Secure Interoperation for Effective Data Mining in Border Control and Homeland Security Applications[1]

Nabil R. Adam, Vijayalakshmi Atluri, Rey Koslowski, Robert Grossman, Vandana P. Janeja, Janice Warner
{adam,atluri,vandana,jwarner}@cimic.rutgers.edu,rkoslowski@earthlink.net,grossman@uic.edu

1. PROJECT SUMMARY

Our NSF funded project 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. Our motivation comes from US Customs, which 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 "high-risk" shipments and containers. This entails a considerable level of communication and data sharing between various government agencies. 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. 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.

2. RESEARCH ACTIVITIES

Secure Data Interoperation: We have proposed a *coalition based access control* (CBAC) model [2], extended it to *dynamic coalition-based access control* (DCBAC) model [15], which allows a user's request to access a resource belonging to another coalition entity to be automatically translated using attributes associated with user credentials and objects. We proposed an approach to handle how local access control policies could be mapped into collaboration level policies using attributes and graphs, and examined the problem of determining appropriate attributes and their values to require from remote users in order to grant them access to a requested resource [16]. Our process transforms Role Based Access Control (RBAC) policies into attribute requirements that must be presented by external users via credentials. We extended the coalition service registry (CSR) architecture which supports DCBAC to include distributed CSR (DCSR) [14]. In a DCSR system, several service registry agents cooperate to provide controlled access to coalition resources. Distribution of the registries results in improved availability, higher concurrency, better response times to user queries, and enhanced flexibility.

Anomaly Detection in Real-time Data Streams: We introduce an algorithm for detecting outliers on streaming data, which relies on computing a dyadic decomposition into cubes in Euclidean space [3]. If we view the dyadic decomposition as a tree with a fixed maximum size, then outliers are naturally defined by cubes containing a small number of points in the cube, or the cube itself and its neighboring cubes. The cumulative sum (CUSUM) algorithm is a standard algorithm for detecting changes in an event stream. CUSUM assumes that both normal and unusual events be modeled using standard distributions, and uses a formula based upon the log odds ratio to sound an alarm when a threshold is passed. Variations on the CUSUM also exist for cases of unknown distributions. A natural extension of CUSUM is to consider collections or ensembles of CUSUM algorithms and a rule that determines one or more members of the ensemble to apply. We consider ensembles of CUSUM algorithms defined by cells in a multidimensional data cube [4], which arise naturally when large data sets have temporal, spatial, or spatial-temporal variations. In the area of change detection in multi-modal streaming data [5], we developed a testbed containing: real time data from over 830 highway traffic sensors in the Chicago region, data about weather, and text data about events. The goal was to detect in real time interesting changes in traffic conditions. Given the size and complexity of the data, we built a separate baseline model for each hour in the day, for each day in the week, and for every 2 or 3 traffic sensors, resulting in over 42,000 separate baseline models. We also built a baseline engine to build the necessary baselines automatically. We modified an open source scoring engine to process in real time each new sensor reading, update the appropriate feature vectors, score the updated feature vectors using the baseline models, and send out real time alerts when deviations from the baselines were detected. The system and architecture we developed should apply more generally when there is a requirement to generate real time alerts from multiple streams of complex data. In [17], we introduce Tukey and Tukey scagnostics and develop graph theoretic methods for implementing their procedures on large data sets. An important advantage of this approach is that the visualization does not suffer the curse of dimensionality that effects many competing approaches. This approach appears to be a fruitful method for visually detecting anomalies in large dimensional data sets.

Anomaly Detection in Spatio-temporal Data: We have proposed a random walk based free-form spatial scan statistic approach for anomalous window detection [6]. A spatial scan statistic considers a scan window, and identifies anomalous windows by moving the scan window in the region. Earlier proposals suffer from two limitations: (i) They restrict the scan window to be of a regular shape (e.g., circle, rectangle, cylinder),

whereas the region of anomaly, in general, is not necessarily of a regular shape. (ii) They take into account autocorrelation among spatial data, but not spatial heterogeneity. As a result, they often result in inaccurate anomalous windows. To address these limitations, we proposed a random walk based Free-Form Spatial Scan Statistic (FS^3). Application of FS^3 on real datasets has shown that it can identify more refined anomalous windows with better likelihood ratio of it being an anomaly, than those identified by earlier spatial scan statistic approaches.

Semantic graph (SG) based knowledge discovery: We use SGs from a set of disparate sources and related ontologies [1]. We took a two-step approach: First, we created a refined enhanced graph by combining multiple relevant SGs and combining relevant knowledge from ontologies. This involved identifying relevant ontologies, reconciling different terminology, inferring new facts, and checking consistency of information in the SGs gathered from different sources. Second, having the enhanced and refined SG, we employed a semantics driven approach to detect patterns.

Diplomacy and politics: We examined how Canada has been using new information technologies to screen terrorists while enabling legitimate travel and trade. We interviewed Canadian Border Services Agency (CBSA) and Foreign Affairs officials in Ottawa and visited border crossings and ports from the Atlantic to the Pacific coasts, met with local CBSA managers and saw the technology at work. The report [7] was presented in Washington at a meeting attended by several Department of Homeland Security officials. Comments were given by Richard Stana, Director, Homeland Security and Justice Issues, Government Accountability Office. It was subsequently discussed in the New York Times, the Congressional Quarterly, The Atlanta Constitution, The Arizona Republic and several other newspapers. Several presentations have been made [8-13].

3. SUCCESS AND IMPACT

As a result of this project, several other collaborations have been generated between Rutgers and SAP, including the RFID, data interoperability and privacy project. This project has resulted in two on-going Ph.D. dissertations. The publications generated during the first two years of funding of this project are available: http://cimic.rutgers.edu/~vandana/BorderControlPublications.htm. N. Adam gave a talk at SAP Research, Kalrsruhe, Germany, in Nov. 2004. During 2006, Dr. Rey Koslowski plans to extend his research to the EU, Australia and New Zealand. His focus is in technologies used at ports of entry to screen passengers and cargo, biometrics-based registered traveler programs, advanced passenger information systems, RFID-enabled biometric visas and ICAO-compliant "e-passports." He will work on EU implementation of EURODAC, the asylum seeker fingerprint database, deployment of the Schengen Information System (SIS) and SISII (information system for sharing data among Schengen Convention signatory states for checking of border crossers), the European Visa Identification System (VIS). In addition to meeting with policymakers and border control officials (interior ministry, customs, immigration, foreign ministry) of EU member states, Australia and New Zealand, he plans to meet with officials at the European Commission and World Customs Organization in Brussels, the Schengen Information System secretariat in Strasbourg, the International Organization for Migration in Geneva, the Budapest Process Secretariat and Austrian Interior ministry in Vienna, Interpol in London, Europol in the Hague and the EU Agency for the Management of Operational Cooperation at the External Borders of the Member States (Frontex) in Warsaw.

REFERENCES

[1] N. Adam et al. "Semantic Graph based Knowledge Discovery from Heterogeneous Information Sources", Working Together: Conference on Public/Private R&D Partnerships in Homeland Security, 2005.

[2] V. Atluri, J. Warner: Automatic Enforcement of Access Control Policies Among Dynamic Coalitions, ICDCIT 2004.

[3] C. Gupta and R. L. Grossman, Outlier Detection in Streams With Dyadic Cubes, will be submitted to KDD 2006.

[4] R. L. Grossman and H. V. Poor, Baselines and Change Detection Using Ensembles of CUSUM Algorithms, submitted to IEEE TKDE.

[5] R. L. Grossman et al. Change Detection and Alerts from Highway Traffic Data, ACM/IEEE SC 2005 Conference.

[6] V.P. Janeja and V. Atluri, "FS3 : A Random Walk based Free-Form Spatial Scan Statistic for Anomalous Window Detection". ICDM 2005, Houston,Texas, USA

[7] R. Koslowski, Real Challenges for Virtual Borders: The Implementation of US-VISIT, Migration Policy Institute Report, June 2005.

[8] R. Koslowski, "Virtual Borders and Homeland Security," Beyond Terror: A New Security Agenda, Watson Institute, Brown University, June, 2005.

[9] R. Koslowski, "Real Challenges for Virtual Borders: The Implementation of US-VISIT," presentation and report release, Migration Policy Institute, Washington, DC June, 9, 2005

[10] R. Koslowski, "Real Challenges for Virtual Borders," Centre on Migration Policy and Society, Oxford University, Mar. 16, 2005.

[11] R. Koslowski, "Toward Virtual Borders: Expanding European Border Control Policy Initiatives and Technology Implementations" An Immigration Policy of Europe? New York University and the European Union Institute, Florence, Italy, March 13-15, 2005.

[12] R. Koslowski, "Possible Steps Towards an International Regime for Mobility and Security", UN, Mar, 2005.

[13] R. Koslowski, "European and Transatlantic Cooperation on Migration, Mobility and Security" Beyond the U.S. War on Terrorism: Comparing Domestic Legal Remedies to an International Dilemma, The J.B. Moore Society Spring 2005 Symposium, University of Virginia Law School, Feb, 2005.

[14] R. Mukkamala, V. Atluri and J. Warner, A distributed Service Registry for Resource Sharing Among Ad-hoc Dynamic Coalitions, IFIP TC-11 WG 11.1 and WG 11.5 Joint Working Conference, 2005

[15] J. Warner, V. Atluri and R. Mukkamala, A Credential-Based Approach for Facilitating Automatic Resource Sharing Among Ad-Hoc Dynamic Coalitions. DBSec 2005: 252-266

[16] J. Warner, V. Atluri and R. Mukkamala, An Attribute Graph Based Approach to Map Local Access Control Policies to Credential Based Access Control Policies, ICISS 2005.

[17] L. Wilkison, A. Anand, R. Grossman, Graph-Theoretic Scagnostics, IEEE Symposium on Information Visualization 2005.

[1]Supported in part by the National Science Foundation under grant IIS-0306838.