Securing Humanitarian Information Exchange: A Mediator-Wrapper Architecture

Abstract

Reliable and secure information exchange, which is crucial for successful response to crisis by humanitarian organizations, requires the responding groups to swiftly organize themselves in new and dynamic ways. Within these resulting impromptu structures, planning, negotiation, and coordination poses significant problems, due to the heterogeneity of the technologies in place. A plethora of technical solutions have been proposed to solve information exchange issues. However, they thought of security as an ad-hoc, especially authentication, authorization, and access control. This paper proposes a conceptual platform, the Secured Humanitarian Information Sharing Architecture (SHISA), that enables heterogeneous humanitarian systems to exchange information while considering authentication, authorization, and access control. SHISA standardizes communication through the exchange of encrypted XML documents. It uses the Privilege Management Infrastructure (PMI) for authentication and authorization. The platform utilizes the mechanisms of indexing and impersonation to control data access so that humanitarian organizations’ users access only the information they need.

Keywords


Introduction

Effective communication is essential for successful coordination and response to crisis, whether natural or man-made. It helps in saving money and time that humanitarian organizations might otherwise spend on duplicate efforts (Huesemann 2006). Effective information exchange also helps humanitarian agencies build partnerships, increase trust, and allows the sharing of resources, which are indispensable characteristics of constructive coordination in response to crisis (Wakolbinger et al. 2013). Most of all, information and communication during crisis play an important role in logistics, organizational learning, health-care delivery, and other humanitarian services (Haselkorn and Walton 2009).

According to King (2005), humanitarian information can be categorized into: 1) situational awareness information which reports the on-the-ground situation in terms of condition of crisis, needs, location, and severity. 2) operational information which help planning and implementing assistance programs such as the stakeholders, the financial reports, among others. 3) background information about the crisis country’s culture, geography, population, and political structure. Finally, 4) analysis information that include evaluations of causes, constraints, efficiency factors, and recommendations.

Problems in humanitarian organizations’ communication are categorized as cultural issues that stem from the environment surrounding the organizations, organizational issues that are related to the operational level inside the organizations, and technical issues that emanate from the variations in the capabilities of the technologies in use by the organizations. Cultural and organizational issues are critical to overcome and troublesome due to the complexity of changing people habits and presuppositions. Technical issues however, can obstruct information exchange utterly. Therefore, we argue that technical issues require to be prioritized due to their inhibiting nature.
In this paper, we focus on the technical issue of heterogeneity of the technologies in place, which hinders the exchange process and access control mechanisms. Heterogeneity may stem from technologies are running on different operating systems, data based on different formats, or software programmed in different programming languages, among others (Bouguetaya et al. 2012). The lack of standardization can cause interoperability issues; furthermore, controlling access within this environment is challenging since it relies on the management software. For example, one organization may use document management software while another organization uses relational database software. The former allows access control at the file or document level, whereas the latter controls access at the individual data element or field level (Lampel et al. 2013). Although technologies that allow heterogeneous systems to communicate have been proposed in the literature, to the best of our knowledge, security and access control have not been studied explicitly. Therefore, the objective of our paper is to develop a platform that allows heterogeneous information technologies to communicate and exchange just the right amount of data required by each humanitarian stakeholder to achieve its tasks. Our proposed system is a loosely coupled mediator-wrapper framework that incorporates mechanisms for standardizing the information to be exchanged through XML. It also authenticates and authorizes participants through the Privilege Management Infrastructure (PMI) (Chadwick et al. 2003) and restricts access for humanitarian information through indexing and impersonation (Lampel et al. 2013).

We address the following research questions: What technical architecture can enable heterogeneous humanitarian organizations to effectively exchange information? And what access control techniques can be utilized so that the organizations only access the information they need in order to make the right decisions and to achieve effective crisis response? We use a design science approach, focusing on solving a problem identified in the literature (Peffers et al. 2007). Defining the problem is the first step in the design, followed by illustrating the objective of the designed artifact (Hevner et al. 2004). Thus, the following section gives the contextual background of the humanitarian information exchange by pointing out the information sharing impediments. The section also presents the architectures for integrating heterogeneous technologies proposed in the literature and their drawbacks. The issues discussed in the section comprise the building blocks for our proposed framework. In the section following, we describe our artifact – the Secured Humanitarian Information Sharing Architecture (SHISA), which synthesizes and builds upon suggested solutions in various silos of the literature. The final section provides some discussion and conclusions.

**Research Background**

Humanitarian information sharing issues can be classified as organizational, technical, and cultural. Technical problems are related to issues in infrastructure (i.e. lack of adequate infrastructure at the time of crisis in the area in which the crisis is taking place), the heterogeneity of information systems (i.e. issues of interoperability of technologies and data structures), deficiencies in information availability (i.e. processing data, consolidating data from different sources, and effective access control), the differences in security level requirements between the organizations involved, and the different levels of information required by the stakeholders. Some organizations may require all available information, while others may need on specific elements. Hence, access should be controlled so that an organization receives all the information it needs and only the information it needs to accomplish its tasks (Haselkorn and Walton 2009; Celik and Corbacioglu 2010; Altay and Labonte 2014; Clark et al. 2015). The disaster response to the 9/11 terrorist attacks in New York City, which involved 1,067 governmental and non-governmental organizations, illustrates these technical difficulties. Incompatible radio systems hindered the information flow in the critical hours following the attacks (Singh et al. 2009, p. 285). Consequently, “in the three months following that event, the death toll went from 6,000 to 3,040, and most of the reduction has been traced to … ‘duplicate reports and confusion in the hours and day immediately following the attack,’ which for our purposes, corresponds to multiple databases and inconsistencies in reporting and updating information” (Phillips et al. 2002, p. 88).

The literature offers various architectures to integrate heterogeneous information systems, classified into layered architectures (e.g. Da Xu et al. 2014; Sanaei et al. 2014; Botta et al. 2016), extension architectures (e.g. Li 2010), and loosely coupled architectures (e.g. Stahl et al. 2013; Lin 2014; Guerrieri et al. 2015). In a layered architecture system, a technology of one type operates over a technology of another type (see Figure 1a). Lower layers technologies provide services to the higher layers. However, higher layer technologies are different and independent of the designs of the lower layer ones (Raghavan and Garcia-Molina 2001). Accordingly, the higher layer technologies employ the lower layers to facilitate concurrency.
control, recovery, caching, index structures, etc. Although layered architectures reduce the need of extensive development time and effort, mapping the data types and operators used by the higher layer technologies to the data types and operators of the lower layer ones is problematic (Raghavan and Garcia-Molina 2001, p. 45).

Figure 1. Common Integration Architectures (Adopted from Raghavan and Garcia-Molina, 2001)

Extension architecture systems are proposed as a solution where the system uses extension modules to enhance the capabilities of the heterogeneous technologies (see Figure 1c). These extension modules provide support for new data types, operators, or query languages that are available in other technologies. In humanitarian organization settings, mapping the data types and operators or adding the extensions are difficult due to the number of organizations involved in the collaboration process and the dynamicity of the coalitions of humanitarian organizations.

Loosely coupled architectures separate the integration process with an isolated integration layer (see Figure 1b). The integration layer provides an integrated access interface to the connected technologies by utilizing special data and query languages. This model does not require complex modifications to the individual technologies. However, it necessitates developing efficient mechanisms to translate the queries expressed in the integration layer for each individual technology. Our choice of the loosely coupled architecture is due to the presumption that the SHISA is ubiquitously available on the time of crisis. It just requires humanitarian organizations to register their users and gain their users' certificates to start communicating among each other.

The Secure Humanitarian Information Sharing Architecture

General Approach

Our proposed artifact comprises a loosely coupled mediator-wrapper that allows the heterogeneous humanitarian technologies to exchange qualitative data and reports. The mediator (the integration layer) provides an abstract integrated view of the data to be exchanged. It acts like the centralized location from which users can query documents. Query responses are provided as secured XML documents that can be mapped to the organization’s local database system automatically (Huesemann 2002). At the humanitarian organization side, the technical implementations are simpler. Therefore, our framework solves the issue of requiring special technical expertise and expensive technologies to collaborate in response to crisis. In terms of mechanisms to translate the queries for each individual technology, our artifact uses a standardized XML schema and enforces access control by using mechanisms of indexing and impersonation as discussed below.

In humanitarian organizations, heterogeneity of technologies is commonplace. For example, one organization might have its data in a relational database and use SQL to tap into the data, while another
organization is using a document management application from which users can access the documents according to their tasks. Furthermore, collaborating organizations may employ multilevel security policies or multilevel security goals. These diverse systems and policies provide obstacles to the mediating system that is used to aggregate the data. For example, the file system will control access at the file level while the relational database will control access at the data element or field level. Consequently, a unified access control model is necessary.

The data aggregator level uses a search engine that crawls each organization’s databases. In order to provide a quick and relevant response to authorized users when inquiring documents, the search engine will index the documents that are to be shared by the humanitarian organization. Once a user submits a query, the search engine will refer to the global index instead of the actual document. This is analogous to searching in a catalogue instead of searching through the actual items (Mudgil et al. 2013).

The documents in the index have varying permission levels. An approach to accommodate this variety of permissions can be to have a global permission set that is identified at the search engine level. This is difficult because it requires a global level administrator to define the access rights for the individual catalogued document. In humanitarian environment, this approach is impractical for three reasons: 1) applying a common permission set to documents among heterogeneous systems is impossible; 2) it requires a duplicate effort of redefining an already defined permission set on the documents collected; and 3) it requires the admin to have knowledge of access control for each organization involved, which is inconceivable (Lempel et al. 2013). Accordingly, the search engine should honor the documents’ permissions set that is identified in the original database. This is achieved by using the original access control list (ACL), which is a table that lists the documents, the project/task that the documents are generated for, and the users’ or groups’ permission on these documents.

A technique for honoring original ACL in our framework is impersonation. Impersonation in this context means that once a query is submitted, the search engine will contact the original database server to check if the requesting user has permission to view the file before returning the results. This helps in reflecting the most recent ACL permissions since the filtering of the query results happen in real time (Lempel et al. 2013). The following two subsection illustrate SHISA in more detail. The first subsection introduces the information exchange standard (IdmlReporting), and the second discusses the data aggregator and its content.

**The Exchange Standard: IdmlReporting**

The standard (IdmlReporting) is used to suffice the information needs of the different technologies and to standardize the access control levels. IdmlReporting (see Figure 2) is based on the *International Development Markup Language (IDML)*. IDML is already in use for qualitative (project) information sharing by organizations such as the *World Bank*, the *Organization for Economic Co-operation and Development (OECD)*, and the *United Nations (UN)*. IDML is based on XML and uses special tags and rules for humanitarian projects. The core activity schema includes project information such as the project titles, the project stakeholders and their roles, the people involved, and funding details. Some of these elements are generic and apply to any project, but others are specific to international development projects such as funding and organizationsInvolved. Funding, for example, can come from different donors and can go to different organizations, as they are involved in the humanitarian project. Accordingly, IDML attributes make it different from profit-oriented projects (Huesemann 2006).

Although IDML was designed to allow the exchanging of high-level descriptions of activities, it cannot give detailed project reports and evaluations. Therefore, idmlReporting schema was proposed, which includes four major reusable components: 1) Detailed Description, 2) Reporting, 3) Evaluation, and 4) Financial Information. Our framework adds a root to idmlReporting, the projectId, since idmlReporting generally allows reporting on one project, whereas our platform is generic for all humanitarian projects and allows multiple projects to be reported at once. The idmlReporting is rooted with the reportsAndEvaluations as the container for the rest of the schema elements. This means that each project will have the main element reportsAndEvaluations, which comprises limitless reportAndEvaluation elements. Each reportAndEvaluation element is a complex type reportType, which can contain mixed contents (broken frame in Figure 2). The reportType includes detailedDescription, evaluation, reporting, financialInformation, and metadata on the report (e.g. number, name, date, description) (Huesemann 2006).
detailedDescription further contains detailed information about a project like background, expectedOutcome, beneficiary, objective, problem, risk, strategicApproach, milestone, and relatedDocument (see Figure 3). The list of complex elements and their components can be found in the detailed documentation for IDML and IdmlReporting provided by Huesemann (Huesemann 2017). Figure 2 and Figure 3 are just snapshots of the schema to show that IdmlReporting contains all the elements/keywords that represent the required data for collaboration among the organizations. The processes of querying and controlling access in SHISA are managed on the fine-grained level of these elements. Therefore, IdmlReporting helps in achieving decomposing a document into those elements and querying only some parts of the document.

The collaborating humanitarian organizations will map their documents to the idmlReporting schema. This can happen automatically if the organization has a relational database (Huesemann 2002). Otherwise, the crawler of the search engine will have a programmed mechanism to go through the documents and find the keywords that match the elements specified in the schema. The details of code are out of the scope of our paper.

In summary, the proposed XML schema will act as the set of comprehensive standardized data elements to be exchanged. IdmlReporting predefines all the elements/keywords that represent the data required for collaboration among the humanitarian organizations. For relational databases, the predetermined elements are mapped to the tables’ fields. Conversely, for document management applications, the
predefined elements will act as the keywords for the search query used for decomposition. Through this process, the access controls that were originally at the file level in the organization’s local database will be converted to the element level at the mediator, thus standardizing the system.

**Architecture of the Data Aggregator**

The data aggregator side represents the global level mediator that facilitates the collaboration and information exchange among the organizations. It includes an attribute authority (AA), a global role manager, search server that includes a global search engine and a search index database, a database that hosts the responses to queries, and the XML generator that translates the documents into idmlReporting schema (see Figure 4).

In order for a user to request information from the data aggregator, the user must be authenticated and authorized to ensure an end-to-end secured communication. A standard that incorporates mechanisms of both authentication and authorization is the 4th edition of X.509, referred to as the Privilege Management Infrastructure (PMI) (SANS Institute 2001). PMI attaches an entity with privileges through Attribute Certificates (ACs). As shown in Table 1, an AC includes information about the version of the certificate, the certificate holder, the issuer (the Attribute Authority AA), a serial number to identify the certificate, a unique identifier of the AA, attributes associated with the holder, extensions (optional) to add information to the AC, and the signature of the AA to ensure the validity of the certificate. In our framework, the extension field holds the projectId element to show for which project this AC is generated and to which project this user is assigned this role. The extension (projectId) is needed because, at any point, an organization can be collaborating for different projects, and in each project a user might have a different role. This is due to the dynamic nature of cooperation between humanitarian organizations.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td></td>
</tr>
<tr>
<td>Serial Number</td>
<td></td>
</tr>
<tr>
<td>Signature ID</td>
<td></td>
</tr>
<tr>
<td>Holder (user) X.500 General Name</td>
<td></td>
</tr>
<tr>
<td>Issuer (AA) X.500 General Name</td>
<td></td>
</tr>
</tbody>
</table>

![Figure 4. SHISA Architecture](image-url)
Verify Period
Attribute (role)
Extension (project ID)

Table 1. Attribute Certificate (AC) Content

For trust purposes, PMI requires an entity with the ultimate authority to be responsible for assigning privileges and will be trusted across the whole system – the so-called Source of Authority (SOA), which is the AA in our model. Since our framework connects disparate organizations that are in different geographical areas, the role management is a hierarchical model of role assignment. At the data aggregator, the role manager is responsible for creating standard roles with standard permissions per schema element. Permissions are “visible and edit.” Visible specifies the visibility of the element to the role, and edit specifies if the role can edit the element information or not. For example, from Table 2, the financial analyst can see and edit risk. Although milestone is an element of detailedDescription, it is set as invisible and is consequently not editable. When access control is enforced, the system will use the lowest level of the element’s (i.e. milestone) permission and will override the higher-level permissions (i.e. detailedDescription). Therefore, the financial analyst will see risk information but not milestone information when requesting a detailed description of the project Hurricane Katrina (see Tables 2 and 3).

Standard roles and their privileges are identified by the global role manager, and each organization assigns its users locally to the standard roles to define their privileges. Once the user is assigned to a role, the organization’s role manager contacts the AA to register the user and to obtain a user certificate. Obtained certificates are saved locally in an LDAP directory so that certificate revocation lists become unnecessary. Simply, a revoked certificate will be deleted from the LDAP folder. The choice of the Role Based Access Control (RBAC) is because of its ability to simplify controlling the access when the number of users is large or when the collaboration is dynamic (i.e. stakeholders can join the project or drop from the project during the project progress).

<table>
<thead>
<tr>
<th>XML Schema Element</th>
<th>Visible</th>
<th>Edit</th>
</tr>
</thead>
<tbody>
<tr>
<td>detailedDescription</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>milestone</td>
<td></td>
<td>•</td>
</tr>
<tr>
<td>risk</td>
<td>•</td>
<td>•</td>
</tr>
</tbody>
</table>

Table 2. ACL for Financial Analyst Project1

Subject constraints are per file in the local ACL and per element in the global search index ACL. As shown in the tables, although the financial analyst can access DOC1 of project1, s/he will not be able to access the milestone information within that document. Since the global (element) ACL does not allow this role to access the element based on the profile privileges specified for project1. The financial analyst is also not allowed to access the risk element of DOC1 in project2 because of their profile. Although the access is granted at the local level, it is not granted at the element level globally. The ACL represents granting the access action only. When access is not explicitly granted, it is assumed to be denied. This embodies the implicit deny rule in security (Carlos et al. 2001).

<table>
<thead>
<tr>
<th>Document ID</th>
<th>Project ID</th>
<th>Issuing Organization</th>
<th>Roles allowed</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOC1</td>
<td>Project 1 (Hurricane Katrina)</td>
<td>Red Cross</td>
<td>administrator, financialAnalyst</td>
</tr>
<tr>
<td>DOC2</td>
<td>Project 1 (Hurricane Katrina)</td>
<td>Red Cross</td>
<td>administrator, projectMngr</td>
</tr>
<tr>
<td>DOC1</td>
<td>Project 2 (Haiti development)</td>
<td>Red Cross</td>
<td>administrator, financialAnalyst</td>
</tr>
</tbody>
</table>

Table 3. Local Organization's ACL

<table>
<thead>
<tr>
<th>Element</th>
<th>Project ID</th>
<th>Roles allowed</th>
</tr>
</thead>
<tbody>
<tr>
<td>risk</td>
<td>Project 1 (Hurricane Katrina)</td>
<td>administrator, financialAnalyst, projectMngr</td>
</tr>
<tr>
<td>milestone</td>
<td>Project 1 (Hurricane Katrina)</td>
<td>administrator, projectMngr</td>
</tr>
<tr>
<td>risk</td>
<td>Project 2 (Haiti development)</td>
<td>administrator, riskAnalyst.</td>
</tr>
</tbody>
</table>

Table 2. ACL for Financial Analyst Project1

Subject constraints are per file in the local ACL and per element in the global search index ACL. As shown in the tables, although the financial analyst can access DOC1 of project1, s/he will not be able to access the milestone information within that document. Since the global (element) ACL does not allow this role to access the element based on the profile privileges specified for project1. The financial analyst is also not allowed to access the risk element of DOC1 in project2 because of their profile. Although the access is granted at the local level, it is not granted at the element level globally. The ACL represents granting the access action only. When access is not explicitly granted, it is assumed to be denied. This embodies the implicit deny rule in security (Carlos et al. 2001).
Table 4. Element Level ACL

As mentioned above, the role of having a search engine is to retrieve the documents from an organizational database and to provide relevant responses to search queries submitted by users. For performance purposes, the search engine builds an index which embodies the corpus of documents that the humanitarian organizations are willing to exchange. Furthermore, the index facilitates avoiding duplicate documents.

To build the index, the search engine uses a crawler to scan the documents in the collaborating organizations’ databases with the role of administrator (or super-user) that is sufficient to access all the repositories. The crawler not only retrieves the content information of a document, but also retrieves metadata about the document such as author, project ID, and project name, among others. This information will be parsed, tokenized, and ingested by an indexer into the index. If the organization uses a relational database, the organization’s fields will be mapped automatically to the idmlReporting schema elements. However, if the organization uses another file management system, the crawler will add another step to searching the text in the document; it will map the document content to the schema elements using a text to XML mapping software. Once the documents and their contents are retrieved, the search engine stores cross reference of the document’s elements in the index data repository. This process of crawling is periodic so that the index repository is always up-to-date with the documents that are in the organizations’ databases.

**SHISA Functioning**

The variety of management packages makes it hard, if not impossible, to come up with a common access control structure. SHISA honors the native original access control policies, but globally it controls the access on the document elements based on users’ profiles.

A collaborating user must be registered in the system and have obtained a certificate specifying their role. When a user submits a query, the search engine matches the query keywords with the indexed elements and generates an interim pre-filtered search result. Once a document is part of a search result, the search engine contacts the originating database with the requesting user’s role to check if the role is authorized. The organization’s server responds to the search engine with an answer of granted or denied access. If denied, the document will be eliminated from the interim result, however, if access is granted, the search engine will then check ACL per each document element. If access to a document is granted but an element is set to not visible, then the document will be returned with the element eliminated. The process of generating the index is what we refer to as indexing, and the process of contacting each organization to filter the interim result based on the user’s role is impersonation.

Once a search result document and its elements are finalized, the resulting documents will be requested from the owner organizations’ databases. At the data aggregator level, the documents will be parsed to the final elements that the requesting user is authorized to retrieve. An XML generator then uses mapping and conversion rules to map the information to the idmlReporting schema to generate an XML document. The document then is encrypted using the public key of the local web server of the organization that represents the requesting user. Finally, the encrypted document is sent to the user.

For the search engine to implement impersonation, the user must present their AC along with the query. The user logs into the local organization’s web server, the web server authenticates the user, and contacts the local role manager to retrieve the user’s certificate from the certificate repository (the LDAP certificate server). Once the user is authenticated and the certificate is retrieved, the local web server will act as the only entry/exit point to and from the organization and will submit the query to the global search engine on behalf of the user. The local web server also attaches the requesting user certificate for the search engine to perform impersonation. The reasoning behind placing certificates in a centralized certificate repository to each organization is to reduce the probability of a certificate being tempered or stolen due to users’ unintentional security threats. Furthermore, the reason of using a single entry/exit point at each organization is to reduce the complexity of key pair management. Instead of having thousands or more of key, each organization will have only a single key pair to be managed in SHISA.

**Conclusion**

Secured information exchange among humanitarian organizations is essential when responding to crisis. Yet, effective communication faces challenges resulting from the heterogeneity of the technologies in place and the dynamicity of the teams. Heterogeneity in this context means that organizations are using
different technologies, different standards, software, protocols, and security levels, among others. Furthermore, coalitions need to be set up swiftly and they are dynamic in that cooperating organization can join or withdraw during the project progress based on their tasks. Consequently, preparation and planning among the organizations is challenging.

Issues that result from heterogeneity and dynamicity can vary, among them are the necessity of standardizing the data elements that need to be exchanged, authenticating and authorizing users to ensure an end-to-end secure communication, and standardizing the level of access control to the data elements. This paper provides a pre-configured platform (SHISA), which humanitarian organizations can utilize to collaborate promptly when responding to crisis. The artifact standardizes the data elements using a predefined XML schema, idmlReporting. For authentication and authorization, the artifact utilizes attribute certificates specified in the 4th edition X.509 standard. Attribute certificates not only validate a user’s identity but attach their roles as well, so that the system can authorize the users before responding to their information queries. Finally, to enforce the access control policies, our artifact uses the processes of indexing and impersonation. Indexing refers to the process of going through each organization’s database and index all their documents corresponding to the standardized elements (keywords) specified in idmlReporting. Impersonation refers to the process of honoring the native access control policies specified at the originating database.

Our artifact closes the gap in the literature of not only facilitating the communication among heterogeneous systems, but also securing the communication, which is as important as facilitating the communication itself. It is also beneficial for humanitarian organizations in solving the issues of secured communication within the limited resources available in place based on available technologies that can be accommodated without requiring unique technical skills.

So far, we progressed through the first three stages of design science as presented by Peffers et al. (2007) that are: (1) identification of the problem and motivation, (2) defining the objectives of the artifact and solution, and (3) presenting the conceptual artifact that synthesizing solutions to the individual problems listed. Our plan is to complete the process in future work by demonstrating detailed scenarios of the artifact and show how it solves the problems mentioned. We then will evaluate the model using Delphi Survey methodology as guided by Hasson et al. (2000).

REFERENCES


