

# Agency Interoperation for Effective Data Mining in Border Control and Homeland Security Applications<sup>1</sup>

Nabil R. Adam, Vijayalakshmi Atluri and Vandana P. Janeja  
CIMIC, Rutgers University, Newark NJ 07102  
{adam,atluri,vandana}@cimic.rutgers.edu

## 1. Project Summary

US Customs has recently embarked on a major modernization initiative of its Information Technology systems. Drawing in data from Customs trade systems, targeting inspectors review manifest information as well as strategic and tactical intelligence to determine which shipments and containers are “high-risk.” This entails a considerable level of communication and data sharing between various government agencies. Our NSF funded project, funded through the digital government program, aims at providing decision makers with the ability to extract and fuse information from multiple, heterogeneous sources in response to a query while operating under a decentralized security administration. Based on the idea of “Smart Borders”, the system will utilize data available from different agencies, ports and customs divisions to supplement the profiling by targeting towards anomalies, and detect various flags raised by non-conforming shipments or abnormal behavior of inbound cargos and raise a combination of alerts. This project proposes generalizable development work, which devises solutions to accomplish secure interoperation among different government agencies. The output of this project would ideally enhance the security aspect of the Automated Commercial Environment (ACE) system by incorporating the concept of semantic interoperability, anomaly detection and subsequent spatial and geographical visualization of information that can help Customs inspectors make better decisions. This project is a collaboration between the industry, government and the academia, providing the opportunity to directly influence the practical needs of the government agency, in this case the US Customs.

## 2. Research Activities

The impact of this project is in many dimensions. (1) It devises solutions to accomplish secure interoperation among different government agencies. (2) It advances fundamental research in Data interoperability, data mining, decentralized security administration, and the diplomacy and politics. (3) As a result of the partnership SAP, the supplier of software to the Customs Modernization Program and contractor to IBM, which is the prime contractor for eCustoms Partnership (eCP) implementing Customs Modernization, it provides the opportunity to directly influence the practical needs of US Customs. (4) The research and development work in this project is generalizable and therefore can serve as a reference model to be adopted by related homeland security agencies. Our work in the area of spatial data mining has investigated various aspects of identification of anomalies in high dimensional spatio-temporal datasets [1,2]. Our work in the area of diplomacy and politics has focused on (1) US-VISIT, which helps to secure our borders and expedite the entry/exit process while enhancing the integrity of our immigration system and respecting the privacy of our visitors [3], and (2) on international cooperation on electronic advanced passenger information transfer and passport biometrics [4]. Our work on semantic interoperability has focused on ontology alignment in which we have proposed domain-independent, statistical ontology construction from text [5]. Our work in data mining has focused on developing web-service based data mining software, which can be used to extract anomalies for the data produced by this project. We have started with network intrusion data since it is readily available, naturally distributed, and heterogeneous. Additionally, our work has examined issues in data mining of homeland defense related data [6,7] and issues in creating web services for mining high volume data streams. [8,9].

## 3. Project Participants

The collaborating researchers include CIMIC, Rutgers Univ.: Nabil R. Adam and Vijay Atluri, Dept. of Political Science, Rutgers Univ.: Rey Koslowski, Univ. of Illinois at Chicago: Robert Grossman, National

---

<sup>1</sup> Supported by the National Science Foundation under grant IIS-0306838.

Center for Data Mining, David Hanley (Technical staff), Columbia Univ.: Vasileios Hatzivassiloglou and Kathleen R. McKeown, Dept. of Computer Science. Industry partners include Tao Lin-SAP Corporate Research Labs, Domain experts include Dr. Stephen E. Flynn-Council on Foreign Relations, C.J. Chang-Foreign Operations Division-U.S. Department of Homeland Security, Luis R. Cortes-Chief of Intelligence, Office of National Risk Assessment (ONRA), Transportation Security Administration (TSA), James R. Sutton-Managing Associate of McManis Associates, Inc. and former Senior Intelligence Research Specialist for the U.S. Department of Justice on the Foreign Terrorist Tracking Task Force. Student participants include Rutgers Univ.: Vandana Janeja, Janice Warner, Columbia Univ.: Melania Degeratu and, Univ. of Illinois at Chicago: Chetan Gupta, Jorge Levera.

#### **4. Data Acquisition**

Acquiring the data has been one of the highly critical tasks of this project. It is essential to understand the security concerns of the agencies in sharing the data, at the same time it is also essential to reflect the real world dynamics in the solutions we provide. We have recently acquired US Customs data from PIERS which has access to import /export data due to the Freedom of Information Act along with U.S. Customs Regulations which authorize press organizations to copy certain shipping documents available to the public. It provides Imports, exports data and US and overseas profiles of companies. We have also been investigating IPUMS datasets, which has similar characterizations to the data gathered by different agencies such as immigration, labor, housing data, etc. C.J. Chang has indicated that the attributes of this data are similar to those of typical customs data. In addition, US Customs is in the process of obtaining approval from the board to share some sanitized data with the project team.

#### **5. Success and Impact of the Project**

As a result of this project, several other collaborations have been generated between Rutgers and SAP. These include the RFID, data interoperability and privacy project, with SAP. This project has resulted in two on-going Ph.D. dissertations. The research publications generated as direct results of the first year of funding of this project are available at <http://civic.rutgers.edu/~vandana/BorderControlPublications.htm>. This project has been a highlight in the Rutgers University Symposium on Homeland Security Research organized by the Rutgers University Homeland Security Research Initiative (RUHSRI). We have actively participated in the two symposia of Homeland Security research at Rutgers University (RUHSRI) (<http://dimacs.rutgers.edu/RUHSRI/>).

#### **References**

- [1] V. P. Janeja, V. Atluri and Nabil R. Adam, "Detecting Anomalous Geospatial Trajectories through Spatial Characterization and Spatio-Semantic Associations," National Conference on Digital Government, 2004,
- [2] N.R. Adam, V.P. Janeja, V. Atluri, "Neighborhood Based Detection of Anomalies in High Dimensional Spatio-temporal Sensor Datasets," ACM Symposium on Applied Computing, March 2004
- [3] R. Koslowski, "International Cooperation to Create Smart Borders" in Ottawa at the Conference on North American Integration: Migration, Trade and Security, April 1, 2004.
- [4] R. Koslowski, "International Cooperation on Electronic Advanced Passenger Information Transfer and Passport Biometrics," at the International Studies Association meeting at Montreal, March 17-20, 2004.
- [5] M. Degeratu and V. Hatzivassiloglou. "An Automatic Method for Constructing Domain-Specific Ontology Resources". In Proc. of the Language Resources and Evaluation Conference (LREC), May 2004, to appear.
- [6] R. L. Grossman, "Alert Management Systems: A Quick Introduction, in Managing Cyber Threats: Issues, Approaches and Challenges," edited by Vipin Kumar, Jaideep Srivastava, Aleksandar Lazarevic, Kluwer Academic Publisher, to appear.
- [7] J. Levera, B. Barán, and Robert Grossman, "Experimental Studies Using Median Polish Procedures to Reduce Alarm Rates in Data Cubes of Intrusion Data, Intelligence and Security Informatics for National and Homeland Security," Hsinchun Chen, Reagan Moore, Daniel Zeng, John Jeavitt, editors, LNCS 3073, Springer Verlag, New York, 2004, pages 482-491.
- [8] C.Gupta and R. L. Grossman, "GenIc: A Single Pass Generalized Incremental Algorithm for Clustering," 2004 SIAM International Conference on Data Mining (SDM 04), to appear.
- [9] R. L. Grossman, Y. Gu, C. Gupta, David Hanley, Xinwei Hong, and Parthasarathy Krishnaswamy, "Web-Service Based Data Mining Middleware for Grid Computing Environments," 7th International Workshop on High Performance and Distributed Mining, to appear.