Protecting Privacy, Civil Rights and Civil Liberties using Data Labeling and Segmentation

Timothy R. Schmoyer, Bill Doyle, Michael Dufel
Jericho Systems Corporation
Dallas, TX
tim.schmoyer@jerichosystems.com
bill.doyle@jerichosolutions.com
michael.dufel@jerichosystems.com

Karyn Higa-Smith
Science and Technology Directorate, Homeland Security
Advanced Research Projects Agency
Department of Homeland Security
Washington, DC
karyn.higa-smith@hq.dhs.gov

Abstract—The Cyber Security Division of Department of Homeland Security (DHS) Science and Technology’s Homeland Security Advanced Research Projects Agency is prototyping data labeling and data segmentation services that leverage dynamic fine-grained access control to protect information considered sensitive and/or related to privacy, civil rights, and civil liberties (P/CRCL). Prototype operationalizes significant work that has already been completed in the Information Sharing Environment (ISE) and the schemas and documentation available in National Information Exchange Model (NIEM) Information Exchange Package Documentation (IEPD) for Critical Infrastructure and Key Resources fusion center information sharing requirements. This paper describes three Unified Modeling Language (UML) use cases, concept of operations, and architecture for using dynamic fine-grained access control, data labeling, and data segmentation to anonymize, mask, or redact data that might be shared by a Fusion Center in accordance with organizational, jurisdictional, and Community of Interest (COI) laws, regulations, and policies.

Keywords-cybersecurity; privacy; fusion center; information sharing; access control; data segmentation; content filtering; de-identification; P/CRCL; NIEM; UML; IAM; IdAM; FICAM; BAE; ABAC; DHS;

I. INTRODUCTION

The DHS Data Labeling and Data Segmentation (DLDS) research and development (R&D) project leverages proven technology demonstrated in [1]. The successful demonstration of this technology to protect data segments contained in a Continuity of Care Document used to exchange health information creates the opportunity to research, develop, and operationalize similar technology that will enable fusion centers and mission partner organizations to better manage access to data and improve information sharing in a manner that protects P/CRCL consistent with applicable law, policy, and mission. The value-added by this project is the enhanced ability of fusion centers to share information resulting in improved data collaboration and situational awareness in near real-time.

The DLDS project focuses on information that might be received, gathered, analyzed, vetted, and shared by a fusion center with private sector, public safety/government, and law enforcement partners. DLDS delivers a robust and secure solution that facilitates the state’s ability to protect its citizens by improving the trusted exchange of data between first responders, the law enforcement community, critical infrastructure providers, other businesses, and citizens. By automating control of P/CRCL sensitive data and minimizing privacy risk in a well-defined, predictable, consistent, and auditable manner, the system will foster confidence that information is being controlled and protected appropriately.

II. HIGH-LEVEL OVERVIEW

A. Fusion Center Information Sharing

Fusion centers have the requirement to collect fact information and then share pertinent information with their mission partners in the public and private sectors within their jurisdiction, and with entities external to their jurisdiction using fusion center resources as shown in Figure 1. Sharing information between fusion centers and with entities external to the fusion center’s jurisdiction is an important mission requirement supporting [2]: Foster a culture that recognizes the importance of fusing “all crimes with national security implications” and “all hazards” information (e.g., criminal investigations, terrorism, public health and safety, all-hazards and emergency response) which often involves identifying criminal activity and other information that might be a precursor to a terrorist plot.

Figure 1. Fusion Center Information Sharing Environment
References [3] and [4] emphasize using a consistent framework for fusion center operations and fusion centers are encouraged to leverage existing systems, databases, and networks. However, fusion centers also have the ability to tailor their scope and mission to meet specific jurisdictional needs which results in unique requirements that must adhere to common jurisdictional constraints. For example, in the area of P/CRC, Fusion Center Guideline 8 states, “Develop, publish, and adhere to a privacy and civil liberties policy” and includes:

- Adherence to applicable state and federal constitutional and statutory privacy and civil liberties provisions.
- Compliance of Fusion center participants with all local, state, tribal and federal privacy laws, when applicable.

As a result, to integrate the DLDS solution into the fusion center information sharing environment, it must facilitate each fusion center’s unique scope and mission; leverage existing systems, databases, and networks; and protect P/CRCL in accordance with local, state, tribal and federal privacy laws.

B. Myriad of Information Systems, Formats, and Constraints

The scope and mission of a fusion center, based on specific jurisdictional needs and operational requirements, might require access to, and responsibility for sensitive and protected information such as health-related data and data contained in motor vehicle records, driver license records, correctional databases, sex offender registries, etc. This data is currently decentralized across many different information systems and in different formats. As a result, fusion center personnel may have access to dozens of information systems through virtual private networks accessed from the fusion center’s internal network.

To operationalize data labeling and data segmentation, standardize the use of metadata, and meet security control requirements for authentication, authorization, and auditing; numerous laws, regulations, policies, procedures, technical guidance, and industry standards must be taken into account. Automating the enforcement of these complex requirements is especially important to dynamically scale as information is collected, stored, and analyzed, while continually complying with jurisdictional policies that can change over time.

During the requirements gathering phase, the research team identified information systems and data formats in fusion centers in Texas and California. Our findings confirm the saying, “If you have seen one fusion center, you have seen one Fusion Center.” Fusion centers have unique solutions that meet their specific jurisdictional needs and must be included in any information sharing solution. Two information systems the research team found commonly accessible from the fusion centers they visited that enable information sharing are the Homeland Security Intelligence Network (HSIN) and Regional Information Sharing System Network (RISS.net). In addition, the National Identity Exchange Federation (NIEF), an outgrowth of the Global Federated Identity and Privilege Management (GFIPM), and Backend Attribute Exchange (BAE), enable solutions that leverage federated identity and attributes for inter-fusion center information sharing. Finally, the research team identified existing NIEM IEPDs and schemas relevant to fusion center information sharing that identify delimited data structures using metadata attributes, including [5], [6], [7], [8], [9], and [10].

Based on these findings and the previous work completed in [1], the DLDS system has been designed to leverage available identity credentials, subject attributes, environmental attributes, resource metadata, and digital policies to perform static and dynamic labeling along with the segmentation of structured and unstructured data. The DLDS system builds on existing authentication and authorization mechanisms, and can be easily integrated with an attribute-based access control (ABAC) capability to determine how to access and distribute data. The DLDS solution leverages privacy fields, dissemination codes, and other “release-to” caveats available in the metadata to enforce digital policy rules that dynamically control access and protect P/CRCL in a well-defined, predictable, consistent, and auditable manner.

III. THREE USE CASES AND A COMMON DLDS

Three UML use cases were developed to inform the software design, engineering and development efforts. An important design constraint reflected in the modeling was that a common DLDS needed to be able to support all three use cases. Each use case consisted of a brief description (title, model name, brief, story, and example), actor descriptions, preconditions, basic flow of events, alternative flows, key scenarios, post-conditions, special requirements (availability and performance metrics), and a description of the “Dynamically Enforce Privacy and Access Control” extension.

A. First Use Case (Fusion Center Centric)

The first use case considered information sharing within the same jurisdiction. In this case, the actors sharing the information and the subject that the information pertains to are covered by the same jurisdictional laws, regulations, and policies.

This use case describes how authorized actors working in the same fusion center jurisdiction would share information, while protecting sensitive and Personally Identifiable Information (PII), by extending existing access control mechanisms. The extension is called “Dynamically Enforce Privacy and Access Control” as shown in Figure 2. A liaison (primary actor) requests information that includes protected information that was previously uploaded to a jurisdictionally-shared data store by an analyst (secondary actor).

This project is sponsored by a Science and Technology Directorate Long-Range Broad Agency Announcement for the Homeland Security Advanced Research Projects Agency, Department of Homeland Security.
The information uploaded by the analyst was based on fact information that the analyst had received, gathered, analyzed, vetted, and properly formatted before sharing. The liaison and analyst (primary and secondary actors) belong to different organizations within the same fusion center jurisdiction and their membership in a COI is not a consideration. Based on access control and privacy policies being enforced for information requested by someone from a different organization within the same jurisdiction (independent of Community of Interest), the requested information is either denied or provided, and if provided, the PII is either visible, anonymized, masked (encrypted), or redacted.

B. Second Use Case

The second use case considered information sharing when one of the actors is in a different jurisdiction and both actors do not belong to the same COI. COIs are common in the national security community and represent a group of users who share common goals, interests, missions, or business processes. Significantly, national security COIs often cross jurisdictional boundaries and are based on formal agreements necessary for the meaningful exchange of information. For example, [11] lists HSIN chartered COIs that are regional, national, and international in scope.

The second use case is similar to the first except the primary and secondary actors belong to different organizations, are in different jurisdictions, and are not members of the same COI. In this use case, the subject to whom the information pertains and the analyst (secondary actor) who received, gathered, analyzed, vetted, and then shared the information are covered by the same jurisdictional laws, regulations, and policies. The liaison (primary actor) belongs to a different organization and is performing a search request from a location that is covered by different jurisdictional laws, regulations, and policies. As a result, a request that was permitted in the first use case might be denied in this use case. In addition, even if the request is permitted in this use case, data that may have been visible in the first use case may be redacted from the search results in this use case. These different results are achieved using the same ABAC and DLDS solutions. Different organizational and jurisdictional policies may be invoked, and the attribute values used by the access control and privacy policies may be different, because the primary actor is from a different jurisdiction and is a member of a different COI than the secondary actor.

C. Third Use Case

The third use case considered information sharing when one of the actors is in a different jurisdiction, but both actors belong to the same COI.

In this use case, common membership in an authorized COI may require a different result than in the second use case. A request that was denied in the second use case might be permitted in this use case. In addition, data that may have been redacted in the second use case may instead be provided but masked (encrypted) in the search results for this use case. This allows members of the same COI to decrypt this information using symmetric or asymmetric algorithms. Again, different results are achieved using the same ABAC and DLDS solutions. Different policies and attribute values may be used because, although the primary actor is from a different organization and jurisdiction, the primary and secondary actors belong to the same COI.
D. A Common DLDS

Figure 3 shows how ABAC and DLDS can be modeled to support all three use cases allowing a common reusable solution for different types of policy sets. In order to operationalize policy-based, privacy-enabled, fine-grained dynamic access control, the Server Application in Figure 2 is augmented by adding a Policy Enforcement Point (PEP) that intercepts requests and responses in order to act on them. The same PEP is shown again near the top of Figure 3 overlapping the ABAC and DLDS portions of the “Dynamically Enforce Privacy and Access Control” model diagram. The request is first sent to the ABAC portion shown on the left, which decides if the search request will be permitted or denied. If the request is permitted, the DLDS portion on the right decides which data segments in the response, if any, need to be anonymized, masked, or redacted.

In each portion of the “Dynamically Enforce Privacy and Access Control” extension, policies are evaluated to make a decision and the decision is enforced by the PEP. Based on the use cases, three types of policies need to be evaluated: organizational, jurisdictional, and COI. The policies being evaluated are in a machine-readable format. For example, the ABAC policies may be formatted in eXtensible Access Control Markup Language (XACML) and the DLDS policies may use the same format or a different one.

At the time of a request received at the PEP, a policy decision may require fetching contextual information about the requestor and the environment in the form of attributes. The attributes may be fetched from several authoritative sources including the local network and federated services such as the NIEF and BAE.

IV. CONCEPT OF OPERATIONS

The DLDS Concept of Operations (CONOPS) identifies the end users, describes their skill levels and environment, provides a description of the DLDS system from the users’ and stakeholders’ perspective, describes how the DLDS will be used in the field to accomplish the intended mission and how the DLDS CONOPS relates to other systems, information sources and destinations, and other relevant CONOPS.

A. Related CONOPS

During the literature review portion of the research, the team reviewed a number of documents related to different fusion center’s unique scope and mission; the operation, policies, and constraints of different fusion centers; and how fusion centers interact with their mission partners. As a result, the DLDS CONOPS was written with the following selected CONOPS in mind: [12], [13], [14], [15], [16].

B. DLDS CONOPS Overview

There are two primary users of the DLDS system. The analyst collects fact information at the fusion center, creates a report, and approves the report for sharing. The liaison requests information from the DLDS protected shared repository. Analysts are not only skilled in recognizing and correlating relevant fact information, they are also trained to recognize, vet, and approve sharing information that may be sensitive, controlled, and auditable. Liaisons are also trained to properly search shared repositories. Analysts and liaisons have authentication credentials based on a unique identity and attributes that are electronically available to determine their organizational membership(s), their jurisdictional boundaries, and their Community of Interest (COI) membership(s). The DLDS can operate on premise or hosted as a cloud service. If on premise, the DLDS operates between the PEP and the shared repository in each fusion center, allowing each of the fusion centers to protect P/CLCR according to that fusion center’s organizational and jurisdictional requirements. If operated in the cloud, the DLDS serves the same function in a multi-tenant environment recognizing that each fusion center may be operating in the same or different jurisdictions, and therefore governed by the same or different jurisdictional laws, regulations, and policies.

C. DLDS in Mission Accomplishment

The combined ABAC and DLDS system improves the ability of analysts to securely share information and for liaisons to access the right information while maintaining the protection
of citizens' privacy, civil rights, and civil liberties. ABAC provides mechanisms to enforce fine-grained dynamic user authorizations based on changing user and/or environmental attributes; manage the federated ingress and egress of data between fusion centers to ensure information sharing does not exceed thresholds for data sensitivity and classification; monitor service level and data use agreements; and ease the management of privileges to manage policy and administer the DLDS. This dynamic data driven approach is heavily influenced by NIST guidance including [17], [18], and [19] and designed to augment the fusion center’s information assurance and risk management capabilities. The DLDS makes per transaction decisions in milliseconds at the time of the request and incorporates user credentials, user attributes, jurisdictional information, community of interest information, and information about the data being requested. The ABAC and DLDS also account for various network and data classification levels needed to support the federated data exchanges.

V. HIGH-LEVEL ARCHITECTURE

In the interests of a robust, modular, and flexible solution, the data labeling and segmentation is being applied in real time as the data is leaving the jurisdictional or organizational boundary. This approach allows for data labeling and segmentation policy changes to be applied independently of the applications that create and manage the underlying data. For large databases, this also allows for policy changes to take effect in real time without requiring an update to the underlying data through a batch job.

A. Federation and an Identity Trust Framework

Figure 4 shows the Identity Trust Framework providing single sign on (NIEF) and federation identity services (BAE) allowing the identity of a single user and their attributes to be passed around multiple system components. The ability to provide identity and attributes in a secure and trusted manner is a key enabler for executing access control and data segmentation policy on a structured document. The federated model allows for users to be managed by their home organizations without requiring a shared monolithic or virtual directory. The use of federated identity systems opens up system services to a wide variety of users while also enabling enhanced security enforcing attribute based access control.

To fully exploit the ability to change policy rules in real time and to provide continuing access control to privacy elements even after documents have been shared, fields can be anonymized, masked, or redacted based on metadata in the structured document. The use of encryption to mask data allows for the P/CRCL to be continuously protected according to the rules of the organization that is providing the data while also allowing for decryption in selective cases under control of the providing organization. In order to achieve these properties, data segments can be encrypted and decrypted using a key in the DLDS trust store and a bundled crypto service. If the encrypted information is needed by a different user, or at a later date, they can then “click” on the encrypted information in the document to request it be decrypted. The user will then be prompted to authenticate using a NIEF token. If the user is allowed access to the encrypted information, the field will be decrypted and the decrypted text presented to the user. The DLDS service may also retrieve new or additional attributes about the user at the time of the request using the BAE.

B. Interoperability across fusion centers

An additional consideration identified during the requirements gathering phase at fusion centers in Texas and California is different fusion centers use different tools to gather, process, analyze, and share information. For example, three tools that the research team identified in use in different fusion centers are Trapwire®, Palantir®, and IBM® i2® Analyst’s Notebook®. In order to share information between these proprietary tools, the data must be transformed to an interoperable data schema. The data will be exported into a NIEM format using the Law Enforcement Information Sharing Program Exchange Specification (LEXS) as a wrapper for the data being exchanged and converting proprietary object types and values to NIEM-conformant IEPD Suspicious Activity Report (SARI) and Infrastructure Protection (IP) elements. This data schema and wrapper allows for the labeling of privacy data and attaching customized display instructions that can be leveraged by the DLDS system. The attachments and display instructions can then be used to communicate usage constraints and obligations to the recipient.

VI. KEY FINDINGS

The development of the prototype revealed some key findings. First, the LEXS schema has been exceptionally well designed. Very complex requirements have been elegantly thought of and accounted for in the existing schemas. The fact that no changes to the schema have been deemed necessary is a testament to the maturity of the specification.

However, allowing for the attachment of an XSLT stylesheet transforming the content into HTML for display in a browser poses an interoperability challenge for DLDS. Attaching an XSLT stylesheet is useful if the content contains any XML extensions that a standard stylesheet would not recognize. However, an attached stylesheet is also needed by the DLDS solution to display labels from the DLDS operation and changes to content to enable encryption/decryption workflows. When implementing DLDS in the context of LEXS messages, it will be important to analyze whether or not it is safe for the DLDS to add its own style sheet. If it is not safe, the display of custom labels will need to be handled in other ways.
VII. CONCLUSIONS

This research developed an architecture, CONOPS, and initial prototype to augment fusion centers’ existing systems with policy-based, privacy-enabled mechanisms allowing organizations to monitor, audit, and commercialize the solution.

VIII. REFERENCES


