KSCO 2007

# Knowledge Systems for Coalition Operation | 2007

EDITORS:
James Lawton
Jitu Patel
Austin Tate

James Lawton
Jitu Patel
Austin Tate
(editors)

# Knowledge Systems for Coalition Operations

Waltham, Massachusetts, USA.
1-2 May 2007

# Contents

# Agent-Based Multi-Layer Collision Avoidance to Unmanned Aerial Vehicles

David Šišlák, Přemysl Volf, Antonín Komenda, Jiří Samek and Michal Pěchouček
Agent Technology Group, Gerstner Laboratory
Department of Cybernetics, Czech Technical University in Prague
Technická 2, Prague 6, 166 27, Czech Republic
Email: sislakd@fel.cvut.cz, {volf|komenda|samek|pechouc}@labe.felk.cvut.cz

***Abstract*—This contribution presents a distributed, multi-layer collision avoidance architecture supporting efficient utilization of air space shared by several autonomous aerial vehicles. Presented multi-layer architecture is based on deliberative deployment of several collision avoidance methods by the aircraft at the same time. Both cooperative and non-cooperative collision avoidance methods are presented in the paper. The robustness of the architecture is justified by means of experimental validation of multi-agent simulation.**

## I. Introduction

Many coalition UAV relief operation (especially in the surveillance and monitoring domains) requires the See & Avoid capability as specified in [1]. The See & Avoid capabilities distributed among several autonomous aircrafts allows to utilize better the benefits of the *free flight* concept [2] – an approach of autonomous routing of the aircrafts based on local collision avoidance mechanisms. Such approach allows efficient operation of dynamically tasked unmanned aerial vehicles (UAV).

The presented work address problem of distributed collision avoidance among autonomous aerial vehicles using multi-agent technology [3] – each UAV is represented by an agent container hosting different functional agents [4]. Each UAV is controlled by a single, dedicated agent. The presented multi-layer collision avoidance architecture provides capability to integrate several different deconfliction algorithms that plans the runtime trajectory of each individual UAV. Such architecture supports operation of the group of cooperative UAV within the environment hosting other non-cooperative flying objects (e.g. civilian air traffic).

Cooperative collision avoidance is based on using different collision metrics [5] and negotiation protocols. One possibility is to shift from centralized solution towards fully distributed solution by deployment of principled negotiation [6]. Another way is to start from fully distributed solution among two aircrafts [7] based on classical agent-based negotiation protocols [8], [9] and extend this algorithm to more UAVs using iteration or making larger groups of negotiating agents. There are also various approaches based on the game theory (e.g. [10]) available in the research community.

Optimization non-cooperative collision avoidance algorithms [11] and [12] allows optimal solving of the collision with non-cooperative flying object (obstacle). These algorithms perform well when coping with a single alien flying object, but they cannot be extended to situation with several flying object, located nearby. Moreover they cannot be used simultaneously with other cooperative algorithms applied for the cooperative collisions at the same place. The research work reported in this contribution was motivated by designing such non-cooperative collision avoidance method that does not suffer these weaknesses.

The implementation and the experiments have been carried out within the framework of the ATC (Air-Traffic Control) system, a multi-agent model of the free-flight UAV operation [4]. The system provides multi-agent flight modeling of huge number of autonomous aircrafts and waypoints and no-fly zones oriented planning of the flight plan.

## II. Multi-layer Collision Avoidance Architecture

The multi-layer collision avoidance module is a part of a special planning agent, hosted by each of the cooperative UAV platforms. This module is capable of solving of the future collisions by means of combination of different avoidance methods. There is no central planner providing collision free flight plan, hence the individual plans are provided by the planning agents[1].

Each planning agent is a self-interested entity which prepares a detailed flight plan for the airplane with respect to current task specified by a series of time-specific way-points. Each UAV is surrounded by a number of concentric spherical zones. Each can have different sized zones: the *communication zone* represent communication range of the data transmitter onboard the aircraft, the *alert zone* defines the operation range of the on-board plane radar, the *safety zone* encapsulates the area around an aircraft that other aircraft are not allowed to enter in order to minimize the mutual influence of the aircraft movements and weather conditions and finally the *collision zone* defining the critical contact area.

The Multi-layer collision avoidance architecture is presented in Figure 1. CSM (*Collision Solver Manager*) is the main controller responsible for the selection of the CS (*Collision Solver*) that will be used for specific collision. CSM is able to combine all the available cooperative and noncooperative

[1]The proposed modular architecture is domain independent. Therefore is ready for deployment on autonomous vehicles like airplanes or ground vehicles.

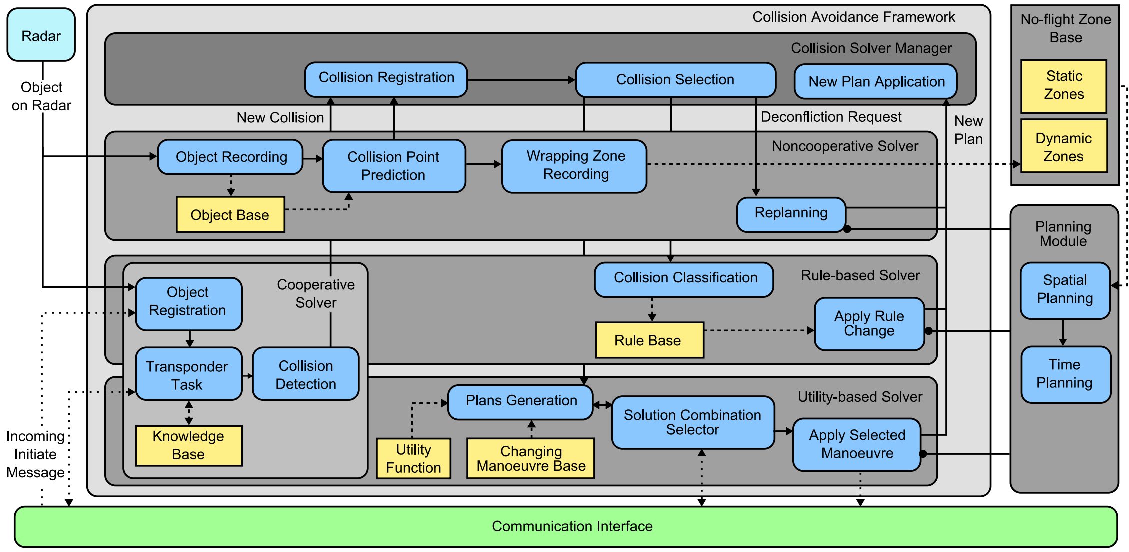


Fig. 1. Multi-layer collision avoidance architecture

algorithms. These CS algorithms are implemented as plug-in solver modules and can be domain dependent or independent. Each collision solvers is responsible for the collision detection (e.g. *Collision Point Prediction* in the non-cooperative CS or *Collision Detection* in cooperative CSs Figure 1) and collision registration with the manager. One collision can be detected by several collision solvers.

Based on priority, CSM assigns each registered collision solver a time slot that can be used for solving by the specific CS. The priority of the solvers is preset, while can be altered during the runtime. Concatenation of these slots creates time axis, see Figure 2 specifying specific, time-oriented switching among the CS operation. Sophisticated switching of the collisions solvers is inevitable in our application as the solvers have different properties. Different solvers provide different quality of the deconfliction solution, while require different amount of time for finding such solution. Specifically, the negotiation oriented solvers may provide better deconfliction solution than non-cooperative solvers, while they may be more time consuming (given by the multi-party interaction). As the time is a very critical factor in our collision avoidance domain, some solvers are not guaranteed to terminate prior a possible collision.

## III. Cooperative Collision Avoidance

Cooperative avoidance is based on communication and negotiations between airplanes. We have implemented two specific types of cooperative collision avoidance mechanisms – *rule-based* and *utility-based* algorithms. Both algorithms share the same collision detector module and the transponder task module, which is used for exchanging local flight plans and other communication between the autonomous agents representing UAVs. Both implemented methods use the peer-to-peer negotiation.

Collision detector works with two flight plans. Each flight path contains time-oriented information. The detector investigates both flight paths in time and tests the condition of safe flights defined by the size of the safety zones. If a collision is detected the common part register the collision in the collision manager.

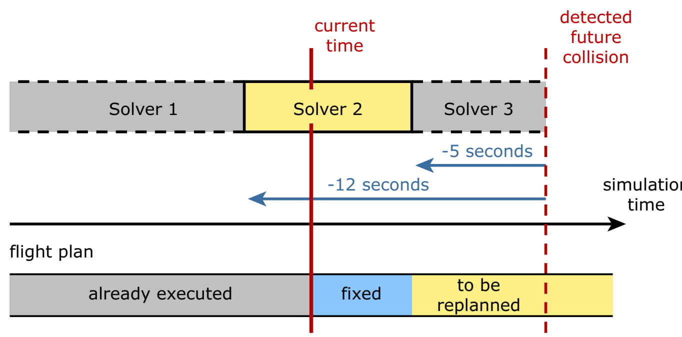


Fig. 2. Selecting process of a collision solver (Solver 3 has time slot 5 seconds, Solver 2 has time slot 7 seconds and for Solver 1 fills the rest of the time axis)

### A. Rule-Based Avoidance Mechanism

The rule-based collision solver (RB) is domain dependent algorithm. It is based on the Visual Flight Rules defined by FAA[2]. First, the type of the collision between the airplanes is identified. The collision type is determined on basis of the angle between direction vectors of the concerned aircrafts projected to the ground plane (defined by X and Y axis). Each collision type has defined fixed manoeuvre which is then applied on the detector flight plan to avoid the collision.

The above rule-based changes to the flight plan are done by both planes independently because the second aircraft detects the possible collision with the first plane from its point of view. More details can be found in [4].

[2] http://www.faa.gov

## B. Utility-Based Avoidance Mechanism

The utility-based algorithm (UB) (extended version of [7]) is a domain independent algorithm, while the manoeuvre implementation is domain dependent. Utility-based avoidance mechanism provides solution for a pair of airplanes. First, the participating airplanes select the master and the slave entities for the detected collision (usually the first entity who identifies a collision is regarded as a master entity). In the Figure 3 there is negotiation flow between both participants.

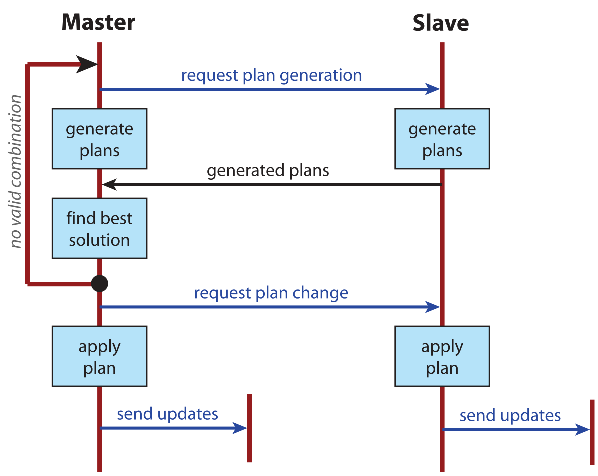


Fig. 3. The negotiation during utility-based cooperative avoidance

Each planning agent generates a set of plans using defined manoeuvres. All pre-defined manoeuvres also allows to apply different level of changes depending on the parameters. The result after applying the manoeuvre is changed flight plan including utility value for this new flight plan. The parameters during the generation process are used to get wider range of solutions in the situations when the solution is not found using flight plans with smaller changes.

The utility function is used to include the aircraft's intention to the proposed solutions of the conflict. The utility value is evaluated as weighted sum of the utility function parts using equation:

$$u = \frac{\sum_{\mathrm{i}} \alpha_{\mathrm{i}} u_{\mathrm{i}}}{\sum_{\mathrm{i}} \alpha_{\mathrm{i}}}, \tag{1}$$

where $\alpha_{\mathrm{i}}$ denotes the weight for the i component of the utility function and $u_{\mathrm{i}}$ is its value. Depending on the configuration there can be components to take into consider total length of the flight plan, time deviations for mission way-points, altitude changes, curvature, flight priority, fuel status, possible damage, type of load. Relative utility value exchanged between planes is computed as quotient of new flight plan to original flight plan. Lower value of utility function suggest the most preferred deconfliction maneuver .

There are 7 parameterized manoeuvres used in the current version of utility-based avoidance mechanism: straight manoeuvre (no change to the flight plan), turn right, turn left, turn up, turn down, speed and slow down manoeuvre.

The best possible deconfliction manoeuvres is identified by a variation of the *monotonic concession protocol* (MCP) [7]. The monotonic concession protocol is a simple protocol developed by Zlotkin and Resenschein for automated agent to agent negotiations [8], [9]. Instead of iterative comparison of the most preferred maneuvers of each party, the complete ordered set o flight plans (and labeled by the utilities) are generated and sent back to the master aircraft. When the master entity generates its own plans and receives plans from the slave entity, it tries to combine all plans together. The collision solution is then selected from cartesian product of the generated plans from both participants. These candidates for solution are ordered in increasing manner by product of utility quotients of flight plan pair. Each pair candidate is tested for a collision. If there is no collision between participants, candidate is selected as collision solution. When there are more pairs with the same sort value without collision, the the solution is selected randomly from these. The slave entity is notified about selected flight plan. This approach turns out to save substantial amount of communication and consequently makes the solver more likely to provide a solution prior a possible collision.

If there is no collision-free pair in the cartesian product, it is necessary to generate more different flight plans. In this case the request for generation of the new set of possible flight plans with greater parameters for manoeuvres is sent to slave entity. The master entity generates its new set of flight plans as well. This new plans are added to flight plans from previous round of generating and the master will repeat to select solution among all flight plans.

# IV. Non-cooperative Collision Avoidance

In the case when the communication between planes is not possible, it is required to solve the future conflict non-cooperatively. Such situation can happen e.g. when the communication device on board of the UAV is broken or if the other aircraft intentionally refuses to communicate (an enemy).

Classical non-cooperative collision avoidance methods running optimization algorithms, [11] and [12], can optimally solve collision with only one non-cooperative object. Such methods can fails when there is more non-cooperative objects simultaneously, as is presented in the experiments (section V). We have designed the method based on the dynamic no-flight zones. The non-cooperative collision avoidance loop (described below) is executed for all objects provided by on-board radar and the wrapping no-flight zones are regularly updated after each radar scan.

Designed method is based on path planning using *A** algorithm which is capable to plan flight path which doesn't intersect any defined no-flight zone. The algorithm is responsible for coordination of all operation needed for avoiding potential future collision of the UAV and an object representing non-cooperative one. It is implemented in the form of a special unit (a solver of the multi-layer collision avoidance framework, described in the section II) that can take part in the process of collision avoidance within the UAV control.

The event that triggers the deconfliction loop is information obtained from the radar providing the position of an unknown objects in the local area, see Figure 1. The observation is used for the update of the solver knowledge base. If there is too enough history available, the prediction of the collision point process is started. The collision point is defined by an intersection of the current UAV's flight plan and the predicted flight trajectory of the non-cooperative object. Current version of the algorithm uses the linear prediction estimating the future object trajectory including current velocity which require two last position with time information. The prediction provides both predicted collision point position and time information. If no such virtual collision point is found, the solver loop ends.

In the opposite case, the collision point is wrapped by a dynamic no-flight zone. All no-flight zones in the system are implemented as binary octant trees, [13]. Leaf cells of trees can be either empty of full (i.e. forbidden), other cells can moreover be mixed if the type of their subcells is not homogenous (all empty or all full). The advantage of using such octant trees is efficient data storage and fast point and line tests. Disadvantages include slow construction of octant trees and rough discretization of cells. To build dynamic no-flight zones as fast as possible, the idea of cached pre-builded versions with transformation matrix is implemented. The transformation includes translation, rotation and deformation of the tree. Such transformation is not applied to the tree itself, but inversely to the tested geometric element (point, line, sphere etc.). In other words, the transformed tree is no projected to the world space, but vice versa the tested element is projected from the world coordinates to the tree space.

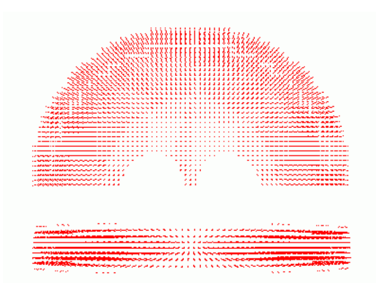

Fig. 4. The shape of the dynamic no-flight zone

The shape of dynamic no-flight zones of non-cooperative object is derived from its possible future flight trajectory, see Figure 4. The trajectory takes into account the minimal turning radius, maximal climbing/descending angle and the prediction time. The predicted dynamic no-flight zone is put to the common knowledge base of the UAV. Zones are used by all components of the UAV control which uses path planning module.

At this point the non-cooperative solver passes information about detected future collision to the collision avoidance manager that will decide when the collision will be solved. It makes decision considering the time of the future collision and avoidance framework configuration.

If the manager module decides that the particular non-cooperative collision should be solved, the last step of the non-cooperative submodule is executed. A new flight plan from the current position fulfilling all mission way-points is generated using planner module. The planner module spatial planning respects all defined no-flight zones in the UAV knowledge base and the produced flight plan doesn't intersect any one. Finally, the new flight plan is applied and handled by the control UAV components.

## V. EXPERIMENTS

The algorithms described in the paper have been evaluated on huge set of experiments: comparison of both co-operative collision avoidance methods, comparison of our dynamic zone-based non-cooperative approach with traditional proportional navigation, and multi-layer collision avoidance architecture validation with real aircraft traffic.

### A. Cooperative Collision Avoidance

The scalability experiments comparing rule-based (RB) and utility-based (UB) collision avoidance have been carried out within the limited square area of 31 x 31 units. To provide appropriate comparison the utility-based method was limited to use only changing maneuvers that do not change the altitude of the UAV. The sequence of 850 runs (configuration using 5, 10, 15 ... 85 simultaneous aircraft, each run 50 times repeated to provide average result values) has been measured for both methods. The size of the UAV's safety zone radius is 0.25 units and radar range radius is 10 units. The flight speed of the UAV can vary between 0.075 and 0.125 units per second with acceleration and deceleration 0.05 of the speed per second. Within this batch of the experiments there were 76490 UAVs simulated within 230 hours of the flight time.

The worst-case scenario was used. The planes were generated randomly on one of the four sides of the experiment area with the intention to fly across the square to the opposite side. All UAVs fly on the same flight level. This setup provides the high number of the collision situation in the central part of the limited area.

The figure 5 presents the simulation results. Top chart plots the average number of the safety zone violations – there is other UAV located nearer than safety zone radius. The UB method provides better result (almost zero violations) than simple RB. We observed that also number of the collision avoidance loops is smaller for the UB (for 80 UAVs it is 4 loops per UAV and 1.07 for UB). The communication data flow used for the negotiation among UAVs was also measured. We observed that the average flow is almost same for both methods (total flow among all for 80 UAVs was 40 MB; communication peek between pair was 13 kB per second). The UB method needs exchange more data is each avoidance loop.

The bottom plot in the figure 5 shows the average sum of differences between the final deconflicted and the initial flight plan for all planes in the specific experiment configuration. We can observe that for the UB method is much better solution

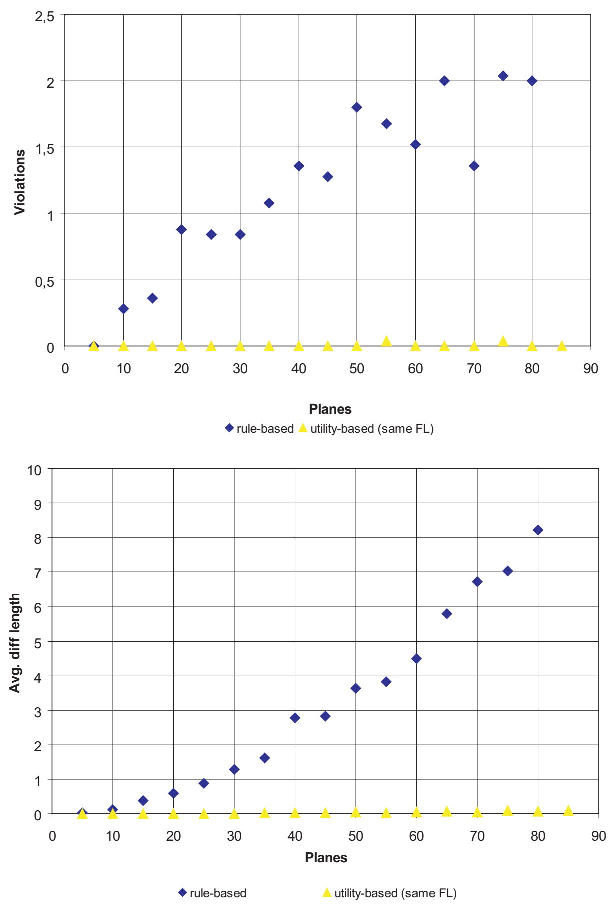


Fig. 5. Average number of safety zone violations (top). Average diff length of flight plan (bottom).

(closer to optimum) than for RB. We have analyzed this result and UB is better because it can keep almost straight flight plan and solve collision by acceleration and deceleration giving the best value for the utility function.

### *B. Non-cooperative Collision Avoidance*

Our non-cooperative algorithm described in this paper have been compared with the optimalization proportional navigation (PN) [12]. Three scenarios with one controlled and one uncontrolled UAV have been carried out. The uncontrolled UAV always flies directly from the starting point to the destination and the controlled one is always heading north and it must avoid collision in the middle of the operation area (scenarios are planar – no altitude changes allowed).

| type | average trajectory length $l[u]$ | | |
|---|---|---|---|
| | proportional navigation | dynamic NFZ | no control |
| perpendicular collision | 33,21 | 33,85 | 30,0 |
| slant collision | 31,17 | 31,52 | 30,0 |
| head-up collision | 30,81 | 30,62 | 30,0 |

TABLE I

Table I shows comparisons of flight path lengths of average trajectories (10 measurements) using PN and dynamic no-flight zone (dynamic NFZ) non-cooperative methods. The proportional navigation in the tests with two aircrafts provides better results due to the fact that PN plans the trajectory it touches the edge of the safety zone. On the other hand, dynamic NFZ leaves greater distance between the trajectory and the object considering possible change of direction of the object.

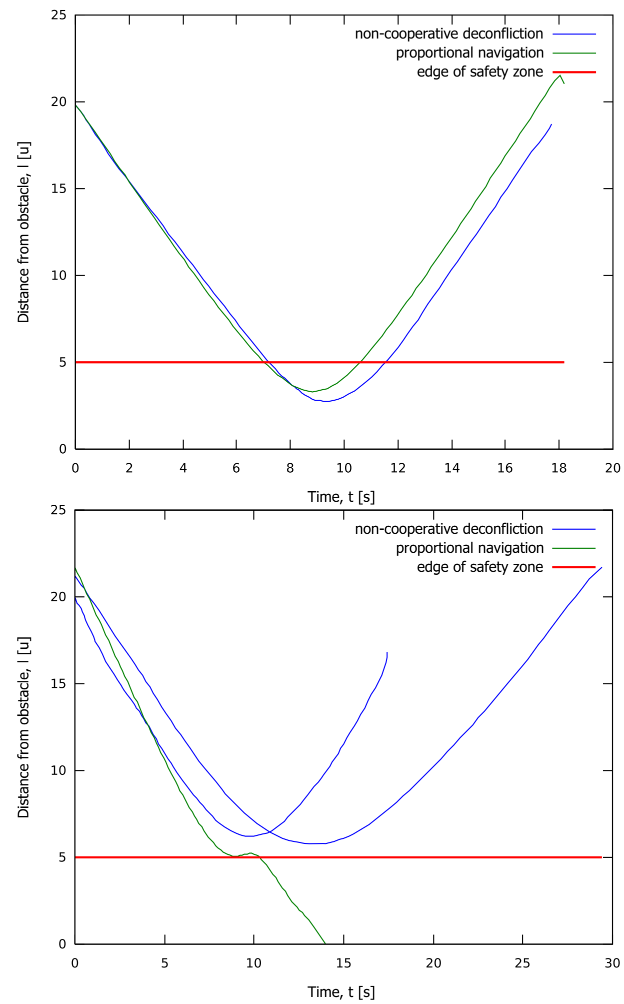


Fig. 6. Two UAV head-up collision scenario (top). Three UAV scenario with proportional navigation failure (bottom).

Figure 6 presents the minimal separation among UAVs with safety zone size highlighted. Top plot provides results for the one controlled and one uncontrolled UAVs head-up scenario. The safety zone for both methods is clearly breached. For PN it is caused by restriction of the maximal acceleration of the airplane, while in the case of dynamic NFZ it is caused by deformation of the dynamic zone not respecting the safety zone size. The bottom plot shows the result for the scenario with two uncontrolled UAVs (obstacles) and PN was configured to take into consideration the nearest obstacle. The PN algorithm fails to the collision and dynamic NFZ works properly. We have performed several other experiments with more UAVs (up to 10) in the restricted area all using described dynamic NFZ algorithm within worst-case scenario. The proposed non-cooperative method handle all situations without any collision.

### C. Multi-layer Collision Avoidance Architecture Validation

Described multi-layer collision avoidance architecture has been validated in the environment where operate real air-traffic over Los Angeles International Airport. Agent controlled UAVs are configured to use utility-based cooperative method with other agent based UAVs and use NFZ-based non-cooperative method against imported air-traffic at the same time.

## VI. Conclusion

In the paper we present multi-layer collision avoidance architecture which is used for distributed air space separation of several agent-controlled autonomous aircrafts in the Air Traffic Control system, [4]. The multi-layer architecture provides to use different collision avoidance methods at the same time. The selection of the appropriate method is managed by the collision solver controller which takes into consideration current configuration and the fact if there can be established communication between planes (trusted aircrafts, same communication protocol, etc.).

There are two cooperative methods implemented: rule-based and utility-based. As is shown in the experiments utility-based method has several advantages and allow possibility for integration of the aircraft intentions in the used utility function. Both methods works in peer-to-peer manner. So, the multi collision situation (collision of more UAVs at the same place) is handled in the iterative way. We are working on the formal model in which we will provide formal proof of the convergence of the iteration to the stable solution. To minimize the number of iteration within the multi collision situation we plan to extent proposed utility-based algorithm to allow negotiation of more that two aircrafts at one iteration.

Within the experiments we have measured the communication flow among the simulated aircrafts using cooperative negotiation-based collision avoidance methods. In some particular applications (operation of the UAVs in the enemy territory) it will be beneficial to minimize the necessary communication data flow among aircrafts to minimize the risk of reveal of their operation, but still profit from the benefits of the cooperative collision avoidance.

During the experiments, validating described no-flight zone-based non-cooperative method, we have found two modules which can be extended to provide better results. Current version of the algorithm uses simple linear prediction for the collision point with non-cooperative plane. The extension of the prediction algorithm which will provide more accurate future position of the observed object will provide better source for the wrapping mechanism of object's future states to the dynamic no-flight zones.

Second problem, which arises during the experiments with designed non-cooperative solution, is related to the improvement of the path planning algorithm and its capability to plan path avoiding no flight zones. Current planner plans the future path without taking into consideration the size of the safety range. The problem can be solved using enlarged defined zones up by UAV's safety zone size. This operation over the octant trees is non-trivial and time expensive. Using a set of pre-builded different zone shapes cannot be used due to various possible parameters and its large memory consumption. Now we are searching for other dynamic zone representation that will provide same or better performance as current system, but which will provide fast transformation with respect to the future dynamically sized safety region.

## Acknowledgment

Effort sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA8655-06-1-3073. The U.S. Government is authorized to reproduce and distribute reprints for Government purpose notwithstanding any copyright notation thereon[3].

## References

[1] DOD, "Unmanned aircraft systems roadmap 2005-2030," 2005.

[2] J. C. Hill, F. R. Johnson, J. K. Archibald, R. L. Frost, and W. C. Stirling, "A cooperative multi-agent approach to free flight." in *AAMAS*, 2005, pp. 1083–1090.

[3] T. Sandholm, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. Cambridge, MA.: MIT Press, 1999, ch. Distributed Rational Decision Making, pp. 201–258.

[4] D. Šišlák, M. Rehák, M. Pěchouček, D. Pavlíček, and M. Uller, "Negotiation-based approach to unmanned aerial vehicles," in *DIS '06: Proceedings of the IEEE Workshop on Distributed Intelligent Systems: Collective Intelligence and Its Applications (DIS'06)*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 279–284.

[5] J. Krozel, M. Peters, K. D.Bilimoria, C. Lee, and J. S. Mitchel, "System performance characteristics of centralized and decentralized air traffic separation strategies," in *4th USA/Europe Air Traffic Management R & D Seminar*, Stanta Fe, NM, December 2001.

[6] J. P. Wangermann and R. F. Stengel, "Optimization and coordination of multiagent systems using principled negotiation," *Journal of Guidance, Control, and Dynamics*, vol. 22, no. 1, pp. 43–50, 1999.

[7] S. Wollkind, J. Valasek, and T. R. Ioerger, "Automated conflict resolution for air traffic management using cooperative multiagent negotiation," in *Proceedings of the American Institue of Aeronautics and Astronautics Conference on Guidance, Navigation, and Control*, Providence, RI, 2004.

[8] G. Zlotkin and J. S. Rosenschein, "Negotiation and task sharing among autonomous agents in cooperative domains," in *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, N. S. Sridharan, Ed. San Mateo, CA: Morgan Kaufmann, 1989, pp. 912–917.

[9] J. S. Rosenschein and G. Zlotkin, *Rules of Encounter*. Cambridge, Massachusetts: The MIT Press, 1994.

[10] J. C. Hill, F. R. Johnson, J. K. Archibald, R. L. Frost, and W. C. Stirling, "A cooperative multi-agent approach to free flight," in *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*. New York, NY, USA: ACM Press, 2005, pp. 1083–1090.

[11] C. Tomlin, G. Pappas, and S. Sastry, "Conflict resolution for air traffic management : A study in muti-agent hybrid systems," 1998.

[12] S.-C. Han and H. Bang, "Proportional navigation-based optimal collision avoidance for uavs," in *Second International Conference on Autonomous Robots and Agents*, S. C. Mukhopadhyay and G. S. Gupta, Eds. Massey University, New Zealand, 2004, pp. 76–81.

[13] S. Frisken and R. Perry, "Simple and efficient traversal methods for quadtrees and octrees," *Journal of Graphics Tools*, vol. 7, no. 3, 2002.

[3]The views and conclusions contained herein are those of the author and should not be interpreted as representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

# Coalition Interoperability Architecture

E. Kuster, Decision Automation Group, Command and Control Division, DSTO, Australia
egon.kuster@dsto.defence.gov.au

**Abstract**– Current military trends indicate that it is highly likely future military operations will require a coalition force. With a majority of data being created or captured in electronic form the importance of a coalition information sharing solution has become paramount to allow coalition partners to exchange and share data in an automated manner. This paper describes a Coalition Interoperability Architecture that allows nations to connect to the coalition network and easily share released national data to other coalition partners, while taking into account differences in policies and requirements of each coalition nation.

## 1. Introduction

Australian defence doctrine states:

*"The sharing of information with potential coalition participants is crucial to building trust and confidence among possible coalition partners."* [1]

Therefore a coalition information sharing solution for nations participating in a coalition operation is extremely important. This is especially true if the solution helps automate the process and allows greater access and audit controls on data released to coalition partners.

Between 2001 and 2004 the Australian Department of Defence and the United States Pacific Command (USPACOM) were engaged in the development of the Coalition Theatre Logistics (CTL) Advanced Concept Technology Demonstrator (ACTD) [2-5], which resulted in a system design for the exchange of logistics data between coalition forces. This paper draws upon the CTL ACTD concepts and provides a discussion on a high level architecture for developing solutions that require Coalition Nations to share structured information.

It widely recognised that when operating in a coalition force a key enabler is the capability of all involved nations to be interoperable [6]. However, the problem is **how** to achieve interoperability. One approach often discussed initially is the purchase of a foreign nation's information systems with the view that if all nations use the same system then interoperability will be achieved. This approach however encounters the following problems:

- Additional training requirements, due to a nation using different systems in their national environment to the coalition environment.
- Difficulty to utilise the foreign nation's system within the national environment due to differences in processes, data, policies, system requirements.
- Using the same software does not guarantee interoperability due to local customisations and configuration details.[1]
- Version updates between ALL nations must be kept sychronised. This is unlikely due to vast differences between acquisition policies and proceedures.

Developing a new single coalition system has the same problems as using a foreign nation's information system; although, a coalition system is better at supporting the coalition data requirements due to its custom development.

Individual system integration is another approach often discussed, where each nation develops an integration layer to another nation's system. This can create fragilty and dependence on maintaining numerous integration components for a plethora of systems owned by other nations as shown in Figure 1.

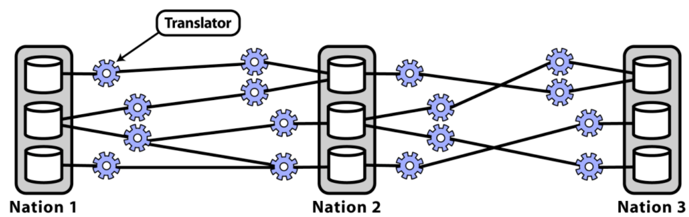


**Figure 1 -** Multiple Systems Integration

None of the previous options deal with the issues of data releasability, authoritative data sources, dynamic coalitions, support to multiple concurrent coalition operations, system ownership/administration/maintenance, coalition agility, integration flexibility and support for future requirements.

This paper describes an architectural approach that endeavors to provide these capabilities. A description of the architecture's main components is provided in section 2. Section 3 discusses data standards, mechanisms for moving data, authoritative data sources, interface standards and data release. In order to maintain and manage the architecture, section 4 explains the architecture's distributed ownership and control mechanisms. Section 5 discusses how the architecture can support a nation's involvement in multiple coalition operations and multiple coalition networks. The architecture's performance and scalability requirements are discussed in section 6. With section 7 providing a conclusion, identifies issues and future work.

[1] For example two Enterprise Resource Planner (ERP) instances may not integrate due to differences in their use and configuration even though both may be implemented with the same software product.

## 2. ARCHITECTURE

The Coalition Interoperability Architecture of Figure 2 contains 3 environments; **National Domain**, **Cross Domain**, and **Coalition Domain**. These three environments allow for the separation of responsibility and provide the necessary abstraction between national information systems, national systems used in coalition operations and coalition maintained information systems. Each domain is described further in the following sub sections.

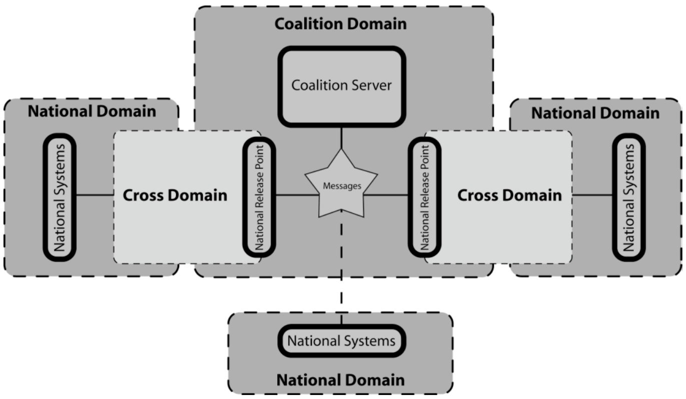


**Figure 2 –** Coalition Interoperability Architecture

### National Domain

Almost all nations involved in a Coalition force will have their own national information systems to support their operations, these are defined as the National Systems in Figure 2. These systems are contained within the National Domain, which is maintained by the owning nation. The National Domain does not specifically need to be located within the owning country's geographical location but may be located within the Area of Operations (AO). The key differentiator is that National Domain systems are only accessed and used by the owning nations personnel to manage their internal operations.

To facilitate coalition partner interactions and align multiple coalition activities there is a need to share data contained in national systems with other coalition partners. This requirement to share is why the owning nation must integrate their national systems with the Cross Domain components, which provide the required release, transformation and expose functions as described below.

### Cross Domain

National systems containing data to be released to coalition partners must be integrated with the Cross Domain infrastructure (see Figure 3), which then provides the means to transfer data between the National and Coalition Domains. Integration can be achieved using any means necessary but is done only once for each national system. The Cross Domain then provides the additional functions to perform data transformations, data release, transfer data between national and coalition networks, and expose the released data to the wider coalition community.

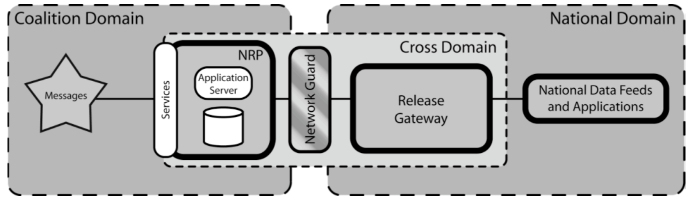


**Figure 3 –** Cross Domain Detail

A key component of the Cross Domain is the National Release Point (NRP), which provides the coalition visible front-end of the national systems and data. All data pushed or requested must go through the NRP. Each nation's NRP provides the necessary services to expose data (data pull), register and produce events (data push) and provide national security rules[2]. Using the NRP to store and expose data so that only interested nations access the information the coalition network's utilization should be reduced as data is not transmitted to all nations. In comparison if only a push mechanism or an alternative architecture using a single centralized server was used then the impact on the coalition network would increase; therefore the use of NRP services is highly recommended. The NRP can also contain an application server, which allows coalition users to access national applications from within the coalition network, if so desired by the owning nation.

Transformation services are required to convert data from national data formats into an agreed coalition data formats to support interoperability. More detail on the coalition data formats are discussed in section 3 (data management). These transformations can occur at multiple locations:

- Integration of nation system – transform as the data is extracted from the national system,
- Release Gateway – transform as part of the security releasing process,
- Entry into the NRP – transform the data after it has crossed between the national and coalition networks and loaded into the NRP,
- Creation of data message – transform data when exposed as a service or sent as a message[3].

As it is the responsibility of the owning nation to design and develop their Cross Domain implementation it is a national decision in how and when the data is transferred as long as coalition data formats are exposed or sent from the NRP. This is also true for the security services on the NRP and on the development of the Release Gateway component.

Most nations involved in a coalition are not willing to heedlessly share all data contained in the national systems to coalition partners; therefore a Release Gateway is most likely required but not mandatory. It is the responsibility of the owning nation to determine the Release Gateway's implementation details and its security and policy details.

[2] See section 3 (Data Management) for more information about push and pull data transfer mechanisms.
[3] See section 3 (Data Management) for more information about services and events within the Coalition Domain.

### Coalition Domain

The Coalition Domain is fundamentally just the coalition network that provides the communication bearers between coalition nation's Cross Domain components. For nations that do not have the resources or requirement for a Cross Domain (and NRP) then it is possible for them to use the services and applications housed on the Coalition Server to interact and share data with partner coalition nations (see Figure 4). Use of the Coalition Server will result in more manual processes, as automation is not easily possible.

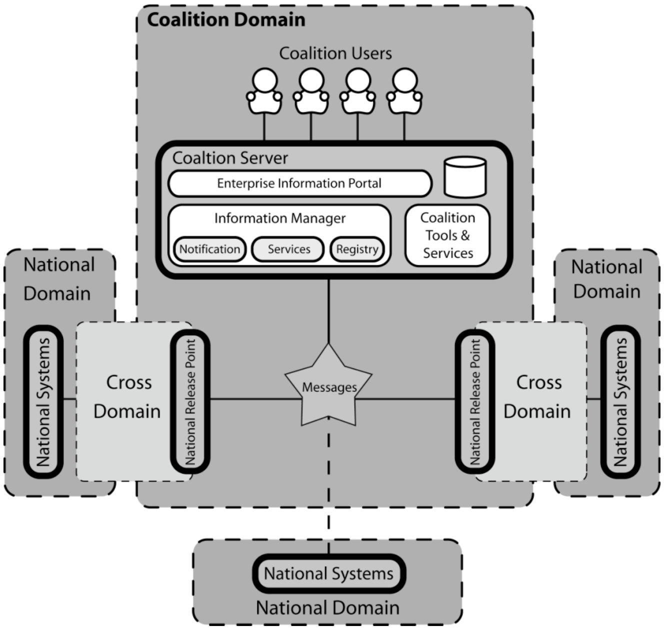


**Figure 4 –** Coalition Domain Detail

The Coalition Server provides additional services to help manage and coordinate interactions between coalition nations such as service registry, centralized data storage, and the aforementioned coalition applications for manual data accesses and entry. All architecture diagrams in this paper mention the existence of the Coalition Server; however this component is **optional** as nations can successfully communicate by sharing messages between services housed on each other's NRPs directly.

## 3. DATA MANAGEMENT

The sole purpose for developing a system described in this paper is to share data with fellow coalition members. This can be in the form of documents/messages or even functionality provided by a nation's own system, such as requesting additional movement support. No matter what functionality, the key to correctly implementing any coalition system is with specifying and managing data correctly.

### Coalition Data Standards

Nations will always have differing data requirements. This may be due to specific systems used, different processes, policy requirements or even procurement approaches. For example, the United States military utilise Unit Type Codes (UTC) to classify units and is used throughout the United States logistics information systems. In Australia there is no concept of a standard UTC as many Australian units are customized and do not conform to a standard type, this is especially true for Australian units involved in coalition operations. Therefore, it is impossible to translate directly between US and Australian data formats that use UTC. Therefore any coalition interoperability architecture will require some form of common language that all coalition nations agree to.

There is current research investigating the use of ontologies and semantics, which hopes to define an approach to automate the connection and relationships between different data sources [7]. The Semantic Web [8] is one name given to the research of defining methods for allowing machines better understand the data that they handle and therefore perform more advanced transformations, data fusion or even automatically respond to events. The use of semantics in a Service Oriented Architecture (SOA) as at that described herein is also known as Semantic Service Oriented Architectures (SSOA) [9]. At the time of writing this research has not developed any useable results directly applicable for coalition integration described herein to automate data relationship development, connections and transformation of data produced by one party and then consumed by another. For this reason this architecture utilizes a standardized coalition data standard that all nations must transform their national data into prior to sharing with other coalition nations.

Coalition data standards must be produced for each data classification item such as a movement request, military units or facilities data. Each data item must support the following requirements:

- All data item shall be **self-describing** with all data contained within to facilitate data transfer and storage (see next section "Message Orientation" for more information).
- All data items shall contain a **core data set** that is mandatory. By maintaining a relatively small core set of data attributes (mandatory) and asserting a majority are kept optional ensures that nations will successfully create the required data either via manual means (eg. using the Coalition Server) or through automated transformation from national systems via the Cross Domain. The smaller the core data set the easier for nations to utilise the coalition data standard; therefore, the core data set must be scrutinized closely so that only required data is included. A simple test of 'if an attribute could be removed without effecting the overall data item's meaning then it must not be included in the core data set' facilitates this. For example, in a movement request the origin and destination attributes must be part of the core data set since without these the movement request would not make sense; while the required delivery date should not be included as it is possible to have a movement request that does not have a defined delivery date and still be valid.
- All data items shall be **extensible**. By allowing extensions in the data system, nations that wish to define additional data will be allowed. These extensions will likely be used to exchange additional data between interested nations, and ignored by others. Ensuring additional data is defined via extensions will help keep the core data item description concise and facilitates quicker uptake.

Extension usage should be reviewed periodically with the purpose of potential incorporation into the data item definition if used frequently by multiple nations.

Any coalition data standards produced must be maintained; therefore a configuration control board must be appointed to ensure the upkeep of the set of data item descriptions. It is possible that an existing multi-national group be charged with this responsibility; however, members involved must be users of the data items to ensure continued relevance and advocacy.

**Message Orientation**

When developing and testing this architecture during CTL ACTD initially a centralized database approach was used where data items were distributed via the network and stored in a centralized coalition relational database structure. This meant that a unit contained within a movement request was the same unit in a coalition movement plan or located on a specific facility and therefore required a unique identifier and all relationships to be created when stored in the central database. This approach quickly became difficult to manage due to numerous data item relationships. Managing data being produced and utilised by multiple parties becomes demanding to track and results in a data integrity degradation and potential data duplication. Trying to maintain data relationships also vastly increased the complexity of the data ingestion components. Therefore instead of relational data structure be used a more document based or **message-oriented** approach be used.

Message-oriented data is where data producers provide all relevant data within the single data item and is sent as a complete message to the receiving party/parties. This means the receiving party can act upon this message without requiring additional information or maintaining a complex database of relationships between data items. Each message sent must also contain the authoritative source of the enclosed data.

**Authoritative Data Sources**

Identifying the source of all data within the coalition is extremely important to ensure the origin of a particular data item. Without this information it is impossible for a nation to know where best to request data updates for information already received. This is especially true if a nation is receiving similar data from multiple nations. To support this requirement all data items should contain a source attribute that identifies the data item's authoritative source.

**Pull versus Push of data**

There are two methods of sharing data after it has transferred between the National and Coalition Domains via the Cross Domain infrastructure; either **push or pull**.

The first method is where data is stored in the producing nation's National Release Point and exposed to other coalition nations in order to **pull** or request the data. This is achieved by exposing a series of data services on the National Release Point, which can be accessed by other nations. This is also known as a Service Oriented Architecture (SOA) as data is accessed via a set of services.

The second method is to **push** or send data to either a series of subscribed nations or predetermined national receivers. As this is essentially the production of events that are either sent to set of subscribed nations or to a list of nations determined by the event producers this approach is known as an Event Based Architecture (EBA).

Both approaches are valid and can be used together. For example a cargo tracking receipt data item could be produced by Australia that is transformed into the coalition data item through the Cross Domain infrastructure and then published as an event and sent to all interested nations that are subscribed to cargo tracking events from Australia, which is a data push approach. This same data item can then be stored within the NRP and then exposed via the cargo tracking receipt data item service so that other coalition nations can gain access to the information and other historical data items, which follows the data pull approach. It is up to the coalition to determine what method of data sharing will be utilised as this architecture supports both approaches.

**Coalition Interface Standards**

In addition to data standards a set of interface standards are also required. These standards define how data is pulled from the NRP or events produced by the NRP. The architecture is implementation agnostic; however it is recommended to use a set of common event and service standards as it is more likely to have both vendor/tool support and be accessible by all coalition nations. The Web Service[4] set of standards [10, 11] is the logical choice for exposing data contained in the NRP as they widely supported and provide easy integration between heterogeneous components. For each data item a set of service interfaces must be agreed upon[5].

For the eventing system there are multiple options including such standards as WebSphere MQ, Java Messaging Service (JMS) and Web Service Eventing (WS-Eventing) [12]. Many of these standards are tied to a specific platform, which contradicts the aim to allow nations to have full control over their NRP development. It is therefore likely that the WS-Eventing standard be used as it is platform agnostic; remarkably this standard is currently not widely used.

**Data Release**

The Cross Domain is a vital component in the architecture when dealing with data release. All nations involved in the coalition will not wish to release all the data they have contained in their national systems, therefore a secure release mechanism should be used. As the National Domain and Cross Domains are the responsibility of the owning nation the details for implementation of the release function will be up to owning nation (ownership and responsibility is discussed more in section 4).

A contract type release mechanism would enforce conforming to a release contract and enable automation. As large amounts of data may transfer the Cross Domain it is highly recommended that an automated release approach be used where possible to help ease the burden of checking data before release.

---

[4] The core set of Web Service standards includes SOAP, WSDL and UDDI. SOAP describes the message structure, WSDL describes the service interface and UDDI provides a registry mechanism.

[5] A discussion on defining services type and its level of detail is outside the scope of this paper.

This architecture assumes that once data has been approved for release from the National Domain to the Coalition Domain that is should be deemed releasable to **all** nations currently connected to the Coalition Domain / Coalition Network. The reason for this assumption is that there is no agreed and approved solution for ensuring that data only goes to selected nations and is not redistributed or exposed via another nation's NRP services. It is possible to enforce data encryption, however the administration overhead could be extremely high in order to manage encryption keys and ensuring identities are current and valid. Releasing data to selected nations in the same coalition force is also against the original idea of sharing data to help support the coalition operation. However, a set of sub-coalitions or bi-lateral connections using the same architecture could be used which operate on separate networks to provide this type of functionality (see section 5 for more information on supporting multiple coalitions).

## 4. Distributed Control

If a centralized solution were to be used for data sharing then a large problem is who owns and maintains this central solution. Even if a nation volunteers to provide support the issue remains as what happens when the other nations are involved in an operation that does not involve the hosting nation? Another potential solution is the use of a commercial third-party; however this differs little to the aforementioned solution as the issue migrates to what happens if the third-party does not have the required resources, or has conflicting requirements or no longer wishes to provide support as it is not commercially viable? An acceptable option, which is utilised in this architecture, is one of distributed control, where all involved and interested parties (the coalition nations) each have part ownership and responsibility in ensuring the working order of the complete solution.

### Ownership and Responsibility

When a nation joins a coalition force it is the responsibility of that nation to provide the required infrastructure and components to connect their Cross Domain to the coalition network. The joining nation also maintains ownership of their NRP and all services and data contained within. A nation is therefore able to remove data quickly and easily from the Coalition Domain by removing it from their NRP.

By ensuring that all nations are responsible for their own Cross Domain and connection with their national systems it ensures that no one nation bears the burden of ensuring the operation of the shared coalition environment. This also means that there is no single failure point. Having each nation responsible for their own NRP design and development is beneficial as each nation can implement their NRP using skills and resources available rather than lock into a particular solution; however the final solution must adhere with the coalition's standards for data and service interfaces.

An additional benefit is as nations have a vested interest and do not just rely on another party to provide the capability it is more likely that nations will remain engaged to further develop and maintain the solution and standards.

## 5. Multiple and Dynamic Coalitions

It is common that coalition operations have nations that join and leave over time. This would mean that for the duration of an operation it is unlikely to have the same set of coalition partners. Therefore the Coalition Interoperability Architecture must support dynamic joining and leaving of partner nations. Additionally any one nation may be involved in multiple simultaneous coalition operations. Each coalition operation is likely to operate using different coalition networks or at least cryptographically separate network enclaves.

### Dynamic Coalitions

As data is kept within the National Release Points in the Cross Domains there is no need to explicitly support the joining and leaving of nations. For example, when a nation joins a coalition force and connects their Cross Domain up to the Coalition Domain then this action would register their NRP as available in the Coalition Domain and data sharing can start immediately. When a nation is no longer involved all that is required is for the nation's Cross Domain and NRP be removed from the Coalition Domain and because the Cross Domain contains all national coalition released data and services they will no longer be present in the coalition system. However, if a nation wishes to leave data behind it is possible for this data be pushed to the Coalition Server for long-term storage.

The caveat to the above statement is that this does not hold true if data has already been exposed and accessed by another nation or sent to another nation as it is possible for that other nation to re-expose or resend the event even if the originating nation has left and removed their Cross Domain components.

If a centralized architecture was used removal of data would be considerably more difficult due to the leaving nation not owning the central data store and therefore having guaranteed access to delete all national data contained within. By storing data only in the Cross Domain that can be removed from the Coalition Domain overcomes these ownership and access issues.

### Multiple Networks

As nations are likely to be involved in multiple concurrent coalition operations or require limited-access bi-lateral or greater communications it is highly likely that the National Domain will need to connect to multiple Coalition Domains on multiple coalition networks.

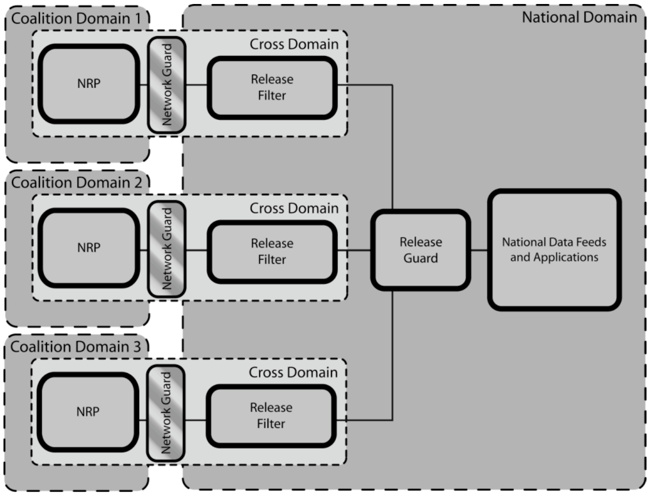


**Figure 5 –**Multiple Coalition Domains

As shown in Figure 5 it is possible for one nation to connect their national information systems to multiple coalition networks. As discussed in section 4 it is the responsibility of the owning nation to design and produce the Cross Domain components. The design depicted in Figure 5 is only one potential solution. By implementing a release guard that automatically tags data on which nations the data is releasable to or which coalition network the data is destined for it is possible to control how data is distributed to multiple Coalition Domains. Once released the Cross Domain component handles the transfer, data exposure and event production necessary to disseminate the data to coalition partners.

## 6. Scalability and Performance

This architecture utilizes a distributed approach for data storage and services by encouraging each nation to maintain their own NRPs. It is believed that this solution will scale to multiple connected coalition nations. The biggest issue will always be with the capacity of the coalition networks. The NRPs are another component that may become stressed if many nations request data simultaneously from the same nation. As it is the responsibility of each joining nation to design and develop their own NRP it is highly recommended that capacity and availability planning be conducted as part of the nation's NRP design activities. It has also been shown that this architecture can be extended to support multiple concurrent coalition nations or even multiple enclaves within a single coalition force through the use of multiple Cross Domain components and a robust release system.

This paper suggests the use of Web Service standards for NRP services and eventing system. As Web Services use XML for the messages sent over the network it is commonly thought to be slow and lacking performance. However the computer speed has been increasing and the costs decreasing which means that any performance issues are no longer an issue. Additionally XML specific compression can extract bandwidth performance gains. Many large commercial and government organisations are already using Web Services for critical services in their enterprise. It is therefore expected that the use of Web Services will not adversely effect the operation of the overall architecture.

## 7. Issues, Future Work, Conclusions and Discussion

When the CTL ACTD commenced implementation of a version of this architecture in 2005 it was found that the coalition networks of that time were unable to support any capability within the Coalition Domain. Therefore CTL ACTD altered its deployed capability to only send data between nations from within the National Domains. This altered solution did not allow for easy handling of multiple nations within the same coalition or support coalition applications run on either the Coalition Server nor national NRPs. Therefore before embarking on implementing the architecture described herein it would be prudent to check that today's coalition networks can support servers located on the coalition side of the network firewall so that the NRPs and Coalition Server can be located and accessed from within the Coalition Domain.

When a nation enters or leaves the coalition force they are expected to bring or remove the NRP, which contains their national data and services. One issue when accessing these NRPs is to first find the location of the other national NRPs. The method employed within CTL ACTD was to statically define the location of each NRP. A better method would be to use a central registry (such as on the Coalition Server) that each national NRP can register itself against. However, this approach is still not optimal as it provides a single point of failure, the coalition server, and does not support instances that do not contain a Coalition Server. Therefore another approach would be to use something like the Web Service Discovery (WS-Discovery) standard [13] which would allow for each NRP's services to be dynamically found by all other NRPs on the coalition network.

Of significance in this paper has been the discussion on the topics and concepts that must be considered when embarking any coalition interoperability exercise such as message oriented data, responsibilities and ownership, support of nations joining and leaving coalition forces and ability to use the same architecture over multiple coalition networks.

A remaining goal of interoperability is the automated management of data as discussed in section 3 which may be ameliorated by concepts as the Semantic Web and SSOA. As these tools and technologies advance it will be increasingly possible to reduce mandating of data standards as it will be possible for translations between national formats to occur on the wire by defining semantic relationships between data items. Unfortunately current semantic tools and engines are in their infancy. Therefore the architecture described herein provides an archetype for implementing coalition interoperability and provides placeholders where technologies such as the Semantic Web and SSOA can be introduced when available.

## References

[1] Australian Defence Force Warfare Centre, "Australian Defence Doctrine Publication 00.3 - Coalition Operations," Australian Defence Headquarters 2002.

[2] E. Kuster, "Coalition Theatre Logistics Interim Operational Capability Spiral 1 Trial," Command and Control Division, DSTO, Technical Note DSTO-TN-0625, 2005.

[3] E. Kuster, "Coalition Theatre Logistics Interim Operational Capability Spiral 2 Trial," Command and Control Division, DSTO, Technical Note DSTO-TN-0666, October 2005 2005.

[4] E. Kuster and A. Baldock, "Coalition Theatre Logistics (CTL) Australian Translation Service (ASTS) User Guide," Command and Control Division, DSTO, Technical Note DSTO-TN-0667, 2005.

[5] E. Kuster, A. Baldock, and S. Geoghegan, "Coalition Theatre Logistics (CTL) Australian Translation Service (ASTS) Developer Guide," Command and Control Division, DSTO, Technical Note DSTO-TN-0665, October 2005 2005.

[6] "Coalition Operations Demand Technology Solutions: Warfighting across hald the globe requires partners that interoperate," *Signal*, pp. 67-72, 2005.

[7] J. T. Pollock and R. Hodgson, *Adaptive Information - Improving Business Through Semantic Interoperability, Grid Computing, and Enterprise Integration*. New Jersey: John Wiley & Sons, 2004.

[8] Wikipedia Community, "Semantic Web," vol. 2007: Wikipedia.

[9] S. Chance, "Semantic Service Oriented Architecture - White Paper Overview," vol. 2007.

[10] E. Christensen, F. Curbera, G. Meredith, and S. Weerawarana, "Web Services Description Language (WSDL) 1.1," W3C, 2001.

[11] M. Gudgin, M. Hadley, N. Mendelsohn, J.-J. Moreau, and H. F. Nielsen, "SOAP Version 1.2 Part 1: Messaging Framework," W3C, 2003.

[12] S. Tarkoma and K. Raatikainen, "State of the Art Review of Distributed Event Systems," Helsinki Institue for Information Technology 19 February 2006 2006.

[13] J. Beatty, G. Kakivaya, D. Kemp, T. Kuehnel, B. Lovering, B. Roe, C. S. John, J. Schlimmer, G. Simonnet, D. Walter, J. Weast, Y. Yarmosh, and P. Yendluri, "Web Services Dynamic Discovery (WS-Discovery)," Microsoft April 2005 2005.

# Supporting Collaborative Operations within a Coalition Personnel Recovery Center

Gerhard Wickler, Austin Tate, and Jeffrey Hansberger

***Abstract*—I-X is a framework that can be used to create an application in which multiple agents adopt a task-centric view of a situation, and which supports the necessary coordination of their activities to respond to that situation. The I-X Process Panel provides the functionality of a to-do list and instant messaging and thus, it is a useful tool when it comes to organizing the response to an emergency. However, I-X goes well beyond this metaphor and provides a number of useful extensions that facilitate the finding and adaptation of plans for teams to respond in dynamic situations.**

**In the Co-OPR (Collaborative Operations for Personnel Recovery) project, the I-X framework has been used to support training exercises for personnel recovery. This paper will describe some of the initial findings that are the result of experiments conducted to evaluate the suitability and extent to which personnel recovery trainees and trainers can be supported by I-X in so-called "Command Post Exercises". The result shows that an I-X application can be useful in such a scenario by eliminating some of the basic problems that often occur.**

***Index Terms*—Decision-making, Emergency Response, Search and Rescue, Personnel Recovery, Planning, Collaboration**

## I. Introduction

Personnel Recovery (PR) is the sum of military, diplomatic and civil efforts to affect the recovery and reintegration of isolated personnel. During any military operation Coalition or Joint Force Commanders and Staff are responsible for being prepared to accomplish the PR execution tasks throughout a specified operational area or determine and accept the risk of not doing so [4]. In order to be prepared, the USJFCOM/JPRA Personnel Recovery Education and Training Center (PRETC) trains military personnel in the execution of PR tasks. This training consists of classroom sessions in which the necessary knowledge is taught, and a series of Command Post Exercises (CPX) in which the students have to perform PR tasks in a simulated fictitious military operation called "Operation Able Sword".

One of the aims of the Co-OPR project is to evaluate the possibility of using the I-X framework to create an application that can be used to support the PR task. In this paper we shall briefly describe the I-X framework including its principal user interface, the I-X Process Panel and underlying ontology, <I-N-C-A>. The Co-OPR application that was developed using the I-X framework is based on requirements that were captured during the observation of a CPX, also described in this paper. How these requirements translated into features of the application will be described next. Finally, we shall describe the results of several experiments that have taken place in AIAI's experimental Emergency Response Coordination Center (e-RCC) and at USJFCOM/J9 to evaluate the Co-OPR application for PR.

G. Wickler and A. Tate are with the Artificial Intelligence Applications Institute, University of Edinburgh, Edinburgh, EH8 9LE, Scotland (+44 131 6502732; fax: +44 131 6506513; e-mail: g.wickler/a.tate@ed.ac.uk).

J. Hansberger is with USJFCOM/J9, Norfolk, VA, USA (e-mail: Jeff.Hansberger@je.jfcom.mil).

## II. The I-X Framework

I-X is a framework that can be used to create an application in which multiple agents, be they human or software, adopt a task-centric view of a situation, and which supports the necessary coordination of their activities to respond to that situation. The I-X Process Panel provides the functionality of a to-do list and thus, it is a useful tool when it comes to organizing the response to an emergency. The idea of using a to-do list as a basis for a distributed task manager is not new [5]. However, I-X goes well beyond this metaphor and provides a number of useful extensions that facilitate the finding and adaptation of a complete and efficient course of action.

### *A. The <I-N-C-A> Ontology*

In <I-N-C-A>, both processes and process products are abstractly considered to be made up of a set of *Issues* (I) which are associated with the processes or process products to represent potential requirements, questions raised as a result of analysis or critiquing [1], etc. They also contain *Nodes* (N) (activities in a process, or parts of a physical product) which may have parts called sub-nodes making up a hierarchical description of the process or product. The nodes are related by a set of detailed *Constraints* (C) of various kinds. Finally there can be *Annotations* (A) related to the processes or products, which provide rationale, information and other useful descriptions. For a more detailed description of these four components see [14].

<I-N-C-A> models, which are generic descriptions of synthesis tasks, are intended to support a number of different uses:

- for automatic and mixed-initiative generation and manipulation of plans and other synthesized artifacts and to act as an ontology to underpin such use;
- as a common basis for human and system communication about plans and other synthesized artifacts;
- as a target for principled and reliable acquisition of knowledge about synthesized artifacts such as plans, process models and process product information;
- to support formal reasoning about plans and other synthesized artifacts.

These cover both formal and practical requirements and encompass the requirements for use by both human and computer-based planning and design systems.

### B. *I-X Process Panels*

I-X Process Panels constitute the primary user interface to an I-X application. A panel more or less directly reflects the <I-N-C-A> ontology underlying the whole I-X system.

When used to describe processes, nodes are the activities that need to be performed in a course of action, thus functioning as the items in an intelligent to-do list. The other elements contain issues as questions remaining for a given course of action, information about the constraints involved and the current state of the world, and annotations or notes such as reports or the rationale behind items in the plan. The user interface to the I-X system, the I-X Process Panel, shows four main parts that reflect the four components of the <I-N-C-A> ontology. They are labeled "Issues", "Activities", "State", and "Annotations", as shown in figure 1.

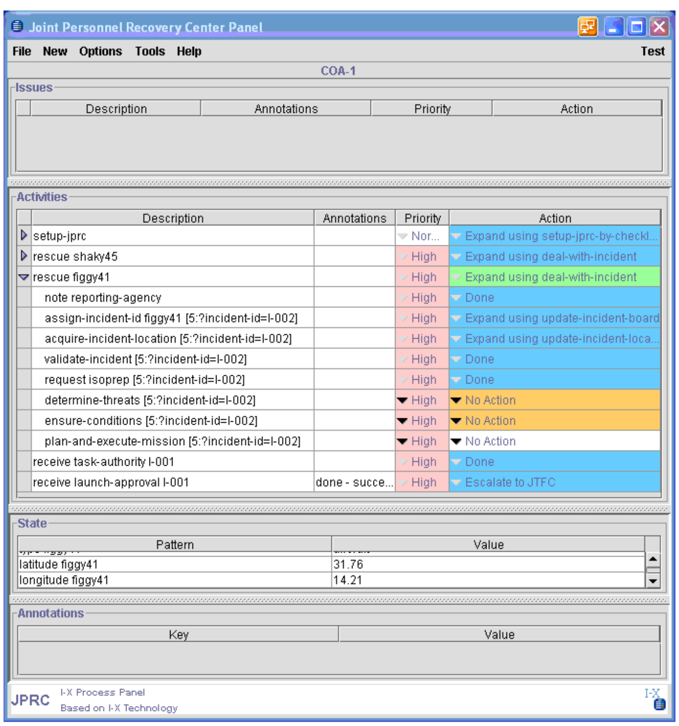


**Figure 1: An I-X Process Panel.**

## III. The Co-OPR Project and Application

The aim of the Co-OPR application was to support trainers and trainees in an emulated half day round of a CPX for a fictitious operation called Operation Able Sword. Such exercises were observed by the project team and researchers in October 2005, and materials were provided to enable research and experimentation.

### A. *Command Post Exercises*

Command Post Exercises are performed at the PRETC as part of the PR course. The course consists of classroom teaching sessions and the CPX in which students are divided into groups, playing the roles of rescue centers that have to respond to some incidents that are emulated by the trainers.

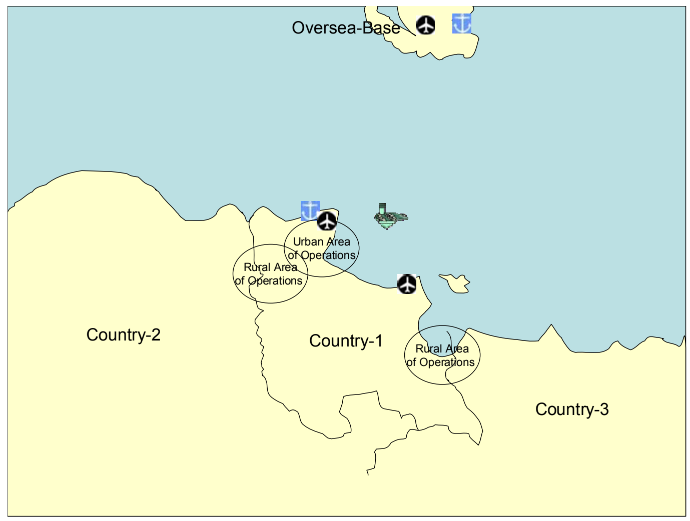


**Figure 2: Generic Scenario Map**

The context for the incidents and rescue missions that need to be launched is Operation Able Sword which nominally takes place in Tunisia on some given dates in June/July 2005. The topology corresponds to the generic map shown in figure 2. In the figure, Country-1 represents the country that is being assisted (Tunisia) and that is in conflict with its immediate neighbors. A shared coastline makes the involvement of the Navy possible. Country-1 also has rural as well as urban areas that make for an interesting variety of potential incidents. Finally, a neutral country provides some oversea base that may play a role.

For a CPX, the students are divided into four groups and placed in different rooms where they act out the activities performed by the different Rescue Component Centers (RCC). In the CPX the Joint Personnel Recovery Center (JPRC) is co-located with the Air Force RCC. All other agents are role-played by the trainers at the PRETC. An overview of the organizational relationships between the different agents is given in figure 3. The first task for the students always consists of setting up the RCCs. Once this is completed the trainers call in incident reports to the different RCCs that have to be collected, analyzed and acted upon, usually by launching an appropriate rescue mission.

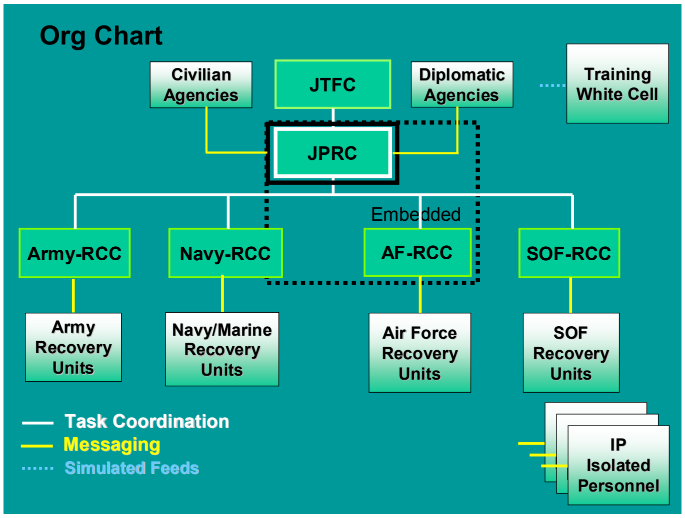


**Figure 3: Organization of Agent in the Scenario**

### B. *I-X for the Co-OPR Application: Requirements*

In observing the Command Post Exercises at the PRETC, we have identified a number of ways in which I-X technology and the user interfaces or tools we can provide may be able to support those involved in search and rescue. I-X uses in a JPRC/RCC could include:

- Communications
  - Simple Chat
  - Structured chat
  - Information sharing
- Task Support
  - Checklists
  - To do list
  - Progress reporting
  - Plan option aids
- Whiteboards
  - Incident
  - Weather/Codes/Info
  - Assets
- Mapboards
  - Terrain and GIS features
  - Routes, ROZs, etc.
  - Town and road plans
  - Sketch maps
- Web Resources
  - Fact Book
  - Phone List
  - Codes
- Mission Folders
  - Attachments

Many of these features are already supported in the I-X framework generically. However, the JPRC and RCCs make heavy use of wall mounted whiteboards, maps, overlays on maps, and pin board material such as codes, phone lists, etc. We have implemented whiteboard and map orientated "viewers" that can all simultaneously share the same state in a single panel for display and sharing. We are now exploring ways in which the state underlying specific views can easily be shared with other users and I-X panels, and ways in which variances between the incoming and current believed state on any panel can be highlighted, such that the changes can initiate issues, activities, constraints or notes that need to be incorporated into the local plan.

We have also created a "white cell" support panel to assist the trainers in a CPX. This will allow:

- Driving a simulation of the world in which the training takes place, including starting and stopping moving assets such as fuel tankers, trucks, planes and ships.
- Setting the world clock as seen by all other I-X panels and users to a simulated time.
- Allowing master scenario event lists (MSELs) to be input and assist in driving the simulation
- Assisting in logging, noting training issues for report back, etc.

All these features are now part of the I-X framework and can be included in any I-X application. The first application to use them is the Co-OPR application described next.

### C. *The Co-OPR Application*

The first step in developing an I-X application consists of deciding which agents to support. For the Co-OPR application it was clear that the most important agent is the JPRC which coordinates the efforts of the different RCCs. Two roles in the JPRC of particular importance are that of the director, who has to manage the centre and make sure everything that needs to be done gets done, and the controller who manages the recovery assets and has to come up with plans for individual recovery missions. Two I-X Process Panels were used to support these two roles. Only the second of these, the one for the controller, had the I-X option management facility enabled which can be used to explore possible courses of action and compare different recovery plans (see figure 4). Other RCCs were supported by a single panel only.

Another agent that plays an important role in the training scenario is the "white cell" that drives the scenarios and simulates the events that lead to the incidents the JPRC has to deal with. An I-X Process Panel was used to support this role by allowing for an additional communication channel with the other agents supported by panels. Finally, some other agents that play only minor roles in the different scenarios were included, e.g. the Joint Task Force Commander (JTFC) that has to give authorization for certain missions. The organization of all the agents in the application is as shown in figure 3.

To implement the task support it was necessary to model a set of standard operating procedures that could be used as refinements in the I-X Process Panel as described above. The refinements used were derived from two sources. Firstly, the U.S. manual for PR [4] was used as a base for knowledge engineering. Secondly, the checklists used by the PRETC during a CPX were imported into I X using a model import facility and manually updated in the I X Domain Editor.

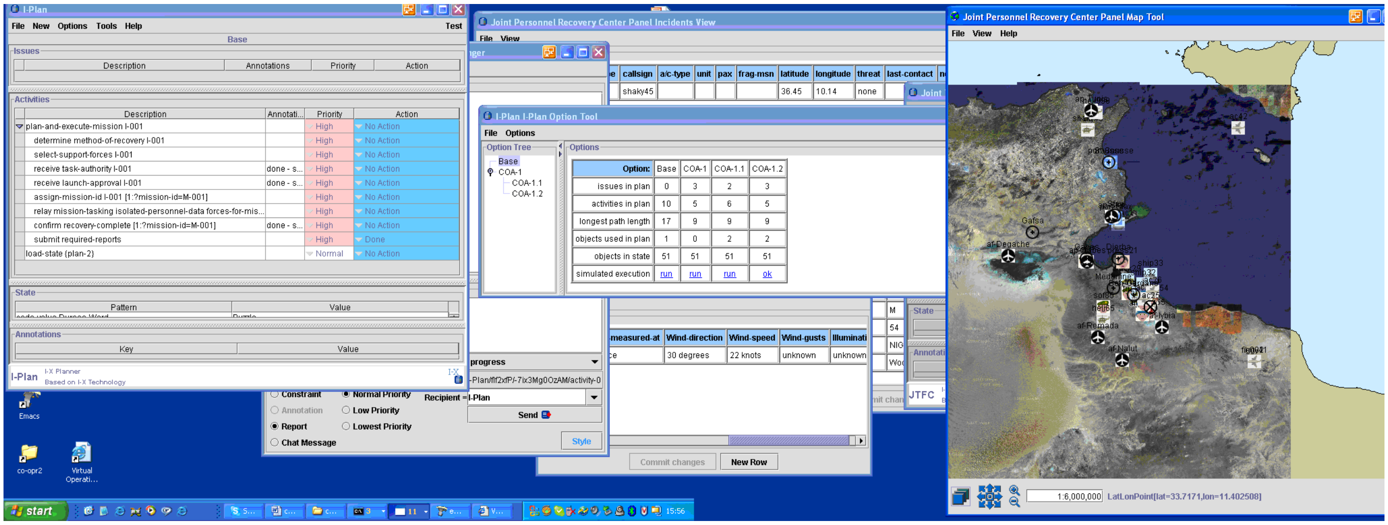


**Figure 4: I-Plan Panel with plan completed; Option Tool with tree structure and comparison matrix; Map Tool showing locations**

## IV. Experiments and Evaluation

The experimentation was designed to demonstrate and stress the value of I-X technology components in response to various individual events in sample incidents and missions provided by the PRETC. Following a number of progressively more realistic trials held in AIAI's experimental Emergency Response Coordination Center, two Co-OPR evaluation experiments were conducted in May and October 2006.

The experiments covered setting up a JPRC which is co-located with an Air Force RCC. Next, incidents of various kinds are dealt with, and a final operation is to prepare a shift change briefing. The aim of the experiment was to allow for an evaluation of the I-X technology as a support tool for both trainers and trainees. At this stage the evaluation was performed by Dr. Hansberger who was remotely observing the experiments from USJFCOM/J9. It is hoped that an evaluation with real users can take place later in the project.

The initial evaluation focused on the cognitive tasks that the JPRC director and JPRC controller performed when working in tandem to respond to the incidents that came into the JPRC as an emergency response coordination centre. This evaluation was necessarily limited in that, without a corresponding analysis of the performance with and use of the current in-situ systems and (manual) processes, a comparative assessment of the influence and worth of the I-X system as a whole is not possible. However, an analysis of the results throws up some interesting insights.

### A. *Evaluation Methodology*

The evaluation methodology was straightforward. The director and the controller roles were played by two members of the I-X development team. In addition to being familiar with the use of I-X systems and with its deployment for this particular domain, these two have gained a basic competence in the objectives, approaches and working practices of the JPRC through observation and completion of basic training courses. An independent observer, a non-participant in the exercise (and also a member of the I-X team), was to observe their behaviour (aided and augmented by self-reporting by the subjects), determine the nature of the task that was currently being performed and the time at which the task began and ended, plus any additional comments or observations. In addition, the exercise was being video-taped, which would allow a retrospective analysis, perhaps with the assistance of the ‘director’ and ‘controller’, of any points during the exercise where the precise nature of the immediate task in hand was not clear. Importantly, the experiment was also observed by a member of the sponsoring organization familiar with PR and with systems evaluation. This was done remotely using Internet collaboration and desktop sharing tools including video teleconferencing.

Once this was done, in an attempt to generalize the various tasks that had been performed where appropriate each task was classified into one of several course-grained ‘cognitive categories’, namely:

*information-gathering*: these tasks involved searching for information that was required before the overall activity of the JPRC could be moved forward. In certain cases, this may involve looking up information in on-line databases, or paper-based manuals, or it may involve, say, (simulated) phone-calls to appropriate colleagues.

*sense-making*: these tasks involved an analysis and interpretation of information with the aim of understanding the problem, enumerating the different options that were available, listing the pros and cons of possible courses of action, and so on.

*decision-making*: these tasks involved the subject making a clear choice from among competing possible activities *that would serve to achieve the objectives of the JPRC by effecting activity in other agents* and then enacting this activity. So, for example, deciding to send a rescue helicopter to a particular destination and issuing the appropriate orders would be an example of a decision point, whereas deciding to look at a map would not, since it has no affect on other agents (and, instead, would probably be an instance of information-gathering).

*housekeeping*: these tasks involved the initial set-up of the JPRC environment, documentation of decisions, logging of calls, etc.

The first three of these categories (the housekeeping category being an artifact arising from the need to manage the JPRC and the 'paperwork' it generates) emerge from consideration of several different 'best practice' approaches to command and control and decision-making in general. For instance, Boyd's well-known OODA loop [8]–Observe, Orient, Decide, Act–can be seen to correspond with these three tasks, observe, orient, and decide. Observation is essentially synonymous in this context with information-gathering and orient is synonymous with sense-making. Since most of the decisions taken by the JPRC staff are done by issuing commands to others (i.e., in I-X terms, sending an activity to another agent) and this is done on the click of a mouse button, we do not attempt to differentiate the decide and act activities for our analysis. We instead conflate these two OODA tasks into the single decision-making category. Similarly, Wohl's SHORe (Stimulus, Hypothesis, Option, Response) framework [16] can be seen as analogous to our categories, with stimulus (Wohl's shorthand term for the information correlation and fusion phase) corresponding to information-gathering, hypothesis (Wohl's situation analysis phase) corresponding to sense-making, and the option and response phases being conflated into the single decision-making task (and for the same reason outlined above).

| Phase | OODA | SHORe | "JPRC Experiment C" Analysis |
|---|---|---|---|
| 1 | *observe* | *stimulus* | *information-gathering* |
| 2 | *orient* | *hypothesis* | *sense-making* |
| 3 | *decide* | *option* | *decision-making* |
| 4 | *act* | *response* | |

**Table 1. Comparison of different Command-and-Control frameworks as they apply in this context; only part of the *act* (OODA) and *response* (SHORe) activities occurs within the context of the JPRC.**

The correspondence between these different models is summarized in Table 1. The fundamental concept underlying all of these models is that a methodical approach to each cycle of the command and control 'loop', based on assembling information, interpreting that information, appraising possible courses of action and making and enacting decisions should lead to clear, consistent, and – ultimately – correct behaviour in situations where the pressure is great and time is short. Our empirical hypothesis here is that the use of the I-X system can encourage its users to adopt such a methodical approach to their task.

### *B. Evaluation Results*

A fragment of the task analysis performed on the activities observed during Experiment C can be seen in figure 5.

| | | | | |
|---|---|---|---|---|
| 37 | 0:51:15 | | decision-making | request launch authority |
| 38 | 0:53:00 | JD | decision-making | mission establish |
| 39 | 0:55:00 | JD | decision-making | send plan to whitecell |
| 40 | 0:55:45 | | | [wait for confirmation from whitecell] |
| 41 | 0:57:02 | JD | housekeeping | updates state assuming mission complete |
| 42 | 1:00:00 | GW | housekeeping | rescue figgy, note reporting agency, establ |
| 43 | 1:01:00 | GW | information-gathering | request info - incident location |
| 44 | 1:02:13 | GW | information-gathering | receive location |
| 45 | 1:02:36 | GW | information-gathering | validate incident (WoD) |
| 46 | 1:03:30 | GW | information-gathering | awaiting message |
| 47 | 1:04:00 | GW | housekeeping | broadcast WoD update to all subordinates |
| 48 | 1:04:46 | GW | information-gathering | ISOPREP - VOC |
| 49 | 1:06:58 | GW | sense-making | identify threats - I-X mesgs |
| 50 | 1:09:47 | GW | sense-making | ensure conditions, ensure OSC |
| 51 | 1:11:10 | | | [interrupt] duress word |
| 52 | 1:13:00 | GW | sense-making | reviews plan (threats) so far |

**Figure 5: Fragment of Co-OPR task analysis.**

Notwithstanding the provisos noted above about the inability at the time of writing to perform a full comparative evaluation, the analysis is encouraging for the use of the I-X in this task. In general, the use of SOPs encouraged a methodical approach to the overall JPRC activity: instances of information-gathering where followed by instances of sense-making which led to decision-making episodes, with no instances of, for instance, a decision-making activity being interrupted or abandoned due to the lack of a crucial piece of information. In addition, at several times during the exercise, important messages arrived which interrupted the current activity and diverted the cognitive attention of the director or controller. Such interruptions can serve to disrupt the flow of the Center, but in the majority of cases, the framework provided by the SOPs allowed a quick resumption of activity once the message had been dealt with.

In addition, the analysis highlighted some areas where further support might prove helpful. In addition to dealing with interruptions, the arrival of new information that demands that the decisions made earlier in the process need to be re-appraised (and, in one case during the experiment, wholly abandoned, with rescue resources 'recalled') is currently not difficult to handle using within the SOP framework (and would seem to require something akin to 'exception-handling' procedures). Successfully dealing with such situations seems to rely heavily on the experience and initiative of the human in question. This would seem to be a general problem with any SOP-based system rather than with I-X per se, but technology that can offer more support would obviously be of great benefit.

Consideration of the time devoted during the experiment to each of the task categories is also interesting. While roughly the same amount of time was spent in information-gathering, sense-making and decision-making during the exercise, a surprisingly large amount of time was spent housekeeping – twice as long, in fact, as the time spent for any of the other categories. This is due, in part, to the time required to initialize the JPRC and check that its procedures and communications are in place, and then later to produce a report summarizing the session activities for the next duty officer. Providing automated assistance for these tasks may reduce the workload of the humans involved while also ensuring a more rapid and efficient establishment of the Center and hand-over of duty.

Aside from an analysis of the cognitive tasks performed by the system users, the experimentation also highlighted a number of open issues with the current prototype. Firstly, support for the white cell was rather limited at this stage. Only the structured messaging feature was a real advantage provided by I-X. However, the way the scenario was driven was adapted to this way of delegating tasks, which does not correspond well to the way the real CPX works. This in effect removes a large part of the sense-making task from the problem and shifts the focus onto the planning activities, an area in which I-X is strong. Secondly, the two panels used by the director and the controller are equipped with independent <I-N-C-A> models which may lead to inconsistent world state representations within the JPRC. While this did not occur during the experiment, it is a potential problem that was noted. Finally, some issues with the user interface need to be addressed for future versions, e.g. the lack of a mechanism to draw the user's attention immediately to new, incoming activities.

## V. Conclusions

The aim of the Co-OPR experimentation was to emulate a half-day round of a CPX usually held at the U.S. Personnel Recovery Education and Training Center. Materials were provided by DARPA and USJFCOM for a search and rescue element of a military mission. The experiments were designed to demonstrate and stress the I-X technology components in response to various individual events in sample training missions and events provided by experience trainers in the US Joint Personnel Recovery Agency.

Initial evaluation indicates that I-X can indeed be used to build applications that support task-centric activities in this domain, and that two main features supported by I-X, namely intelligence through integrated standard operation procedures, and coordination support through linked process panels, are useful in supporting the overall activity of a Coalition or Joint Personnel Recovery Center. More specifically, an analysis of the experiment shows that the hierarchical structure of the tasks in the to-do list helps users to focus their efforts and avoid distractions, and if interrupted, it helps them to quickly continue with important decision making without having to repeat information-gathering or sense-making activities that have already been completed. It can also help in handover between personnel when staffs change as it presents a clear status picture of tasks and actions that have been accomplished in the past, on-going current activities, and future tasks needing completion.

## Acknowledgment

The I-X project is sponsored by the Defense Advanced Research Projects Agency (DARPA) under agreement number F30602-03-2-0014. The University of Edinburgh, the U.S. government and research sponsors are authorized to reproduce and distribute reprints and on-line copies for their purposes notwithstanding any copyright annotation hereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of other parties.

## References

[1] Buckingham Shum, S., Selvin, A., Sierhuis, M., Conklin, J., Haley, C. and Nuseibeh, B. (2006). Hypermedia Support for Argumentation-Based Rationale: 15 Years on from gIBIS and QOC. In: Rationale Management in Software Engineering (Eds.) A.H. Dutoit, R. McCall, I. Mistrik, and B. Paech. Springer-Verlag: Berlin

[2] Conklin, J. (2003) Dialog Mapping: Reflections on an Industrial Strength Case Study. In: P.A. Kirschner, et al. (eds.) Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making. Springer-Verlag: London, pp. 117-136.

[3] Horling, B., Lesser, V., Vincent, R. and Wagner, T. (2006) The Soft Real-Time Agent Control Architecture. *Autonomous Agents and Multi-Agent Systems*, Volume 12, Number 1, Springer Science+Business Media , pp. 35-92.

[4] Joint Publication 3-50 (2005) Joint Doctrine for Personnel Recovery, 2nd Draft, March 2005.

[5] Kreifelts, Th., Hinrichs, E., and Woetzel, G. (1993) Sharing To-Do Lists with a Distributed Task Manager. In: de Michelis, G. and Simone, C. (eds.) Proceedings of the 3rd European Conference on Computer Supported Cooperative Work, pp 31-46, Milano, 13-17 September 1993, Kluwer, Dordrecht.

[6] MacLean, A., Young, R., Bellotti, V. and Moran, T. (1991) Design space analysis: Bridging from theory to practice via design rationale. In Proceedings of Esprit '91, Brussels, November 1991, pp 720-730.

[7] Nau, D., Cao, Y., Lotem, A., Muñoz-Avila, H. (1999) SHOP: A Simple Hierarchical Ordered Planner. Proc IJCAI 1999, pp. 968-973.

[8] Osinga, F. (2006) Science Strategy and War, The Strategic Theory of John Boyd, Abingdon, UK: Routledge, ISBN 0-415-37103-1.

[9] Polyak, S. and Tate, A. (1998) Rationale in Planning: Causality, Dependencies and Decisions. Knowledge Engineering Review, Vol.13(3), pp 247-262.

[10] Selvin, A.M. (1999) Supporting Collaborative Analysis and Design with Hypertext Functionality, Journal of Digital Information, Vol. 1 Issue 4.

[11] Selvin, A.M., Buckingham Shum, S.J., Sierhuis, M., Conklin, J., Zimmermann, B., Palus, C., Drath, W., Horth, D., Domingue, J., Motta, E. and Li, G. (2001) Compendium: Making Meetings into Knowledge Events. Knowledge Technologies 2001, Austin TX, USA, pp 4-7.

[12] Tate, A. (1995) Integrating Constraint Management into an AI Planner. Journal of Artificial Intelligence in Engineering, Vol. 9, pp 221-228.

[13] Tate, A., Dalton, J. and Stader, J. (2002) I-P2- Intelligent Process Panels to Support Coalition Operations. In Proceedings of the Second International Conference on Knowledge Systems for Coalition Operations (KSCO-2002). Toulouse, France, April 2002.

[14] Tate, A. (2003) <I-N-C-A>: An Ontology for Mixed-initiative Synthesis Tasks. Proceedings of the Workshop on Mixed-Initiative Intelligent Systems (MIIS) at the International Joint Conference on Artificial Intelligence (IJCAI-03), Acapulco, Mexico, August 2003, pp 125-130.

[15] Wickler, G., Tate, A. and Potter, S. (2006) Using the <I-N-C-A> Constraint Model as a Shared Representation of Intentions for Emergency Response, Proc. 1st Int. Workshop on Agent Technology for Disaster Management (ATDM), at the 5th Int. Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2006), Hakodate, Japan, May 8-12, 2006.

[16] Wohl, J.G. (1981) Force Management Decision Requirements for Air Force Tactical Command and Control. IEEE Transactions on Systems, Man, and Cybernetics, 1981, 11.

# Autonomous Protection Mechanism for Joint Networks in Coalition Operations

Martin Rehák, Jan Tožička, Michal Pěchouček, Magdalena Prokopová and Lukáš Foltýn
Department of Cybernetics, Czech Technical University in Prague
Technická 2, Prague 6, 166 27 Czech Republic
email:{mrehak,pechouc,prokopova,lfoltyn,tozicka}@labe.felk.cvut.cz

***Abstract*—Any successful coalition cooperation requires efficient communication network connecting the coalition members. Protection of this joint network requires special techniques as it is highly dynamic, heterogenous and a joint network management team can not always be established. To address the requirements for joint network protection, we propose a design of a highly autonomous, adaptive and decentralized agent-based mechanism for network intrusion detection and self-protection. Detection process is based on correlation of anomalies in network traffic with simple alarms raised by host-based intrusion detection components, in order to achieve a low false positive rate. The self-protection mechanism features distributed, policy driven deployment of automatically generated filters. Our approach doesn't require any direct operator oversight, but all components are subject to policies established by their owners to prevent undesirable behavior or system misuse. The whole approach is validated in a high-level network model with worm propagation scenario.**[1]

## I. Introduction

This work addresses a pressing need related to integration of information networks during coalition operations. While the ad-hoc network interconnections across the filed can significantly increase the efficiency of operations by allowing the partners to exploit the data from local sensors and assets, the complexity of network management increases significantly. This work addresses a specific part of the network management – collaborative **network intrusion detection and response** (intrusion prevention). In order to protect the network against the attacks[6], we propose to deploy an automatic, adaptive, collaborative mechanism that detects the attacks and proposes and appropriate response. Such response allows to keep the network operational, without requiring direct human involvement in the process. Each of the coalition members can control its devices and Intrusion Detection Agents (IDS agents) by means methods of adjustable autonomy [1].

The requirements on the presented network protection mechanism are significantly different from the classic network protection due to the different nature of the coalition network: (*i*) There is **no** trivially identifiable **perimeter** between the hosts and devices of coalition members, as most network connections between the hosts of the same organization can pass through the network segment managed by another member. (*ii*) The network must be **set-up rapidly**, with minimal overhead and human intervention. The same requirement applies to reconfiguration, as the network needs to cope with **frequent changes** – new coalition members and/or assets connecting in real-time, disconnected or destroyed assets or communication inaccessibility. (*iii*) The network is likely to be a highly **heterogenous** composition of high bandwidth fixed links in the headquarters of the individual members, fixed and higher latency links between the member's headquarters and a mesh network in the field where the assets of all (or some of) the members can form a joint ad-hoc wireless network. (*iv*) Each coalition member must be able to define, adapt and enforce its **policies** regarding network and system security, without explicitly disclosing some of these policies to other coalition members. (*v*) Oversight of the network by a dedicated joint **CIRT** (Computer Incident Response Team) can not be assumed due to the operational limitations.

Therefore, the IPS system shall be autonomic [15]: it shall require no or minimal initial configuration, shall be able to observe the network traffic and associated reaction of the hosts to identify the potentially malicious flows and to autonomously create and deploy filters on network devices to contain the threat, without disturbance of legitimate flows. Furthermore, it shall impose no limitations on network topology and shall autonomously adapt to network changes. It shall also require only minimal real-time supervision by human operators, especially in its ad-hoc part where we assume the deployment in field. On the other hand, operators must be able to control the system behavior, to analyze and confirm its decision, and to change the policies and priorities.

In the next Section, we will present a tentative architecture of the system, before describing core system components in dedicated sections. Our approach to the problem is based on established multi-agent techniques: trust modelling (in Sec. III), policy-based dynamic service composition (Sec. II) and negotiation with distributed task allocation via peer-to-peer auctions in Section IV.

## II. Architecture

In order to satisfy the requirements stated above, the system shall be decentralized, autonomous and shall feature an efficient and effective learning component. Our architecture, as shown in Fig. 1 consists of three functionalities, each of

[1]**Acknowledgment:** We gratefully acknowledge the support of the presented research by Army Research Laboratory project N62558-05-C-0028.The first author would also like to acknowledge the support of the National Institute of Informatics in Tokyo during his inspiring visit.

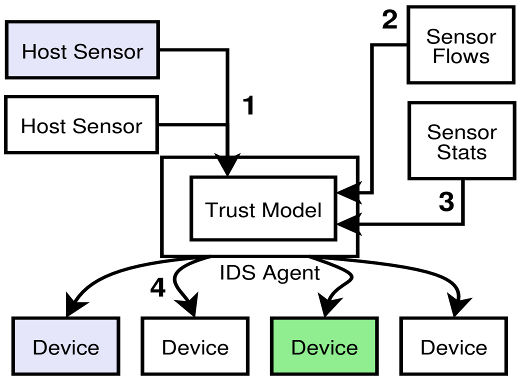


Fig. 1. Architecture of the IPS system. Colors of components represent ownership. Only a single IDS agent is represented for clarity.

which implements one of the protection phases from sensor to actual defensive action:

• **Observation** components are of two basic types - **Host sensors** are deployed on selected hosts in the network (protected hosts) and raise alarms (1 in Fig. 1) when an exceptional behavior suggests a possibility of attack or intrusion attempt. **Network sensors** are located on network devices and capture the relevant information about the network flows (2 in Fig. 1), together with the associated statistics (3 in Fig. 1). The information we observe is outlined in Table I.

• **Detection** components, denoted as **IDS agents** processes the information from the Observation components using its trust model [13] in order to correctly correlate the alarms received from hosts with the current network flows. Its role is to correctly identify the malicious traffic and to generate a description of the malicious flow(s) for filter creation in reaction phase.

• **Reaction** functionality is based on a cooperation of IDS agents with autonomous, adaptive network devices. IDS agents generate the filter(s) (4 in Fig. 1) describing the attack and pass these filters onto the network devices. These devices are reflective [8], able to autonomously adapt to changing environment. In our specific case, they would incorporate the traffic filtering capacity, either locally or in a collaboration with other devices, while respecting their predefined policies.

When deployed on the network, all system components identify each other and select the best data and service providers. This selection is neither final nor exclusive – sensors can provide the information to more than one IDS agent (1,2,3 in Fig. 1), IDS agents shall gather their information from at least several sensors and communicate between them using a reputation mechanism. Similarly, the reflective network devices that actually react to attacks can dynamically select between the software filters and associated threat evaluations provided by several IDS agents (4 in Fig. 1). All these interactions are dynamic, but regulated by strict policies [1] to prevent the misuse of network protection mechanism, and to regulate the cooperation with Sensors and IDS agents belonging to other coalition members.

When the joint network is deployed, we assume that the majority of hosts is equipped with host sensors that can range from slightly enhanced personal firewalls [7] with event reporting, through log file processors to dedicated Intrusion detection applications[2]. Observation and detection phase of our architecture is based on assumption that the hosts in the coalition network are highly heterogenous, with different OS, application versions, security policies and local protection. Therefore, in order to successfully compromise the operation of the network, the attacker must either perform a detailed fingerprint of the network, with a very high risk of being discovered before an actual attack, or attack using indiscriminate, epidemics-style methods of attack like Distributed Denial of Service (DDoS) or Worms[6]. As the first type of threat is being sufficiently addressed by current methods [7], this paper will concentrate on protection against the massive, indiscriminate attacks only. In such cases, the heterogeneity of the network, typically considered as a security risk, becomes the strength – an attack is likely to fail on some hosts, being subsequently detected by their Host sensors and the alarm is passed to the IDS agents. In our system, we don't require any supplementary information regarding the origin of the alarm to be associated – we only use the alarms as a stimulus for the dedicated IDS agents. This choice significantly facilitates the integration of the system (and limits misuse), as besides the established multi-agent techniques, sensor agents can either use any method ranging from SNMP trap to invoking HTTP connection to pass the information.

In the same time, Network-based sensors (typically based on routers/switches) observe and sample the traffic, and pass the statistics and samples to IDS agents presented in the next section. IDS agents use enhanced trust modeling to identify the malicious flows and to create their efficient description in a form of a filter. This filter is then distributed to network components for allocation by decentralized allocation process (Sec. IV). The goal of this process is to protect the path between any two hosts on the network by each relevant filter. Perimeter or zone protection would be irrelevant as there is no well defined perimeter to protect, and also because by its nature, the system is likely to detect the threat once it has spread over at least some part of the system.

## III. Detection Mechanism

Based on our previous work, the IDS agents use an extended *trust model* [11], [10], [12] to correlate the alarm received from the hosts with the traffic and to identify the malicious flows. In general, trust models [13] are used to identify malicious elements by observing the system behavior and to infer conclusions about the actors. In our system, IDS agents evaluate the trustfulness of *network flows*, unidirectional series of TCP (or UDP/ICMP) network connections [7].

The trust model inside each IDS agent processes several inputs provided by sensors: (*i*) information about the *network flows* in the network that contains the TCP/IP headers and

[2] www.tripwire.org

| Feature | Description |
|---|---|
| ***Connection Identity*** | |
| srcIP | Source IP Address. |
| destIP | Destination IP |
| srcPort | Source Port |
| destPort | Destination Port |
| Protocol | Protocol (TCP/UDP/ICMP) |
| Payload Signature | First 256 bytes of the flow content (application headers) |
| ***Connection Context*** | |
| count-dest | Number of flows to unique destinations from the same source. |
| count-src | Number of flows from the unique sources toward the same destination. |
| count-serv-src | Number of flows from the same IP to the same port. |
| count-serv-dest | Number of flows to the same destination IP using the same source port. |

TABLE I
CHARACTERISTICS OF THE FLOW, SING THE NETFLOW-LIKE IDENTITY FORMAT AND CONTEXT (ADAPTED FROM [3]).

**Algorithm 1**: Processing of the new observation.

**Input**: *flow,situat,trustObs*
*closest* $\leftarrow$ *nil*;
*mindist* $\leftarrow \infty$;
*id* $\leftarrow$ identityCx(*flow*,*situat*)
**foreach** *rc* $\in$ *rclist* **do**
  *dist* $\leftarrow$ distance(*rc*,*id*)
  **if** *dist* $<$ *mindist* **then**
    *closest* $\leftarrow$ *rc*
  *wei* = weight(*dist*)
  **if** *wei* $>$ *threshold* **then**
    *rc*.updateTrust(*trustObs*,*wei*)
**end**
**if** *mindist* $>$ *clustsize* **then**
  rclist.append(id)
  *id*.updateTrust(*trustObs*,*wei*)
**else**
  *closest*.updatePosition(id)

first 256 bytes of the flow content (typically containing an application header [4]), (*ii*) *statistics* of the existing flows in the network, with the parameters identified by the MINDS team [3], listed in the context part of Table I, and (*iii*) the *alarms* raised by Host sensors as detailed in Sec. II.

The trust model correlates the occurrence of alarms received from the hosts with the features of the relevant captured flows and the statistics associated with a particular flow. The features and the statistics form an *Identity-Context* feature space [12]. As the number of flows in an average network can be significant, it is important to process, represent, and maintain the data efficiently, and to keep the model synchronized with the most recent alerts and traffic. Therefore, instead of associating the information with individual flows in the *Identity-Context* space, the flows are aggregated into clusters. For each cluster, the IDS maintain its trustfulness, expressed as a fuzzy number [11]. The learning of trustfulness in our model is iterative, and each iteration consists of two stages. In the first stage, IDS agents generate and update the clusters using the Leader-Follower[3] algorithm [2], while in the second stage IDS agents update the trustworthiness that is associated with the centroids of the clusters that are adjacent to the observation representation in the Identity-Context space. For sake of efficiency, our implementation actually performs both stages in the same time, as we can see in Alg. 1.

When IDS agent **observes** a flow, it extracts the identity features (see Tab. I), then retrieves the associated statistics to determine the position of this vector in the Identity-Context space (*id* in Alg. 1) . Then, it computes the distance of the observed flow vector to each of the existing centroids (*rc* $\in$ *rclist*), and update the trustfulness of these centroids with a weight *wei* that decreases with growing distance. If the closest centroid is farther away than a predetermined threshold *clustsize*, a new cluster is defined around the new observation. When the model is **queried** regarding the trustfulness of a specific flow vector, it aggregates the trustfulness associated with the centroids, with a weight that decreases with the distance from the centroid to the flow representation in the Identity-Context space. Once it has has determined the trustfulness, in a form of a quasi-triangular fuzzy number, it calculates inferences with a high trust and low trust fuzzy intervals. These intervals represent the knowledge about the normal level of trustfulness in the environment, and allow the IDS agent to infer whether the flow is trusted or not under current conditions [11]. The detection model assumes:

• **Host population heterogeneity**: As the IDS agents rely on feedbacks from the attack-resistant, protected hosts to raise an alarm upon unsuccessful intrusion attempt, we have to assume that the successful intrusion attempts will not raise an alarm immediately. Therefore, a highly heterogeneous population of systems and applications ensures optimal detection.

• **Immediate attack manifestation**: The current version of the model correlates only the recent traffic with the actual alarms received from the hosts. If the alarms are delayed (e.g. by delaying visible actions of the malicious code), or a significant proportion of alarms is false, the model no longer works. Therefore, while we insist on the fact that collecting the alarms from as many sources as possible is desirable, the mechanism shall treat them as reputation sources and can assess their trustfulness as well [5].

## IV. REACTION MECHANISM

This section describes the deployment of a traffic filter in the network, assuming tree network topology. In this section we present the algorithm and we prove its *correctness* (i.e. it protects as many vulnerable host as possible) and *optimality* (i.e. it uses minimal amount of resources) in the situations where a single filter is deployed. We also explain and illustrate that this algorithm is not correct for a sequence of filters. This fact will be demonstrated on a specific example. We present an extension of this algorithm so that the agents are allowed to outsource the process of traffic filtering to the other agents.

[3]Our current implementation uses LF clustering as it is efficient in terms of computational and memory cost, and allows on-line processing. Any other clustering algorithm can also be used instead, the reference points can even be placed arbitrarily [10].

### A. Filter Deployment Algorithm – FDA

The algorithm is initiated by the IDS agent by creating new filter and sending a request for filtering to all agents on neighboring network devices. These agents further propagate the filter through the network by sending the same requests to their neighbors. The requested agent replies whether the traffic through this network node is already filtered by this filter (and thus safe) or not. According to the answers decision is made whether the filter needs to be deployed on the requestor's network node. To formalize the algorithm we introduce following replies to the filter requests:

• `Safe`: the traffic is **safe** according to this filter, i.e. the traffic is filtered or does not require filtering,

• `Opt`: the deployment of filter by requested agent is **optional**; it can perform filtering of the traffic, but for the sake of optimality it might be done by some other agent; it's up to the requestor to choose, and

• `Needed`: it is **necessary** to filter the traffic for this connection (this reply is always sent by a vulnerable host).

Algorithm description uses following notation:

• $\{\mathtt{Opt}\}$: set of requested agents that replied with specified type of answer (same for `Safe` and `Needed`).

•$\{\mathtt{Opt}-\mathtt{1}\}$: $\{\mathtt{Opt}\}$ without the $\mathtt{Opt}^0$ agent

• $|\mathtt{Opt}|$: number of answers of given type

Agent that receives a request from requestor handles it using the FDA algorithm (Algorithm 2). It waits for the replies of subsequent agents based on them it and available resources it decides how to react for the request. This reaction consists of three separate decisions:

• **Filter deployment:** whether to deploy the filter locally

• **Reply to $\{\mathtt{Opt}\}$:** what to react from $\{\mathtt{Opt}\}$ agents

• **Reply to requestor:** how to reply the original requestor (except the IDS agent that initiated the filter deployment).

**Algorithm 2**: FDA Algorithm

**Send** request to all neighbors without the requestor.
**Wait** for all replies.
**Process** replies:
**if** $|\mathtt{Needed}| + |\mathtt{Opt}| = 0$ **then**
  **Filter deployment:** No.
  **Reply to $\{\mathtt{Opt}\}$:** N/A
  **Reply to requestor:** `Safe`.
**else if** *having enough resources to deploy filter* **then**
**(i)** **if** $|\mathtt{Needed}| + |\mathtt{Opt}| > 1$ **then**
  **Filter deployment:** Yes.
  **Reply to $\{\mathtt{Opt}\}$:** *Do-not-filter*.
  **Reply to requestor:** `Safe`.
**(ii)** **else** /* $|\mathtt{Needed}| + |\mathtt{Opt}| = 1$ */
  **Filter deployment:** Yes, if the reply for `Opt` is *Do-filter*. No, otherwise.
  **Reply to $\{\mathtt{Opt}\}$:** *Do-not-filter*.
  **Reply to requestor:** `Opt`.
**else if** *not having needed resources* **then**
**(iii)** **if** $|\mathtt{Needed}| > 1$ **then**
  Perfect deployment does not exist
  (to find near-to-perfect solution: $|\mathtt{Needed}| = 1$)
**(iv)** **else if** $|\mathtt{Needed}| = 1$ **then**
  **Filter deployment:** No resources.
  **Reply to $\{\mathtt{Opt}\}$:** *Do-filter*.
  **Reply to requestor:** `Needed`.
**(v)** **else** /* $|\mathtt{Opt}| > 0$ & $|\mathtt{Needed}| = 0$ */
  **Filter deployment:** No resources.
  **Reply to $\{\mathtt{Opt}\}$:** Send $\{\mathtt{Opt}-1\}$ *Do-filter*.
  **Reply to requestor:** `Opt`. Forward the received reply to $\mathtt{Opt}^0$ agent.

***FDA is Correct:*** Algorithm FDA filters the communication between each pair of vulnerable leaves.
*The proof:* It is obvious that path between every pair $\langle A, B\rangle$ of nodes will be covered by filter in their common ancestor $R$ or sooner if there is enough resources. Let us suppose that the ancestor $R$ has not enough resources and that the path between $A$ and $B$ is not covered. Then there has to exist a node $F$ between the nodes $R$ and $A$ (or $R$ and $B$ analogously) that can filter (let us suppose that $F$ is such a node that is the closest to $R$). The reply from $F$ to its ancestor is `Safe` (condition *(i)*), and the path is covered), or `Opt` *(ii)*. The `Opt` message going to $R$ cannot be changed to `Needed` without the filter deployment, therefore $R$ receives at least one `Opt` reply which will be replied by *Do-filter (iv), (v)*. The *Do-filter* reply will spur filter deployment on $F$ and the path between $A$ and $B$ will be covered. □

***FDA is Optimal:*** Optimality of FDA is defined by using minimal number of deployed filters.
*The proof:* Filter is deployed in nodes only if it has enough resources and at least two nodes require it (condition *(i)*). If such a node does not use filter, then at least $\{\mathtt{Needed}\}+\{\mathtt{Opt}\}-1$ of subsequent nodes would have to filter, or otherwise there would exist an uncovered path between two of its children. Since $\{\mathtt{Needed}\} + \{\mathtt{Opt}\} - 1 \geq 1$ the number of filter would not be lower. If the node has not enough resources the situation is similar. Each node that does not filter sends *Do-filter* to $\{\mathtt{Needed}\} + \{\mathtt{Opt}\} - 1$ of his children as you can see in the condition *(iv)*. Slightly more complicated situation appears when the agent receives *Do-filter* in the condition *(v)*. It can happen only if his ancestor has not enough resources and it used the condition *(iv)* or *(v)*. This situation can repeat till the root of the tree where the requestor selects only $\{\mathtt{Opt}\} - 1$ of its children (condition *(iii)*). Therefore the condition *(v)* does not affect the optimality neither. □

### B. FDA example

The properties listed and proved above do not hold for the situation where a sequence of more than one filter is to be deployed. Our argument is based on the following counterexample demonstrating that optimality of the filters' allocation depends on the location of the IDS agent. The Figure 2 shows a situation when the algorithm provides locally optimal deployment of each of the three filters (1, 2, and 3) but longer term filter distribution is not globally optimal. The deployment of the filter 1 in the upper graph has caused the necessity to deploy three instances of the filter of a type 2.

### C. Extended Filter Deployment Algorithm EFDA

In practical applications, the requests for filters' deployment may exceed the hardware resources of the specific node. We propose to solve such a situations by extending the FDA algorithm in order to allow delegation of the filtering process among the nodes. The filter delegation is based on negotiation

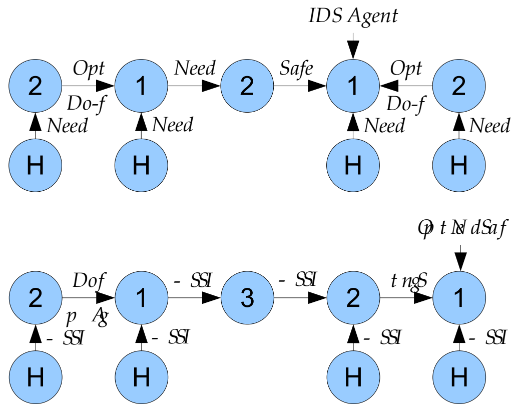


Fig. 2. FDA optimally deploys each of the filters but the result depends on the position of IDS agent (i.e. where the algorithm starts). Labels illustrate the replies for requests.

among two or more nodes. As a result of the negotiation one node would be dedicated to provide filtering services to the other nodes. Such extension can reduces the number of deployed filters, but increases the communication traffic (as the flows must be tunneled to filtering devices). FDA algorithm is modified in such a manner that instead of local filter deployment, the agent uses Contract Net Protocol (CNP) to ask trusted agents (determined by policy or a trust model) to perform filtering on its behalf. Requested agents reply with the proposals where the price for the service is specified. The price includes the distance of the agents (this distance corresponds to the growth of the traffic) and computational costs. Requesting agent then chooses the best proposal to accept.

Similarly to above, optimality can be achieved only for a single filter deployment, while for the sequence of locally optimally deployed filters, global optimality is not guaranteed. We suggest using specific auctioning techniques such as ECNP (Extended CNP) or PAP (Provisional Agreement Protocol) [9] that allows backtracking in the service composition space using *provisional accepts* and *provisional rejects*. By this the delegating agent can to find the optimal solution, on the cost of substantially higher communication traffic.

## V. Simulation and Experiments

The above-presented mechanism is evaluated in a very high-level, scripted model of the worm attack following [6], where we simulate the network only as a set of hosts with assigned IP addresses that can freely communicate with each other through an abstract network component with IDS and filtering functionality.

Our simulation is modeled after the propagation strategies of real worms, as investigated in [6] – with a disabled protection, we actually obtain very similar results regarding the dynamics of the worm spread. Our simulated network consists of two C-type IP4 subnets, with roughly 125 hosts each, assigned randomly over the IP range if the subnet, and two smaller C-type subnets, with about 12 and 6 nodes. The scanning strategy of the worm is simple: one half of the explored addresses is selected within the same C subnet, while the other half is selected randomly over the whole IP address space. In the experiments, we assume that approximately one half of the hosts is vulnerable to the threat and can serve for further worm spread once infected. All other hosts raise an alarm upon infection attempt.

In our experiments, we evaluate the ability of the system to detect the threat as soon as possible and to filter the malicious flows in attempt to prohibit the propagation of the threat. In this model, we assume immediate filter deployment, and filter efficiency identical to the trust model of the IDS agent upon which it is based. Two principal questions are addressed: The first question is whether our learning approach performs fast enough to counter the threat before it affects the whole network. The other question is closely related to the deployment in dynamic, multi-owner networks; how will the system perform with significant fraction of false alarms from host-based sensors. Such alarms can appear either due to the misconfigurations, human and system errors and other "natural" causes, or can be a result of a subversive action of an adversarial coalition member.

In Figure 3, we can see the trustfulness and distrustfulness of worm flows with three different levels of background noise. The blue (dotted) curves correspond to the lowest level of background noise, (background alarm level is uniformly distributed between $0$ and $1/32$ on the $[0, 1]$ scale). The second case (magenta, dash-dotted lines) corresponds to the scenario with the noise in $[0, 1/16]$ interval, while the solid red lines correspond to noise up to $1/4$ of the scale. Alarm levels over the time can be seen as three rising curves in the lower part of the graph. The curves in the upper part of the graph represent the degree of trustfulness (round points) and distrustfulness (crosses) of the worm traffic as estimated by the IDS agent's trust model. Once the distrustfulness of any flow is higher than trustfulness, the flow is considered as malicious and removed. We can see that the dotted/blue lines corresponding to the lowest level of noise intersect around $t = 23$, soon after the worm was introduced in step 20. The models with higher levels of noise (solid lines) take more time to take the decision, the most noisy one reaching the conclusion around $5$ steps later. This results in an important difference that we can clearly see in Fig 4, where we show a ratio of false positives: worm flows classified as non-malicious (circles), together with rate of infection of vulnerable hosts (crosses). We can see that the blue dotted line is the fastest to fall, while the others take more time, and more hosts are infected as a consequence. On the other hand, as the implicit rule generated by the model with lowest noise was the most specific one, the repeated attack in step 35 was more successful than for the models with higher noise levels, even if the new filter was generated rather quickly. All the experiments are averaged over 10 runs for each configuration.

Our results suggest that a trust model that combines network anomaly detection with feedback from host-based IDS can improve network security in a coalition settings, where the reliability, integration simplicity and robustness are the most

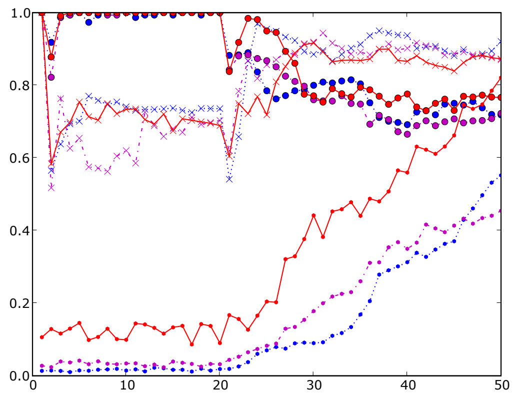


Fig. 3. Inference process and alarm levels for various noise levels.

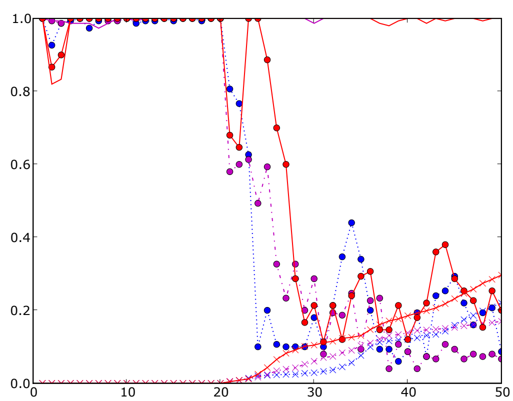


Fig. 4. Ratio of flows passing through the filter for changing noise levels, with ratio of infected vulnerable hosts.

important aspects of any autonomous security system.

In order to better evaluate the actual deployment of the filters in the network, and to verify the impact of network topology and hardware limitations on filtering efficiency, we have implemented a flow-level network simulation: a-net, using the A-globe [14] platform and we are currently working on the evaluation of the reaction phase.

## VI. Conclusions and Current Work

This work presents a hybrid, both network-based and host-based intrusion prevention system oriented towards an application in dynamic network environments without clearly defined security perimeters. Such environments are typical for coalition operations, where the assets of different coalition members are interconnected in the field, in order to cooperate locally, without intervention of command and staff entities. Therefore, we propose a dynamic distributed approach to the problem, where the behavior of all types of components (host and network sensors, IDS agents and reactive components) is autonomous, but controlled by clearly defined policies. Furthermore, with an assumption that the network devices are reconfigurable and can be used for content filtering, the system closes the loop from sensor through decision to autonomous until deployment of defence mechanism. The work is based on two relevant types of techniques from multi-agent filed: dynamic trust modeling and reflective agents.

In the current status, the technology is validated on a high-level network simulation, on the networks with hundreds of nodes. We aim to extend this simulation in two directions: in the first case, we are working with the real-time network data to identify the attacks in the realistic network traffic. On the other hand, we are increasing the scale of the current low-level simulation to validate the response concepts and to evaluate the effects of the real-time filter deployment on worm-like threat mitigation.

## References


[1] J. M. Bradshaw, M. Sierhuis, A. Acquisti, P. Feltovich, R. Hoffman, R. Jeffers, D. Prescott, N. Suri, A. Uszok, and R. Van Hoof. *Agent Autonomy*, chapter Adjustable Autonomy and Human-Agent Teamwork in Practice: An Interim Report on Space Applications, page 296. Springer, 2004.

[2] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, New York, 2nd edition, 2001.

[3] L. Ertoz, E. Eilertson, A. Lazarevic, P.-N. Tan, V. Kumar, J. Srivastava, and P. Dokas. Minds - minnesota intrusion detection system. In *Next Generation Data Mining*. MIT Press, 2004.

[4] P. Haffner, S. Sen, O. Spatscheck, and D. Wang. Acas: automated construction of application signatures. In *MineNet '05: Proceeding of the 2005 ACM SIGCOMM workshop on Mining network data*, pages 197–202, New York, NY, USA, 2005. ACM Press.

[5] T. D. Huynh, N. R. Jennings, and N. Shadbolt. On handling inaccurate witness reports. In *Proc. 8th International Workshop on Trust in Agent Societies*, pages 63–77, Utrecht, The Netherlands, 2005.

[6] D. Moore, C. Shannon, G. M. Voelker, and S. Savage. Internet quarantine: Requirements for containing self-propagating code. In *INFOCOM*, 2003.

[7] S. Northcutt and J. Novak. *Network Intrusion Detection: An Analyst's Handbook*. New Riders Publishing, Thousand Oaks, CA, USA, 2002.

[8] F. Ortin and J. M. Cueva. Non-restrictive computational reflection. *Comput. Stand. Interfaces*, 25(3):241–251, 2003.

[9] D. Perugini, D. Lambert, L. Sterling, and A. Pearce. A distributed agent approach to global transportation scheduling. In *The 2003 IEEE/WIC International Conference on Intelligent Agent Technology (IAT 2003)*, pages 18–24, Halifax, Canada, 2003.

[10] M. Rehak, M. Gregor, M. Pechoucek, and J. M. Bradshaw. Representing context for multiagent trust modeling. In *IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2006 Main Conference Proceedings) (IAT'06)*, pages 737–746, Los Alamitos, CA, USA, 2006. IEEE Computer Society.

[11] M. Rehák, Lukáš Foltýn, M. Pěchouček, and P. Benda. Trust model for open ubiquitous agent systems. In *Intelligent Agent Technology, 2005 IEEE/WIC/ACM International Conference*, number PR2416 in IEEE, 2005.

[12] M. Rehak, M. Pechoucek, and M. Gregor. Trust Modeling with Context Representation and Generalized Identities. Technical report, Gerstner Laboratory, CTU in Prague, 2007.

[13] J. Sabater and C. Sierra. Review on computational trust and reputation models. *Artif. Intell. Rev.*, 24(1):33–60, 2005.

[14] D. Šišlák, M. Rehák, M. Pěchouček, M. Rollo, and D. Pavlíček. *A*-**globe**: Agent development platform with inaccessibility and mobility support. In R. Unland, M. Klusch, and M. Calisti, editors, *Software Agent-Based Applications, Platforms and Development Kits*, pages 21–46, Berlin, 2005. Birkhauser Verlag.

[15] S. R. White, J. E. Hanson, I. Whalley, D. M. Chess, and J. O. Kephart. An architectural approach to autonomic computing. In *ICAC*, pages 2–9. IEEE Computer Society, 2004.

# The Semantic Battlespace Infosphere: A Knowledge Infrastructure for Improved Coalition Inter-Operability

Paul R. Smart and Nigel R. Shadbolt
School of Electronics and Computing Science, University of Southampton, Southampton, SO17 1BJ, UK
{ps02v, nrs}@ecs.soton.ac.uk

**Abstract**—*Operational effectiveness in coalition environments is based on the need for inter-operability at a variety of levels. While inter-operability concerns are most easily thought of in terms of technology, this paper emphasizes the importance of consensual interpretations of the semantic significance of exchanged information. In this paper we outline some of the challenges to effective modes of information exchange in coalition operational contexts. We also discuss potential approaches to these challenges in the context of a semantically-enabled technological framework for information exploitation – the Semantic Battlespace Infosphere (SBI). Relevant aspects of this framework are introduced and some of the socio-technical challenges that are likely to be encountered are discussed.*

## 1. Introduction

Both the UK and the US recognize the importance of trans-national alliances as the basis for future military operations. The vision is that military activities spread across all levels of the operational spectrum (from large-scale war-fighting to peace support and humanitarian assistance) will assume the form of 'coalitions of the willing' [1]. Such coalitions will demand the close inter-operation, but not necessarily integration, of multi-national forces, each of which brings its own set of technological, ideological, organizational, procedural and cultural idiosyncrasies to the theatre of operations. Issues of inter-operability therefore sit at the heart of many research efforts concerned with the future effectiveness of military coalitions. Inter-operability is particularly important in contexts that involve non-military agencies, e.g. diplomatic, humanitarian and civil authorities. In addition, the rapid, opportunistic exploitation of situation contingencies, the need to self-synchronize and the requirement to synergistically marshal diverse military assets in the context of agile force structures, requires the ability to exploit and share information in ways that transcend the traditional boundaries of national affiliation and operational environment (i.e. land, sea and air).

This paper discusses some of the issues related to improved coalition inter-operability that are being researched in the context of the International Technology Alliance[1] (ITA) program. Our approach to coalition inter-operability is grounded in the use of Semantic Web technologies and is organized around a framework for advanced modes of information integration, exploitation and exchange in coalition military contexts. We refer to this framework as the Semantic Battlespace Infosphere (SBI) and present it as an extension to an existing information management system known as the Joint Battlespace Infosphere (JBI) [2, 3]. This paper summarizes the basis for the SBI and describes some of the socio-technical issues that may limit the acceptability or viability of proposed solutions.

## 2. Coalition Inter-operability

Inter-operability issues are most easily thought of in terms of technology, but they can actually assume a variety of forms. Inter-operability issues may, in fact, arise on a number of levels, including, the technological, the socio-cultural, the operational and the epistemic or knowledge-based. While this paper does not attempt to belittle the importance of any of these levels, we argue that many of the most difficult problems, with regard to inter-operability, lie in the realm of 'epistemic inter-operability', i.e. the compatibility that exists between coalition partners with respect to conceptual models, inference processes, reasoning strategies, etc. In these cases, our concern is not merely limited to issues regarding the physical exchange of information; we also have to consider the meaning assigned to information items and the interpretive biases possessed by coalition partners. Such factors can influence the interpretation of information content in terms of both its semantic referents (what the information is about) and its semantic significance (the implications of the information for current and future action). Issues of epistemic inter-operability also extend beyond simple information exchange contexts; they subsume the ability to integrate and fuse information from physically disparate and semantically heterogeneous information sources, as well as the ability to coordinate and orchestrate the activities of distributed problem-solving agents. A key element of these abilities, we argue, lies in the capacity to exploit semantically-enriched representations and to establish mappings between ostensibly disparate representations by virtue of their semantic similarity. The framework we propose, the SBI, presents one strategy for the realization of these semantically-enabled capabilities.

## 3. Semantic Battlespace Infosphere

The JBI is a combat information management system designed to acquire and represent information from a wide

[1] http://www.usukita.org/

variety of information sources with the express purpose of supporting enhanced situation awareness, coalition inter-operability and operational flexibility [4]. The JBI exploits information available in existing C2 (Command & Control) systems, but it does not aim to replace them. Rather, the emphasis is on information exchange between these systems. The JBI also aims to support improved situation awareness based on its ability to integrate (fuse) information from different sources and to make inferences based on environmental data. Situation awareness is further supported by the proposed ability of the JBI to tailor information content to suit the needs of individual end-users: "the commander gets high-level coverage of the campaign, while the soldier in the field gets a detailed description of a nearby hostile base" [2; pg. iii].

It is not the purpose of this paper to review the relative merits or demerits of the approach advocated by the JBI. We suggest that the basic vision proposed by the JBI is sound, but that its applicability to coalition inter-operability is undermined by its failure to fully embrace semantically-enriched representational schemes. Such schemes would seem indispensable for a number of reasons, not least because of the difficulty of agreeing a common language or vocabulary that is accepted across all elements of a coalition formation. Even if a common vocabulary could be agreed, it would not necessarily extend to the variety of humanitarian, diplomatic and government agencies with which coalition partners must often inter-operate. Moreover, semantically-enriched forms of representation support the easy revision and maintenance of existing vocabularies and conceptual models. New terms, such as 'Qandahar', can easily be added and associated with existing terms, such as 'Kandahar', without causing major disruption to existing capabilities. To address the potential shortcomings of the JBI we propose the notion of the SBI, which attempts to take the basic vision of the JBI and extend it with respect to semantically-enabled capabilities[2].

The notion of the SBI does not refer to a specific system or capability so much as a framework for inter-operability-focused modes of technology exploitation and development. Like the JBI, the SBI attempts to provide a capability for inter-operability between previously disparate C2 systems; it also aims to facilitate situation awareness with regard to information integration and publish/subscribe mechanisms. Unlike the JBI, however, the technological commitments endorsed by the SBI concept differ from those characterizing the JBI initiative. In particular, the notion of the SBI countenances a framework within which a variety of semantic technologies are used to support key capabilities (e.g. meaning-preserving modes of information exchange, enhanced situation awareness, a common understanding of the significance of events and information items, etc.) by virtue of their capacity to support and exploit semantically-enriched representations within highly distributed network environments. It builds on many of the guiding principles and technologies of the Semantic Web initiative [5], but the aim is to adapt and extend these principles and technologies to suit the military environment.

The core technological ingredients of the SBI are domain ontologies, created using languages such as the Web Ontology Language (OWL) [6, 7] and Resource Description Framework (RDF) [8]. These ontologies constitute the representational bedrock for inter-operability-related capabilities such as mediation services between federated systems of systems, meaning-preserving modes of information exchange, and query execution at semantic levels of abstraction. Although the technological approach we advocate relies on the use of Semantic Web languages such as RDF, RDFS and OWL, the choice between these languages is often a trade-off between the relative simplicity of languages such as RDF (which supports ease of development and maintenance) versus the greater semantic expressivity of languages such as OWL (which supports greater reasoning capabilities). Although simplicity is to be countenanced in situations where we want to accomplish the widespread adoption of a representational language, it is questionable whether RDF will support the kinds of capabilities we desire in respect of coalition inter-operability solutions. Such solutions are founded on the ability to identify semantic correspondences between ostensibly disparate vocabularies [9] (a capability variously referred to as semantic integration, ontology mapping or ontology alignment), but languages such as RDF may not provide enough semantic constraints in order to effect this mapping (at least not one that can support wholly automated alignment solutions). Ideally, what we seek to accomplish with respect to ontology mapping, is an automated approach that capitalizes on the capacity of semantically-enriched representational formalisms to denote the semantic similarity between ontology elements (i.e. classes, properties and instances). With its greater degree of semantic expressivity, OWL is much better equipped to provide this capability, but a reliance on OWL may affect the tractability of a knowledge capture solution (see Section 4.1), especially in a coalition environment where the extent of the conceptual space and the cultural heterogeneity of participating agencies may undermine any effort to develop a single, all-encompassing ontology.

Semantic queries are another essential element of the SBI framework. Semantic query languages, such as SPARQL, work in conjunction with ontologies to increase the accessibility of information content in a way that is more likely to address the goal-relevant epistemic needs and requirements of a particular problem-solving agent. The point is that unlike conventional query languages, such as SQL, semantic query languages operate at the semantic level – they focus on the conceptual structure of the domain and they do not make any assumptions about the underlying structure of the dataset used to store data within the domain. This shift in focus is of potential benefit because it allows queries to be expressed in a form that makes sensible contact with the kinds of conceptual abstractions and

[2] Note that our aim here is not to argue for the replacement of the JBI, so much as its extension with semantically-enabled capabilities.

relationships identified within a domain of discourse[3].

Languages such as SPARQL provide an effective semantic query solution, and many knowledge repositories cater for such queries, e.g. 3Store [10]; nevertheless, it is not entirely clear how such queries should be executed in an environment where knowledge content is distributed across multiple nodes of a dynamic (and sometimes *ad hoc*) network environment. One particular problem concerns the likely volatility of semantic query results in military contexts characterized by *ad hoc*, mobile and wireless communication infrastructures. The key problem is that time-variant changes in network connectivity (or the differences in connectivity apparent from the perspective of physically distributed military agencies[4]) results in the differential availability of nodes and their associated knowledge resources. This can contribute to a confusing situation picture because query results executed from one location in the network need not coincide with the results of the same query executed elsewhere. Moreover, the same query may return different results at different times based on the physical distribution of knowledge resources and the extent of intervening changes in network topology. The distributed nature of knowledge resources is a potential problem here because it complicates the possibility of establishing a common collective representation about the nature and implications of the current situation picture. Ultimately, we argue, this can attenuate shared situation understanding and situation awareness and undermine the potential for coalition inter-operability.

These concerns about query execution in distributed environments also apply, to some extent, to reasoning processes. To the extent that reasoning processes subtend multiple, physically distributed resources, then reasoning outcomes will depend on the relative stability of the network infrastructure across multiple invocations of the same reasoning process (either from different points in the network or from the same point at different times). As with query capabilities, the dynamic nature of the military network environment (in contrast, perhaps, to the situation with the World Wide Web), potentially undermines the possibility for a common understanding of the operational picture, especially when coalition elements are geographically distributed and connectivity privileges are non-uniform.

Like the JBI, the SBI does not aim to replace existing systems; rather the idea is that the SBI will act as the middleman within a federated system of systems. The SBI thus aims to serve as a mediator between previously disparate systems, enabling applications and services to exchange information in ways that preserves the original semantics of information content. Note that this does not mean that information will necessarily remain the same (e.g. with respect to its physical form) across information exchange contexts. The point about using domain ontologies, in conjunction with ontology alignment solutions, is that the emphasis is on the semantics of the information content, not the information content *per se*. Thus, if information was required to be transformed in the course of an exchange involving culturally-disparate user communities, then such a transformation would be undertaken in an effort to avoid semantic ambiguity and misinterpretation by the target community. The key point here is that our mediation solution does not simply aim to provide a mechanism for information exchange; rather the emphasis is on enabling information exchange with respect to common semantic frames of reference, frames of reference that make explicit the meaning of information content to all coalition partners.

The emphasis on explicit semantics and the capability for flexible modes of information transformation to support meaning-preserving modes of information exchange is something which tends to be overlooked by many other approaches to information exchange in the military domain. Conventional IEDM solutions [11], for example, may be poorly suited to meaning-preserving modes of information exchange simply because such approaches do not avail themselves of a sufficiently rich repertoire of semantically-enriched representational formalisms. The JBI is also deficient in this respect. The definition of battlespace objects in the JBI is based on XML Schema language, but such languages are potentially inadequate in a coalition environment where data standards may be difficult to establish, enforce and sustain. Perhaps most importantly, an XML schema provides a syntactic specification that defines the structural organization of data. While this is valuable as a mechanism for data exchange, it does not provide any means for the effective representation of semantic information, i.e. what the data means in terms of its actual relationships to other data within the same dataset and potential relationships to data that may be defined elsewhere and received at a different time. To capture the semantic significance of data requires knowledge about classes of data objects and how these objects relate to one another. This type of information is precisely what domain ontologies within the SBI framework are intended to capture.

## 4. (SOCIO-)TECHNICAL CHALLENGES

The realization of capabilities implied by the notion of the SBI depends on the resolution on a number of key challenges. Some of these challenges derive from the idiosyncratic nature of the military network environment,

[3] One concern, of course, from an inter-operability perspective is that such languages, despite exploiting domain ontologies, are likely to be largely ineffectual in a situation where different coalition partners are making distinct ontological commitments (i.e. using different, albeit overlapping, ontologies within the same problem domain). However, such concerns can be allayed, to some extent, by the use of ontology alignment solutions. In this case, semantically-equivalent entities in separate ontologies can be targeted by the query, despite the fact that they may be identified using different linguistic labels.

[4] For example, imagine a sub-component of the larger network infrastructure becomes temporarily isolated or has limited connectivity to the remainder of the network. In this case, the networked resources available to the same query executed at different points in the network could yield radically different query results.

especially the prevalence of *ad hoc*, mobile, wireless, networks; others relate to more general issues that pervade the Semantic Web and user-centred design communities. Above all we suspect that many of these challenges encompass the social, cultural and psychological domains - they are not merely related to the provision of a technological and representational substrate for meaning-preserving modes of information exchange. The point is that we may need to pay careful attention to the cultural differences between coalition partners and the organizational, psychological and social contexts that influence the acceptability and/or usability of proposed solutions. The following sections provide an overview of (some of) the potential challenges we face with respect to a realization of semantically-enabled capabilities supported by the SBI.

### 4.1. Knowledge Capture

A key limiting factor in the widespread adoption and use of ontologies is the overhead associated with their initial development and subsequent maintenance. Ontology development relies on the capture and formalization of domain knowledge, and this can sometimes serve as a significant bottleneck in the knowledge engineering process [12]. Even when individuals or institutions can marshal the resources to overcome this bottleneck, problems of completeness (does the ontology achieve adequate coverage of the target domain?), consistency (does the ontology align itself with the representational strategies adopted by other agencies?) and topicality/currency (does the ontology reflect the current conceptual focus and complexity of target user communities?) all threaten to undermine the potential utility of an ontology.

One approach to overcoming these problems focuses on the extent to which some aspects of the semantic infrastructure of a domain can be derived from the actions and interactions of agent communities within the domain. The idea here is that information networks serve as a substrate for patterns of activity that contribute to the (automatic) pooling of knowledge and expertise within specific communities of interest. One example, of this phenomenon is represented by the notion of ‘collaborative filtering’ wherein, for example, patterns of consumer activity associated with the purchase of books, CDs, DVDs, etc., allows for the generation of product recommendations that then influence subsequent purchase patterns [13]. Another example is provided by customizable tagging schemes, such as those used by Flickr. Such schemes devolve much of the responsibility for initial ontology (or at least taxonomy) development to the user community, which avoids the upfront cost for agreeing upon a taxonomy when, perhaps, the nature of the information to be collected and its use are not yet known. It also allows the taxonomy to emerge and change dynamically as additional information is accumulated. One aspect of our research in the ITA aims to explore these and other techniques as a means for understanding the extent to which networked patterns of activity can be exploited to infer or derive some insight into the semantic infrastructure of a domain.

### 4.2. Distributed Knowledge Infrastructure

A second major challenge to semantically-enabled capabilities, in the context of the SBI, is the notion of a distributed knowledge infrastructure. Just as the vision of the Semantic Web is a vision of an extended Web of machine-readable information and automated services [5], so the notion of the SBI builds on the capabilities (and limitations) of existing military information and communication infrastructures – infrastructures in which knowledge (and the services they support) will often be distributed across multiple nodes of a large-scale information network. Decentralization of resources within such networks contributes to the resilience of higher-order capabilities in the face of network disruption (i.e. the failure of network elements), but it also implies a number of distinct technical challenges, including (but certainly not limited to) the following[5]:

1. How are we to orchestrate and coordinate the activity of services in a dynamic network environment, an environment in which different service-related capabilities are dependent on the time-variant topological organization of the network?
2. How are we to fully exploit knowledge and services in an environment with no centralized standards agency? For example, how are we to deal with the tendency of novel coalition elements to describe knowledge and services in a manner that reflects their idiosyncratic (perhaps culturally-entrenched) modes of conceptualizing problem domains and describing service-capabilities?
3. How can we cope with the various performance constraints associated with network infrastructures (e.g. the latency required to aggregate knowledge across multiple network nodes) to yield knowledge services (e.g. reasoning capabilities) that complete within an operationally useful timeframe?

There are a variety of potential responses to these challenges, not all of which fall within the scope of the current paper. Tentative solution strategies include the active (knowledge-driven?) configuration of network topologies to better suit specific knowledge processing requirements, the duplication of knowledge resources throughout the network in order to support a form of epistemic redundancy in the face of network dynamics, and the use of query caching strategies to achieve virtual knowledge-processing stability in the face of network disruption. However, many of the inter-operability challenges to be tackled in the context of our SBI research programme relate to the need to align disparate ontologies and integrate information content from multiple semantically-heterogeneous information sources. This is the challenge of semantic integration.

[5] These are in addition to the problems associated with the potential volatility of semantic query results as discussed in Section 3.

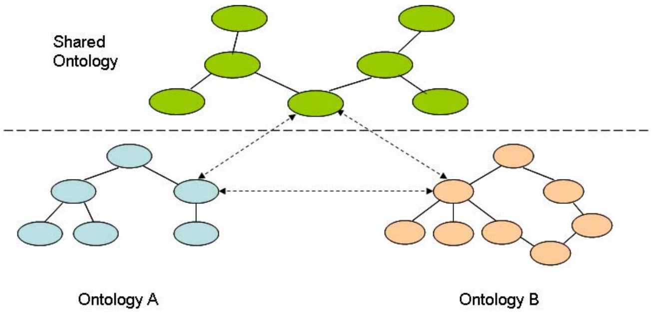


**Figure 1:** Ontology mapping/alignment

### 4.3. Semantic Integration

As the global information environment becomes increasingly pervasive and spans ideologically, culturally and ethno-linguistically diverse communities, so the information exchange challenge becomes ever harder. The battlefield network currently consists of many information sources, including in-theatre sensors, platforms, intelligence reports and remote information such as archival intelligence and satellite data. In future coalition contexts, strategies for operationally-effective modes of information exchange and exploitation will need to target a wider variety of disparate information repositories and communication systems, including digital datalinks, military information repositories, and the totality of the information space available via internet-enabled and peer-to-peer computing environments. In situations such as these the potential for semantic ambiguity is rife because the meaning of symbolic information often reflects the experiential, epistemic, cultural and task-specific biases of the information provider. Both the semantic referents and semantic significance of information is not invariant with respect to information exchange contexts, rather one sees a degree of semantic specificity - a community specific interpretation of meaning that may not necessarily transcend cultural, organizational and/or national boundaries. The point is that once we encounter distributed network environments that subtend a wide variety of information systems, sources and user communities (as is often the case in coalition contexts), we face a critical challenge in terms our ability to integrate and share information in a semantically-sensible manner (one that respects the meaning assigned to information content by the originating agent or agency). We refer to this as the semantic integration challenge.

Our approach to this challenge, in the context of the ITA, is grounded in the use of semantically-enriched domain knowledge models (ontologies) and the notion of ontology alignment. Ontology alignment (see Figure 1) is a key element of semantic integration [9]. Its aim, in essence, is to establish mappings between the elements of multiple (ostensibly disparate) domain ontologies as a means of identifying semantically-equivalent sub-components. Once established, these mappings can be used to drive information aggregation and integration activities (as might be required for goal-relevant information processing [see 14], for example), but most of all we see such mappings as a mechanism for ontology-based mediation of information exchange between conceptually or linguistically disparate communities.

A wide variety of tools and techniques have emerged to support ontology alignment [9, 15] all of which rely, to a greater or lesser extent, on the similarity between ontology elements, e.g. in terms of common relationships, linguistic labels or instance sets [15]. Nevertheless, there is fertile ground for research here both in terms of the accuracy of the techniques and the degree of automation that each technique supports. One potentially interesting strategy is to explore cognitive science techniques for concept mapping, which, in general, have received little attention in the Semantic Web community. Goldstone et al [16] thus describe a technique that is grounded in the use of neural networks to establish mappings between the elements of a conceptual system. Such approaches, they argue, could be successfully applied to problems in ontology alignment.

One aspect of our approach in developing the SBI, and one that most strongly discriminates our work from previous work in ontology mapping/alignment, concerns the need to dynamically integrate and align ontology fragments in specific task contexts (see Figure 2). In essence this approach eschews the idea of large-scale, global ontology alignment independent of task context; rather it countenances the idea that ontologies (or relevant sub-components thereof) should be dynamically aligned to reflect goal and task-relevant processing. Such capabilities rely on effective mechanisms to represent the epistemic requirements of tasks, and to prune larger ontologies in light of these requirements, i.e. to extract just those elements that are relevant for current problem-solving activity.

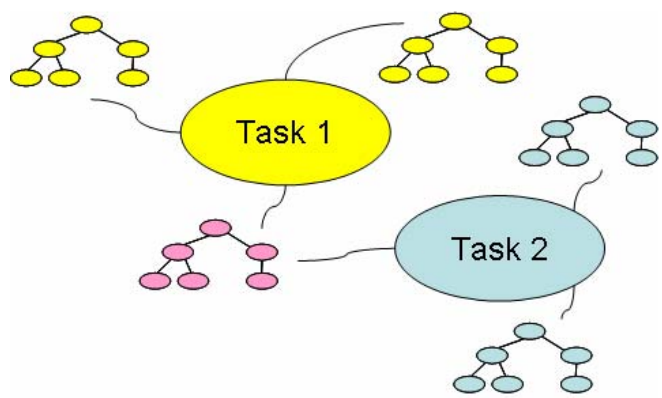


**Figure 2**: Task-specific semantic integration

### 4.4. Visualization & Interaction

The human end-user is a key element of the SBI. Even in the case of automated services, the outcome of such services needs to be carefully geared to suite the needs and requirements of the human operator in relation to problem-solving objectives. The SBI is, in essence, intended to provide a framework for assisting coalition members with respect to knowledge processing, and thus its success is ultimately predicated on the ability of application interfaces to support operationally-effective modes of working (and thinking).

User-centred design approaches are likely to be of paramount importance in understanding the opportunities for operationally-effective information exploitation within coalition contexts. One approach we have adopted in the ITA is the notion of Goal-Directed Task Analysis [14]. Such an approach embraces both user-centred design principles and also focuses on the goals of the human operator. In essence it provides a framework for understanding the information requirements of the decision-maker in specific operational contexts and helps to drive information aggregation, semantic integration and service coordination processes in respect of goal-oriented information processing. Such analyses are critical in terms of understanding the (sometimes subtle) differences in problem-solving strategies adopted by different coalition partners, and they help to customize semantically-enabled information aggregation and integration processes in ways that best support shared situation awareness and collaborative problem solving [14].

## 5. Conclusion

In this paper we have attempted to highlight the potential role played by semantic technologies with regard to coalition inter-operability issues. We have introduced the notion of the SBI, which provides a vision of semantically-enabled capabilities in the future battlespace environment, and suggested ways in which this framework could facilitate inter-operability in military coalition contexts. The aim of the SBI is, in essence, to develop a common semantic frame of reference to support consensual interpretations of entities, events and actions across force elements and between coalition partners. This is important because future coalition capabilities will critically depend on an ability to exchange information in ways that preserve the meaning assigned to information content, especially when the operational context demands close cooperation with non-military agencies (e.g. diplomatic, humanitarian and civil authorities in the case of peace support and humanitarian relief contexts). A key objective of the SBI is therefore to provide a foundation for coalition inter-operability, enabling semantic integration with respect to both digital datalink networks and unstructured, non-military information sources. Notwithstanding the ready availability of extant semantic technologies (e.g. OWL, RDF, SPARQL, etc.), a number of aspects of the military coalition environment make the solution space somewhat different from that which we encounter in the case of the Semantic Web. Such challenges include the need to deal with the idiosyncrasies of the military network infrastructure (e.g. its mobile, *ad hoc* and wireless nature), the need to engender effective information exchange solutions via ontology alignment/mapping mechanisms and the need to deal with the potential (perhaps culturally-entrenched) differences between coalition partners with respect to the interpretation of situation-relevant information. These issues impact directly on coalition inter-operability issues and form a key component of our ongoing research in the ITA.

## Acknowledgment

This research was sponsored by the US Army Research laboratory and the UK Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Army Research Laboratory, the U.S. Government, the UK Ministry of Defense, or the UK Government. The US and UK Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon. The authors would like to express their thanks to Dave Braines at IBM UK Ltd for his comments on an earlier draft of this paper.

## References

[1] "Delivering Security in a Changing World: Defence White Paper," Ministry of Defence, UK 2003.

[2] "Building the Joint Battlespace Infosphere: Volume 1 - Summary," United States Air Force - Scientific Advisory Board, USA SAB-TR-99-02, 2000.

[3] "Building the Joint Battlespace Infosphere: Volume 2 - Interactive Information Technologies," United States Air Force - Scientific Advisory Board, USA SAB-TR-99-02, 1999.

[4] "Information Management to Support the Warrior," United States Air Force - Scientific Advisory Board, USA SAB-TR-98-02, 1998.

[5] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, vol. 284, pp. 34-43, 2001.

[6] G. Antoniou and F. van Harmelen, "Web Ontology Language: OWL," in *Handbook on Ontologies*, S. Steffen and R. Studer, Eds. Berlin: Springer Verlag, 2003.

[7] D. L. McGuiness and F. van Harmelen, "OWL Web Ontology Language Overview," World Wide Web Consortium, 2004.

[8] B. McBride, "The Resource Description Framework (RDF) and its Vocabulary Description Language RDFS," in *Handbook on Ontologies*, S. Steffen and R. Studer, Eds. Berlin: Springer Verlag, 2003.

[9] Y. Kalfoglou and M. Schorlemmer, "Ontology mapping: the state of the art," *Knowledge Engineering Review*, vol. 18, pp. 1-31, 2003.

[10] S. Harris and N. Gibbins, "3store: Efficient Bulk RDF Storage," presented at 1st International Workshop on Practical and Scalable Semantic Web Systems, Sanibel Island, Florida, USA, 2003.

[11] E. Lasschuyt, M. van Hekken, W. Treurniet, and M. Visser, "How to make and effective information exchange data model," NATO RTO IST-042 Symposium on Coalition C4ISR Architectures and Information Exchange Capabilities, The Hague, Netherlands JWP 3-52, 2004.

[12] F. Hayes-Roth, D. A. Waterman, and D. B. Lenat, *Building Expert Systems*. Reading, Massachusetts: Addison-Wesley, 1983.

[13] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Internet Computing*, vol. 7, pp. 76-80, 2003.

[14] M. Endsley, B. Bolte, and D. G. Jones, *Designing for Situation Awareness: An Approach to User-Centered Design*. London: Taylor & Francis, 2003.

[15] "State of the art on ontology alignment," University of Karlsruhe, Karlsruhe, Germany Knowledge Web Deliverable 2.2.3, 2004.

[16] R. L. Goldstone, Y. Feng, and B. J. Rogosky, "Connecting concepts to each and the world," in *Grounding Cognition: The Role of Perception and Action in Memory, Language and Thinking*, D. Pecher and R. Zwaan, Eds. Cambridge, UK.: Cambridge University Press, 2005.

# Cooperative Self-Organizing Coalitions: Technological Framework

Alexander Smirnov, Nikolay Shilov, Tatiana Levashova
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Russia
E-mail: {smir, nick, oleg}@iias.spb.su

**Abstract**—*The paper proposes an approach aimed at solving a problem of organizing ad-hoc self-configuring networks for such tasks as disaster relief and evacuation. Usually the tasks of this type involve a number of different heterogeneous teams, which have to collaborate in order to succeed. Centralized control is not always possible due to damages in local infrastructure, different subordination of participating teams, etc. The approach assumes decentralized communication and ad-hoc decision making based on the current situation state and its possible future development in order to avoid the above obstacles. There are a number of problems to be resolved within the approach. They include interoperability at both technological and semantic level, situation understanding by the members via information exchange, protocols of ad-hoc decision making for self-organization and other. It is planned to use such technologies as situation management, knowledge and ontology management, profiling, intelligent agents, Web-services, decision making support, negotiation protocols, etc.*

## 1. Introduction

Usually, coalition-based disaster relief and evacuation operations involve a large number of different heterogeneous teams (sometimes multinational), which have to collaborate in order to succeed. Such teams might include medical brigades, firefighters, rescuers, military personnel, commercial / governmental / non-commercial organizations, volunteers, etc. Besides, during such operations it might be necessary to use external sources to get required information (e.g., medical databases, transport availability, weather forecasts). Coalition-based organization requires intensive information exchange in order to achieve necessary level of the situational awareness, create ad-hoc action plans, have continuously updated information.

Centralized control is not always possible due to probable damages in local infrastructure, different subordination of participating teams, etc. Another disadvantage of the centralized control is its possible failure that would cause stopping of the entire operation. Possible solution for this is organization of a decentralized self-organizing coalitions consisting of the operation members. However, in order for this coalition-based network to operate it is necessary to solve a number of problems that can be divided into technical (hardware-related) and methodological constituents.

The paper proposes an approach aimed at solving a problem of organizing ad-hoc self-configuring networks [1-3] for such tasks as disaster relief and evacuation. Proposed approach assumes decentralized communication and ad-hoc decision making based on the current situation state and its possible future development in order to avoid the above obstacles. There are a number of problems to be resolved within the approach. They include interoperability at both technological and semantic level, situation understanding by the members via information exchange, protocols of ad-hoc decision making for self-organization and other. Proposed technological framework incorporates such technologies as situation management, knowledge and ontology management, profiling, intelligent agents, Web-services, decision making support, negotiation protocols, etc.

The paper covers such subproblems as providing for semantic interoperability between the members and definition of the standards and protocols to be used. Proposed methodology is based on the earlier developed concept of knowledge logistics [4] and includes such technologies as situation management, ontology management, profiling and intelligent agents [4]. Standards of information exchange (e.g., Web-service standards), negotiation protocols, decision making rules, etc. are used for information / knowledge exchange and rapid establishing of ad-hoc partnerships and agreements between the operation members.

## 2. Proposed Approach

At the first stage of the research the lifecycle phases of the self-configuring network and major requirements to them were defined (Table 1). Based on these requirements the main ideas the approach is based on were formulated:

1. A common shared top-level ontology (application ontology) serves for terminology unification. Each member has a fragment of this ontology corresponding to

Table 1. Lifecycle phases for the self-configuring network, its needs and services to fulfill them

| Life cycle phase | Needs | Services |
|---|---|---|
| Community building (once, new members are added on a continuous basis) | Common infrastructure<br><br>Common communication standards and protocols | Modelling goals and objectives<br><br>Identification, qualification, registration of members<br><br>Common knowledge representation<br><br>Common modeling for community members |
| Formation (continuous, initiated by the situation, or a task as a part of the situation) | Task definition model (context)<br><br>Partner selection | Task modelling<br><br>Rules of partner selection |
| Operation (continuous) | Coordination and synchronization | Rules of re-negotiation and solution modification if necessary |
| Discontinuation (continuous, initiated by members) | Termination of the established agreements | Update of the current solution |

his / her capabilities / responsibilities. This fragment is synchronized automatically when necessary (not during the operation). The ontology and rules for fragment definition are to be defined. Here there are two possible directions: (i) building a common "heavy" detailed ontology that is to be used by all the community members, and (ii) peer-to-peer partner search and negotiation with dynamic update of the members' knowledge depending on their partners (no common detailed ontology). The approach presented proposes a combination of these two directions.

2. Each member has a profile describing his/her capabilities, appropriate ontological model.

3. Each member is assigned an intelligent agent, representing him/her. The agent collects information required for situational understanding by the member, negotiates with other agents to create ad-hoc action plans. The agent should be defined rules to be followed during negotiation processes. These rules depend on the role of the appropriate member.

4. Web-service standards are used for interactions. External sources (e.g., medical databases, transport availability, weather forecasts) should also support these standards and the terminology defined by the common shared ontology.

To estimate the applicability of the approach a case study for disaster relief and evacuation operation will be used. The scenario includes the following tasks: portable hospital configuration and allocation, delivery of the portable hospital supplies, delivery of required supplies to the disaster site, evacuation of people from the disaster site. The detailed description of the case study can be found in [5]. Evaluation of the approach is planned to be done via development of a research prototype implementing the approach and performing experiments with this prototype based on the above case study.

## 3. TECHNOLOGICAL FRAMEWORK

The generic scheme of a self-organizing network is presented in (Figure 1). Each operation member is represented as an intelligent agent acting in the system. The architecture of the agent is presented in Figure 2. Each agent has his own knowledge stored in his knowledge base. This knowledge is described by a potion of the common shared ontology related to the current agent's (and member's) tasks and capabilities and called context. Capabilities, preferences and other information about the agent are stored in his profile that is available for viewing by other agents of the community. It facilitates communication, which is performed via the communication module responsible for meeting protocols and standards that are used within the community.

The agents communicate with other agents with two main purposes: (1) they establish links and exchange information for better situation awareness, (2) they negotiate and make agreements for coordination of their activities during the operation. The agents may also get information from various information sources, for example, local road network can be acquired from a geographical information system (GIS).

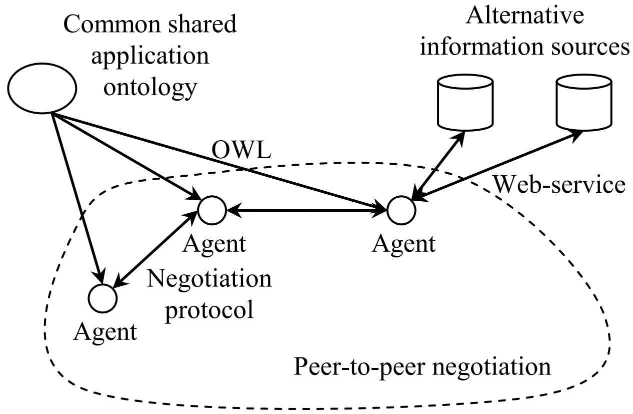


Figure 1. Generic scheme of a self-organising network

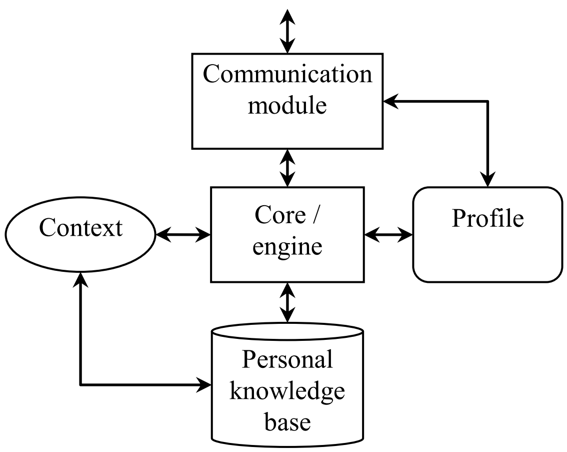


Figure 2. Agent architecture

## 4. APPLICATION ONTOLOGY

The main idea actualized in the approach is capturing information and knowledge relevant to a task at hand, organization them into contexts, and solving this task as a constraint satisfaction problem.

Two types of context are used: 1) *abstract context* that is an ontology-driven model integrating information and knowledge relevant to the problem, and 2) *operational context* that is an instantiation of the abstract context with data provided by the information sources including users or calculated based on functions specified in the abstract context. Resources providing information and knowledge to a context are the application ontology, information sources, and users.

Relevance of information and knowledge is evaluated on a basis how they are related to the modeling of the task of the acting unit. The application ontology serves as a knowledge source. It is created for a macro-situation by subject experts. The application ontology specifies domain-related and task-related knowledge. This knowledge appears in the application ontology as domain and tasks & methods constituents respectively. Within the considered case study the application ontology was created to specify the knowledge required in disaster events. Some parts of the application ontology were imported from ontologies found in Internet ontology libraries.

Top-level classes of the domain constituent of the application ontology and relationships between them are shown in Figure 3. Below, brief descriptions for these classes and relationships between them are given.

Class *Situation* is a state or event consisting of one or more objects having certain properties, or bearing certain relations to each other. This class comprises a set of states (e.g., *Weather*) and situations (e.g., *Emergency*) considered being related to the disaster event.

Class *Helping Operations* represents *Helping Actions* and *Helping Services* involved in disaster aid operations. Class *Helping Actions* comprises a set of classes representing operations (e.g., *Relief Operations*) and events (e.g., *Rescuing*) in which some actor helps someone, either by doing something that directly benefits that one, or by preventing something that would harm him/her. Class *Helping Services* comprises a set of classes representing services providing while the disaster aid operations (e.g., *Informing*, *Transportation Service*).

Class *Role* represents persons, organizations, and resources whose instances are "involved" in situations in various ways. E.g., subclass *Resource* introduced in this class represents *Transportation Devices* that can be used for disaster relief purposes. As well, class *Role* specifies roles filled by members in an event while their job activities. The purpose of this is to reduce the amount of information to be analyzed by the decision makers and to provide them with the information that is relevant for a particular user (for a particular user role). For job role specification subclass *Job Role* is introduced. Referring to the disaster relief task following job roles have been introduced in the class *Emergency Worker* (a subclass of the class *Member Role*): *Emergency Decision Maker*, *Emergency Dispatcher*, *Emergency Medical Technician*, and *Firefighter*.

Class *Locus* is used to describe the region (area) where the situation takes place. This class is used to model the region. The class comprises a set of subclasses for modeling region infrastructure, geographical area, object coordinates, region (object) boundaries, etc.

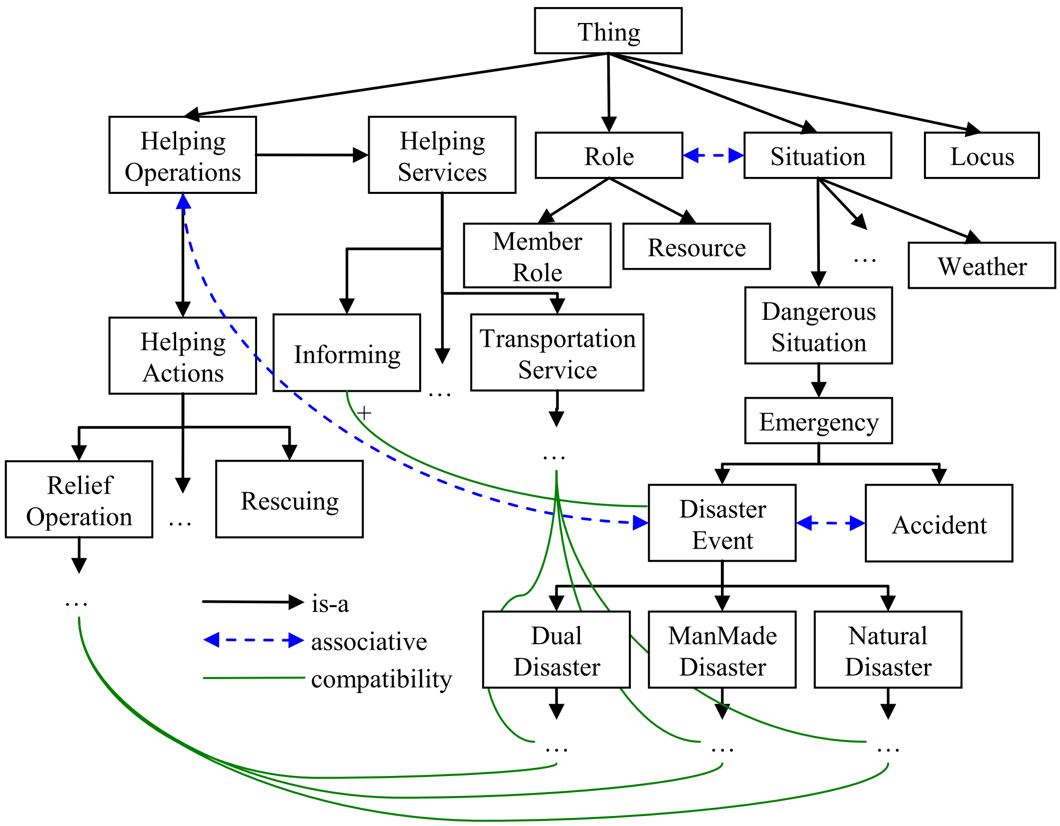


Figure 3. Application ontology: top-level classes of the domain ontology constituent

Associative relationships between classes *Role* and *Situation* relate situations to instances that are "involved" in them in various ways. The instances fill different roles. Associative relationships between classes *Disaster Event* and *Accident* indicate that disasters may cause accidents. Associative relationships between classes *Helping Operations* and *Disaster Event* show that if a disaster happens helping operations are needed. Kinds of the operations that are needed for helping in different disasters are specified as class compatibility relationships. In the figure this is depicted as the compatibility relationships between subclasses of the class *Helping Operations* and subclasses of the class *Disaster Event*. If a compatibility relationship is specified as compatible then the helping action related by this relationship to the disaster (a class corresponding to the disaster) may be required for disaster relief. If this relationship is specified as incompatible then there is no necessity in this action. For instance, the class *Founding An Organization* that is a subclass of the class *Helping Operations* is specified as incompatible to the subclass *Fire* that is a subclass of the class *Disaster Event* (the classes *Founding An Organization* and *Fire* are not appeared in the figure since they are bottom-level classes). Compatibility relationships in the figure are shown between the class *Informing* and the class *Disaster Event* ("+" indicates that the related classes are compatible). The relationship means that all kinds of informing (announcing, reporting, etc.) can be required when a disaster happens. All kinds of the relationships are inherited by the lower-level subclasses starting from the class they are specified for.

## 5. NEGOTIATION PROTOCOL

Making feasible decisions fast is very important in disaster response and evacuation operations. Since time limits often do not allow performing optimization to find optimal solutions for problems it is reasonable to speak about feasible solutions. Below, requirements for choosing a negotiation protocol are presented. The requirements are based on the described above principles for building a cooperative self-organising network.

In order to choose a protocol the main specifics to the approach were formulated as follows:

1. *Contribution*: the agents have to cooperate with each other to make the best contribution into the overall system's benefit – not into the agents' (members') own benefits;

2. *Task performance*: the main goal is to complete the task performance – not to get profit out of it;

3. *Non-mediated interaction*: the agents operate in a decentralized community and in most of the

negotiation processes there are no agents managing the negotiation process and making a final decision;

4. *Common terms*: since the agents work in the same system they use common terms for communication. This is achieved via usage of the common shared ontology.

5. *Trust*: since the agents work in the same system they can completely trust each other (the agents do not have to verify information received from other agents);

Classical protocols [6] have been analyzed in order to choose one most suitable for the approach (Table 2). Based on the analysis of these protocols and the above requirements to them, the contract nets and bargaining can be chosen as a basis for development of the negotiation model in the approach.

Table 2. Comparison of negotiation protocols for self-organizing networks

| Protocols / Criteria | Voting | Bargaining | Auctions | General Equilibrium Market Mechanisms | Coalition Games | Contract Nets |
|---|---|---|---|---|---|---|
| Contribution | ☑ | ☑ | ☐ | ☐ | ☐ / ☑ | ☑ |
| Task performance | ☐ / ☑ | ☐ / ☑ | ☐ | ☐ | ☐ | ☑ |
| Non-mediated negotiation | ☐ | ☐ / ☑ | ☑ | ☐ | ☐ | ☐ |
| Common terms | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ |
| Trust | ☑ | ☑ | ☐ | ☑ | ☑ | ☑ |

## 6. Conclusions

The paper represents a technological framework for the approach to disaster relief and evacuation based on self-organizing coalition networks. It is proposed that self-organization can resolve problems arising from failures of centralized control due to probable damages in local infrastructure, different subordination of participating teams, etc. The approach is at its early stages and only some of its parts have been developed. The paper describes the common shared application ontology that is to be used by the agents representing operation members. Then, the criteria identified for choosing the basic negotiation protocol and its further modifications are discussed. The authors believe that once completed the proposed architecture could efficiently work for a range of the real world problems.

## Acknowledgements

This research was carried out as a part of CRDF partner project # RUM2-1554-ST-05 with US ONR and US AFRL, projects funded by grants # 05-01-00151 and # 06-07-89242 of the Russian Foundation for Basic Research, as well as projects # 16.2.35 of the research program "Mathematical Modelling and Intelligent Systems", # 1.9 of the research program “Fundamental Basics of Information Technologies and Computer Systems” of the Russian Academy of Sciences.

## References

[1] A. C. Viana, M. D. Amorim, and S. Fdida and J. F. Rezende, “Self-organization in spontaneous networks: the approach of DHT-based routing protocols”. *Ad Hoc Networks Journal*, special issue on Data Communications and Topology Control in Ad Hoc Networks, Vol. 3, No. 5, 2005, pp. :589-606.

[2] B. Hammer, A. Micheli, A. Sperduti, M. Strickert. “Recursive self-organizing network models”. *Neural Networks*. Vol. 17, No. 8-9, 2004, pp. 1061-85.

[3] T. Nakano and T. Suda, “Self-Organizing Network Services with Evolutionary Adaptation”, *IEEE Transactions on Neural Networks*, Vol. 16, No. 5, 2005

[4] A. Smirnov, M. Pashkin, T. Levashova, N. Chilov, “Fusion-Based Knowledge Logistics for Intelligent Decision Support in Network-Centric Environment”. George J Klir (Ed.) *International Journal of General Systems*. Taylor & Francis, ISSN: 0308-1079, Vol. 34, No 6, 2005, pp. 673-690.

[5] A. Smirnov, T. Levashova, N. Shilov. Context-Driven Decision Support for Megadisaster Relief. *Journal of Emergency Management*, Prime National Publishing Corporation, ISSN 1543-5865, Vol. 4-5, 2006, pp. 51-56.

[6] T. Sandholm, “Distributed Rational Decision Making”, *Multiagent Systems: A Modern Introduction to Distributed Artificial Intelligence (*G.Weiss, ed.), MIT Press, pp. 201-258, 1999.

# Dynamic Planning & Execution

Maj Andy Aitken, Defence Science and Technology Laboratory (Dstl), UK. aaitken@dstl.gov.uk
Todd Humiston, Air Force Research Laboratory (AFRL), US. Todd.Humiston@rl.af.mil
Jitu Patel, Defence Science and Technology Laboratory (Dstl), UK. jmpatel@dstl.gov.uk

**Abstract**—*In order to gain tactical and operational advantage, the military must deliver effects by conducting activities at optimum tempo; thus staying inside an adversary's decision-action cycle. To date, planners have used various methods to deal with changes in situation: contingency planning, commander's intent and inbuilt flexibility in plans. However, some recent operational experiences indicate that the dynamism of the battlespace has outstripped current ability to respond effectively. The challenge is how to plan and execute continuously in the face of a rapidly changing situation. The Technical Cooperation Program (TTCP)[1] Action Group 1 (AG1) has been investigating both the theory and practical application of Dynamic Planning and Execution (DP&E). Having examined the problem in a Defence context, the findings suggest that Representation of Plans and Coordination are two of the key technical enablers in realising DP&E. This paper is a summation of AG1's work on DP&E to date, and will present the findings from studies and explain the rationale and nature of technical challenges.*

## 1. Introduction

The environment in which our Armed Forces operate has changed greatly since the end of the Cold War. The order of the bi-polar world has been replaced with a contemporary operating environment which is typified by disorder, complexity and uncertainty. Above all, this environment is dynamic, where constant change and a rapidly evolving situation are the norm. The Armed Forces have had to adapt to face the extremely high tempo of this new operating environment [1]; they must have the agility to create desired effects against enduring uncertainty [2]. As the Armed Forces deliver effects by planning and then conducting activities, the military planning and execution processes must become as agile (responsive, resilient, flexible and adaptable) [3] as the application of force itself.

AG1 has been investigating DP&E since 2003 [4]. Its initial efforts were devoted to identifying collaborative activities within national research programmes on DP&E. This revealed that the concept is not well understood and that there has been very little current research into the area. However, there are a number of military capabilities that would benefit from having the ability to dynamically plan and execute [5]. To understand the relevance of DP&E to the Armed Forces, a short study investigated whether the concept is recognised in current military doctrine, and examined recent operational experience [6]. Over the last three years, AG1 has held number of workshops to discuss the issues and technical challenges in realising DP&E.

The next section presents findings on the military relevance of DP&E, which enabled the development of a conceptual model. This is followed by a brief description of AG1's examination of potential areas for the practical application of DP&E. The final section presents the issues and technical challenges in realising DP&E.

## 2. A Military Perspective of DP&E

The first task for the study was to identify whether Defence doctrine already recognised DP&E, subject to which the meaning of the phrase 'dynamic planning and execution' would have to be defined. Without specifying what was meant by DP&E, it would be difficult to elicit opinions from military planners. The next step was to assess whether there was a need for a DP&E capability.

### Definitions

It was established that UK concepts and doctrine contained no official definition of DP&E. The definitions of *planning* and *execution* are well understood, but the dictionary definition of *dynamic* highlighted the relevance of change. [2] DP&E is concerned with ongoing change and activity during planning and execution cycles across all levels of command. The study therefore focused upon how planners and executors deal with a change in the situation during the execution of a mission.

When attempting to develop a conceptual model of DP&E, the study differentiated between DP&E and other types of planning. DP&E is not *Contingency Planning*,[3] where the

[1] TTCP was formed in 1957 with the aim of fostering cooperation in the science and technology needed for conventional, i.e. non-atomic, defence.

[2] Dynamic – (adj) (of a process or system) characterised by constant change or activity. Oxford English Dictionary.

[3] 'Developing a plan for possible operations where the planning factors have been identified or can be assumed. This plan is produced in as much detail as possible, including the resources needed and deployment options, as a basis for subsequent planning.' AAP-6.

potential for change has been identified prior to the execution phase. DP&E is more akin to *Crisis Response Planning,*[4] due to the presence of similar conditions: reaction to current events; pressure of time sensitive situations; unforeseen events; and the absence of a prepared contingency plan. Although not enshrined in doctrine, various definitions of DP&E have evolved over the last few years. The following definition of *Continuous Adaptive Planning* captures the essence of DP&E:

*'Continuous Adaptive Planning is the systematic, on-demand creation and revision of executable plans, with up-to-date options as circumstances require.'* [7]

Having investigated other potential definitions of DP&E, the study confirmed the applicability of the working definition which AG1 had adopted upon forming:

*'The capability to continuously develop, analyse, select and flexibly execute a robust course of action based on a commander's evolving intent, situational awareness and capabilities'* [5]

### Doctrine

Although the UK has no recognised definition or concept of DP&E, UK doctrine contains various references to the requirement for agility and responsiveness in planning and execution. The requirement for simultaneous (as opposed to sequential) planning and execution processes is also recognised in the Effects Based Approach (EBA)[5] doctrine:

*"The 'Effects-Based Operations (Analytical) Concept' highlighted the aspiration to apply an effects-based methodology to the* ***concurrent, continuous analysis, planning, execution and assessment activities*** *required for modern operations…considerable benefit would be gained from the early adoption of an effects-based 'way of thinking' within the extant planning and execution processes".* [8]

The study established the presence of a firm doctrinal link between EBA and DP&E. The recognition of this link is important to the military community as it affords the DP&E concept a secure doctrinal foundation.

### Recent Operational Experience

Analysis of recent operations confirmed the extremely dynamic nature of today's battlespace, with the time available for planning being shortened significantly. This suggests that the current planning process needs adaptation if it is to continue to deliver operational advantage in the future. Recent operations have clearly identified the requirement to be able to plan quickly in response to a developing situation.

## 3. Conceptual Model

Having confirmed the military relevance of DP&E, AG1 developed a conceptual model to show the logic that military capability is enhanced by an ability to conduct DP&E.

Within a given Campaign Plan, the Armed Forces must achieve intended Effects (represented in Figure 1 by the triangle). Effects are the consequences of one or more activities, therefore to enhance Effects, DP&E must support the delivery of activities at optimum tempo (represented in Figure 1 by the horizontal rectangle).

A number of capabilities indirectly support the achievement of effect, including the ability to plan. These supporting capabilities must combine to deliver appropriate activities at optimum tempo.

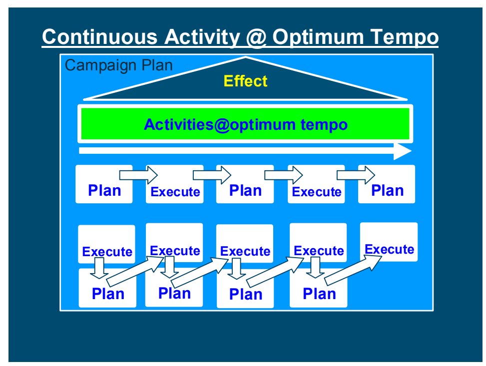


**Figure 1** – Current Planning & Execution Processes

In order to continually stay within our adversary's decision action cycle, and thereby maximise operational advantage, we must continually deliver activities at optimum tempo for the entire duration of a campaign.

### Current Planning and Execution Processes

How do we currently achieve continuous activities at optimum tempo in order to contribute to our desired effects? The study assumed that the planning and execution processes in question related to one of several missions within an extant Campaign Plan. The Endstate, Objectives and Effects of this Campaign Plan remain unchanged. This example could be viewed as the planning and execution processes supporting a tactical level mission within the operational level Campaign Plan.

Currently we plan and execute sequentially (the refine, monitor and assess stages have been omitted for simplicity).

[4] '…an activity based on current events and conducted in time sensitive situations. Essentially, unforeseen events for which there may be no specific Contingency Plan (although it could also be based upon an existing Joint Planning Guide, Joint Contingency Plan or Joint Operation Plan). Formerly known as Crisis Action Planning.' UK JWP 0-01.1.
[5] The Effects Based Approach (EBA) is also referred to as the Effects Based Approach to Operations (EBAO).

The planning directs the execution, which in turn informs subsequent planning. This is not strictly true, as we achieve continuous execution by planning and executing concurrently. Initial planning directs subsequent execution, which informs future planning, and so on.

**Change in Situation**

The current ability to execute continually remains effective as long as our understanding of the situation is accurate. However the situation will change (represented in Figure 2 by the flash), and unforeseen and challenging events will occur. Although the execution will continue, it is unlikely to be as effective as prior to the change in situation; the direction and coordination provided in the plan was based on situational understanding which no longer exists. The subsequent execution phase will be less effective until relevant and timely direction is provided, so this will not deliver continuous activities at optimum tempo. This, in turn, will degrade the desired effects we are trying to achieve.

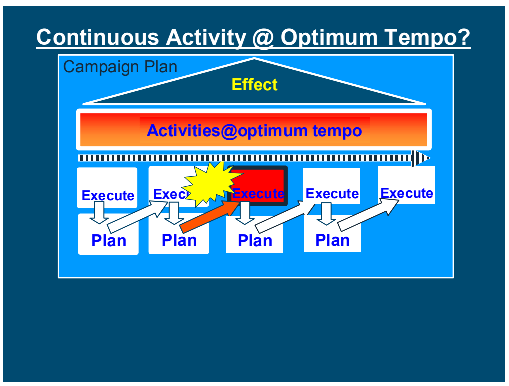


**Figure 2** – Impact of Change in Situation

**Impact of a change in situation**

The impact of a change in situation upon execution raises a number of issues:

(1) We currently alter execution following an unforeseen event, but could this be done better, and how can technology support this? Being able to dynamically plan and execute would allow this change in execution to be more effective.

(2) Although we currently alter execution, this is generally done using mission command, working within the superior commanders' intent, and the timely committal of reserves. Could this execution be enhanced with timely direction and coordination which has been based upon up to date situational awareness? How do we improve the information exchange between planners and executors?

(3) After a situation change, further planning (or replanning) currently takes time, which has the potential to reduce the effectiveness of execution until this new or amended plan is issued. How can we speed up this planning?

(4) Any recovery option will have an impact on future execution due to the coordination and synchronisation burden of deviating from an extant plan. The deduction from this is that maximum use of the 'old' plan will reduce this burden.

In sum, how do we provide the executors with the timely, effective direction and coordination, which will optimise actions and tempo? How can technology support the human in planning better and quicker, and facilitate coordination among geographically distributed entities?

**Continuous and Simultaneous Planning and Execution**

The ultimate solution has the planning and execution functions being undertaken continuously and simultaneously. Up to date world understanding informs the planning process, which in turn adjusts the plan as required. Under these circumstances, an unforeseen and challenging change in the situation has minimal impact on the effectiveness and continuity of execution. We can cope with a change in situation by being able to continue to conduct activities at optimum tempo, which in turn achieves our desired effects.

## 4. Example Application Domains

As part of AG1's remit to identify potential areas for collaboration, AG1 conducted a hard-copy questionnaire survey of relevant research projects in the TTCP participating nations [5]. The approach adopted by AG1 is depicted in Figure 3 below. To ensure a common understanding of terminology used to describe relevant project characteristics, three standard taxonomies were created: Military Capability (desired capability for conducting DP&E); Problem Types (well-defined basic problem classes); and Technologies (general categories of information technology that may be relevant to the goals of DP&E).

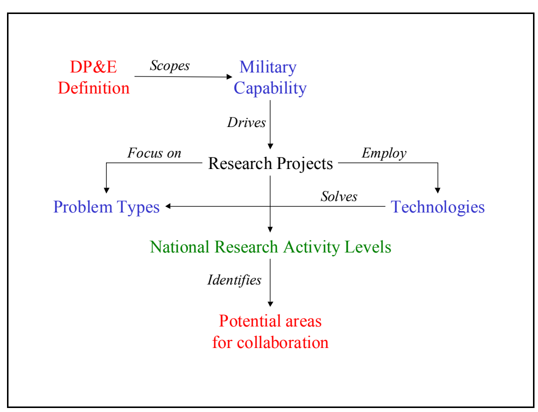


**Figure 3 -** AG1 Approach

The survey suggested that all four nations were interested in the following three broad application domains that have DP&E elements:

(1) Effects Based Operations (EBO) - a process for obtaining a desired strategic outcome or "effect" on the enemy, through the synergistic, multiplicative, and cumulative application of the full range of military and non-military capabilities at the tactical, operational, and strategic levels.

(2) Logistics for Agile Forces – a military force is considered to be agile if it is able to adapt and respond to changes in its environment in real-time. Agile formations make transitions quickly between changes in task, purpose, and direction, manoeuvring into and out of contact - without sapping operational momentum. A critical component is its logistics support element which needs to be able to dynamically plan and execute for agility.

(3) Wide Area Time Sensitive Airspace Coordination (WATSAC) - encompasses the constructive interaction of airborne platforms across the entire battlespace in minutes/seconds from a decision point. It includes aspects of airspace deconfliction as well as cooperative Intelligence, Surveillance and Reconnaissance (ISR) concepts.

For purposes of brevity, only the WATSAC concept and its relationship to DP&E challenges will be explored in further detail below.

**WATSAC and DP&E**

Airspace management is key to successful operation in both civilian and military environments. It consists of effective management and coordination of airspace use to support various users in enabling flight safety, defence, security, monitoring, and air traffic management. Emerging challenges require immediate attention: Increasing demands on airspace utilization; contention imposed by user and operational multiplicity and diversity; participation of heterogeneous organizations in various dynamic uncertain environments.

In the military context, airspace control refers to theatre airspace management. Its aim is to enhance combat effectiveness by promoting the safe, efficient, and flexible use of airspace. There are two types of control currently employed by the military; positive control and procedural control. Positive control is based on either visual or radar identification of the aircraft operating in the airspace. Although used heavily in the commercial aviation industry, positive control can be difficult to achieve in the military environment as aircraft may not want to 'squawk' their transponder information in a hostile environment. Procedural control dictates specific actions for a pilot to execute based on flight routes, altitudes and time, in order to maintain this separation when the aircraft is outside the limits of positive control.

Airspace deconfliction problems can be viewed as a subtask of cooperative ISR platform management. It currently consists of resolving conflicts and maintaining collision avoidance conditions by dynamically (re-) assigning airspace variables (time, space, user refusal, risk acceptance) while sharing flight activity information between key military organizations and aviation agencies. Conflict resolution is first achieved during initial planning, while real-time airspace deconfliction is handled dynamically for time-critical airspace requests.

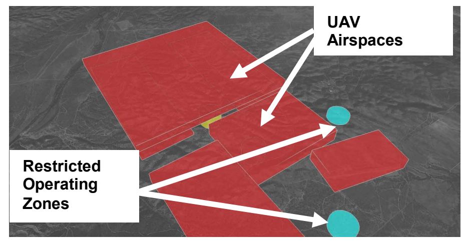


**Figure 4** – Current Planning for UAV Airspace

Current practice sees volumes of airspace deconflicted over time without necessarily deconflicting the objects within that airspace. It tends to be a fairly static, centralized process in planning, with limited distribution in execution. Figure 4 above illustrates a typical approach for UAV airspace planning. An entire block of airspace is reserved for the entire flight time of the UAV even though the UAV will only occupy a very small portion of that space at any given time. Planners at a higher echelon will default to this sort of planning in absence of coordination with lower echelons doing detailed route planning. This approach can limit the use of that airspace by other users in the area.

With a more detailed definition of the planned route for the UAVs (perhaps from lower echelons), the airspaces required may look more like Figure 5. Note the UAV airspaces have been defined by a series of waypoints based on their planned route and a volume around each waypoint. This is a more efficient use of airspace as contrasted with the previous approach and allows for much finer grained control over the battlespace.

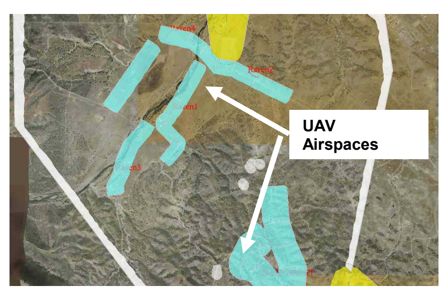


**Figure 5** – Future UAV Airspace Planning

With further improvements in near real time coordination and plan representation, it is conceivable to think about airspace deconfliction as occurring dynamically without any 'reserved' airspace done in preplanning. This is a concept that requires a great deal of coordination and situational awareness on the part of each aircraft (manned or unmanned) and weapon or the entities that control them. It also assumes a fairly rich communications environment and a high degree of information sharing. Similar assumptions can be found in many net-centric operations discussions. There are times, however, where this approach is not desirable for the military and a more procedural approach is preferred.

The prior discussion has concentrated on aspects of airspace deconfliction. A similar or perhaps greater level of coordination is likely required in a cooperative ISR mission where airborne entities are not just deconflicting operations, but actually trying to synergise their activities. Entities may share much more than the minimal information for deconfliction such as mission goals, negotiation of roles, and information obtained toward the accomplishment of those goals. AG1 also continues to pursue these broader aspects of WATSAC.

## 5. ISSUES & CHALLENGES

One of the key characteristics of military planning is the decoupling of planning and execution activities. Specialist planning staff initiate plans, before handing them off for refinement by different staff. After receiving the refined plan, yet another staff will be responsible for its implementation, including provision of timely direction to subordinates as the situation develops. As each staff branch hands off a plan, it starts planning for the subsequent phase of the operation. At various stages of maturity, this plan will be disseminated to subordinate HQs and Units to allow concurrent activity. Each subordinate will conduct further mission specific planning before issuing direction to its own subordinates, as shown in Figure 6.

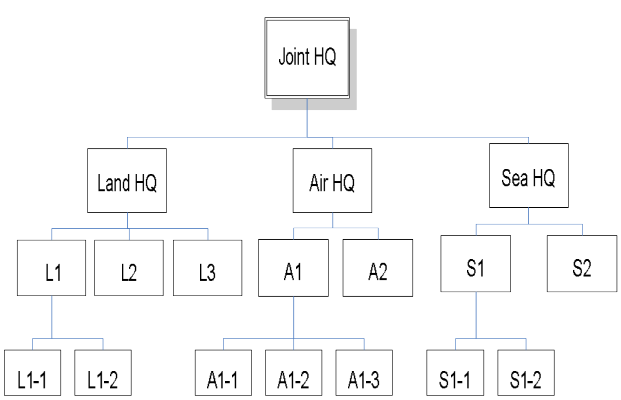


**Figure 6 –** Example of Command Hierarchy

The plans are normally communicated in printed text and diagrams.[6] At each level of command, as one progress down the tree from the root node (Joint HQ), the size of the overall plan documentation increases as each subordinate appends details of their role in the higher level plan. Thus, the plan received by L1-1 leaf node would be a plan from parent unit (L1) but including detail and context from the Land HQ and Joint HQ plans.

The following are some of the issues that need to be addressed in order to realise DP&E:

(1) Subjectivity in Plans: Plans are generated and interpreted by humans, each of whom is shaped by their knowledge, training, experience and individual human characteristics.[7]

(2) Timeliness: Static representations of plans are not easy to update and tend to take a lot more time than is normally available.

(3) Replanning: At each level of command, plans are often amended in light of changing situation before and during execution. Static representations of plans do not contain any of the reasoning, logic and interdependencies which remain of value when considering how to alter a plan.[8]

(4) Coordination: Plans issued by a particular HQ contain the coordination to synchronise the activities of subordinates. Should a subordinate deviate from the given plan due to a changing situation then this is typically done in consultation with peers (ie. L1 would talk to L2, L3 and if required A1).

**Plan Representation**

In order to dynamically change a plan effectively and efficiently, an executor must understand the logic and semantic meaning of the original plan, and must be able to rapidly manage the interdependencies affected by the change. Currently this requires access to the original human planner, 'on the fly' planning, and rapid consultation. A solution would be to have a dynamic representation of a plan with some automated (planning) support. Such a representation would help address the issues identified above.

In the last decade, there has been some effort in developing representation formalisms for planning, for example, <I-N-O-V-A> [9], SPAR [10], XPDDL [11]. From the discussions at a recent Representation Workshop,[9] it was apparent that are still outstanding challenges to be

[6] Quick Battle Orders or 'hasty' plans may be communicated by voice.
[7] Examples of human characteristics are confidence, morale, aggression, attitude to risk, etc.
[8] Dwight D Eisenhower: "Plans are nothing; planning is everything".
[9] Organised by AG1 and held 11-12 October 2006 at AFRL, Rome, New York.

addressed before dynamic representation can be usefully employed in military planning. Particularly, the representation needs to be able to facilitate exchanges between human-human, human–machine and machine–machine. Critical to this is the ability to capture and represent context and assumptions. Another military specific problem is the linking of various components of the plans, and from high level (Strategic-Operational) to lower levels (Operational-Tactical).

### Coordination

For successful outcome, military operations have to be very well coordinated. However, in light of a changing situation, there is an enduring need to coordinate actions among peers. This is typically done manually, though some rudimentary automation is beginning to be fielded on the battlefield. These tools allow for distributed workflow-like coordination and awareness while the work done 'behind the scenes' of these tools is a primarily manual process. With digital representation of plans, there is the possibility of some automated help in coordination. The issue (4) identified above is central to DP&E.

The discussion during the AG1 Coordination Workshop[10] revealed several key areas that require additional research: modelling adversarial behaviours to support more realistic coordination strategies, richer plan representation and trust in agent-based systems; truly open environments so that agents are not built within stove piped architectures, but are built to be interoperable and can leverage existing capabilities instead of duplicating efforts in the coordination process; the ability to coordinate in uncertain environments [12].

## 6. Conclusions

AG1's investigations into DP&E have highlighted that the concept is not very well understood. Despite this, evidence is emerging that future military capability would be enhanced by an ability to conduct DP&E; and WATSAC is an example of this. A key finding during AG1's workshops was that richer plan representations are needed to enable autonomous systems and large scale groups to coordinate effectively. As was highlighted previously in this paper, as one progresses to lower echelons of planning, the level of detail required in the plan continues to increase. Plan representations must be able to accommodate these details and furthermore, these details need to be common (or at least interoperable) among autonomous entities. The ability to interconnect formal coordination mechanisms and protocols within the formal plan representation was viewed as a novel capability and should be investigated.

[10] Held 12-13 October 2006 at AFRL.

## Acknowledgement

The work presented in this paper is a collaborative effort of the AG1 Team, which includes: Lin Zhang and Don Gossink, DSTO, Australia; Jean Berger, DRDC, Canada; Ranjeev Mittu, NRL, US; Gary Yerace and Alison Clark, ARL, US; James Lawton, AFRL, US.

## References

[1] Alberts, D., *Information Age Transformation: Getting to a 21st Century Military*, CCRP, 1996.

[2] Van Creveld, M., *Command in War*, Harvard University Press, Cambridge, MA., 1985.

[3] UK Joint High Level Operational Concept.

[4] Berger, J. et al., TTCP/C3I/AG1 Position Paper. In *Proceedings of the Third Knowledge Systems for Coalition Operations*, pages 119 – 123, 2004.

[5] TTCP C3I AG1 DP&E Projects Survey Report, Jan 2005.

[6] Aitken, A., The Military perspective of DP&E. In *AG1 Third Annual Report*, May 2006.

[7] Hill, Surdu & Pooch, *A Methodology to support Anticipatory Planning* , Texas A & M University.

[8] Joint Doctrine Note 1/05 - *The UK Military Effects-Based Approach*, Sep 2005. Developed and refined by Joint Doctrine Note 7/06 – *Incorporating and Extending the UK Military Effects-Based Approach*, Sep 2006.

[9] Tate, A., Representing Plans as a Set of Constraints – the <I-N-O-V-A> model. In the proceedings of the 3rd International Conference on Artificial Intelligence Planning Systems, May 1996.

10] Tate, A., Roots of SPAR – Shared Planning and Activity Representation, in *The Knowledge Engineering Review*, Vol 13(1), pp 121-128, March 1998.

[11] XPDDL – www.cis.strath.ac.uk/~jg/XPDDL.

[12] Mittu, R., and Berger, J., *Coordination Workshop Report*, TTCP AG1 Internal Publication, October 2006.

# Distributed Deliberative Planning with Partial Observability: Heuristic Approaches

D. Perugini[1], D. Jarvis[2], S. Reschke[1], D. Gossink[1]
[1]Decision Automation Group, Command and Control Division, DSTO, Australia
[2] School of Computing Sciences, Central Queensland University, Australia

***Abstract*- Military operations typically involve cooperation of various military, government and commercial organizations from various nations. In order to coordinate these autonomous organizations, a social mechanism is required that facilitates deliberative planning and task allocation in decentralized, open and dynamic environments, and enables agreements via a legal contracting process. In this paper, we present (a component of) such a mechanism, called the Legal Agreement Protocol (LAP). Agents that plan using LAP must plan with partial observability – that is the customer is only aware of proposals (capabilities) that suppliers choose to send. This makes it difficult for the customer to determine the (minimum/average) expected cost of any unallocated sub-tasks in its search. In this paper, we present and compare various heuristics that allow the customer to dynamically determine the expected cost for sub-tasks as proposals are received during planning. We show that different heuristics have tradeoffs in terms of quality of solution and search effort (efficiency of search and quantity of communication). The number of distributed agents involved in planning also influences the effort required to search. More agents increase communication, but provide more information (observability) about agents' capabilities to be utilized by the heuristics.**

## I. Introduction

Military operations typically involve cooperation of various military, government and commercial organizations from various nations. Whether it is a military conflict or a peaceful operation such as emergency response, many *decentralised* autonomous organizations need to interact and coordinate their actions to achieve operational goals. The autonomous organizations are typically self-interested, and thus make their own decisions which benefit themselves, and may not be willing to be *controlled* (or tasked) by other organizations. Additionally, organisations may contain proprietary or classified information that describe their capabilities, such as schedules, which they may be reluctant to release to a central planner. In any case, information to be released may be extensive, difficult to describe, and too complex for others to comprehend. As a result, central or hierarchical planning and task allocation approaches are not well suited to a predominantly decentralized environment.

The military operations environment is also *dynamic* and *open*. An organization's capabilities may change throughout the operation, or even throughout the planning process. Organizations with indeterminate capabilities may come and go at any time, even during the planning process. Finally, in order for organizations (businesses) to allocate tasks among each other, they may need to follow contract law procedures to enable the formation of *legal agreements* for services/capabilities. Even if the domain is cooperative and contracts are not necessary, following contract law procedures enables individuals and organizations to formally establish commitments and agreements. For example, if one party receives an offer, it knows that the other party has formally committed to the agreement as opposed to just informing of the willingness to participate (a proposal).

The Defence Science and Technology Organisation (DSTO) has developed an agent (or organizational) interaction protocol called the Legal Agreement Protocol (LAP). It is an extension of the Provisional Agreement Protocol [1-4], which is based on the Contract Net Protocol (CNP) [5]. LAP facilitates deliberative planning and task allocation in decentralised, dynamic and open environments. LAP also conforms to the principles of Contract Law. LAP enables task allocation and agreements between organizations (e.g. in coalition command and control) without the need for rigid hierarchies and delegation among organizations. The protocol enables an iterative interaction process in which *customer* agents extract, match and negotiate capabilities from other *supplier* agents, and assemble capabilities via deliberation (e.g. distributed A* search, see later), in order to achieve the customer's tasks. Agents do not need to release all information regarding their capabilities, and can release only those capabilities that they are willing to perform. LAP also allows agents to adapt to a changing and open environment (during and after the planning process) via updating, withdrawing and backtracking mechanisms.

LAP consists of a messaging component and a reasoning component. The reasoning component of LAP provides a generic capability for control of the messaging component. However, the actual reasoning strategies that are employed will be largely application specific and must therefore be provided by the application developer. It is the provision of reasoning strategies that employ LAP and not LAP itself that will be the focus of this paper. Therefore, the complete LAP specification will not be presented. We present only those elements that facilitate deliberative planning, and thus demonstrate a core component of LAP.

In using LAP, customer agents which need to deliberate have partial observability – that is, a customer is only aware of the *proposals* (capabilities) that suppliers choose to send. This makes it difficult for customer agents to perform distributed search as effective as centralized approaches that have all proposals available for processing. In this paper, we present and compare some heuristics to address one of various search issues with partial observability, namely expected cost evaluation. Expected cost evaluation is used in informed searches to provide an estimate of the cost of achieving any unallocated/unachieved tasks.

In the next section we discuss the LAP deliberative planning process and the various heuristics that we have devised. In section III we present experimental results, which we use the set partitioning datasets from [6]. We show that different heuristics have tradeoffs in terms of quality of solution and search effort (efficiency of search and quantity of communication). The number distributed agents involved in planning also influences the effort required to search. More agents increase communication, but provide more information (observability) about agents' capabilities to be utilized by the heuristics. We conclude in section IV.

## II. LAP Deliberative Planning

### A. The Protocol

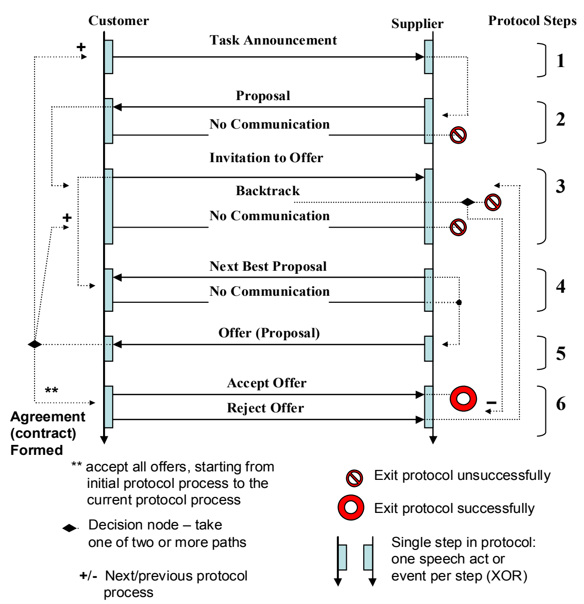


Figure 1. Representation of the deliberative component of LAP.

A diagrammatic representation of the deliberative component of LAP, using the Protocol Flow Diagram representation [1], is shown in Figure 1. Updating, withdrawing, negotiating and contract termination/discharging components are not shown or used. The two vertical lines represent the customer and supplier that are cooperating. There are 6 steps in the protocol, represented by the pairs of boxes along the vertical lines. The box pairs represent an XOR relationship, and therefore only one message (speech act) may be communicated, or one event may occur, at each step. Therefore, at step 3, the customer may either send an **invitation to offer** or **backtrack**, but not both. The dotted lines emanating from each message or event are control lines, which indicates the next step of the protocol to proceed to if that speech act or event occurs. Therefore, after a **task announcement** at step one, the protocol proceeds to step 2. The side (i.e. customer or supplier) that the control line appears is irrelevant. Note that the customer may be at different steps in the protocol between different suppliers. A diamond along a control line represents a decision node, and thus the agent can take one of two or more paths, depending on the situation and its reasoning.

The last two elements of the Protocol Flow Diagram are the "+" and "-" symbols, which represent LAP proceeding to the following (new or existing) or previous protocol process, respectively. A single protocol process is the execution of the protocol for a single task, and thus contains its associated state variables (such as proposals sent/received) and the position in the protocol. In LAP (unlike CNP), a proposal may be submitted that *partially* achieves the announced task. As will be discussed below, if a proposal is selected which does not completely achieve the task in a current protocol process, a new protocol process is spawned in order to achieve the remaining task that the proposal did not achieve. In backtracking, a customer may proceed to the previous protocol process to select another proposal (option/path). Note that each protocol process corresponds to a node and its branches discussed in the next section. The previous or following protocol process corresponds to a nodes parent or child node (and their associated branches), respectively. We denote an announced task for the current, previous and following protocol process as $T$, $T^-$, $T^+$, respectively.

We now describe LAP (its steps) in detail:

**Step 1:** LAP is initiated at step 1 with a **task announcement** $T = T^{init}$. LAP also arrives at step 1 from step 6 of the previous protocol process for task $T^-$ if the offer (proposal) did not fully achieve $T^-$. The task announcement is the tuple $<T, f(p), \varphi, d, s>$ where: $T$ is a set of tasks; $f(p)$ is the proposal evaluation function which customer will evaluate the proposals $p$ (allowing suppliers to submit their best proposal); $\varphi$ is the set of offers currently accepted in previous protocol processes that are from the supplier whom is being sent the task announcement; $d$ is the *proposing deadline* which is the earliest time that the customer may select a proposal for its set of tasks $T$; and $s$ is the task shutdown deadline which is the time that the task is no longer available/valid. The protocol proceeds to step 2.

**Step 2:** At step 2, each supplier may submit their *single best **proposal*** that they believe can fully of partially achieve $T$, based on $f$, which does not conflict with other offers/proposals $\varphi$ that were submitted by the supplier. Note that suppliers are not committed to proposals. Proposals are used to inform customers of potential capability/services. After a proposal is submitted, control proceeds to step 3. Suppliers need not submit any proposals – **no communication**, and thus exit the protocol process unsuccessfully (without a contract).

**Step 3:** After the proposing deadline $d$, the best proposal for $T$, based on $f(p)$, is given an **invitation to offer**. This essentially asks the supplier to commit to the proposal (i.e. offer it formally). The protocol proceeds to step 4. Suppliers that do not have their proposals selected (invited to offer) receive **no communication** and may exit the protocol after task shutdown deadline $s$. If the customer finds the path in its search to be unsuitable, for example, because it appears that there are better options on different paths or no proposals were submitted (and hence, assume that the current path leads to an infeasible solution), then the customer may **backtrack**. If this occurs in the initial set of tasks $T^{init}$, then no solutions exists and thus exit the protocol. Otherwise, for current task $T$, take the control path to step 6 of the previous protocol process associated with task $T^{-}$.

**Step 4:** At step 4, the supplier may submit its *next best proposal* (based on $f(p)$) if it has one (suppliers may comprise many proposals), in addition to submitting an offer in the next step. This allows the customer to determine if another proposal option by the supplier could be better than the selected proposal, facilitating deliberative planning (e.g. distributed A*). The supplier may either submit an updated **proposal**, or **no communication** (indicated by only an offer in the next step). The protocol proceeds to step 5.

**Step 5:** The supplier submits the proposal as an **offer**, committing the supplier to the proposal. If the task $T$ is completely achieved by the offer, then proceed to step 6. Otherwise, there are two options. **(i)** If it is the first time that the offer has been submitted, take the control path to step 1 of a newly spawned protocol process, to achieve the remaining task $T^{+}$ that was not achieved by the offer. **(ii)** If the offer has been previously submitted, and thus there already exists proceeding protocol processes as a result of this offer (LAP is back at this protocol process after backtracking), then the customer may proceed to step 3 of the following protocol process in order to select a proposal for it.

**Step 6:** If arrived from step 6, the customer has found a solution (plan) to achieve its initial task $T^{init}$. The customer may then accept all the offers by submitting **accept offers** messages, starting from the offer associated with the initial task $T^{init}$ (initial protocol process) to the offer associated with the current task $T$ (current protocol process). The customer has formed $N$ contracts/agreements with up to $N$ suppliers in order to achieve $T^{init}$. Thus both the customer and supplier are committed to the contract. If arrived from backtracking at step 3 of the following protocol process (for $T^{+}$), the customer can *backtrack* by **rejecting** the offer for $T$, and control proceeds to step 3 allowing the customer to either select a new proposal for $T$, or backtrack further. If arrived from step 6 *and* the customer is not satisfied with the final plan, then the customer does not have to accept the final plan. Rather, the customer can **reject** the offer for $T$, and go to step 3 to either select another proposal for $T$ or backtrack further.

Note that in this paper we ignore any communication problems and agent failures – our focus is on the planning process and its heuristics.

*B. LAP Distributed Deliberative (A*) Search & Example*

LAP facilitates a distributed deliberative search for the customer. Thus LAP can facilitate distributed A* (or best-first), branch-and-bound and depth-first searches. We use A* search [7] in this paper. We will use Figure 2 and a set partitioning example [6] (used in our experiments in the next section) to describe LAP's A* deliberative search. The initial task announcement $T^{init}$ = {1, 2, 3, 4, 5} is the root node. In a military coalition setting, integers 1 to 5 may represent sub-tasks that need to be achieved by coalition elements (suppliers) in order to achieve the military operation. The proposal evaluation function sent with the task could be $f(p)$ = *[minimize, price, k = 0, e = <2, 1, 6, 4, 2>, a = 2]*, where *minimize* states that the customer wants to minimize the solution, based on the *price* to achieve the task (could use time or distance, or a combination). We will discuss $a$ later. The value $k = 0$ is the sum of the costs of proposals that have been offered/selected in previous protocol processes (cost of the path in the search tree up to this task), which for the root node is zero. The ordered set $e$ = *<2, 10, 6, 4, 8>* is the expected price/cost for task 1 to 5, respectively, that the suppliers must use in order to calculate the heuristic cost for their proposals, which is $f = g + h$ (as used in A* search). $g$ is the current path cost, which includes a supplier's proposal, so $g = k + j$, where $j$ is the cost of the suppliers proposal. $h$ is the expected cost of any sub-tasks that the proposal does not achieve. In order to guarantee optimality, the expected cost must used be less than the actual optimal cost (an under-estimate).

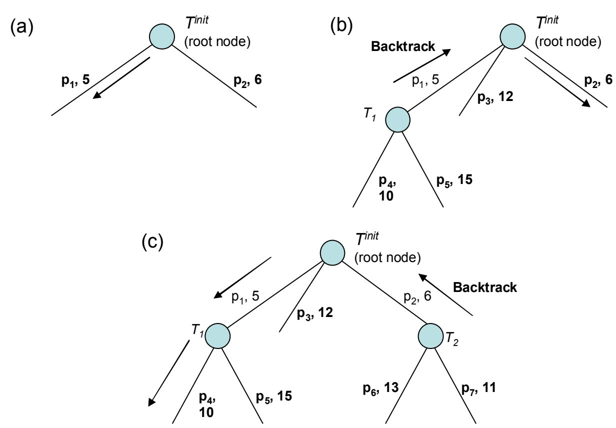


Figure 2. Search tree facilitated by LAP for the customer.

An example, say agent A has two proposals {1, 3, 4} and {3, 5} at cost 3 and 6, respectively (i.e. agent A can achieve tasks 1, 3 and 4 at a cost of 3, or achieve tasks 3 and 5 at a cost of 6). Agent A evaluates it proposals using $f(p)$. For proposal {1, 3, 4}, $g = 0 + 3$, unachieved sub-tasks are $T^{init}$/{1, 3, 4} = {2, 5}, and from the set $e$, the expected cost to achieve 2 is 10 and for 5 is 8, resulting in $h = 10 + 8$. The cost of the proposal, which is the cost $f$ that customer will also evaluate to, is $f = 21$. Similarly, for proposal {2, 5}, $g = 0 + 6$, $T^{init}$-{2, 5} = {1, 3, 4}, so $h = 2 + 6 + 4$, so we have $f = 18$. Therefore, agent A will submit it proposal {2, 5}, which will be preferred by the customer. We assume that competition between suppliers, and their desire to have their proposal allocated will motivate them submit their best proposal based on *f(p)*.

Proposals received from suppliers for a task are the branches in the search tree. In the example in Figure 2 (a), the customer receives proposals $p_1$ and $p_2$ for $T^{init}$. The customer may search the most appropriate path by selecting the leaf branch (proposal) in the search tree with the lowest *f* value (*f* value is shown as 5 for $p_1$ and 6 for $p_2$ Figure 2). Therefore, the customer invites to offer the proposal $p_1$ = {2, 3, 4}. In Figure 2 (b), the supplier submits an offer for $p_1$, as well as an updated (next best) proposal $p_3$ for $T^{init}$. Since the offer $p_1$ does not fully achieve $T^{init}$, the remaining task $T_1$ that the proposal does not achieve, $T_1 = T^{init} - \{2, 3, 4\} = \{1, 5\}$, is announced, which is a child node of branch $p_1$. In the figure, proposals $p_4$ and $p_5$ are submitted for $T_1$. The customer discovers that the path $p_2$ for $T^{init}$ (the root node) looks more promising than the proposals $p_4$ and $p_5$ at the current node (current protocol process). The customer backtracks by moving to the root node (proceeds to the previous protocol process, for $T^{init}$), rejecting $p_1$, and then selects (invite to offer) proposal $p_2$. Similarly, in Figure 2 (c), an offer is submitted for $p_2$ (no updated proposal is sent with the offer in this case), a new task $T_2 = T^{init} - p_2$ is announced, proposals $p_6$ and $p_7$ are received, and the customer backtracks in order to follow the lowest cost path $p_1$ and $p_4$. The search continues until a plan of proposals/offers is found, and the task allocated to the respective suppliers (by accepting the offers).

In our example, the order in which proposals are selected in the search tree is irrelevant. Therefore, path $p_1$ and $p_2$ is the same as path $p_2$ and $p_1$. In order to reduce the search space, by searching combinations rather than permutations, we use a strategy which *we* term called *anchoring* (discussed in [8]). At each node (task announcement), the customer selects a sub-task which is "anchored". Suppliers are only able to submit proposals which can achieve that sub-task. In the *f(p)* presented above, *a = 2* indicates that the customer wants sub-task 3 (in task {1, 2, **3**, 4, 5}) to be anchored, requiring suppliers to only consider and submit proposals that contain 3. In order to reduce the search space further, it is best anchor on sub-tasks which are constrained (do not occur in many proposals) first, at the top of the search tree. This reduces the branching factor at the top of the search tree.

### C. *Issues Due to Partial Observability*

The set partitioning problem comprises a set of tasks $T = \{1, 2, \ldots, m\}$, and a collection of package proposals $B = \{B_1, B_2, \ldots, B_n\}$ (from [6, 8]). A package proposal is a tuple $P_j = \langle p_j, c_j \rangle$, where $p_j \subseteq T$ is a set of achieving capabilities/services at cost $c_j$. The aim is to:

$$\min \sum_{j=1}^{n} c_j x_j \quad \text{subject to} \quad \sum_{j|i \in p_j} x_j = 1 \quad \forall i \in \{1 \ldots m\} \quad \text{and} \quad x_j \in \{0,1\}$$

When solving the set partitioning problem in a centralized fashion, where the agent contains all the proposals, the minimum expected cost for each task and the anchoring ordering can be obtained by scanning all the available proposals. The anchoring ordering can be determined by counting the number of proposals that each sub-task appears in. Sub-tasks that appear less often are placed higher in the order. The minimum expected cost for each sub-task is determined naïvely by $\min_{j|i \in p_j} \frac{c_j}{|p_j|}$.

In a decentralized environment with LAP, the customer agent does not have access to all the proposals. The customer may potentially start planning with no proposals and no information about suppliers' capabilities and their costs. Therefore, current centralized approaches to expected cost and anchoring ordering formulation are not appropriate.

In this paper we focus on heuristics to determine the expected cost. Therefore, in our experiments in the next section, we processed all the proposals to find the anchoring order in order to improve the efficiency of the search. Heuristics to determine the anchoring order dynamically in the presence of partial observability is future work.

### D. *Expected Cost Heuristics*

In order to address the partial observability problem with expected cost, we allow the customer to dynamically determine these values during planning as proposals are received. We assume that the customer knows nothing about suppliers' capabilities when planning commences. By receiving proposals, the customer gains some observabilty, and thus can potentially obtain better estimates as LAP planning proceeds.

The customer may commence with an expected cost of zero for each sub-task, which is an underestimate and thus guarantees optimality, and refine the values as planning proceeds. An issue with expected cost determination is that the customer may overestimate, and thus may not receive an optimal solution. Some of our proposed heuristics aim to minimize the chance of over-estimation. For each sub-task in a received proposal, we define the sub-task cost as $c_j/|p_j|$. Proposed heuristics for expected cost are:

- *Minimum cost:* Store the minimum sub-task cost seen for each sub-task, and use this for the expected cost.
- *Alpha factor on difference, limited:* The aim is to slowly increase the expected cost in order to prevent exceeding the actual minimum cost. $v$ is the current estimate for a sub-task, and $s$ is a new observed sub-task cost. If $s < v$, then set $v = s$ and use the minimum cost heuristics (above) for the rest of the planning process. Otherwise, $\Delta = v - s$, and the new estimate $v = v + \alpha \cdot \Delta$.
- *Average over all sub-tasks:* For each sub-task, store all observed sub-task costs in a list. The expected cost for each sub-task is the average cost.
- *Average of current average:* The "average" cost $v$ for each sub-task is stored. When a new sub-task cost $s$ is observed, the new average cost is $(v + s)/2$.

The aim of the first two expected cost heuristics is to find a suitable expected cost estimate while still trying to guarantee optimality. The last two heuristics aim to satisfice the solution.

## III. EXPERIMENTATION

In this section we present experimental results using our heuristics on the set partitioning datasets from [6]. Only the small problems from the datasets were used with our distributed planning approach. Centralized approaches can solve much larger problems in a reasonable amount of time because they do not have the added communication time during each step of the search. For very large problems, a greedy (depth-first) LAP distributed search is more appropriate. Centralized search approaches may make quick bad decisions at each node in order to allow an agent to search many paths very quickly. Communication time at each node, and the cost of communication at each node, restricts this with a LAP approach. Therefore, with a LAP approach, the aim is to make the best possible decision at each node in order to reduce the number of nodes traversed (increase the efficiency of the search).

Solutions (plans) obtained are evaluated on three factors: quality of solution obtained; the number of nodes traversed; and the number of branches received. The number of nodes traversed is indicative of the efficiency of the search (less nodes, more efficient), and both the number of nodes traversed and the number of branches received is indicative of the communication required.

TABLE 1
EXPERIMENTAL RESULTS FOR EXPECTED COST HEURISTICS.
*$Average \pm Standard\ Deviation.$*

| Heuristic | Quality (% *above* optimal) | Nodes Normalized | Branches Normalized | # Opt |
|---|---|---|---|---|
| **Minimum Cost** | 2.7 ± 6.9* | 33.0 ± 26.7 | 28.7 ± 30.7 | 57 |
| **Alpha Factor (α=0.2)** | 0.07 ± 0.52 | 77.5 ± 22.5 | 62.2 ± 30.0 | 85 |
| **Alpha Factor (α=0.4)** | 0.36 ± 1.4 | 58.5 ± 23.8 | 46.3 ± 28.5 | 76 |
| **Alpha Factor (α=0.6)** | 0.51 ± 1.7 | 50.1 ± 25.9 | 41.1 ± 29.3 | 69 |
| **Alpha Factor (α=0.8)** | 0.71 ± 1.9 | 44.2 ± 25.0 | 36.6 ± 29.2 | 63 |
| **Average Cost** | 22.5 ± 20.5 | 1.0 ± 2.6 | 0.54 ± 0.83 | 5 |
| **Average of Average** | 30.9 ± 27.3 | 0.65 ± 1.8 | 0.41 ± 0.69 | 4 |

In our experiments, we ran a total of 90 scenarios: 18 datasets; 1, 2, 5, 50 and 100 suppliers; and the four expected cost heuristics, using α values of 0.2, 0.4, 0.6 and 0.8 for the alpha factoring heuristic. Results are presented in TABLE 1 and Figure 3 and Figure 4. TABLE 1 presents the average and standard deviation for the quality (percent that the solution cost is above optimal, and hence a larger value implies less quality) of solution obtained, and number of nodes and branches traversed/received *normalized* (between 0 and 100), in all the scenarios for each heuristic. In order to normalize the number of nodes and branches, we used $((n\text{-}L)/(H\text{-}L))\times 100$, where $n$ is the number of nodes/branches, and $L$ and $H$ are the lowest and highest number of nodes/branches observed for the particular dataset across all heuristics, respectively. The number of optimal solutions (labeled "# Opt" in the table) obtained for each heuristic is also presented in the table.

From the results it can be seen that the two average cost heuristics find a solution with an order of magnitude less nodes and branches than the minimum cost and alpha factoring heuristics, but the quality of solution is considerably worse. Therefore, the two average cost heuristics will be preferred when minimal effort (time and communication) is a priority. The other heuristics are preferred when quality of solution is a priority.

The average cost heuristics will tend to overestimate the optimal solution. This enables the search to reach many sub-optimal (average) solutions, which can be found quickly. The average cost heuristic gave a 27% improvement in the solution quality than average of current average heuristic, but at a greater effort (time and communication) with 35% more nodes traversed and 24% more branches received.

With the alpha factor heuristic, the lower the value of α, the better the quality of solution (and number of optimal solutions). A low α causes the expected cost to increase slowly in reaching the actual minimum expected cost, minimizing the chances of overshooting it, and thus maximizing the chances of obtaining an optimal solution. A large α may result in the expected cost to climb too quickly and overshoot the actual expected minimum cost for some parts of the search. Using a low α results in greater effort to find a solution because for part of the search, the expected cost will be grossly under-estimated, and thus causes an inefficient and uninformed search.

The quality of minimum cost heuristic was 280% worse that the worst alpha factor heuristic (α = 0.8), but the reduction in effort to find a solution was only 25% for nodes traversed and 22% for branches received. Therefore it is worthwhile putting in a little extra effort and use the alpha factor heuristic in order to obtain a much better solution. The solution for the minimum cost heuristic is considerably worse because the expected cost is initially (at the top of search tree, and through most of the initial search) set to the minimum sub-task cost of the received proposals, and this cost is likely to be an overestimate. The expected cost will approach the minimum cost as more proposals are received, potentially a considerable way through the search.

Figure 3 and Figure 4 shows the number of branches and nodes (normalized), respectively, for each heuristic as the number of suppliers increase. Values for average cost and average of current average are scaled up by a factor of 20 so that the trend is visible compared to the scale of the other heuristics. The number of branches received increases with the number of suppliers (for all heuristics) because there are more suppliers to submit proposals for each task announcement (node). The number of branches from 50 to 100 suppliers does not increase significantly for two reasons. Firstly, the maximum number of proposals that can be submitted for each task announcement (the branching factor) may be approximately 50 or less, and thus more suppliers does not result in more proposals submitted for each task. Secondly, as will be discussed next, more suppliers results in a more efficient search for most heuristics.

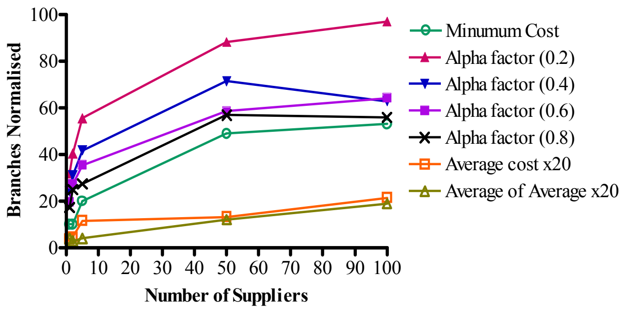


Figure 3. Number of branches for varying number of suppliers.

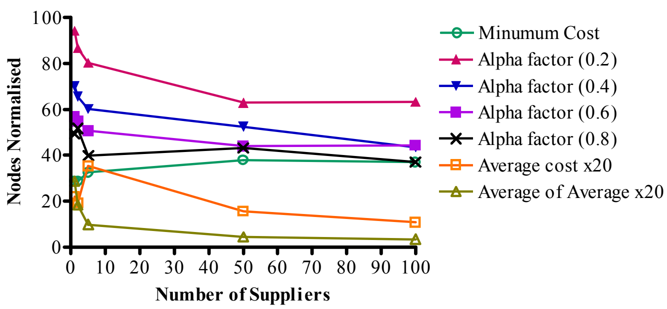


Figure 4. Number of nodes with varying number of suppliers.

The trend for the number of nodes traversed in the search decreases (more efficient search) as suppliers increase for all heuristics but the minimum cost heuristic. The decreasing trend with the alpha factor heuristic is due to the expected cost for each sub-task increasing quickly, and closer to, the actual minimum cost early in the search, as a result of many proposals being submitted for each task announcement when there are many suppliers. The decreasing trend with the average cost heuristic is due to a stable, accurate and likely larger average expected cost (than an average obtained from one/few lowest cost proposals with few suppliers) at the start of the search, obtained from many submitted proposals for each task announcement. Similarly, with the average of the current average heuristic, although the new expected cost for each sub-task is highly dependent only on the last few observed proposals and not previously observed proposals, receiving more proposals increases the chances that the last observed proposals have a higher expected cost than receiving only one/few lowest cost proposals from few suppliers.

The increasing trend with the minimum cost heuristic is due to the increased likelihood that the expected cost will be lower with many proposals, since the expected cost is the minimum of all sub-task expected costs observed so far. A lower expected cost reduces the number of potential solutions that can be reached (only better solutions can be found).

Our results did not show any significant trend with quality of solution as the number of suppliers increased.

## IV. Conclusion

In this paper we presented four heuristics to enable agents to plan and allocate tasks using LAP in the presence of partial observability. The heuristics enabled agents to dynamically determine an expected cost for each sub-task as proposals are received by suppliers. We show that different heuristics have tradeoffs in terms of quality of solution and search effort (efficiency of search and quantity of communication). The two average cost heuristics allowed more efficient (quick) search than the other minimum cost heuristics, but at the expense of the quality of solution. Although the alpha factor heuristic ($\alpha$ = 0.8) required more effort than the minimum cost heuristic, the increase in the quality of solution was significant.

The number suppliers involved in planning influences the effort required to search. The number of branches received in the search increases with the number of suppliers due to more proposals being received for each task announcement. Increasing the number of suppliers decreased the number of nodes traversed for all but the minimum cost heuristic due primarily to larger expected cost estimates, resulting in more efficient search. The expected cost for the minimum cost heuristic decreased with more suppliers.

Investigating heuristics for anchor ordering in the presence of partial observability is future work.

## Acknowledgment


We would like to thank Peter Smet, Dale Lambert and Jason Scholz for their assistance.


## References


[1] D. Perugini, "Agents for Logistics: A Provisional Agreement Approach," in *Department of Computer Science and Software Engineering*, vol. PhD. Melbourne: The University of Melbourne, 2006.

[2] D. Perugini, D. Lambert, L. Sterling, and A. Pearce, "A Distributed Agent Approach to Global Transportation Scheduling," presented at Proceedings of IEEE/WIC International Conference on Intelligent Agent Technology (IAT 2003), Halifax, Canada, 2003.

[3] D. Perugini, D. Lambert, L. Sterling, and A. Pearce, "Agent-Based Transport Scheduling in Military Logistics," presented at Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'04), New York, U.S.A., 2004.

[4] D. Perugini, D. Lambert, L. Sterling, and A. Pearce, "From Single Static to Multiple Dynamic Combinatorial Auctions," presented at Proceedings of IEEE/WIC International Conference on Intelligent Agent Technology (IAT 2005), Compiegne University of Technology, France, 2005.

[5] R. G. Smith, "The contract net protocol: high level communication and control in a distributed problem solver," *IEEE Transactions on Computers*, vol. C-29, pp. 1104-13, 1980.

[6] P. Chu and J. Beasley, "Constraint Handling in Gentetic Algorithms: The Set Partitioning Problem," *Journal of Heuristics*, vol. 11, pp. 323-357, 1998.

[7] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*: Prentice-Hall International, 1995.

[8] P. Cramton, Y. Shoham, and R. Steinberg, "Introduction to Combinatorial Auctions," in *Combinatorial Auctions (forthcoming)*: MIT Press, 2006.

# Assessing the transmission of Commanders Intent

Ian R Whitworth  Geoffrey N Hone
Cranfield University at the Defence Academy of the United Kingdom, Shrivenham, SN6 8LA  UK

## *Abstract*

**For any military operation to succeed, it is essential that the Intentions of the Officer in Command are accurately transmitted down through the command structure. These intentions should convey the commander's requirements for the proposed operation, and should include a statement of the purpose, and required outcome, of the activity about to be undertaken. If the intent is not transmitted accurately, the desired effect will probably not be achieved. This paper will describe a two-stage process directed at assessment of the transmission of Commander's Intent.**

## 1. INTRODUCTION

The progress of orders, through any command structure, has been described by Bateman [1] as the RUDE Cycle (Receive, Understand, Disseminate, Execute), with the same RUDE process occurring at each stage down the hierarchy. It follows from this that at each level below the top command level, a sub-commander must carry out the first three components and then prepare for the fourth. Thus, if a Commander carries out an assessment of the orders handed down by his direct subordinates to their subordinates – that is to say: at two levels down from his own – then the Reception, Understanding and Dissemination components can form a framework for that assessment. If the hierarchy is sufficiently large, "three-down" assessment may be possible. A tool for the assessment by the Commander of lower level orders is currently being developed, but validation of this tool is essential.

Given a validated tool with which a commander can assess the RU&D components of command transmission, it will then be practicable to move to the second stage. This is envisaged as the establishment of measurable variables in the order communication process. There are several potential quantitative measures – taken from cognitive and social psychology - which could be applied to any communication (from length or frequency of messages, to volume of words, sentences, etc, or the number of pages used for an order). The work of English and Guppy [2] suggests that the more effective tank crews use fewer communications, but it is far from clear if this, or the more general measures, can be directly applied to all military activity.

Given a validated method of assessing the communications as "Orders", sampling a range of potential measures should enable some correlations to be made to establish applicability. When the applicable measures have been identified, suitable tools can be generated to facilitate the rapid application of these measures. It should be noted that a generic toolset for this form of knowledge elicitation has already been prototyped, and examples of how this could be applied to the two stages above will be given.

A further stage – envisaged but, as yet, unplanned – is dependent on the two stages above. This would relate communications events to an operational timeline. By working from "Start" to "H-hour", a plot of the exact timings for the issue of Warning, CONOPS and Confirmatory orders should throw further light on the order process.

### 1.1 Command Structure

If we refer to our top-level, or starting-point, Commander as "CMD", then his direct subordinates will be SUB1, their direct subordinates will be SUB2 and so on. In reality, there will be only a few levels below our CMD, but these may vary from nation to nation. As an example, a British Brigade Commander may have two or three Battle-groups (BG) to command. Each Battle-group Commander will have (say) two Infantry Companies and two Squadrons of Armour. Each Company or Squadron will have two or three Platoons or Troops, each made up of individual infantry sections or tanks. With regard to formalised orders, there are only four levels of command, in the UK, that may need to be considered in detail, if we assume a Brigade Commander as CDM.

If we now consider our CMD as having issued a set of orders to his SUB1s, then this CMD is the best person to establish if his intent has been correctly transmitted, it seem logical that the CMD can assess this by a study of the orders passed down from SUB1s to the SUB2s (and SUB2s to SUB3s). In the case of our British command structure (and taking CMD to be Brigade level), this offers three points – with Brigade as the top level, the points are 1: BG, 2: Company / Squadron, 3: Platoon / Troop - at which orders can be assessed for "transmission of intent" on the basis of how the original intent has been passed down the command structure. This resembles the "Chinese Whispers" game beloved of children's parties, but with the originator being able to check (but without any facility to amend) the message as it is relayed. The command hierarchy is shown in Figure 1, and the potential check (or assessment) points in Figure 2. A tool for the qualitative assessment, by a Commander, of orders generated at a lower level, but which relate to his own is currently being developed, but is not yet validated.

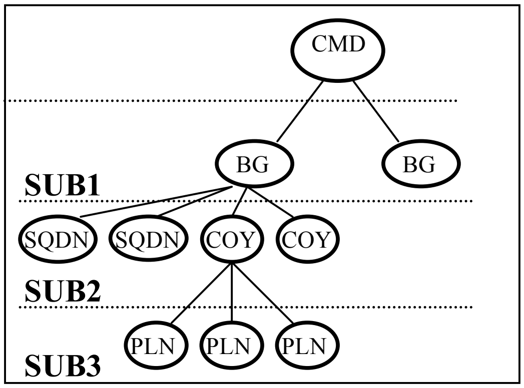


Figure 1: Command Structure

In practical terms, we envisage a situation as in Figure 2, where a Brigade Commander (CMD) issues orders to the Commanders (SUB1) of each of two Battle-groups (or BGs). Each B-G Commander (SUB1) will then issue orders to probably four Company or Squadron Commanders (SUB2); these will in turn issue orders to a number of SUB3s. Our originating CMD can now assess:

Orders SUB1 to SUB2
Orders SUB2 to SUB3

Additionally, SUB1 can assess the orders from SUB2 to SUB3. This is in accord with the concept that a (sub)commander is required to understand the intent of the commander two levels above him [3].

This enables the checks as shown in Figure 2. The CMD can assess 8 sets of orders issued by the BG Commanders, and 24 sets issued by the Company or Squadron Commanders. The BG Commanders will also be able to assess the 24 sets of orders issued at Company or Squadron Commander level.

1.2 Assessment

While this may serve to show the accuracy of the transmission of the original Orders, and will probably give a reasonable indication of where errors may have entered the order system, it may not indicate why those errors have crept in. A typical question in this area is:

*Did SUBs show clear commitment to the Orders passed down to them?*

Two points should now be noted:

a. This will require a military Subject Matter Expert (SME) opinion – and a Brigade Commander should be such an SME.

b. The phrasing of this question will require some revision, as is illustrated in the discussion of the assessment tool.

1.3 Quantitative Measures

In the case of British forces it is considered that any Commander takes 1/3rd of the available time (to H-Hour) for his own order generation process, while leaving 2/3s for his subordinates. This 1/3-2/3 rule has a ripple effect down the command structure, with the end of each 2/3rd period being aligned at H-Hour. During a Commander's 1/3rd period, one or more Warning orders may be issued (the first warning order is to be issued in a "timely manner"), followed by expanded warnings or Operational Concept (CONOP) orders, ending with a Confirmatory or Final Order. This process offers the opportunity to obtain values on a number of measures.

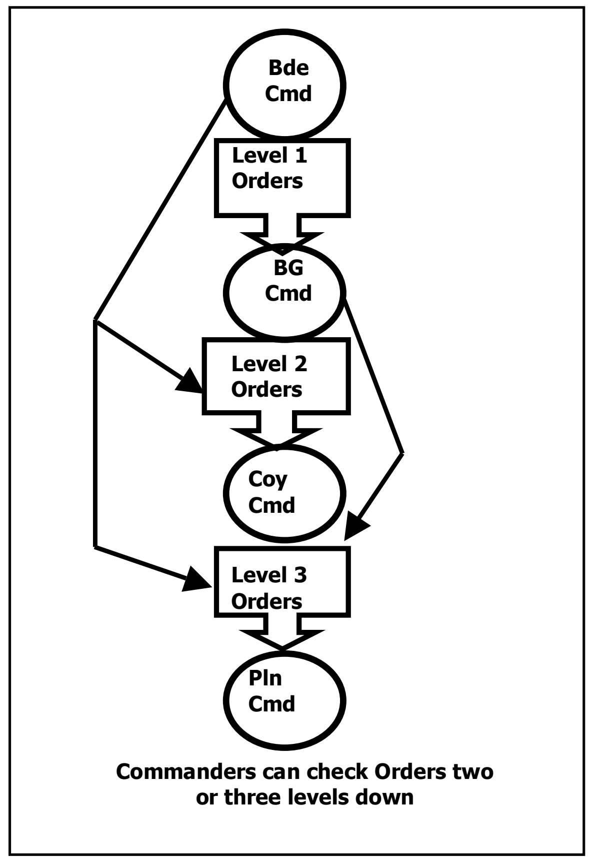


Figure 2: Scope for Order Assessment

If we assume that some Higher Authority has instructed our Brigade Commander (CMD) to achieve an effect by a given time, we have a practical duration time for the line from "Start" to "H-hour". Along this time-line, a number of events can be plotted:

a. Exact time for the issue of each Warning order.
b. Exact time for the issue of a CONOP order.
c. Exact time for the issue of the Confirmatory order.

This can be done at each level of command. Further, each instance of a SUB, at any level, seeking explanation of any point in his orders can also be plotted, and related to the basic timeline events. Combined with the evaluation of orders issued at one or two levels down (as Figure 1) this would provide data on the best use of available time as related to the transmission of intent. It may also prove instructive to relate any upward queries to the timing, frequency and volume of intermediate orders.

There are some further quantitative measures (measurable variables) that can be used. These relate to the actual transmission of orders, and should be considered as independent of the actual method of order transmission. Typically, these could include:

a. Length of the order (in pages, words, characters, or transmission time, for example).

b. Length of each sentence, and the number of sentences.

c. Time for each query (if any).

The work of English and Guppy [2] as mentioned above, suggests that the more effective tank crews use fewer communications, but it is far from clear if this, or the more general measures, can be directly applied to all military activity. Another tool, to facilitate this approach (and which can be used to confirm or refute the effectiveness of any proposed measures) is also being considered for development.

## 2 THE RESEARCH PROTOCOL

2.1 Terminology:

The top-level commander is termed CMD, those sub-commanders – one, two or three levels down the command hierarchy – are termed SUB-1, SUB-2 or SUB-3. The Orders given by CMD are termed O, those by the sub-commanders becoming O-sub-1, O-sub-2 and O-sub-3 respectively.

| | | | |
|---|---|---|---|
| Brigade<br>**Bde-CMD** | | Battlegroup<br>**BdeSub-1** | |
| | **O** | | **O** |
| Battlegroup<br>**BdeSub-1** | | Company<br>**Bde SUB-2** | |
| | **O-sub-1** | | **O-sub-1** |
| Company<br>**Bde SUB-2** | | Platoon/Troop<br>**Bde SUB-3** | |
| | **O-sub-2** | | **O-sub-2** |
| Platoon/Troop<br>**Bde SUB-3** | | | |
| | **O-sub-3** | | |

Figure 3: The proposed notation.

Some form of notation is required to apply this: thus, if a Brigade Commander is called Bde-CMD then :

- the Battle-group Commanders are Bde-SUB-1s
- the Company/Squadron Commanders are Bde-SUB-2s, and -
the Platoon/Troop Commanders become Bde-SUB-3s.

If a Battlegroup Commander is BG-CMD, then the Company and Platoon Commanders become BG-SUB-1s and BG-SUB-2s respectively. The Orders given by each of them are numbered to reflect this.

From this, references to “Two levels down” or “Three levels down” must always be taken as referring to the specified CMD, independent of the rank held by any particular person.

There are several different types of Order, issued at different stages of an operation, and it can be regarded as critical that the assessment of different levels must always be at the same stage. It will probably be best to start with the initial Warning Order. Thus, CMD can assess O-sub-1, O-sub-2 (and perhaps O-sub-3), while SUB-1 can only assess O-sub-2 (and perhaps O-sub-3). In the context of an exercise, the assessment should be done as soon as possible after the respective Orders have been issued. It is also important that elapsed time at which each set of orders is issued is recorded. A form of notation, as outlined above, to distinguish the individual orders and the person issuing them will be required, and it is envisaged that this may have to reflect the type of combat structure being assessed.

2.2 Assessment:

This is carried out by each commander at 2, 3 (and perhaps 4) levels down, and is effected by use of the Assessment Tool being developed from the Cranfield Cognitive Toolset. This tool – based on the principle of the Osgood Semantic Differential [4]- enables a short computerised survey, in which the Commander’s responses are not constrained. The two-step survey takes the form of a single question assessment of each Order (2 or 3 levels down as appropriate) followed by a short detailed question set (probably 5-7 questions in total) which is again applied to the issued Orders. The assessing Commander will not be able to see any of his previous responses.

**The first step** is a single assessment of each Order (at all relevant levels) by the appropriate CMD, with one question only being asked:

***“Do these Orders transmit my Intent to lower echelons in a satisfactory manner?”***

Note that the question may not be phrased in this exact manner (see below). The answers form a baseline assessment to which all subsequent assessments can be compared. When this has been done:

**The second step** is a detailed assessment of each Order, where answers to a standard set of questions are sought. It is hoped to limit the question set to 7 items – and preferably to link these items to the headings of the Five-Paragraph Model of Orders (Situation, Mission, Execution, Service Support, and

Command & Signal) used by both the UK and US forces, and a NATO standard [5]. The secondary aim will be to make the question set independent of the arm of service being assessed.

The question posed above (technically a dichotomous choice question) has a potentially limited response (YES or NO), while the Assessment Tool seeks to elicit shades of opinion. This requires that the questions be phrased in a different manner. The question above:

> ***"Do these Orders transmit my Intent to lower echelons in a satisfactory manner?"***

would need to be re-cast so that it cannot be answered YES or NO, and will take a form like:

> ***"To what degree do these orders accurately convey your intent to a lower echelon?"***

2.3 The proposed Assessment Tool

One of the components of the Cranfield Cognitive Toolset is a survey tool (the OSD Tool) based on the principle of the Osgood Semantic Differential [4], in which the respondent is asked to indicate where their opinion (or position) lies on a continuum between two opposing descriptors. It is essential that the question be properly phrased, and that – together with the two descriptors – it offers the potential for a wide range of responses. Thus, the recast question above could have two descriptors:

***"Badly"*** ***"Very well"***

thus offering a wide choice of position between them.

In use, the OSD Tool presents the basic question above a continuum between the two descriptors. The respondent is asked to indicate his/her position by dragging a pointer (of the normal Windows form) along the continuum by using a mouse function, and clicking on a button when they are satisfied that the pointer is correctly positioned. The starting position is shown below in Figure 4.a below

Since the respondent is asked to indicate a position between the two descriptors, and not to choose a given point on an arbitrary scale, the response will be fast, and no less accurate than a forced choice.

From the viewpoint of the researcher, however, the continuum shown below (Figure 4.a) is actually hiding a multi-point scale. This scale can have a range of intervals from 2 to 100, permitting the use of several statistical analysis approaches. Further, while the scale is originally an equal interval scale, the data can be exported into a spreadsheet (e.g. Excel) and related to an unequal interval scale; this reflecting a non-linear relationship between the two descriptors.

It is envisaged that a number of pilot studies will be required before any unequal interval scale can be determined. While no firm relationship between any scale labels and scale points can be pre established, we believe that the two examples shown above may well be representative. The establishment of a generic scale can be attempted when the second stage (detailed questions) has been completed

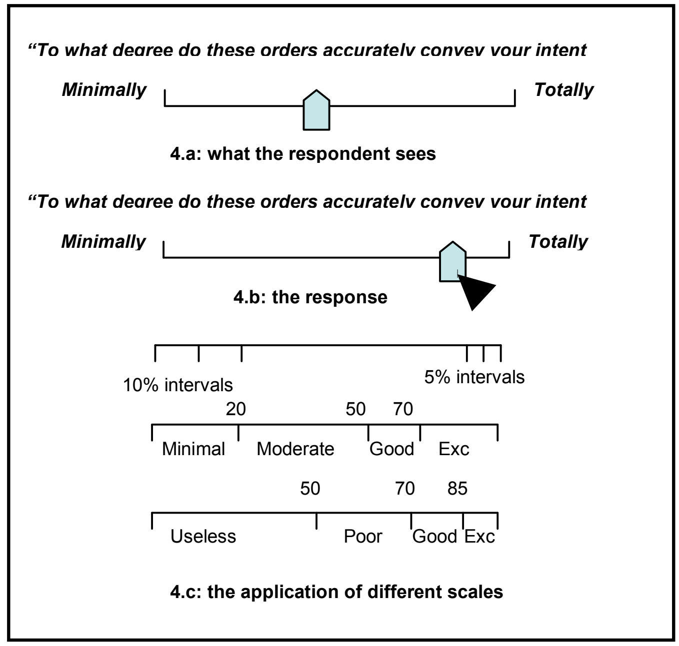


Figure 4: Survey input and hidden scales

.The ability to export to a spreadsheet has a number of benefits:

First: This allows the use of templates that can carry out multiple correlations. This will be of particular value in the comparison of overall assessments to detailed assessments. Since the scale intervals and their labels are determined *post hoc*, the process can be repeated until a good fit with the original responses is obtained. This, in turn, will provide some measure of validation. It should be particularly easy for the researcher to compare the overall assessment with the detailed assessment on a numerical basis, and then to identify potential anomalies, within the individual CMDs assessments (2 levels down), and between the assessments for level 2 and level 3.

Second: This may also serve to identify any particular problem areas at the O-sub-2 level. There are several ways of presenting data in a graphical form from a spreadsheet; a bar-chart, with a superimposed line denoting the boundary between acceptable and non- acceptable values has found favour in other applications.

Third: This enables the use of other templates that can grade each instance of O-sub-2 separately, even to the extent of generating a colour-coded rating. Such gradings and ratings will have to be related to an unequal interval scale, and the validation of such a scale would be required before any high degree of confidence could be placed in the rating. The use of colour coding (and particularly the Red-Amber-Green, or Traffic Light model) is used elsewhere in the military, as a way of offering a fast warning without the need for an observer to evaluate numerical data.

## 3 THE SECOND STAGE

Given data from the first stage assessments, it will become possible to evaluate a number of measures (as quantitative variables) based on the cognitive and communications psychology literature, as well as that from the Command and Control field. Research by Klein [6] revealed a dramatic difference in word-count in almost 100 Command Intent statements, and it is not clear how much of this variation was due to differences in context. As an example of how this could be approached, all communications "up" one command level, seeking clarification or amplification of an Order, could be logged and correlated against the CMD assessment. One would expect that there would be a high correlation of the numerical values of this (or of one or more of the other types of measure outlined in the introduction) with good or bad CMD assessments. For example, a SUB receiving a good CMD assessment, may have made fewer clarification requests than another SUB. This would, in turn, enable the identification of any events that would serve as indicators that could show potential problems in transmission of Intent.

Provided that the CMD assessments and the data from identified variables is kept separate, the variables would almost certainly be usable on direct CMD-SUB orders. Interpretation of the relationships would, however require some degree of SME input. The judgement of the military value and correctness of an order must be a matter for military judgement. Quantitative measures that can be correlated, both with each other, and with the success or failure in respect of obtaining the required effect, should not need any military validation, although this will remain desirable.

## 4 CONCLUSION

The procedure outlined above offers a two-stage approach to the measurement of the Transmission of Command Intent. While it has not yet been validated using realistic command or planning exercises, it would appear to offer a useful framework for extending the present largely qualitative approach to assessing command intent. The tool mentioned above has been prototyped using a generic toolset. This – the Cranfield Cognitive Toolset - was primarily developed for the assessment of Human-System Integration, but has already been used for course assessment at the Defence Academy of the UK, and been shown to be an effective way of collecting data.

**Acknowledgements:**

Research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

The authors would like to thank Ms J de Looy-Hyde of DSTL for her review of Command Intent, used in the preparation of this paper.

## REFERENCES

[1] Robert Bateman III, Avoiding Information Overload, *Military Review, Jul-Aug 1998* downloadable from http://www-cgsc.army.mil/milrev

[2] English, N. and Guppy, A. (1994) *The adequacy of AGPT to support tactical training: a user evaluation survey.* Farnborough, Hampshire: The Army Personnel Research Establishment, Working Paper 8/94

[3] Army Doctrine Publication, Volume 2, COMMAND, 2005

[4]Osgood, C.E., Suci, G., & Tannenbaum, P. (1957) *The measurement of meaning.* Urbana, IL: University of Illinois Press

[5] Nato (1994) *STANAG 2014: Operations Orders, Annexes to Operations Orders, Administrative and Logistics Orders* Brussels, NATO (relates to US Army Form AF71-2, and the UK Staff Officers Handbook)

[6] Klein, G., (1993) Characteristics of Commander's Intent Statements. Proceedings of Command and Control Conference, National Defense University, Ft McNair, Washington, DC

# Coordination of Highly Contingent Plans

David J. Musliner, Honeywell Laboratories, Minneapolis, MN, david.musliner@honeywell.com
Robert P. Goldman, SIFT, LLC, Minneapolis, MN, rpgoldman@sift.info
Edmund H. Durfee, Jianhui Wu, Dmitri A. Dolgov,
University of Michigan, Ann Arbor, MI, {durfee, jianhuiw, ddolgov}@umich.edu
Mark S. Boddy, Adventium Labs, Minneapolis, MN, mark.boddy@adventiumlabs.org

**Abstract**—*Conventional military planning systems construct plans with very limited flexibility. In the future, military plans will evolve into a much more expressive, contingent form. This paper describes how Honeywell's distributed Coordinator agents reason about complex domains to construct and execute highly contingent plans. The agents operate in a very dynamic environment in which complex hierarchical tasks can arrive unpredictably and the agents have to build coordinated joint plans on the fly, while execution proceeds. Using carefully limited forms of inter-agent communication, the agents develop agreements on their future coordinated behavior and rely on those agreements to build highly contingent plans (partial policies) that specify what actions they should take in a wide variety of possible futures. As mission execution proceeds and the tasks yield varying outcomes, the agents must rapidly, continually coordinate and adapt their plans. The result is a distributed multi-agent system capable of building and flexibly executing complex, highly-contingent coordinated mission plans.*

## 1. INTRODUCTION

Conventional military planning systems construct plans with very limited flexibility; often there is only a baseline operational plan with a few hand-crafted contingency branches that essentially amount to deploying reserved assets which otherwise remain unused. In the future, as automated planning systems become more sophisticated and military operations become more automated, military plans will evolve into a much more expressive, contingent form. Plans will be built to account, ahead of time, for operational tasks that take varying time, have varying levels of success, and should be combined in widely different ways depending on what earlier tasks (and adversaries) have accomplished. Constructing such contingent plans in a distributed coalition environment and coordinating the distributed execution of those plans will become much harder than current coalition activities. The DARPA COORDINATORS program is exploring the core computational issues that underlie exactly these problems.

This paper describes how Honeywell's distributed COORDINATOR agents reason about complex domains to construct and execute highly contingent plans. The agents operate in a very dynamic environment in which complex hierarchical tasks can arrive unpredictably and the agents have to build coordinated joint plans on the fly, while execution proceeds. Using carefully limited forms of inter-agent communication, the agents develop agreements on their future coordinated behavior and also develop highly contingent plans (partial policies) that specify what actions they should take in a wide variety of possible futures. As mission execution proceeds and the tasks yield varying outcomes, the agents must rapidly, continually coordinate and adapt their plans.

Our current solution combines restricted forms of inter-agent coordination agreements with dynamic, probabilistic projections of possible future worlds in the form of Markov Decision Problems (MDPs). By carefully guiding and pruning the projection, or unrolling, of the MDP model of possible future states, our COORDINATOR agents attempt to focus their decision-making attention on the partial plans with the highest probability of being useful. The MDP formulation allows our COORDINATOR agents to produce highly contingent plans in the form of partial policies, specifying what actions to take in all the possible future states explored so far. Novel technical elements of our COORDINATOR agents include their ability to exploit problem structure to dramatically reduce the complexity of future planning, their methods for guiding the MDP unrolling process, and their ability to continue unrolling during mission execution.

The result is a distributed multi-agent system capable of building and flexibly executing complex, highly-contingent coordinated mission plans.

## 2. THE COORDINATORS PROBLEM

Our work is being done in the context of the DARPA-funded COORDINATORS program, which aims to identify, prototype, and evaluate technical approaches to scheduling and managing distributed activity plans in dynamic environments. As a motivating example, consider the following scenario. A hostage has been taken and might be held in one of two possible locations. Rescuing the hostage requires that both possible locations are entered by special forces *simultaneously*. As the activities to move personnel and materiel into place are pursued, delays may occur or actions intended to achieve precursor objectives may have unexpected results (e.g., failure). COORDINATOR agent systems will be associated with the various human participants. COORDINATOR agents should monitor the distributed plans and manage them as the situation evolves, to increase their effectiveness and make them more likely to succeed.

In general, a set of COORDINATOR agents is meant to work together to maximize the reward gained by the group as a whole. In other words, the problem is to compute an effective *joint* policy for the agent society, in which the actions taken by one agent can depend on the state of the group as a whole, not just the local state of that agent. The agents are time-pressured: each agent must make timely action decisions during execution. Furthermore, the problem must be solved in a distributed fashion.

Although this is a problem of joint action, the problem solving is necessarily distributed, for reasons both definitional and practical. The definitional reasons include the fact that each agent has only a partial, local model of the problem, and the agents are prohibited (for organizational reasons) from building a complete joint model of the situation. The practical reasons include the sheer scope of the problem to be solved.

Each agent's partial problem model (aka domain model) includes the actions that the agent can execute, which are stochastic, rather than deterministic, and some of the actions its peers can perform. The problem model also provides *partial* information about the rewards that the society as a whole will receive for reaching various states. This model is not static: the agent can receive information about action outcomes and problem model updates during execution. Therefore, agents must be able to manage and reformulate policies reactively.

## 3. C-TÆMS

COORDINATORs researchers have jointly defined a common problem domain representation based on the original TÆMS language [1]. The new language, C-TÆMS [2], provides a semantically sound subset of the original language, representing multi-agent hierarchical tasks with stochastic outcomes and complex hard and soft interactions. Unlike other hierarchical task representations, C-TÆMS emphasizes complex reasoning about the utility of tasks, rather than emphasizing interactions between agents and the state of their environment.

C-TÆMS permits a modeler to describe hierarchically-structured tasks executed by multiple agents. A C-TÆMS task network has *nodes* representing *tasks* (complex actions) and *methods* (primitives).[1] Nodes are temporally extended: they have durations (which may vary probabilistically), and may be constrained by release times (earliest possible starts) and deadlines. Methods that violate their temporal constraints yield zero quality (and are said to have *failed*). At any time, each C-TÆMS agent can be executing at most one of its methods, and no method can be executed more than once.

A C-TÆMS model is a discrete stochastic model: methods have multiple possible outcomes. Outcomes dictate the *duration* of the method, its *quality*, and its *cost*. Quality is constrained to be non-negative, and duration must be an integer greater than zero. Cost is not being used in the current work. Quality and cost are unitless, and there is no fixed scheme for combining them into utilities. For the initial COORDINATORS experiments, we treat quality as non-normalized utility (we will use the terms "utility" and "quality" pretty much interchangeably).

To determine the overall utility of a C-TÆMS execution trace, we must have a mechanism for computing the quality of tasks (composite actions) from the quality of their children. Every task in the hierarchy has associated with it a "quality accumulation function" (QAF) that describes how the quality of its children are aggregated up the hierarchy. The QAFs combine both logical constraints on subtask execution and how quality accumulates. For example, a :MIN QAF specifies that all subtasks must be executed and must achieve some non-zero quality in order for the task itself to achieve quality, and the quality it achieves is equal to the minimum achieved by its subtasks. The :SYNCSUM QAF is an even more interesting case. Designed to capture one form of synchronization across agents, a :SYNCSUM task achieves quality that is the sum of all of its subtasks that start at the same time the earliest subtask starts. Any subtasks that start after the first one(s) cannot contribute quality to the parent task.

The quality of a given execution of a C-TÆMS task network is the quality the execution assigns to the root node of the task network. C-TÆMS task networks are constrained to be trees along the subtask relationships, so there is a unique root whose quality is to be evaluated. C-TÆMS task networks are required to have a deadline on their root nodes, so the notion of the end of a trace is well-defined. One may be able to determine bounds on the final quality of a task network before the end of the trace, but it is not in general possible to determine the quality prior to the end, and it may not even be possible to compute useful bounds.

Traditional planning languages model interactions between agents and the state of their environment through preconditions and postconditions. In contrast, C-TÆMS does not model environmental state change at all: the only thing that changes state is the task network. Without a notion of environment state, in C-TÆMS task interactions are modeled by "non-local effect" (NLE) links indicating inter-node relationships such as enablement, disablement, facilitation, and hindrance.

Figure 1 illustrates a simple version of the two-agent hostage-rescue problem described earlier. The whole diagram shows a global "objective" view of the problem, capturing primitive methods that can be executed by different agents (A and B). The COORDINATORs agents are *not* given this view. Instead, each is given a (typically) incomplete "subjective" view corresponding to what that individual agent would be aware of in the overall problem. The subjective view specifies a subset of the overall C-TÆMS problem, corresponding to the parts of the problem that the local agent can directly contribute to (e.g., a method the agent can execute or can enable for another agent) or that the local agent is directly affected by (e.g., a task that another agent can execute to enable one of the local agent's tasks). In Figure 1, the unshaded boxes indicate the subjective view of agent-A, who can perform the primitive methods Move-into-Position-A and Engage-A. The "enable" link indicates a non-local effect dictating that the Move-into-

[1] The terminology is somewhat unfortunate, since conventional HTN planners refer to their composite actions as *methods* and their primitives as operators.

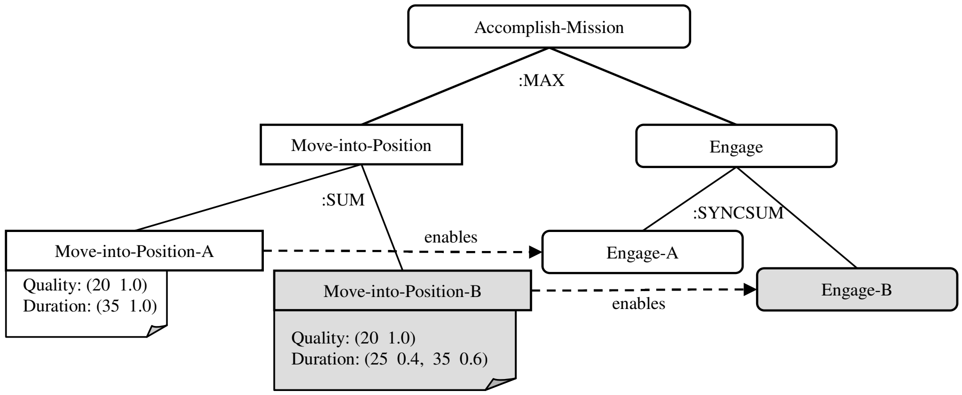


**Figure 1:** A simple C-TÆMS task network for two agents, illustrating some of the representation features. Some details have been omitted for brevity.

Position-A method must be completed successfully before the agent can begin the Engage-A method. The diagram also illustrates that methods may have stochastic expected outcomes; for example, agent-B's Move-into-Position-B method has a 40% chance of taking 25 time units and a 60% chance of taking 35 time units. The :SYNCSUM QAF on the Engage task encourages the agents to perform their subtasks starting at the same time (to retain the element of surprise).

## 4. Solution Approach: Markov Decision Processes

Given a C-TÆMS task network with stochastic method outcomes, we can frame the objective COORDINATORs problem as a multi-agent Markov Decision Process (MDP) [3]. Briefly, an MDP is akin to a finite state machine, except that transitions are probabilistic, rather than deterministic or nondeterministic. Agents may also receive reward (which may be either positive or negative) for entering some states. Typically, this reward is additive over any trajectory through the state space (some adjustments are needed in the case of MDPs of infinite duration). The solution to an MDP is a *policy* — an assignment of action choice to every state in the MDP — that maximizes *expected utility*. Expressing the COORDINATORs problem as an MDP provides a sound theoretical basis for decision-making and action under uncertainty. Furthermore, there are relatively simple, efficient algorithms for finding optimal policies. However, the state space size of the MDPs can be enormous.

A single COORDINATOR agent's C-TÆMS task model specifies a *finite-horizon* MDP. The problems are finite-horizon because C-TÆMS problems have finite duration, with no looping or method retries. However, the MDP tends to be quite large for even modest-sized C-TÆMS problems because of the branching factor associated with uncertain outcomes, and because of the temporal component of the problem. For example, even a single applicable method with three possible durations and three possible quality levels gives us a branching factor of nine. In addition, time is a critical aspect of TÆMS problems: methods consume time and NLEs can have associated delays (so WAIT is often a useful action alternative). Furthermore, an agent can always abort a method that it is executing, and choose to start a different method. So the branching factor is never less than two at every time tick, in a full consideration of the (single-agent) problem.

Multi-agent C-TÆMS MDPs are even worse. If one were to formulate a centralized COORDINATORs problem directly as an MDP, the action space would have to be a tuple of assignments of actions to each agent. Each agent's policy could be dependent on all the possible actions that the other agents could choose, and all the outcomes they could receive. Naturally this causes an explosion in the state space of the problem. Beyond complexity, there are other reasons we cannot construct the optimal multi-agent MDP. COORDINATORs problems are time-constrained and truly distributed: each COORDINATOR agent gets only a limited subjective view and a limited time to build and execute its plans, so forming a perfectly optimal, centralized joint policy is not feasible. Furthermore, information security policies may prevent the agents from sharing their local views completely.

Therefore, we have a developed a distributed COORDINATOR agent system that tries to retain the principled advantages of an MDP-based approach while supporting truly distributed operations and information hiding, in a time-adaptive manner. Each agent builds a *partial* MDP for its local subjective problem, to support its own decision-making about what actions (methods) it should perform. The partial MDP is incrementally extended as more deliberation time is available to the agent, so that it becomes complete and locally-optimal if sufficient time is available.

Because each agent's subjective view may not accurately convey how local method quality contributes to the overall team mission quality, simply solving local MDPs for optimal policies is not sufficient. We must have the agents communicate to share information about their plans and expectations, so that agents whose problems interact can coordinate effectively. To that end, our agents also have a coordination/negotiation capability that allows them to efficiently reach joint agreements about how they will coordinate over interactions portions of the full C-TÆMS problem.

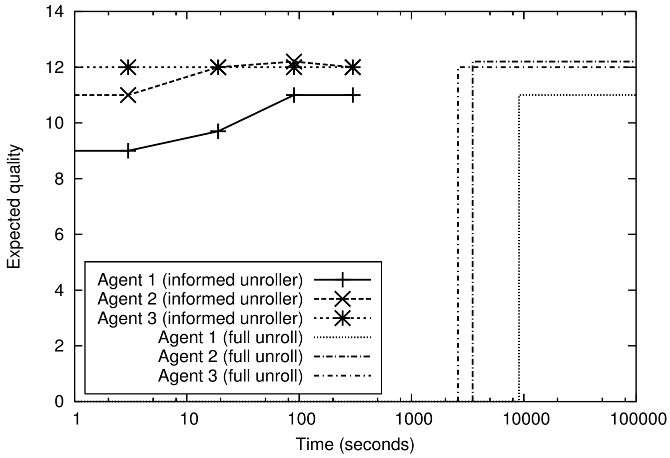


**Figure 2:** The Informed Unroller can find near-optimal policies much faster than building the complete MDP.

## 5. Partial MDPs: "Informed Unrolling"

We refer to the process of converting a C-TÆMS problem into an MDP problem as "unrolling," because it involves projecting forward from an initial state (where no methods have been executed) to imagine future possible states of the C-TÆMS network in which some methods have been executed at particular times and have received particular outcomes. The core unrolling algorithm is thus a simple state-space enumeration process where an MDP state is expanded by creating the successor stats that result from each of the possible action choices and their outcomes. These successor states are added to an *openlist* of un-expanded states, and the process ideally continues until the openlist is empty and the full MDP state space has been enumerated.

Since full enumeration of even single-agent C-TÆMS MDPs is often impractical, we have developed a technique for heuristically-guiding the enumeration of a subspace of the full MDP. Our *informed unroller* (IU) algorithm prioritizes the openlist of states waiting to be unrolled based on an estimate of the likelihood that the state would be encountered when executing the optimal policy from the initial state. The intent is to guide the unrolling algorithm to explore the most-probable states first.

One cannot determine the probability of reaching a state without considering the policy followed by the agent. Therefore, the IU intersperses policy-formulation (using the Bellman backup algorithm) with unrolling. This means that we must be able to find an (approximately) optimal policy for partial MDP state spaces, which means we must have a heuristic to use to assign a quality estimate to leaf nodes in our search that do not represent complete execution traces. We have developed a suite of alternative heuristics for estimating intermediate state quality, since the problem of finding a good heuristic is quite difficult.

Early results from our evaluation of the IU algorithm against a complete solution of (small) MDPs are promising. For example, in Figure 2 we show a comparison of the performance of the informed unroller against the complete unrolling process. In these small test problems, the informed unroller is able to find a high-quality policy quickly and to return increasingly effective policies given more time. This allows the IU-agent to flexibly trade off the quality and timeliness of its policies.

The IU approach is related to the "approximate dynamic programming" algorithms discussed in the control theory and operations research literature [4]. These approaches derive approximate solutions to MDP-type problems by estimating, in various ways, the "cost to go" in leaf nodes of a limited-horizon portion of the full state space. While our exploration of the literature is not yet complete, initially we believe that a key difference in our IU approach is the notion of time-dependent horizon control and unrolling-guidance (vs. just estimation of leaf-node reward for policy derivation).

The IU method is a special case of the find-and-revise algorithm schema [5] (which is a generalization of algorithms such as $LAO^*$ [6]). $LDFS$-family algorithms use knowledge of the initial state(s) and heuristics to generate a state subspace from which a policy can be abstracted. A find-and-revise algorithm finds a state in the network for which the current value estimate is inaccurate, and revises the value for that state (e.g., by generating successors, and propagating the value functions backwards in standard MDP fashion).

Our technique differs from the general case, and its instances, in substantial ways. $LAO^*$ generates a state subspace from which the optimal policy can be provably derived. The IU, on the other hand, executes online, and might lack enough time to enumerate such a state subspace even if it knew exactly which states to include. The IU is an anytime algorithm, unlike $LAO^*$, which runs offline. For this reason, the IU makes no claims about policy optimality; indeed, it is not even guaranteed to generate a closed policy.

The general find-and-revise algorithm family can provide guarantees weaker than those of $LAO^*$, but those guarantees rely on having an admissible heuristic value function for states that have not been fully explored. However, even if we had an admissible heuristic, it is not at all clear that the IU should use it. An admissible heuristic will tend to push the policy expansion to explore states where it is **possible** that the optimum will be found, in order that we not miss the optimum. However, the IU is operating in a time-pressured domain. So we should not be encouraging the system to move towards promising unexplored areas — that will tend to leave the agent with a policy that is broad but shallow, and virtually guarantee that it will "fall off policy" during execution. Instead of admissibility, we must find a heuristic function that will cause the agent to tend to build policies that trade off considerations of optimal choice against completeness/robustness of the policy. It is possible that this heuristic should be time-dependent — as the agent runs out of time for policy development, the IU's heuristic should focus more on robustness and less on optimality.

## 6. Coordination

When we consider multiple COORDINATOR agents, the problem expands to finding an optimal *joint policy*. This problem is challenging because:

- The number of possible local policies for agents is in general very large, so the product space of joint policies to search through can be astronomical.
- The size and distribution of the problem makes reasoning about the global behavior of the system impossible.

To address these practical limitations, our COORDINATOR agents do not try to solve the full optimal joint policy problem. Instead, they make several simplifying assumptions and restrict the forms of solutions they will be able to find, making the search for an approximately-optimal joint solution more tractable. Our agents use limited forms of negotiation to establish a set of inter-agent commitments. These commitments represent a partial set of agreements about which agent is performing which methods, at what times. The agents then rely on those commitments when generating their partial MDP policies. The commitments are used as both assumptions (e.g., another agent has agreed to perform a method that will enable my action) and as obligations (e.g., I have agreed to perform a method that will enable another agent). Assumptions such as remote enablement agreements can be built into the local problem model by including "proxy" methods that enable the local method at the agreed-upon time. Obligations to execute methods by a particular time are met by adding extra reward to the MDP in states that satisfy the commitment. These two mechanisms bias the MDP policy-generation process towards policies that rely upon and satisfy the agent's commitments.

There are several ways in which this approach may result in sub-optimal behavior. For example, the actual optimal policy set may not adhere to a static set of commitments: to behave optimally, agents may have to adjust which enablements they will accomplish depending on how prior methods execute. To mitigate this weakness, our agents deliberate and negotiate continually, so that they can manage and adapt their commitment set and policies on the fly as methods execute.

## 7. Conclusions

Our multi-agent coordination system uses limited forms of negotiated commitments to bias partial-MDP policy derivation. The resulting agents are able to very quickly create initial coordinated policies, improve those policies given more deliberation time, and adapt the policies as new information arrives, including both method outcomes and new C-TÆMS problems.

In the context of coalition operations, where different agents may not be able to share some portions of their intentions, these techniques can still be applied. Enforcing information security or privacy policies could be done on a local-agent level, preventing the agent from establishing commitments about private intentions (e.g., not telling other agents that it intends to execute a particular method or task). The resulting system would be expected to perform less-optimally, given the restrictions on its search for joint policies, but the system should still be robust and capable of establishing coordinated behavior on the portions of the problem over which agents are willing to communicate.

## Acknowledgments

This material is based upon work supported by the DARPA/IPTO COORDINATORs program and the Air Force Research Laboratory under Contract No. FA8750–05–C–0030. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA, the U.S. Government, or the Air Force Research Laboratory.

## References

[1] B. Horling, V. Lesser, R. Vincent, T. Wagner, A. Raja, S. Zhang, K. Decker, and A. Garvey, "The TÆMS white paper," University of Massachussetts, Amherst, Computer Science Department, Tech. Rep., Jan. 1999.

[2] M. Boddy, B. Horling, J. Phelps, R. P. Goldman, and R. Vincent, "C-TÆMS language specification," Apr. 2005, unpublished; available from this paper's authors.

[3] M. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 1994.

[4] D. P. Bertsekas, "Dynamic programming and suboptimal control: A survey from ADP to MPC," in *Proc. Conf. on Decision and Control*, 2005.

[5] B. Bonet and H. Geffner, "Learning depth-first search: A unified approach to heuristic search in deterministic and non-deterministic settings, and its application to MDPs," in *Proceedings of the Sixteenth International Conference on Automated Planning and Scheduling*, D. Long, S. F. Smith, D. Borrajo, and L. McCluskey, Eds., Jun. 2006, pp. 142–151.

[6] E. A. Hansen and S. Zilberstein, "LAO: a heuristic search algorithm that finds solutions with loops," *Artificial Intelligence*, vol. 129, no. 1-2, pp. 35–62, 2001.

# Holistan: A Futuristic Scenario for International Coalition Operations

David Roberts, Gavin Lock, and Dinesh C. Verma, *Fellow, IEEE*

***Abstract*— Scenarios play an important role in driving the research agenda for programs focused on improving effectiveness in network centric operations. Several excellent scenarios such as Binni (www.binni.org) have played a key role in motivating research projects. However, as the geopolitical situation in the world has evolved, several new factors have become prominent that require the creation of a new scenario. This paper presents a new scenario for coalition operations that was developed to drive the research in the International Technology Alliance program, a multi-disciplinary multi-year program between the US and UK.**

***Index Terms*— Network Centric Operations; Military Scenarios; Coalition Operations;**

## I. Introduction

In order to derive the technical requirements of network centric operations, a well-developed scenario can play an important role. While determining the technical requirements and scope of an international multi-area research program jointly sponsored by the US Army and UK Ministry of Defence, we needed to identify scenarios for this purpose. This paper discusses the scenario that we developed, the reason we opted to develop a new scenario instead of reusing an existing one, and how we are using the scenario to develop the technical requirements for the research program.

US/UK International Technology Alliance (ITA) is a multi-year fundamental research program that looks at research activities spanning four technical areas: (i) network theory (ii) security across system of systems (iii) sensor information processing and delivery and (iv) distributed decision making and coalition planning. The program is structured to address twelve important technical projects, (three projects per technical area) with significant collaboration and linkages across projects in different technical areas. The program is designed to foster collaboration between the US and UK research organizations and entities. A more detailed overview of the research program can be found at URL http://www.usukita.org.

In order to derive the technical requirement of the program, we looked at existing scenarios. However, it became clear that we would need to develop a new scenario to fully address the requirements of the program, even though several excellent scenarios already exist in the community. Section II of this paper discusses the reasons why we felt that defining a new scenario would be necessary.

Section III describes the theater of operation that we developed for the scenario, which is a fictional country located in the Middle East. Sections IV, V and VI describe some of the vignettes that we have developed in the context of the program, and describe how we have used those vignettes to bridge the gap between technical requirements of our research program and the military context provided by the scenario. Finally, Section VII provides our experience with the scenarios and areas for future work.

## II. Need for a New Scenario

As we launched into the task of selecting a suitable scenario, our first instinct was to reuse some of the excellent existing scenarios like that of Binni (http://www.binni.org) for the purposes of our program. However, on further research, we felt that the existing scenarios might lead us into the trap of planning for "the last war". We will use Binni as the basis for comparison, but the reasoning can be easily seen to be applicable to other scenarios as well. On a basic research program like that of ITA, we would expect our research to become useful to the military in the course of next 5-10 years. As we tried to extrapolate the future requirements, and the changes in the geopolitical situations, we felt that there were several key changes that needed to be captured into the new scenario.

Existing scenarios have made several implicit assumptions which were valid at the time of the development of that scenario, but may not be valid 5 years from now, just as our

Manuscript received January 15, 2007. Research was sponsored by US Army Research laboratory and the UK Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Army Research Laboratory, the U.S. Government, the UK Ministry of Defense, or the UK Government. The US and UK Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon..

D. Verma is with the IBM TJ Watson Research Center, 19 Skyline Drive, Hawthorne, NY 10532, phone: 914-784-7466; fax: 914-784-6205; (e-mail: dverma@us.ibm.com).

G. Lock, is with Logica CMG, Stephenson House, 75 Hampstead Road, London, NW1 2PL, UK. (e-mail: gslock@hotmail.co.uk).

D. Roberts is with the IBM United Kingdom, Hursley Park, Winchester, Hants, UK SO21 2JN (e-mail: daveroberts@uk.ibm.com).

extrapolations are not likely to be valid a couple of decades from this paper. Some of the implicit assumptions that are no longer valid include coalition trust relationships, the root cause for conflict, and the sophistication levels of the enemy.

Existing scenarios typically include a coalition of two or more advanced nations (e.g. NATO countries) cooperating with a local government in Asia/Africa to counter insurgents or aggressors. The trust relationship between the coalition members is high, and all parties of the coalition are committed to the common cause of defeating the insurgents. As an example, in Binni scenario, it is clear that the UN backed forces have a mission that is fairly well aligned with that of the Gao and Binni governments, and Agadez is clearly on the other side. If we compare that to the real geopolitics in contexts such as Afghanistan, neither the concept of a reasonably well-trusted set of coalition members, nor a clearly identified sets of good guys versus bad guys are valid. Loyalties and sentiments of coalition members vary widely; some have apprehensions about a repetition of the colonial era; and a distribution of population across national boundaries that bear no relationships with the tribal and ethnic boundaries creates divided loyalties among individuals. Even among states, a varying level of autonomy among different arms of the states makes it unclear whether full support from all agencies of a coalition member is available, e.g. conflicts among the positions of different branches of Pakistan in different issues related to Afghanistan have been reported in the media. The trust relationships in such type of coalition are much more complex than that of existing scenarios, and are likely to remain valid in the next decade of coalition operations.

Another key factor that has changed from Binni to the present time is the root cause of the conflict and insurgency. Binni was a conflict caused by contention for resources. On the other hand, the prevailing conflict in the current war on terror is a conflict caused due to ideology and misguided beliefs of a small but violent set of people. While resource contention conflicts can be resolved and settled through various settlements among the feuding parties, resolving ideological disputes is a much trickier endeavor. One particularly difficult aspect of an ideological conflict is that even beneficiaries of a humanitarian mission may end up harboring a deep-seated resentment against their benefactors. Suicide bombers rarely emerge in conflicts based on contention for resources, yet are a common phenomenon is conflicts that are fueled by ideology – never mind the fact that the ideology may be misguided or be fanatically observed by only one side in the conflict.

Yet another implicit assumption in the existing scenarios has been the lack of technical sophistication of the enemy. The Agadez soldiers and their supporters lack any deep knowledge of the technology possessed by the UN coalition forces, and are unlikely to tamper with a smart sensor deployed in the field. On the other hand, the enemy in the war against terror has people with significant technical sophistication, and is likely to be able to launch a variety of sophisticated attacks against the infrastructure deployed in the field. While the enemy may not have the same level of sophistication as a coalition of US and UK armed forces, this gap can often be surmounted with human ingenuity and ideology-driven passion to do harm.

Another key aspect of the future war would be the importance of non-military modes of engagement. The future insurgent may try to utilize the Internet and globally connected media to mobilize world opinion against coalition operations. What is gained in military tactical operations can easily be lost by a poor plan of engagement with the civilian forces in the theater of operations. These nuances of future decision making are not reflected in the current scenarios.

Because of these important changes in the geopolitics of global conflict, we felt that a new scenario that would reflect the situations of a potential conflict 5-10 year sin the future needs to be developed.

## III. THE THEATER OF OPERATIONS

The theater of operations for our scenario is set in the hypothetical country of Holistan located somewhere in the middle-east, with a population which has a strong underlying resentment fuelled by religious fundamentalists against the western civilization. The government of the country has obtained weapons of mass destruction (nuclear and chemical) against the wishes of US and UK foreign policy, but is generally friendly towards the two nations. However, due to an impending crisis in the country, which strongly increases the risks of nuclear/chemical technology falling into the hands of religious militants, the two nations are forced to form an uneasy coalition with a somewhat reluctant government of Holistan with the objective of saving the surrounding area from a larger crisis. The terrain, environment and the operational environment of the coalition operations provide new challenges for the research community, including the topics of managing limited trust and risk among coalition partners, hostile environments without characteristics conducive for wireless networks, innovative use of sensors in parasitic modes, decision making with partial information, and synthesis of environment data into simpler situational contexts.

Holistan, officially the Malek Republic of Holistan, is a country located in the Greater Middle East. It has a thousand-kilometer coastline along the Arabian Sea in the south and borders the country of Rugistan on the west, the Democratic Republic of Weighat to the east and the Republic of Sugaria in the far northeast. The map of Holistan and neighbors is shown in Figure 1.


Figure 1

Holistan was established as a modern state in middle of twentieth century, after gaining freedom from Britain which had colonial control over it. The region, however, has a long history of settlement and civilization. The region was invaded by Berbers, Mongols, Afghans, Greeks, Persians, and Arabs, and is a melting pot of different cultures. It was incorporated into the British colonies in the nineteenth century. Since independence, Holistan has experienced times of significant military and economic growth, and times of instability. Holistan has a large, but perhaps not quite modern, armed force and is a declared nuclear weapons state.

Maulana Bismillah Malek was an Islamic spiritual leader in the fourteenth century who gave a new interpretation to Islam in the Sufi tradition. The variant of Islam preached by Malek became the dominant religion of Holistan. Maulana Malek rose to become the counselor and spiritual guide of the reigning monarch who officially declared himself a Malek, and promoted the conversion of Holistan to Malek beliefs.

After independence, Holistan had periods of democracy intermingled with periods of military rule. The current ruler is military dictator, but has allowed the growth and operation of political parties in elected bodies at provincial and national level. The most successful party is led by religious fundamentalists. Holistan acquired nuclear technology in the 1990s, which is ostensibly for power generation, but the country also possesses nuclear and chemical weapons.

Holistan is a federation of three provinces and a capital territory. The provinces and the capital territory are subdivided into a total of 110 districts. Each district contains several district blocks and villages with local government. The provinces are Bhalustan lying in the south western corner of Holistan - a sandy desert with a significant number of mineral resources; Cincodoab a fertile plain area of Holistan, home to 60% of its population and housing its nuclear power plants; and Mantristan is a rugged mountainous territory which is populated by tribes that maintain a traditional lifestyle, and is a hot-bed of insurgency supported by Malek religious fundamentalists. Because of diversity of population among different regions of Holistan, there have been various sedition movements calling for independence of various regions at times, but most of them have been repressed by the military regime. A significant number of Maleks have emigrated to the West and work in the Middle East, US and Europe (including UK).

Holistan is known to house several fundamentalists of Malek religion, which have provided safe harbor and an operational base for cross-border terrorist operations into Weightan and Rugstan, neighbors to east and west which have significant Malek populations. Although the official military government position is that they do not support terrorists, covert support is alleged by neighboring nations. The situation was considered under control until the uncovering of a significant plot by the terrorists to take over the control of nuclear installations and to explode such bombs simultaneously in US, UK, Rugistan and Weightan. Fortunately, the plan was intercepted in the nick of time. Several members of this plot belonged to the Holistan army, and serious questions were raised in international circles about the ability of the local army to prevent such actions in the future. The uncovering of this plot was accompanied with a significant increase in sedition insurgency in Mantristan by Malek fundamentalists. Unsure about the ability of the military government to be able to defend its installations, and to prevent nuclear capabilities to fall into the hands of religious fundamentalists, the US and UK governments persuaded the military dictator of the country to form a coalition between the three counties to protect the nuclear installations from falling into the wrong hands.

As part of this operation, US and UK military forces have taken over the protection and security of the nuclear installations, and the townships associated with the installations. Each such installation has a dedicated township which houses the employees of those installations. As part of this coalition, US and UK troops have also accompanied Holistan forces into operations against fundamentalist insurgents in the regions of Mantristan and Bhalustan.

Several sections of Holistan society have been resentful of the coalition arrangement, and view this as a half-disguised reemergence of colonism. While no overt objections have been made by any of the official agencies, several members of the Holistan armed forces are known to view this as a violation of their sovereignty. The president of Holistan welcomes the presence of coalition forces which help him in curtailing the power of the fundamentalists (who have maligned him as an apostate for inviting infidel forces in the region), yet resents the fact that he can not overtly support insurgents in Weightan and Rugistan, and that his nuclear installations are not totally under his control. Nevertheless, the stability and additional security brought by the engagement of US and UK forces has been welcomed. Religious leaders and a wide variety of insurgents have, on the other hand, vociferously criticized the coalition and deployment as a sell-out of the Malek principles, and have vowed to keep on fighting until the western forces are driven off the land of Holistan and nuclear installations returned fully under the control of Malek people. They have threatened to launch peaceful demonstrations as well as guerilla attacks on the coalition forces. Although all coalition members, US, UK and Holistan, have stated repeatedly that the goal of the coalition is limited to safeguarding the nuclear installations and stabilizing the country, the insurgents and fundamentalists have ignored or dismissed these statements. Some of the employees of the nuclear installations are also known to be sympathetic to the causes of the insurgents.

In this theater of operation, we now discuss some vignettes and describe how these vignettes have been used to derive the technical requirements of our research program.

## IV. VIGNETTE 1: SEARCH AND RESCUE IN URBAN TERRITORY

In order to win hearts and minds of the local population, the US and UK forces conduct frequent humanitarian missions in areas neighboring the nuclear industry townships. While supervising a mobile health clinic set by the coalition forces, the team receives notification that a US Special Forces helicopter has crashed near a site of cultural significance. The area around the crash site is known to have a nucleus of fundamentalists opposed to the coalition presence on Holistan soil. The helicopter crew is moving from the immediate vicinity of the crash site to a predefined emergency rendezvous. Members on the clinic are tasked to change their current mission objectives to join the Special Forces at the rendezvous point and ensure their safe passage to the base. Sensor networks in the immediate vicinity of the crash have been detected and destroyed by insurgent forces, and the terrain of the area prevents any other form of communication near the crash site.

For effective execution of this rescue, the US/UK future defense visions require the infrastructure in the field to be aware of the goals of the tactical operation being performed by the rescue team. Once the rescue mission is defined, the networking assets in the field should reconfigure themselves to optimally support the operation of the mission. Any assets in the field must be rapidly redeployed and configured to support the mission. The trust and security of devices available to the rescue team for its operations should be ascertained, and appropriate privileges assigned to rescue team members to access those devices. Any information needed for the mission should be transparently available to the rescue team members by downloading into their devices when they are in areas of good network connectivity before they reach into the rendezvous area where connectivity might be unavailable or compromised. The security policies and access to information required by the rescue force must be altered dynamically as the objective of the team changes. Cultural sensitivities must be taken into account to develop a complete situational awareness of the operation and contact with health clinic personnel must be maintained to prevent them being drawn into any engagement.

A key decision involved in the rescue operation is the level of engagement desired from the local Holistan authorities (military as well as civil). If the Holistan local authorities are sympathetic to the fundamentalists, engaging them would endanger the members of the Special Forces, whereas if they end up cooperating, their involvement could help the situation significantly.

Accordingly, we can identify the following technical requirements of the various technical areas of ITA from this vignette:

**Network Theory**: (i) A theory must be developed that can characterize the impact on computer communication effectively when significant portions of a mobile ad-hoc network are destroyed. (ii) A theory of self-management and self-configuration of mobile networks needs to be developed to adapt to sudden failures.

**Security**: (i) An architecture for managing security policies across coalition operations that can adapt to dynamic changes needs to be developed (ii) A quantification of trust and risk among different security actions needs to be developed; and (iii) A mechanism to conserve battery power while Special Forces communicate with other coalition members needs to be developed.

**Sensor Information Processing**: (i) Algorithms that allow proactive deployment of required information of a mission to the rescue team need to be developed; (ii) Algorithms to assess the quality of information from different sensor networks in the field (some of US, some of UK, some owned by Holistan authorities, and some possibly tampered by insurgents).

**Distributed Coalition Planning**: (i) Schemes to assess reliability of different members need to be developed (ii) An understanding of cultural impact of any operation needs to be developed; and (iii) planning models that allow examination of different approaches in the field need to be developed.

## V. VIGNETTE II: DIRTY BOMB

Coalition intelligence agencies have received information about a plan to smuggle nuclear material from a facility in Holistan across the border into Rugistan to carry out a dirty bomb attack in the capital. This plan has been named by the insurgents as ‘The sword of Jihad’.

The Human Intelligence (HUMINT) source of this information is well placed and has good provenance to date. Signal Intelligence confirms the surge in communication between suspect elements in both countries and other HUMINT sources confirm that known suspects and unknown personnel are gathering in both urban and rural locations along the proposed route across the border and into Rugistan.

Co-operation with the coalition intelligence community and the Holistani Intelligence Service confirms that there are elements within a nuclear facility in Holistan that are sympathetic to the insurgents cause and actively support them in the financial, logistical and personnel areas. Holistani Intelligence Service agrees to carry out surveillance operations in order to track all movement of the insurgents and their allies. All information and intelligence from these operations will relayed via a real time link to the coalition intelligence community.

The Joint Coalition Commander, with the support of the coalition intelligence community, directs that an intervention operation to stop the attack in the capital by the insurgents be carried out. Final approval of the operational plan will lie with him. The commander’s intent is “to mount an intervention operation in order to deny the insurgents the opportunity to carry out ‘The sword of Jihad’.’’

The plan consists of four phases: (a) deployment of a sensor network, intelligence teams and other assets to track any movement of radioactive material away from the installations; (b) planning for an intervention force to intercept any bomb delivery in rural as well as urban areas around the installation (c) capture of the involved insurgents by the intervention team and (d) recovery of the nuclear material in a safe manner. In order to carry out all of the phases successfully, the following technical challenges will need to be solved in the various technical areas:

**Network Theory**: (i) An approach must be developed for rapid deployment of communication services in locations without significant infrastructure presence. (ii) Self-organization of assets in order to support a mission at hand needs to be developed.

**Security**: (i) The trust and risk relationships among different coalition partners and their systems ought to be established and (ii) automatic security policies deployments enabling exchange of information to all of the coalition members in support of missions need to be developed;

**Sensor Information Processing**: (i) Sensor deployment techniques for best tracking of dirty bombs needed to be developed and (ii) trustworthiness and quality of information of sensors deployed into various contexts need to be determined and propagated.

**Distributed Coalition Planning**: (i) An analysis of the cultural differences between the different coalition members needs to be taken (ii) differences in command and control procedures of different members ought to be analyzed and any risks discovered ought to be planned against; and (iii) different possible collaboration models among the coalition partners need to be identified and analyzed.

## VI. VIGNETTE III: TRACKING INSURGENTS

Coalition intelligence agencies have been informed that a covert operation to cause unrest in the civilian population has been put into place. The town of VritaSajan, a uranium enrichment site, is the intended target. The plan is to foment riots among the two sects of Malek religion in VritaSajan township, and use the rioting as a background to incite a mob to storm the enrichment facility. The protective measures taken by US or UK forces will be used as a pretext to launch a civil disobedience movement forcing the military dictatorship of Holistan to request withdrawal of western forces from the nuclear facilities and turn them back over to Holistani authorities. The plan is to be initiated by the infiltration of some covert insurgent leaders, but the adequate physical description of those leaders is not available.

In order to thwart this insidious and long-term plan by the insurgents, the coalition commander has directed an operation to track any suspicious activities among the population of the nuclear township. The commander's intent is to identify any rabble rousers and have them be deported from the neighborhood. The public face of operations will be conducted by the Holistani forces, but the intelligence and decision planning support would be provided by the US and UK coalition forces.

The plan to counter this threat consists of three phases (a) installation of sensors and intelligence devices which can track the movements of the people in and around the township, specially around the nuclear facilities; (b) the deployment of human intelligence operations among Holistani intelligence community so that any unusual social gatherings can be identified and its consequences analyzed; and (c) collection and processing of information available from human and sensor intelligence and checking on the status of the progress of the operation.

In order to perform the operations, the following challenges need to be address in each of the technical areas.

**Network Theory**: (i) A model for adequate communication among nodes in an urban environment without adequate communication infrastructure needs to be developed; (ii) a model for interoperation between sensors deployed by the three members of coalition needs to be developed.

**Security**: (i) The policies for engagement in difference aspects of the operation need to be translated for proper operation among the different network infrastructure; (iii) since some of the sensors are deployed into possibly hostile hands and tampered with, anti-tampering techniques need to be developed; and (iii) since insurgents are technically sophisticated, advanced models for how a deployment can be abused in various manners need to be developed.

**Sensor Information Processing**: (i) A method for tracking movements of suspect individuals, e.g. by issuing special identity cards with RFID tags, or other mechanisms need to be developed (ii) the quality of information and trustworthiness needs to be determined and (iii) the information from human intelligence must be fed to drive the operations of sensor network intelligence.

**Distributed Coalition Planning**: (i) Techniques to develop the military plan for what would normally be civilian engagement needs to be developed (ii) Analysis of the cultural reaction of different societies to specific action needs to be developed and incorporated and (iii) techniques to model collaboration among different types of people and analysis of the nuances of the various interactions need to be developed.

## VII. CONCLUSION

Predicting the future is always a difficult task, and we can not be certain that the coalition would be engaged in operations similar to the one outlined in the future. However, given the geopolitical situation in the world today, such a scenario and the vignettes outlined above do not appear to be too unreasonable, and can be conceived to occur in some real countries as well.

As the military is called upon to perform delicate operations which may not conform into neatly defined roles of conventional warfare, complex decision making situations will arise, and network centricity would be key to ensuring success in operations. The technical challenges required in addressing the vignettes described above drive interesting research problems, and the applicability of the research problems extends into many scenarios which may be quite distinct from the approach outlined in this paper.

In developing this paper, we have focused on the technical requirements that will be driven into the four technical areas of interest to the alliance, but the same vignettes can be used to drive technical requirements of other technical areas as well.

## ACKNOWLEDGMENT

The authors thank several members of the alliance for ideas leading to the scenario and vignette development and refinement including Bruce Wynn, Judith Spearing, Jitu Patel, Gavin Pearson, Jim Hendler, Thomas LaPorta and Nigel Shadbolt.

# Coordination Challenges and Issues in Stability, Security, Transition and Reconstruction and Cooperative Unmanned Aerial Vehicle Scenarios

**Myriam Abramson**[1], **Ranjeev Mittu**[1], **Jean Berger**[2,]

Myriam.Abramson@nrl.navy.mil, Ranjeev.Mittu@nrl.navy.mil, Jean.Berger@drdc-rddc.gc.ca

[1]*US Naval Research Laboratory, Washington, DC USA*

[2]*Defense Research Development Canada, Quebec, CANADA*

**Abstract**—*Through the emergence of new doctrine, stability operations are becoming a core U.S. military mission that the Department of Defense (DoD) must be prepared to conduct and support. These operations are now given priority comparable to combat operations. The immediate goal often is to provide the local populace with security, restore essential services, and meet humanitarian needs. The long-term goal is to help develop indigenous capacity for securing and providing essential services. Many stability operations tasks are best performed by indigenous, foreign or U.S. civilian professionals. Large scale disasters are an example where Stability, Security, Transition and Reconstruction (SSTR) operations can provide value to foreign governments and non-governmental institutions which are under great stress to respond in a timely and effective manner. Without the means to properly coordinate these efforts, basic assistance and relief operations would be severely impeded.*

*The use of Unmanned Aerial Vehicles (UAVs) to support Intelligence, Surveillance and Reconnaissance (ISR) is becoming increasingly important. These assets can enable the collection of needed information for the execution of a given set of tasks. In large scale operations, however, the ability for the UAVs to self-coordinate may be needed as it will be difficult for human operators to effectively control large teams of UAVs.*

*This paper will begin by introducing some of the key aspects of multiagent coordination, with a focus on the operational challenges with regard to SSTR such as disaster management response as well as UAV coordination. We will then discuss the coordination challenges and gaps in order to motivate an adaptive, multiagent based approach to coordination as well as additional opportunities for research. We will conclude with a brief summary.*

## 1. INTRODUCTION

Coordination is the cornerstone of multi-agent systems, and various theoretical frameworks and limited views toward characterizing its very essence have been proposed. Models include interdependency management-based theory of coordination, organizational structuring, reference model, and multiagent frameworks. According to the theory of coordination proposed by Malone and Crowston [7], coordination is defined as the act of managing/mediating interdependencies between activities. A dependency is a relation among activities mediated by producing or consuming resources. They identify three types of dependencies: flow dependencies (e.g. goals), in which an activity produces a resource to be used by another activity; sharing dependencies, in which multiple activities can use the same resource, and fit dependencies, where multiple activities collectively produce/consume the same resource.

Organizational structuring [6] as a framework for activity interaction aims at modeling and capturing direct supervision, standardization skills, processes, outputs, mutual adjustment; authority structure, roles and responsibilities. The reference coordination model [9] is a meta-model multi-layered structure proposed to describe various coordination models. The model hierarchically composes object and activity levels, and an activity management level (described through a set of rules, specific mechanisms, programs or a selection of interaction patterns), ultimately leading to a meta-model defining an emergent coordination model, and then to a more abstract level (meta-meta model of coordination models) defining the so-called reference model. It contains terminologies and other concepts required to describe these coordination models. On the other hand, Tolksdorf recognizes the lack of consensus on the relations between coordination, communication and cooperation and outlines the need to work towards a standardized terminology which contains definitions and clarifications of basic notions including the term "coordination". Agent framework [8] is also presented to form "the foundation for the development of a complete theory of coordination". The authors suggest that future extensions to the proposed formalism should evolve toward providing formal proofs of the best coordination schemes associated with different scenarios, and develop formal methods to derive coordination mechanisms suitable for any given scenario based on the interdependencies among agents.

### 1.1 Coordination Taxonomy

Recently, Storms and Grant [10] proposed a simple taxonomy for coordination, capturing relevant, but basic properties relating to some popular metaphors

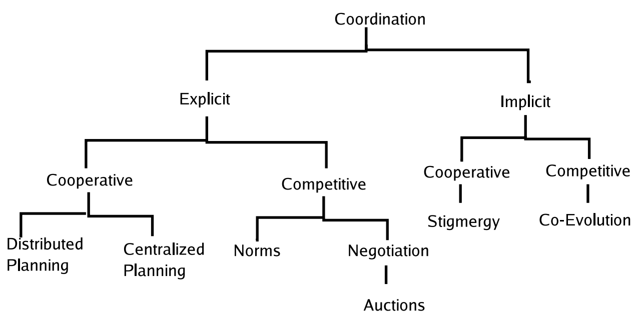


*Figure 1: Coordination Taxonomy and Related Coordination Mechanism*

to describe currently known approaches. Figure 1 shows a compact taxonomy highlighting these properties while exhibiting links to computational approaches based on those metaphors.

Coordination may first be explicit or implicit referring to communication. Implicit coordination is based on predefined or learned agreements shared by interacting agents as opposed to explicitly resorting to communication (explicit) means to mediate interactions. Agreements may derive from social laws or conventions (means of managing commitment in changing circumstances) in which agents operate under local sensing and control allowing information-sharing and interaction through multi-level pattern (intent, plan) recognition and local environment changes (markers). Coordination may be cooperative or competitive in terms of agent behaviours. Cooperative behaviours specify a common shared goal whereas competitive or self-interested attitudes emphasize individual goals. We contend that possible state-dependent behaviour/attitude may coexist at the same level as well, leading to a third mixed or semi-cooperative form. Coordination may be static or dynamic. In addition, coordination may be centralized or decentralized in which single (dedicated agent with specialized coordination capability) or multiple (e.g. all agents having coordination capabilities) entities are responsible to mediate interactions, defining the control property. Finally coordination strategies may be static or dynamic, that is, determined at design time or at run-time respectively.

### 1.2 Coordination Metaphors and Mechanisms

Based on that taxonomy, a variety of well-known metaphors for agent behaviour and communication toward coordination may be easily mapped, such as organizational (authority structure, role - cooperative), biological (living systems, colony/swarms, stigmergy – cooperative) and market (negotiation, auction, mechanism design – competitive). A real-world problem domain involving systems with specific organizational and problem decomposition structures and constraints may also expand complexity to multi-level and cross-level coordination issues, resulting in the composition or combination of coexisting metaphors exhibiting a variety of properties.

Widely used coordination mechanisms can be generally summarized in various classes and variants [11]. Such mechanisms include organizational structuring, defining social laws, agent responsibilities, capabilities, authority relationship, connectivity and control flow; market-based (negotiation, auction variants, mechanism design, argumentation); contract net, where a manager assumes the role of dividing a problem into sub-problems and searching for contractors to tackle them (bid), then evaluates bids, select and awards contracts; stigmergy (interaction between agents through their environment (markers recognition), emergent behaviour/intelligence - ant colony, swarms; as well as frameworks (distributed constraint satisfaction and/or optimization, decision theory and reinforcement learning, co-evolution, etc.). Any alternate interaction protocols/schemes may ultimately be derived or inspired from those variants.

Despite all proposed frameworks, a unified approach for coordination remains elusive as there is still no single best way to coordinate due to problem space properties, domain, system and state characteristic dependencies, required frequency of interaction and, respective intrinsic strengths and weaknesses of various approaches.

### 1.3 Coordination Metrics

Because coordination is an emergent property of interactive systems, it can only be measured indirectly through the performance of the agents in accomplishing a task where a task is decomposed in sub-goals. The more complex the task, the higher the number of sub-goals needed to be achieved. While performance is ultimately defined in domain-dependent terms, there are some common

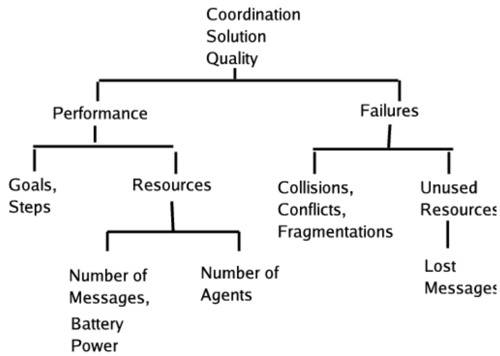


*Figure 2: Coordination Quality Metric*

characteristics. Performance in a task can be measured either as the number of steps taken to reach the goal, i.e. its time complexity, or as the amount of resources required, i.e. its space complexity. An alternative evaluation for coordination is the absence of failures or negative interactions such as collisions or lost messages. Figure 2 illustrates a simple taxonomy of coordination solution quality in pursuit games. A coordination metric can be obtained using multiple attribute decision-making methods such as a harmonic mean of appropriately weighted goals achieved, resource expanded, and conflicts [1] or a linear weighting combination of resource expanded and conflicts to evaluate coordination costs alone [4]. To show the scalability of a solution, the evaluation must linearly increase with the complexity of the task [2].

## 2. COORDINATION CHALLENGES AND ISSUES

We now briefly describe the SSTR and UAV problem domains, and then discuss coordination challenges and issues in these domains, in order to motivate an adaptive multi-agent based approach to coordination.

### 2.1 SSTR Example

Through the emergence of new doctrine, stability operations are becoming a core U.S. military mission that the Department of Defense (DoD) must be prepared to conduct and support. These operations are now given priority comparable to combat operations. The immediate goal often is to provide the local populace with security, restore essential services, and meet humanitarian needs. The long-term goal is to help develop indigenous capacity for securing and providing essential services. Many stability operations tasks are best performed by indigenous, foreign or U.S. civilian professionals [13]. Large scale disasters are an example where SSTR operations can provide value to foreign governments and non-governmental institutions which are under great stress to respond in a timely and effective manner. Without the means to properly coordinate these efforts, basic assistance and relief operations would be severely impeded.

By definition, SSTR operations are conducted outside the boundaries of US lands and territories. While there are similarities at the systems level for the employment of automated information systems regardless of whether the operations are conducted outside US boundaries or domestically for homeland defense (Defense Support to Civil Authorities), there are generally more legal restrictions that must be considered when DoD is responding domestically. This includes a distinction between National Guard forces that are acting in a State role on orders from their Governor (Title 32), and those that have been called-up in a Federal role (Title 10) on orders from the President. This also includes restrictions on the collection and sharing of law enforcement data and intelligence related information between other Federal Agencies and DoD. For these reasons we will limit our scope to examples of military operations outside of US borders.

The U.S. military may be tasked to lead and manage efforts involving non-DoD participating partners, which may include select military units of other nations and/or non-governmental organizations (NGO) such as the United Nations, Doctors Without Borders, International Red Cross/Red Crescent, and other international relief organizations.

Large scale natural disasters are one example where proper coordination could provide value. Notional examples include

> ***Disaster Relief:*** *Following a tsunami in the western Pacific, the U.S. Navy is appointed Combined/Joint Task Force Commander for disaster relief operations involving an island nation that experienced severe destruction from several 50-foot waves. Coalition partners include naval elements from various Pacific Rim nations, e.g., Australia, Thailand, Japan, China, South Korea, and India. Ground/air elements*

*from these same countries are involved in delivering relief supplies and distribution of those supplies is being managed by a combination of efforts by the host nation, the World Bank, USAID, and international relief organizations such as the Red Cross.*

***Humanitarian Assistance:*** *Following a period of severe drought and dislocation of local peoples, the U.S. Army is appointed Combined/Joint Task Force Commander for humanitarian assistance operations in a region of sub-Saharan Africa. Coalition partners include the United Nations, Doctors Without Borders, and the International Red Cross.*

These examples demonstrate the range of SSTR operations. Finding a unified approach is a key problem that is particularly acute where a cooperative approach in the preparedness phase has to be complemented with a competitive approach in the response phase due to life-threatening situations.

The National Response Plan [12] is used by Federal agencies and departments domestically and not for SSTR operations, but it provides a national-level framework that could bridge other coordination gaps that exist for an international response. The NRP provides a unified framework with detailed protocols for a comprehensive approach to all phases of disaster management, namely preparedness, prevention, response, recovery and mitigation. Those guidelines seek to improve the coordination and integration of federal, state, local and private sectors and incorporate lessons learned and best practices. The coordination efforts are as follows:

***Coordination of plans:*** *To execute mitigation efforts of future disasters.*

***Coordination of public information:*** *To combat fear and the spread of misinformation.*

***Multi-agency coordination system:*** *Between public health, housing and transportation agencies, etc.*

Computational research issues in this framework involve multiagent planning, replanning and scheduling between heterogeneous coordination entities. The context of a plan ensures that the desired results will be obtained with minimal costs. Distributed techniques such as automated plan merging and negotiation tools between responders may resolve local conflicts issues without an entire replanning effort. The degree of interdependence (coupling) in capabilities and resources is a factor in the complexity of the coordination task. While coordination tools have been directed towards assisting human-to-human collaboration, agents can be introduced to reduce interdependence by providing fast and robust solutions bypassing delays in human response such as information gathering tasks. Specifically, coordination support assistant agents can help incident commanders in directing large-scale teams and gather information for situational awareness. Human-computer interactions have also become critical in flexible robot-agent-person teams to smooth out the cognitive demands of such interactions.

### 2.2 UAV Example

Network centric automated decision support capabilities for operations and mission planning in tactical military domains and environments may involve a heterogeneous group of sensors and effector agents drawn from distinct classes. These assets are generally engaged over a variety of mission tasks including ISR and response/service tasks evolving in a potentially dynamic, uncertain, dense and congested environment with both known and unknown targets and threats (a mix of moving/static, evading/non-evading behaviors).

These “agents” must cooperatively and/or non-cooperatively search and act on the environment to carry out a collection of distributed continual planning ISR and response/service management tasks. These include information gathering, exploration, target search: detect, locate, track, identify, classify/confirm, assess outcome, monitor, track and move, engage, destroy, etc.

Tasks may be naturally determined or dynamically dictated as a result of agents’ actions, emerging goals or changes in current state estimation, requiring proper dynamic resource management and coordination. It should be noted that picture compilation and exploitation are not mutually exclusive or loosely coupled, and interdependencies due to resource contention or goal dependencies may generally be quite complex. For instance, a distributed information gathering task may explicitly serve the purpose of picture compilation. This would help in further refining the strategy used to collect additional information needed for continual refinement of the picture. A reconnaissance mission is such an example, in which shared cognitive maps translating probability of target/threat locations or identity declarations may be exploited to optimize heterogeneous resource allocation in gathering additional information while updating/improving state estimation (picture compilation quality) in dynamic uncertain environments. The same observation on resource sharing and goals interdependencies holds for inter- and intra-picture exploitation tasks.

In these problems, resources must be allocated and coordinated in a timely manner to dynamically schedule and visit targets/threats, determine suitable routes among obstacles and manage airspace utilization and resource sharing.

A key enabler of a sustainable military force is the notion of a tiered system. A tiered system is an integrated, multi-tier intelligence system encompassing space and air-based sensors linked to close-in and intrusive lower tiers. The lower tiers (e.g., UAVs) are not only the critical source of intelligence; they can also serve as a key cueing device for other sensors. There is active research and exploration within the US DoD to understand the technical challenges in building tiered systems.

Multiagent (human and computational, cooperative, self-interested, or a mixture of both) coordination to achieve coalition formation, task allocation, path planning and other activities represent key areas to be explored. In that respect, coordination through learned behaviors and through human interactions offers a major challenge.

## 3. CHALLENGES AND GAPS

There exist similar technical challenges with regard to coordination in both problem domains, such as cooperative information-sharing in partially observable dynamic environments. As an example, in SSTR operations the communications infrastructure may be severely degraded or completely destroyed, preventing the first responders to effectively communicate. Similarly in scenarios requiring multiple UAVs to coordinate, distance and environment factors may prevent reliable communication.

Recent technological advances in mobile ad-hoc networks (MANET) are key enablers in the deployment of net-centric cooperative multiagent systems on the battlefield and in natural disaster areas. The limited communication range in MANET provides only a partial knowledge of the global environment but is not necessarily restricted to the immediate neighbors. Those constraints make it advantageous for agents to self-organize within their communication range using multicast, while the absence of centralized control requires a distributed control policy to manage joint distributed beliefs. The uncertainty that a message will arrive at its destination in a finite amount of time violates one of the basic communication assumptions of distributed constraint satisfaction algorithms [5]. How to extend those algorithms to open and uncertain environments is still an active area of research [3]? Coordination strategies have to be robust against message loss and equipment failures. The concept of network-aware coordination, in which agent-based coordination algorithms can utilize network state information in order to communicate more effectively by understanding each others communications constraints, is an area that has not received much attention. Additionally, human-computer interactions have become critical in flexible robot-agent-person teams to smooth out the cognitive demands of such interactions and need to be explored further.

Some deficiencies in surveillance and reconnaissance persistence, penetration and identification, battle damage assessment, and data processing, exploitation, and dissemination are due to serious limits [assets] to penetrate foliage, track individuals, identify Weapons of Mass Destruction components, defeat camouflage, and identify decoys. Dealing with these surveillance and reconnaissance challenges will require lower tiers (UAVs) of close-in and intrusive sensors. However, even as the DoD becomes more dependent on networked C3ISR, no dedicated ‘red team’ effort exists which concerns itself with camouflage, concealment, and deception; vulnerabilities; and tactics which might be used by adversary against our emerging C3ISR system. This is an area where recent advances in game theory can play a significant role in understanding adversarial behaviors, which can be encoded in simulations to aid in the development of tiered systems, particularly from the perspective of how these assets will coordinate in response to such behaviors.

It should be noted that tiered-system components such as UAVs or space-based assets are not only useful for ISR activities supporting more traditional combat operations, but may also enable effective SSTR operations.

Given the diversity of the assets, and the fact that coordination must be achieved both in the horizontal and vertical planes, and the environments in which the components of a tiered system will operate; it is not likely that a single coordination approach or even a family of coordination approaches will work well from a static perspective. It is more reasonable to expect that systems should learn which approaches work well and under which circumstances, and adapt appropriately.

## 4. TOWARDS ADAPTIVE MULTI-AGENT SYSTEMS COORDINATION

A suitable framework (or multiple frameworks) is required to address current challenges and issues in agent-based coordination. The proposed multiagent coordination approach should be flexible enough to adequately address resource constraints imposed by limits in the communication, computational and temporal dimension (should exhibit adaptability in time-constrained environments); handle information constraints such as security and privacy in information exchange; permit run-time reasoning regarding the selection of particular coordination mechanism/protocol; tradeoff between the cost of reasoning versus value of coordination, and attempt to dynamically choose between centralized and decentralized mechanisms.

The framework should support the investigation of coordination concepts in net-centric problem settings/environments. It should provide flexibility for problem definition, and allow for studying different concepts, including models, algorithms, or agent-mediated decision support capabilities. The framework should permit basic simulation in order to validate advanced multi-agent coordination concepts in order to asses the value of coordination.

## 5. CONCLUSION

Coordination is a key requirement underlying distributed continual planning to satisfactorily improve net-centric decision support components characterizing dynamic planning and execution. In this paper we briefly overviewed the basic elements and aspects of coordination and focused on some of the issues, gaps and challenges lying ahead for the defense research community. As a result, research areas to be further investigated have been identified in relation to SSTR such as disaster management response and the cooperative UAV problem domains.

### References


[1] M. Abramson, W. Chao, and R. Mittu, Design and Evaluation of Distributed Role Allocation Algorithms in Open Environments, International Conference on Artificial Intelligence, Las Vegas, NV, 2005.

[2] E. H. Durfee, Scaling up Agent Coordination Strategies, IEEE Computer, 2001.

[3] P. J. Modi, Distributed Constraint Optimization for Multiagent Systems, PhD thesis, USC, 2003.

[4] A. Rosenfeld, G. A. Kaminka, and S. Kraus, Adaptive Robotic Communication using Coordination Costs for Improved Trajectory Planning, in Proceedings of the American Association of Artificial Intelligence, 2006.

[5] M. Yokoo, Distributed Constraint Satisfaction, Springer-Verlag, 1998.

[6] Henry Mintzberg, "The Structuring of Organizations: A Synthesis of the Research, Prentice Hall, Englewood Cliffs, N. J., 1979.

[7] Thomas W. Malone and Kevin Crowston, "What is Coordination Theory and How Can it Help Design Cooperative Work Systems", In Readings in Groupware and Computer-Supported Cooperative Work, Assisting Human-Human Collaboration, Ed. Ronald M. Baecker, Morgan Kaufmann Publishers, Inc., SF, California, 1993.

[8] Riyaz Sikora and Michael J. Shaw, "A Multi-Agent Framework for the Coordination and Integration of Information Systems", In Management Science, Vol. 11, Issue 11, Part 2 of 2, November 1998.

[9] Robert Tolksdorf, "Models of Coordination", In Engineering Societies in the Agent' World, A. Omicini, R. Tolksdorf, and F. Zambonelli, Editors, ESAW 2000, LNAI 1972, Springer-Verlag, Berlin Heidelberg, 2000.

[10] P.P.A. Storms and T.J. Grant. Agent coordination mechanisms for multi-national network enabled capabilities. In Proceedings of the 11th International Command and Control Research and Technology Symposium (ICCRTS) on Coalition Command and Control in the Networked Era, Cambridge, UK, Sept 2006.

[11] Gerhard Weiss, "Multiagent Systems and Distributed Artificial Intelligence", In Multiagent Systems – A Modern Approach to Distributed Artificial Intelligence", Gerhard Weiss, Editor, the MIT Press, Cambridge, Massachusetts, 1999.

[12] US Department of Homeland Security, "National Response Plan", [Web Site] http://www.dhs.gov/xprepresp/publications, last accessed 01/23/2007.

[13] US Department of Defense, "DoD Directive 3000.05-Military Support for Stability, Security, Transition, and Reconstruction (SSTR) Operations", http://www.dtic.mil/whs/directives/corres/html/300005.htm, last accessed 02/07/2007.

# AKTiveSA: Supporting Civil-Military Information Integration in Military Operations Other than War

Paul R. Smart, Alistair Russell and Nigel R. Shadbolt
School of Electronics and Computing Science, University of Southampton, Southampton, SO17 1BJ, UK
{ps02v, ar5, nrs}@ecs.soton.ac.uk

**Abstract**—*The ability to integrate information from a variety of sources is a key ingredient of enhanced situation awareness in both conflict and non-conflict situations. Information integration is particularly important in Military Operations Other Than War (MOOTW) contexts because the range of relevant information sources is considerably more diverse than that seen in more conventional war-fighting operations. In this paper we provide an overview of a technical demonstrator system (the AKTiveSA TDS), which was developed as part of the UK MoD's Defence Technology Centre (DTC) initiative. The demonstrator combines a variety of semantic technologies with advanced modes of visualization and interaction in order to highlight how extant Semantic Web technologies can be used to improve situation awareness and facilitate information integration in a simulated humanitarian relief scenario.*

## 1. Introduction

Military Operations Other Than War (MOOTW) present a number of challenges to military agencies in terms of both information exploitation and inter-agency collaboration. MOOTW operations often require the integration of information from a variety of physically disparate and semantically heterogeneous information resources and, in many MOOTW operational contexts, such as humanitarian relief, military agencies are often required to work alongside other, non-military agencies, which raises issues concerning the effectiveness of inter-agency communication and information exchange mechanisms. In the context of the UK Data and Information Fusion (DIF) Defence Technology Centre (DTC) initiative we have been engaged in a program of research, called AKTiveSA, which aims to investigate issues relating to the effective inter-operation of both military and non-military agencies in a variety of MOOTW contexts. As part of our work we have developed a simulated humanitarian relief scenario to support subsequent knowledge engineering activities and also to provide a basis for technical demonstration. The scenario features a number of humanitarian incidents based around an earthquake in the Ghazni region of Afghanistan, and these events are superimposed on a backdrop of ongoing military conflict involving air strikes and ground manoeuvres by coalition forces against elements of the former Taliban regime. Formal ontologies, developed using the Web Ontology Language (OWL), were used to support the semantic annotation of various scenario-relevant information resources and to provide a representational substrate for inference-mediated information fusion processes. In addition to providing an effective solution for search and retrieval operations using semantic query languages such as SPARQL[1], formal ontological representations were also used to support a number of visualization capabilities including the generation of MIL 2525B military symbols [1] and map overlays that juxtapose both military and non-military information. In this paper, we describe the ontologies that were developed in the context of the project and demonstrate how these can be used to support integrated information displays in the context of a technical demonstrator system: the AKTiveSA TDS. The system described herein exploits a variety of semantic technology components to highlight strategies for semantically-mediated information integration and display, especially in situations where information about the disposition of coalition military assets needs to be aligned with situation-relevant civil-intelligence. Although, our development work with respect to the AKTiveSA TDS is ongoing, this paper provides an overview of current system capabilities and describes how these capabilities will be extended in future development efforts. The overview of system capabilities presented here extends an earlier characterization of a prototype system, which was described in Smart et al [2].

## 2. Ontology Development

The basis of our approach in developing semantically-mediated approaches to information integration consists in the provision of formal ontologically-motivated characterizations of the problem domain. Due to the desired support for a variety of knowledge-oriented processes, we required the use of a suitably rich and expressive medium for knowledge representation. Recently, attempts to provide a set of representational formalisms for the communication of ontological structures within the framework of the Semantic Web have coalesced around the Web Ontology Language (OWL) [3, 4]. We elected to use this language as the representational medium for our ontology engineering activities, in part due to its endorsement by the World Wide Web Consortium (W3C), its close alignment with Resource Description Framework (RDF) and its level of semantic expressivity (which supports a variety of types of automated reasoning, e.g. subsumption reasoning).

A number of ontologies were developed as part of the knowledge engineering initiative for the AKTiveSA project. These included, but were not limited to:

[1] http://www.w3.org/TR/rdf-sparql-query/

1. **Geography:** This ontology deals with all the geographical aspects of the problem domain. It encompasses a wide variety of conceptualisations including terrain features, transport routes, rivers, shorelines, terrain elevation data, etc.

2. **Transportation:** This ontology covers all aspects of transportation in the problem domain. This overlaps, to some extent, with the geography ontology in the sense that transportation routes, e.g. airways and roads, may also be considered elements of the geographical (geo-spatial) domain.

3. **Meteorology:** This ontology deals with all aspects of the climate and weather. The meteorology ontology is important in enabling the system to interpret and utilize information derived from local weather reports and forecasts as well as long term data about regional rainfall, snowfall, seasonal temperature, etc.

4. **Humanitarian Aid:** This ontology covers information of relevance to humanitarian operations, i.e. humanitarian hazards (e.g. floods), humanitarian organizations, humanitarian aid programs, humanitarian aid workers, etc.

5. **Military Entities:** This ontology includes all relevant conceptualisations in the military domain, including tactical operational areas and zones, military platforms, intelligence information, weapons, etc.

The results of the knowledge engineering initiative were presented in the form of a knowledge web (see Figure 1), which provided a web-based medium for browsing both domain ontologies and knowledge sources used as part of the ontology engineering initiative. The knowledge web was used by military Subject Matter Experts (SMEs) as part of the knowledge validation process.

## 3. Technical Demonstrator System

To showcase the role played by semantically-enriched representations in supporting information integration and situation awareness, we developed a technical demonstrator system using a variety of technology components. These components are described in subsequent sections.

### 3.1. Knowledge Repository

The AKTiveSA Knowledge Repository is a key element of the AKTiveSA TDS and corresponds to an instance of the AKT 3Store [5]. The 3Store combines an RDF tripleStore with a SPARQL query engine that promotes the efficient storage and retrieval of RDF metadata. The 3Store is implemented on top of a MySQL database engine, which can be manipulated using conventional queries formulated in SQL. However, in order to provide more sophisticated query capabilities, the 3Store also incorporates a SPARQL interface. The 3Store SPARQL engine transforms a

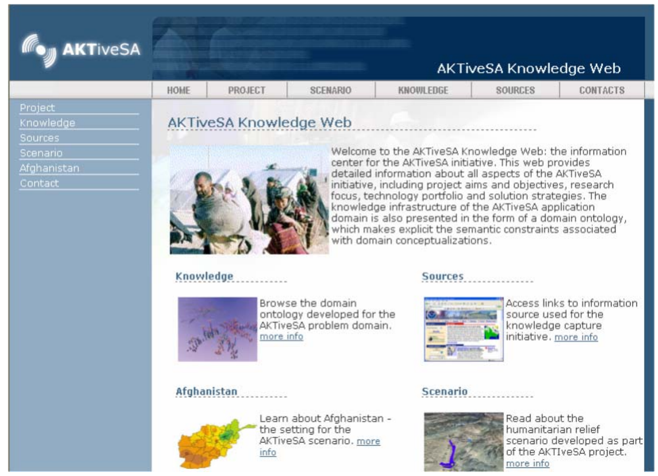


**Figure 1:** AKTiveSA Knowledge Web

SPARQL query into a SQL query, which can then be executed against the relational database representation of the RDF data to return a query result. The AKTiveSA Knowledge Repository is used as the storage medium for the entire knowledge infrastructure of the AKTiveSA application domain. It contains all the ontologies and meta-data associated with the application, including both schematic knowledge (classes) and knowledge objects (individuals).

### 3.2. Military Symbology Browser

To provide a common representation of the operational environment, and to support a common understanding of the situation picture, we aimed to exploit standard military symbols for the graphical representation of military entities. The MIL 2525B symbol specification [1] was used as the basis for creating military symbols using ontological characterizations of specific military entities. Essentially, all military entities defined within the scenario were associated with ontology elements in the Knowledge Repository, and this information served as the basis for the dynamic creation of graphic symbols via a web service interface. The web service in this case accepted, as input, a Uniform Resource Identifier (URI), which uniquely identified the target entity in the context of the AKTiveSA ontology infrastructure. The

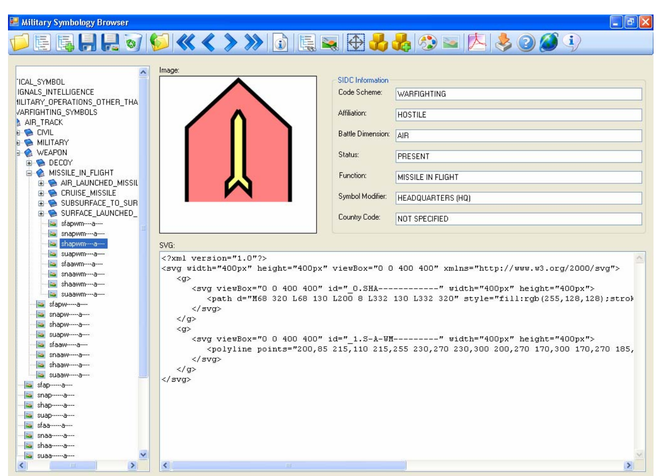


**Figure 2:** Military Symbology Browser

service then retrieved relevant information about the entity from the Knowledge Repository and used this information to determine the Symbology Identification Code (SIDC) for the entity[2]. This information was in turn used to programmatically create an appropriate graphic symbol using the Military Symbology API, a code component specifically engineered for military symbol generation in the context of the AKTiveSA project. Once created, the web service returned an image to the AKTiveSA Client Application (see Section 3.5) for display in the user interface. The technology portfolio of the AKTiveSA initiative includes the Military Symbology Browser (see Figure 2), a stand-alone application that exposes the functionality of the aforementioned Military Symbology API. This application demonstrates the automatic generation of military symbol hierarchies, the use of Scalable Vector Graphics (SVG) to represent graphic elements, the alignment of graphic symbols with SIDC information, and the automatic rendering of SVG elements to image files using GDI+. The application and the associated web service wrapper to the Military Symbology API could be used, in a highly reusable fashion, to support ontology-driven military symbol generation in a variety of application contexts.

### 3.3. AKTive8 Semantic Web API

Many of the knowledge and reasoning capabilities of the AKTiveSA TDS rely on the AKTive8 Semantic Web API. This is a .NET class library that provides a generic and reusable framework for processing Semantic Web data. It was developed in the context of the AKTiveSA initiative and encapsulates the functionality for knowledge representation, publishing and reasoning in the context of the Semantic Web. Within the context of the AKTiveSA TDS, the AKTive8 API is used to provide access to the AKTiveSA Knowledge Repository, to execute semantic queries, and to provide an interface between the AKTiveSA Client Application and other technology components, e.g. the AKTiveSA Scenario Generator.

### 3.4. Scenario Generator

The AKTiveSA TDS was developed to showcase how semantic technologies could be used to improve information integration in respect of civil-military operational contexts. As part of the development effort towards this goal it was necessary to define a scenario that could be used for visualization, demonstration and evaluation purposes. The chosen scenario features a number of humanitarian incidents, including the collapse of the Band Sultan dam in Ghazni province. These humanitarian incidents are superimposed on an ongoing backdrop of military conflict (based on coalition military operations against Taliban insurgents) in the South-Eastern region of Afghanistan. Although the scenario was based on real-world events, most of the information resources for our scenario were largely contrived. In particular, we manually created the web pages,

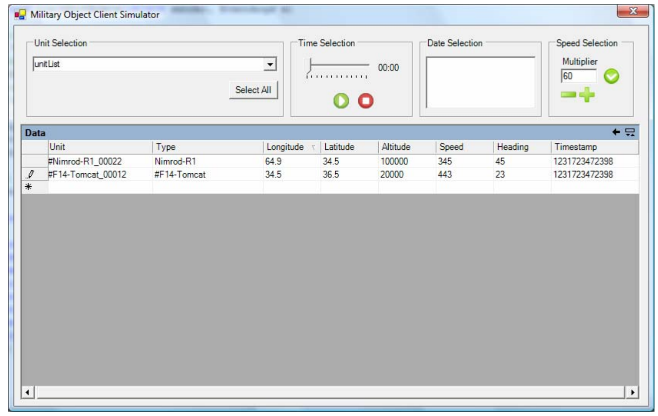


**Figure 3:** Scenario Generator Tool

text reports, RSS feeds, etc., which were necessary to support the scenario timeline. In most cases, however, these resources were derived from actual information resources publicly available via the web.

A scenario generator tool was developed to coordinate the generation of information feeds and event triggers in respect of the scenario. Figure 3 shows a screenshot of the scenario generator tool. The tool reads information from a flat relational database, which contains information about the disposition of military and non-military units. It then feeds this information into an AKTive8 API web service for use within the AKTiveSA Client Application. The scenario generator tool can be used to control the speed of the scenario timeline as well as the current position within the timeline - changes to the current time cause the Scenario Generator to update information about all scenario objects (military units, refugee convoys, etc.) visible via the user interface.

### 3.5. AKTiveSA Client Application

The AKTiveSA Client Application (see Figure 4) is the centrepiece of the AKTiveSA TDS. It hosts the NASA Worldwind[3] visualization component, which co-opts both photorealistic satellite imagery with digital terrain elevation data to provide, what is in effect, a three dimensional model of the Earth's surface. The AKTiveSA Client Application extends the Worldwind component (called the WorldView, in the context of our AKTiveSA component hierarchy) by providing a number of advanced interface components, including overlays that can host multiple of types of filtered information content (e.g. military symbols, maps, images, etc.).

*Architecture*

The AKTiveSA TDS architecture is based on a client-server model, wherein the AKTive8 API mediates access to a shared AKTiveSA Knowledge Repository that is used by (perhaps) multiple instances of the AKTiveSA Client Application. This architecture allows for the provision of shared views of the current operational picture by both military and civilian agencies (i.e. views can be shared

[2] The SIDC is a unique alphanumeric identifier, which serves to uniquely identify each symbol within the MIL 2525B symbology specification.

[3] http://worldwind.arc.nasa.gov/

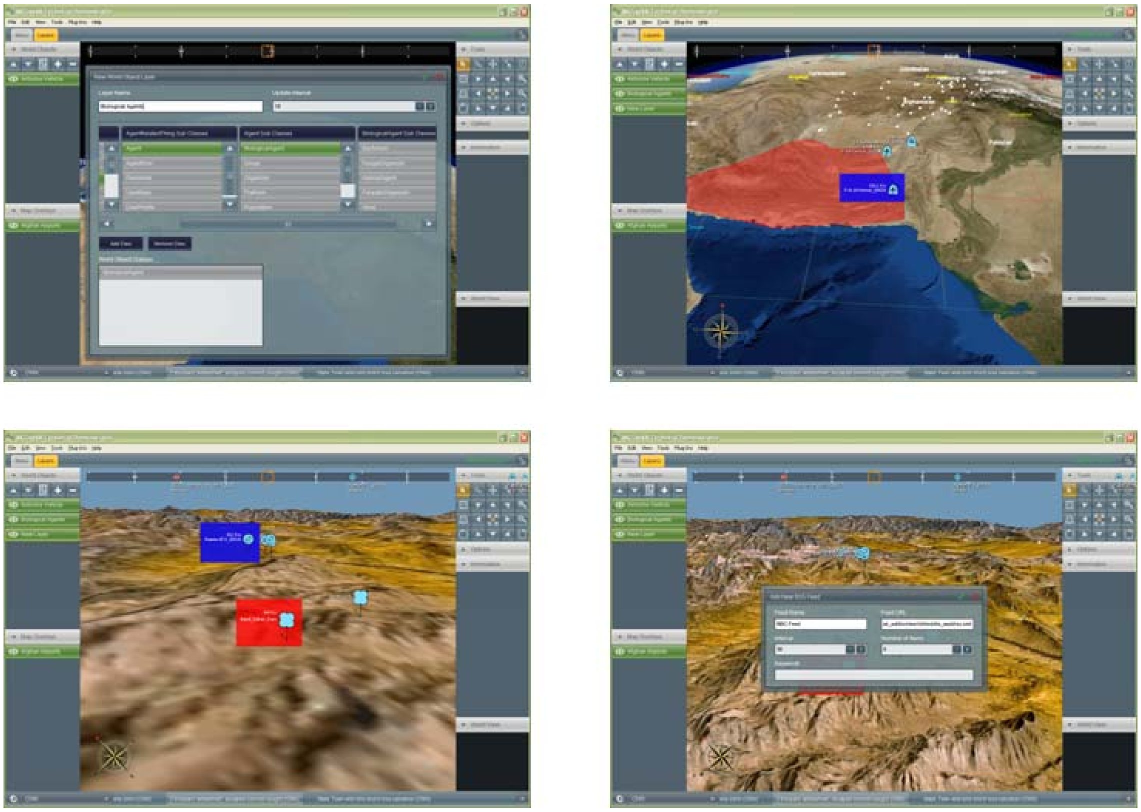

**Figure 4:** AKTiveSA Client Application

between multiple instances of the AKTiveSA Client Application running within a single networked environment).

Each application instance, in this case, provides editing capabilities, which allows authorized users to update the contents of the Knowledge Repository, thus changing the knowledge infrastructure of the situation picture.

*View Manipulation*

Each AKTiveSA Client Application provides multiple opportunities for manipulation of the WorldView. These actions enable the user to change the camera view angle, tilt-level, elevation level, etc. The current implementation assumes that user interaction is primarily mediated via standard input devices, such as keyboard and mouse; however, alternative modes of user interaction could be achieved, and in some situations these might be considered preferable. One notable example here concerns the use of large multi-touch screens which provide users with intuitive modes of interaction and, in some cases at least, also permit multiple users to work simultaneously with different parts of the application interface [6]. Touch-screen control of the user interface will often emulate that achieved using the mouse, so, for example, touching and dragging the WorldView will emulate a panning operation, while touching and dragging the fingers in a radial direction will emulate a zoom-in or zoom-out operation.

*RSS Ticker*

The AKTiveSA Client Application includes an RSS ticker control, which is designed to display information from various, user-selected news feeds. The summary for each news item is displayed in a scrolling marquee at the bottom of the user interface and each summary is associated with a 'More Information Button', which, when clicked, will open a dialog box to display the entire news article. Items that have already been displayed in this way are rendered using a white font, while items that have not previously been displayed are rendered using a red font.

*Image Overlays*

Image overlays represent visual information about areas of the terrain as rendered in the WorldView component. Image overlays can include both vector based graphics (for region-based information, such as population density or language distribution maps), or additional satellite imagery layers. Such overlays provide a means to easily extend the types of information content that can be displayed using the WorldView component.

*Temporal Dynamics*

The AKTiveSA Client Application provides two means to manipulate the scenario timeline in order to gain a view of past and future (predicted) states. The movement toolbar displays a visible movement trail, which indicates the past and predicted movement of a selected entity (e.g. a military unit) within the WorldView display (see Figure 5). This

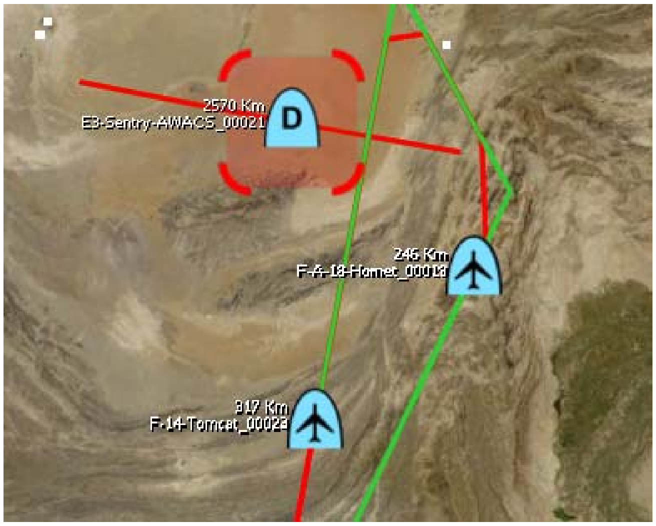

**Figure 5:** Entity Movements

visual aid facilitates an understanding of the temporal evolution of the current situation picture both in terms of past situations and (likely) future situations. A second tool, the time gesture tool, compliments the movement toolbar by allowing users to use simple input gestures to alter the current time setting. Horizontal movements from left to right across the screen, using either the mouse (or fingers on a touch-screen interface) advances the temporal position forward in time, whereas a gesture from right to left will cause the time setting to regress. Combining horizontal gestures with vertical gestures allows the speed of temporal progress to be manipulated: a movement from the bottom of the screen to the top will increase the speed at which the time changes, allowing for large jumps in temporal position; movements from the top of the screen to the bottom will decrease the speed at which the time changes, allowing for fine-tuning of the temporal setting.

*Knowledge Monitors*

Knowledge Monitors are technology components that serve as daemons responsible for the detection of knowledge-rich contingencies and the execution of knowledge processing actions. As their name suggests, Knowledge Monitors constantly monitor the evolving knowledge infrastructure of the AKTiveSA application domain. Each instantiated monitor is associated with a set of conditions and actions. When the conditions have been satisfied, the monitor invokes its associated actions, which typically serve to increase the situation awareness of the end user via alerts, notifications, status reports, emails or RSS feeds. Knowledge monitors can be created by end-users to detect and monitor interesting or important situation contingencies and they therefore serve as useful devices for event-driven processing and improved situation awareness.

## 4. CONCLUSION

The AKTiveSA initiative focuses on the development of technology to improve information integration and situation awareness in a variety of civil-military operational contexts. To this end we have adapted a range of semantic technologies to demonstrate how semantically-enriched representational schemes can be used to assimilate information, facilitate search and retrieval and deliver operationally-useful decision outcomes in the context of a simulated humanitarian relief scenario. In addition to its ability to provide a sufficiently expressive medium for the representation of knowledge-rich contingencies, we also argue that the provision of semantically-enriched information can be used to filter information so as to avoid situations of information overload that might otherwise result from unrestricted access to large-scale, semantically heterogeneous information environments.

The AKTiveSA TDS provides a platform for demonstrating semantically-mediated modes of information integration and aggregation in the context a simulated humanitarian relief scenario. Operational contexts that require civil-military cooperation necessitate the exploitation (and exchange) of information that is considerably more diverse than that seen in conventional war-fighting contexts. For example, humanitarian relief operations, like that described in our scenario, necessitate the integration and juxtaposition of a variety of forms of information, including (but not necessarily limited to) information about refugee movements, the disposition of hostile forces, the status of ongoing offensive operations, the intentions and activities of humanitarian aid agencies, weather conditions, terrain features and the navigational status of key transport routes. The exploitation of such information for the purposes of operationally-effective modes of planning and decision-making is facilitated, we argue, by the use of semantically-enriched representational schemes (that support the rapid search and retrieval of information) in conjunction with flexible modes of visualization and interaction (that present the user with an opportunity to overlay and juxtapose multiple types of information in the context of an integrated display environment).

The technology development effort associated with the AKTiveSA initiative is ongoing, and the current paper has provided only a snapshot of system capabilities at an intermediate point in the development timeline. Our future development efforts will focus on a number of enhancements to the current system, including the provision of a virtual adviser component (an animated virtual agent providing vocal prompts and alerts), a reasoning subsystem component (providing decision support with regard to selected knowledge-intensive tasks, e.g. humanitarian needs assessment), a resource annotation component (which will use natural language processing technologies to assist with the classification and semantic annotation of textual resources), and a graphical query designer component (to enable end-users to construct semantic queries in an intuitive and visually-oriented manner). These enhancements will, we hope, extend the applicability of the system to situations and domains beyond civil-military operational contexts. For example, we believe the current research could be usefully applied to any domain requiring information integration with respect to heterogeneous information sources for the purposes of enhanced situation awareness. These could include the emergency services, search and rescue operations [7], e-Health [8] and homeland security [9].

## ACKNOWLEDGMENT

This work was funded by a grant awarded to General Dynamics UK Ltd and the University of Southampton as part of the Data and Information Fusion Defence Technology Centre initiative.

## REFERENCES


[1] "Department of Defense Interface Standard: Common Warfighting Symbology," Department of Defense, USA MIL STD 2525B, 2005.

[2] P. R. Smart, N. R. Shadbolt, L. A. Carr, and m. c. schraefel, "Knowledge-Based Information Fusion for Improved Situational Awareness," presented at 8th International Conference on Information Fusion, Philadelphia, USA., 2005.

[3] G. Antoniou and F. van Harmelen, "Web Ontology Language: OWL," in *Handbook on Ontologies*, S. Steffen and R. Studer, Eds. Berlin: Springer Verlag, 2003.

[4] D. L. McGuiness and F. van Harmelen, "OWL Web Ontology Language Overview," World Wide Web Consortium, 2004. http://www.w3.org/TR/owl-features/

[5] S. Harris and N. Gibbins, "3store: Efficient Bulk RDF Storage," presented at 1st International Workshop on Practical and Scalable Semantic Web Systems, Sanibel Island, Florida, USA, 2003.

[6] J. Y. Han, "Low-cost multi-touch sensing through frustrated total internal reflection," presented at 18th Annual ACM Symposium on User Interface Software and Technology, Seattle, Washington, USA, 2005.

[7] "CoSAR-TS - Coalition Search and Rescue - Task Support Website." http://www.aiai.ed.ac.uk/project/cosar-ts/

[8] N. R. Shadbolt, P. Lewis, S. Dasmahaptra, D. Dupplaw, B. Hu, and H. Lewis, "MIAKT: combining grid and web services for collaborative medical decision making," presented at 2nd e-Science All Hands Meeting, Nottingham, England, UK, 2004.

[9] "Semantic Web Applications for National Security - Conference Website." http://www.schafertmd.com/swans/

# Evaluating the Impact of Culture on Planning and Executing Multinational Joint Force Stability, Security, Transition and Reconstruction Operations

Kenneth Sharpe, Vienna, VA (kenneth.w.sharpe@saic.com)
Keith Gremban, Englewood, CO (keith.d.gremban@saic.com)
Kimberly Holloman, Sterling, VA (kimberly.a.holloman@saic.com)
**Science Applications International Corporation**

Abstract—This paper describes a recent effort, sponsored by the Defense Research Advanced Projects Agency (DARPA), to evaluate the need for a cultural assessment tool to mitigate the negative effects of cultural differences on planning and executing multinational stability and reconstruction operations.

## 1.0 INTRODUCTION

*If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle. – Sun Tzu*

Sun Tzu's ancient words are particularly relevant today as the U.S. and coalition partners battle insurgents and terrorists in Iraq and Afghanistan and prepare to fight tomorrow's terrorists around the world. These adversaries bear little resemblance to the 'enemy' that the U.S. prepared for and trained to fight for the last fifty years. U.S. military planning tools and processes were optimized to conduct major combat operations against a Soviet threat on Northern European terrain using well equipped ground forces and superior air support.

The threat has changed. Today's adversaries and enemies use different tactics and operate in diverse and complex environments. They also live in a virtual world facilitated by global communications networks. This environment is multidimensional and more often than not, the outcomes of struggles are determined by religious, ethnic, tribal, political and/or class conflicts and allegiances rather than military strength. Increasingly, success is less about winning wars and more about winning the peace. In order to effectively combat the enemy of today and tomorrow, the government, not just the military, must develop capabilities and competencies that allow us to gain a true understanding of world views, values and beliefs.

A Combined Joint Task Force (CJTF) charged with conducting stability and reconstruction operations typically must coordinate and plan the activities of foreign military forces, foreign government organizations, as well as non-governmental organizations. Two specific capabilities and competencies are needed for a CJTF to operate effectively. First, military personnel must develop in-depth cultural knowledge and awareness of adversaries, as well as neutrals. As retired Major General Scales stated, the type of conflict we are experiencing in Iraq requires "an exceptional ability to understand people, their culture, and their motivations." [1] Today, we face a vast array of potential adversaries that differ greatly in terms of their objectives, strengths and vulnerabilities, and reflect very different cultures. While increasing the cultural awareness of service men and women is essential, it is insufficient to fill the entire cultural knowledge void in the near or long term. Not everyone has the aptitude or the time to become a cultural specialist. Additionally, cultural knowledge of one society does not apply to others that we may engage in the future. To provide tomorrow's warfighters with enhanced cultural knowledge across the spectrum of potential adversaries the U.S. military must explore alternative, tailorable solutions.

The second, key competency needed is the ability to form a CJTF quickly, potentially with unfamiliar forces, and retain the capacity to plan and execute effectively. Military personnel must be able to effectively communicate, collaborate and synchronize with personnel who have different cultural backgrounds, specializations and experiences. While cultural differences between multinational forces have always been present, in the past they were largely mitigated by alliance structures and extensive multinational training exercises. Today, extensive training with every potential partner is not feasible. The U.S. must develop alternative solutions to better develop capabilities and to overcome potential negative effects of cultural differences on CJTF operations.

This paper discusses our recent work to explore weaknesses in current CJTF planning and to take a first look at solutions which provide these two core capabilities and competencies. The Cultural Integration Experiment (CIE), sponsored by DARPA, conducted an experiment in which subject matter experts (SMEs) role-played commanders and staff of a rapidly constituted CJTF performing a mission in Africa. The CJTF effort focuses on multinational planning for Stabiliity, Security, Transition and Reconstruction (SSTR) operations. The hypothesis was that cultural factors play a critical role in multinational planning and execution of SSTR operations. Additionally, a secondary hypothesis held that, because of cultural differences, not all members of a CJTF are equally qualified to conduct different missions.

The logic of the primary hypothesis flows like this: different nationalities and sub-cultures exhibit different behaviors and capabilities based on their unique culture; logically, the military culture of these groups differ; military culture drives the way a military organizes and trains for war; these differences result in differences in the way militaries educate, field, equip, train, and employ their forces. American military culture, sometimes called the American Way of War, resulted in highly trained, flexible soldiers equipped with sophisticated equipment and robust communications that is typically employed with massive firepower and maneuver to destroy the enemy.

A UK approach, using their military culture, might feature a more subtle application of force to “persuade” the opponent to start behaving in an acceptable manner. The UK military is comprised of highly trained, professional soldiers but have fewer forces, less capacity to move massive firepower and equipment to remote locations, and generally do not enjoy the advantage of overwhelming combat force. The difference in these two military cultures was demonstrated in the manner in which the US and UK forces approached their stabilization duties in Iraq.

The CIE program posits the notion that cultural primitives exist across cultures, though each culture expresses the primitive in a different way. For example, all cultures have a concept of honor; however, the way honor is expressed in different cultures varies greatly. Another example is family: the role and prominence of the concept of family differs widely between cultures. These primitives are by nature long term and relatively stable; the specific way cultural primitives express is affected by more transient factors in the environments such as the state of the economy, a drought or famine, etc. Cultural primitives guide and shape the tendencies for behavior and attitude of a culture; however, they do not determine and predict individual behavior. Cultural primitives are accurate across large numbers but are not predictive for individual actions.

The CIE program explored these primitives and how they can be identified and leveraged to improve multinational planning and execution of SSTR operations. A specific point of interest was the use of planning tools among different nationalities of a CJTF and which, if any, cultural effects degraded smooth integration of planning tools across cultural boundaries.

During CIE, a series of limited objective experiments were conducted to explore the role of culture in multinational planning of stability and support operations: there were two with the United Kingdom, one with the Singapore Defense Force, and the previously described experiment with a CJTF in Africa. This paper describes the CIE program, its progress to date, and plans for future development.

## 2.0 BACKGROUND

The U.S. and its coalition face adversaries that challenge the ability of the military to leverage its overwhelming conventional military superiority to achieve desired effects. The U.S. National Defense Strategy (NDS) [2] identifies four types of challenges that present significant threats to U.S. national interests:

- *Traditional* challenges are posed by states employing recognized military capabilities and forces in well understood forms of military competition and conflict
- *Irregular* challenges come from those employing unconventional methods to counter the *traditional* advantages of stronger opponents
- *Catastrophic* challenges involve the acquisition, possession, and use of WMD or methods producing WMD-like effects
- *Disruptive* challenges may come from adversaries who develop and use breakthrough technologies to negate current U.S. advantages in key operational domains.

While the U.S. military is well equipped to counter traditional challenges, preparing for and responding to the other challenges requires new approaches and the development of new capabilities and competencies, such as stability and support operations. The Department of Defense Directive 3000.5, dated November 2005 (Military Support for Stability, Security, Transition and Reconstruction (SSTR) Operations) explicitly elevates stability operations to a priority level comparable to major combat operations. The Directive states that stability operations are a core mission area of the U.S. military, are essential to ensuring U.S. national interests, and that integrated civilian and military teams are critical to successful stability operations.

In June of 2006 the Joint Staff released the SSTR Joint Operating Concept (JOC), that describes how the future Joint Force Commander (JFC) will provide military support to SSTR operations. SSTR operations run the gamut from counterinsurgency operations to humanitarian relief efforts. The SSTR JOC emphasizes the need for ‘unified action’ in SSTR operations. Unified action is defined as “the successful integration and synchronization of the multidimensional efforts of the U.S. military, U.S. Government agencies, coalition partners, as well as multinational and private sector actors, along with host nation agencies in pursuit of success.” [3] The JOC specifies that management of the coalition is a first order responsibility of a JFC. A critical component of this management is cultural knowledge of the CJTF team. The JOC states that “the commander should leverage cultural and historical ties between coalition forces and the host nation that could positively impact the mission.” [4]

The SSTR JOC also describes the need for the JFC to develop in-depth cultural knowledge of potential adversaries, neutrals, allies as well as others interacting in the environment. The most recent U.S. Army Field Manual (FM) 3-24, authored by General David Petraeus and Lieutenant General James Amos, repeatedly states that cultural knowledge is critical to counterinsurgency operations.

*Cultural knowledge is essential to waging a successful counterinsurgency. American ideas of what is 'normal' or 'rational' are not universal. To the contrary, members of societies often have different notions of rationality, appropriate behavior, level of religious devotion, and norms concerning gender… For this reason, counterinsurgents—especially commanders, planners, and small-unit leaders—should strive to avoid imposing their ideals of normalcy on a foreign cultural problem.* [5]

Currently, the level of cultural knowledge necessary to plan and execute effective SSTR operations far exceeds existing capabilities. At the strategic level, cultural ignorance and misunderstanding results in policies that perpetuate differences, making conflict more likely and stability and reconstruction less likely to occur. The lack of strategic level understanding of Iraqi culture, and, especially, ignorance of the tribal and sectarian divisions within Iraq, is partly responsible for the emergence and intensity of the current insurgency. The recently published Iraqi Study Group Report explicitly cited the lack of cultural knowledge as a key shortfall in Iraq:

*All of our efforts in Iraq, military and civilian, are handicapped by Americans' lack of language and cultural understanding. Our embassy of 1,000 has 33 Arabic speakers, just six of whom are at the level of fluency. In a conflict that demands effective and efficient communication with Iraqis, we are often at a disadvantage.*[6]

At the operational level, cultural ignorance can result in lost opportunities to leverage public opinion and exacerbation of existing conflicts. If not remedied, operational failures can contribute to overall mission failure. For example, when the U.S. military closed Muqtada al Sadr's anti-American *al HawzaI* news paper in 2004, it became a turning point in the insurgency. While closing the paper may have served an immediate military objective of removing what was perceived to be a security threat, it exacerbated the larger cultural conflict between the Americans and Iraqis by reinforcing negative perceptions of Americans as hypocrites.

At the tactical level, cultural ignorance can be deadly. Unfamiliarity with language, customs, and gestures has resulted in multiple deaths in Iraq, where, for instance, the American gesture for stop (arm straight out with palm up) signifies welcome to Iraqis. One tactical commander described the lack of cultural knowledge as follows:

*I had perfect situational awareness. What I lacked was cultural awareness. I knew where every tank was dug in on the outskirts of Tallil. Only problem was, my soldiers had to fight fanatics charging on foot or in pickups and firing AK-47s and RPGs [rocket-propelled grenades].Great technical intelligence. Wrong enemy.* [7]

In addition to a lack of cultural knowledge of our adversaries, we also lack understanding of the cultures of the countries with which we serve together in SSTR operations. Even cultures as closely aligned as the U.S. and UK experience cultural disconnects that can impact CJTF collaboration and effectiveness. Despite longstanding cultural ties, important differences between U.S. and UK forces made collaboration a challenge during Operation Iraqi Freedom. Major General Robin Brims, General Office Commanding (GOC) of the UK 1st Armoured Division, explained, "Being an ally is a two-way street. When you find someone or something odd, reflect with certainty that someone finds you and your people very odd, too." [8]

Differences in culture can manifest themselves in a variety of ways that impact multinational planning. For instance, perceptions of what is appropriate in terms of the degree of formality within a team and the disparity in power between superiors and subordinates are both influenced to some extent by culture. Team collaboration can suffer if cultural disconnects result in poor communications and role and responsibility assignments. Poor collaboration can impact the ability of a CJTF to perform its mission.

Given that a gap in our cultural knowledge exists, what can be done to improve cultural capabilities and competencies?

## 3.0 SOLUTION STRATEGY

The CIE objective was to identify areas within the domain of multinational SSTR planning that are vulnerable to errors or inefficiencies due to cultural differences and demonstrate a proof of concept that culturally informed planning tools can be effectively integrated and utilized. The overarching objective of the project is to explore ways for improving automated decision tools to reduce the negative effects and enhance the positive effects of cultural differences so as to develop and execute more effective SSTR plans.

### CIE Approach

The CIE effort is based on the premise that culture has a significant impact on human behavior. Culture is defined as the shared attitudes, values and beliefs of a collective. Culture is a social phenomenon. It provides a group with tools that help its members "maintain group solidarity, coordinate behavior with others, bargain effectively, manage conflicts of interest, and predict the consequences of their actions." [9] The nature of a group's underlying culture is reflected in its observable artifacts and behaviors;

however, much of culture is ‘below the tip of the iceberg.”

Cultural artifacts, such as flags, emblems, songs, etc. are powerful purveyors of cultural meaning, but do not capture the entirety of a group’s culture. Behaviors, such as rituals, customs, as well as what is considered to be good versus bad behavior, are often driven by specific, socially accepted roles, rules and relations, which constrain and enable people’s choice of actions. These roles, rules and relations have developed over time. These in turn reflect foundational, often unarticulated attitudes and beliefs. The relationship amongst observable behavior and cultural artifacts and underlying and unobservable cultural factors is dynamic, with feedback loops flowing in both directions. Figure 1 illustrates the different levels of culture.

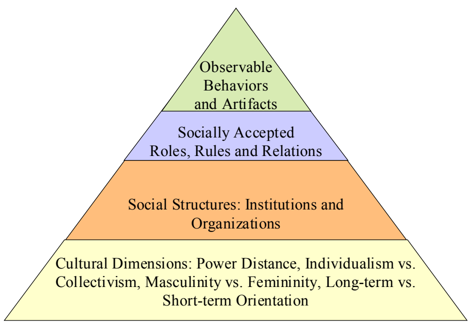


Figure 1. Levels of Culture

To gain insight into a group’s underlying culture, researchers have examined people’s deep seated and often unarticulated attitudes, values and beliefs. Some have explored cross-cultural differences in fundamental cultural attitudes or dimensions and found stable variations. Geert Hofstede, based on extensive empirical evidence from tens of thousands of people, identified five dimensions that reflect fundamental cultural differences across societies. [10] These are power distance, uncertainty avoidance, individualism versus collectivism, masculinity versus femininity, and long-term versus short-term orientation.

- *Power distance* reflects the extent to which the less powerful members of the group accept and expect that power is distributed unequally. Societies that have a high power distance are more accepting of absolute power differences and members tend to not question authority. In low power distance societies, authority is often questioned and authority relations are often contextually determined.
- *Uncertainty avoidance* reflects the extent to which a group is tolerant of ambiguity. Societies with high uncertainty avoidance often construct rigid rules and regulations in an attempt to clearly define right and wrong to minimize the effects of uncertainty. Conversely, societies with low uncertainty avoidance are comfortable with fewer rules and tend to be tolerant of different interpretations of what is appropriate.
- *Individualism versus collectivism* reflects the degree to which individuals are integrated into the group. In individualist societies, people are expected to take care of themselves. In collectivist societies, people are integrated closely into the group and are often very loyal to the collective.
- *Masculinity versus femininity* reflects the distribution of emotional roles between the sexes. Societies that are highly masculine tend to strongly differentiation emotional roles between the sexes so that assertiveness and competitiveness are appropriate for men, and modesty and caring are appropriate for women.
- Long-term versus short-term orientation reflects variations in people’s attitude toward time and cause and effect relationships. Societies with a long-term orientation tend to value thrift and perseverance and are focused on the future; societies with a short-term orientation tend to value respect for tradition and fulfilling social obligations and are focused on the present and the past.

In studies examining business professionals, Hofstede found that attitudes toward these dimensions varied dramatically across countries. Other researchers have validated these results using more representative samples, providing support to the assertion that cultural beliefs vary across countries. Figure 2 illustrates cultural differences across two cultural dimensions for several countries. [11]

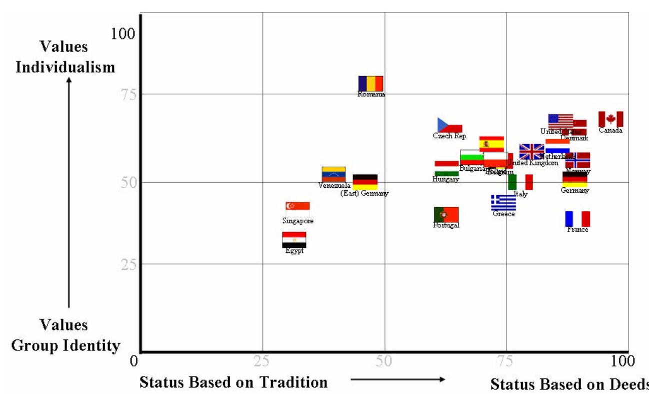


Figure 2. Cultural Differences Across Countries

While differences across cultural dimensions exist within countries, the differences across countries are significant and reflect patterns that are manifested in social institutions, organizations, roles, rules and relations as well as behaviors and artifacts. Researchers and practitioners have studied how these underlying dimensions relate to higher cultural levels and have found patterns that are consistent across time and countries. The business community has extensively explored the relationship between cultural dimensions and higher level social structures and behaviors. The results of this research indicate that, for instance, the dominant organizational structure of firms, the propensity to enter new markets and the adaptability of firms to changing market conditions are all influenced by underlying cultural attitudes

.Other researchers have focused on how culture affects the behavior of nations. For instance, culture affects a society's propensity toward violence and war. The "Democratic Peace Theory' posits that democracies are less prone to war than authoritarian regimes. Numerous studies have found support for the proposition that countries with democratic norms valuing participation in decision making processes, peaceful conflict resolution and leadership accountability are less likely to engage in international conflicts than societies with norms which favor absolute authority and limited participation in decision making. [12] Researchers have found that a "culture of mistrust, fear, and harshness in social relations" as well as a high level of differentiation between gender roles, positively correlates with frequency and intensity of war engagements. [13]

The CIE project intends to build on the extant research related to cultural dimensions and seeks to extend this research into the area of multinational planning and execution of SSTR operations. To make this extension, it is necessary to first understand the planning and execution process and how culture can influence this process.

U.S. military organizations have traditionally followed a doctrinal process to produce plans. The military decision making process (MDMP) is hierarchical and integrated; commanders receive mission tasks from their superior organization, analyze requirements, employ the staff to generate various courses of actions to accomplish mission requirements, and then select a plan. This plan is then sent down to the subordinate units that, in turn, follow the same process to produce their plans.

Recent SSTR and counterinsurgency doctrine, however, emphasize that the complexity of these operating environment requires new approach to planning. The U.S. Army Counterinsurgency FM states,

*The complexity of insurgency presents problems that have incomplete, contradictory, and changing requirements. The solutions to these intensely challenging and complex problems are often difficult to recognize as such because of complex interdependencies. While attempting to solve an intensely complex problem, the solution of one of its aspects may reveal or create another, even more complex, problem.* [14]

In order to plan in such an environment, the Counterinsurgency FM states that a CJTF must first develop a comprehensive understanding of the problem before it can develop a plan. Effective planning requires that decision makers have critical thinking skills, that they can conceptualize complex interdependencies and systems, that they are comfortable with intuitive decision making and continuous learning. These attributes, however, are not universal and are likely influenced by cultural factors. The potential for cultural differences to make a difference is particularly high during the design phase of SSTR operations. It is during this process that differences in assumptions arise amongst multinational coalitions partners. The CIE effort aims to explore how cultural differences impact the design phase of SSTR planning and develop appropriate assessment tools that mitigate the possible negative effects of cultural differences.

The SSTR JOC emphasizes that a primary objective of preparing and planning for SSTR missions is to 'harmonize the many diverse civilian and military efforts within a comprehensive, integrated strategy." [15] This necessarily entails understanding the cultural differences of the CJTF and developing strategies to mitigate them.

Recently, joint doctrine has established that joint planning should be based on the principles of Effects Based Operations (EBO). Effects based operations are, "planned, executed, assessed, and adapted based on a holistic understanding of the operational environment in order to influence or change system behavior or capabilities using the integrated application of selected instruments of power to achieve directed policy aims**.** " [16]

Planning for EBO requires that planners identify actions to nodes to effects: actions that can be taken by a force (a resource), applied to nodes that exist in an environment, and result in effects that bring conditions closer to the desired end state. Planners using this technique break assigned missions or requirements into those effects required to be applied to specific nodes. Given this linkage, planners can develop different courses of action (COA) that achieve the mission.

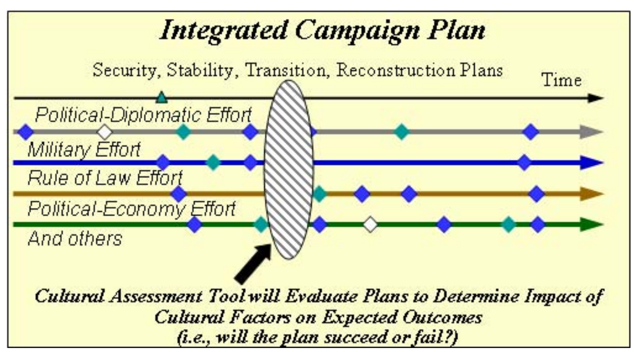


Figure 3. Cultural Assessment Tool

In executing EBO, planners may apply a full range of action options (resources) from any available national resource. These resources are generally categorized in to diplomatic, information, military and economic (DIME) domains. The objectives of EBO are to affect the desired changes in the operating environment across political, military, economic, social, information, and infrastructure systems (PMESII). Creating the DIME and PMESII structure for resource identification and effects definition, reflects the enhanced awareness by military strategists and planners that to succeed in future struggles they must incorporate more than traditional military factors.

The next step in the evolution of CIE is to identify what capabilities and characteristics are required of future decision support tools, based on an underlying cultural ontology and associated knowledge base, to improve the planning and execution of multinational SSTR operations. This will provide architecture for and capacity to give military planners the tools needed to utilize cultural information in building SSTR plans and to assess such plans with respect to cultural sensitivities.

The basis for these tools must be a cultural ontology that defines the taxonomy of cultural primitives of interest to the planning and execution of multicultural SSTR operations. A useful knowledge base associated with the ontology must contain facts and rules that define the relationships across cultural primitives for specific cultures. Any sense-making engine that attempts to reason over cultural knowledge will face many instances of ambiguous or contradictory data bits. This will be one of the most difficult problems to overcome in developing a reasoning engine for cultural understanding.

The challenge will be integrating the cultural ontology and knowledge base with the current planning and situational awareness (SA) tools to provide an effective decision support suite for SSTR operations. Current planning and SA tools are not well-configured for SSTR operations. Additional planning and assessment functionality that uses the cultural knowledge from the reasoning engine and feeds the current planning and SA tools would provide the rapid, tailored, cultural data required to meet this demand. It is anticipated that any fully functional system will require a hybrid design philosophy.

Any cultural assessment planning aid must be "plan aware." This essential functionality may best be implemented by integrating with existing which use traditional rule-based inferencing and other techniques.

## 4.0 CONCLUSION

The CIE effort aims to augment cultural education and training efforts aimed at improving SSTR operations. It seeks to develop a technical approach that will provide CJTF commanders and staff with an assessment tool that will allow them to minimize the negative effects of cultural differences amongst CJTF team members. It will also assess how cultural factors will impact the likelihood that the plans generated by automatic decision support tools will result in the desired effects.

## REFERENCES

[1] Quoted in Montgomery McFate, "Anthropology and Counterinsurgency: The Strange Story of their Curious Relationship." *Military Review* (March-April 2005), p. 24. Original source, MG Robert H. Scales, "Culture-Centric Warfare," *Proceedings* (October 2004).

[2] "National Defense Strategy of the United States of America." DoD, March 2005, http://www.globalsecurity.org/military/library/policy/dod/nds-usa_mar2005.htm accessed July 11, 2006.

[3] U.S. Department of Defense, Military Support to Stabilization, Security, Transition, and Reconstruction Operations, Joint Operating Concept. December 2006.

[4] Ibid.

[5] U.S. Department of Defense, Counterinsurgency Field Manual, FM 3-24, December 2006.

[6] Iraqi Study Group Report, December 2006.

[7] Montgomery McFate, "The Military Utility of Understanding Adversary Culture," *Joint Forces Quarterly*, 38, 2005.

[8] Andrew Gregory, "1st (UK) Armoured Division in Iraq: January to April 2003," *Field Artillery*, January 1, 2004.

[9] Jack Snyder, "Anarchy and Culture: Insights from the Anthropology of War," *International Organization*, 56, 1, Winter 2002, pp. 7-45.

[10] Geert Hofstede, Culture's Consequences (2nd, fully revised edition, 2001) and Cultures and Organizations, Software of the Mind (2nd, revised edition 2005, with Gert Jan Hofstede).

[11] Fons Trompenaars (1989). Riding the Waves of Culture, Seven Cultures of Capitalism, Building Cross-Cultural Competence and 21 Leaders for the 21st Century.

[12] Russett, Bruce (1993). *Grasping the Democratic Peace*. Princeton University Press.

[13] Jack Snyder, "Anarchy and Culture: Insights from the Anthropology of War," *International Organization*, 56, 1, Winter 2002, pp. 7-45.

[14] U.S. Department of Defense, Counterinsurgency Field Manual, FM 3-24, December 2006.

[15] U.S. Department of Defense, Military Support to Stabilization, Security, Transition, and Reconstruction Operations, Joint Operating Concept. December 2006.

[16] U.S. Joint Forces Command, Standing Joint Force Headquarters (Core Element) SOP, 14 Jul 04.

# Understanding Process Differences: Agreeing Upon a Single Way to Skin a Cat

Steven Poltrock and Mark Handel
The Boeing Company
Seattle, WA 98124
425-373-{2729; 2883}
{steven.poltrock; mark.j.handel}@boeing.com

Mark Klein
Massachusetts Institute of Technology
Cambridge, MA 02139
617-253-6796
m_klein@mit.edu

**Abstract**—*Supporting human collaboration is challenging partly because of variability in how people collaborate. Even within a single organization, there can be many variants of processes which have the same purpose. When diverse coalition members must work together, the differences can be especially large, baffling and disruptive. Coordination theory provides a method and vocabulary for modeling complex collaborative activities in a way that makes both the similarities and differences between them more visible. To demonstrate this, we modeled three very different engineering change management processes and found: (1) most of the work is coordination-related; (2) despite large apparent differences, a coordination-theoretic analysis revealed substantial commonalities among the three processes; and (3) differences in the processes were due to choices regarding coordination mechanisms. This approach has promise for helping to merge or integrate different processes and to suggest ways that agents can participate in complex collaborative processes.*

## 1. Introduction

Coalition members must work together despite significant differences in language, culture, policies, and organizational procedures or practices. When two or more organizations follow different processes or procedures to accomplish the same goals, these differences can become obstacles to effective collaboration. A method for analyzing processes that identifies the root causes of their differences may help coalitions eliminate, reconcile, or at least understand how and why their processes differ, and lead to more effective joint efforts.

This paper presents a top-down approach for modeling and analyzing the similarities and differences between related complex processes. It is based on the insight that much of collaborative work involves activities for coordinating inter-dependent "core" tasks. Different groups, even if they have the same core tasks, may chose different ways for coordinating them. These choices can yield work processes that appear widely divergent, even though their purposes are essentially identical. Our approach is based on making these shared core tasks and differing coordination choices readily visible.

In the following sections we introduce coordination theory and a method, based thereon, for modeling processes. We apply this method to *change management*, a complex collaborative process frequently performed in engineering organizations. Although the basic change management process is relatively constant, there is great variation in its implementation in different contexts. We compare three different change management processes, identifying their differences in terms of the coordination mechanisms they invoke. We conclude by addressing the relevance of this work to knowledge systems for coalition operations.

## 2. Coordination Theory

Coordination theory [4, 5, 6] is the general body of theory about how people or software agents coordinate their activities, and it has been the subject of research in both computer science where the focus is on coordinating software agents, and the social sciences, where the focus is on describing how people coordinate.

A key concept in coordination theory is that collaboration occurs in order to manage the dependencies between tasks. A *flow* dependency exists when one person creates a product required by another person. A *sharing* dependency exists when a task requires a shared resource such as the labor of people who are involved in other tasks. A *fit* dependency exists when two or more people create products that must integrate. There are, of course, many ways to manage each type of dependency. People communicate, share information with one another, and use collaboration technologies in order to manage these dependencies. Variation in complex activities is due largely to different choices regarding how to manage these dependencies.

Malone and his colleagues [2,3,5] have developed a top-down approach to modeling complex activities. In this approach, one defines a process by identifying the core tasks and key dependencies in that process, and then selecting the coordination mechanisms that will be used to manage each dependency. These mechanisms may introduce new dependencies and exceptions that will in turn require additional mechanisms and handlers. This decomposition can continue to any desired level of detail. A key element in this approach is a large taxonomically-organized repository, known as the Process Handbook, which captures the substeps of these mechanisms, the exceptions commonly encountered with each mechanism, as well as handlers for resolving these exceptions. These mechanisms represent, as we shall see, high-level building blocks for creating models of collaborative processes.

## 3. CHANGE MANAGEMENT

Change management is a key process in engineering organizations. A large aerospace program, for example, may have hundreds of change management processes that govern changes to software applications, plans, requirements, costs, schedules, configurations, and any other attribute of importance to the program. There is widespread agreement, depicted in Figure 1, about the basic change management process, but the details vary widely from one instance of the process to another. A change is proposed, it is authorized (or not) based on an assessment of its impact, and if authorized it is then implemented.

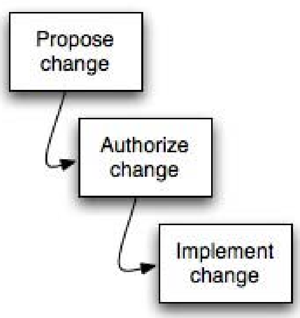


**Figure 1** - Basic change management process

**A Coordination Theoretic Model for Change Processes**
In Figure 2 we illustrate the first steps of developing a model of the change process using coordination theory. We first must identify the "deep structure" for the process, i.e. the core tasks and key dependencies. The change process consists of three core tasks (propose changes, authorize changes, and implement changes) as well as two key dependencies (a change request (CR) flows from the first task to the second, and an authorizing change notice (CN) flows from the second task to the third):

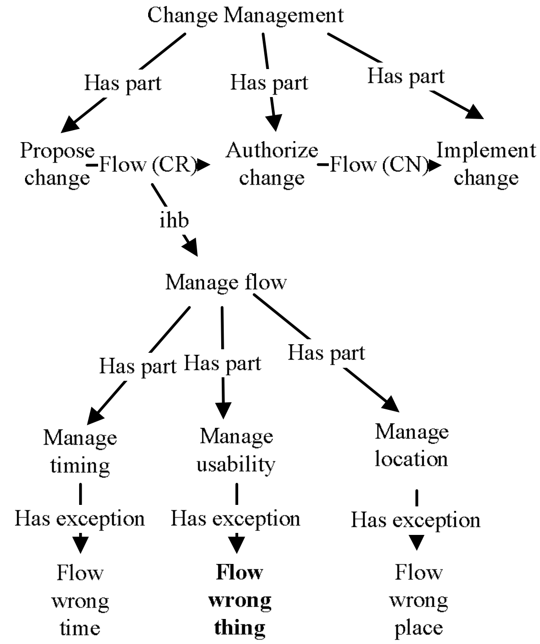


**Figure 2** - Top level of the change process

Next, we define a coordination mechanism for the first flow dependency. Any flow is managed by some variation of the generic "manage flow" building block in the Handbook repository. This template captures the fact that managing a flow always involves managing the timing, usability, and location of the resource that is flowing. Each of these subtasks, furthermore, have their own characteristic exceptions (the manage usability step has, for example, the exception "flow wrong thing"), and each of these exceptions has a range of processes (not shown) for handling them.

**Avoiding Inappropriate Changes**

The key challenge in change management is to avoid implementing the wrong change request or, in other words, to avoid the "flow wrong thing" exception shown in Figure 2. Table 1 lists some of the mechanisms in the Handbook repository suited for handling this exception.

**Table 1** - Handlers for "flow wrong thing"

| | |
|---|---|
| Avoided by | Filter out unwanted elements (by individual or team judgment) |
| Resolved by | Filter out bad agents |
| Detected by | Monitor agents for commitment violations |
| Anticipated by | Track reputation information |
| Avoided by | Provide incentives |

**Table 2 -** Tradeoffs for specializations of filtering out unwanted elements

| Alternative | Best for | Cost | Quality | Speed |
|---|---|---|---|---|
| Filter by individual | Initial pruning of easy-to-find problems, such as missing data | Low | Low | Fast |
| Filter by team | Careful evaluation of resource from multiple perspectives | High | High | Slow |

Most engineering organizations use variants of the "filter out unwanted elements" handler. One variant (filter by individual) involves asking an individual to reject the change requests that in his or her judgment are not viable, while the second involves asking a team to serve this role. The Handbook repository includes tradeoff tables that describe the relative strengths and weaknesses of alternative processes for fulfilling a given function. The tradeoffs for the two variants of "filter out unwanted elements" are shown in Table 2. Filtering by individual is fast and cheap but of low quality, whereas filtering by teams is slow and more expensive but higher in quality. In many engineering organizations, filter-by-individual is used as an initial screening step, followed by filter-by-team. The generic "filter by team" handler in the Handbook repository consists of several steps (Figure 3). The first is to get reviews or assessments of the impact of a CR, and this requires creating a review request that is sent to all reviewers, performing the reviews, and then consolidating the reviews into a coherent package. The filter-by-team handler also includes making an accept/reject decision, which requires first reviewing the completed package and then making a decision. Note that there are dependencies between these lowest level parts, and each dependency in turn requires a coordination mechanism. Figures 2 and 3, taken together, represent a coordination-theoretic model of the change management process used by many engineering organizations.

## 4. PROCESS VARIATION

When activities are collaborative, the way they are performed varies greatly from organization to organization, from one team of participants to another, and from one time to another [1]. This is certainly true for change management. We investigated whether such differences can be explained as a consequence of selecting different coordination mechanisms and/or different exception handling mechanisms and, if so, whether this perspective on process variation can help us understand and, if necessary, eliminate it.

To do so, we compared three change management processes used within one aerospace program: one for managing change to cost and schedule (BMW), another for managing change to product configuration (CCP), and a third for managing change to processes and tools (SIP). These are complex processes involving many people in varying roles. Applications that automate the flow of work have been implemented for all three processes, and a portion of the BMW workflow management process model is depicted in Figure 4 to illustrate their complexity. The complexity of these processes prevents us from presenting a complete coordination theory model; the simplest model (BMW) includes 48 steps. Even though the processes all have essentially the same goal, the way they were modeled by their users and managers were widely divergent, the workflow applications that support them are different, and commonalities are far from obvious.

Our first key finding was that most of the steps in these processes involved coordination. Of the 48 tasks in BMW, for example, 41 are coordination mechanisms (e.g. sending change requests to the reviewers, collecting and consolidating the reviews, distributing them, holding a review meeting, and notifying the requestor about the outcome) or exception handlers (e.g. filtering CRs and handling review requests that are sent to the wrong person or not returned on time).

A second key finding is that the differences between these processes concerned how they perform coordination and exception handling. In order to make these differences readily visible, we created a process representation we call a "derivation tree". This tree captures the refinements (i.e. additions of coordination mechanisms or exception handlers)

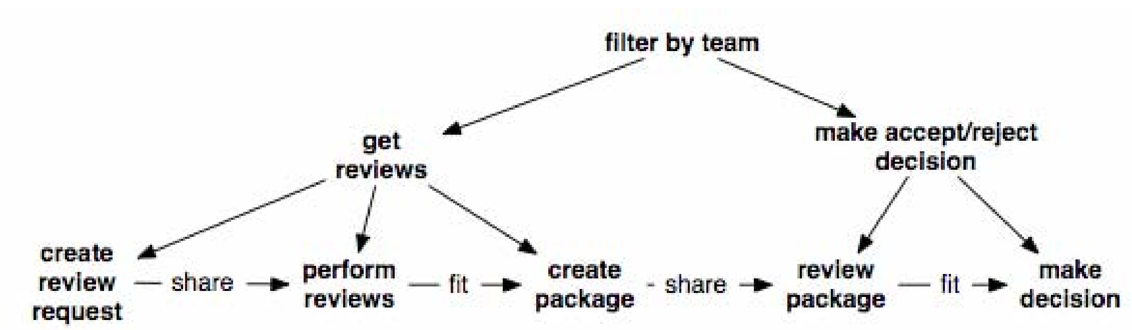


**Figure 3**.-The "filter by team" exception handler.

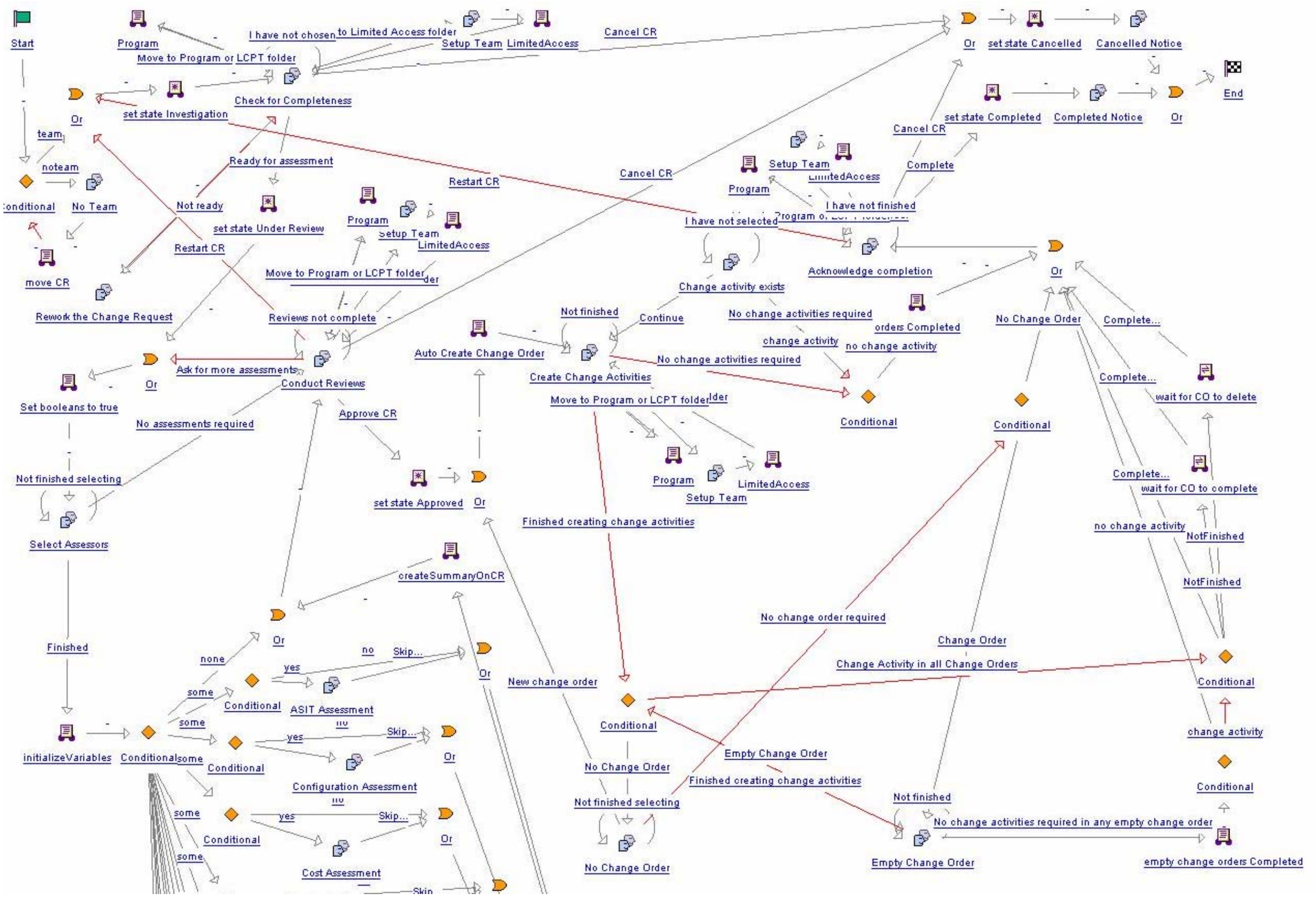

**Figure 4** – A portion of the workflow model that implements the BMW process

used to create a process model using the top-down analysis introduced above. The derivation tree for the generic change management process described above, for example, is presented in Figure 5.

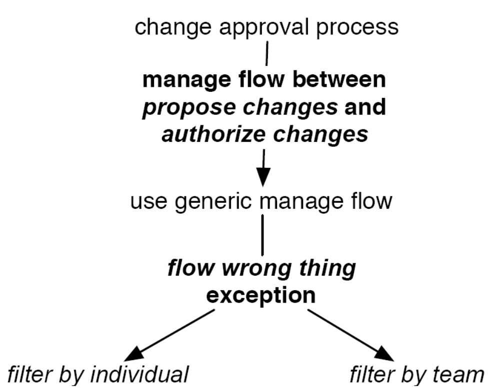


**Figure 5**.-Derivation tree for the generic change management process model.

Each arrow describes (in bold text) what aspect (dependency or exception) of the model is to be refined. The target of each arrow captures the coordination or exception handling process selected for this purpose. These processes can in turn have dependencies and exceptions that need to be refined. A derivation tree is generally quite compact, because a refinement often represents the addition of a relatively large building block (i.e. a coordination mechanism or exception handler) from the Handbook repository. The derivation tree for the BMW change process, for example, consists of 11 refinement operations, while the conventional flowchart model for this process includes 48 steps.

Derivation trees can be used to highlight the similarities and differences between related processes. The trick is to consolidate, into a single tree, the derivation trees for the processes being compared. Differences between the processes become immediately evident as alternative refinements for a given dependency or exception. Figure 6 shows a consolidated derivation tree for the BMW, CCP, and SIP change processes. Much of the tree, we can see, is shared by all three processes. All start with the generic change approval process, and use individual and team reviews to avoid authorizing the wrong change requests.

The differences between these processes appear in bold in Figure 6. One difference concerns *who* can generate change requests. In the CCP process, any engineer can request a

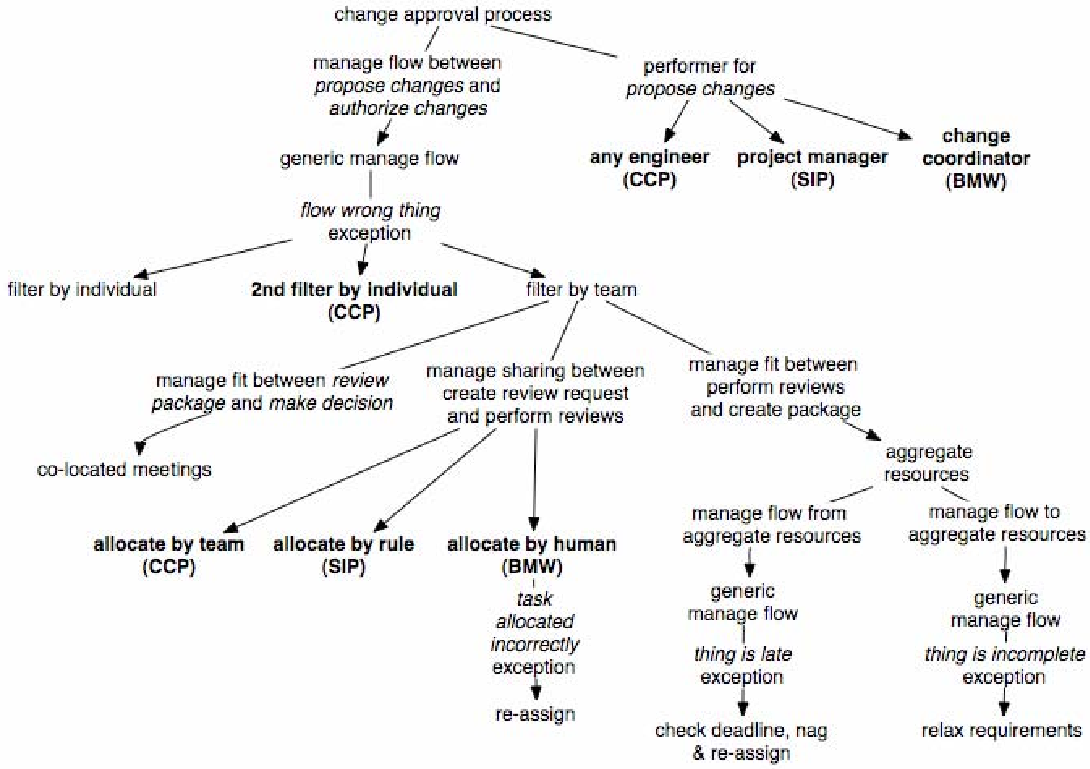


**Figure 6 -** Consolidated derivation tree comparing the BMW, CCP and SIP change processes.

change. In the BMW process, proposed changes must be submitted by engineers via a change coordinator, so he or she can filter/revise inappropriate change requests before they enter the change process. In SIP, change requests must be submitted via the project manager.

A second difference is that the CCP process includes a second filter-by-individual step, performed by the requestor's manager. This additional step avoids the cost of a team review for CRs that are unlikely to be authorized.

The processes also differ, finally, in how people are assigned to review change requests (i.e. in how we refine the sharing dependency between "create review request" and "perform reviews"). A fixed set of reviewers are expected to review every CR in the SIP process, which corresponds to an allocate-via-rule mechanism. The change coordinator in the BMW process decides who reviews CRs, which corresponds to an allocate-via-human-judgment mechanism. For the CCP process there are thousands of potential reviewers, and no one individual can be expected to have the knowledge required to determine who should review each CR. Instead, they use an 'allocate by team' mechanism, wherein members of a team make suggestions concerning how to allocate each resource (change request), and these recommendations are consolidated somehow (e.g. concatenated) to produce the final list of recipients (change request reviewers).

## 5. CONCLUSION

In this research we studied just three of hundreds of change management processes within a single aerospace program. Although all three processes are intended to accomplish the same thing, they use different tools and procedures, and the documentation that guides participants in these three processes looks radically different. This makes it difficult to understand and, if desired, eliminate differences between process variants. By applying top-down coordination-theoretic modeling, supported by a Handbook of generic coordination mechanisms and exceptions handlers, we were able to create derivation trees for all three processes that made the source of their similarities and differences much easier to identify. With this information in hand, we can then begin asking *why* the processes differ in these ways, and whether/how we want to change them in order to make the processes more effective and more consistent.

Coalition operations require that people work together despite differences in culture, organization, experience, resources, and skills. Each coalition member is likely to have a different process for accomplishing essentially the same tasks. Consider a task such as mission planning that requires collaboration among people with a range of knowledge, experience, skills, and responsibilities. Coalition members with their own mission planning process are likely to have difficulty appreciating why other members follow different processes. Collaboration between coalition members could suffer because of this failure of understanding. They may not recognize how collaboration is intended to occur, and consequently fail to provide information at appropriate times or take appropriate action when information is provided to them.

Models of mission planning based on coordination theory could help understand the common structure of the activity and identify differences and the reasons for those differences. This analysis can be the foundation for defining a common process that spans all coalition members or points of integration between their different processes.

These models also help identify tasks that agents can perform and how agents can most effectively collaborate with people. Many of the tasks in a change process could be performed by agents. For example, an agent could perform the filter-by-individual task. More importantly, agents can serve as exception handlers. There are many possible exceptions defined for mechanisms in the knowledge base, and handlers may anticipate, avoid, detect, or resolve these exceptions. An agent handler could attempt to avoid an exception, anticipate and detect its occurrence, and either resolve it independently or escalate it to a human participant.

## REFERENCES

[1] Dourish, P. *Where the action is: The foundations of embodied interaction.* MIT Press, 2001.

[2] Klein, M. and Dellarocas, C. Designing robust business processes. *Organizing business knowledge: The MIT process handbook.* The MIT Press, 2003, 423-439.

[3] Klein, M. and Petti, C. A handbook-based methodology for redesigning business processes. *Knowledge and Process Management, 13,* 2006, 108-119.

[4] Malone, T.W. and Crowston, K. What is coordination theory and how can it help design cooperative work systems? *Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW'90),* Los Angeles, October 7-10, 1990, pp. 357-370.

[5] Malone, T.W., Crowston, K. and Herman, G.A. *Organizing business knowledge: The MIT process handbook.* The MIT Press, 2003.

[6] Malyankar, R.M. and Findler, N.V. A methodology for modeling coordination in intelligent agent societies. *Computational and Mathematical Organization Theory, 4,* 1998, 317-345.

## ACKNOWLEDGMENT

*MIT's contribution to this work was funded by The Boeing Company, award MIT-BA-GTA-1.*

*Research was sponsored by US Army Research laboratory and the UK Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Army Research Laboratory, the U.S. Government, the UK Ministry of Defense, or the UK Government. The US and UK Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.*

# Cultural Issues in Coalition Planning

Winston R. Sieck, Klein Associates, wsieck@decisionmaking.com
Jitu Patel, Dstl, jmpatel@dstl.gov.uk

**Abstract**—*This paper outlines the cultural issues that manifest in the planning and decision-making phases of coalition operations. It then summarizes the current status of emerging tools for identifying cultural differences, with respect to supporting coalition planning activities. The emphasis is on cultural variations in cognition, language, distributed social interaction, as well as their interrelationships. In particular, the paper describes Cultural Network Analysis for extracting and representing culturally shared, complex mental representations that drive decisions; tools for the assessment of commander's intent across coalition boundaries; and methods for investigating social interactions and language in multicultural distributed collaboration settings.*

## 1. Introduction

Coalition operations have emerged as a key feature of military operations in the post-Cold War era. Indeed, it is now the standard mode across the spectrum of military interventions, with virtually all U.S. and U.K. operations over the last two decades having been conducted as part of a coalition [1]. A coalition typically consists of a collection of distinct state militaries[1], as well as non-military organizations such as departments of state, multinational agencies (e.g., UN), non-governmental groups (NGOs, media, relatives), and sometimes agents of the host nation. Each of these groups differ in size, composition in terms of age and gender, activities performed on the ground and ways of performing them, reasons for participation, and success criteria of the operation. Furthermore, they bring their own cultural backgrounds that influence their decision-making and collaborative styles, natural and doctrinal languages, technological capabilities, and core competencies.

Cultural differences can and do lead to tensions between coalition partners. For instance, Winslow's [3] study of UN peace operations identified the seven areas of tensions between military and other groups in theater (including other national militaries): organizational composition, different tasks and ways of accomplishing them, differing time frames, different definitions of success, different abilities to exert influence, control of information, and control of resources. These cultural issues come into play at all stages of operations: planning, giving and taking orders, executing orders and reporting. The focus of the current paper, however, is on the cultural issues that arise in coalition planning and decision making.

## 2. Cultural Issues in Planning

Culture has come to the fore as a critical area for developing current and future capabilities. For example, Petraeus reported that recent experiences in Afghanistan and Iraq show that "cultural awareness is a force multiplier." That is, knowledge of the cultural "terrain" can be as important as, and sometimes even more important than, knowledge of the geographic terrain [4]. Scales cast the situation even more starkly: "it is more important to understand motivation, intent, method, and culture than to have a few more meters of precision, knots of speed, or bits of bandwidth" [5].

Much of this discussion about cultural issues rightfully emphasizes the cultures of the adversary and populace from the host nation. For example, Yates' analysis of the historical record points out a need to look at not only what resources the military can bring to bear to solve anticipated problems, but how the local population and its leaders will respond to

[1] For example, "the military presence established in Bosnia-Herzegovina following the end of their civil war included participants from over 38 nations" [2].

actions the military undertakes to address those problems [6].

However, cultural differences exist even among the closest of coalition partners, such as between the U.S. and U.K. militaries. Understanding culture within the core of a coalition amounts to developing a deep understanding of "ourselves" rather than our host, and in many cases raises a distinct set of cultural issues that matter in that particular context. For example, eating behaviors may have an impact on relationships between members of the military and host nation, but have little bearing on the coalition team putting together a campaign plan.

The increased attention to culture has not arisen in a vacuum, but rather in response to a shift in the requirements for current and envisioned future warfare. Chiarelli and Michaelis described requirements for full-spectrum operations, including the simultaneous pursuit of the development of economic pluralism, promoting governance, restoration and improvement of essential services, and training and employing host nation security forces, all in addition to combat operations and with a primary emphasis on "winning the peace" [7]. Pierce and Dixon have identified these new challenges as constituting a special case of a "wicked problem" [8]. Wicked problems have no known solutions or single right answers, but instead include myriad evolving issues and constraints that demand solutions that satisfice for the relevant stakeholders. In such cases, multiple stakeholders with distinct agendas must come together to discuss the elements of the ill-defined and changing problem space in order to develop generally acceptable and adaptable plans. Hence, coalition planning and decision making emerges as the center of gravity for successfully handling wicked problems. The significance is that these core processes must be adapted to be truly integrative and inclusive of the represented coalition members in a way that capitalizes on the benefits of their cultural diversity. Wicked problems cannot be solved by one major player driving the planning and decision-making processes.

A first step in realizing the required level of inclusiveness is to model the diversity of coalition planning and decision-making processes in terms of their meaning, expectations, rules, and norms. In order to support a team of coalition planners who are planning a campaign involving full spectrum operations, we need to first understand the differences between the partners' concepts of plans, plan quality, and the tools and processes believed to yield high-quality plans. These concepts can differ in potentially significant ways, including: 1) whether goals are seen as fundamentally clear or vague in specification; 2) whether the ideal is a highly detailed plan such that execution cannot begin until all ambiguities are removed, or the ideal is a general concept to which details will be added once the action starts; 3) whether plan revisions are viewed as necessary for adaptability or indicators of poor planning; 4) what the basic functions of a plan are considered to be; 5) how the plans are structured and the way they are communicated to sub-units. This initial list of potential differences in plans and planning needs to be expanded upon and validated. It also needs to be more deeply understood in terms of the relevant cultural and contextual pressures.

Though being "culturally aware" is now seen to be essential, military have always been cognizant of differences between service arms (air, land, and sea) and possible confusions are minimized through use of Liaison Officers and joint training. This strategy is also successfully employed with close coalition partners who participate in multinational training exercises and exchange Liaison Officers to mitigate the problem of language and potential cultural misunderstandings. Unfortunately, this strategy works only in well-defined conflict scenarios in which friends and enemies are constant, and there is ample time to get to know the coalition partners. Conflicts in the past decade have been marked by asymmetric opponents and very fluid coalitions put together in short time frames. As a consequence, the benefits of Liaison Officers and multinational training exercises cannot be relied upon for "new" coalition partners.

## 3. RESEARCH ON CULTURAL ANALYSIS

Current methods for understanding cultural differences include ethnographic methods that focus on qualitative analysis of a single cultural group, and psychological methods that attempt to capture cultural differences in a few generic dimensions, such as individualism/collectivism. Such methods are limited in their ability to capture and represent cultural commonalities and disconnects with the precision needed to enable the design of complex cognitive systems, such as coalition command. Hence, there is a need for rigorous methods to capture, analyze, represent, and compare the cognitive, linguistic, and social dynamics of cultures.

Recently, several efforts have gotten underway to research, develop, and evaluate new approaches to enhance multicultural collaboration in coalitions. For example, Pierce and Dixon describe several technologies under development by the U.S. Army Research Laboratory [8]. More recently, the International Technology Alliance (ITA) between the U.K. Ministry of Defense and the U.S. Army Research Laboratory has devoted one of its twelve interdisciplinary projects to culture. Specifically, the overall objective of the "Cultural Analysis" project in the ITA is to explore the development of methods to advance the state of the art of cultural analysis. A second objective is to employ these methods to analyze key cultural issues in coalition operations, especially planning, decision making, and negotiation. The challenge is to develop methods that can detect and represent the more subtle differences that do exist between culturally similar national partners engaging in collaborative planning and decision making, than could be accounted for using existing approaches.

### Cultural Network Analysis

Cultural Network Analysis (CNA) refers to a collection of methodologies for building cultural models [9]. Cultural Network Analysis includes methods to first elicit and analyze the mental models of a sample of individuals within the population, and second measure the degree to which elements of the mental models are shared across individuals. Lastly, CNA provides a framework for representing the culture. A fully developed, analytical cultural model represents the statistical distribution of mental models for a particular cultural group and domain. Formal representation makes it possible to use cultural models in a variety of applied contexts. Cultural Network Analysis can be used to build cultural models of collaborative planning and decision making, in addition to other macrocognitive functions.

Cultural Network Analysis is an outgrowth of the cultural epidemiology theoretical view of culture as comprising distributions of networks of causally-interconnected ideas within populations (i.e. mental models). Cultural Network Analysis builds on a synthesis of conceptually related methods for knowledge elicitation, analysis, and representation that stem from the diverse fields of naturalistic decision making, cognitive anthropology, cognitive psychology, and decision analysis. Prior to the development of CNA, none of these fields alone has offered a comprehensive, end-to-end approach for cultural modeling.

### Tools for the Assessment of Command Intent

A second methodology being proposed is a tool for improving the assessment of Commander's Intent in coalition settings [10]. The process of a subordinate commander receiving and interpreting orders from a higher level in the command structure, and developing and issuing his own orders from them, is well documented for both U.S. and U.K. forces, but assessment of the process, especially when U.S.-U.K. coalition operations are considered, will be greatly assisted by a well-defined method, and measurement tool based upon it. Development of the tool for improving the assessment of Commander's Intent will be based on measuring the effectiveness of commanders and sub-commanders in interpreting and developing orders. Applied to coalition force command structures, it will allow assessment of the influence of command language and culture on U.S.-U.K. coalition operational effectiveness. It will also provide the first step in the development of quantitative measures related to actual transmission of orders.

### Measures for Analysis of Distributed Social Interactions

The third emphasis is on communication patterns. There are at least two different communication means of interest: face-to-face and facilitated by electronic communications. The primary concern in the development of the current measures is with mediated communications. Specifically, a variety of communication applications will be instrumented and used to analyze and measure the cross-cultural variability of communications at several levels, including: social network analysis, communication pattern analysis, communication behavior analysis, conversation and discourse analysis, and communication content analysis. A special emphasis area within the discourse analysis is to improve on the understanding of pragmatics. Parallel work in computational language analysis is focusing on meaning using techniques such as Latent Semantic Analysis [8]. In the ITA effort, the emphasis is at the level of communicating intent through what is said, as well as how it is said [11]. This focus will provide the type of "conversation analysis" required to address the needs of US-UK coalition C2 requirements.

## 4. Summary

U.S.-U.K. coalition and joint operations often face operationally and environmentally complex and dynamic scenarios, which require highly synchronized coordination and adaptive capability. New methodologies for cultural analysis that are required to meet these challenges are currently under development under the International Technology Alliance.

## Acknowledgment

Research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

## 5. References

[1] Bensahel, N. (2003). Preparing for coalition operations. In L. E. Davis & J. Shapiro (Eds.), *The U.S. Army and the new national security strategy* (Vol. MR+1657-A, pp. 111-128).

[2] Pierce, L.G. (2002). Barriers to adaptability in a multinational team. *Proceedings of the 46th Human Factors and Ergonomics Society Annual Meeting,* pp. 225-229, Baltimore, MD.

[3] Winslow, D. (1999). Military Culture in Complex Cultural Encounter. *Proceedings of the conference on Ethics in Canadian Defence*, Ottawa, 2-3 November 1999.

[4] Petraeus, D. H. (2006). Learning Counterinsurgency: Observations from Soldiering in Iraq. *Military Review*, Jan-Feb 2006.

[5] Scales, R. H. (2004, September). Culture-centric warfare. *Proceedings*.

[6] Yates, L.A. (2006). The US military's experience in stability operations, *1789-2005: Global war on terrorism occasional paper 15*. Fort Leavenworth, KS: Combat Studies Institute Press.

[7] Chiarelli, P. W., & Michaelis, P. R. (2005, July-August). Winning the peace: The requirement for full-spectrum operations. *Military Review*, 4-17.

[8] Pierce, L. G., & Dixon, M. W. (2006, May). Improving multicultural teamwork to combat terrorism. In *Tactical Decision-Making and Situational Awareness for Defense Against Terrorism*. Turin, Italy.

[9] Sieck, W. R., & Rasmussen, L. J. (2006). *Cultural Network Analysis*. (Technical Report 06TA4-P11-T1). Fairborn, OH: Klein Associates division of ARA.

[10] Whitworth, I. R., Hone, G. N. & De Looy-Hyde, J. (2007, May). Assessing the transmission and understanding of

Commander’s Intent. *Proceedings of Knowledge Systems in Coalition Operations*.

[11] Kao, A., Poteet, S., & Xue, P. (2007). *Towards a computational pragmatics approach to analyzing coalition communication processes*. (Technical Report 06TA4-P11-T3). Boeing Phantom Works.

# Neurodynamics of Consciousness and Cultures

Leonid I. Perlovsky, leonid@deas.harvard.edu, telephone: 781-377-1728
Harvard University, Cambridge, MA, and the US Air Force Research Laboratory, Sensors Directorate, Hanscom AFB, MA 01731

*Abstract* - **The knowledge instinct drives higher cognitive functions of the mind and determines evolution of consciousness and cultures. Dynamic logic mathematically describes these mechanisms, including concepts, emotions, instincts, consciousness and unconscious. The two main aspects of the knowledge instinct are differentiation and synthesis. They are in complex relationship of symbiosis and opposition, leading to complex non-linear evolution. Mathematical modeling of these dynamics in a population leads to predictions for the evolution of languages, consciousness, and cultures.**

## 1. Dynamic Logic and The Knowledge Instinct

To satisfy any instinctual need—for food, survival, and procreation—first and foremost we need to understand what's going on around us. The knowledge instinct (KI) is an inborn mechanism in our minds, an instinctual drive for cognition which compels us to constantly improve our knowledge of the world.

Biologists and psychologists have discussed various aspects of this mechanism, a need for positive stimulations, curiosity, cognitive dissonance [1,2,3]. Until recently, however, this drive was not mentioned among 'basic instincts' on a par with instincts for food and procreation.

The fundamental nature of this mechanism became clear during mathematical modeling of workings of the mind. Our knowledge always has to be modified to fit the current situations. We don't usually see exactly the same objects as in the past: angles, illumination, and surrounding contexts are different. Therefore, our internal representations have to be modified; adaptation-learning is required [4,5,6].

Virtually all learning and adaptive algorithms maximize correspondence between the algorithm internal structure (knowledge in a wide sense) and objects of recognition. Internal mind representations, or models, which our mind uses for understanding the world, are in constant need of adaptation. Otherwise we would not be able to understand the world, to orient ourselves, or satisfy any of the bodily needs. Therefore, we have an inborn need, a drive, an instinct to improve our knowledge, and we call it *the KI*. It is a foundation of our higher cognitive abilities, and it defines the evolution of consciousness and cultures. Psychologically we perceive satisfaction or dissatisfaction of the KI as aesthetic emotions of harmony or disharmony between our knowledge and the world.

Dynamic logic (DL) [7], mathematically implements the KI and basic mechanisms of the mind identified by many authors [8,9,10,11]. DL is a part of Neural modeling fields (NMF) network, a multi-level, hetero-hierarchical system. Interactions between adjacent hierarchical levels include bottom-up and top-down signals (fields of neural activation). Top-down signals are generated by adaptive models. At each level, output signals are concepts recognized in (or formed from) input signals. This general structure of NMF corresponds to our knowledge of neural structures in the brain. In NMF, the KI maximizes a similarity measure between the models and signals.

## 2. Conscious, Unconscious, and Differentiation

DL satisfies the KI and improves knowledge by evolving vague models toward crisp models; this maximizes similarity between models and data [7]. This knowledge accumulation proceeds in the minds of every member in a society and constitutes an essential aspect of cultural evolution. Vague and uncertain models are less accessible to consciousness, whereas crisp and concrete models are more conscious.

Conscious concepts are developed by the mind, according to Jung, based on genetically inherited structures, archetypes, which are inaccessible to consciousness [12,13]. Grossberg [4] suggested that only signals and models attaining a resonant state (that is signals matching models) can reach consciousness. It was further detailed by Taylor [14]; we may be conscious of differences between the mind's prediction model and sensory observations. Dynamic logic evolves fuzzy models into crisp models. Fuzzy models are not accessible to consciousness. Final results of DL, resonant states characterized by crisp models and corresponding signals are accessible to consciousness. Increase in knowledge and improved cognition results in better, more diverse, more differentiated consciousness.

In evolution, the original state of consciousness is undifferentiated unity. This initial unity of psyche limited the abilities of the mind, and further development proceeded through the differentiation of psychic functions. Differentiation of consciousness began millions of years ago. It has accelerated recently, and still continues today [15,12,16].

In pre-scientific time there was a popular idea of homunculus, a little mind inside our mind, which perceived our perceptions and made them available to the mind. This naive view is amazingly close to NMF mechanisms. The fundamental difference is that NMF does not need an infinite chain of homunculi. Instead, there is a hierarchy of the mind models. The conscious differentiated aspect of the models decreases at higher levels, and they are more uncertain and fuzzy. Until at the top of the hierarchy there are mostly unconscious models of the meaning of existence.

The origin of concepts is from two sources, inborn archetypes and cultural models transmitted by language. Vague and unconscious models-concepts evolve into more crisp and conscious ones. Psychologically this process was called by Carl Jung *differentiation* of psychic content [12].

## 3. HIERARCHY AND SYNTHESIS

At each level of the NMF hierarchy there are input signals from lower levels, models, similarity measures, emotions, and actions. The activated models send input signals to the next level, where more general concept-models are recognized or created [19]. Each model finds its mental meaning and purpose at a higher level. For example, consider a model "chair." It has a "behavioral" purpose for sitting. In addition, "chair" has a "purely mental" purpose at a higher level, a purpose of recognizing a more general concept, say of a "concert hall," which model contains rows of chairs.

Models at higher levels are more general than at lower levels [4,17]. At higher cognitive levels, models correspond to objects, to relationships, to situations, etc. At still higher levels, even more general models reside, corresponding to complex cultural notions such as family, love, friendship, and abstract concepts such as law, rationality, etc. According to Kantian analysis [18], at the top of the hierarchy of the mind are models of the meaning and purpose of our existence, unifying our knowledge.

Models at the bottom of the hierarchy correspond to directly perceived objects. These models are more evolutionary than cultural. They are "grounded" in "real" objects in the world. This is not true for concept-models at higher levels. Abstract models cannot be perceived directly in the world (e.g., "rationality," or "purpose of life"). They accumulate in cultures due to languages. They are not automatically related to events in the world. For example, every five-year-old knows about "good guys" and "bad guys." Yet, nobody claim to perfectly use these models to understand the world. The study of mechanisms relating concepts of language and cognition have just begun [16,19,20].

As mentioned, models acquire additional meanings and purposes at higher levels. We enjoy solving complex problems. This emotional feel of harmony is because high level concepts unify many lower level concepts and increase the overall meaning and purpose of our diverse knowledge. Jung called this synthesis, which he emphasized is essential for psychological well being.

Synthesis, the feel of overall meaning and purpose of knowledge, is related to the meaning and purpose of life, which we perceive at the highest levels of the mind hierarchy. At those high levels models are intrinsically vague and undifferentiated; their conceptual and emotional contents are not quite separable. This inseparability, which we sometimes feel as a meaning and purpose of our existence, is essential for evolution and survival. If knowledge does not support this feel, the entire hierarchy would crumble, which was an important mechanism of decay of old civilizations. The KI demands satisfaction at the lowest levels of concrete objects, and also at the highest levels of the mind hierarchy, understanding of the entire knowledge in its unity, which we feel as meaning and purpose of our existence. This is the other side of the KI, a mechanism of *synthesis* [12].

## 4. EVOLUTION OF CONSCIOUSNESS AND CULTURES

Estimating many models from limited data is unreliable; many solutions are possible, one no better than the other. Psychologically, the emotional investment in each concept decreases with an increase in the number of concepts, and a drive for differentiation and creating more concepts subsides. Emotional investment in a concept is a measure of the purpose of this concept within the mind system, a measure of synthesis. Thus, the drive for differentiation requires synthesis. More synthesis leads to faster differentiation, whereas more differentiation decreases synthesis.

Synthesis is related to language: cognitive models are developed from language models. Another aspect of synthesis is hierarchy. Some concepts are used more often than other; they acquire multiple meanings, which is opposite to differentiation. These more general concepts "move" to higher levels. These more general, higher-level concepts are invested with more emotion. Synthesis increases.

To summarize, differentiation and synthesis are in complex relationships, at once symbiotic and antagonistic. Synthesis is related to hierarchical structure of knowledge and values. It leads to spiritual inspiration, creativity, fast differentiation, creation of knowledge, science and technology. At the same time, "too" high level of synthesis stifles differentiation. Emotional attachments to concepts make them difficult to modify or create new ones. Differentiation discounts psychological emotional values of individual concepts, and destroys synthesis, which was the basis for differentiation. In this paper we develop dynamic models of neural mechanisms of differentiation, synthesis, and hierarchy using measures averaged over population of interacting agents.

We characterize accumulated knowledge, or differentiation, by a "mean field" averaged quantity, D, the average number of concept-models used in a population. Differentiation involves developing new, more detailed models from the old ones. Therefore the speed of differentiation is proportional to accumulated knowledge,

$$dD/dt = aD. \tag{1}$$

Here, a is a constant. This equation leads to an exponential growth of knowledge,

$$D(t) = D_0 \exp(at). \tag{2}$$

This is a kind of model considered by Kurzweil [21]. The exponential growth in knowledge predicted by this model he calls singularity. However knowledge does not grow continuously in long term. In all cultures differentiation growth is sometimes interrupted and culture disintegrates or stagnates. E.g., Western culture disintegrated and stagnated during the Middle Ages. Whereas some researchers have attributed the disintegration of Roman Empire to barbarians or to lead poisoning [22], here we would like to search for possible intrinsic spiritual, neural mechanisms.

A realistic dynamic of knowledge accumulation involves synthesis, S. An instrumental measure available for sociological research [23] is an average emotional investment per concept in a society. With the growth of differentiation, the emotional value of every individual concept diminishes. The opposite process of synthesis growth is created in hierarchies [24]. Diverse, differentiated, knowledge at particular level in a hierarchy acquires meaning and purpose at the next level. The simplest measure of hierarchy, H, is the number of hierarchical levels, on average, in the minds of the population. Accounting for hierarchical synthesis, can be written as

$$dS/dt = -bD + dH. \quad (3)$$

Here, b and d are constants.

A realistic equation for differentiation would account for the following. The speed of differentiation is proportional to accumulated knowledge, D, and is enhanced by synthesis, S, and is therefore proportional to D*S. We have to take into account that, psychologically, synthesis is a measure of the meaning and purpose in knowledge and culture, it is a necessary condition for human existence, and it has to remain positive. When synthesis falls below certain positive value, $S_0$, knowledge loses any value, culture disintegrates, and differentiation reverses its course, i.e.,

$$dD/dt = a\,D\,(S - S_0). \quad (4)$$

If the hierarchy, H, is genetically or culturally fixed to a constant value, eqs. (3) and (4) have several joint solutions. Let us explore them. First, there is a long-term solution with constant knowledge and synthesis:

$$D = (b/d)\,H; \quad S = S_0. \quad (5)$$

Here, differentiation and synthesis reach constant values and do not change with time. The hierarchy of concepts (and values) is rigidly fixed. This could be a reasonable solution, describing highly conservative, traditional societies in a state of cultural stagnation. The conceptual hierarchy, H, reaches a certain level, then remains unchanged, and this level forever determines the amount of accumulated knowledge or conceptual differentiation. Synthesis is at a low level $S_0$. All cultural energy is devoted to maintaining this synthesis, and further accumulation of knowledge or differentiation is not possible. Nevertheless, such a society might be stable for a long time. Some Polynesian and New Guinean cultures, lacking writing or complex religion and practicing cannibalism, still maintained stability and survived for millennia [25]. Chinese culture had stagnated since early BCE until recent times, although at much higher level of the hierarchy. It would be up to cultural historians and social scientists to evaluate whether such cultures are described by the above mathematical solution and, if so, what particular values of model parameters are appropriate.

Alternatively, if evolution starts with $S > S_{0,}$ differentiation first grows exponentially ~ exp( a $(S-S_0)$ t ). This eventually leads to the term –bD in (3) overtaking dH, so that synthesis diminishes and the differentiation growth exponent is reduced. When $S < S_{0,}$ differentiation falls until bD = dH, at which point differentiation grows again, and the cycle continue, leading to oscillating differentiation and synthesis.

Expanding knowledge in the long term requires expanding hierarchical levels. Knowledge accumulating at a particular level in the hierarchy may lead to certain concept-models being used more often than others. These concepts used by many agents in a population in slightly different ways acquire more general meanings and give rise to concepts at a higher level. Thus, increasing differentiation may induce more complex hierarchy, i.e.,

$$dH/dt = e\,dD/dt. \quad (6)$$

Eqs. (6), (4), and (3) describe a culture expanding in its knowledge content and in its hierarchical complexity. For example, a solution with fixed high level of synthesis can be described by

$$D(t) = D_0 \exp(\,a(S - S_0)t\,), \quad S = \text{const} > S_0,$$
$$H(t) = H_0 + e_c\,D_0 \exp(\,a(S - S_0)t\,). \quad (7)$$

This implies the “critical” value for parameter e,

$$e_c = b / d\,. \quad (8)$$

If $e > e_c$, then synthesis, differentiation, and hierarchy grow indefinitely. This unbounded growth is too optimistic compared to the actual evolution of human societies.

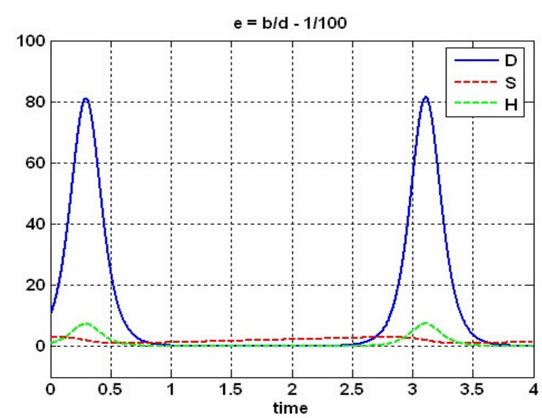


Fig. 1. Evolution of differentiation and synthesis described by eqs. (3, 4, 6) with parameter values a = 10, b = 1, d = 10, $S_0$=2, $H_0$ = 3, e = 0.99 < b/d, and initial values D(t=0) = 10, S(t=0) = 3, Alternating periods of cultural growth and stagnation.

If $e<e_c$, then synthesis and knowledge hierarchy collapse when differentiation destructs synthesis. However, when

differentiation falls, $H_0 > e_c D_0 \exp( a(S - S_0)t )$, synthesis again starts growing, leading to the growth of differentiation. After a fast flourishing period, synthesis again is destructed by differentiation when its influence on synthesis overtakes that of the hierarchy, and culture collapses. These periods of collapse and growth alternate, as shown in Fig. 1.

Assumption (6) of the hierarchy growing in sync with differentiation is too optimistic. The growth of hierarchy involves the differentiation of models at the highest level, concepts of the meaning and purpose of life. They cannot be made fully conscious, and in many societies they involve theological and religious concepts of the Highest. Changes in these concepts involve changes of religion, such as from Catholicism to Reformation, they involve national upheavals and wars, and they do not always proceed smoothly as in (6). Currently we do not have theory adequate to describe these changes; therefore we proceed within a single fixed religious paradigm. This can be approximately described as constant hierarchy H, as in the previous section. Alternatively we can consider slowly expanding hierarchy,

$$H(t) = H_0 + e^*t. \qquad (9)$$

The solution of eqs. (3, 4, 9) is illustrated in Fig. 2.

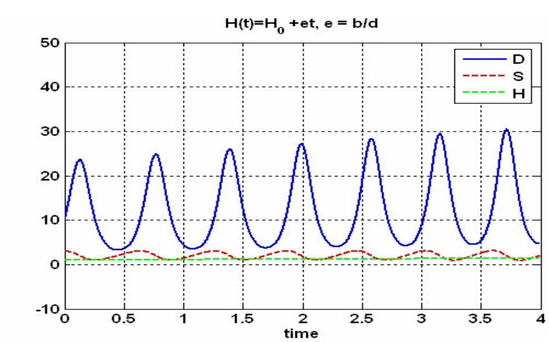


Fig. 2. Oscillating and growing differentiation and synthesis (eqs. 3, 4, 9); slow growth corresponds to slowly growing hierarchy, e = 0.1.

This growing and oscillating solution might describe Judeo-Christian culture over the long period of its cultural evolution. Note, the evolution and recoveries from periods of stagnation in Western culture were sustained by the growing hierarchy of knowledge and values. This stable, slow growing hierarchy was supported by religion. However, science has been replacing religion in many people's minds (in Europe more so than in the US) approximately since the Enlightenment (the 18th c.). The current cultural neurodynamics in Western culture are characterized by the predominance of scientific differentiation and the lack of synthesis. More and more people have difficulty connecting scientific highly-differentiated concepts to their instinctual needs. Many turn to psychiatrists and take medications to compensate for a lack of synthesis. The stability of Western hierarchical values is precarious, and during the next down-swing of synthesis hierarchy may begin to disintegrate, leading to cultural collapse. Many think that this process is already happening, more so in Europe than in the US.

So far we considered only the inspirational role of synthesis. The effect of synthesis, as discussed previously, is more complex: high investment of emotional value in every concept makes concepts "stable" and difficult to modify or differentiate [26]. Therefore, a high level of synthesis leads to stable and stagnating culture. We account for this by changing the effect of synthesis on differentiation as follows:

$$dD/dt = a\,D\,G(S), \quad G(S) = (S - S_0) \exp(-(S-S_0)/ S_1) \qquad (10)$$
$$dS/dt = -b\,D + d\,H \qquad (11)$$
$$H(t) = H_0, \text{ or } H(t) = H_0 + e^*t. \qquad (12)$$

Solutions similar to those previously considered are possible: a solution with a constant value of synthesis similar to (7), as well as oscillating and oscillating-growing solutions.

A new type solution possible here involves a high level of synthesis with stagnating differentiation. If dH > bD, then according to (11) synthesis grows exponentially, whereas differentiation levels off, and synthesis continues growing. This leads to a more and more stable society with high synthesis, with high emotional values attached to every concept, while knowledge accumulation stops.

Cultural historians might find examples of stagnating internally stable societies. Candidates are Ancient Egypt and contemporary Arab Moslem societies. Of course, these are only suggestions for future studies. Levels of differentiation, synthesis, and hierarchy can be measured by scientific means, and these data should be compared to the model. This would lead to model improvement, as well as to developing more detailed models [27]. And we hope that understanding of the processes of cultural stagnation will lead to overcoming these predicaments and to improvement of human condition.

Let us now study the interaction of cultures with different levels of differentiation and synthesis. Both are populations of agents characterized by NMF-minds and evolutionary eqs. (10, 11, 12). Culture k=1 is characterized by parameters leading to oscillating, potentially fast growing, differentiation and a medium oscillating level of synthesis ("dynamic" culture). Culture k=2 is characterized by slow growing, or stagnating, differentiation and high synthesis ("traditional" culture). In addition, there is a slow exchange by differentiation and synthesis among these two cultures (examples: the US and Mexico; or in general, immigrants to the US from more traditional societies; or academic-media culture within the US and "the rest" of the population). Evolutionary equations modified to account for the inflow and outflow of differentiation and synthesis can be written as

$$dD_k/dt = a_k D_k G(S_k) + x_k D_{\underline{k}} \qquad (13)$$
$$dS_k/dt = -b_k D_k + d_k H_k + y_k S_{\underline{k}} \qquad (14)$$
$$H_k = H_{0k} + e_k{}^*t \qquad (15)$$

Here, the index $\underline{k}$ denotes the opposite culture, i.e., for k=1, $\underline{k}$ = 2, and v.v. Fig. 3 illustrates sample solutions to these equations.

In Fig. 3 the evolution starts with two interacting cultures, one traditional and another dynamic. Due to the exchange of differentiation and synthesis among the cultures, traditional culture acquires differentiation, looses much of its synthesis, and becomes a dynamic culture. Let us emphasize that although we tried to find parameter values leading to less oscillations in differentiation and more stability, we did not find such solutions. Although parameters determining the exchange of differentiation and synthesis are symmetrical in two directions among cultures, it is interesting to note that traditional culture does not initially "stabilize" the dynamic one, the effect is mainly one-directional, that is, traditional culture acquires differentiated knowledge and dynamics. Wild swings of differentiation and synthesis subside a bit only after $t > 5$, when both cultures acquire a similar level of differentiated knowledge; then oscillations can partly counterweigh and stabilize each other at relatively high level of differentiation. It would be up to cultural historians and social psychologists, to judge if the beginning of this plot represents contemporary influence of American culture on the traditional societies. And if this figure explains why the influence of differentiation-knowledge and not highly-emotional stability-synthesis dominates cultural exchanges (unless "emotional-traditionalists" physically eliminate "knowledge-acquiring ones" during one of their period of weakness). Does partial stabilization beyond $t > 5$ represent the effect of multiculturalism and explain the vigor of contemporary American society?

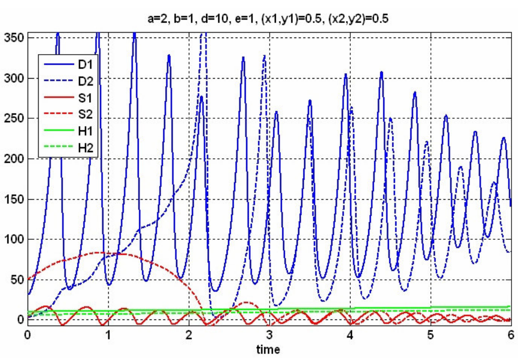


Fig. 3. Effects of cultural exchange (k=1, solid lines: D(t=0)= 30, $H_0$ = 12, S(t=0) = 2, $S_0$ = 1, $S_1$ = 10, a = 2, b = 1, d = 10, e=1, x = 0.5, y = 0.5; k=2, dotted lines: D(t=0)= 3, $H_0$ = 10, S(t=0) = 50, $S_0$ = 1, $S_1$ = 10, a = 2, b = 1, d = 10, e=1, x = 0.5, y = 0.5). Transfer of differentiated knowledge to less-differentiated culture dominates exchange during $t < 2$ (dashed blue curve). In long run ($t > 6$) cultures stabilize each other and swings of differentiation and synthesis subside (note however, that in this example hierarchies were maintained at different levels; exchange of hierarchical structure would lead to the two cultures becoming identical).

## V. FUTURE DIRECTIONS

High levels of differentiation, according to models in the previous section, are not stable. By destroying synthesis, differentiation undermines the very basis for knowledge accumulation. This led in the previous section to wild oscillations in differentiation and synthesis. Here we analyze an important mechanism of preserving synthesis along with high level of differentiation, which will have to be accounted for in future models.

Since time immemorial, art and religion have connected conceptual knowledge with emotions and values, and these provided cultural means for maintaining synthesis along with differentiation. A particularly important role in this process belongs to music, since music directly appeals to emotions [28,29].

Music appeared from the sounds of voice, i.e., from singing. Images of neural activity show that the human brain has *two centers controlling melody of speech*; an ancient center located in the limbic system, and a recent one in the cerebral cortex. The ancient center is connected with direct uncontrollable emotions, whereas the recent center is connected with concepts and consciously controlled emotions [29,30].

Prosody of speech in animals is governed by a single ancient emotional center in the limbic system. Sounds of animal cries engage the entire psyche, rather than concepts and emotions separately. A cry of danger is *inseparably* fused with recognition of a dangerous situation, and with a command to oneself and to the entire flock: "Fly!" An evaluation (emotion of fear), understanding (concept of danger), and behavior (cry and wing sweep) – are not differentiated. The conscious and unconscious are not separated: recognizing danger, crying, and flying away is a fused concept-emotion-behavioral *synthetic* form of thought-action. Animals can not control their larynx muscles *voluntarily*.

Emotions-evaluations in humans have separated from concepts-representations and from behavior. For example, when sitting around the table and discussing snakes, we do not jump on the table uncontrollably in fear every time "snakes" are mentioned. This *differentiation* of concepts and emotions is driven by language. Prosody or melody of speech is related to cognition and emotions through aesthetic emotions. *The human voice engages concepts and emotions. Melody of voice is perceived by ancient neural centers involved with archetypes, whereas conceptual contents of language involves conscious concepts. Human voice, therefore, involves both concepts and emotions; its melody is perceived by both conscious and unconscious; it maintains synthesis and creates wholeness in psyche.* [31]

Over thousands of years of cultural evolution, music perfected this inborn ability. *Musical sound engages the human being as a whole*—such is the nature of archetypes, ancient, vague, undifferentiated emotions-concepts of the mind. By turning to archetypes, music gets to the most ancient unconscious depths as well as to the loftiest ideas of the meaning of existence. This is why folk songs, popular songs, or opera airs might affect a person more strongly than words or music separately.. *High forms* of art effect synthesis of the most important models touching the meaning of human existence. *Popular songs*, through interaction of words and sounds, connect the usual words of everyday life with the depths of the unconscious. This explains why in contemporary culture, with its tremendous number of differentiated concepts and lack of meaning, such an important role is taken by popular songs. [7,16,32].

Whereas language evolved as the main mechanism for the differentiation of concepts, music evolved as the main mechanism for the differentiation of emotions (conscious emotions in the cortex). This differentiation of emotions is necessary for unifying differentiated consciousness: synthesis of differentiated knowledge entails emotional interactions among concepts [33]. This mechanism may remedy a disturbing aspect of the oscillating solutions considered in the previous section, the wild oscillations of differentiation and synthesis. Future research will have to make the next step, and define the mechanism by which differentiated aesthetic emotions unify contradictory aspects of knowledge. We will have to understand processes in which the KI differentiates itself and the synthesis of differentiated knowledge is achieved.

Future experimental research will need to examine, in detail, the nature of hierarchical interactions, including the mechanisms of emerging hierarchy: to what extent the hierarchy is inborn vs. adaptively learned. Studies of the neurodynamics of interacting language and cognition have already begun [7,26,,34]. Future research will need to model the differentiated nature of the KI. Unsolved problems include: neural mechanisms of emerging hierarchy, interactions between cognitive hierarchy and language hierarchy [35]; differentiated forms of the KI accounting for emotional interactions among concepts, the infinite variety of aesthetic emotions perceived in music, their relationships to mechanisms of synthesis [16,31,32].

Cultural historians can use the results of this chapter as a tool for understanding the psychological mechanisms of cultural evolution. The results may explain how differentiation and synthesis have interacted with language, religion, art, and especially music, and how these interactions have shaped the evolution of various cultures. Social psychologists can use the results of this chapter as a tool for understanding the psychological mechanisms governing present conditions. It is possible to measure the levels of differentiation and synthesis in various societies, and to use this knowledge for improving human conditions around the world. It will also be possible to predict future cultural developments, and to use this knowledge for preventing strife and stagnation, and for stimulating wellbeing.

## ACKNOWLEDGMENTS

I am thankful to D. Levine, R. Deming, B. Weijers, and R. Kozma for discussions, help and advice, and to AFOSR for supporting part of this research under the Lab. Task 05SN02COR, PM Dr. Jon Sjogren.

## REFERENCES

[1] Harlow, H.F., & Mears, C. (1979). The Human Model: Primate Perspectives, Washington, DC: V. H. Winston and Sons.

[2] Berlyne, D. E. (1960). Conflict, Arousal, And Curiosity, McGraw-Hill, New York, NY; Berlyne, D. E. (1973). Pleasure, Reward, Preference: Their Nature, Determinants, And Role In Behavior, Academic Press, New York, NY.

[3] Festinger, L. (1957). A Theory of Cognitive Dissonance, Stanford, CA: Stanford University Press.

[4] Grossberg, S. (1988). *Neural Networks and Natural Intelligence.* MIT Press, Cambridge, MA.

[5] Zeki, S. (1993). *A Vision of the Brain* Blackwell, Oxford, England.

[6] G. Ganis and S. M. Kosslyn, 2007, Multiple mechanisms of top-down processing in vision.

[7] Perlovsky, L.I. (2006). Toward Physics of the Mind: Concepts, Emotions, Consciousness, and Symbols. Phys. Life Rev. 3(1), pp.22-55.

[8] Grossberg, S. (1988). *Neural Networks and Natural Intelligence.* MIT Press, Cambridge, MA.

[9] Zeki, S. (1993). *A Vision of the Brain* Blackwell, Oxford, England.

[10] Grossberg, S. & Levine, D.S. (1987). Neural dynamics of attentionally modulated Pavlovian conditioning: blocking, inter-stimulus interval, and secondary reinforcement. *Psychobiology*, **15**(3), pp.195-240;

[11] Perlovsky, L.I. 2001. Neural Networks and Intellect: using model based concepts. New York: Oxford University Press

[12] Jung, C.G., 1921, *Psychological Types.* In the Collected Works, v.6, Bollingen Series XX, 1971, Princeton University Press, Princeton, NJ.

[13] Jung, C.G. (1934). *Archetypes of the Collective Unconscious.* In the Collected Works, v.9,II, Bollingen Series XX, 1969, Princeton University Press, Princeton, NJ.

[14] Taylor, J. G. (2005). Mind And Consciousness: Towards A Final Answer? Physics of Life Reviews, **2**(1), p.57.

[15] Jaynes, J. (1976). The Origin of Consciousness in the Breakdown of the Bicameral mind. Houghton Mifflin Co., Boston, MA; 2nd edition 2000.

[16] Perlovsky, L.I. (2007). *The Knowledge Instinct.* Basic Books. New York, NY.

[17] Zeki, S. (1993). *A Vision of the Brain* Blackwell, Oxford, England.

[18] Kant, I. (1798) Anthropologie in pragmatischer Hinsicht, see Anthropology from a Pragmatic Point of View, Cambridge University Press (2006), Cambridge, England.

[19] Perlovsky, L.I. (2006). *Modeling Field Theory of Higher Cognitive Functions*. Chapter in A. Loula, R. Gudwin, J. Queiroz, eds., Artificial Cognition Systems. Idea Group, Hershey, PA, pp.64-105.

[20] Perlovsky, L.I. (2006). *Neural Networks, Fuzzy Models and Dynamic Logic*. Chapter in R. Köhler and A. Mehler, eds., Aspects of Automatic Text Analysis (Festschrift in Honor of Burghard Rieger), Springer, Germany, pp.363-386.

[21] Kurzweil, R. (2005). The Singularity Is Near: When Humans Transcend Biology, Viking.

[22] Alexander Demandt: 210 Theories, from Crooked Timber weblog entry August 25, 2003.

[23] C. L. Harris, A. Ayçiçegi, and J. B. Gleason, Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language, Applied Psycholinguistics (2003), 24, pp. 561-579

[24] Perlovsky, L.I. (2007). Neural Dynamic Logic of Consciousness: the Knowledge Instinct. Chapter in Neurodynamics of Higher-Level Cognition and Consciousness, Eds. Perlovsky, L.I., Kozma, R. Springer, accepted for publication.

[25] J. Diamond, Collapse: How Societies Choose to Fail or Succeed, Viking, New York, NY, 2004.

[26] Perlovsky, L.I. (2006). Symbols: Integrated Cognition and Language. Chapter in A. Loula, R. Gudwin, J. Queiroz, eds. Semiotics and Intelligent Systems Development. Idea Group, Hershey, PA, pp.121-151.

[27] Perlovsky, L.I. (2005). Evolving Agents: Communication and Cognition. Chapter in Autonomous Intelligent Systems, Eds: V. Gorodetsky, J. Liu, V.A. Skormin. Lecture Notes in Computer Science, 3505 / 2005. Springer-Verlag GmbH.

[28] Crystal, D. (1997). The Cambridge encyclopedia of language, second edition. Cambridge: Cambridge University Press.

[29] L.I. Perlovsky, Evolution of Consciousness and Music, Zvezda, 2005 (8), pp. 192-223 (Russian); http://magazines.russ.ru/zvezda/2005/8/pe13.html

[30] Damasio, A.R. (1994). Descartes' Error: Emotion, Reason, and the Human Brain. Grosset/Putnam, New York, NY.

[31] Perlovsky, L.I. (2006). Co-evolution of Consciousness, Cognition, Language, and Music. Tutorial lecture course at Biannual Cognitive Science Conference, St. Petersburg, Russia.

[32] L.I. Perlovsky, Music – The First Principles, 2006, http://www.ceo.spb.ru/libretto/kon_lan/ogl.shtml.

[33] Perlovsky, L.I. (2006). Joint Evolution of Cognition, Consciousness, and Music. Lectures in Musicology, School of Music, University of Ohio, Columbus.

[34] Fontanari, J.F. and Perlovsky, L.I. (2006). Meaning creation and communication in a community of agents. World Congress on Computational Intelligence (WCCI). Vancouver, Canada.

[35] Perlovsky, L.I. (2004). Integrating Language and Cognition. *IEEE Connections*, Feature Article, **2**(2), pp. 8-12.