

Submission to the National Conference on Digital Government, 2001
Submission Type: PAPER (SHORT PAPER preferred)

**Building an Infrastructure for Law Enforcement Information
Sharing and Collaboration: Design Issues and Challenges**

Michael Chau

PhD Student/Research Associate, Dept. of MIS
The University of Arizona
1130 E Helen, Rm 430X
Tucson, AZ 85721
mchau@bpa.arizona.edu
<http://ai.bpa.arizona.edu/~mchau/>

Homa Atabakhsh

Associate Director, COPLINK Center for Excellence
The University of Arizona
1130 E Helen, Rm 430KK
Tucson, AZ 85721
homa@bpa.arizona.edu

Daniel Zeng

Assistant Professor, Dept. of MIS
Associate Director, Hoffman E-Commerce Lab
The University of Arizona
1130 E Helen, Rm 430K
Tucson, AZ 85721
zeng@bpa.arizona.edu
<http://bpa.arizona.edu/~zeng/>

Hsinchun Chen

McClelland Endowed Professor, Dept. of MIS
Director, Artificial Intelligence Lab
The University of Arizona
1130 E Helen, Rm 430Z
Tucson, AZ 85721
hchen@bpa.arizona.edu
<http://ai.bpa.arizona.edu/hchen/>

Keywords: Delivery of Information, Information Filtering, Data Integration, Collaboration, Security.

Introduction

The rapid advancement of information technologies and the Internet provides great opportunities as well as challenges for government agencies. On one hand, these technologies allow better use and sharing of information. With the growing popularity of the Web, information can be shared among agencies or made accessible to the public more easily than ever. On the other hand, however, there are many issues and challenges which are not yet completely resolved. In this paper, we report on our experience in designing a collaborative information sharing infrastructure in the law enforcement domain. We also review some of the major challenges we encountered in this project.

Our project aims at efficiently satisfying the various information needs of law enforcement personnel. We have collaborated with the Tucson Police Department (TPD) which provides us with an excellent testbed for our infrastructure. TPD has more than 1000 employees and serves a population of 475,000 citizens.

Problems in Information-Seeking in Law Enforcement

Information is heavily needed in various processes in law enforcement, such as crime analysis and investigation. Crime analysts often need to inquire different distributed data sources to obtain the required information on a criminal case. Based on our previous successful implementation of the COPLINK system at TPD [Hauck & Chen, 1999], we identified a few areas that can be further improved.

Information Access

Many different data sources are available at TPD. Besides several internally maintained databases, external databases maintained at the county level (the Pima County) and the state level (the Arizona State) also need to be inquired frequently by police officers. In addition, there are many other useful data sources on the publicly accessible Web, that are useful to law enforcement personnel. These include search services such as Property Assessor Information, Map Guide, People Search and Reverse Lookup. In order to find the desired information, police officers have to know where such data sources are located, and whether such sources have good authority and credibility. Without substantial experience and knowledge, it becomes difficult to locate important and accurate information. Also, police officers have to query each of the data sources individually. This requires a large amount of effort to query all the relevant data sources, each with a different search interface and display. Users should be able to get the required information through a single, integrated interface.

Information Monitoring

Another problem is the reactive nature of the data sources. Police officers often have to track the activities of a particular suspect or the whereabouts of a vehicle. These monitoring functions are usually not available in most search systems. As a result, the data sources have to be checked manually on a regular basis, giving the user a heavy mental effort.

Personalization

Personnel with different job duties and geographical locations have very different information needs. Given the large number of data sources, police officers face the problem of finding the right data sources that are relevant to their jobs. The current system is not customizable to satisfy personalized information needs.

Collaboration

Police officers with different job functions and at different locations have acquired a vast amount of knowledge. A police officer may obtain some knowledge from the observations during patrol, while a crime analyst may get some particular insights when doing an investigation. Such knowledge however, is tacit and not efficiently shared. When a police officer needs some particular information, he or she does not know whom to contact. We also realize that there are situations where two different units are working on two closely related cases (e.g., related to the same person), without knowing that another person is working on a similar case. As a result the two units are not able to collaborate and share the knowledge. The issues of sharing knowledge in a collaborative manner and linking together people working on similar cases have to be addressed.

Design Challenges

In order to address the above problems, we propose to build an infrastructure for information sharing and collaboration for TPD. During our design, we face some design challenges that we address in this section. While these issues are of definite importance in the law enforcement domain, we believe they are also applicable to digital government research in other domains.

Distributed, Heterogeneous Data sources

As mentioned above, there are many data sources that a police officer, for example a crime analyst, needs to frequently inquire in order to perform an investigation. These data sources are mostly in the form of databases that have been in operation for years or decades. Other data sources include information or search services which are available on the Web. Because the data sources are funded and maintained by multiple agencies, they employ a wide range of different hardware platforms, database systems, network protocols, data schemas, and user interfaces. This poses a great challenge for managing these data sources. The question is how the system can provide a single interface to satisfy a user's information needs. Several research groups have proposed the use of distributed ontology or agent technology database to deal with this problem (e.g., Arens, Knoblock & Shen, 1996; Bayardo et al, 1997; Ambite et al, 2001; Bouguettaya et al, 2001). Although these systems can effectively manage multiple data sources, they do not allow users to add a new resource easily or to rate the quality or credibility of an existing data source.

Security and Confidentiality

In the law enforcement domain, security is of great concern. Most of the data we deal with are highly sensitive. Improper use of data could lead to fatal consequences. Data security therefore requires very careful considerations in our infrastructure. Police officers of different ranks and job functions have different levels of access to each of the data sources. We also need to deal with the varied security requirements of the data sources maintained at different levels in the government hierarchy, namely the city level, the county level, the state level, and the federal level.

Another issue here is the confidentiality of an investigation. Some criminal cases, such as an internal investigation, are highly confidential with only a few authorized personnel having access to them. With the rising rate of computer crimes and data abuse in recent years, the authorization and authentication process have to be carefully designed to ensure integrity.

Trust and Willingness to Share Information

In a recent workshop sponsored by the National Science Foundation and the University of Arizona COPLINK Center (COPLINK 2001), trust and willingness to share information has been voted as one of the most critical issues in information sharing in the law enforcement domain. Agencies are not motivated to share information and knowledge if there is no immediate gain. They may also fear that information being shared would be misused, resulting in legal liabilities.

This becomes an interesting design issue in building an information sharing and collaboration infrastructure.

Proposed Architecture

Taking into consideration the above issues and challenges, we propose the COPLINK Collaboration system, an infrastructure which tries to address those problems to a certain extent.

Linking People Together

We adapt an approach based on collaborative filtering in our system. Collaborative filtering is defined as the collaboration among people to help one another perform filtering by recording their reactions to the information they read (Goldberg et al, 1992). While traditional collaborative filtering relies on documents read (e.g., Konstan et al, 1997) or items purchased by users (e.g., Amazon.com), we make use of the users' search actions and search histories.

The rationale behind this is that when two users search for the same information, it is likely that the users have similar information needs and that they may possibly be working on two related cases. When a user performs a search query through COPLINK Collaboration, the search query will be forwarded to the corresponding data sources. At this stage we have not introduced the use of ontology; users have to choose which data sources they want to inquire. Our system will then return the search results to the user and store the search session in the user profile database for further analysis. Collaboration will be performed by applying data mining techniques on these user search profiles. Our previous research has shown that machine learning and data mining techniques such as ID3, genetic algorithms and relevance feedback can be applied in inductive search query analysis (Chen et al, 1998). Using these techniques, the system tries to (1) recommend similar cases to users and (2) identify police officers with similar information needs. If the system finds that two police officers have performed similar searches (e.g., similar search keywords or search criteria), it will alert both users and provide a way for them to contact each other for further collaboration.

Since no extra effort is required from the users, we expect them to be more willing to share information. Figure 1 shows a simplified version of the system architecture.

Finding Useful Data Sources

Another measure we use for collaboration is users' rating of data sources. Users can give ratings to or write reviews for the data sources they use. Our algorithm will be based on a hybrid approach which combines content-based filtering and collaborative filtering (Balabanovic & Shoham, 1997). Our filtering algorithm will also apply data mining techniques. Useful information sources will be recommended to a user based on one's past search sessions and the recommendations of other users who have similar information needs. This allows users to have a personalized list of data sources they need while irrelevant sources are filtered.

Information Monitoring

Users can also set up monitoring tasks through the system. If a user wants to monitor a particular database for a particular query, the system will store the monitor task in the user profile database and check the corresponding database periodically. When relevant data is updated or inserted into the database, the system will send an alert message to the user through the user interface, email, voice message or cellular phone.

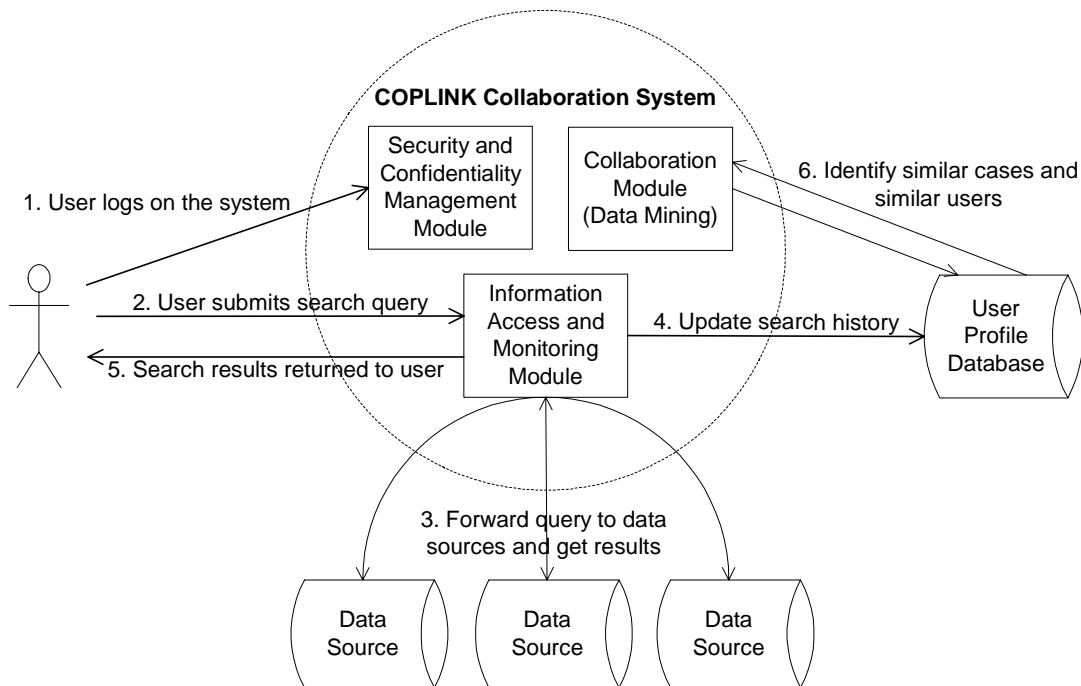


Figure 1. Proposed system architecture to support information sharing and collaboration

Security and Confidentiality

Security is enforced at two levels. First, a user has to supply username and password for logging on the system. This ensures the personal information and search histories are not accessible by others. When a user needs to access data sources that require further authorization and authentication, the user has to provide the corresponding username and password.

When a user performs a search, he or she can specify the confidential level of the search. From our user requirement studies, we found that user would like to have fine-grained control on the confidential level. Some cases may be accessible by personnel within the same unit, while some other cases may only be accessible by several specific individuals.

Conclusion

In this paper, we have identified the major challenges in designing an information sharing and collaboration infrastructure within the law enforcement domain. We have proposed and are implementing an architecture that addresses some of the major issues. We also hope our experience can provide useful insight to other digital government research projects.

References

- J. Ambite, Y. Arens, E. Hovy, A. Philpot, L. Gravano, V. Hatzivassiloglou & J. Klavans (2001). Simplifying Data Access: The Energy Data Collection Project, *IEEE Computer*, 34(2), pages 47-54, February 2001.
- Y. Arens, C. Knoblock and W. Shen (1996). Query Reformulation for Dynamic Information Integration, *Journal of Intelligent Information Systems*, 1996.
- M. Balabanovic and Y. Shoham (1997). Fab: Content-based, Collaborative Recommendation, *Communications of the ACM*, 40(3), pages 66-72, March 1997.
- R. J. Bayardo, W. Bohrer, R. Brice, et. al. (1997). InfoSleuth: Agent-based Semantic Integration of Information in Open and Dynamic Environments, *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 195-206, Tucson, AZ, USA, May 11-15, 1997
- A. Bouguettaya, M. Ouzzani, B. Medjahed, & J. Cameron (2001). Managing Government Databases, *IEEE Computer*, 34(2), pages 56-64, February 2001.
- H. Chen, G. Shankaranarayanan, A. Iyer, and L. She (1998). A Machine Learning Approach to Inductive Query by Examples: An Experiment Using Relevance Feedback, ID3, Genetic Algorithms, and Simulated Annealing, *Journal of the American Society for Information Science*, 49(8), pages 693-705, June 1998.
- COPLINK (2001). *Conference on Law Enforcement Information Sharing and Analysis: Technologies and Practices for the 21st Century*, Tucson, AZ, January 24-27, 2001.
<http://ai.bpa.arizona.edu/workshop/CoplinkCenter/>
- D. Goldberg, D. Nichols, B. Oki, & D. Terry (1992). Using Collaborative Filtering to Weave an Information Tapestry, *Communications of the ACM*, 35(12), pages 61-69, December 1992.
- R. Hauck and H. Chen (1999). Coplink: A Case of Intelligent Analysis and Knowledge Management, *Proceedings of the International Conference on Information Systems(ICIS)*, pages 15-28, Charlotte, NC, December 1999.
- J. Konstan, B. Miller, D. Maltz, J. Herlocker, L. Gordon & J. Riedl (1997). GroupLens: Applying Collaborative Filtering to Usenet News, *Communications of the ACM* 40(3), pages 77-87, March 1997.