ALFUS Workshop #6

July 28-29, 2004
SAIC, Huntsville, Alabama

Meeting Report

A. Summary

Two main objectives for this workshop were exit strategy and framework development. A significant part of the discussions was devoted to the product of the ALFUS effort, as the product was considered an integral part of the overall exit strategy. Several components were identified, namely:

- Tool (currently conceptualized in spreadsheet with macro functions) for calculation and integration of autonomy level metrics.
- User guide.
- White paper describing the framework and its rationale. Plans and recommendations for testing and validation are also to be described.
- Summary/Executive model (e.g., concise autonomy level descriptors) and Terminology to facilitate communication at the levels of combat developers, managers, and users.

The group identified several exit strategy targets, including SAE AS-4 Committee (Aerospace Standards), JTA/DISR, and JRP[1]. Interaction with Jeff Kotora, representing AS-4, resulted in the identification of several possible approaches for collaboration, including offering the ALFUS products of work as recommended practices under AS-4 and participating as an AS-4 subcommittee. The starting point for further exploration and collaboration should be spring of 2005 when the SAE Committee is expected to be approved and operational.

We planned to continue following the DISR evolution process to determine an appropriate approach for interaction. We also planned to ramp up interaction with JRP per its revision cycle.

On framework development, a metrics measurement process, various metrics integration approaches, and associated examples were presented and discussed. These helped the tool development effort. All the three metrics development subgroups welcomed new volunteers for future participation.

Many programs/initiatives/organizations began exploring applying the ALFUS results.

B. Action Items

Please:

- Forward thoughts on the tool either to Charlie B. (cc. Hui) or through the email distribution list for discussion.
- Continue participating in metrics development work. This includes contributing to the other components of the identified product.
- Charlie/Hui to decide whether to submit an abstract for tool demonstration/presentation in the next year's SPIE Symposium on Defense and Security.

[1] ANSI, also suggested earlier but not discussed due to the time constraint, could also be explored.

- Help disseminating ALFUS results in your organization, up the chain of command and horizontally.
- Look into new scenarios (section C, subsection 2).
- NIST to clarify standards issues as described in section C, subsection 4.2.

C. Discussions

1. Content of the ALFUS product

The group agreed that we should stress delivering concrete, tangible products as opposed to serving as a forum for discussions. One aspect of the exit strategy is the content of the product. The identified product components include a tool for the metrics, user manual, and a white paper describing the rationale of the framework. The product should also include vocabulary which, in turn, includes the terminology and descriptors for the three metric axes and for the summary autonomy scale.

1.1 Tool and Manuals

Charlie B. conceptualized a spreadsheet based tool that listed metrics on the top row and task decomposition on the left column. The idea is to allow mission to be decomposed into subtasks at sufficient levels of detail as programs desire. Each of the tasks can be measured against the metrics. We should be able to design the spreadsheets such that the individual metric scores for a mission or a task can be automatically "summed" up according to the desirable integration algorithms (discussed separately).

Charlie volunteered to head the tool effort. Participants were urged to send him comments or ideas. On task decomposition, Charlie presented the concept of applying "automating human cognitive skills" as a guideline. Hui H. brought up that the NIST 4D/RCS architecture described a method for task decomposition. Ray H. also commented that he could contribute his experience. All of these could be consolidated and incorporated in the tool.

User community, such as UAMBL, needs to participate in this effort, as it is desirable that the tool is compatible with the current practices.

To describe how to use the tool, the manuals should include a set of examples or use cases from various domains.

The tool might evolve into a system specification tool.

1.2 Executive/Summary Model

This model aimed at providing a set of representations for autonomy levels. The issues include clarity, simplicity, and subjectivity vs. objectivity. Is it better to use a single number (or descriptor) or three numbers (or descriptors) to indicate autonomy levels? An advantage of using a single number would be simplicity, which might facilitate focused discussions. But, a disadvantage might be the ambiguity in the information that the number is to convey. Kerry commented that executive or operational users may desire simple and descriptive vocabulary. The vocabulary does not have to be precise nor does it have to be fully objective ("educated subjectivity"). Using one number may be more desirable, but three numbers should be fine (or "one number plus three caveats"), too. In either case, it would be discouraging to those types of users if they are required to go through the details of metrics before they can talk about the autonomy levels of their systems.

Ambiguity can also arise when different sets of the three numbers can be computed for the same autonomy level. For example, the triplets of (3,4,5) and (1,7,4), representing (HRI, Environmental Difficulty, Mission

Complexity), could both represent autonomy level 4 when we average the numbers, un-weighted. Kerry commented that this should be OK from an executive user's point of view.

Several formats could be considered for the model, including:

(a) descriptor based. Suggestions included:
    1. developing the descriptors for several key levels and not for all the but ones,
    2. developing specific descriptor sets for particular domains,
    3. descriptor vocabulary should be chosen to reflect the expected measurements from the involved metrics.

    Certain autonomy levels may be required to meet performance requirements, but, should the descriptors be developed based on performance? Performance, capability, requirement specifications, and autonomy levels are related but distinguishable terms.

(b) A method in which parameters and/or guidelines for the levels are given, guiding questions will be asked, and examples or references are provided to help users determining the levels. Examples:
    1. If a UMS is required to perform the X and Y missions under the Z and W environmental conditions and subject to the U and V types of HRI, then the UMS requires autonomy level M.
    2. The test course X is evaluated to be at the Environmental Difficulty level N.

Brian N. presented a first attempt HRI descriptors in a previous workshop. Woody E. and Hui H. showed their attempts, on Environmental Difficulty and Mission Complexity, respectively, in this workshop. Additional work is required and the suggested focus would be to reflect the metric measurement processes.

1.3 Framework White Paper

The white paper is to describe the framework and explain the product and the process. Plans for system testing and validation processes (see section 2) and for the dissemination of ALFUS are also to be included.

The SPIE paper looks appropriate to serve as a good foundation for this white paper. Hui H. should continue coordinating the effort and incorporate peoples' input.

1.4 In Summary (post meeting notes)

The ALFUS product includes the following volumes:

- Volume 1: Terminology
- Volume 2: Framework White Paper
- Volume 3: Executive/Summary Model: descriptors, tables, guidelines
- Volume 4: Detailed Model: Tool and Manuals

1.5 Scheduling and Approach

On the Tool:

- From now through prior to the next meeting: suggestions are solicited, please either use the email distribution list for open discussions or forward them directly to Charlie (cc. Hui for recording purposes).
- By the end of September, Charlie and Hui should determine whether the concept of the tool has progressed enough to warrant submitting an abstract for a presentation/demonstration in the next year's SPIE Defense and Security Symposium.
- By the next meeting, Charlie planned to complete the design of the tool and hoped that a prototype could be demonstrated (threshold and objective, respectively, using Kerry's terms).

On the white paper and the executive/summary model:

- Comments are requested, a draft should be shown in the next meeting.

2. Testing and Validation

An example question was how does one define and measure a HRI duration? This would involve designing testing methods: how many runs are required to calculate the statistics, how or whether to hold other factors constant, how is the boundary of a human interaction period identified, etc. We agreed that the group, being volunteer based, probably does not have sufficient resources and expertises to fully develop this testing and validation process. As such, this process should not be a part of the ALFUS framework as we are configuring. Rather, we should describe a set of recommendations, an approach, and/or a plan in the white paper. We certainly could pursue this issue if funding could be secured.

Also discussed was the need for scenarios during the framework development stage. Besides what we already have, additional scenarios could include, possibly: Charlie on an ANS scenario, GDRS and SPAWAR might have scenarios on warehouse security, Honeywell might have some on scout missions for ATTD.

3. Customers

Not in any order:

Kerry began to use some of the ALFUS concepts and terms in the Glossary, ORD, and O & O evolution processes in FCS. Would continue exploring further applying the ALFUS results. He commented that the timing for the ALFUS work is good when a lot of programs have the needs.

Charlie B. is heading the tool effort with the objective that the product would be helpful in evaluating the FCS ANS autonomy levels and in providing explanations on why the system has achieved contractual autonomy level requirements. He also suggested that our product might be helpful to some other FCS manned vehicles and he might be able to facilitate interactions.

Brian N. presented that TARADEC is beginning to experiment with the HRI metrics in their programs. He presented how the metrics may be applied to a selected scenario to measure operator interaction.

Bob S. (in an pre meeting message) had his team in AFRL beginning to use the metrics to measure the complexity of some of their UAV activities.

Woody E. suggested that he might be able to help disseminating the results in the OSD JPO/JRP area, including, potentially, NUSE2.

Keith and Ray planned to look into applying the metrics to their UAV activities.

Steve S. suggested that TRADOC is scheduling a meeting to discuss various robotic roadmaps that various Army agencies have produced. A suggestion was for this group to attend the meeting and present the autonomy level framework, including the terminology, to help facilitating a common reference.

David Knichel reported, in a pre-meeting message, that many other agencies/programs are in the need for a standard autonomy level framework.

NIST plans to use the ALFUS framework as a foundation for a planned framework for the performance measurement of intelligent unmanned systems.

It is encouraging that many programs/initiatives have started exploring applying the ALFUS framework. As much as we all believe in the framework's value, it could only be demonstrated by successfully applications.

4. Exit Strategy/Standardization/Dissemination of Results

4.1 Targeted Standards/Major Communities

1. JTA (proposed by Dave and concurred by many)
   Is being transitioned into Defense Information Standards Repository (DISR). Efforts are ongoing within DOD in defining a process for adopting standards into DISR and we are monitoring the progress.

2. SAE / JAUS as a documentation vehicle (proposed by Kerry, Steve and concurred by many)
   Jeff Kotora briefed that JAUS is being petitioned as Committee AS-4 under the SAE Aerospace council. (AS-2 produced GOA). DoD converted a majority (or all?) of its MIL standards to SAE.

   Discussed multiple ways of collaborations. The emerging preferences were, first, to offer the ALFUS products of work as recommended practices under AS-4. This would not involve the ALFUS group to be a part of the SAE committee. The other preference would be for the ALFUS group to participate as a subcommittee under AS-4, parallel to the other JAUS subcommittees.

   Jeff suggested that we wait six months or so when the AS-4 is up and running before interacting with them for a proper path.

   Jeff requested that ALFUS to give an update in the upcoming JAUS meeting.

3. JRP as a documentation vehicle (proposed by Kerry)
   Discussed with Jeff. Will continue exploring a suitable path.

4. Steve S. proposed, in a post workshop message, for ALFUS to be presented in RDECOM Robotics IPTs to help with the roadmap efforts, to be viewed as recommended practices for FCS UMS developers, to be provided to developers of UMS Reliability and Maintainability Failure Definition and Scoring Criteria (FDSC) documents, and to be provided to agencies such as the TRADOC Futures Center, the Rapid Equipping Force (REF) and those supporting Training, Exercise and Military Operations (TEMO) for use in articulating general unmanned system domain capabilities and requirements.

4.2 Issues

The ALFUS process could serve as a reference model (as GOA) as opposed to a rigid standard. People refer to it but there would not be a compliance issue (proposed by Dave and others). SAE and other standards bodies maintain multiple levels of standards. The ALFUS process could fall in the category of recommended practices, while its terminology could be submitted as a standard (Woody, Steve, others). A common reference process would help reducing the amount of variations in the resulting autonomy levels. A commonly accepted and proven process also encourages practitioners to use the framework and helps dissemination of the ALFUS product.

It was commented that NATO STANAG 4586 might need a task decomposition process. Collaborations might be helpful.

Jeff raised the point that standards bodies copyright their standards. It is unclear whether such a policy would constrain ALFUS from pursuing wider adoptions beyond the single standards body and whether such

a policy may conflict with the Government public domain policy (e.g., such that it prevents the work from being published as NIST Special Publications). These issues need to be clarified before we approach the standards bodies.

5. Framework Development

5.1 Bandwidth

It was brought up that bandwidth requirements are major concerns (in FCS). In a situation when a UMS requests permission to fire weapon on a moving target, the response time is critical. Teleop also requires sufficient bandwidth to be effective. The question to ALFUS would be how does the framework take these requirements into consideration? Could partially be handled in Mission Complexity, where the Collaboration/Interface requirement may dictate the bandwidth requirement? Bandwidth could also be considered as an environmental constraint. Further investigation is required.

5.2 Configuration of a UMS, Consistent and scalable process

It was commented that we should develop the metric evaluation process properly such that:

- the application of the metrics at the subsystem levels and at the system level would yield consistent results,
- applying the process by different people would yield comparable results.

A UMS may employ an off-board autonomous subsystem, such as a mission planner that is embedded in a battle command. In this situation, what would be the configuration of the UMS? Could we consider the planner as a part of the UMS? If not, how is the autonomy level of the vehicle determined, when it may interface with a non-human and autonomous (planning) system? In the newly focused net-centric environmental model, data are perceived to be produced and shared from distributed locations. How does this affect the autonomy level measure? How is such a system configured?

Is “continuity of geometric structure” a good criterion for system boundary? How about using human interface as the boundary?

A interesting question would be whether a system with a high level autonomy could be composed of components or subsystems with only low or mid levels of autonomy.

5.3 Metrics

Brian N. updated the HRI metrics. He also described a set of FCS mission characteristics and how the HRI metrics may be applied to measure the missions. The effort could be regarded as parallel to the technology readiness level (TRL) measurement effort. One of the key metrics is the amount of time operators have to spend to fix problems. Common statistics seemed to be that operators have to spend 20% to 30% of the time to teleop or fix mobility problems and 50 -60% of the time on payload.

NAVSEA might have data and effort for testing HRI that Keith and Ray might be able to access and help with the HRI effort.

What we should stress would be guidelines and not mission specific. System specific task decomposition can serve only as use cases for the ALFUS effort.

Hui H. suggested that a simple way of categorize mission complexity might be to just use the commanding levels as described in the NIST 4D/RCS reference architecture, with modification/expansion as needed. Bill K. concurred. We will investigate this further.

6. Integration of Metrics

Hui H. presented a metrics measurement process. Included were the various options, such as weighted min or max. Example used was to use remote control (100 % HRI) to perform complex missions in a difficult environment. Apply the weighted minimum approach on the three axes would result in autonomy level 1. Weighted average would yield higher autonomy levels, which might be deserving, arguably.

There could be numerous ways of assigning weights, linear or nonlinear, for example, (0 through 10), (1, 2, 3), (1, 5, 10), and (1, 2, 4, 8). It should be a part of our objective to provide some guidelines on the weight factor. Further investigation would be needed to find out what kind of imbalance in the system may be caused by putting too much weight on particular metrics.

The issue of metric interdependency was presented and acknowledged. John S. commented that the success rate might be affected by the effectiveness of HRI. This is another open issue.

Yet another open issue was that the autonomy level might be higher in some portion of a mission and lower in the others.

7. Terminology

The definition for Autonomy was discussed. A question was whether to uniquely rename the term to avoid possible confusion with other autonomy level efforts which might have different interests, such as biological or philosophical, whereas ALFUS focuses for the unmanned systems to perform human assigned missions. Names such as unmanned system autonomy and robotic autonomy were mentioned. Most of the participants felt that there was not an urgent need to make such a change.

We plan to change the title of the terminology document from:
Terminology for Specifying the Autonomy Levels for Unmanned Systems
Version 1.0

To

Autonomy Levels for Unmanned Systems (ALFUS) Framework
Volume I: Terminology
Version 1.0

in order to make the document more in line with the overall construct of the framework product.

8. Consistency with Existent Autonomy Level Work

Bob S., a principal with the AFRL ACL, has begun relating the ALFUS metrics with ACL. Ray H., a principal with the levels of autonomy behavior chart (published by Army Science Board) also planned to compare it with the ALFUS metrics. These would, hopefully motivate the ALFUS framework to achieve a higher level of genericity such that it can either help providing the metric level details or serve as a foundation for some of the existent work.

9. Workshop Issues

9.1 Next Meeting

Tentatively identified for the week of Oct. 18 and located at AFRL, Dayton, Ohio. Hui H. to work with Bob Smith and the Group to finalize the issue.

9.2 Meeting format

The current format is favored if the meeting location allows for most of the participants to travel in the morning. A single day meeting should be considered otherwise.

9.3 Miscellaneous

It should be contributors' responsibility to notify that the contributed material is restricted in its distribution. Dennis Overstreet is welcome to resume his participation (switched job to another government contract). Tom A. agreed to help in the Environmental difficulty and HRI metrics. John S. agreed to work with Charlie B. on the tool and related issues. Steve S. agreed to help in the Mission Complexity metrics.

D. ATTACHMENTS:

- Background slides.
- Integration Issues slides
- HRI slides
- Mission Complexity example
- Environmental Difficulty metrics

POC: hui-min.huang@nist.gov