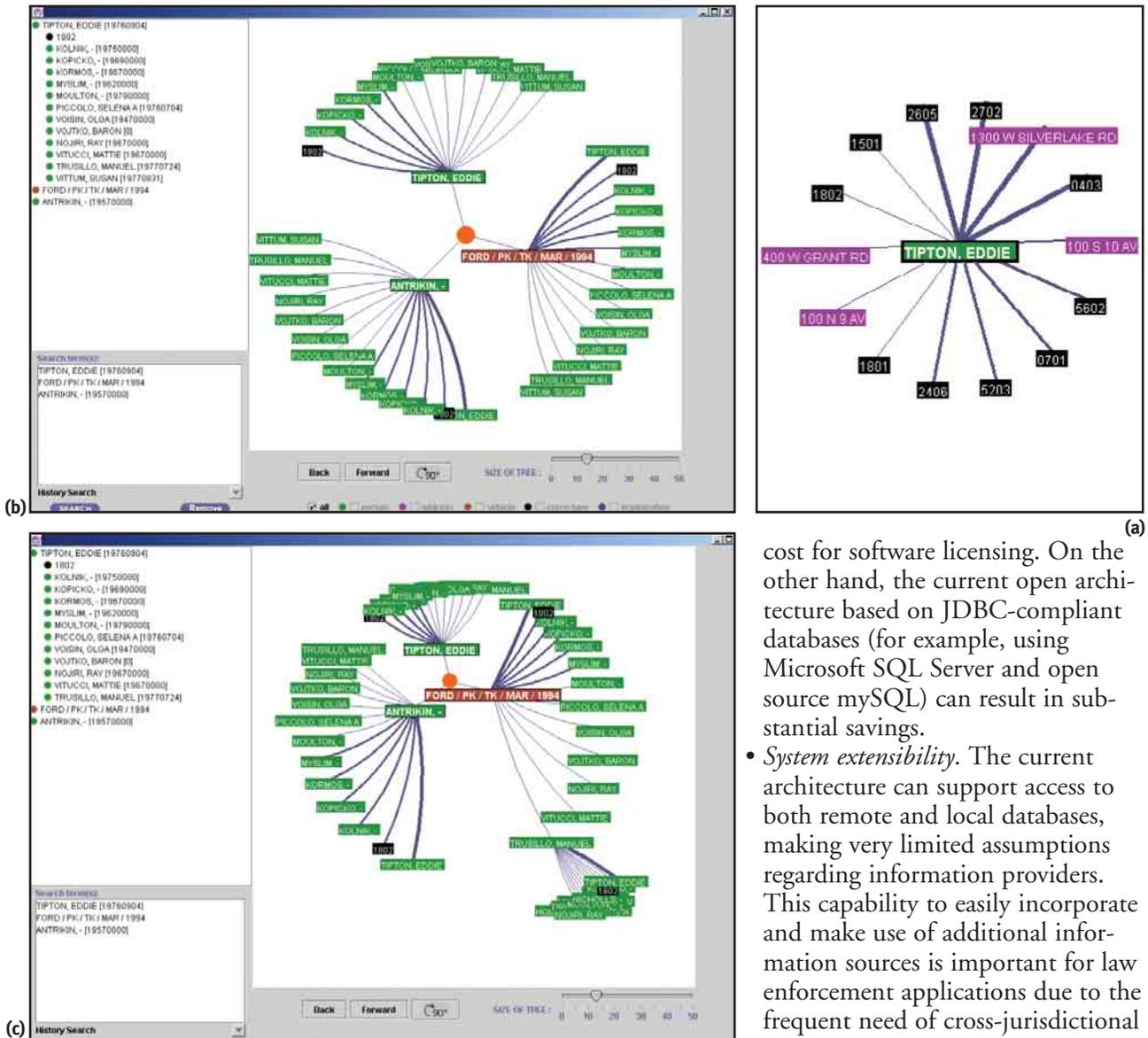


Figure 3. Hyperbolic tree view of associations in COPLINK. An officer searches for all entities related to a suspect named Eddie Tipton. All entities related to this suspect are displayed (a). The thickness of the arcs indicates the entity type as depicted at the bottom of (c). A witness has seen this suspect in a 1994 Ford truck. Another witness thinks this suspect knows a person whose last name is Antrikin. The officer then enters these two terms in the list of search terms. The screen is displayed (b) showing all entities associated with each of these search terms. The officer notices another suspect's name (Trusillo Manuel) appearing as one of the relationships. By selecting that entity, the tree expands an additional level (c) and displays all entities related to it. The officer may choose to view the hyperbolic tree using filters given on the bottom of the screen. These include filters on different entity types as well as a slider that allows the viewing of a given number of results.

ture based on Oracle products, which had the advantages of excellent system performance and the availability of a rich set of integrated development tools. It could, however, incur significant

cost for software licensing. On the other hand, the current open architecture based on JDBC-compliant databases (for example, using Microsoft SQL Server and open source mySQL) can result in substantial savings.

- *System extensibility.* The current architecture can support access to both remote and local databases, making very limited assumptions regarding information providers. This capability to easily incorporate and make use of additional information sources is important for law enforcement applications due to the frequent need of cross-jurisdictional collaborations in dealing with crimes that are typically not confined to one geographical location.

There are still some open issues with COPLINK Connect as well as some areas in need of improvement. These include:

- Power users such as detectives and specialized crime analysts have needs beyond what COPLINK Connect offers. For instance, these users often need to perform sophisticated searches combining information items across the four pre-defined search categories. This motivated the development of the second COPLINK module, COPLINK Detect.
- Selecting a suitable system architecture can be difficult. Besides the tradeoffs discussed earlier, issues such the stability and maturity of the

emerging open-standard technologies and protocols need to be carefully examined.

- Police officers spend much of their time in the field investigating crimes. To accommodate such user mobility, multimodal access to the system and delivery of contents (for example, wireless access) is important.

COPLINK Detect

Targeted for detectives and crime analysts, COPLINK Detect shares the same incident record information as the Connect module and utilizes the database indexes it generates. However, the Detect system has a completely redesigned user interface, and employs a new set of intelligence analysis tools to meet its user needs. Figure 2 shows a sample search session.

Much of crime analysis is concerned with creating associations or linkages among various aspects of a crime. COPLINK Detect uses a technique called Concept Space [3] to identify such associations from existing crime data automatically. In general, a concept space is a network of terms and weighted associations that represent the concepts and their associations within an underlying information space. COPLINK Detect uses statistical techniques such as co-occurrence analysis and clustering functions to weight relationships between all possible pairs of concepts. No hand-coded domain knowledge is necessary for COPLINK Detect to perform the Concept Space analysis.

In COPLINK Detect, detailed criminal case reports are the underlying information space and concepts are meaningful terms occurring in each case [5]. These case reports contain both structured (for example, database fields for incidents containing the case number, names of people involved, address, date, and so on) and unstructured data (narratives written by officers commenting on an incident, for example, “witness1 said he saw suspect1 run away in a white truck”). The deployed version of COPLINK Detect uses only structured data. However, we are currently experimenting with our Named Entity Extraction program to extract meaningful criminal elements (places, suspects, properties) from unstructured police narratives with some initial success [2]. Using COPLINK Detect, investigators can link known objects (for example, a given suspect) to other related objects (people and vehicles related to this suspect) to create criminal social networks that might contain useful information for further investigation. COPLINK Detect has access to a collection of 1.5 million TPD case reports, spanning a time frame from 1986 to 1999. The system can automatically

identify relationships among Person, Organization, Location, Vehicle, and Incident and Crime type.

The areas for improvement and open issues associated with COPLINK Detect include:

- The system can identify linkages among specific concepts or objects residing in existing databases. However, it does not support crime prediction. A data mining approach to identification of selected criminal patterns (for example, drug trafficking, theft) and anomalies (identification frauds) is currently under development.
- Even with user-controllable filtering, the amount of link and association information available can be overwhelming, requiring better visualization and summarization tools.
- Geographical locations play a central role in police work. COPLINK Detect identifies associations but does not map them. A map of locations could be of tremendous value to detectives and crime analysts. Along similar lines, timing also plays an important role in investigations. The current version of COPLINK Detect does not support temporal reasoning or visualization.
- Due to the complexity of COPLINK Detect, user training becomes essential. Our experience has demonstrated how critical it is to articulate to the user the differences between hard facts and soft statistical inferences (such as all system-generated associations).
- Both COPLINK Connect and COPLINK Detect could be potentially misused. The user might be tempted to go on fishing expeditions for questionable purposes. The automatically generated criminal social networks might be misinterpreted and lead to incorrect inferences. It remains an important open research question how to safeguard these potential misuses without affecting user productivity.

WE ARE currently working on a module called COPLINK Collaboration that will enable sharing of crime data and information search experience among team members. One of the most intriguing aspects of developing such a collaborative system in law enforcement concerns information privacy, security, and the legal ramifications of having to keep track of information search logs for an extended period of time. COPLINK Collaboration will also include a text-based wireless access and alerting component using cell phones and pagers to meet the needs of mobile law enforcement officers.

We are also experimenting with visualization tech-

niques such as using hyperbolic trees to better present identified associations in COPLINK Detect. A user can search all entities having a relationship with a given search term (using the concept space technique) and view the relationships in the form of a hyperbolic tree as well as in a hierarchical tree structure, as illustrated in Figure 3.

A Summary of User Studies

Several field user studies have been conducted to evaluate the COPLINK system. Detailed reports are available in [4, 8]. We summarize two studies here.

A group of 52 law enforcement personnel from TPD representing a number of different job classifications and backgrounds were recruited to participate in a study to evaluate COPLINK Connect. Both interview data and survey-data analyses support a conclusion that use of COPLINK Connect provided performance superior to using the legacy police Records Management System (RMS). In addition to the statistical data, these findings were supported by qualitative data collected from participant interviews. Comments collected from interviews indicate that COPLINK Connect was rated higher than RMS in terms of interface design, performance, and functionality. Participants indicated the quality and quantity of information from COPLINK Connect surpassed those of RMS. During the period of user evaluation, use of COPLINK Connect had led to the investigation of cases that otherwise might not have been picked up, as well as aided in making multiple arrests [4].

A pilot user study to evaluate COPLINK Detect was conducted with 33 self-selected participants, including seven crime analysts and 17 detectives from TPD. Data was collected by three methods: verbal reports, search notes from searches they performed, and electronic transaction logs. Participants indicated that COPLINK Detect required minimum training (eight participants started to use the system effectively even without any training); the system improved case closure and crime solvability through uncovering critical associations; and the system was very responsive (all the associations and related incident reports were identified in less than one minute) [8].

Conclusion

COPLINK Connect was formally deployed at TPD in the spring of 2001. There are currently about 500 COPLINK Connect users in TPD, covering almost all TPD job classification and ranks. COPLINK Detect was recently deployed at TPD and is being used by 120 detectives. The deployment of COPLINK in several law enforcement agencies in the Phoenix area is under development. Agencies

from other states also have shown strong interest in using COPLINK.

Developed to facilitate federal, state, and local law enforcement information sharing and knowledge management, COPLINK serves as a model for the next-generation information systems aimed at improving the government's crime fighting capabilities and facilitating its homeland security effort.

For more information, visit ai.bpa.arizona.edu/go/datawarehousing. **C**

REFERENCES

1. Atabakhsh, H., Schroeder, J. Chen, H., Chau, M., Xu, J., Zhang, J., and Bi, H. COPLINK knowledge management for law enforcement: Text analysis, visualization and collaboration. *National Conference on Digital Government*. (Los Angeles, CA, May 21–23, 2001).
2. Chau, M., Xu, J., and Chen, H. Extracting meaningful entities from police narrative reports. In *Proceedings of the Second National Conference on Digital Government Research*. (Los Angeles, CA, May 20–22, 2002).
3. Chen, H., Schatz, B., Ng, T., Martinez, J., Kirchoff, A., and Lin, C. A parallel computing approach to creating engineering concept spaces for semantic retrieval: The Illinois Digital Library Initiative Project. *IEEE Trans. Pattern Analysis and Machine Intelligence* 18, 8 (1996), 771–782.
4. Hauck, R. and Chen, H. COPLINK: A case of intelligent analysis and knowledge management. In *Proceedings of the 20th Annual International Conference on Information Systems*. (1999), 15–28.
5. Grishman, R. Information extraction: Techniques and challenges. M.T. Pazienza, Ed. *Information Extraction*. Springer-Verlag, New York, NY, 1997.
6. Hoogeveen, M. J. and van der Meer, K. Integration of information retrieval and database management in support of multimedia police work. *J. Information Science* 20, 2 (1994), 79–87.
7. Lingerfelt, J. Technology as a force multiplier. In *Proceedings of the Conference in Technology Community Policing*. National Law Enforcement and Corrections Technology Center, 1997.
8. Martinez, J. and Moosman, A. COPLINK Detect User Study Report. University of Arizona, AI Lab Technical Report, 2001.
9. Pliant, L. High-technology solutions. *The Police Chief* 5, 38 (1996), 38–51.

HSINCHUN CHEN (hchen@eller.arizona.edu) is McClelland Endowed Professor in the MIS department at the University of Arizona, Tuscon.

DANIEL ZENG (zeng@eller.arizona.edu) is an assistant professor in the MIS department at the University of Arizona, Tuscon.

HOMA ATABAKHSH (homa@eller.arizona.edu) is a principal research specialist in the MIS department at the University of Arizona, Tuscon.

WOJCIECH WYZGA (wwtza@knowledgecc.com) is Vice President, Engineering at the Knowledge Computing Corporation, Tuscon, AZ.

JENNY SCHROEDER (jschroel@ci.tuscon.az.us) is a lieutenant in the Tuscon Police Department, AZ.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
