10TH INTERNATIONAL COMMAND AND CONTROL RESEARCH AND TECHNOLOGY SYMPOSIUM

THE FUTURE OF C2

Command and Control Forensics

Topic: Decisionmaking and Cognitive Analysis

Kirk Dunkelberger, Mark Akey, Michael Downs, Maj. Ryan Paterson; USMC

Point of Contact: Kirk A. Dunkelberger

Northrop Grumman Corporation

6112 Constitution Drive, Fort Wayne, IN 46804

260 435-4603 (voice), 260 435-4639 (fax)

kirk.dunkelberger@ngc.com

ABSTRACT

The Command Post of the Future (CPOF) represents a significant step towards network-centric command and control (C2) that can provide a seamless, collaborative problem-solving environment spanning the command structure from division to company commanders. Currently deployed in Operation Iraqi Freedom (OIF), a complete snapshot, a *repository*, of the CPOF database has been archived at regular intervals for the past year. Never before in the history of warfare has a real time operational data been collected with this granularity or for this length of time. By the time DARPA transitions this technology to the Army we will have collected data on the deployment of three divisions during their tours of duty in Iraq.

The data collected during operations presents a unique opportunity to begin researching adaptive technology for command and control Analyses of the CPOF data can be leveraged to guide component and service development in the areas of learning and reasoning (accelerating the C2 processes for which CPOF is intended), knowledge management (increasing the efficiency of the data, information, and knowledge resident therein), team formation (accelerating the discovery and matching of problem to problem solver), as well as too many others to list. This paper documents early analysis of the CPOF data, provides a context for those analyses, and recommends current and future opportunities for adaptive command based on the results of those analyses.

PROBLEM

The Command Post of the Future (CPOF) represents a significant step towards network-centric command and control (C2). Sponsored by DARPA, CPOF is one of the first tools of its kind to provide a seamless, collaborative problem-solving environment that currently spans the command structure from division commanders to company leaders and has the potential to reach down to the squad level. The system is currently deployed in Operation Iraqi Freedom (OIF) with seventy-five peer workstations distributed throughout Baghdad. CPOF in and of itself is receiving rave reviews from users, decision-makers, and sponsoring agencies, but it is also providing another, perhaps even more important service that has the potential to impact C2 for decades to come.

Behind every collaboration architecture lays a shared database; at regular intervals (typically hourly) for the past year, a complete snapshot, a *repository*, of the CPOF database has been archived. Each snapshot contains not only the database contents, but also a log of the transactions occurring since the previous archive. In other words, for every hour of the last year of OIF, we have available to us a representation of the collective cognitive state of an actively deployed force engaged real time in a politically, tactically, and strategically challenging mission. The greatest problem facing the development of cognitive systems, aids for collaborative C2 being a prime example, is the lack of data on which to base the underlying model. Never before in the history of warfare has a real time operational data been collected with this granularity or for this length of time. By the time DARPA transitions this technology to the Army we will have collected data on the deployment of three divisions during their tours of duty in Iraq.

The data collected during operations presents a unique opportunity to begin researching adaptive technology for command and control Early exploration leads researchers to believe that a command system could adapt over time by monitoring how the user uses his system, monitoring the user's collaboration with other users, and monitoring the operational environment. The archived data (measured currently in terabytes) presents an almost unlimited opportunity…

RELEVANCE TO COMMAND AND CONTROL

CPOF is a system designed by its users. The underlying data model had few intrinsic schema and the users defined schema and sharing schemes to suit the problem at hand in real time. So through time, there has been a second, implicit collaboration scheme underway: the design of a C2 system. We have an audit trail of that design since its inception through all its versions to the present in the form of the collective repositories and the data therein. Analyses of the CPOF data can be leveraged to guide component and service development in the areas of learning and reasoning (accelerating the C2 processes for which CPOF is intended), knowledge management (increasing the efficiency of the data, information, and knowledge resident therein), team formation (accelerating the discovery and matching of problem to problem solver), as well as too many others to list.

The challenge is to identify the data and control patterns hidden in the archives and leverage them into tools for adaptive command and control. CPOF forensics has two primary foci: *technical* performance optimization (CPOF user, system) and *operational* effectiveness optimization (planning time, plan completeness, resulting battle tempo, etc.). Technical performance optimization is based on tools that are primarily task-independent. These tools are designed to minimize fighting with the system and maximize fighting the battle. Operational effectiveness optimization is based on tools that are primarily role-specific. These tools enable fluid cross-positional collaboration of the entire tactical operation center team; and should have different perspectives based on whether they are being used by organic (intel doing intel) or inorganic (maneuver using intel) personnel. MG Peter Chiarelli, Commanding General, 1st Cavalry Division (1CD), states, "We have just begun to mine the potential of this application, it will change the way we fight, change not only what our staffs do but also their size and composition." The traditional staff functions are transforming before our eyes and this transformation has been captured for analysis.

Dynamic Ad-hoc Teams. Will it be possible to leave behind the rigors of Napoleonic staff structures for a more agile, better informed team of professionals? The type of information found in CPOF and others envisioned to be available soon may be able to impact this area. Already evident in the operational data returning from country are emergent teams that form and dissolve as the situation dictates. Information is beginning to flow horizontally and leaving its traditional vertical pathways.

Anticipatory Support of Warfighters. Interviews with officers throughout 1CD give hints to their excitement about the ability to look over the boss' shoulder while he is in the early stages of thought formation. The S-3 officer from DISCOM stated, "For the first time in my career I feel like we are able to provide truly anticipatory logistics support to the division." Early CPOF experiments showed that one can easily change the word logistics to intelligence, fire, maneuver, communications… Now aid the human with context and patterns collected locally on his system and globally through out the operational environment. The term self-synchronized is no longer the latest buzz word but a near reality.

Proactive Information Retrieval. Today's command and control systems still provide little in the way of automated support for the warfighter. Many have gone to great lengths to digitize the analogue process on which most leaders learned, but few have done anything to really provide automation. The type of information found in the CPOF repositories begins to allow this to happen. Decision-makers at all levels will soon have to spend much less time thinking about where to find the right information and will rather spend their time on the executive decision processes that only humans can do.

CPOF Overview

The goal of Command Post of the Future (CPOF) is to improve the speed and quality of command decision making. In military decision making, as it is currently practiced, approximately 70% of commanders' time is spent on acquiring and maintaining situation awareness. The current practice of decompositional planning as codified in the Army's MDMP (Military Decision Making Process) and the Marine Corps MCPP (Marine Corps Planning Process) lead to decisions that have been shown to be inconsistent with the time-driven, uncertain nature of warfare and the cognitive processes of expert warfighters. Instead, military planning and decision-making is an intuitive process based on experiential pattern-recognition and mental simulation. CPOF will develop tools that leverage the strengths of expert commanders in achieving situation awareness, developing courses of action (COA), and making effective command decisions. CPOF has identified three functions of the command & planning process where technology can be applied to achieve these goals: 1) improved situation awareness (SA); 2) more natural capture of COA development; and 3) improved COA communication and collaboration.

The Command Post of the Future is currently deployed in support of the 1st Cavalry Division (1CD) in Iraq, as well as with the 3rd Infantry Division (3ID) now en-route to Baghdad. MG Chiarelli says of CPOF, *"I can't explain to you what you have done for me, because you don't know what it was like before we had CPOF. You have saved soldiers' lives…more of my soldiers will return home unharmed because of this technology*."

There are 75 CPOF client systems located in Iraq with the 1st Cavalry Division (1CD), primarily at the Division Headquarters (HQ) and subordinate Brigade Headquarters. In addition, we have a small number of clients supporting a liaison officer at Corps Headquarters and with selected aviation battalion units. These clients all communicate with a common repository server located at Division HQ. The Division uses CPOF to enhance situational awareness and operational/tactical planning and execution. The Division uses CPOF as the backbone of distributed BattleStaff Update Briefs (BUB) and Commander's Update Briefs (CUB). This results in a very high level of shared situational awareness between the Division and Brigade staffs without imposing a requirement for travel through dangerous areas to a central briefing location. CPOF enables multiple, distributed groups to quickly and easily collaborate on planning and execution issues. CPOF has become their primary C2 system. Finally, CPOF has provided the Commander the ability to easily mine historical operational data to allow him to predict likely enemy future actions.

The 1CD is demonstrating today the value of information collection, manipulation, and application in ways not previously possible. The use of CPOF has been identified as one of the first true applications of Network Centric Operations, where we can see how information can change the way we think and fight. While the Army will continue to support the use of CPOF in Iraq by fielding it to 3rd Infantry Division (3ID) and then 4th Infantry Division (4ID), there is yet more opportunity to impact that fight.

MACE / TACC / BattleZone Overview

During the summer of 2004 DARPA funded a small study to begin exploring the data repositories from Iraq. Dubbed the Multi-echelon Adaptive Command Environment or MACE study a team of researchers from DARPA, Northrop Grumman, Lockheed Martin and SET Associates began asking questions and exploring the data. The team started with the hypothesis that emerging learning technologies could be applied to real time operational data to provide commanders with proactive and adaptive technologies to enable, among other things, dynamic ad-hoc teams and anticipatory information retrieval. Two workshops were held during the study that brought together experts in the fields of learning, search and social network analysis. As part of the due diligence or proof of confidence the team from Northrop Grumman applied signal processing techniques to verify that indeed interesting and useful patterns existed in the repositories. The results of these investigations follow as the basis of this paper.

The most significant outcome of the MACE study was the realization that the development of command technologies was about to enter a new paradigm. All involved in the process began to realize the power of "Capturing Live Combat." CPOF was not only capturing vehicle positions, operational orders and graphics it was also capturing the very human side of command. How do the commander and his staff solve problems? What information do the individuals in the organization access, or not access, in the execution of their duties? Down to the individual ink strokes we can recreate the problem solving process of these units. Look forward now to a day when these types of collaboration tools get into the hands of leaders at the lowest tactical level. Look forward to a day when the young officer fresh out of The Basic School begins to build a profile of the types of problems he has solved, the way he handles information, the places he has been and the people he has worked with. We imagine a BattleZone. The BattleZone is more a concept than a piece of technology, it is comprised of multiple technical programs as well as ideas on the process of how we develop leaders for combat. The BattleZone is made up of three parts; BattleNet, BattleSchool and BattlePlex.

Phase 1: The **BattleNet** is an environment where military professionals conduct their trade. The environment extends to the entire operational cycle and moves from, a content unaware to a content aware environment. This will enable its users to observe and understand as they train, plan, rehearse, execute and review in their operational environment. All phases captured, all phases informing the others. Moving towards a true capability to capture live combat and make it accessible throughout operations.

In his monograph, Dr. Leonard Wong espouses the need for the military to capitalize on the chaos that is Iraq.

> This monograph examines the Operation IRAQI FREEDOM environment and concludes that the complexity, unpredictability, and ambiguity of postwar Iraq is producing a cohort of innovative, confident, and adaptable junior officers. Lieutenants and captains are learning to make decisions in chaotic conditions and to be mentally agile in executing counterinsurgency and nation-building operations simultaneously. As a result, the Army will soon have a cohort of company grade officers who are accustomed to operating independently, taking the initiative, and adapting to changes. (Wong, Developing Adaptive Leaders: The Crucible Experience of Operation Iraqi Freedom, July 2004)

Through the capture of live combat at ever increasing levels of granularity we should be able to harness some of the experience of these "adaptable junior officers."

Phase 2: BattleSchool embarks to be a fundamental shift in the way we prepare combat leaders at all levels of experience. As the captains of tomorrow enter their command level courses they leave behind the crates of text books and begin interacting with the operational environment of yesterday. Students at the Army's Command and General Staff College will be able to train and practice on an interactive encyclopedia of the operations that took Fallujah.

Phases 3 & 4: BattlePlex moves from an interactive environment to an immersive environment, students are now teamed with other students and virtual peers actually commanding the forces of yesterday in a simulated environment. Causality can be investigated and extreme scenarios can be practiced in realistic environments so that they can "learn to make decisions in chaotic conditions and to be mentally agile."

APPROACH

The sheer size, breadth, and depth of the CPOF data intuitively imply a layered approach to analysis. The MACE study focused on the initial stage of defining, parameterizing, and detecting discriminating characteristics of the data using statistical features; more specifically, the signal processing domain was mined for techniques for detecting, classifying, and localizing both local and global trends in the data. Identification of these local characterizations of the data led directly to inverse filters which discriminated abnormalities, anomalies, and features of interest within the data. Detailed, content sensitive analysis of areas of interest would logically follow. These analyses are the focus of the DARPA CPOF Pattern Discovery study currently underway.

Each CPOF user is represented by a workspace. This workspace contains (has references to) a time varying set of uforms. This set of uforms represents both the data of interest to the user (implicitly) and his/her organization of that data within a palette of views (explicitly). This initial set of analyses was based on gross ordinal and cardinal statistics of this data distribution, i.e.

- the number of uforms in a workspace over time,
- the number of uforms created, modified, and/or deleted over time,
- the overlap of (a count of common) uforms between workspaces over time, and
- the trajectories of both specific uforms and uform types through workspaces.

These core features were then aggregated into pairwise, arbitrary subset, and global measurements as appropriate for analyses.

These features can be likened to those commonly used in information retrieval (i.e. internet search engine) applications: term frequency, inverse document frequency, co-occurrence, inverse reference count, etc. The discriminating aspect of this investigation is that we have treated these features as vectors in time. These *signals* have frequency domain characteristics that indicate command and control patterns. They have impulse and step responses that indicate a system response to events. They have flow characteristics that indicate necessary, adaptive, ad-hoc network connectivity. Coming to grips with this set of analogies is critical in understanding our core approach to the analyses.

Six classes of analysis were performed under the MACE study. For each data analysis investigation, the table below lists these parameters:

- Warfighter Requirement: A hypothesis couched in terms of commander and staff C2 observations and reflections.
- Derived Analytical Hypothesis: Derived from the warfighter requirement, a derived hypothesis couched in terms of system analysis.
- Data Analysis Objective: Given the warfighter requirement and analytical hypothesis, the results we intended to discover, identify and capture.

The sections following the table provide an overview of each analysis in turn.

Warfighter Requirement	Derived Analytical Hypothesis	Data Analysis Objective
Infer and adapt to both formal and emergent informal organization	There exist matching activity patterns across individual workspaces	Discover inter-workspace workflow and infer intra-workspace procedure through cross correlation
Adaptive support for periodic activities as well as aperiodic events	There exist periodicities and variations thereof within a single workspace	Develop event detection filters and identify periodic events through Fourier analysis
Team formation acceleration	There exist temporal knowledge sharing patterns	Generalize inter-workspace content patterns and infer/recognize emergent patterns through linear prediction
Reduced latency in data access and presentation	(Variable order) Markov models can predict knowledge flow	Learn and generalize knowledge propagation and build knowledge flow predictors using compression techniques
Effective, efficient situation awareness	There exist general life cycle models for data (based on type, content, …)	Identify type/time clusters and characterize movement of clusters through type density moments
Protect blue information and identify red information vulnerabilities	Social network models and metrics can cue information operations	Identify gatekeepers, bottlenecks, gurus, etc. and characterize time-variant informal organization through SNA

Correlation Analyses: Matching Activity Patterns across Workspace Pairs

Warfighter Requirement	Derived Analytical Hypothesis	Data Analysis Objective
Infer and adapt to both formal and emergent informal organization	There exist matching activity patterns across individual workspaces	Discover inter-workspace workflow and infer intra-workspace procedure through cross correlation

Workspace activity was characterized as a count of the uforms created, added to the local workspace, modified, and deleted from the local workspace over time. In this and the other analyses, time was quantized to facilitate the analysis goal, i.e. weeks for system learning curve assessment, days for operational organization analysis, hours (or minutes) for tactical response characterization. In this particular case study, time was quantized to the hour and several days' worth of data was analyzed; this was appropriate for the analysis goal of identifying diurnal system usage patterns. The greatest contributor to the expected activity pattern was expected to be surges related to preparation for the BUB and CUB. Secondary contributors were expected to be (re)orientation at shift changes (at unknown and times varying between organizational units). Significant anomalies other than these would have to be manually associated with environmental events or otherwise resolved through consultation with the subject matter experts (SMEs).

Figure 1 displays a characteristic result of this class of analysis. In the left portion of the figure, the raw activity of two workspaces is displayed. Each is seen to have ebbs and flows, but both appear to have two major activity surges (albeit at different times). Intuition tells us to consider preparation for the briefings as the underlying cause. The right portion of the figure displays the correlation of activity with varying degrees of phase shift (time delay) imposed. It can clearly be seen that peaks in correlation occur at two specific times. In other words, the activity in one workspace is clearly offset from the other by one of these times. What we are seeing is that both workspaces are preparing for the BUB and CUB, however, one regularly prepares before the other (or the other prepares after, depending on perspective).

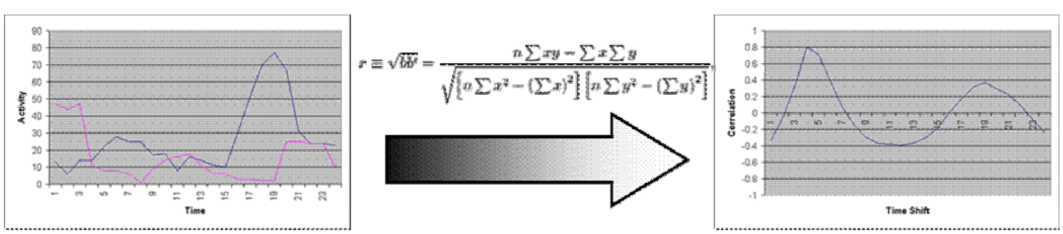

Figure 1. Pairwise cross correlation of activity to identify relative time shifts.

To what can we attribute this? Perhaps the shift change of the one workspace is offset from the other. Perhaps local operating procedure dictates that the briefing material be prepared early due to a dependency on content from outside sources (which are on an offset schedule). Once this type of effect (anomaly?) is identified by cross correlation analysis, the underlying cause can be identified through other methods. Again, recall the common goal of (computationally) cheaply identifying effects of interest then applying more expensive techniques in the critical areas.

Coordination of this analysis with other approaches (some of which are addressed in later sections of this paper) may identify that products from the first workspace owners are required by the second in order to complete their briefing preparation. In this case, this timeline could perhaps be shortened by pushing the products as they are completed rather than waiting for a pull. Consultation with the SMEs and consideration of the overall shift scheduling may indicate that an awkward sequence (i.e. data consumers changing shift just as products required become available) causes unnecessary delays in serial workflow. In this case, reorganization may be in order.

In either of these (or other) cases, this type of analysis helps to identify inter-workspace workflow through temporal dependencies. Understanding the workflow can provide cues for team formation and technical optimization in the form of data anticipation and active data push. Intra-workspace procedure can also be inferred. From this, bottlenecks due to global organizational issues can be identified and addressed. Keep in mind that individually, each of the discussed analysis techniques provides indicators for more in depth investigation. The combination of multiple techniques (or the combination of a technique with externally available information) can provide actionable guidance for adaptive command automation.

Fourier Series Analyses: Determining Periodicity Within A Single Workspace

Warfighter Requirement	Derived Analytical Hypothesis	Data Analysis Objective
Adaptive support for periodic activities as well as aperiodic events	There exist periodicities and variations thereof within a single workspace	Develop event detection filters and identify periodic events through Fourier analysis

While the previous analysis considered the relationship of diurnal activity between pairs of workspaces, this case study focused on identifying characteristic diurnal activity patterns within a single workspace. The (sampled) time sequence of workspace activity (determined from the count of the uforms created, added to the local workspace, modified, and deleted from the local workspace as previously) was converted to the frequency domain using a Fourier transform. The Fourier series is an approximation of the original time sequence built from a summation of sinusoidal basis functions at multiple frequencies; the Fourier domain representation consists of the coefficients on (weights of) the components in the approximation.

While it's not possible to completely describe the Fourier representation herein, what is important is that significant *periodic* events will be indicated by dominant components in the Fourier (frequency) domain. The left portion of Figure 2 displays the raw activity of a single workspace over time. The right portion displays the magnitude of the frequency domain representation (via a Fourier transform) of the activity. So, focusing again on the BUB and CUB, we should see two significant contributions. Unfortunately, they are not symmetric so the effects are not as apparent as one would like (since their frequencies must be represented by multiple coefficients).

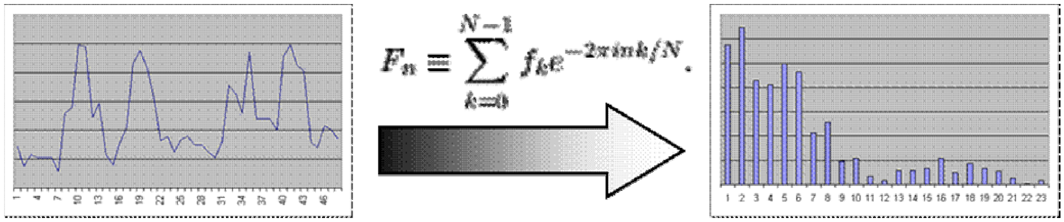

Figure 2. Fourier domain representation of activity within a single workspace.

From a more pragmatic viewpoint, efficient filtering operations can be performed in the frequency domain (via Fourier transforms). In Figure 3, the first row displays the feature that we'd like to detect (a spike in activity) and its representation in the frequency domain. Convolution (application of the filter) becomes a simple multiplication in the frequency domain (the second row). Finally, converting back to the time domain, we see the filtered activity; peaks represent matches and event detection is enabled.

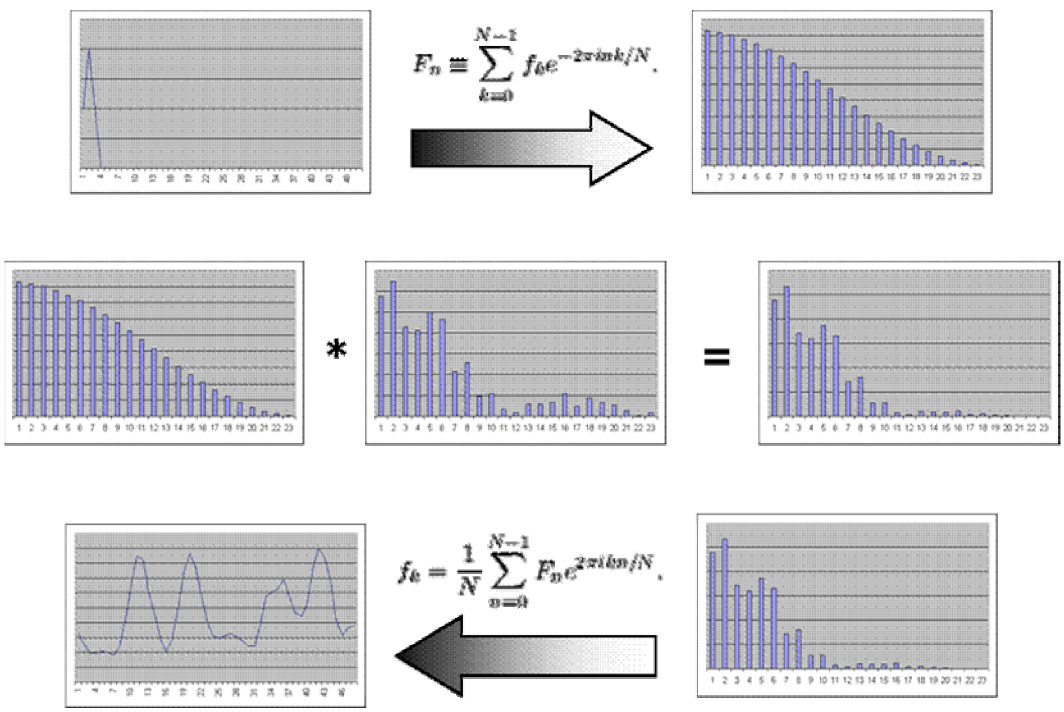

Figure 3. Matched filter for event detection in the frequency domain.

Analysis of activity levels through time in the frequency domain provides two core capabilities: the detection of periodic (and therefore aperiodic) activities and efficient search for particular activity patterns. Both of these capabilities can be leveraged into team formation recognition and intuitive knowledge adaptation through the continuous inference of informal periodic activities. Anticipatory retrieval automation, another direct result of this technique, can be leveraged to manage bandwidth and optimize proactive search and anticipatory retrieval efficiency.

Knowledge Sharing Analyses: Overlap of UUIDs between Workspaces over Time

Warfighter Requirement	Derived Analytical Hypothesis	Data Analysis Objective
Team formation acceleration	There exist temporal knowledge sharing patterns	Generalize inter-workspace content patterns and infer/recognize emergent patterns through linear prediction

In this analysis, the sharing of CPOF information over time was measured from the overlap of workspace content. Each uform in the system has a universal unique identifier (UUID) that is assigned on creation, is associated with that uform for its entire life cycle, and is guaranteed to be unique throughout its existence and beyond its deletion (they are not reused). While the database structure underlying CPOF is actually a directed, cyclic graph of uforms (which reference each other by UUID), a more traditional key-value access method is possible using UUIDs. Separating a repository snapshot into workspaces and indexing the contained uforms by UUID allows simple set operations (i.e. union, intersection, etc.) to determine knowledge sharing.

The sharing patterns that emerge (see Figure 4) fall into two intuitive categories: consistent and anomalous. These categories map directly to activities implied by formal organization (the consistent components) and activities initiated by events in the environment (the anomalous components). We also observe that there are sets of core knowledge (unit associations, geography, etc.) that are pulled on the creation of the workspace and persist through all workspace activities; further analyses along these lines should consider separation of and normalization against these baseline data (which naturally leads into the analyses discussed in *Type Density Analyses: Life Cycle Model of Uform Types over All Workspaces* later in this paper).

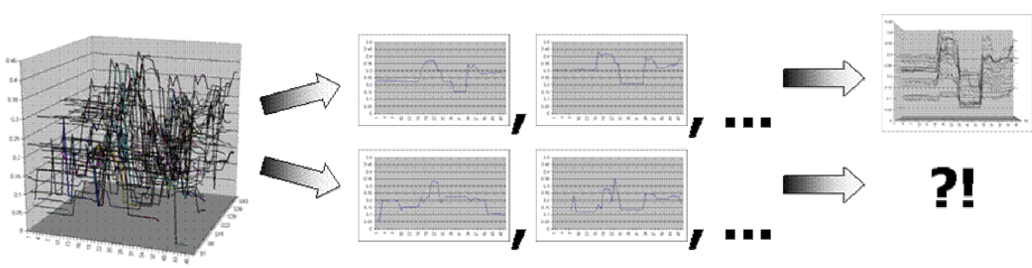

Figure 4. Workspace sharing patterns: consistency and anomalies.

Addressing Figure 4 in more detail, the left panel displays the percentage of common uforms across workspace pairs through time (uforms common to all workspaces have been removed). Within these time trajectories, we see common patterns (extracted through one of many common clustering algorithms) displayed notionally in the upper center portion of the figure, and trajectories which lay several deviations outside the common clusters displayed notionally in the lower center portion of the figure. The implication is that these two classes (consistent and anomalous) do indeed represent standard organizational and reactive event activity; this can be verified through examination of the identified uforms.

Based on the hypothesis (qualitatively demonstrated through this initial analysis) that there are (at least) these two types of knowledge sharing patterns, we infer that the CPOF data provides an opportunity to learn consistent collaboration patterns in order to optimize storage and bandwidth via anticipatory retrieval. Additionally, we infer the need to focus on the inference and recognition of new sharing patterns in order to anticipate team formation.

Discrete Sequence Analyses: (Variable Order) Markov Models of Uform Traffic through Workspaces

Warfighter Requirement	Derived Analytical Hypothesis	Data Analysis Objective
Reduced latency in data access and presentation	(Variable order) Markov models can predict knowledge flow	Learn and generalize knowledge propagation and build knowledge flow predictors using compression techniques

Workflow estimation can be based on the inference (estimation) of generalized event sequences based on (automated) measurements of observed activity. The analysis of the data streams formed by the ordered event measurements was formally couched as a *discrete sequence prediction* problem; more generally speaking this is *sequence learning*. Sequence learning can be decomposed into four canonical problems:[1]

- *sequence prediction* is concerned with predicting future sequence elements based on previous elements;
- *sequence generation* is concerned with generating elements in a sequence (based on a class exemplar);
- *sequence recognition* is concerned with the classification of (portions of) sequences into classes (based on exemplars and a distance metric) ;
- *sequence decision-making* consists of selecting a sequence of actions to fulfill some goal (such as minimizing some cost).

(The long-term vision of) adaptive command involves multiple instantiations of each canonical problem. For instance, accelerating collaboration consists of *recognizing* multi-user / multi-element action sequences in progress; recognition of the front handle of a sequence leads to estimations of the *predicted* future actions. Future actions must be *generated* or more precisely, specified, based on actual slot data in the front handle and appropriate exemplar facets. Finally, optimal *decisions* on internal system responses supporting and/or accelerating the near future collaboration activity must be made and facilitated.

The classic application of discrete sequence recognition / prediction is lossless compression; it is from this corpus of algorithms that we drew our algorithms for the discrete sequence analyses. Each is a form of a variable order Markov model (VMM); the algorithms investigated included:

---

[1] A. Ekárt: Genetic Algorithms for Discrete Sequence Prediction. J. Bullinaria (Editor). Proceedings of the UK Workshop on Computational Intelligence UKCI'2002, Birmingham, 2-4 September 2002, pp. 31-36.

- Lempel-Ziv 78[2]
- Prediction by Partial Match[3]
- Context Tree Weighting[4]
- Context Tree Weighting for Multi-Alphabets (DE-CTW)[5]
- Probabilistic Suffix Trees[6], and
- LZ-MS (Nisenson improvements to Lempel-Ziv)[7]

These algorithms were applied to the training data in the form of containment of a specific UUID in a given workspace through time and evaluated via average log loss

$$\ell\left(\hat{P}, x_1^T\right) = -\frac{1}{T}\sum_{i=1}^{T}\log_2 \hat{P}\left(x_i \middle| x_1 \ldots x_{i-1}\right)$$

for optimality against the problem at hand. While the details of each method vary, each in its own way implies a probability tree specifying the conditional probabilities of each succeeding symbol given a sequence of preceding symbols.

Technical challenges experienced during model construction and application included:

- computational complexity (of both training and run-time)[8]
- the (unknown) utility / completeness of data for training, discriminability, etc.;
- the selection of metrics (micro, macro, local, global, partial, complete, …) in both the estimation and application contexts[9]

Figure 5 notionally displays the analysis process. The left portion of the diagram indicates the tabular input into the process with each group of rows detailing the path of a single UUID through workspaces together with its arrival time. Application of one of the several VMM algorithms constructs a tree of probabilities that translates to estimations of time-dependent probabilities of a particular UUID (parameterized by its type and perhaps other qualities) arriving in another workspace (the right portion of the diagram). In simpler terms, given a uform, we can predict in which workspace it will show up next.

---

[2] Ziv, J. and Lempel, A., "Compression of individual sequences via variable-rate coding", IEEE Transactions on Information Theory, 24:530-536, 1978.

[3] Cleary, J.G. and Witten, I.H., "Data compression using adaptive coding and partial string matching", IEEE Transactions on Information Theory, 24(4):413-421, July 1978.

[4] Willems et al., The context-tree weighting method: Basic properties, IEEE Transactions on Information Theory, pgs. 653-664, 1995.

[5] Volf, P.A., *Weighting Techniques in Data compression Theory and Algorithms*, PhD thesis, Technische Universiteit Eindhoven, 2002.

[6] Ron et al., "The power of amnesia: Learning probabilistic automata with variable memory length", Machine Learning, 25(2-3):117-149, 1996.

[7] Nisenson et al, "Towards behaviometric security systems: Learning to identify a typist", 7th European Conference on Principles and Practice of Knowledge Discovery in Databases, 2003.

[8] J. R. Quinlan: Boosting First-Order Learning. Algorithmic Learning Theory, 7th International Workshop, ALT '96, Sydney, Australia, October 1996, Proceedings.

[9] B.D. Davison and H. Hirsh. Predicting Sequences of User Actions. In Proceedings of the AAAIC 1998 Workshop on Predicting the Future: AI Approaches to Time-Series Analysis 1, AAAI Press, pp. 5-12, 1998.

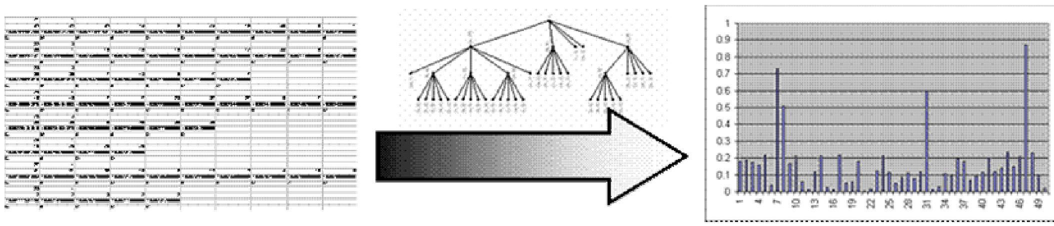

Figure 5. UUID trajectory estimation through discrete sequence prediction.

Given the ability to learn and generalize knowledge propagation enables the inference of team formation from "next step" indications as well as the ability to optimize anticipatory retrieval from these generalizations. Provided that we can build predictors for knowledge flow (as evidenced from this initial study), it follows that we can learn collaborative workflow and informal organization and then provide cognitive pre-fetch i.e. smart anticipatory retrieval.

Type Density Analyses: Life Cycle Model of Uform Types over All Workspaces

Warfighter Requirement	Derived Analytical Hypothesis	Data Analysis Objective
Effective, efficient situation awareness	There exist general life cycle models for data (based on type, content, …)	Identify type/time clusters and characterize movement of clusters through type density moments

The previous analyses purposely had limited dependencies on and interactions with any type of uform content (i.e. relied primarily on cardinal and ordinal statistics). This analysis takes the next step towards uform content understanding by considering the properties of *classes* of uforms, in particular, their life cycle characteristics. As can be seen in Figure 6, three general life cycle types are observed, each having distinguishing temporal characteristics.

Certain types of uforms are created, incorporated into workspaces, and endure for the life of the system deployment. Intuitively, these include organizational information, terrain, strategic intelligence, etc. The next most long-lived uforms have life spans of hours or days and intuitively correspond to state information on current thrusts, events, and/or operations. Finally, there are the most ephemeral uforms that intuitively represent cognitive processing and the message traffic between team members that enables collaborative cognitive processing. These types are notionally portrayed in the three right segments of Figure 6.

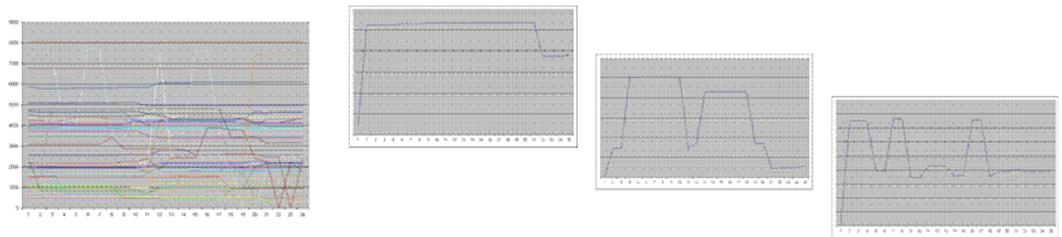

Figure 6. Characterization of three types of uform life cycles.

The CPOF data repositories contain clusters in the joint uform type and life span space. Given that we can infer type/time clusters (as demonstrated by this initial analysis), it flows logically that we can dynamically adapt anticipatory retrieval (space and time). Furthermore, given that we can characterize movement of clusters through workspaces, then it follows that we can learn knowledge requirements, cue team formation, and provide local, context-aware proactive search / retrieval.

Social Network Analyses: Societal Activity and Relations as a Model for C2

Warfighter Requirement	Derived Analytical Hypothesis	Data Analysis Objective
Protect blue information and identify red information vulnerabilities	Social network models and metrics can cue information operations	Identify gatekeepers, bottlenecks, gurus, etc. and characterize time-variant informal organization through SNA

In the most briefly addressed of the MACE analyses, social network modeling techniques were applied to the CPOF repository data. The foundational data was identical to that used in *Discrete Sequence Analyses: (Variable Order) Markov Models of Uform Traffic through Workspaces*, however, rather than focusing on content-blind network flow techniques, the content parameters source, sink, and content summary (subject) information were leveraged.

We are currently teamed with Carnegie Mellon University in the development and deployment of several sparse graph theory approaches to identifying min cut / max flow properties of (social) networks. These map directly to the identification of gate keepers, bottlenecks, gurus, etc. in the underlying social network (command and control organizational) model represented. Direct implications to the adaptive command environment include the ability to aid team formation through problem / solver matching and tune proactive search to increase situational awareness. The ability to characterize (time-variant) informal organization, i.e. the actual real-time functional organization of the force, provides the capability to learn coalition formation and dissolution (cue team formation), focus local anticipatory retrieval on rate limiters, and learn future optimal, context-sensitive formal organization doctrine.

Additional research is strongly indicated in this technical area and we are actively seeking teammates with which to execute.

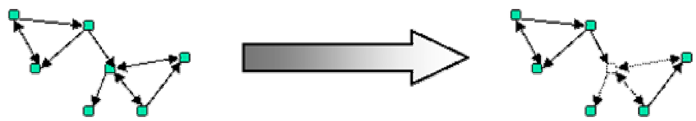

Figure 7. Dynamic, real time social network analyses provides cues for both technical and operational optimization.

RESULT SUMMARY

Generalizing the results, we see that:

- Global organization drives activity, but local standard operating procedure drives the details.
- Foundational activity patterns exist, but outliers represent tactically significant activity.
- Dynamic team formation is the key to emergent problem solving.
- Organization implies workflow, but circumstances drive ad-hoc organization.
- Temporal interest patterns can be generalized.
- Information exchange is key to command and control acceleration and exploitation.

While none of these results is unexpected, it is extremely significant that each can be inferred, derived, and verified from the CPOF data. The power is not in the generalizations, but that specific instances of these conditions can be detected not only retrospectively from archived repository data, but in real time during mission execution. Real time recognition on live CPOF data provides formerly unavailable indications on which true adaptive command can be based.

FUTURE RESEARCH

Based on the analyses performed to date, the following additional analyses are suggested:

Warfighter Requirement	Derived Analytical Hypothesis	Data Analysis Objective
Leverage Activity Surges	Detect and identify C2 decision boundary conditions	Discover tripwires for system adaptation and knowledge push.
Capitalize on Data Overlap	Identify common knowledge, forms and use	Uncover user profiles, default views, templates, etc. to move focus from tool to decisions and provide database design cues.
Take Cues from Anomalies	Detect and identify C2 environmental (sensed data) boundary conditions	Discover approaches for recognizing inferred areas of interest, activities of interest, patterns of interest, etc. and for enabling knowledge-based compression.
Identify Critical Resources	Identify “critical path” and “critical path people” in the operational flow	Discover cues for efficient, effective team formation and elicit requirements for adaptive networking.

Integrate Analysis Tools	Determine the statistical relevance (MOPs & MOEs) of potential analytic tool integration using the CPOF data, e.g., can the tool maximally extract value from the data?	Facilitate fully integrated TOC workstations and assess the value of current and potential future data bridges.
Improve Knowledge Interfaces	Detect and identify knowledge search “miscues” and “near hits” by the user	Discover opportunities for workload reduction, decision timeline reduction, dialogue pattern recognition, and query language improvement.
Tune Knowledge Management	Determine the efficiency of the emerging user-defined ontology; identify optimization/reduction of the semantics	Validate semantics to expedite situation awareness and enable efficient representation and distribution.
Exploit Outside Knowledge Sources	Determine the increase in team formation efficacy using additional, external databases	Identify methods for improving team formation and prioritizing additional data bridges.
Model the Architectures	Detect and identify causal and behavioral dependencies	Discover informal organizational cues, emerging doctrine and (system, information, network, etc.) architectural cues.

With the proposed analyses above as the technical context, we turn now to the operational context: the BattleZone concept. Some potential areas for future research are found in leveraging the chaos of the collected cognitive state found in CPOF. Others deal with leveraging the collected data to allow for simulation and “What If” analysis.

Cross-domain Fusion: Relating information from disparate sources to create a single coherent picture. Real-time capture of battle space information occurs in a highly unconstrained system context, requiring cross-domain fusion. Digital data comes from disparate sources: sensor systems, human communications, environmental descriptions, order-of-battle, and cultural and political estimates. While adding complexity to an already difficult fusion requirement, the scope provides a rich set of information and increases the number of available clues that might lead to success.

Cross-domain Reasoning: Enabling reasoning systems to apply domain-specific expertise in novel contexts. Fully exploiting a rich set of information sources will require development of reasoning technologies in each of the separable information domains. These new capabilities must also be applicable to cross-domain problems. Many first-principle reasoning systems are available for building more comprehensive, cross-domain analysis systems.

Concept Bridging: Discerning key information that builds a bridge, linking two partial hypotheses. The advantage of real-time data capture and analysis depends on automated processes that reason about the integrated data and uncover missing information, also in real-time. There is a body of automated planning work that is applicable to this task if improved to discern evidence useful in linking partial hypotheses.

Dynamic, Automatic Indexing: Automating the indexing problem to enable application of analogical reasoning. The indexing problem for case-based reasoning (CBR) represents the difficulty in distilling signature characteristics of the stored data (the case) to be used as indices into that case during later retrieval for matching to similar circumstances. A potential approach is automated real-time indexing techniques based on the ability to request missing data (also in real-time) so that the indexing algorithms will be provided complete descriptions i.e. A Combat Google™.

Case-based Reasoning: Exploitation of electronically captured battle space information. Complete realization of the potential value requires advance on several fronts and those we exemplify here all depend on advancing the state of the possible in applying case-based reasoning (CBR) strategies. Novel applications of the technology could address improving autonomous agent behaviors in entity level simulations, refinement of plans during course of action analyses, and even serve as metrics during evaluation of different C2 strategies as but a few examples.

Dynamically Composable Simulation Federations. Just as the collaborative environments now connect humans representing various domains of expertise, there is a requirement to create simulation federations that permit many diverse, automated "experts" to join in collaborative analyses. Current technology allows rudimentary federation of simulation systems; complex agreements for cooperation must be established well before system interaction. Even given these interchange agreements, compatibility issues remain. As an example, different simulation systems usually include their own unique representations of the environment, leading to multiple answers to identical questions.

THE CHALLENGE

Future command technologies must be able to learn and adapt to the user, his collaborative partners and the operational environment. The work documented herein has demonstrated the feasibility of such technologies using *retrospection*; the emerging technical challenge is adaptive command using *precognition*. Regardless of the technical level of sophistication targeted or achieved, the primary goal must always be to enable the human to adapt more quickly than his enemy and practice his art in a way that leaves him never surprised, not replace him or make him a mere weapons system. The challenge is to develop technologies that will transform the way we think about command and control and produce technologies that amplify intuitive knowledge. This challenge can be met only by a multi-disciplinary team of technologists, cognitive psychologists and military operators that is directed by the principle of Recognition Primed Decision Making[10,11] and enabled by the Capture of Live Combat.

---

[10] Gary Klein, *Sources of Power: How People Make Decisions*, MIT Press, Cambridge, MA, 1998.
[11] Karol G. Ross, Gary A. Klein, Peter Thunholm, John F. Schmitt, and Holly C. Baxter, "The Recognition-Primed Decision Model", MILITARY REVIEW, July-August, 2004.