Artificial Intelligence for Homeland Security

Hsinchun Chen and Fei-Yue Wang, *University of Arizona*

The tragic events of 11 September 2001, and the subsequent anthrax letter contaminations had drastic effects on many aspects of US society. Terrorism became the most significant threat to national security because of its real and potential damage to our infrastructure, economy, and people. In response to this challenge, federal authorities began actively implementing comprehensive strategies and measures to achieve three homeland security objectives:[1]

1541-1672/05/$20.00 © 2005 IEEE
Published by the IEEE Computer Society

- prevent future terrorist attacks,
- reduce the nation's vulnerability, and
- minimize the damage and recovery from attacks that occur.

State and local law enforcement agencies, likewise, have become more vigilant about criminal activities that harm public safety and threaten national security.

Researchers in the natural, computational, and social sciences as well as engineering, medicine, and many other fields have responded to the government's call for science and technology to help enhance its capabilities in fighting the new counterterrorism war. Information technology is cited as an indispensable part in making our nation safer.[2] IT can support intelligence and knowledge discovery by collecting, processing, analyzing, and developing applications for terrorism- and crime-related data.[3] Federal, state, and local authorities can use the results to make timely decisions, select effective strategies and tactics, and allocate appropriate resources to detect, prevent, and respond to future attacks.

Critical mission areas

In its report, *National Strategy for Homeland Security*, the US Department of Homeland Security (DHS) identifies six critical mission areas where IT can contribute to accomplishing the three strategic national security objectives.[1]

Intelligence and warning

Although terrorism depends on surprise in its attacks, terrorist activities are not random. Nor are they impossible to track. Terrorists must plan and prepare before executing an attack by selecting a target, recruiting and training executors, acquiring financial support, and traveling to the targeted country.[4] To avoid being preempted by authorities, they might hide their true identities and disguise attack-related activities, just as other criminals do.[5]

By analyzing communication and activity patterns, IT makes it possible to detect deceptive identities. By employing other surveillance and monitoring techniques, we can issue timely, critical alerts through intelligence and warning systems and prevent attacks or crimes.

Border and transportation security

Terrorists enter a targeted country through an air, land, or sea port, where customs and immigration authorities gather information daily. This information includes traveler identities, images, fingerprints, vehicles used, and other characteristics.

By sharing and analyzing information from multiple sources, we can create *smart borders* that greatly improve counterterrorism and crime-fighting capabilities. Smart borders depend on technologies such as information sharing and integration, collaboration and communication, biometrics, and image and speech recognition.

Domestic counterterrorism

Because both international and domestic terrorists might participate in local crimes, state and local law enforcement agencies can contribute to homeland security missions when they investigate and prosecute crimes.

> By sharing and analyzing information from multiple sources, we can create smart borders that greatly improve counterterrorism and crime-fighting capabilities.

Experts regard terrorism as a type of organized crime, like gang activity and narcotics trafficking, in which multiple offenders cooperate to carry out offenses. IT that helps find cooperative relationships and interactive patterns among criminals can also help in analyzing terrorism. In addition, monitoring criminal use of advanced IT can help public-safety personnel and policy makers.

Protecting critical infrastructure and key assets

Roads, bridges, water supplies, and many other physical service systems are critical infrastructure and key national assets. Their vulnerabilities make them potential targets of terrorist attacks. Virtual (cyber) infrastructures such as the Internet are also vulnerable to intrusions and other threats.[6]

To monitor these assets, we need not only physical devices, such as sensors and detectors, but also advanced information technologies that can model normal use behaviors and distinguish abnormal behaviors from them. Such information can guide the selection of protective or reactive measures to secure these assets from attacks.

Defending against catastrophic terrorism

Terrorists attacks that use weapons of mass destruction or other means—like hijacking commercial airlines for suicide missions—to kill thousands of civilians at a time can devastate a society. Information systems that effectively collect, access, analyze, and report data relevant to catastrophic events are critical to helping prevent, detect, and manage responses to these attacks.[7]

Emergency preparedness and responses

Prompt and effective responses reduce the damage in national emergencies. In addition to helping defend against catastrophic terrorism, IT can help design and test optimized response plans,[8] identify experts, train response professionals, and manage the consequences of an attack. Moreover, information systems that facilitate social and psychological support for attack victims can help society recover from disasters.

Security domain challenges

Intelligence and security agencies already gather large amounts of data from various sources. In addition to the usual difficulties of processing and analyzing large data stores, counterterrorism and crime-fighting applications pose some unique IT problems and challenges.

Distributed criminal enterprises

Terrorism and other kinds of organized crime are often geographically and temporally dispersed. As a result, investigations must cover multiple offenders and criminal activities in different places at different times. This can be fairly difficult, given limited intelligence and security agency resources. Moreover, as computer and Internet technologies advance, criminals are using cyberspace to commit various types of cybercrimes under the disguise of ordinary online transactions and communications.

Diverse sources and formats

Large data volumes and diverse sources and formats create significant challenges, including information stovepipes and information overloads.

Intelligence and Security Informatics Resources

The US National Science Foundation and Department of Justice's National Institute of Justice funded the First NSF/NIJ Symposium on Intelligence and Security Informatics (ISI) in June 2003. The two-day program included five keynote speakers, 14 invited speakers, 34 regular papers, and six posters. It has grown each year since. In 2005, the IEEE took over the sponsorship. The following proceedings document the conferences:

- H. Chen et al., eds., *Intelligence and Security Informatics: Proc. 1st NSF/NIJ Symp. Intelligence and Security Informatics*, Springer, 2003.
- H. Chen et al., eds., *Intelligence and Security Informatics: Proc. 2nd Symp. Intelligence and Security Informatics*, Springer, 2004.
- P. Kantor et al., eds., *Intelligence and Security Informatics: Proc. IEEE Int'l Conf. Intelligence and Security Informatics*, Springer, 2005.

ISI 2006 is planned for 22–24 May 2006, in California (www.isiconference.org).

Many federal research and funding agencies have established new research programs that aim to address different facets of national security research. Not intending to be comprehensive, we summarize some significant programs here, especially relevant to ISI academic researchers in universities and research institutes:

- *National Science Foundation*.The NSF has issued several Information Technology Research program announcements with a national security focus. The NSF directorates for Computer and Information Science and Engineering (www.nsf.gov/dir/index.jsp?org=CISE) and for Social, Behavioral, and Economic Sciences (www.nsf.gov/dir/index.jsp?org=SBE) are encouraging multidisciplinary research projects relevant to ISI. The NSF/CIA Knowledge Discovery and Dissemination program is a good example of joint NSF and intelligence community funding initiatives. Most NSF-funded projects stress scientific innovation.
- *Department of Homeland Security*. The DHS has funded research centers at four universities through its Centers of Excellence program (www.dhs.gov/dhspublic/interapp/editorial/editorial_0498.xml). Many new ad hoc initiatives are also in development in areas such as terrorism informatics research, bioagent surveillance, smart borders, biometrics, and critical infrastructure protection.
- *Department of Defense and intelligence community*. Since the Total Information Awareness program ended in disaster, the DOD and the intelligence community haven't publicized many new research activities relevant to ISI. The Advanced Research and Development Activity is one exception. The ARDA program aims to develop advanced information technologies for the intelligence community.
- *Centers for Disease Control and Prevention and National Institutes of Health*. The CDC and NIH support national security research relevant to infectious diseases and bioagent surveillance. The NIH's National Library of Medicine has issued solicitations in crisis management relevant to public health. Like the DHS, the CDC appropriates significant funding to state and local jurisdictions for their public health, disease surveillance, and emergency response needs. Such projects are often short-term and for implementation purposes.
- *Department of Justice*. The DOJ and its research arm, the National Institute of Justice (www.ojp.usdoj.gov/nij/), traditionally fund R&D projects relevant to local and state public safety and law enforcement. More recently, however, the NIJ has focused some research programs on counterterrorism. The NIJ has limited funding ability, so most of its projects are smaller scale than other agencies and oriented toward single investigators. Most projects must demonstrate significant public safety relevance and value.
- *Office of Naval Research, Air Force Research Labs, and other funding agencies*. Funding opportunities exist with traditional armed-forces research agencies as they adjust existing programs to the new homeland security mission. Several private foundations have also begun to support selected homeland-security-related research, particularly in citizen responses, disaster relief, and education.

The intelligence and security domain differs from domains such as marketing, finance, and medicine. These domains can collect data from particular sources, such as a company's sales records or a patient's hospital medical histories, but the intelligence and security domain doesn't have a well-defined data source. Investigators must gather both authoritative information (for example, crime incident reports, telephone records, financial statements, and immigration and customs records) and open source information (news stories, journal articles, books, and Web pages). Data formats range from structured database records to unstructured text, image, audio, and video files. Important information such as criminal associations might be available but only in unstructured, multilingual texts, which are difficult to access and retrieve.

Moreover, as data volume increases, extracting intelligence and knowledge from it becomes more difficult.

Crime and intelligence analysis techniques

Current research on counterterrorism and crime-fighting technologies has only begun to address homeland security challenges.

IT has developed several potentially helpful tools and methodologies, including data integration, data analysis, text mining, image and video processing, and evidence combination.[2] However, how to employ them in the intelligence and security domain remains an unanswered question, as does how to use them effectively in national security mission areas.

Turning raw homeland security data into actionable intelligence requires significant research and advancement in various subdisciplines of artificial intelligence. These subdisciplines include data mining, text mining, Web mining, natural language processing, planning, reasoning, conflict resolution, link analysis, and search algorithms.

ISI: Emergence of a discipline

Facing critical national security missions in the context of various data and technical-domain challenges, we see a pressing need to develop the science of *intelligence and security informatics*. To this end, we've organized an annual conference to define ISI (see the "Intelligence and Security Informatics Resources" sidebar for conference proceedings). Its main objective is to develop advanced information technologies, systems, algorithms, and databases for national security applications through

Table 1. Comparing biomedical informatics and ISI.

Comparison areas		Biomedical informatics	Intelligence and security informatics
Challenges	Domain-specific	Complexity and uncertainty associated with organisms and diseases Critical decisions regarding patient well-being and biomedical discoveries	Geographically and temporally distributed organized crimes Cybercrimes on the Internet Critical decisions related to public safety and homeland security
	Data	Information stovepipes and overload • HL7 XML standard • Public Health Info. Network Messaging System • Patient records, diseases data, medical images	Information stovepipes and overload • Justice XML standard • Criminal incident records • Multilingual intelligence open sources
	Technology	Ontologies and linguistic parsing Information integration Data and text mining Medical decision-support systems and techniques	Information integration Criminal network analysis Data, text, and Web mining Identity management and deception detection
Methodology		Knowledge discovery from databases	Knowledge discovery from databases
Contributions	Scientific	Computer and information science, sociology, policy, legal Clinical medicine and biology	Computer and information science, sociology, policy, legal Criminology, terrorism research
	Practical	Public health Patient well-being Biomedical treatment and discovery	Crime investigation and counterterrorism National and homeland security

an integrated technological, organizational, and policy-based approach.[3]

ISI vs. biomedical informatics

We compared ISI with biomedical informatics, an established academic discipline addressing information management issues in biological and medical applications.[9-11] Table 1 summarizes the similarities and differences.

The two disciplines have much in common. In terms of data characteristics, both face information stovepipe and overload problems. In terms of technology development, both are searching for new approaches and methods as well as innovative uses of existing techniques. In terms of scientific contributions, both can add insights and new knowledge to various academic disciplines.

Most importantly, both need a research framework to guide their development. A framework based on knowledge management and data mining has begun to emerge in biomedical informatics.[11] ISI can benefit from this work. We believe the *knowledge discovery from databases* methodology might prove critical in addressing unique ISI challenges.[12] KDD has already proved successful in other information-intensive, knowledge-critical domains including business, engineering, biology, and medicine.

Caveats for data mining

The potential negative effects of intelligence gathering and analysis on individual privacy and civil liberties are well publicized.[13,14] Many laws, regulations, and policy agreements govern data collection, confidentiality, and reporting. They can directly impact ISI technology development and applications.

Intelligence and security agencies as well as ISI researchers must be aware of these laws and regulations in their work. Moreover, we recommend a hypothesis-guided, evidence-based approach to crime and intelligence analysis research. In other words, probable and reasonable causes and evidence should exist before targeting particular individuals or data sets for analysis.

Researchers must strictly follow investigative and legal procedures. It is neither ethical nor legal to "fish" for potential criminals from diverse and mixed crime, intelligence, and citizen-related data sources. For example, the US Congress shut down DARPA's Total Information Awareness program and the Multistate Antiterrorism Information Exchange (Matrix) system because of their potential for misusing citizen data and impairing civil liberties.[14,15]

In this issue

This special issue consists of seven articles, which we summarize here in light of the six critical mission areas identified in DHS's *National Strategy for Homeland Security* report.[1]

"Service-Based Computing on Manets: Enabling Dynamic Interoperability of First Responders," by Joseph Kopena, Evan Sultanik, Gaurav Naik, Iris Howley, Maxim Peysakhov, Vincent A. Cicirello, Moshe Kam, and William Regli, relates directly to the emergency-preparedness-and-response mission area. The authors describe the development of the Philadelphia Area Urban Wireless Network Testbed. The PA-UWNT is a mobile ad hoc network consisting of PDAs, tablet computers, and laptops. Potential applications include coordinated police presence at large public events, medical personnel at an accident scene, emergency responders to a natural disaster, and other homeland security scenarios.

"Ontology-Centered Syndromic Surveillance for Bioterrorism," by Monica Crubézy, Martin O'Connor, David L. Buckeridge, Zachary Pincus, and Mark A. Musen, contributes to the mission of defending against catastrophic terrorism. The authors describe a system architecture that supports knowledge-based data integration and problem solving for monitoring prediagnostic health-related data for nascent disease outbreaks. A set of reference ontologies supports semantic data integration, and a parallelizable blackboard architecture implements a way to invoke appropriate problem-solving methods. The authors demonstrate their approach with BioSTORM, an experimental system that offers an end-to-end solution to syndromic surveillance.

"Deception Detection through Automatic, Unobtrusive Analysis of Nonverbal Behavior," by Thomas O. Meservy, Matthew L.

Jensen, John Kruse, Douglas P. Twitchell, Gabriel Tsechpenakis, Judee K. Burgoon, Dimitris N. Metaxas, and Jay F. Nunamaker Jr., supports intelligence interrogations and passenger screening. The authors present a system for detecting deception from nonverbal behavioral cues. The system extracts information about hand and head movements and automatically identifies behavioral patterns that indicate deception. In tests, the system classifies deception and truth with greater accuracy than humans.

"US Domestic Extremist Groups on the Web: Link and Content Analysis," by Yilu Zhou, Edna Reid, Jialun Qin, Hsinchun Chen, and Guanpi Lai, describes the development of methodologies for capturing, classifying, and organizing domestic extremist Web site data and using it for content and link analysis. The authors show results indicating that such analyses can help identify interorganizational structure and cluster affinities among domestic extremist groups. This work could contribute to the intelligence-and-warning mission area as well as domestic counterterrorism.

"Rule + Exception Strategies for Security Information Analysis," by Yiyu Yao, Fei-Yue Wang, Jue Wang, and Daniel Zeng, discusses the knowledge representation and data mining framework based on the "rule + exception" methodology and its potential application in security-related systems. Such techniques would be appropriate for database applications in the intelligence-and-warning mission area.

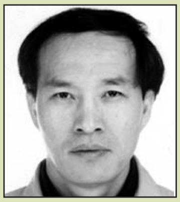
"Distributed Interactive Video Arrays for Event Capture and Enhanced Situational Awareness," by Mohan M. Trivedi, Tarak L. Gandhi, and Kohsia S. Huang, describes a multicamera video surveillance approach called DIVA, which can be used for vehicle tracking, perimeter monitoring, and bridge structure monitoring as well as for people tracking and activity analysis. Such video surveillance techniques support the protecting-critical-infrastructure mission as well as emergency preparedness and response.

"Applying Authorship Analysis to Extremist-Group Web Forum Messages," by Ahmed Abbasi and Hsinchun Chen, has strong relevance to the intelligence-and-warning mission. The authors study the problem of online anonymity by using techniques derived from literary analysis. In applying these techniques to English and Arabic extremist-group Web forum messages, they successfully evaluated the performance impact of different language features and classification techniques across both languages.

A new discipline such as ISI requires careful cultivation and development by researchers and practitioners from many different disciplines—from computer science to electrical engineering and public health. Many opportunities exist for developing innovative, high-impact ISI projects. Active research will help improve knowledge discovery and dissemination and enhance information sharing and collaboration among academics, government agencies, and industry to address a new and changing threat to national security. ■

The Authors

Hsinchun Chen is McClelland Professor of Management Information Systems at the University of Arizona. His research interests include intelligence analysis; data, text, and Web mining; digital libraries, knowledge management, medical informatics, and Web computing. He received his PhD from information systems from New York University. Contact him at MIS Dept., Univ. of Arizona, Tucson, AZ 85721; hchen@eller.arizona.edu; http://ai.arizona.edu.



Fei-Yue Wang is a professor in the University of Arizona's Systems & Industrial Engineering Department and director of the University's Program for Advanced Research for Complex Systems. He is also director of the Intelligent Control and Systems Engineering Center at the Chinese Academy of Sciences' Institute of Automation. His research interests include intelligent control, computational intelligence, and complex systems. He received his PhD in computer and systems engineering from Rensselaer Polytechnic Institute. Contact him at the Systems & Industrial Eng. Dept., Univ. of Arizona, Tucson, AZ 85721; feiyue@sie.arizona.edu.

Acknowledgments

We would like to acknowledge the US National Science Foundation, Intelligence Technology Innovation Center/Central Intelligence Agency, Department of Homeland Security, and Department of Justice/National Institute of Justice for sponsoring the Intelligence and Security Informatics conferences, which have helped foster and build a new community of ISI researchers.

References

1. *National Strategy for Homeland Security*, Office of Homeland Security, 2002.
2. National Research Council, *Making the Nation Safer: The Role of Science and Technology in Countering Terrorism*, Nat'l Academy Press, 2002.
3. H. Chen, *Intelligence and Security Informatics for National Security: Information Sharing and Data Mining*, to be published, Springer, 2005.
4. M. Sageman, *Understanding Terror Networks*, Univ. Pennsylvania Press, 2004.
5. G. Wang, H. Chen, and H. Atabakhsh, "Automatically Detecting Deceptive Criminal Identities," *Comm. ACM*, vol. 47, no. 3, 2004, pp. 71–76.
6. W. Lee and S. Stolfo, "Data Mining Approaches for Intrusion Detection," *Proc. 7th Usenix Security Symp.*, 1998, Usenix Assoc.; www.usenix.org/publications/library/proceedings/sec98/lee.html.
7. L. Damianos et al., "MiTAP for Biosecurity: A Case Study, *AI Magazine*, vol. 23, no. 4, 2002, pp. 13–29.
8. Q. Lu, Y. Huang, and S. Shekhar, "Evacuation Planning: A Capacity Constrained Routing Approach," *Proc. First NSF/NIJ Symp. Intelligence and Security Informatics* (ISI 03), Springer, 2003, pp. 111–125.
9. E.H. Shortliffe et al., *Medical Informatics: Computer Applications in Health Care and Biomedicine*, Springer, 2004.
10. H. Chen and J. Xu, "Intelligence and Security Informatics for National Security: A Knowledge Discovery Perspective," *Ann. Rev. Information Science and Technology* (ARIST), vol. 40, Information Today, 2005; www.asis.org/Publications/ARIST/vol40.html.
11. H. Chen et al., eds., *Medical Informatics: Knowledge Management and Data Mining in Biomedicine*, Springer, 2005.
12. U.M. Fayyad and R. Uthurusamy, "Evolving Data Mining into Solutions for Insights," *Comm. ACM*, vol. 45, no. 8, 2002, pp. 28–31.
13. J.S. Cook and L.L. Cook, "Social, Ethical, and Legal Issues of Data Mining," *Data Mining: Opportunities and Challenges*, J. Wang, ed., Idea Group Publishing, 2003, pp. 395–420.
14. R. O'Harrow, *No Place to Hide*, Free Press, 2005.
15. American Civil Liberties Union, "Matrix: Myths and Reality," white paper, ACLU, 2004; www.aclu.org/Privacy/Privacy.cfm?ID=14894&c=130.