Middleware Platform for Recruiting and Proactively Managing Virtual Panels of Intelligence Experts[1, 2]

Hyong-Sop Shim, Clifford Behrens, and Devasis Bassu

Applied Research
Telcordia Technologies, Inc.

One Telcordia Drive
Piscataway, NJ 08854
732-699-2609
{hyongsop, cliff, dbassu}@research.telcordia.com

Abstract

The potential gained from "tapping" the knowledge of domain experts in war-gaming and decision-modeling has yet to be fully realized in the Intelligence Community. While there may be a number of reasons for this, perhaps chief among them is the cost of involving human experts, and the belief that intelligence based on expert opinion is poor quality and unreliable. To this end, we are working on a middleware platform, called Collaborative Panel Administrator (CPA). CPA is specifically designed to support virtual panel management and asynchronous data collection. It also allows for data aggregation and imputation to facilitate incremental validation of intelligence models. To maximize objectivity of panelist input, the CPA panel management facilitates blind collaboration, in which identities of panelists in the same panel can selectively be hidden from each other, even while working on the same intelligence model, thus reducing the effect of "group think" and adverse social dynamics. CPA mainly works as a data hub between intelligence modeling tools and corresponding model validation services. In this paper, we discuss in detail motivation and design of the CPA. Where appropriate, we also discuss how distributed agents would provide an effective means of realizing some of CPA functionalities.

1. Introduction

Good intelligence analysis begins with good intelligence analysts. However, traditional approaches to recruiting intelligence experts for panels do not always ensure involvement of the best and brightest. Moreover, modeling approaches that require bringing together expert panelists in one place to meet face-to-face is not only impractical and costly but also unproductive; personalities and inter-personal communications skills, rather than individual expertise and well-reasoned argumentation, often significantly influence the output of such a panel. Furthermore, same-time, same-place arrangements for convening panels do not allow for systematic qualification of panelists and validation of their knowledge or post hoc refinement of intelligence models as new information becomes available.

To address these issues, we are advancing the concept of *virtual panels* for distributed, collaborative model-building by intelligence experts. By virtual, we mean panels whose members may be distributed in space and time. Individual panelists may submit their model inputs incrementally and asynchronously, i.e., at their own convenience. By collaborative, we mean that panelists, if desired and further supported by policy, may collaborate with each other over distance and time by using both synchronous and asynchronous means of communication and information sharing.

Providing support for virtual panels poses many challenges to IT developers. For example, while panelist recruitment within a virtual space opens the possibility for recruiting the best and brightest experts, regardless of their physical location, it also requires support in the way of extensive directory and search services. In addition, if information is collected from panelists incrementally and at *their* convenience, this implies that, at any moment in time, the amount of data needed to compute a consensus model may be incomplete, so missing data must be imputed. Obviously, data imputation methods work best when the existing data they exploit have been collected in a rigorous and systematic manner, i.e., according to a data acquisition plan based on sound experimental design. Otherwise, only a small, unbalanced, and unrepresentative subsample of data may be available for model estimation, which can result in severely biased models or may make model estimation computationally intractable.

What is needed is new infrastructure that dramatically reduces the cost of involving human experts in collaborative war-gaming and decision-modeling activities, so that these may include the best and brightest experts, regardless of their physical location. This infrastructure should help a human Panel Administrator to rapidly identify and recruit the most capable panelists, and acquire information from them in a timely manner, using the most appropriate data-acquisition and modeling tools for the task. It should also provide a seamless interface to

[2] This material is based upon work supported by the Space and Naval Warfare Systems Center, San Diego under Contract N66001-03-C-8005.

collaboration tools, enabling cost-effective and rapid communication among panelists. Most of all, this infrastructure needs to provide intelligent control over available resources in the local IT environment so that they are used most effectively to produce valid and reliable intelligence.

To this end, we are developing a multi-agent middleware platform, called the Collaborative Panel Administrator (CPA), designed to provide human Panel Administrators with systematic and extensible support for panelist recruitment, data acquisition and processing, and panel lifecycle management. Currently, CPA is being developed to support virtual panelists use of a Web Service (WS), called the Schemer (Behrens and Shim 2004), which provides knowledge validation and collaboration services to client war-gaming and decision-modeling tools. Wherever appropriate, CPA management functions are designed to make use of distributed agents to more proactively deliver new services and meet new requirements.

In this paper, we motivate and discuss our design of key CPA management services in detail. Section 2 examines our motivation for middleware technologies for supporting virtual panels in distributed intelligence data collection and analysis. Section 3 discusses in detail functional requirements for our middleware. Section 4 briefly describes CPA's architecture. Section 5 discusses in detail the CPA panel management process. Section 6 briefly describes Schemer Web Service and its methodology for data acquisition plan generation. Section 7 reports the current implementation status and highlights key challenges. Section 8 reviews related work, and Section 9 draws conclusions from the paper.

2. Motivation

In general, the CPA functions as a "bridge" between online information analysis services, e.g., the Schemer, and intelligence modeling tools used by virtual panels of domain experts to analyze panel data and produce models with it. Online services provide application-specific functionalities, e.g., entity extraction and knowledge validation (Behrens and Kashyap 2002), which is often critical for enabling virtual panels to refine and enhance their knowledge base, and produce accurate intelligence in a timely and efficient manner. As such, these services are constantly "tweaked" with new algorithms, and architectural and operational enhancements to provide the best performance possible, regardless of application. However, managing panels in virtual environments is a different issue, which has not received (and probably will not receive) much attention from service providers.

Support for managing virtual panels is also lacking in most existing software tools used by intelligence experts for modeling and analysis. Typically designed as single-user applications, these tools (rightfully) focus on easing data acquisition with easy-to-use graphical user interfaces (GUIs), and by optimizing analysis and model computation algorithms. Use of such tools necessarily implies panels with "same-time, same-place" contexts. Each tool can individually be re-designed and re-implemented to support virtual panels. However, doing so is not trivial and may introduce undue usage and management overhead as the same virtual panel may employ multiple tools.

Orthogonal to the issue of enabling virtual panels, but equally important, is support (or lack of support) for integrating online information analysis services with intelligence modeling tools. Simply put, the prevailing practice of making online services available to individual tools has been point-to-point, which is costly and leads to tightly-coupled and closed systems that cannot easily scale to new functional and operational requirements.

CPA aims to close the "gap" between online information analysis services and individual intelligence modeling tools. In particular, the CPA provides a set of management services commonly required by virtual panels to assist with this integration of services and tools.

3. Functional Requirements

Virtual panel management and service integration in distributed and asynchronous environments requires the following functions:

- *Panel Lifecycle Management* – Systematic support should be provided for recruiting, convening and disbanding panels. Panel recruitment involves identifying potential qualified experts. When possible, recruitment should be automated by agents that exploit available algorithms such as "snowball" sampling, tracing citation networks, or discovering associations in concept spaces computed from knowledge products. In addition, long-lived panels should be allowed, which involves keeping track of the current state of panel membership and allowing panelists to log in (possibly from different locations), perform their work, and log out at their own convenience without losing prior work.
- *Service Data Model Management* – Online information analysis services often require input data to conform to a predefined data model. For example, Schemer exploits a data model that defines *meta* information about panel and instrument, where panel information includes panelist identifiers and (possibly) their roles, and instrument refers to a set of data items on which Schemer performs its consensus analysis. By meta, we mean the details of instrument structure and format, e.g., number of data items, item identifiers, and their data types, and panel information, e.g., number of panelists and their identifiers. These meta data are defined for each specific client modeling tool and panel. In general, information analysis services determine meta data models from which specific data models are generated only when they are integrated with specific client tools. Therefore, a flexible and easy-to-use means should be provided for allowing creation, maintenance, and distribution of client-specific data models in distributed environments. In addition, as multiple panels may exist that use the same

client tool, cross-panel sharing of client-specific data models and per-panel customization of a data model should also be supported.

- *Service Data Acquisition Management* – Virtual panelists may submit their modeling data incrementally and asynchronously. At the same time, many services, including Schemer, require complete data sets, i.e., all the panelists should have provided a model fully-populated with data, before they can apply their processing. Therefore, support should be provided so that individual panelist data are stored, version-controlled, and aggregated for submission to services.

On the other hand, populating a model fully often takes a long time. Thus it may be detrimental to get virtual panels "going," especially during the early stages of model-building, to always require complete response sets from every panelist before getting analytical results back. Furthermore, it may be desirable to have analytical results computed from partial data which, while perhaps not as valid, may still be useful for gaining early insights, or for further refining an instrument or panel membership. To this end, Schemer accepts partially-populated models for analysis by way of imputing missing data values. So that a consensus model is computed from balanced data and representative of a panel, Schemer also provides a service that generates data acquisition plans based on current panel instrument and membership information. Specifically, a data acquisition plan is generated for a panel upon request from Panel Administrators, and this plan assigns each panelist a subset of items to "answer." The size of each item subset is smaller than the complete set of items, and no panelist is assigned more items than any other, thus sharing the data acquisition burden equally amongst panelists. Moreover, these assignments are determined in such a way that, when every panelist has answered his/her assigned items, each item is answered at least n times, where n is dependent on both size of the complete set of items and size of the panel membership.

Therefore, a scalable and flexible means should be available by which data acquisition plans can be asynchronously distributed among panelists and be enforced in a timely manner. The latter requires that the middleware platform have an understanding of the contents of a data acquisition plan, keep track of each panelist's progress, and proactively generate "reminders" to those who are delinquent.

- *Security Policy Management* – Sometimes analytical results from online services may contain sensitive information which should not be accessible to all panelists. For example, Schemer's consensus analysis results contain computed metrics on panelist competency which, if distributed to every panelist, may create tension. Furthermore, there may be cases where the exact identities of panelists should not necessarily be known to all panelists. Therefore, a general and flexible means of creating and enforcing security policies for identifying sensitive data and assigning appropriate access and presentation rights for panelists should be provided.

- *Collaboration Management* – To support communication and data sharing needs of distributed panelists, groupware applications (Ellis and Gibbs 1991) are used. To both minimize manual synchronization efforts and further promote collaboration and knowledge sharing among distributed panelists, analytical results from online services should be collaboration-aware and integrated with locally available groupware applications. In general, different panels use different groupware applications and communications tools. Thus, an adaptive means should be provided so that the same analytical results can be used in different collaboration and communication environments without placing undue overhead on service providers.

4. Collaborative Panel Administrator

In this section, we describe the architecture of the Collaborative Panel Administrator (CPA). Functionally, it aims to provide virtual panel management services in Section 3 to a wide variety of online analysis services and client modeling tools. Architecturally, as shown in Figure 1, the CPA functions as a proxy to client model-building tools by intercepting and relaying client requests to online services and then storing and distributing analytical results to clients according to panel security policies.

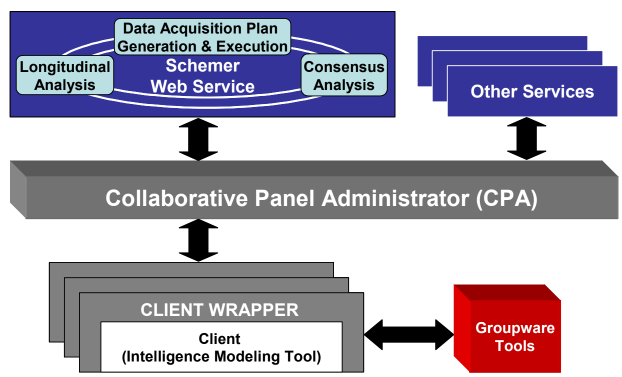

Figure 1. CPA Architectural Overview

As shown in Figure 1, the CPA works in conjunction with the Client Wrapper. It is designed to facilitate CPA←→client integration by providing customizable utilities (with GUIs) for defining per-service and per-panel instruments, security policies, and data entry forms. Furthermore, the Client Wrapper integrates analytical results with groupware tools in use on the client host by defining a set of APIs that encapsulate integration details. In (Behrens and Shim 2004), we have discussed how this strategy is used to integrate Schemer's consensus analysis results with a commercial groupware system. We are currently working to extend our support for other groupware systems.

The CPA is designed so that the Client Wrapper works as a CPA-aware personal assistant, while some of the CPA's services, i.e., panelist recruitment, data acquisition plan interpretation and execution, and interaction with online services, are designed as more proactive agents. With this design, general inter-agent interactions required for realizing virtual panel management services can be defined and built independent of specific services and clients. Information specific to online services and client tools can be captured and stored in a knowledge base for agent use. Furthermore, new online services can be supported on an as-needed basis, e.g., by creating service-specific agents, without affecting operations of the rest of the system.

5. CPA Panel Management

In this section, we present the functional overview of CPA virtual panel management process.

Recruit Panel

The first step in the process is panel recruitment, in which the Panel Administrator chooses candidate panelists for his or her panel. Specifically, the Panel Administrator determines panel metadata, which includes panel name, subject matter in which panelists should have expertise, and panel size (number of required panelists). The panel size may depend on data analysis and validation services to be used in the panel. For example, the Schemer Web Service (see Section 6) requires a minimum of 6 panelists to perform its consensus analysis. Therefore, the Panel Administrator should also determine validation services to use in his/her panel during the recruitment process. This implies that an interface is required that allows the CPA to dynamically discover operational requirements of services and generate appropriate data entry forms. To this end, we are defining an XML schema for service deployment in the CPA.

The Panel Administrator should also determine the panel type. CPA supports two types of panel: LEAD_PANEL and EXPERT_PANEL. The main focus of a LEAD_PANEL is to bootstrap the work of the EXPERT_PANEL by creating a straw-man model, while that of the corresponding EXPERT_PANEL is to analyze the straw-man model. What constitutes a straw-man model is dependent on the information modeling tool to be used in the EXPERT_PANEL. For example, it may be as simple as a questionnaire or as complex as a large influence network of nodes and interconnecting links, where a node represents some event and a link denotes the (presumed) influence of one event on another, e.g., see (Rosen and Wayne 2000). In the latter case, the panelists in the LEAD_PANEL would determine the nodes and links of an influence net, i.e., its topology, while those in the EXPERT_PANEL would provide their opinions on the likelihood of a given event happening and degree of its influence on other interconnected events. In the former case, the panelists in the LEAD_PANEL would generate questions to be included in the questionnaire, while those in the EXPERT_PANEL would answer those questions. We envision that the size of a LEAD_PANEL will be much smaller than that of an EXPERT_PANEL, i.e., around 1-3 experts.

The above discussion assumes availability of *CPA-enabled* information modeling tool on the host of each panelist. To this end, the Client Wrapper (see Section 4) provides a set of application programming interfaces (APIs), through which the information modeling tool can interact with the CPA.

During the recruitment process, the Panel Administrator also determines the recruitment method by which candidate panelists should be selected. As previously discussed, a number of methods can be used, including manual selection from an existing user directory, snowball sampling, and tracing of citation networks. It is up to the Panel Administrator to choose which method to use. The choice may depend on a number of factors, including availability of in-house expertise and trusted experts in the subject matter. Regardless of the exact method used, the Panel Administrator has a list of candidate panelists at the end of the recruitment process.

Create Panel

Subsequently, the Panel Administrator invites candidate panelists to the panel. Currently, the invitation is sent in the form of an email message, which includes the Panel Administrator's invitation message and the deadline by which the recipient should accept the invitation. It also contains a link to a Web page where the recipient can accept or reject the invitation.

Internally, the CPA creates a panel object for storing the accept/reject status of candidate panelists. Associated with the panel object is a monitor agent that keeps track of and proactively notifies the Panel Administrator (via email) of invitation status updates. The history of who has been invited and who has accepted / rejected the invitation is archived, so that, for example, the same candidate panelist is not invited multiple times.

Convene Panel

When a desired number of candidate panelists have accepted the invitation, the Panel Administrator can convene the panel. Specifically, s/he contacts and debriefs the panelists on the goals and other pertinent information about the panel. The Panel Administrator also gives instructions on the logistics of participating in the panel, e.g., creation of aliases and passwords and download and installation of client software modules, which include the Client Wrapper and intelligence modeling tools used in the panel.

By design, the Panel Administrator does not reveal panelist identities to the panel membership, and panelists are expected to interact with each other via aliases. To this end, we are working on the concept of *proxy agents* for panels and panelists, which would work as intermediaries in both inter-panelist collaboration and panelist ←→ panel interactions, e.g., data collection, retrieval of analytical results, and event notification.

Data Acquisition

Once convened, the members of a LEAD_PANEL get to work on a straw-man model. Because straw-man models should often be created from scratch, a high degree of interactivity and information sharing among panelists is expected. Use of proxy agents would enable anonymous collaboration among panelists, which, in turn, would increase the level of objectivity in the straw-man model. However, it may be counter-productive to always insist on collaboration “in the dark”; sometimes, face-to-face meetings are required in order to solve complex issues. It is up to the Panel Administrator if and when to reveal panelist identities.

Regardless of how the straw-man model is created, the Panel Administrator may choose to analyze and validate it using an online service. For example, the Schemer Web Service (Behrens and Shim 2004) may be used to compute consensus analysis on a collection of panelists’ edits or revisions of straw-man model, the results of which may be used to guide inter-panelist collaboration and further refine the model. We describe a similar process for stepwise model refinement using Schemer consensus analysis results shortly.

The main goal of an EXPERT_PANEL is for the panelists to assess straw-man models created by the corresponding LEAD_PANEL. To this end, the Panel Administrator generates a data acquisition plan, in which each panelist is assigned some part of a straw-man model to work on, which may, for example, entail answering a subset of questions in a questionnaire or assigning values to a subset of nodes and links in an influence network. The algorithm used to generate data acquisition plans is dependent on the data models of online data analysis/validation services in use. See Section 6 for an overview on Schemer Web Service’s algorithm. Data acquisition plans facilitate incremental data collection, analysis, and validation, which, in turn, helps refine models, as described shortly.

Associated with each acquisition plan is a Panel Administrator-defined deadline, by which panelists should complete their assignments. The data acquisition plan is sent to each panelist, whose Client Wrapper interprets and executes the plan; it parses the plan to determine the part of the straw-man model assigned to its panelist and proactively alerts the panelist when s/he falls behind the schedule. It is responsible for storing the panelist input on the local host and sending it to the CPA upon panelist request. This way, the Client Wrapper works as a personal assistant in the data acquisition process.

Data Analysis and Validation

Once a given data acquisition plan has been fulfilled, the Panel Administrator can aggregate the collected data and submit it to an online service for analysis and validation. Currently, the Schemer Web Service (Behrens and Shim 2004) is available for computing consensus and longitudinal analysis (see Section 6 for a brief introduction). The Panel Administrator can use analytical results from Schemer to refine intelligence models as follows. First, Schemer generates its consensus analysis results and encapsulates them in an artifact, called Schemer Knowledge Object (SKO), per user request. SKO contains the consensus model of the submitted set of panelist responses. For example, if data is collected using a questionnaire, the consensus model would provide a consensus answer for each question in the questionnaire as computed by Schemer based on the submitted response set; see (Behrens and Shim 2004) for detailed algorithm for Schemer’s consensus analysis. In addition, the SKO includes a knowledge map (or KMap) showing how knowledge is distributed amongst panelists, based on analysis of their responses. The larger the spread in this distribution is, the greater the variability in the panelist response set, and the less reliable the computed consensus model. Figure 2 shows an example SKO and KMap.

Upon completion of a consensus analysis request, the Panel Administrator can retrieve the corresponding SKO and examine its contents (via his or her Client Wrapper). If the consensus model is deemed less reliable than desired, the Panel Administrator can spur them to collaborate by distributing the SKO. The Client Wrapper includes an SKO Wrapper that provides graphical means of displaying the SKO, giving panelists insights as to who seems to know the most/least about the target domain, and who possess the most similar/different points of view, as shown in Figure 2.

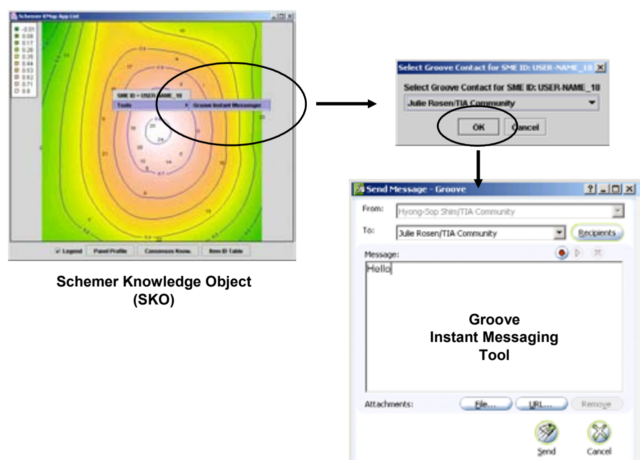

Figure 2. Schemer Knowledge Object (SKO). The local panelist can initiate communication with other panelists by right-clicking on panelist aliases (shown as numbers) on the contour map of the SKO and selecting desired tools.

The reason for providing these insights is to provoke collaboration among panelists, with the intent of fostering consensus formation and knowledge-building, while also reducing "group think" and the adverse effects of social dynamics. To this end, we allow the Panel Administrator to actively mediate / monitor / archive inter-panelist communications. In addition, by default, panelists interact with each other via proxy agents, hiding their identities (note that panelist identities are shown in Figure 2). We call this type of collaboration, which is motivated by results of analysis and validation of panel-generated data *knowledge-based collaboration*.

Subsequently, the Panel Administrator may repeat the entire process of data acquisition plan generation, data collection, and data analysis and validation, at the end of which a new SKO is created. The consensus model in the new SKO is a refined version of the consensus model in the earlier SKO. At this point, the Panel Administrator can compute the degree of changes in panelist responses by making a request to the Schemer Web Service to perform a longitudinal analysis on the current set of SKOs (see Section 6). This provides the Panel Administrator with an effective means of tracking changes in panelist opinions and monitoring consensus formation from one iteration to the next.

Disband Panel

The Panel Administrator may repeat the process of data acquisition plan generation, data collection, and data analysis and validation as many times as s/he deems necessary. With each iteration, the consensus model of the panel may be further refined, and the panel knowledge in the subject matter may be enhanced via collaboration. However, it is also possible that panelists may not converge on a single consensus model even after many iterations. Either way, when the Panel Administrator feels no further refinement can be made to the consensus model, s/he disbands the panel, at which point the CPA removes the panel object from the system and performs other clean-up operations. The CPA may archive all the artifacts generated during the lifetime of the panel for future reference.

6. Schemer Web Service (WS)

In this section, we introduce the Schemer Web Service as an example of online analysis services that the CPA is designed to support. See (Behrens and Kashyap 2002) for in-depth discussion on Schemer.

Consensus and Longitudinal Analysis

Schemer is a flexible knowledge-driven technology that motivates collaboration through a heightened awareness of "who knows what." Schemer provides this capability by imposing a rigorous scientific methodology, known as "consensus analysis," on collaborative modeling, which yields (1) timely and relevant knowledge validation and collaboration metrics, (2) visual representations of collaboration processes and the distribution of knowledge within expert panels, and (3) real-time model estimation from information provided by panels of subject matter experts. Schemer services have been designed and implemented as a Web Service (WS) that derives consensus knowledge further supported by metrics to validate the derived knowledge and competency of human collaborators. Unlike previous approaches that force consensus, often yielding results that only replicate the biases promoted by a dominant panelist, Schemer derives its models from the consensus discovered in the inputs provided by all panelists, each weighted by an objective estimate of their respective expertise. Longitudinal analyses of knowledge distributions within panels by Schemer also provides a means of measuring consensus evolution and knowledge building over time, and the contribution of collaboration to these processes.

Data Acquisition Plan Generation

As previously discussed, incremental and asynchronous data acquisition should be monitored and properly guided in order to ensure balanced data collection, which is critical for meaningful analysis. We address this issue by putting a data acquisition plan in place, which essentially has two primary functions: (1) provide a prioritized data collection strategy given the present data and (2) provide imputation strategies so that analysis may be carried out on incomplete data.

The data acquisition strategy generated by Schemer (and carried out by the CPA) is based on the principles of "balanced incomplete block design." The idea is to *grow* the incomplete data in a balanced fashion (i.e., each instrument item/data slot and each panelist get proper representation) towards a fully complete dataset.

In order to impute missing item values in Schemer, we propose a k-NN (k nearest neighbors) imputation scheme with appropriate thresholds (Troyanskaya 2001). The underlying principle is that a panelist will tend to respond (for the missing data) in a similar fashion as other panelists who seem to *match on most responses* (for the collected data). Based on the thresholds, the scheme may even choose to drop a panelist or a piece of the panel instrument if the collected data is insufficient to yield any meaningful results. From Panel Administrator's point of view, having the ability to set these thresholds provides some control over the quality of the collected data, hence also on the quality of the derived consensus model. Data acquisition plans are the central component of the CPA because they provide the "intelligence" which manages data acquisition, data analysis and collaboration in a manner that ensures timely production of the best models possible, given available human and IT resources.

7. Implementation Status and Issues

While some of constituent components of the CPA have been implemented, e.g., the Schemer Web Service and SKO Wrapper, the work on the CPA is ongoing. One key issue to CPA implementation is how to make the CPA a general-purpose panel management platform that can not only support the Schemer Web Service but also other online data analysis and validation services that may become available in the future. A common way of achieving this would be to define a set of interfaces, which would-be services must implement. In the context of Web Services, such interfaces will be specified in Web Service Description Language (WSDL).

However, one unique challenge in implementing the CPA stems from having to support *partial-sharing of data models*. For example, in order to aggregate panelist response data for Schemer's consensus analysis, the CPA should know the details of Schemer's response data model. However, hard-coding the Schemer's data model into implementation makes the CPA Schemer-specific and makes it difficult to extend the CPA to other services. Similar issues exist in supporting data acquisition plans, which, by definition, include some elements of service-specific data models. To address these issues, we are defining a set of meta models that would be shared between the CPA and CPA-enabled services. These meta models define data elements and attributes that any service-specific data model should include. Service-specific details will be specified in a service-specific data model, which is an instance of the meta model. Then, the CPA can be a service-agnostic platform by limiting its implementation to the elements of meta models.

8. Related Work

Virtual panel management has been an active area of research, especially in the field of computer-supported cooperative work (CSCW); see (Ellis and Gibbs 1991) for introduction on CSCW. For example, Corona (Hall et. al. 1996) is a server-based system for managing and transferring shared application state and data among groups of distributed end users. The system is part of a distributed system, called Upper Atmospheric Research Collaboratory (UARC), which provides resource sharing and synchronous / asynchronous collaboration capabilities for distributed communities of space scientists (Subramanian et. al. 1999). UARC has Web-based shared workspaces, where scientists can determine presence of each other in the same workspace, communicate using a chat tool, and receive help from system administrators. TeamRooms (Roseman and Greenberg 1996) is one pioneering work that uses group/social awareness as a means of inducing collaboration. It provides a shared workspace on the desktop that allows distributed collaborators to determine each other's current location (in the workspace) and work context by way of graphical user interface (GUI) components specifically designed to provide awareness information. Multi-User Dungeons (MUDs) have advanced the concept of online user communities by providing an ability to create interconnected rooms that can be navigated, and also define shared objects and possible operations on them via a command line interface (Curtis and Nichols 1993; Evard 1993).

The CPA is distinguished from the above and other similar works in its assumption about the membership of virtual panels. Specifically, in a CPA virtual panel, panelists may not necessarily know each other's identity. In fact, except for the Panel Administrator, it may often be the case that a panelist may not know the full membership of his/her panel for the duration of the panel's lifetime. This is by design in order to obtain as objective expert opinion from individual panelists as possible.

This assumption leads to our ongoing work on providing support for "blind collaboration," in which panelists contribute their input towards a common goal independently and asynchronously of each other. In addition to proxy agents (see Section 5), support mechanisms may range from email with hidden recipient identities to "sanitizing" analytical results based on access privileges of individual panelists prior to panel-wide distribution. Knowledge-based collaboration (see Section 5 and Figure 2) also enables the Panel Administrator to determine whether or not the identities of involved panelists should be revealed to each other on a case-by-case basis. A major part of further work on the CPA concerns design and development of "wrapper" technologies so that the artifacts of any information analysis and modeling services can be used as collaboration media.

In contrast, in many CSCW systems for distributed collaboration, it is often assumed that collaborators know the identities of each other. In fact, the identities of collaborators, along with information about their presence in shared workspaces, are often used as the main conduit for inducing collaboration. This has led to active research in generation and use of awareness information as collaboration media and proliferation of "awareness widgets," e.g., (Gutwin et. al. 1996; Molli et. al. 2001; Hill and Gutwin 2003). The ubiquitous "buddy lists" in instant messaging applications have their origin in awareness research in CSCW and are being effectively used for inducing "water-cooler" conversations and informal teamwork in workplaces, e.g., (Handel and Herbsleb 2002).

We believe that the issue of allowing for strong objectivity required of panelists engaged in collaborative information analysis and model building activities has not yet been well addressed in traditional CSCW research. CPA represents our approach to addressing this issue by providing support for carefully mediated and blind inter-panelist collaboration.

9. Conclusion

Good intelligence analysis requires domain expertise and well-reasoned argumentation by panelists. We have introduced the concept of a virtual panel as a means for providing more effective recruitment of panelists and data acquisition from them, allowing panelists to contribute their model input asynchronously, i.e., at their own convenience. However, little information technology currently exists for supporting virtual panels, forcing individual modeling tools to create ad-hoc solutions or even worse, requiring expert panelists to always convene at the same time in the same place. This situation significantly reduces both productivity and quality of their work.

To this end, we propose a multi-agent middleware system, called the Collaborative Panel Administrator (CPA), designed specifically to support management of virtual panels of intelligence experts in distributed and asynchronous environments. It offers panel administration features, including panelist recruitment agents, generation of data acquisition plans and corresponding agents that execute them in a timely manner, and analytical results that are improved through collaboration. CPA has a distributed, multi-agent architecture for maximum flexibility and scalability to new requirements and services. As an initial proof-of-concept, we are developing CPA to support the Schemer, a knowledge validation and collaboration Web service, and its client modeling tools. With CPA, we believe that we are addressing a general need within the intelligence community for infrastructure that makes it easier to integrate client modeling tools, analytical services, and collaboration tools deployed in local IT environments.

References

Behrens, C. and V. Kashyap. 2002. The "Emergent" Semantic Web: A Consensus Approach to Deriving Semantic Knowledge on the Web. *In V. Kashyap and L. Shklar (eds.), Real World Semantic Web Applications. Frontiers in Artificial Intelligence Applications*, vol. 2. pp. 69-90. Amsterdam: IAO Press.

Behrens, C., and Shim, H.S. 2004. Web Services for Knowledge-Driven Collaborative Modeling. *IT for Counterterrorism, 2004 IEEE Aerospace Conference*.

Troyanskaya, O., et.al., 2001. Missing value estimation methods for dna microarrays. *Bioinformatics*, 17(6), pp. 520-525.

Ellis, C. A., and Gibbs, S. J. 1991. Groupware Some Issues and Experience*, Communications of the ACM*, 34(1), pp. 38-58.

Hall, R., A. Mathur, G., Jahanian, F., Prakash, A., and Rasmussen, C., 1996. Corona: A Communication Service for Scalable, Reliable Group Collaboration Systems, *Proceedings of ACM 1996 Conference on Computer Supported Cooperative Work (CSCW 96)*, pp. 140-149.

Subramanian, S., Malan, G.R., Shim, H.S., Lee, J.H., Knoop, P., Weymouth, T.E., Jahanian, F., and Prakash, A., 1999. Software Architecture for the UARC Web-based Collaboratory, *Internet Computing*, IEEE, 3(2), pp. 46-54.

Roseman, M. and Greenberg, S. 1996. TeamRooms: Network Places for Collaboration, *Proceedings of ACM 1996 Conference on Computer Supported Cooperative Work (CSCW 96)*, pp. 325-333.

Gutwin, C., Roseman, M., and Greenberg, S. 1996, A usability study of awareness widgets in a shared workspace groupware system, *Proceedings of the ACM 1996 Conference on Computer Supported Work (CSCW 96)*, pp. 258–267.

Hill, J., and Gutwin, C. 2003. Awareness Support in a Groupware Widget Toolkit, *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work*, pp. 258-267.

Curtis, P. and Nichols, D. 1993. MUDs Grow Up: Social Virtual Reality in the Real World. *Proceedings of the Third International Conference on Cyberspace*.

Evard, R. 1993. Collaborative Networked Communication: MUDs as Systems Tools, *Proceedings of the* 7^{th} *USENIX Systems Administration Conference (LISA VII)*, pp. 1-8.

Molli, P., Skaf-Molli, H., and Bouthier, C. 2001. State Treemap: an Awareness Widget for Multi-Synchronous Groupware, *Proceedings of the* 7^{th} *International Workshop on Groupware (CRIWG 2001)*, pp. 106-114.

Handel, M. and Herbsleb, J. 2002. What is Chat Doing in the Workplace, *Proceedings of ACM 2002 Conference on Computer Supported Cooperative Work (CSCW 02)*, pp. 1-10.

Rosen, J. and Smith, W. 2000. Influence Net Modeling for Strategic Planning: A Structured Approach to Information Operations, *Phalanx*, 33(4).