

# Intelligence and Security Informatics for Homeland Security: Information, Communication, and Transportation

Hsinchun Chen, *Senior Member, IEEE*, Fei-Yue Wang, *Fellow, IEEE*, and Daniel Zeng, *Member, IEEE*

**Abstract**—Intelligence and security informatics (ISI) is an emerging field of study aimed at developing advanced information technologies, systems, algorithms, and databases for national- and homeland-security-related applications, through an integrated technological, organizational, and policy-based approach. This paper summarizes the broad application and policy context for this emerging field. Three detailed case studies are presented to illustrate several key ISI research areas, including cross-jurisdiction information sharing; terrorism information collection, analysis, and visualization; and “smart-border” and bioterrorism applications. A specific emphasis of this paper is to note various homeland-security-related applications that have direct relevance to transportation researchers and to advocate security informatics studies that tightly integrate transportation research and information technologies.

**Index Terms**—Border and transportation security, homeland security, infectious disease informatics, intelligence and security informatics.

## I. INTRODUCTION

### A. National Strategy for Homeland Security and the Role of Science and Technology

In response to the September 11, 2001 terrorist attacks, federal, state, and local governments; legislative bodies; the private sector; and private citizens across the nation have all been mobilized at an unprecedented scale to contribute to homeland security. The scientific and engineering research communities are no exception and have been called upon to play an important role in this national effort.

Manuscript received December 1, 2003; revised July 16, 2004 and July 31, 2004. This work was supported in part by the National Science Foundation under Grants EIA-9983304, EIA-0326348, and KDD 9983304; by the Department of Homeland Security and Corporation for National Research Initiatives (CNRI) under the “Border Safe” initiative; by the Outstanding Young Scientist Research Program under Grant 60125310; by the Key Project on Networked Systems under Grant 60334020; by the National Natural Science Foundation; by a 973 Project under Grant 2002CB312200 from the Ministry of Science and Technology; by a Shandong 863 Project under Grant 030335 from Shandong Provincial Government; and by Grant #ORP-0303 for open research projects from the Key Laboratory of Complex Systems and Intelligence Science, Chinese Academy of Sciences, China. The Associate Editor for this paper was S. Tang.

H. Chen is with the Management Information Systems Department, University of Arizona, Tucson, AZ 85721 USA (e-mail: hchen@eller.arizona.edu).

F.-Y. Wang is with the Key Laboratory of Complex Systems and Intelligent Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China, and also with the Systems and Industrial Engineering Department, University of Arizona, Tucson, AZ 85721 USA (e-mail: feiyue@email.arizona.edu).

D. Zeng is with the Management Information Systems Department, University of Arizona, Tucson, AZ 85721 USA (e-mail: zeng@eller.arizona.edu).

Digital Object Identifier 10.1109/TITS.2004.837824

The federal government defines homeland security as “a concerted national effort to prevent terrorist attacks within the United States, reduce America’s vulnerability to terrorism, and minimize the damage and recover from attacks that do occur” [1, p. 2]. In 2002, two important reports were published to provide strategic directions for homeland-security-related efforts. The first report is entitled “National Strategy for Homeland Security” and was issued by the Office of Homeland Security at the White House in July 2002 [1]. It aims to establish strategic objectives around which “the entire society can mobilize to secure U.S. homeland from the dangerous and evolving threat of terrorism” [1, p. 4]. It aligns and focuses homeland security functions into several critical mission areas and identifies foundations that cut across all these mission areas. These foundations represent main sources of strength to draw upon in supporting and enabling terrorism-fighting activities. The second report is entitled “Making the Nation Safer: The Role of Science and Technology in Countering Terrorism” and was published by the U.S. National Research Council [2]. Developed by leading scientific, engineering, medical, and policy experts in the U.S., this report articulates the commitments of the scientific, engineering, and health communities to helping the U.S. and the world respond to security challenges and identifies critical national- and homeland-security research areas in specific terms.

The purpose of this paper is to introduce some of the emerging national- and homeland-security research from information technology (IT) and transportation systems and technology (TST) perspectives. We coined the term *intelligence and security informatics* (ISI) to refer to such emerging research [3], [4]. Before presenting our definition of ISI and its main technical components and research areas, we discuss the broad application and policy context of ISI studies based on the homeland-security research and operations objectives and frameworks laid out in [1] and [2], with specific emphasis on IT- and TST-related issues and challenges.

Six mission-critical areas are identified in [1]. For almost all of them, IT and TST research can be highly relevant and become part of the integrated solution to address corresponding challenges arising in each of these respective areas.

- 1) *Intelligence and Warning*: IT researchers can help build new information- and intelligence-gathering and analysis capabilities to detect future terrorist activities.
- 2) *Border and Transportation Security*: IT and TST researchers can help to develop identity-management and deception-detection techniques for creating “smart

borders.” In addition, TST and IT researchers can help to enhance the security of international shipping and develop strategic and operational models for effective and efficient transportation-related infrastructure protection policies.

- 3) *Domestic Counterterrorism*: IT researchers can help to improve information access and sharing, as well as the crime-analysis abilities of local, state, and federal law-enforcement officers.
- 4) *Critical Infrastructure and Key Assets*: TST and IT researchers can help to develop analytic, modeling, and simulation tools for critical infrastructure (including transportation systems), cyberspace vulnerability and risk analysis, and protection.
- 5) *Defending against Catastrophic Threats*: IT researchers can help to develop simulation, detection, and alerting techniques for potential catastrophic threats, such as chemical and biological attacks. TST researchers can also assist in such efforts by focusing on the necessary transportation aspect of carrying out such attacks.
- 6) *Emergency Preparedness and Response*: IT researchers can help to improve information sharing and communication interoperability for the first responders before and during emergencies. TST researchers can contribute by providing logistics decision-aiding systems to improve the operational efficiency of responses.

The “National Strategy for Homeland Security” report also describes four foundations that enable and support terrorism-fighting activities: law; science and technology; information sharing and systems; and international collaboration. New laws and policies to facilitate interagency and international information sharing and collaboration have continued to develop and evolve. It is not surprising to see that science and technology (including TST) and information sharing and systems (including IT as its technical component) are viewed as two major advantages that the U.S. possesses to counter terrorism.

The “Making the Nation Safer” report notes that “the vulnerability of societies to terrorist attacks results in part from the proliferation of chemical, biological, and nuclear weapons of mass destruction, but it also is a consequence of the highly efficient and interconnected systems that we rely on for key services such as transportation, information, energy, and health care. The efficient functioning of these systems reflects great technological achievements of the past century, but interconnectedness within and across systems also means that infrastructures are vulnerable to local disruptions, which could lead to widespread or catastrophic failures. As terrorists seek to exploit these vulnerabilities, it is fitting that we harness the nation’s exceptional scientific and technological capabilities to counter terrorist threats” [2, p. 1]. This report focuses on eight research areas that are of critical relevance to homeland security: 1) biological sciences; 2) chemical issues; 3) nuclear and radiological issues; 4) information technology; 5) transportation; 6) energy facilities, cities, and fixed infrastructure; 7) behavioral, social, and institutional issues; and 8) systems analysis and systems engineering.

The discussions on IT [item 4)] and TST [item 5)] are particularly relevant to this paper. For IT, the report recommends three primary counterterrorism-related areas: 1) information and net-

work security; 2) the IT needs of emergency responders; and 3) information fusion and management. Information and network security research develops approaches and architectures for prevention, identification, and containment of cyberintrusions and recovery from them. The research to address the IT needs of emergency responders is expected to deal with issues such as ensuring interoperability, maintaining and expanding communications capabilities during an emergency, communicating with the public during an emergency, and providing support for decision makers. Information fusion and management research for the intelligence, law-enforcement, and emergency-response communities includes data mining, data integration, language technologies, and processing of image and audio data.

As to TST, recognizing transportation systems’ common characteristics, such as openness and accessibility; entwinement in society and the global economy; diversity of owners, operators, users, and overseers; and the fact that transportation systems are recurrent targets of terrorist attacks as well as tempting weapons themselves, the report advocates a layered security-system approach, long used to secure communications and information systems, to protect transportation security. Such an approach relies on multiple interconnected security features that provide backup for one another. From a research perspective, the report identifies the critical needs for research in areas such as operations (e.g., understanding normal patterns of transportation activity and behavior; identification of anomalous and suspect activities) and human factors (e.g., design of security devices and facilities, procedures that are efficient and reliable, understanding how to obscure the risk of getting caught, and understanding how technology can complement and supplement humans). The significant overlaps between some of these TST challenges and the IT issues discussed previously are obvious.

### B. Developing the Science of ISI

A running theme in many of the homeland-security-related applications and challenges discussed in the previous section is concerned with information processing. Similar to medical and biological research, law-enforcement, criminal-analysis, and intelligence communities face significant information overload and yet also have tremendous opportunities for innovation. We believe that, as in “medical informatics” and “bioinformatics,” there is a pressing need to develop the science of ISI—the study of the use and development of advanced information technologies, systems, algorithms, and databases for national- and homeland-security-related applications through an integrated technological, organizational, and policy-based approach.

Many existing computer and information systems techniques need to be re-examined and adapted for homeland-security applications. New insights from this unique domain could result in significant breakthroughs in data mining, visualization, knowledge management, and information security techniques and systems. For example, social network analysis technologies and methodologies could be adopted to help the intelligence community to detect planned future attacks and to uncover and understand Bin Laden’s terrorist networks. Visual data-mining techniques, such as association rules and multidimensional information visualization, could be used to identify criminal

relationships. Record linkage and string comparator algorithms could be useful for criminal identity-deception detection.

Practical and novel information technologies, techniques, methods, practices, and systems that can contribute to knowledge in this important emerging field are critically needed, including but not limited to areas such as

- information sharing and system interoperability;
- knowledge discovery and knowledge management;
- criminal data mining, social network analysis, and event detection;
- multimedia and multilingual intelligence and security information analysis;
- Web-based intelligence monitoring and analysis;
- deception detection systems;
- intrusion detection systems and information awareness;
- cybercrime detection and analysis;
- agents and collaborative systems for intelligence sharing;
- crime and intelligence visualization;
- bioterrorism tracking, alerting, and analysis;
- major disaster prevention, detection, and management (including related intelligent transportation systems applications and route planning).

Academic meetings have begun to emerge to encourage research and discussions among previous disparate research and practitioner communities in ISI-related areas. The symposium series on Intelligence and Security Informatics (<http://ecom.arizona.edu/ISI>) is such an example. The proceedings of these symposia and several special issues on ISI in academic journals also provide snapshots of the current state of the art in ISI research [3]–[6]. ISI-related research also is the focus of several federal programs, including the Knowledge Discovery and Dissemination Program, jointly managed by the National Science Foundation (NSF) and the Intelligence Technology Innovation Center, and the Information Technology Research program/National and Homeland Security priority area, funded by the NSF.

### C. Paper Organization

In this introduction section, we summarized the needs of conducting IT and TST research in the context of national and homeland security. We also provided a general definition of ISI, an emerging field of study with significant practical relevance, and discussed sample research topics in this field.

In the rest of this paper, we will present three case studies conducted at the University of Arizona, Tuscon, that aim to develop and apply ISI techniques and methods to meet several of the pressing national- and homeland-security challenges. These case studies were chosen to cover homeland-security applications of different types and to present the readers with a variety of ISI research methodologies and technical approaches. We hope that this “sampling” approach will help the readers to appreciate the scope of ISI research and the diversity of homeland-security application settings. Each of these case studies represents a major mission critical area discussed in [1].

In the first case study, presented in Section II, we describe the BorderSafe project, an information integration and cross-jurisdictional criminal network analysis study aimed to help create

a “smart” and safe border [7]. The second case study, presented in Section III, summarizes the Dark Web portal project that is motivated to facilitate counter-terrorism intelligence gathering and analysis [8]. In the third and last case study reported in Section IV, we present our design and prototype implementation of a national infectious-disease information infrastructure, called the West Nile Virus and Botulism Portal, which can help to detect and respond to bioterrorism threats [9].

We conclude this paper in Section V by highlighting future ISI research directions, especially those involving integration between IT and TST research.

## II. BORDERSAFE PROJECT: BORDER AND TRANSPORTATION SAFETY

### A. Project Background

The Department of Homeland Security (DHS) has been called upon to provide accurate information on inbound goods and passengers to support risk-based management tools for various border-management agencies [1]. For instance, as vehicles enter the country, DHS records each license plate with a crossing date and time. Agents also search vehicles for drugs and other contraband goods and related records are kept by DHS. The benefits of triangulating such border-crossing data with local law-enforcement records are obvious. However, integrating data from law-enforcement and border-crossing sources poses many challenges.

The BorderSafe project aims to develop a framework to meet such cross-jurisdictional information integration needs. In addition, we are interested in identifying investigative leads in support of border transportation and security by analyzing known relationships between known criminals, vehicles, criminal incidents, and border-crossing activities through the use of criminal-activity networks (CANs).

### B. Cross-Jurisdictional Information Integration and Analysis Framework

The key to a cross-jurisdictional information integration and CAN analysis framework is identification of three classes of data: 1) *base data* with overlapping information from multiple jurisdictions with multiple object and relation types; 2) high volume but relatively simple *supplementary data* to enhance CAN information content; and 3) case-specific or *ad hoc* query-specific data expressing important relationships or features. Given these classes of data, integration can proceed in three steps: schema-level transformation of base data; entity-matching to align objects across data sets; and normalization and matching of supplementary data.

In our proposed information-integration and CAN-analysis framework, base data should be semantically aligned and mapped to support CAN generation following the classic data-integration techniques that require reconciliation of legacy data into a common schema and instance-level entity matching [10], [11]. Police records are a prime example of this kind of data, because multiple jurisdictions keep similar types of information for a common set of objects (e.g., vehicles, people, and incidents).

Entity matching at the instance level in this ISI context is expected to heavily rely on heuristics. Primary objects will include people, locations, and vehicles. More research into appropriate identity-matching algorithms for cross-jurisdictional data sets is needed [12], [13]. For instance, input from domain experts suggests an initial match for people using first name, last name, and date of birth. These heuristics are, however, not perfect. Other alternatives, such as FBI and state numbers, may be useful, but are not consistently available. Locations can be matched based on geo-codes and vehicles can be matched by license plate and vehicle-identification number. In effect, license-plate data have some interesting and useful characteristics. Plate numbers can be recorded in an unobtrusive fashion and, while criminals frequently try to avoid identification by lying about their names in routine interactions with law-enforcement officials, license-plate numbers can be directly observed. In addition, vehicles used by criminals are often registered in someone else's name. Even if a criminal uses an alias in incidents involving a particular vehicle, the resulting person-vehicle data can implicitly link the incidents.

In addition to the base data, investigators use many additional supplementary or query-specific information resources to identify criminals' activities and associations. Such additional data, however, may not be readily available for a variety of reasons. Frequently, not all useful data are directly accounted for in the global schema. For example, police systems do not usually store border-crossing events. Also, information such as jail visitation histories is important and could be included, but typically has not been included in a police agency's data system. Furthermore, some information that is of a sensitive nature or provides useful contextual backgrounds is often not available (or cannot be made accessible) through police systems because of privacy and security concerns.

Our proposed framework allows for the inclusion of this kind of data by treating it as supplementary or as query specific. A data source is appropriate for supplementary integration when: 1) it is available in quantity and can be appropriately organized; 2) its sensitivity level allows it to be shared across multiple investigations; and 3) it is contextually appropriate outside of a single investigation. Data can be used as supplementary data if it can be reduced to one or more lists of features or events directly associated with identifiable objects in the base data set. For example, mug shots of people, border-crossing records, or jail visitations can all be recorded associated with particular individuals already contained in a base data set of criminal incidents. Query-specific data can also be used to guide CAN analysis. For example, if phone records indicate that a suspect called 19 different people, a CAN could query for relationships involving any of these 20 people to arrive at a context-specific result without storing subpoenaed data in the general investigation data set. Both supplementary and query-specific data have to be normalized and matched to the objects and entities from the base data.

### C. Research Test Bed: BorderSafe

The BorderSafe project is a collaborative effort involving the University of Arizona's Artificial Intelligence (AI) Laboratory; law-enforcement agencies including the Tucson Police Department (TPD), Phoenix Police Department (PPD), Pima County

TABLE I  
KEY STATISTICS ON THE TPD AND PCSD DATA SETS

	TPD	PCSD
Recorded Incidents	2.84 million	2.18 million
Persons	1.35 million	1.31 million
Vehicles	623,656	520,539

TABLE II  
CBP BORDER-CROSSING DATA

1,125,155	Records: plate, state, date, & time
226,207	Distinct vehicles
209	Days of information out of 18 months
130,195	Plates issued in AZ
5,546	Plates issued in CA
90,466	Plates issued in Mexico

TABLE III  
BORDER-CROSSING LICENSE PLATES BY DATA SET

	TPD	PCSD	Combined
Border Crossing Plates found in Police Records	8,300	6,619	13,111
Crossings Associated with Those Vehicles	34,632	31,075	59,275

Sheriff's Department (PCSD), and Tucson Customs and Border Protection (CBP), as well as San Diego Automated Regional Justice Information Systems (ARJIS) and the San Diego Super Computer Center (SDSC). We briefly present one cross-jurisdictional data-sharing initiative within the BorderSafe project.

In this initiative, we integrated TPD, PCSD, and CBP data sets. Table I summarizes the TPD and PCSD data sets, covering adjacent areas in and around the city of Tucson, AZ, with a combined population approaching 1 000 000. The CBP data set, summarized in Table II, was handled as a supplementary source that identifies border-crossing vehicles. Video equipment automatically extracts license-plate numbers of cars as they cross into or out of the country. Camera errors are manually corrected as incoming vehicles pass through the port of entry.

Integration of these three cross-jurisdictional data sets proceeded in three steps: 1) TBP records were mapped to a common schema; 2) cross-jurisdictional identities were matched; and 3) CBP data were imported as a supplementary source.

We began analyzing these integrated data sets by evaluating the overlaps between them. In one such analysis, we tallied the number of out-of-state vehicles in the TPD records, finding that 7% of the vehicles involved in gang-related, violent, and narcotics crimes are registered outside of Arizona. Table III shows the number of license plates found in both the CBP crossing data and TPD/PCSD data sets, confirming that a lot of cross-border activities can be identified with vehicles connected to crime incidents.

### D. Criminal Activity Network Analysis

Several systems currently support the visualization of criminal activity using a network-based representation. Analyst's Notebook has been widely employed by law-enforcement officers in the U.S. and The Netherlands [14]. Notebook relies on a human analyst to detect criminal relationships in data, then

TABLE IV  
AVERAGE NUMBER OF ASSOCIATED PEOPLE, VEHICLES,  
AND LINKS FOR 50 SELECTED INDIVIDUALS

	People	Vehicles	Associations
TPD associations only	193.3	1.44	659.92
PCSD associations only	120.36	1.08	1,016.14
All associations	389.92	6.92	2,487.14

generates a link chart based on manual input or relational data stored in a spreadsheet or text file. Similarly, the Netmap system provides network visualization functionality by laying out entities of various types on the perimeter of a circle and placing straight lines between entities to represent links. Netmap has been adopted in the FinCEN system by the U.S. Department of the Treasury to analyze patterns of financial-transactional data to detect money laundering [15]. The Watson system can search and identify possible associations between persons by querying databases [16]. Given a person’s name, Watson can automatically form a database query to search for related persons and present the results in a link chart.

These systems focus on presenting data in a network while our privacy-sensitive approach to CAN analysis and visualization emphasizes domain-appropriate interaction between a user and a visualization tool. We argue that an interactive tool allowing for inclusion of base, supplementary, and case-specific sources would be helpful to law-enforcement practitioners, given the nature of available data. We now briefly summarize a case study motivated to evaluate the usefulness of cross-jurisdictional information on CAN analysis. First, we randomly chose 50 people from a combined list of wanted suspects and known drug traffickers. We only selected people that appear in both TPD and PCSD records. We then used the associations or links that occur when individuals or vehicles are listed together in an incident report to perform CAN analysis.

For each person, we followed all known person-to-person associations and compiled a list of connected people. Links for each person in this list were also followed by using the same approach. We then followed person-to-vehicle links to identify plate numbers. The result was a network of all people within two “hops” of the individual under study and all associated vehicles known to have cross-border activities. We created three networks for each person: one with links from the TPD data set; one with links found in the PCSD data set; and one using the links in both data sets. Table IV reports the average number of associated people, associated vehicles, and associational links found for the 50 selected individuals. It is not surprising that combining the data sets allowed us to connect people and border-crossing vehicles for this list of known criminals.

Based on these identified associations, we then created a set of CAN visualizations that was intended to be used by law-enforcement officers to pursue investigative leads. To reduce users’ potential cognitive overload, a variety of visual cues were used in our preliminary implementation. For instance, we differentiated entity types by shape, key attributes by node color, degree of activity as node size, connection source by link color, and some details in link text or rollover tool tips. Fig. 1 shows a network connecting narcotics traffickers and border-crossing license plates.

### III. DARK WEB PORTAL PROJECT: INTELLIGENCE AND WARNING

#### A. Project Background

Terrorists across different jurisdictions heavily utilize modern transportation and communication systems for relocation, propaganda, recruitment, and communication purposes. Thus, addressing issues such as how to trace the dynamic evolution, communication, and movement of terrorist groups across different jurisdictions and how to analyze and predict terrorists’ activities, associations, and threats becomes an urgent and challenging issue.

While the Web has brought about tremendous benefits to mankind, many terrorists, extremist groups, hate groups, and racial-supremacy groups are also using Web sites and other online tools. We refer to the part of the Web used for such illegitimate and malicious purposes as the *Dark Web*.

In traditional terrorism research, the major data sources used by researchers are limited to news stories, journal articles, books, and media-derived databases. The data-analysis methods are primarily limited to manual approaches [17], [18]. The problem of lacking reliable data sources and advanced analysis methodologies has been hindering terrorism researchers from producing more research of genuine explanatory and predictive value [18]. To address this challenge, many terrorism-specific knowledge portals have been built by specialized research centers such as the Center for the Study of Terrorism and Political Violence, located at St. Andrews University, Scotland. However, Kennedy and Lunn’s study [19] on 28 different sources of terrorism databases and archives showed that none of these existing portals provided Dark Web information. There has been no general methodology for collecting and analyzing Dark Web information.

Dark Web contents could be analyzed to enable a better understanding and analysis of terrorist activities. However, there are several problems that prevent the effective and efficient discovery of Dark Web intelligence. The amount of data available on the Web is often overwhelming to counterterrorism experts, a problem commonly referred to as *information overload*. Also, data posted on the Dark Web are typically not persistent and sometimes are intentionally misleading. The language barrier is another problem faced by the counterterrorism experts when dealing with the multilingual information generated by terrorists from different parts of the world.

In our recent research, we have been developing a Web-based counterterrorism knowledge portal, called the Dark Web Portal, to support the discovery and analysis of Dark Web information. Specifically, the Dark Web Portal integrates terrorist-generated multilingual data sets on the Web and uses them to study advanced and new methodologies for predictive modeling, terrorist network analysis, and visualization of terrorists’ activities, linkages, and relationships.

Fig. 2 summarizes the three major components of our Dark Web Portal project: Dark Web test-bed building; Dark Web link analysis and social network analysis; and Dark Web Portal building.

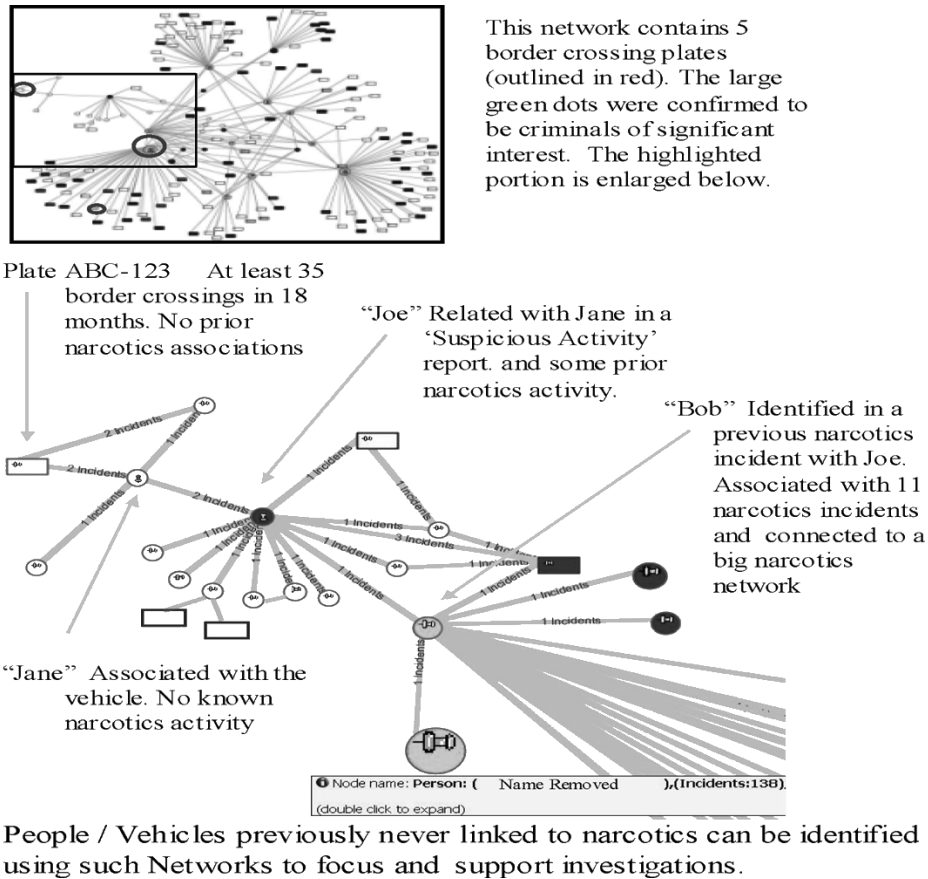


Fig. 1. CAN visual analysis: border crossings and narcotics.

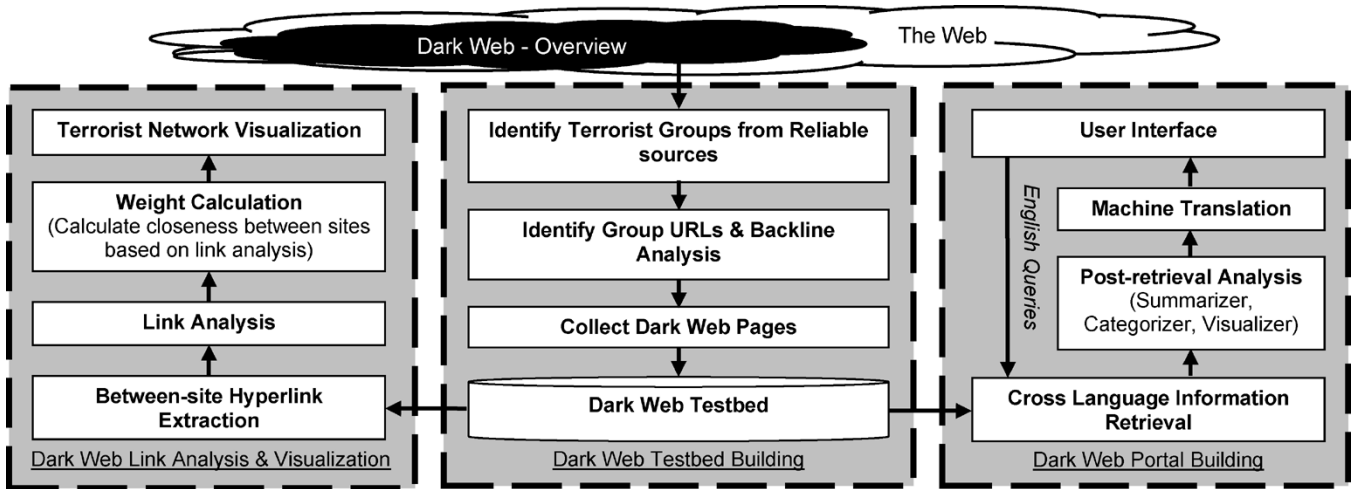


Fig. 2. Overview of the Dark Web Portal project.

**B. Dark Web Test-Bed Building**

The goal of the Dark Web test-bed-building effort is to build a high-quality up-to-date test bed that contains multilingual information created by major terrorist groups around the world. A two-step process was followed in this effort.

1) *Identify Terrorist Groups:* We started the test-bed-building process by identifying the groups that are considered by reliable sources to be terrorist groups. The main sources

we used to identify U.S. domestic terrorist groups include the Anti-Defamation League (ADL), Federal Bureau of Investigations (FBI), Southern Poverty Law Center (SPLC), Militia Watchdog (MD), Google Directory (GD), and Hate Directory (HD). To identify international terrorist groups, we relied on the following sources: the United States Committee For A Free Lebanon (USCFAFL); Counter-Terrorism Committee (CTC) of the UN Security Council (UN); U.S. State Department report (U.S.); *Official Journal of the European Union* (EU); and

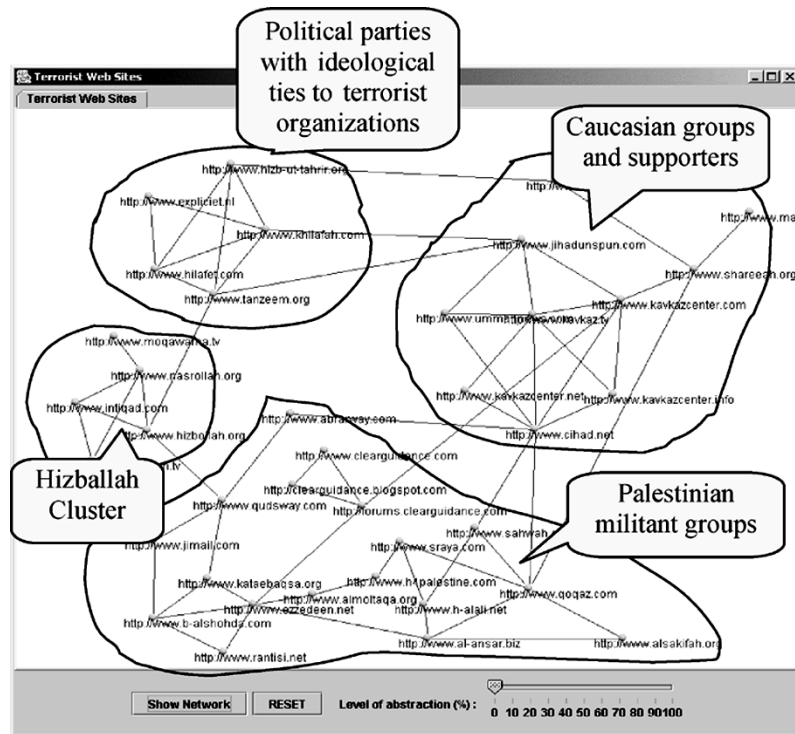


Fig. 3. Terrorist-group relationship network diagram generated based on the hyperlink structure between their group Web sites.

government reports from the United Kingdom (U.K.), Australia (AUS), Japan (JPN), and P.R. China (CHN). These sources have been identified following the recommendations of core terrorism authors. A total of 224 US domestic terrorist groups and 440 international terrorist groups have been identified.

2) *Identify Terrorist Group Uniform Resource Locators (URLs) and Collect Terrorist-Generated Web Pages:* We manually identified the URLs created by the terrorist groups. For U.S. domestic groups, group-generated URLs can be found in FBI reports and the GD. For international groups, we used the group names as query terms to search major search engines such as Google and manually identified the group-created URLs from the result lists. To ensure that our test bed covers all the major regions in the world, we sought the assistance of language experts in English, Arabic, Spanish, Japanese, and Chinese to help us collect URLs in different regions. All URLs collected were manually checked by experts to make sure that they were created by terrorist groups. After the URL of a group is identified, we used the SpidersRU's toolkit, a multilingual digital-library-building tool developed by our own group, to collect all the Web pages under that URL and store them into our test bed.

To maintain the freshness of data, we are repeating the above process of identifying group URLs and collecting terrorist-generated Web pages on a monthly basis. Also, the information collected during different time periods can be analyzed to study the dynamic evolution of the terrorist groups over time. So far, we have collected 500 000 Web pages created by 94 U.S. domestic groups, 400 000 Web pages created by 41 Arabic-speaking groups, 100 000 Web pages created by Spanish-speaking groups, and 2200 Web pages created by Japanese-speaking groups. This test-bed-building effort is ongoing.

### C. Link and Network Analysis

Terrorist groups consist of actors linked to each other through complex networks of direct or mediated exchanges [20]. Identifying how relationships between groups are formed and dissolved in the terrorist-group network would enable us to decipher the social milieu and communication channels among terrorist groups across different jurisdictions. Previous studies have shown that the link structure of the Web represents a considerable amount of latent human annotation [21]. Thus, by analyzing and visualizing hyperlink structures between terrorist-generated Web sites in our test bed, we could discover the structure and organization of terrorist group networks, capture network dynamics, and predict trends of terrorist attacks.

To test our ideas, we conducted an experiment in which we analyzed and visualized the hyperlink structure between approximately 100 000 Web pages from 46 Web sites in our current test bed. These 46 Web sites were created by four major Arabic-speaking terrorist groups, including Al-Gama'a al-Islamiyya (Islamic Group, IG), Hizballa (Party of God), Al-Jihad (Egyptian Islamic Jihad), Palestinian Islamic Jihad (PIJ), and their supporters. Hyperlinks between each pair of the 46 Web sites were extracted from the Web pages and a closeness value was calculated for each pair of the 46 Web sites. Two Web sites were considered to be closer to each other when there are more hyperlinks between them. A terrorist-group relationship network diagram was generated based on the closeness values between the group Web sites using multidimensional scaling algorithm (see Fig. 3).

We presented the network illustrated in Fig. 3 to the domain experts and confirmed that the structure of the diagram matches

the knowledge of the experts as to how the groups relate to each other. For instance, domain experts identified four main clusters: the Palestinian Organizations cluster; Caucasian Jihad groups and their supporters' cluster; the Hizballah cluster; and a fourth cluster containing Web sites of political parties that have ideological connections with other terrorist groups. These clusters represent a logical mapping of the existing relations between the 46 groups. An example of this kind of mapping is the Palestinian terrorist group's cluster, where many of the Palestinian terrorist groups' Web sites, as well as their leaders' Web sites, are clustered together.

The use of link analysis and visualization on the Dark Web data clearly provides significant value-added services. First, the network-based diagram helps experts to easily identify patterns in terrorist-group relationships. Closely related terrorist groups can be easily identified as a cluster of nodes that are close to each other. Relationships between any two groups can be easily tracked by following the links in the diagram. Second, the network can provide new insights as to relations between terrorist organizations. For instance, the link between the Hizballah cluster and the Palestinian cluster shown in Fig. 3 suggests some kind of connection between the two entities that has been overlooked by the domain experts.

#### D. Dark Web Portal Building

We are currently developing an intelligent Web portal system to support user-friendly access to Dark Web information and to provide advanced analysis and visualization functions. In this section, we briefly survey some of the IT research that is relevant to this ongoing portal-building effort.

To address the information-overload problem, the Dark Web Portal will be fitted with post-retrieval components. A modified version of a text summarizer called TXTRACTOR, which uses sentence-selection heuristics to rank text segments [22], will be added into the Dark Web Portal. This summarizer can flexibly summarize Web pages using three or five sentence(s) such that experts can quickly get the main idea of a Web page without having to read through it.

A categorizer that organizes the search results into various folders labeled by the key phrases extracted by the Arizona Noun Phraser (AZNP) [23] from the page summaries or titles will also be integrated into the Dark Web Portal system to facilitate the understanding of different groups of Web pages.

In addition, in the Portal we are planning to include a visualizer that clusters Web pages into colored regions using the Kohonen self-organizing map (SOM) algorithm [24], further reducing the information-overload problem when a large number of search results are obtained.

To address the language-barrier problem, we plan to add a dictionary-based cross-language information retrieval (CLIR) component into the portal [25]. Our initial implementation of this component can take English queries and retrieve documents in English, Spanish, Chinese, Japanese, and Arabic. Another component that we will add to the Dark Web Portal is a Machine Translation component, which will translate the multilingual information retrieved by the CLIR component back into the experts' native languages.

## IV. WEST NILE VIRUS–BOTULISM (WNV–BOT) PORTAL PROJECT: DEFENDING AGAINST CATASTROPHIC THREATS

### A. Project Background

Infectious disease outbreaks are critical threats to public health and national security [26], [27]. With greatly expanded trade and travel, infectious diseases, either naturally occurring or caused by biological terror attacks, can spread at a fast pace within and across country borders, resulting in potentially significant loss of life, major economic crises, and political instability.

Currently, a large amount of infectious-disease data is being collected by various laboratories; health-care providers; and government agencies at the local, state, national, and international levels. However, there exist a number of technical and policy-related challenges that hinder the effective use and sharing of infectious-disease data, especially data sets across species and across jurisdictions, in regional, national, and global contexts [28]. Existing infectious-disease information systems do not fully interoperate. The information-management environment used to analyze large amounts of infectious disease data and develop predictive models needs major improvements. In addition, an efficient reporting and alerting mechanism across organizational boundaries within and beyond the public health systems (including law enforcement and first responders) is lacking. Furthermore, data ownership, confidentiality, security, and other legal and policy-related issues need to be closely examined.

The WNV–BOT Portal project is a collaborative effort to meet some of these challenges. Jointly conducted by the University of Arizona AI Laboratory, the New York State Department of Health and its partner Health Research, Inc., and the California State Department of Health Services and its partner PHFE Management solutions, this project is aimed at developing scalable technologies and related standards and protocols that are needed by the full implementation of the national infectious-disease information infrastructure and at studying related policy issues. Our project has been focused on two prominent infectious diseases: West Nile Virus (WNV) and Botulism (BOT). These two diseases were chosen as our first target because of their significant public health and homeland-security implications and the availability of related data sets in both New York and California. In this section, we provide a brief overview of the WNV–BOT Portal system, a completed research prototype that resulted from our project. We also discuss related spatio-temporal hotspot analysis research that has major applications in not only public health and counter-bioterrorism, but also on other ISI applications such as crime analysis.

### B. WNV–BOT Portal System

The WNV–BOT Portal system has been developed to integrate infectious disease data sets on WNV and BOT from New York, California, and several federal data sources. It also provides a set of data-analysis, predictive-modeling, and information-visualization tools that are tailored for these two diseases. Fig. 4 summarizes these data sets and intended users of the WNV–BOT Portal.

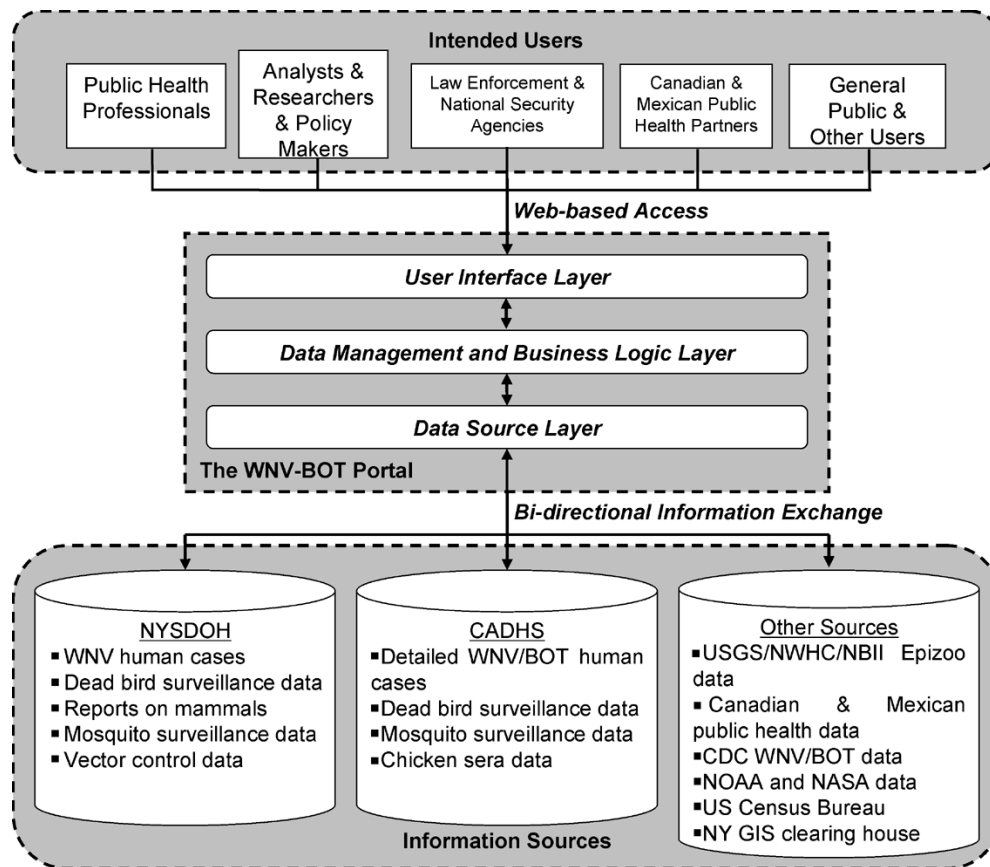


Fig. 4. WNV-BOT Portal system overview.

As illustrated in Fig. 5, from a systems perspective, the WNV-BOT Portal is loosely coupled with the state public health information systems in that the state systems will transmit WNV/BOT information through secure links to the portal system using mutually agreed-upon protocols. The system also automatically retrieves data items from sources such as those from the U.S. Geological Survey (USGS) and stores them in the internal data store.

Architecturally, the WNV-BOT Portal consists of three major components: a communication backbone; a data store; and a Web portal. The communication backbone module implements data-transmission protocols. It normalizes data from various participating sources based on HL7-compliant XML standards ([www.hl7.org](http://www.hl7.org)) and is responsible for converting incoming data into a data format internal to the Portal system. The data-store module receives normalized data from the communication backbone and stores it in a relational database. The data-store module also implements a set of Java application programming interfaces (APIs) that serves data requests from the user and imposes data-access rules. The Web portal module implements the user interface and provides the following main functionalities:

- searching and querying available WNV/BOT data sets;
- visualizing WNV/BOT data sets by using spatial-temporal visualization;
- accessing analysis and prediction functions;
- accessing the alerting mechanism (currently under development).

A typical data query process is comprised of five steps, as follows.

- Step 1) User selects the disease of interest.
- Step 2) User selects interested data sets.
- Step 3) User specifies the time and geographic ranges of interest.
- Step 4) System displays the returned query results.
- Step 5) User uses a visualization tool to summarize and interactively explore the returned results.

The role of visualization techniques in the context of large and complex data-set exploration is to organize and characterize the data visually to assist users in overcoming the information-overload problem. The WNV-BOT Portal makes available an advanced visualization module, called the spatial-temporal visualizer (STV), to facilitate the exploration of infectious disease case data and to summarize query results. STV is a generic visualization environment that can be used to visualize a number of spatial-temporal data sets simultaneously. It allows the user to load and save spatial-temporal data in a dynamic manner for exploration and dissemination. STV has three integrated and synchronized views: periodic, timeline, and geographical information system (GIS). The periodic view provides the user with an intuitive display to identify periodic temporal patterns. The timeline view provides a two-dimensional (2-D) timeline along with a hierarchical display of the data elements organized as a tree. The GIS view displays cases and sightings on a map. Fig. 6 illustrates how these three views can be used to explore infectious-disease data sets.

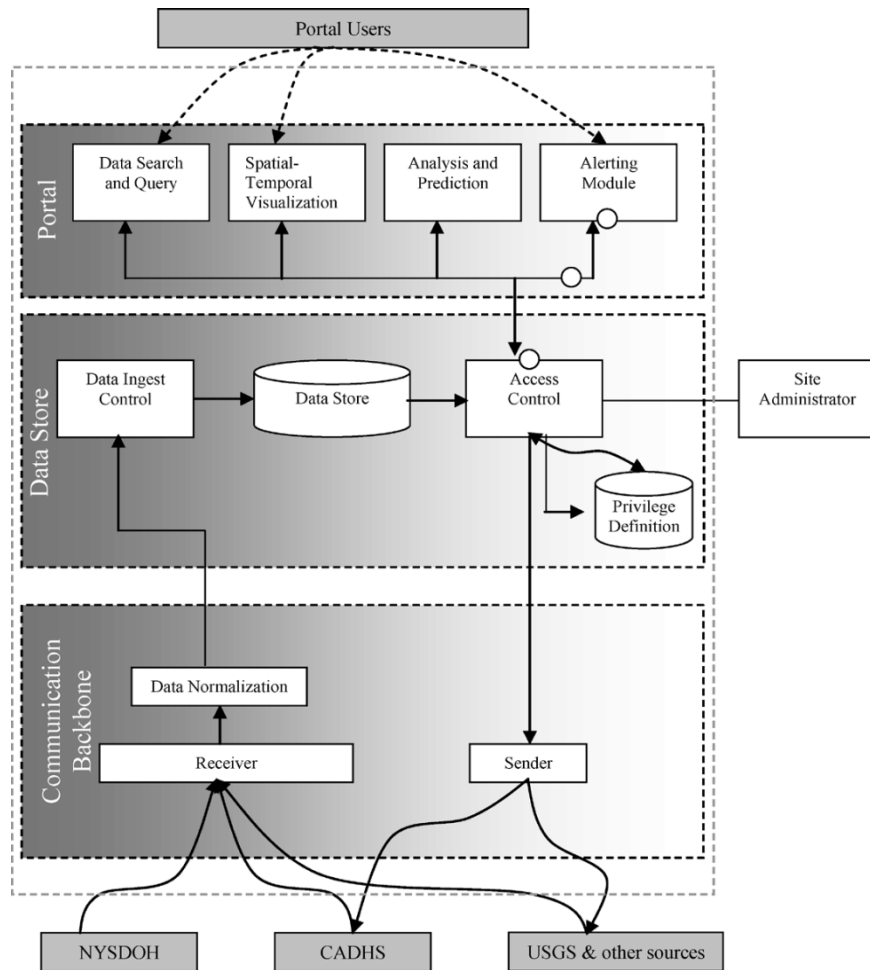


Fig. 5. Overall architecture of the WNV-BOT Portal system.

### C. Spatio-Temporal Hotspot Analysis

Spatio-temporal hotspot analysis is an important tool to facilitate analysis and model building for data with spatial and temporal attributes, which are ubiquitous in many ISI applications [29]. A hotspot is a condition indicating some form of clustering in a spatial and temporal distribution. For WNV, localized clusters of dead birds typically identify high-risk disease areas. Automatic detection of dead-bird clusters using hotspot analysis can help to predict disease outbreaks and to effectively allocate prevention and control resources [30]. In a broader application context, recent years have seen increasing interest in answering the following central questions of great practical importance arising in spatio-temporal data analysis and related predictive modeling.

- How to identify areas having exceptionally high or low measures?
- How to determine whether the unusual measures can be attributed to known random variations or are statistically significant? In the latter case, how to assess the explanatory factors?
- How to identify any statistically significant changes (e.g., in rates of health syndromes or crime occurrences) in a timely manner in geographic areas?

Two types of approaches have been developed in the literature to address some of these questions. The first type falls under the general umbrella of *retrospective* models [31]. It is aimed at statistically testing whether a disease is randomly distributed over space and time for a predefined geographical region during a predetermined time period. The second type of approach is *prospective* in nature, with repeated time-periodic analyses targeted at the identification of statistically significant changes in an online context [32].

Most of the existing disease informatics research uses the spatial scan statistic techniques to perform hotspot analysis [31], [33], [34]. In the WNV-BOT Portal project, we have experimented with other hotspot analysis techniques (e.g., risk-adjusted nearest neighbor hierarchical clustering) that have been developed and successfully applied in crime analysis to disease informatics [35]. We have also developed a new technique based on support-vector clustering [29]. Initial experimental results indicate that these techniques are complementary to the spatial-scan techniques in many regards. In a broader context, we are pursuing research in vector-borne emerging infection predictive modeling. In particular, we are: 1) augmenting existing predictive models by taking additional factors (e.g., weather information, bird migration patterns) into consideration and 2) tai-

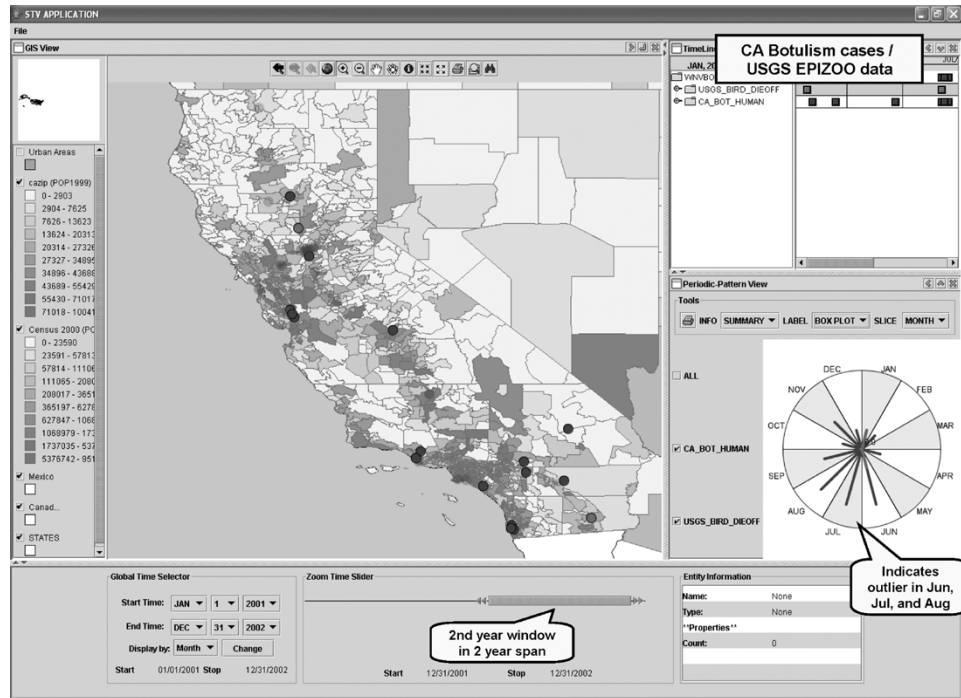


Fig. 6. Visualizing BOT cases using STV.

loring data-mining techniques for infectious disease data sets that have prominent temporal features.

#### D. Ongoing and Future Research

The WNV–BOT Portal project has supported exploration of and experimentation with technological infrastructures needed for the full-fledged implementation of a national infectious-disease information infrastructure and helped to foster information sharing and collaboration among related government agencies at state and federal levels. In addition, we have obtained important insights and hands-on experience with various important policy-related challenges faced by developing a national infrastructure. For example, a nontrivial part of our project activity has been centered around developing data-sharing agreements between project partners from different states.

One of the major foci of our ongoing effort is cross-jurisdictional alert generation and dissemination. For the past decades, alert-dissemination networks such as health-alert network (HAN) systems are being developed and deployed in state public health agencies. However, there is a critical need to create cross-jurisdiction alerts and to automate the dissemination process. We are developing an advanced alerting module as part of the WNV–BOT Portal to complement alerting and surveillance systems that already exist in various states.

We conclude this section by discussing the pathway leading to the national infectious-disease information infrastructure based on the lessons learned from our current project. Due to the complexity of such an infrastructure from both technical and policy standpoints, we envision that its development path will follow a bottom-up evolutionary approach. Initially, each individual state

will develop its own integrated infectious-disease infrastructure for a limited number of diseases. Following the successful deployment of such systems, regional nodes linking neighboring states can be established. Such regional nodes will leverage both state sources and data from federal agencies such as the Centers for Disease Control and Prevention (CDC), USGS, and U.S. Department of Agriculture (USDA). National and international infrastructures will then become a natural extension and integration of these regional nodes, covering most infectious disease types.

#### V. CONCLUSION AND FUTURE DIRECTIONS

The scientific and engineering research communities are playing a significant role in the national effort to fight terrorism and to secure the homeland. ISI is an emerging field of study motivated by national- and homeland-security-related applications. This paper summarizes the broad application and policy context for this emerging field. It focuses on homeland-security research issues from the perspectives of information, communication, and transportation and presents three case studies that illustrate several key ISI research areas.

ISI as a field is still in its early stage of formation and development. We conclude this paper by discussing ISI research topics related to both IT and TST research that promise to lead to fruitful findings and practical applications.

- Communications and information systems protection has received significant attention from the research community. Although protection of such systems is far from a solved problem, much research and systems experience has been accumulated with mature modeling and analysis

support. Transportation-systems protection can greatly benefit from such experience.

- IT and TST researchers share many similar analytical and simulation tools. There exist great opportunities to foster cross-fertilization and to enable complementarity and reusability among these models and tools that were initially developed in different research contexts. One example is the development of a new type of structured search processes for collecting and analyzing information for research in transportation systems [36].
- There is a pressing need to develop models and techniques that consider IT and TST issues in an integrated manner in various ISI settings. For instance, IT infrastructure protection and transportation-systems protection are sometimes intertwined. In another example, both IT and transportation/logistic support issues need to be considered to help first responders to form effective rescue plans. Research at the intersection of IT and TST can be critically relevant and potentially fruitful.

#### ACKNOWLEDGMENT

The authors would like to thank B. Marshall, S. Kaza, J. Xu, H. Atabakhsh, J. Qin, E. Reid, W. Chung, Y. Zhou, W. Xi, G. Lai, T. Elhourani, A. A. Bonillas, C. Tseng, and C. Larson, all from the AI Laboratory, University of Arizona, Tucson; M. Sageman from the University of Pennsylvania, Philadelphia; M. Eidson and I. Gotham from the New York State Department of Health; C. Lynch from the California Department of Health Services; M. Ascher from the Lawrence Livermore National Laboratory; and T. Petersen and C. Violette from the Tucson Police Department for their contributions to the three case studies described in this paper.

#### REFERENCES

- [1] Office of Homeland Security, The White House, National strategy for homeland security, July 2002.
- [2] National Research Council, *Making the Nation Safer: The Role of Science and Technology in Countering Terrorism*. Washington, DC: Committee on Science and Technology for Countering Terrorism, U.S. National Research Council, The National Academies Press, 2002.
- [3] H. Chen, R. Miranda, D. Zeng, T. Madhusudan, C. Demchak, and J. Schroeder, Eds., *Intelligence and Security Informatics. Proceedings of the First Symposium on Intelligence and Security Informatics (ISI'03)*. ser. Lecture Notes in Computer Science (LNCS 2665). New York: Springer-Verlag, 2003.
- [4] H. Chen, R. Moore, D. Zeng, and J. J. Leavitt, Eds., *Intelligence and Security Informatics for National And Homeland Security. Proceedings of the Second Symposium on Intelligence and Security Informatics (ISI'04)*. ser. Lecture Notes in Computer Science (LNCS) 3073. New York: Springer-Verlag, 2004.
- [5] *Decision Support Syst.*, H. Chen, Ed., 2004. Special issue on intelligence and security informatics, to be published.
- [6] *J. Amer. Assoc. Inform. Sci. Technol.*, H. Chen, Ed., 2004. Special issue on intelligence and security informatics: A system perspective, to be published.
- [7] B. Marshall, S. Kaza, J. Xu, H. Atabakhsh, T. Petersen, C. Violette, and H. Chen, "Cross-jurisdictional criminal activity networks to support border and transportation security," in *Proc. 7th IEEE Int. Conf. Intelligent Transportation Systems*, Washington, DC, Oct. 2004, pp. 100–105.
- [8] H. Chen, J. Qin, E. Reid, W. Chung, Y. Zhou, W. Xi, G. Lai, T. Elhourani, A. Bonillas, F.-Y. Wang, and M. Sageman, "The dark web portal: Collecting and analyzing the presence of domestic and international terrorist groups on the web," in *Proc. 7th IEEE Int. Conf. Intelligent Transportation Systems*, Washington, DC, Oct. 2004, pp. 106–111.
- [9] D. Zeng, H. Chen, L. Tseng, C. Larson, M. Eidson, I. Gotham, C. Lynch, and M. Ascher, "Toward a national infectious disease information infrastructure: A case study in West Nile Virus and Botulism," in *Proc. 5th Annu. Nat. Conf. Digital Government Research*, Seattle, WA, May 2004, pp. 45–54.
- [10] H. Garcia-Molina, J. D. Ullman, and J. Widom, *Database Systems, the Complete Book*. Englewood Cliffs, NJ: Prentice-Hall, 2002.
- [11] E.-P. Lim, J. Srivastava, S. Prabhakar, and J. Richardson, "Entity identification in database integration," *Inform. Sci.*, vol. 89, pp. 1–38, 1996.
- [12] G. Wang, H. Chen, and H. Atabakhsh, "Automatically detecting deceptive criminal identities," *Commun. ACM*, vol. 47, pp. 70–76, 2004.
- [13] H. Chen, D. Zeng, H. Atabakhsh, W. Wyzga, and J. Schroeder, "COPLINK managing law enforcement data and knowledge," *Commun. ACM*, vol. 46, pp. 28–34, 2003.
- [14] P. Klerks, "The network paradigm applied to criminal organizations: Theoretical nitpicking or a relevant doctrine for investigators? Recent developments in the Netherlands," *Connections*, vol. 24, no. 3, pp. 53–65, 2001.
- [15] H. Goldberg and T. Senator, "Restructuring databases for knowledge discovery by consolidation and link formation," in *Proc. 1998 AAAI Fall Symp. Artificial Intelligence and Link Analysis*, 1998, pp. 47–52.
- [16] T. Anderson, L. Arbetter, A. Benawides, and A. Longmore-Etheridge, "Security works," *Security Manage.*, vol. 38, no. 17, pp. 17–20, 1994.
- [17] E. Reid, An Analysis of terrorism literature: A bibliometric and content analysis study, working paper, School of Library and Inform. Manage., Univ. Southern California, Los Angeles, 1983.
- [18] A. Silke, "Devil you know: Continuing problems with research on terrorism," *Terrorism Political Violence*, vol. 13, no. 4, pp. 1–14, 2001.
- [19] L. Kennedy and C. Lunn, Developing a foundation for policy relevant terrorism research in criminology, working paper, Rutgers, The State Univ. New Jersey, New Brunswick, 2003.
- [20] M. Sageman, *Understanding Terror Networks*. Philadelphia, PA: Univ. of Pennsylvania Press, 2004.
- [21] D. Gibson, J. Kleinberg, and P. Raghavan, "Inferring web communities from link topology," in *Proc. 9th ACM Conf. Hypertext and Hypermedia*, 1998, pp. 225–234.
- [22] D. McDonald and H. Chen, "Using sentence-selection heuristics to rank text segments in TXTRACTOR," in *Proc. 2nd Joint Conf. Digital Libraries*, 2002, pp. 28–35.
- [23] K. M. Tolle and H. Chen, "Comparing noun phrasing techniques for use with medical digital library tools," *J. Amer. Assoc. Inform. Sci. Technol.*, vol. 51, no. 4, pp. 352–370, 2000.
- [24] T. Kohonen, *Self-Organizing Maps*. Berlin, Germany: Springer-Verlag, 1995.
- [25] L. Ballesteros and B. Croft, "Dictionary methods for cross-lingual information retrieval," in *Proc. 7th DEXA Conf. Database and Expert Systems Applications*, 1996, pp. 791–801.
- [26] D. Berndt, A. Hevner, and J. Studnicki, "Bioterrorism surveillance with real-time data warehousing," in *Proc. NSF/NIJ Symp. Intelligence and Security Informatics*, 2003, pp. 322–335.
- [27] L. Damianos, J. Ponte, S. Wohlever, F. Reeder, D. Day, G. Wilson, and L. Hirschman, "MiTAP for bio-security: A case study," *AI Mag.*, vol. 23, no. 4, pp. 13–29, 2002.
- [28] E. Conrad, Developing digital neural networks for worldwide disease tracking and prevention. presented at *Proc. Computer Science and Telecommunications, Board of the National Research Council Workshop*. [Online] [http://www7.nationalacademies.org/cstb/project\\_geospatial\\_papers.html](http://www7.nationalacademies.org/cstb/project_geospatial_papers.html)
- [29] D. Zeng, W. Chang, and H. Chen, "A comparative study of spatio-temporal hotspot analysis techniques in security informatics," in *Proc. 7th IEEE Int. Conf. Intelligent Transportation Systems*, Washington, DC, Oct. 2004, pp. 112–117.
- [30] I. J. Gotham, M. Eidson, D. J. White, B. J. Wallace, H. G. Chang, G. S. Johnson, J. P. Napoli, D. L. Sottolano, G. S. Birkhead, D. L. Morse, and P. F. Smith, "West Nile virus: A case study in how NY state health information infrastructure facilitates preparation and response to disease outbreaks," *J. Public Health Manag. Pract.*, vol. 7, no. 5, pp. 75–86, 2001.

- [31] M. Kulldorff, "Prospective time periodic geographical disease surveillance using a scan statistic," *J. R. Stat. Soc. A*, vol. 166, no. 1, pp. 61–72, 2001.
- [32] C. Sonesson and D. Bock, "A review and discussion of prospective statistical surveillance in public health," *J. R. Stat. Soc. A*, vol. 166, pp. 5–12, 2003.
- [33] M. Kulldorff, T. Tangoc, and P. Parkd, "Power comparisons for disease clustering tests," *Comput. Stat. Data Anal.*, vol. 42, no. 4, pp. 665–684, 2003.
- [34] G. P. Patil and C. Tailie, "Upper level set scan statistics for detecting arbitrarily shaped hotspots," *Environ. Ecol. Stat.*, vol. 11, no. 2, pp. 183–197, 2004.
- [35] N. Levine, *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 2.0)*. Houston, TX: Ned Levine, 2002.
- [36] F.-Y. Wang, G. Lai, and S. Tang, "An application specific knowledge engine for researches in intelligent transportation systems," in *Proc. IEEE Conf. Intelligent Transportation Systems*, Washington, DC, Oct. 2004, pp. 960–964.



**Hsinchun Chen** (M'91–SM'04) received the B.S. degree from the National Chiao-Tung University, Hsinchu, Taiwan, R.O.C., the M.B.A. degree from the State University of New York, Buffalo, and the Ph.D. degree in information systems from New York University, New York.

He is a McClelland Professor of Management Information Systems at the University of Arizona (UA), Tucson. He is the author of seven books and more than 150 papers, covering intelligence analysis, data/text/web mining, digital library, knowledge management, medical informatics, and Web computing. He is on the Editorial Boards of the *Journal of the American Society for Information Science and Technology*, *ACM Transactions on Information Systems*, and *Decision Support Systems*. He is a Scientific Counselor for the National Library of Medicine (NLM) and has served as an Advisor for the National Science Foundation (NSF), Department of Justice (DOJ), NLM, and other international research programs in digital library, digital government, medical informatics, and intelligence analysis. He is the Founding Director of the UA Artificial Intelligence Laboratory and the Hoffman E-Commerce Laboratory. The UA Artificial Intelligence Laboratory, which houses over 40 researchers, has received more than \$15 million in research funding from NSF, NIH, NLM, DOJ, CIA, and other agencies over the past 15 years. The Hoffman E-Commerce Laboratory, funded primarily by major IT industry partners, features state-of-the-art e-commerce hardware and software in a cutting-edge research and education environment. He has been a Principal Investigator (PI) of many major NSF Digital Library and Digital Government research programs.

Dr. Chen was the Andersen Consulting Professor of the Year in 1999. He serves on the Editorial Boards of IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, and IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS. He served as Conference Co-Chair of the ACM/IEEE Joint Conference on Digital Libraries (JCDL) in 2004 and of the NSF/NII Symposium on Intelligence and Security Informatics (ISI) in 2003 and 2004. He has served as the Conference General Chair or International Program Committee Chair for the past six International Conferences of Asian Digital Libraries (ICADL) from 1998 to 2003. His COPLINK system has been widely adopted in law enforcement (in CA, AZ, TX, MI, MA, WA, etc.) and the intelligence community (CIA, NSA, and Department of Homeland Security) in the United States. He has received numerous industry awards in knowledge-management education and research including the AT&T Foundation Award, SAP Award, and the Andersen Consulting Professor of the Year Award.



**Fei-Yue Wang** (S'87–M'89–SM'94–F'03) received the B.S. degree in chemical engineering from Qingdao University of Science and Technology, Qingdao, China, in 1982, the M.S. degree in mechanics from Zhejiang University, Hangzhou, China, in 1984, and the Ph.D. degree in electrical, computer, and systems engineering from the Rensselaer Polytechnic Institute, Troy, NY, in 1990.

He joined the University of Arizona, Tucson, in 1990, where he became a Full Professor of Systems and Industrial Engineering in 1999 and currently is Director of the Program for Advanced Research in Complex Systems. In 1999, he founded the Intelligent Control and Systems Engineering Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China, with the support of the Outstanding Oversea Chinese Talents Program. Since 2002, he has been Director of the Key Laboratory of Complex Systems and Intelligence Science, Chinese Academy of Sciences. He was Editor-in-Chief of the *International Journal of Intelligent Control and Systems* from 1995 to 2000 and currently is the Editor-in-Charge of the Series in Intelligent Control and Intelligent Automation, Editor for the ITS Department of the *IEEE Intelligent Systems* and *IEEE Transactions*, and on the Editorial Board of several other international journals. He has published more than 200 books, book chapters, and papers since 1984 and has received more than \$20 million and over ¥ 50M RMB from the NSF, DOE, DOT, NNSF, CAS, Caterpillar, IBM, HP, AT&T, GM, BHP, RSVI, ABB, and Kelon. His current research interests include modeling, analysis, and control mechanism of complex systems; agent-based control systems; intelligent control systems; real-time embedded systems, application-specific operating systems (ASOS); applications in intelligent transportation systems, intelligent vehicles, and telematics; web and service caching; smart appliances and home systems; and network-based automation systems.

Dr. Wang received the Caterpillar Research Invention Award with Dr. P. J. A. Lever in 1996 for his work in robotic excavation and the National Outstanding Young Scientist Research Award from the National Natural Science Foundation of China in 2001, as well as various industrial awards for his applied research from major corporations. He is an Associate Editor for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION, and IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS. He is an Elected Member of the IEEE SMC Board of Governors and IEE ITSC AdCom and Secretary of the IEEE Intelligent Transportation Systems Council and President-Elect of the IEEE Intelligent Transportation Systems Society. He was Program Chair of the 1998 IEEE International Symposium on Intelligent Control; the 2001 IEEE International Conference on Systems, Man, and Cybernetics; General Chair of the 2003 IEEE International Conference on Intelligent Transportation Systems; and will be Co-Program Chair of the 2004 IEEE International Symposium on Intelligent Vehicles and General Chair for the same conference in 2005. He was Vice President and one of the major contributors of the American Zhu Kezhen Education Foundation and a Member of the Boards of Directors of five companies in information technology and automation.



**Daniel Zeng** (S'99–M'04) received the B.S. degree in economics and operations research from the University of Science and Technology of China, Hefei, China, and the M.S. and Ph.D. degrees in industrial administration from Carnegie Mellon University, Pittsburgh, PA.

He is an Assistant Professor and Honeywell Fellow in the Department of Management Information Systems, University of Arizona, Tucson. He also directs four National Science Foundation (NSF)-funded research projects as Principal Investigator or co-Principal Investigator. He has co-edited two books and published approximately 55 peer-reviewed papers in management information systems and computer science journals, edited books, and conference proceedings. He serves on the Editorial Boards of the *Journal of Database Management* and *International Journal of Intelligent Information Technologies*. His research interests include software agents and their applications, distributed optimization, computational support for auctions and negotiations, intelligent information integration and caching, and recommender systems.

Dr. Zeng is a Member of Institute for Operations Research and the Management Sciences (INFORMS), Association for Information Systems (AIS), American Association for Artificial Intelligence (AAAI), and the Association for Computing Machinery (ACM). He served as Program Co-Chair of the first NSF/NII Symposium on Intelligence and Security Informatics (ISI'03) and as Conference Co-Chair of ISI'04.