DIMACS Programs in Homeland Security

Introductory Comments

DIMACS, founded in 1989 with a prestigious NSF ”science and technology center” grant, is a partnership of Rutgers and Princeton Universities, AT&T Labs, Bell Labs, NEC Laboratories America, and Telcordia Technologies, with partner organizations Avaya Labs, HP Labs, IBM Research, Microsoft Research and Stevens Institute of Technology. It is headquartered at Rutgers University. There are some 250 scientists who are ”permanent members” of DIMACS, most located at one of the partner organizations. These permanent members are all paid by their own organizations, but are involved as appropriate in DIMACS projects, workshops, and educational programs.

DIMACS projects take several forms. There are large, multi-disciplinary, multi-institutional group projects where the time of participants (faculty, industry researchers, graduate students, postdocs) is supported. There are smaller, more informal ”working groups” that meet occasionally, where usually no one is paid directly or only one or two people have their time paid for. There are also lots of workshops (maybe 30 a year) that we use to gather interdisciplinary groups from around the country and that often spawn research collaborations. We organize a lot of our programs around themes, call them ”special focus programs,” and include in the special focus programs workshops, tutorials, and working groups, and support postdocs and graduate students.

For more about DIMACS, see http://dimacs.rutgers.edu/

Selected DIMACS Projects

- 1. **Monitoring Message Streams.** This large multi-player research project is sponsored by the intelligence community. It is concerned with developing methods to monitor huge streams of text data for messages and to identify clusters of messages on events of interest. There are three themes to both our past and proposed work. a) We attack all five stages of message filtering (compression/indexing, text representation, matching/similarity measures, learning from data and prior knowledge, and fusing of results from multiple sources) simultaneously, with particular attention to interactions between them. b) We pursue algorithmic improvements that allow tunable tradeoffs between increased effectiveness and reduced resource usage (memory, CPU). c) We seek ways to use all available information resources to reduce the need for the users of filtering systems to provide large amounts of training data or other labor-intensive input.

- 2. **Author Identification.** Homeland security and the criminal and civil justice systems require reliable and valid methods to identify the authors of anonymous documents. Our objective is to bring state-of-the-art statistical and computing technologies to bear on problems of author identification. We are working on techniques for identifying authors in large collections of textual artifacts (e-mails, communiques, transcribed speech, etc.). Our approach focuses on very high-dimensional, topic-free document representations and particular attribution problems such as: (1) Which one of these K authors wrote this particular document? (2) Did any of these K authors write this particular document? (3) Were these two documents written by the same author? (4) Are these two putative authors really the same person? This project is also a large, multi-player project supported by the intelligence community.

- 3. **“Special Focus” on Computational and Mathematical Epidemiology.** This elaborate program involves modeling of infectious diseases, including human and plant diseases. The large number of working groups and workshops include topics such as large data sets, data mining and epidemiology, analogies between human and computer viruses and immune systems, analysis of large models, social networks, vaccination strategies, spatio-temporal modeling of disease, climate and disease, botanical epidemiology, and co-evolution of hosts, pathogens, and vectors. There is some emphasis on exploring challenges for discrete mathematics and theoretical computer science in these domains, methods that have not been widely used before.

URL: http://dimacs.rutgers.edu/SpecialYears/2002_Epid/

- 4. **Working Group on Adverse Event/Disease Reporting, Surveillance, and Analysis.** This group deals with problems of syndromic surveillance and works closely with the CDC. It deals with such data as managed care patient encounter data, pre-diagnostic chief complaint data, over-the-counter sales transactions, 911 emergency calls, ambulance dispatch data, absenteeism data, ED discharge summaries, prescription data, and adverse event reports. The group is investigating such methodologies as spatial-temporal scan statistics, statistical process control, Bayesian applications, market-basket association analysis, text mining, rule-based surveillance, and change-point techniques. More on the group can be found at

  http://dimacs.rutgers.edu/Workshops/AdverseEvent/

- 5. **Bioterrorism Working Group**. This group meets from time to time informally. It has had one formal meeting that led to a report (cited below) and its work has led to involvement with the HHS smallpox modeling group, the beginnings of work on agroterrorism threats, and a list of research challenges that has informed other working groups at DIMACS and elsewhere. Relevant report:

  http://dimacs.rutgers.edu/Workshops/WGDeliberate/FinalReport5-20-02.doc.

- 6. **Working Group on Modeling Social Responses to Bioterrorism.** This group, which was an offshoot of the group in number 5, has been challenging the traditional assumption in bioterrorist and disease response planning that there is a fixed social landscape in which the public consists of passive bystanders and rational actors who comply with health authorities. Issues considered by the group have included risk communication, measurement of social disruptions and economic effects of disease events, social stigmata, etc. The group is also concerned with modeling public health decision making and, generally, incorporating social behavior into epidemiological models. Relevant report: http://dimacs.rutgers.edu/Workshops/Modeling/Report.doc

- 7. **Bioterrorism Sensor Location.** The government's plans to develop and place sensors that might quickly provide warning of a possible bioterrorist attack leads to challenging questions. Networks of sensors are expensive to purchase and ways to locate them to maximize "coverage" and expedite an alarm in case of an attack are not easy to determine. This project is aimed at formulating the sensor location problem (SLP) carefully and at developing algorithmic approaches for solving it. When sensors set off an alarm, decision makers need to decide whether or not a bioterrorist attack has taken place, estimate its extent and location, and choose among a variety of possible responses. They can be guided by the pattern of sensor alarms. But the problem of how to interpret this pattern, the Pattern Interpretation Problem (PIP), presents a major challenge that is intertwined with the SLP. This project aims to formulate the PIP precisely, investigate its implications for finding good solutions to the SLP, and to find methods for solving the PIP and attacking the PIP and SLP together. We are concentrating on combinatorial optimization approaches to the SLP and new statistical approaches to the PIP. The project came to us via MITRE Corporation and the Defense Threat Reduction Agency and at present is being supported through the Computational and Mathematical Epidemiology Special Focus.

- 8. **Predicting Disease Outbreaks from Remote Sensing and Media Data.** Outbreaks of disease in other parts of the world have the potential to adversely affect the security of the United States. They lead to local social disruption, which may escalate to the point where it could politically destabilize an entire region. With the increasing interconnectivity of the US with the rest of the world via the air traffic grid, these transnational bio-threats quickly become issues for US security. Recent work has shown that it is possible to predict disease outbreaks in distant parts of the world using remotely sensed satellite data, thus allowing the possibility of implementing preventive measures as well as providing the opportunity for earlier intervention through more efficient mobilization of response. A similar advantage can be obtained in accelerating the management of international health related social disruptions, if they could be predicted in a like manner, either from enviro-climatic data obtained from remote sensing or from other indicators such as media reports. This is the goal of the project. That is, to predict the probability of disease-related social unrest in an area based on remotely measured indications similar to those used in forecasting infectious disease outbreaks. The project, a multi-player, multi-institutional project, is just getting under way.

- 9. **"Special Focus" on Communication Security and Information Privacy.** Vitally important aspects of our modern society have become dependent on rapid and secure communication, which is increasingly electronic. The new electronic age offers vast potential for new services and applications, but gives rise to serious new vulnerabilities and security threats. Moreover, many of the most important new applications come at the price of threats to privacy. This special focus is exploring the new vulnerabilities and threats and new methods for dealing with them. Working Groups are investigating On-line Privacy (Threats and Tools), Privacy/Confidentiality of Health Care Data, Intrusion Detection and Network Security Management Systems, the Secure and Efficient Extraction of Joint Information from Multiple Data Sets, and Mobile Code Security. Workshops are being held on Large-scale Internet Attacks, Intellectual Property Protection, Security of Web Services and E-Commerce, Cryptography (Theory Meets Practice), Security Analysis of Protocols, and Security and Trust Issues Associated with Ad-Hoc Computing/Pervasive Networking. Additional workshops are on the topics Database Security: Query Authorization and Information Inference, Mobile and Wireless Security, and Electronic Voting (Theory and Practice). Of special relevance to homeland security are the efforts in this special focus on privacy-preserving data mining, confidentiality of health data, and secure information sharing. For general information about this special focus, see http://dimacs.rutgers.edu/SpecialYears/2003_CSIP/

- 10. **Port-of-Entry Inspection Algorithms.** Finding ways to intercept illicit nuclear materials and weapons destined for the U.S. via the maritime transportation system is an exceedingly difficult task. At the present time, only about 2% of ships entering U.S. ports have their cargoes inspected. The purpose of this study is to develop decision support algorithms and tools that will optimally intercept illicit materials and weapons. No single "silver bullet" approach will suffice. Instead, a full system-of-systems approach that examines opportunities for detection from the material's point of origin to its intended target must be considered. Existing algorithms for optimally intercepting illicit cargo assume that sensor performance, operating characteristics of ports, and overall threat level are all fixed. The new approaches involve sophisticated decision logics and are built around physical models, simulations, operational models, and decision architectures of various kinds. Complications involve economic impacts of surveillance activities, errors and inconsistencies in available data on shipping and import terminal facilities, and the tradeoffs between combinations of sensors. This project will investigate the decision problem of when to initiate different levels of inspection if there are seasonal variations in cargo flows and cargo types, sensor reliability effects, and changing threat levels. It will explore new sensor deployment methods and sensor configurations, the problem of false alarms from naturally occurring radiation sources (which vary spatially) and from innocent cargos (such as medical waste), and models of "information sensors." The project is joint with Los Alamos National Laboratory and is funded by Los Alamos and by the Office of Naval Research.