HumanID with Composable Capability on Demand

Gail Hamilton, Keith J. Miller, Andrew P. Founds,
MITRE Corp., McLean, VA,
ghamilton@mitre.org, keith@mitre.org, afounds@mitre.org

Abstract— *In this paper we describe a prototype Composable Capability on Demand (CCoD) system with a focus on disaster relief and non-combatant evacuation, and the integration of a HumanID component into that system.*

We will describe the basic tenets of a CCoD system and the methodology we used to create this system. We will describe a middleware component, HumanID, that was integrated with little effort into the CCoD framework.

The HumanID capability described in this paper provides a middleware component that enables search and management of biometric and biographic person identity data across disjoint data sources. We discuss our technical approach and challenges encountered in creating this middleware layer, as well as the robustness needed in a search capability designed to compensate for varying degrees of quality of identity attributes across data sources.

Having described the CCoD and HumanID components, we go on to discuss our strategy for a rapid integration of the HumanID system into the CCoD framework. This integration enabled the composition of identity information about subjects of interest with the other situational awareness information available from other feeds in the CCoD system. For example, a user can search on a subject using biographic information, and the resulting information can then be disseminated and visualized via a CCoD system that may contain further relevant information about the individual from other loosely-coupled components.

We continue with consideration of a use case for this integration, focused on the aggregation and tracking of identity information concerning an individual in a high-security / high-threat zone.

1. INTRODUCTION

The underlying tenet of a Composable Capability on Demand (CCoD) system is that it enables non-technical operators to rapidly combine, adapt, and extend capabilities to respond to evolving threats and mission needs [1]. A CCoD system provides relevant, timely and trusted information to aid cognitive understanding and decision support by drawing mission information from traditional and non-traditional data sources to enhance situational awareness, collaboration, social networking and decision support. It focuses on rapid development of new capabilities, carried out by combining small, reusable chunks of code referred to as components. These components may include elements from traditional programs of record (PORs) and non-traditional, Web 2.0 technologies.

A CCoD component must be interoperable with other components through a 'loose coupler', and can be easily reused or repurposed. Its functionality is self-contained, i.e. it is aware only of its own capabilities and resources, and can function independently of all other components connected to the CCoD framework.

Although CCoD is not, itself, a system, it does require a framework to develop and integrate components. This framework enables the construction of a consolidated view that provides greater situational awareness (SA) than would be possible by viewing the individual systems in isolation.

Situational awareness involves being aware of what is happening around you to understand how information, events, and your own actions will impact your goals and objectives, both now and in the near future. Providing tools for enhanced situational awareness is a key element to improving to Homeland Security.

One of the CCoD components is a MITRE developed proof of concept system called HumanID. HumanID involves accessing multiple disjoint data sources using biographic and biometric information. While the data sources in question may contain varying sets of identity attributes, as long as the types of identity attributes are not pair-wise disjoint among all of the data sources, HumanID enables searching, linking, and aggregation of identity data across data sources, enabling a view of an individual that is not possible within any single data source.

2. SYSTEM DESCRIPTION

CCoD Concepts

A CCoD system consists of two major elements:

(1) User-composable interface.
(2) Loose coupler around which components are integrated.

Composable Interface— A user-composable interface provides non-technical operators with the capability to configure and use a system according to their precise needs. This flexibility can be achieved by providing user-facing 'widgets', which are small, modular components, to allow for complimentary views of the same data or complimentary data about the same scenario.

Loose Couplers—A CCoD system allows for complimentary views of the same data by enabling integration of disparate systems through loose coupling among participants. Loose coupling describes an architecture where integration interfaces are developed with minimal assumptions between two or more sending/receiving parties, thus reducing the risk that a change in one module will force a change in another application/ module [5]. The integration interfaces communicate using information exchange structures which apply to a variety of circumstances. This information exchange structure supports the loose coupling architecture and is referred to as a 'loose coupler' [2]. The definition of a loose coupler is an information object design that optimizes data utilization among a community of independent participants to achieve global cost savings [2].

The characteristics of a loose coupler are that it captures the core common data of a domain. It is therefore optimized to be used by the greatest number of systems but, conversely, is not optimized for use in a single system. A commonly used example of a loose coupler is keyhole markup language (KML) for geographic data. Other examples exist such as CoT (Cursor on Target), a MITRE-developed data format that enables DoD systems to communicate [3]. This data captures the minimal amount of information about an event, the 'what, when, where' information [4] [6]. Other efforts are being made to define more detailed common formats for data formats, notably UCORE for DoD. However, the other component to a loose coupler is that it is 'commonly accepted'. KML became a commonly accepted data format because it was a standard that was used in Google Earth, a ubiquitously useful application for visualizing of geographic data.

One of the issues of a loose coupler is that it is impossible to predict which standard will become commonly accepted. Therefore a CCoD system should allow for flexibility around the actual loose coupler. That is, if another format is more useful, then the CCoD needs to be able to adapt. The flexibility can be achieved by providing user-facing 'widgets' to allow for quick translation between formats.

Choosing an industry standard data format has the additional benefit that most systems may already provide their data in this format. So integration with an existing system is more streamlined.

Case Study: Geographic Situational Awareness about Individuals in a High Threat Zone

The case study used for the CCoD is a NEO (Non Combatant Evacuation Operation) in a hypothetical mid-size African country, with friendly relations with the United States [1].

Noncombatant evacuation operations (NEOs) pose a unique difficulty from a mission assurance perspective. Not only is there substantial variability between operations in the nature of the mission, there are also significant complications resulting from the fact that any NEO will possibly involve both military and civilian organizations, and include foreign entities and multinational/transnational organizations. In addition, the situation on the ground is highly dynamic, involving a high level of threat to American citizens. Up-to-the-minute situational awareness is, therefore, critical in enabling the commanders to make effective decisions.

Often the most up-to-date data being received in such a location is through non-traditional data formats, such as RSS feeds, Twitter, Flickr and other open-source/ Web 2.0 technologies. This information, though vital, is typically not captured in traditional C2 (Command and Control) systems. Using CCoD architecture, through the use of Web 2.0 technologies, each person on the ground can provide additional information to the system.

In this scenario, the subjects fall into two categories, known evacuees and subjects of interest. The evacuees are state employees typically tracked through an RFID tag. In an NEO, a Neo Tracking System (NTS) is typically used to track these signals and gain up-to-date information about known evacuee location. In addition are subjects of interest, people known to be in the area and requesting evacuation to the US, but whose identity and intent has not been verified.

The use of HumanID as a component to the CCoD framework, in this scenario, allows for verification through biometric and biographic identifiers to establish the identity of a potential evacuee. In addition, merging the HumanID component via CCoD using subject name as well as KML as a loose coupler allows for rapid enhancement of the situational awareness with minimal effort. Also, knowledge about potentially hostile subjects known to be in the area could be added to the system by integrating border-crossing information.

Design of CCoD System

The overarching goal of CCoD is to enable transition from a traditional system to a technical standard that emphasizes reusable small, single (or simple) purpose widgets, which can be easily reconfigured by a non-technical user as the situation evolves.

Technical areas that this case study focused on were

- Establishing, and converting all data to a common format.
- Integration of open source tools/ data sources that could be leveraged.
- Reusable mechanisms for data capture from open source data, such as RRS feeds, KML feeds, HTML screen scraping.
- Integration of Programs of Record with nontraditional sources.
- Leveraging of research efforts to show how CCoD enables rapid enhancement of capability.
- Building reusable components that can be easily repurposed for different domains.

Integration of the HumanID component allowed for concentration on the last three of these technical areas.

CCoD Implementation

The system uses an iGoogle-like interface, as shown in Figure 1, but is also implemented in SharePoint, to underscore that the components themselves can be ported to different frameworks.

The display framework is a customizable web-based AJAX interface, and is built upon the JavaScript framework, called Prototype. This framework was called iCCoD.

Like Gadgets for iGoogle, the iCCoD interface allows for CCoD components (widgets) to be added to the system. In this case, the system retrieves a list of currently listed widgets from the MySQL database, and allows for the user to add widgets through clicking a checkbox. The widget is then added to the screen, in a position specified through preferences, and can be removed and re-added to the screen at any point.

To allow integration of data from multiple sources in a format that is domain-agnostic, it was determined that the KML, being an industry standard, is the optimum data format to be used in CCoD. KML is a file format used to display geographic data in an ‘earth browser’ such as Google Earth or Google Maps. KML uses a tag-based structure with nested elements and attributes and is based on the XML standard.

Both traditional and non-traditional data sources can be added to this map, through a mechanism described below. The traditional feeds are typically from Programs of Record (PORs) that use proprietary data formats. Components, which are transparent to the user, are written to provide translations between these custom formats and the common core format, KML. This allows the user access to the SA information but insulates them from needing to know or understand the original data format.

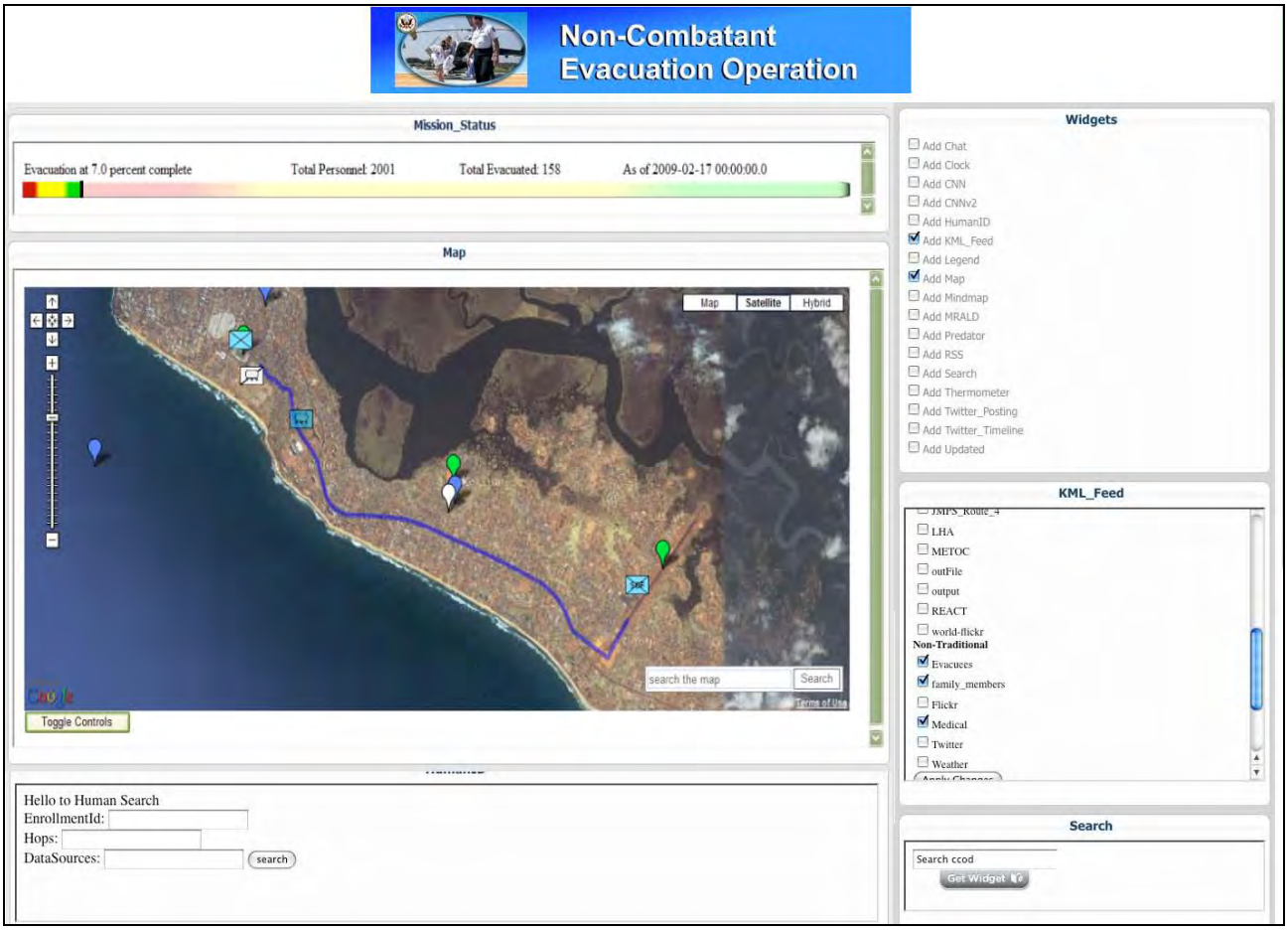

Figure 1 - iCCoD Interface. Showing List of Components and List of Data Feeds

These CCoD KML components are available when the 'Add KML Feeds' checkbox is selected from the list of widgets. A list of KML data feeds is then displayed, also with a corresponding checkbox, to allow for the data to be added dynamically to the map. The non-traditional data feeds were GeoRSS feeds from Web 2.0 sites such as Twitter and Flickr.

In addition, other feeds such as weather and satellite data, as well as evacuee data from the Neo Tracking System, were displayed on the map through KML.

CCoD uses many programs of record currently in use at MITRE, as well as a simulation of a predator feed. These all produce data in multiple disparate formats. One of the data formats used is a MITRE-developed data standard, known as Cursor on Target (CoT) [6]. This standard is based on XML, and is primarily for military systems . It is, therefore, required that the data be transformed from CoT to KML. For the PORs this transformation, takes place within the tool itself, using a plug-in written in Java, running on the servers hosting the PORs.

CCoD then pulls in this information via a KML Listener, written in Python. Data is then parsed to determine the source of the feed, and written to respective KML files within a folder on CCoD for further processing (see section below). Data update rates from each POR vary. Since the data from POR represents real-world scenarios, the data received will become stale over time as the situation changes. This is represented in the system as a 'graying out' of the icons representing elements, such as planes, vehicles or army units. The rate of graying out varies according to the rate of data decay of the corresponding POR.

Initially, standard 'lollipop' icons were used to denote POR elements on the map. These were then changed to standard military icons, defined by MIL Standard 2525, using an XSLT form, to transform the images used.

Data is not only ingested, but can be produced by CCoD, and returned to the PORs. In this scenario, the data representing the subjects are returned. This data is then stored in a MySQL database, which simulates the data from a NEO Tracking System (NTS). NTS is a system used by the Department of State (DoS), to track State employees via RFID tags. In this case 1,500 subjects are used. As 1,500 'lollipops' on the map will quickly overwhelm the display, and also slow performance of the system, it was decided that each 'lollipop' will represent 100 subjects, with a program determining a 'center of mass' of the group of 100 subjects, and returning it for display via KML.

The POR data feeds need to be converted to CoT. CoT

contains meta-data not currently recorded within the database, and not displayed within the evacuee KML. Care is taken to monitor that 'round tripping' of data does not occur, i.e. that subject data ingested by the POR is then returned to the map via the same POR tool, and that this subject data is not returned multiple times by multiple tools [1].

Another system that is used for display is the Navy Meteorology and Oceanography (METOC) system, which displays satellite and weather information, primarily for military use. METOC uses another XML format, Joint METOC Broker Language (JMBL), which also requires conversion to KML to be useful to CCoD. This translation is accomplished by a python program, weather2kml.py. This program makes two SOAP requests to the METOC web service, and parses the information into a KML format. The first SOAP request retrieves images appropriate for display; the second retrieves any weather alerts. The data is then output to metoc.kml for further processing using the file weather_template.kml, as a template for formatting (see below).

CCoD uses the Google Maps API to display the KML information. Initially an attempt was made to use an online version of Google Earth, however, at the time the performance overhead was thought to be prohibitive.

The disadvantage of using Google maps is that the resultant KML needs to be visible to the KML Map rendering service, which resides at Google -- outside the MITRE firewall. Any data produced by CCoD within the MITRE firewall is therefore not visible to the KML renderer. The solution is to use Amazon Simple Storage Services (S3) (or Cloud Computing) to host the resultant KML files. As all data produced is simulated (and is nonetheless encrypted to simulate real-world operating conditions) this solution is thought to not pose any unacceptable security issues.

All data that is published using Amazon Web Service has to reside in a predetermined folder on the CCoD server. A service monitors any files residing within this folder, and on detection of an update, uploads the modified files to the Amazon Simple Storage platform.

If any files are added or changed within this folder, a call is made using Amazon Web Services, which updates the KML that resides on the Amazon Simple Storage platform. As the philosophy of CCoD is to build capabilities rather than systems, CCoD has capability to interact with Amazon S3 in multiple ways, depending upon user preference and system architecture. Thus, the following two methods are provided to publish the KML to the Amazon S3:

- A perl script, kml_upload.pl with a cron file that schedules this script to be executed once every 2 minutes
- A Java object, S3Uploader.java is called to update the data on the Amazon web server if a change in any of the files within the KML directory is detected.

To add a further dimension to the data, we integrate biometric and biographic data into the system, through the KML loose coupler.

Integration of HumanID with CCoD

Border crossing data contains biometric information of the individual, as well as the location of the border crossing. To integrate this data through the KML loose coupler the geographic information is translated to KML format. Hence information about potentially hostile individuals currently present in the area can be integrated into the system, augmenting the capability of the system to provide situational awareness information.

This allows for searches of multiple databases to be carried out on individuals based upon one of the following types of biometric or biographic information:

- Name data
- Finger-print images
- Iris images
- Face images

HumanID Description

A search query submitted from a user is passed to the HumanID middleware, which is used to orchestrate a federated search across various data sources. Each data source participating in the federation has, internal to its infrastructure, a set of name and biometric (fingerprint, face and/or iris) matchers. The matchers are used to process a query, submitted in the form of a name or biometric modality or both, and return the most similar results to the input query. For this prototype system, each data source returns the top results which matched at 90 percent confidence or higher. The results returned to the middleware consist of the pair [name + match confidence] and [biometric(s) + match confidence].

Search results retrieved from each data source are used as new search input terms to other data sources in the federation. This concept is referred to as a "hop". As previously mentioned, the idea is to collect information from each data source and query other data sources using the newly learned identity information to build a more complete representation of the initial identity that was used in the search query. Finally, the HumanID middleware orders the results by data source by decreasing confidence of the match. The search result is returned to the user for displaying.

Limitations

CCoD—There are several limitations in the current state of the practice that limit the immediate widespread use of CCoD solutions.

First, the CCoD approach is limited at the moment by the need for connected, exposed, discoverable, well-understood data sources. Few data sources currently exist that meet these criteria. Often adaptors need to be written by experts of the programs of record, so there is work needed behind the scenes to make the interoperability seamless.

Second, CCoD has dependencies on services of others, such that their infrastructure, performance, and security risk impacts the network and performance of the resulting CCoD system. Further, this is a new business paradigm, which is still gaining acceptance. As with other technology and standards adoption, this will change slowly over time as the concept proves its utility.

Finally, since the loose coupler is optimized to provide the broadest applicability over a maximum number of systems, it is not optimized for any particular system. Hence, some data loss will result on translation to the common core format.

HumanID—The current state of the HumanID middleware is constrained to searching in a serial fashion. The current architecture could be extended to facilitate searching in parallel to reduce the overall query time Furthermore, the restrictiveness of returning results that are matched at 90% confidence or higher may not yield any results for certain data sources. In order to obtain results from these data sources, if applicable, the middleware could be enhanced to enable the user to specify the confidence of matches to be returned from each data source. Enhancing HumanID to overcome this limitation would also require additional work to the matchers which return results to the middleware.

Future Work

CCoD
The CCoD is concept is currently being applied to multiple areas of work within MITRE.

HumanID
Enhanced search capabilities could be extended across the temporal and geospatial dimensions to make more course-grained search queries possible. Additionally, the HumanID middleware could prompt the user during the search to identify which of the results they would like to use in the next "hop" to give the user greater control over the total number of results returned by the system.

3. CONCLUSIONS

The goal of CCoD is to enable non-technical operators to rapidly extend current capabilities to respond to evolving mission needs. Using the CCoD framework, multiple disparate sources can be integrated into a single interface around a loose coupler.

In this case study, this provides enhanced situational awareness in a high security/ high threat zone. The interface allows for non-technical users to include and remove data sources as required. Data from both traditional programs of record, and non-traditional, technologies can be combined together to provide a more accurate understanding of current conditions.

Integration of the HumanID system within the CCoD framework enables two complimentary information sources to be used to enhance situational awareness in this highly dynamic and high threat operation.

REFERENCES

[1] J. High, P. Barry, B. Hirsch, G. Hamilton, D. DeMouplied, M. Heller, C. Jillson, R. McKee, R. Miller, *Composable Capabilities on Demand*, MITRE Technical Report (Working Draft), 2009.

[2] R. Miller, D. Winkowski, *Loose Couplers as an Information Design Strategy*, Military Communications Conference, 2007. MILCOM 2007. IEEE , vol., no., pp.1-6, 29-31 Oct. 2007

[3] A. Rosenthal, L. Seligman and S. Renner, *From Semantic Integration to Semantics, Management: Case Studies and a Way Forward*, ACM SIGMOD Record, 33(4), 44-50, 2004.

[4] R. Cherinka, J. Mathews, R. Miller, D. Pitcher, W. R. Sears and T. Semanchik, *Agile Capability development assessment and Transition in Support of The Global War on Terror*, Military Communications Conference, 2007. MILCOM 2007. IEEE

[5] D. Kaye, *Loosely Coupled, The Missing Pieces of Web Services*, RDS Press, 2003.

[6] *Cursor on Target,* Military Information Technology Online Edition, vol. 8, issue 7, September 2004.