Delivery of Enriched Information to Deployed Forces

Philip Hawthorne, Glen Coomber and Mark Burnett
Military Information Enterprise Branch
Information Technology Division
Defence Science and Technology Organisation

Abstract

Connectivity limitations necessitate careful management in delivering information to deployed forces. Not only must the quantity of information be appropriate to client needs and communication links, but quality, timeliness and assurance must also be maintained. The Australian Defence Force's Theatre Broadcast System (TBS) provides efficient broadcast of data, but much of this data will not be relevant to individual deployed users. PeaceKEEPA Phase 1 (PK1) demonstrated how appropriate information management could deliver a "smart push" capability over TBS.

DSTO is further enhancing TBS with new modes of information delivery to individual deployed clients. This paper describes PeaceKEEPA Phase 2, a joint initiative between DSTO's Information Technology and Communications Divisions. PK2 aims in part to take unstructured textual data (such as situational awareness information) and enrich the basic text with data from other sources and of other types, including reference, imagery, track and geospatial data. The information gathering side of the system employs a federation mechanism called MANIFOLD (**MAN**agement of **I**nformation by **F**ederation **O**f **L**ogical **D**ata). Sets of "enriched" briefs, personalised according to the roles and locations of intended recipients, are prepared in accordance with operational needs. These are then disseminated via shared use of the TBS broadcast medium to implement point-to-point functionality similar to conventional e-mail.

MANIFOLD provides a unified view and single point of access into a wide variety of data sources. This allows deployed clients to make best use of any limited back-link connectivity to formulate queries whose results can be delivered directly or over TBS (according to data volumes). In summary, PK2 helps to reduce the impacts of poor connectivity on deployed forces by codifying the basic information requirements of individual clients, and using TBS to deliver targeted source material with supplementary information of likely interest.

Introduction

Command centres require timely delivery of information relevant to their current situation, with sufficient detail but not too much. With fully-staffed fixed headquarters and adequately provisioned communications, needs can be met… Within an area of operations, with smaller deployed command centres and relatively impoverished communications, the challenges are greater. PK1 [??] demonstrated that deployed clients with a portable satellite receiving dish and laptop PC could automatically select information intended for them out of the bulk data broadcast by TBS. With poor or no back-link connectivity (eg. during EMCON), the question is how to determine what information should be packaged for delivery to individual deployed clients in the absence of specific requests. Some guidance here is that generally clients are most concerned with information relating to their area of responsibility and with supply lines. A further division occurs according to the role of the client such as planning, personnel, logistics, intelligence, etc. Finally, individuals may want to personalise their user profile to specify additional information requirements by, for example, specifying "standing queries" that are executed whenever some set of trigger conditions is met and whose results are forwarded to them.

Also, comms and manual labour are precious commodities - we aim to increase comms efficiency, while reducing the unnecessary effort in a deployed headquarters that might be caused by information overload. These requirements mean that good information management is essential in order for a resource like Theatre Broadcast System to be effective.

What are we doing about it?

Improving the Information Management aspects of TBS (as was begun with PK1) to enhance comms utilisation, as well as adding a new capability of creating an information federate that spans the fixed/deployed contexts. The eventual goal is that staff in a deployed command centre can access information that will be tailored and enriched in an appropriate manner, based on operational tempo, communications availability and user needs.

The initial phase aims to deliver to deployed command centres message feeds that have been enriched with information from other sources (databases, Web sites, information services, etc). The following phase will extend the MANIFOLD information federate (the source of all PK2 data) to bridge the air gap between fixed and deployed sites, so that a commander can access a far larger range of information than would otherwise be available.

The components of PeaceKEEPA Phase 2

From a technology perspective, PK2 is a set of existing components combined in an innovative manner. This design also serves to validate DSTO's EXC3ITE (Experimental C3I Technology Environment) concept of component-based C2 systems. These components include:

- MANIFOLD
- TBS
- The EXC3ITE Profile Service
- DSTO's FX (Fact Extractor / Information Extraction) framework
- COTS products and technologies such as servlets, CORBA, etc.

These components are explained separately, and an example serves to explain the operation of PK2 as a whole.

MANIFOLD

DSTO is developing the MANIFOLD architecture to meet various information management challenges, through unifying a wide range of data sources and taking account of real-world constraints such as bandwidth, footprint in theatre, cost, etc. MANIFOLD can be broadly be classed as a mediation architecture, as first explored in IBM's Garlic project [1].

Many existing information systems have been developed for specific requirements and only subsequently has it become clear that value can be obtained by answering queries that depend on relating data across individual systems. One approach here is to combine individual systems into one larger overall system. Changing or replacing existing systems ahead of their planned lifetimes is, however, both expensive and risky. Aside from the direct costs, there are flow-on effects such as the costs of retraining and disruptions to activities during the transitional period. Perhaps most importantly though are the issues of complexity and scale. A system designed to meet well-understood local needs is easier to design, implement and manage than a larger system with broader application (see below). There are good reasons therefore to find methods of combining data from individual holdings without major changes to those holdings. To this end, MANIFOLD assumes that existing systems will not be changed substantially in the process of achieving federation. The mediation approach encapsulates each source of data with a layer that provides access to its capabilities (via an ontology, or logical model) in a manner consistent with every other source; this is known as a unifying abstraction. In MANIFOLD, the unifying abstraction is very simple. A system of any size or complexity (from a single component to a data warehouse) can be wrapped (*adapted*) by a black box that supplies information in a defined way and is (optionally) able to supply result sets that match any queries posed. As an example, a track service might be adapted as a box that produces "Track" entities, that each contain a series of "points", each of which in turn contains a "latitude" and "longitude". Furthermore, we may be able to ask for these entities based on some constraints such as where the "track" is "within distance X of point Y". The point here is that the actual underlying technologies are hidden.

The second critical component of MANIFOLD is the “hub”. This component presents a collection of data source adaptors as a single source, marshalling all information flows to and from the constituent adaptors, and presenting a unified merged view of those adaptors.. This in turn makes a hub behave externally as just another data source adaptor, allowing a single point of access to multiple sources. A greater advantage, however, is the ability to exploit synergies between systems through both implicit and explicit relationships between data elements, thus creating new information where previously there was only disparate data. As an example, assume we have a gazetteer service alongside our track service that provides the name and location of towns and cities. Through use of the MANIFOLD hub, we can now ask for tracks that are within a given distance of a certain town.

Part of MANIFOLD's strength stems from its simplicity – there is no required pre-existing, enterprise data model to adhere to. Historical experience demonstrates that as the size and dynamism of constituent information systems grows, an all-encompassing enterprise model becomes increasingly difficult to maintain. As with avoidance of changes to existing systems, we see this as a key underlying assumption for successful data federation. The MANIFOLD alternative is for information systems to describe their own capabilities as they come together to form an information system "community" – that is, sets of related information systems in an operational domain (perhaps financial systems, or personnel systems) that can usefully be linked into a common logical representation. Only part of the content of information systems implemented to meet well-understood local needs will be relevant in a broader context. By identifying information exchange requirements between domains and catering only for these, overall management overheads can be reduced significantly.

Ideally, a hierarchy of MANIFOLD domains corresponds to such management domains. Information communities can be set up in a variety of systems domains with only limited centralised coordination, creating logical models that are appropriate for the systems involved and the benefits accruing from federation of their data. This is important for two reasons: the likelihood of the data integration *actually being realised* is increased (due to the reduced problem complexity); and, more importantly, the personnel in each systems domain can see direct benefit as the integration is focused on their domain and the benefits accruing, as opposed to a more remote strategically-oriented process. An example of an information community in this context might be a deployed headquarters, or even cells within that HQ.

An obvious question at this point is why tolerate (and even encourage) the development of stovepiped information systems with what appear to be stovepiped federations? The principle here is to avoid management overheads and effective inertia that increase dramatically as the size of the domain being managed goes beyond a practical limit. Buy-in to standardisation and coordination efforts is easier to achieve when the local benefits are apparent and standardisation decisions make sense in the local context. We use the concept of *schema translation* to facilitate effective exchange of information between domains that are more or less independently managed (with a degree of coordination across domains forming the next level in a hierarchy).

Schema translation involves linking two or more federates together by reconciling the differences between their representations of information and processing capabilities. A simple example would be a Navy federate describing 'ships' while an Army federate describes 'vehicles'. A query that spans the common elements of both can be resolved through a common description of 'platforms'. As with all other MANIFOLD components, the technical emphasis is on the need to deploy schema translators rapidly and easily by configuring them for a domain of application, rather than building them as a specific point-wise solution.

Why is MANIFOLD so important in the context of PK2? Because PK2 retrieves all its data from MANIFOLD, it is completely decoupled from the source of this data, and the underlying technology required to access that data. This also means that anything MANIFOLD can talk to is also a source of information for broadcast via TBS (or any other system that requests data from MANIFOLD). Also, in the deployed context, multiple sources of the same *sort* of information can be made available seamlessly, and chosen according to circumstances. For example, if communications are lost, the MANIFOLD hub at a deployed location can fail-over to a local store of information, which while not being the most up to date, is better than nothing (given that its lack of currency is identified in some manner).

TBS

The Theatre Broadcast System is a combination of hardware and software built to distribute information via satellite to deployed forces. TBS's focus is on a smaller scale broadcast system than the US Global Broadcast System. To maximise the efficiency of the bandwidth, TBS provides a scheduling and priority system to manage use of the medium.

EXC3ITE Profile Service

This service is based on a concept similar to LDAP (Lightweight Directory Access Protocol) [2], but is designed to be of even lighter weight and focussed less on directory management and more on distributed user preferences. The profile service is designed to be pervasive, potentially being consulted at every level of a system. At the upper layers, the service might dictate how information should be presented to a user, for example whether a set of intelligence report links should be placed in a table, or geo-located on a map. At the information management layer (where MANIFOLD sits) the profile service might inform the information service as to how to seek information. Based on who is seeking the information, and what their tasks are, should we seek assurance over timeliness, currency over communications expense, etc. At the lower levels the profile service could inform data services of the degree of detail required, etc. Imagery retrieved from an imagery service might be presented at a lower resolution, or have gaps in regions known to be of little interest. Track data may intentionally be simplified where appropriate (if we are casually interested in the shipping activity in a given sector, it may be sufficient to show the path of ship as a straight line between two points, rather than a more-complete record of acquired track points).

The challenge in this area is to strike a balance between relieving users from information overload and omitting information that turns out to be critical.

Fact Extractor Framework

DSTO's FX framework is designed to allow the rapid extraction of information from unstructured textual sources. Currently, FX processing relies on a combination of pattern and context matching. In order to better support projects such as PK2, we are extending this basic framework into a more-capable Document Markup Service. One function will be to map the basic FX templates into a MANIFOLD information federate's ontology, to decouple the service from reliance on the specifics of FX templates. The aim here is to take advantage of other technologies (such as natural language processing) as they become more capable, or as domains are identified where certain technologies perform better than others. This service will process documents from a variety of sources including message feeds, web servers, document repositories, etc.

Once facts are identified in a source document, a copy is created in which the facts are highlighted and related back to the underlying ontology. These facts are also stored in a database (keyed to both the source and enriched documents) for later searching and trend analysis. An example of the type of analysis that could be performed, given such a store of information, might be detecting abnormal platform activity in a region, or identifying otherwise hidden relationships between entities mentioned in reports.

PK2 as a Whole

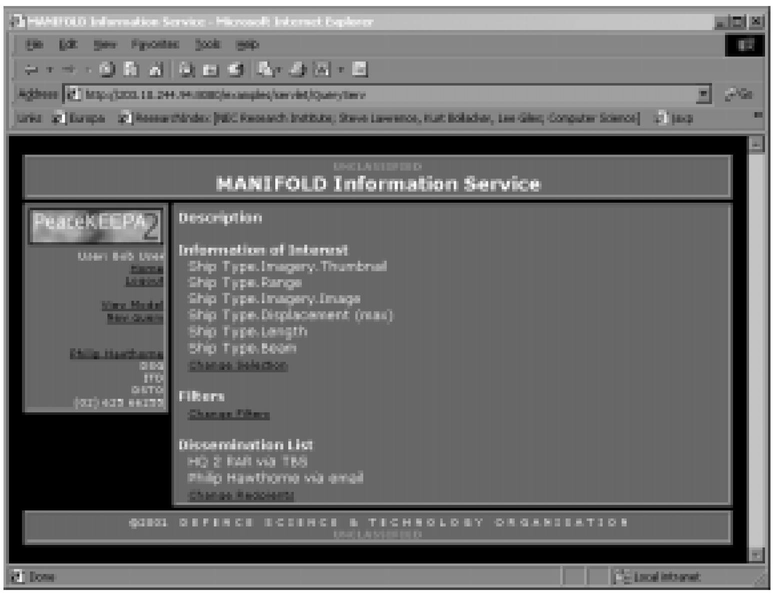
As mentioned above, PK2 builds on several other DSTO initiatives. This section explains how various technologies are combined to give a new capability to commanders in the field, using as an example the path followed by an intelligence report as it moves from the operational level fixed network down to various deployed brigade intelligence cells.

```
MEKO200 FFG-151 ARUNTA 050416Z SYDNEY
-- (U) DPTD MELBOUNRE, ENR BRISBANE
```

An intelligence report is published onto the message feed, where it is picked up by the Document Markup Service, and processed according to FX templates written for that specific domain. The Document Markup Service has a MANIFOLD adaptor interface that contains a number of subscriptions (or standing queries) entered by an operator via a query interface to the hub, which forwards the subscriptions to the document service.

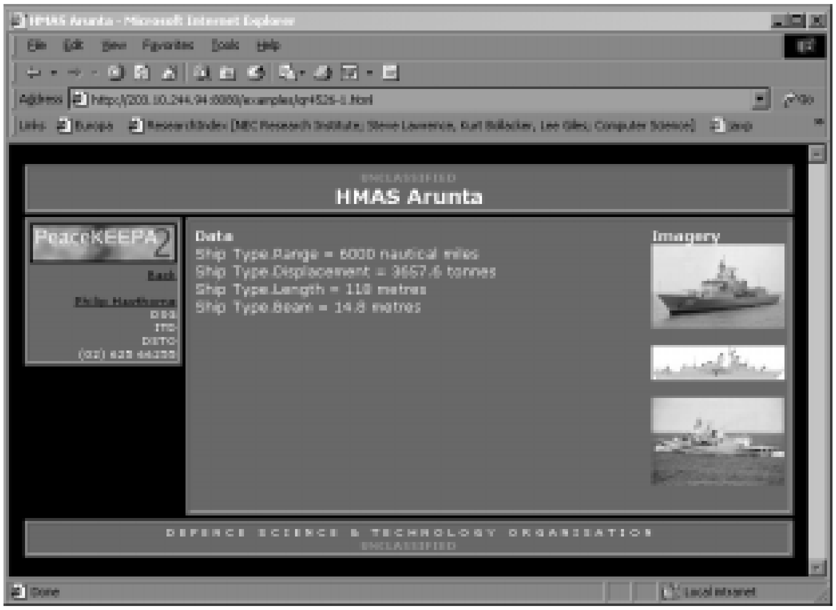
```
MEKO200 FFG-151
<ShipName>ARUNTA</ShipName> 050416Z
<LocationName>SYDNEY</LocationName>
-- (U) DPTD
<LocationName>MELBOUNRE</LocationName>,
ENR
<LocationName>BRISBANE</LocationName>
```

Suppose a subscription specifies that any documents or messages mentioning shipping activity in a given region should be sent to the MANIFOLD hub. The current INTREP fits that description, so it is forwarded. MANIFOLD itself knows nothing about intelligence reports other than where they come from and where they go to. However, a PK2 module loaded in the hub knows how to read marked up documents, so it receives the message for further processing. The PK2 module "reads" the message, and for each fact found, checks the MANIFOLD hub for queries that can find related information to enrich the document. These queries are created at an earlier preparatory stage, but can be added to or altered to fit changing operational needs. An example “enriching” query is shown below.



In the given INTREP, a particular ship is mentioned and the PK2 module sees that it has been instructed to retrieve the dimensions, range, and related imagery of any ship mentioned (as shown by the picture at left). PK2 needs only to fill in the ship's name in the filering part of this query and MANIFOLD finds and retrieves the information from the information space (this information might be accessible from a single point such as Janes Fighting Ships, or as a composite from a range of other sources). MANIFOLD returns this information to PK2, along with information for any other requests the module made in regard to this message. PK2 now renders this information into a brief (perhaps in HTML) based around the original text, but with "links" to pages and/or tables of related information, including a table of ship data in this case). Images, and other non-textual data (such as geospatial or track data) can be written out in appropriate file formats, and the whole brief is packaged up and handed back to MANIFOLD.

Again MANIFOLD is ignorant as to the meaning of this data, but because of where it came from, and who asked for it, it knows to route this package of information to the TBS transmission component. This consults the profile service to determine various properties such as the importance of the message, and which clients to send it to. TBS broadcasts the information and the PK1 receive software determines for each receive point whether this information is relevant, and if so, where to send it. The intelligence brief makes its way to the G2 staff and is presented in a browser window. Clicking on the highlighted ship name "Arunta" brings up the window above. Importantly, the underlying metadata has been retained such that if staff wish to request further information about some aspect of the brief (assuming they have a limited backlink), the task will be much simpler than starting afresh.



Discussion

We have described above initial steps in delivery of appropriate information to commanders in theatre. The goal here is that packaged with the targeted source material will be all supplementary information required by the deployed user (not visible except when accessed via internal hyperlinks), such that no further requests have to be processed. This work forms part of a larger initiative aimed at giving effect to the RMA concepts of mobility and minimal footprint in theatre.

In conclusion, appropriate information management, based on unification of a wide range of data sources and types, used in conjunction with a high-bandwidth broadcast capability and detailed user profiles, allows delivery of targeted operational information to individual deployed clients. This both adds value to existing information sources and significantly reduces the connectivity disadvantages experienced by deployed forces.

References:

[1] http://www.almaden.ibm.com/cs/garlic/homepage.html
[2] http://www.ietf.org/rfc/rfc2251.txt