Leveraging Structural Characteristics of Interdependent Networks to Model Non-linear Cascading Risks

Anita Raja*, Mohammad Rashedul Hasan*, Shalini Rajanna, Ansaf Salleb-Aoussi+
Department of Software and Information Systems*
University of North Carolina at Charlotte
Charlotte, NC 28223
Center for Computational Learning Systems+
Columbia University
NY, NY 100126
Email: {anraja, mhasan8, srajanna}@uncc.edu, ansaf@ccls.columbia.edu

Abstract

This paper describes our continuing efforts to forge new ground in identifying the effects of interdependency on acquisition and, if needed, uncovering early indicators of interdependency risk so that appropriate governance oversight methods can then be isolated. Specifically, we seek to study the topologies of Major Defense Acquisition Programs (MDAPs) networks and associated cascading consequences of interdependencies in such highly dependent networks. Since the start of this new project phase a couple of months ago, we have begun harnessing the extensive data that has been collected over the years in the form of Defense Acquisition Execution Summary (DAES) documents for the MDAPs. We present a road map of our research plan and our preliminary results in our ongoing efforts on leveraging network structure and automatic data extraction to study cascading risks. We will also identify the challenges to data acquisition.

1. Introduction

This research seeks to study the structures of the Major Defense Acquisition Programs (MDAPs) networks and the associated cascading consequences of interdependencies in such highly dependent networks. It involves identifying the effects of interdependency on the acquisition process and, if needed, uncovering early indicators of interdependency risk so appropriate governance oversight methods can then be isolated. Hence, this research seeks to address the problem that there is little insight on the effects of interdependencies and a lack of tested metrics to provide early indication of the acquisition risks of interdependent programs. It breaks ground in the area of (i) studying non-linear cascading effects in the context of a network of MDAPs consisting of some not-so-successful programs (that which experiences cost growth) as compared to (ii) the study of the decision mechanisms of successful programs. Lessons learned from this comparative analysis would help model the behavior of other MDAP programs. The project will use the extensive data that we have collected over the years in the form of Defense Acquisition Execution Summary (DAES) documents for the MDAPs

This work builds on our previous results (Raja et. al., 2012) obtained from a manual analysis of data belonging to a small network of MDAPs representing a case study. Our goal was to model “what-if” analyses that would help decision-makers to gain insight on the cascading effects of perturbations among interdependent networks and take appropriate measures to handle them. We used the case study to first determine whether the data required to build a decision theoretic model is available and then study whether this decision-theoretic model captures the cascading interdependencies that are of interest to us. We also examined the data investigation process to identify the challenges that were encountered. Our results showed that MDAP-related data characteristics support the multiple perspective study of perturbations and it is possible to recast the study of cascading effects as a sequential decision problem. We identified local and non-local issues that when left unmitigated led to performance breaches in the MDAPs. We also observed that it is crucial to consider the uncertainty in action outcomes in the decision-making process and that a non-local perspective may help explain a performance breach in situations where a solely local perspective does not. These observations supported our conjecture that a decision-theoretic model is a good methodology to study interdependencies in the MDAP network and to capture early indicators of interdependency risk. Finally, we captured the informational value in the existing data and the challenges inherent in the data collection process with respect to their role in isolating risks and initiating appropriate government oversight methods.

The sheer volume and complexity of the data required to populate our decision theoretic models effectively has led us to identify methods for **automating** the data extraction, network analysis and construction of the decision model that is the focus of our current work. This project, initiated a couple of months ago, has the following research goals: 1) Examine and compare the network structure characteristics of interdependent regions belonging to successful and not so successful MDAP programs to augment our current work in "what-if" analyses. 2) Automate the data extraction and analysis process by leveraging algorithms for decision support as well as image and text analysis. 3) Continue to identify the challenges in acquiring the data from the government and program managers. In this paper, we will discuss our proposed ideas for this year long project and the initial work we have done to achieve the above mentioned research goals.

2. Background:

It has been shown that data are the foundation for decision-making in the acquisition environment. The Department of Defense (DoD) has spent a significant amount of effort working across the organization to identify useful sources of data and to conduct analyses. The importance to acquisition research of studying MDAP interdependencies was emphasized during the 2012 Annual Acquisition Research Symposium by the introduction of a new panel titled Predicting Performance and Interdependencies in Complex Systems Development. Prior research has established that MDAPs are demonstrably interdependent and that they can be thought of as networks of interdependent programs (Lewin, 1999; Flowe, Brown, & Hardin, 2009; Flowe, Kasunic & Brown, 2010). Also, the acquisition paradigm established in statute (10 U.S.C. 2434; Defense Acquisition Workforce Act, 1990), in policy (DoD 5000.02; Under Secretary of Defense for Acquisition, Technology, and Logistics [USD(AT&L)], 2008), and in regulation tends to favor the notion of MDAPs as being independent, which would cause exogenous factors caused by interdependence to be overlooked or misinterpreted.

Although it is critically important to understand the program interfaces and interdependencies, there are few tested and proven tools for program managers and acquisition executives to probe the joint space or to track the cascading effects that the joint space might trigger. There is reason to believe that the exogenous issues generated from the shared domains remain unnoticed to the extent of causing the program to potentially experience severe performance degradation (Brown, 2011). The complexity of the joint environment is likely to have a direct bearing on acquisition activities. The precise effect on acquisition, and its resulting managerial implications, are, as of yet, unknown. We believe that given the frequency with which government agencies are moving toward joint initiatives, the findings of this research project based on DoD programs may prove instrumental to a wide-ranging audience.

Furthermore, at the 2012 Acquisition Symposium, Dr. Frank Kendall III, the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]), discussed the DoD's strategic priorities, especially around acquisition. These priorities included achieving affordable programs that execute well and improving efficiency (via Better Buying Power and other initiatives). We believe the work described in this paper will help us understand the performance of the programs in various scenarios and contribute directly to the above priorities by achieving affordable programs that are successful as well as improving overall efficiency.

Along with other researchers (Brown & Owen, 2012), we have begun to harness a network-centric approach to study DoD acquisition and focus on an MDAP network of interrelated programs that exchange and share resources for the purpose of establishing joint capabilities. Some work (Zhao, Gallup, & MacKinnon, 2012) has been done to analyze the unstructured and unformatted acquisition program data using a data-driven automation system called Lexical Link Analysis (LLA). LLA is used to determine the correlation between system interdependency and development costs in an effort to enable acquisition researchers and decision-makers to recognize important connections that form patterns derived from dynamic data collection. In other work (Han, Fang, & DeLaurentis, 2012), a Bayesian Network (BN) method is used to assess the cascading effects of requirement and systems interdependencies on risk in an effort to effectively analyze alternatives in a capability-based acquisition strategy. The technique is evaluated within a synthetic network and identifies critical systems and requirements.

We believe our work will help us understand the performance programs in various scenarios and contribute directly to the above priorities by achieving affordable programs that are successful as well as improving overall efficiency.

3. Research Methodology:

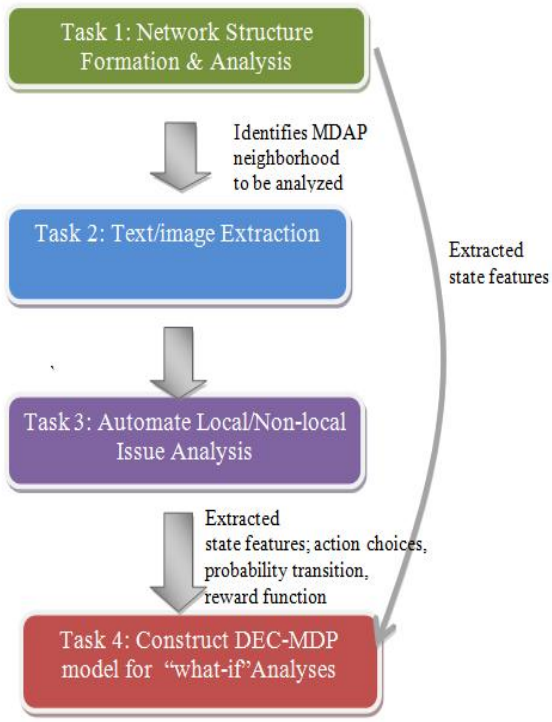


Figure 1: Research Goals

The overall goal of this research is to continue our efforts to forge new ground on identifying the effects of interdependency on acquisition and, if needed, uncovering early indicators of interdependency risk so appropriate governance oversight methods can then be isolated. Hence, this research seeks to address the problem that there is little insight on the effects of interdependencies and a lack of tested metrics to provide early indication of the acquisition risks of interdependent programs. It breaks ground in the area of (i) studying non-linear cascading effects in the context of a network of MDAPs consisting of some not-so-successful programs (that which experiences cost growth) as compared to (ii) the study of the decision mechanisms of successful programs.

The information pertaining to acquisition research is overwhelming and multifarious. It appears to be a daunting task for the acquisition researchers, let alone the program managers, to integrate and understand the vast and dynamic data in a coherent way. To define the interrelationship among the MDAPs from a network centric viewpoint and to identify different network dependencies within the domain of MDAPs, following set of data resources are useful:

- Monthly DAES reports that provide an early-warning report on the status of some program features such as cost, schedule, performance, funding etc.
- SARs that summarize the latest estimates of cost, schedule, and technical status to be reported annually in conjunction with the President's budget
- Program Element (PE) documents (called PE docs or R-docs) that are used to justify congressional budgeting process.
- Program Objective Memoranda (POMs) which are submitted by the Components (Military Departments and DoD Agencies) to OSD Comptroller

Below, we describe the main tenets of the four research tasks illustrated in Figure 1. Since we are in the very early stages of this project, we will describe our proposed research for each of the tasks and also discuss initial progress we have made so far.

3.1 Task 1: Network structure formation and analysis

We plan to address the following two questions as part of this task:

- What are the essential features of the network that reveal the joint space dynamics?
- What are the relative priorities associated with these features and how do they affect the network relationship?

Network Structure Formation: From our previous study, we identified both successful and not-so-successful programs with respect to performance breaches. For the current study, we plan to build funding networks for these two types of MDAPs. We will study the Program Element (PE) accounts of these programs from their "Track to Budget" files and would find their first-order funding neighbors. This process would enable us to define the network topology for the analysis of its properties.

Network Structure Analysis: Network theory (Ahuja et al., 1993; Albert et al., 2000) provides useful tools to calculate and understand quantities or measures that capture significant features of the network topology. These measures help analyze the network data based on the structure of the network and also help to understand how those properties are related to the practical issues that we care about. In other words, network theory provides a rich set of measures and metrics that can help understand what the network data may tell. A key metric for network data analysis is various types of centrality measures. Centrality quantifies how important are the nodes (or edges) in a networked system. There are a wide variety of mathematical measures of node centrality (Bonacich, 1987; Borgatti, 2005; Freeman et al., 1991) that focus on different concepts and definitions of what it means to be central in a network. A simple but very useful example is the measure called degree. The degree of a node in a network is the number of edges attached to it.

In case of an MDAP funding network, degree-centrality would show how many funding neighbors a particular MDAP has and how it could be related to the performance of the program. For example, having many funding partners incurs more risk in terms of being affected by the cascading consequences. Many of the standard algorithms for the study of networks are already available, ready-made, in the form of professional network analysis software packages. Some of the software packages for analysis of network data are Paejk (http://vlado.fmf.uni-lj.si/pub/networks/Pajek/), Netminer (http://www.netminer.com/index.php), yEd (http://www.yworks.com/en/products_yed_about.html); JUNG (http://jung.sourceforge.net/) etc.

State of the program in our decision-theoretic DEC-MDP model captures the critical information at a specific point in time that will support the decision-making to guarantee good performance. To describe the state space and to identify some of the key state features we will employ an appropriate network analysis tool for the MDAP networks. We plan to address the following question: what are the network properties that essentially contribute towards the good/poor performance of the respective MDAPs? Our goal is to measure some of the important centrality measures for the network and correlate it with the performance of the node (the program). Centrality measures help us to determine (i) which nodes are important in the network and (ii) to assess their importance with respect to their performance.

We plan to first define an undirected funding network for a chosen MDAP. We will then measure the following network centralities for 5/10 years time-span for all MDAPs: degree, betweenness, closeness, similarity, local clustering coefficient etc. We discuss these metrics in greater detail in the following paragraphs. We also plan to calculate the performance factor for 5/10 years time-span for all MDAPs, based on a composite metric (it may include the breach factors, %PAUC, funding delta etc. from SAR files). This will help us to determine how each of the centrality measures affects the performance of the programs over time.

The above methodology will enable us to identify additional state features to describe the state space of the program within the DEC-MDP model. The following is the list of features of interest. **Feature 1**: Program ID; **Feature 2:** Current Year **Feature 3:** Current Month; **Feature 4:** Cost (APB) Status: for 9 months, starting from the current month; **Feature 5:** Cost (Contract) Status: for 9 months, starting from the current month; **Feature 6:** Schedule (APB) Status: for 9 months, starting from the current month; **Feature 7:** Schedule (Contract) Status: for 9 months, starting from the current month; Feature **8:** Performance (APB) Status: for 9 months, starting from the current month; **Feature 9:** Performance (Contract) Status: for 9 months, starting from the current month; **Feature 10:** Funding (APB) Status: for 9 months, starting from the current month; **Feature 11:** Funding (Contract) Status: for 9 months, starting from the current month; **Feature 12:** Degree centrality; **Feature 13:** Closeness Centrality; **Feature 14:** Betweenness Centrality; **Feature 15:** Local Clustering coefficient; **Feature 16:** Commodity Type ; **Feature 17:** Partner Abandonment.

We have identified features **Feature 1** through **Feature 11** to be useful features based on our past work. As part of this project, we propose to continue studying these features and introduce more network-centric features in the context of studying the role of interdependencies on performance. Features 12-17 capture some of the key network centric features for the MDAP of interest. For example, **Feature 12** (degree centrality) measures the connectivity of a program with other programs.

A higher connectivity might incur higher risk because of its sharing of funding with many partners. Feature 13 (closeness centrality) measures the mean distance of a program from other programs. These centrality measures could offer better understanding about the propagation speed of the cascading effects. Feature 14 (betweenness centrality) measures the importance of the program that may reside in the overlapping region of more than one sub-networks and which is able to control the flow of influence among different sub-networks. Feature 15 (local clustering coefficient) measures the formation of groups among the member nodes and it may be related to the degree distribution of the network. For example, typically nodes with higher degree have lower local clustering coefficient on average. Therefore, a node with higher local clustering coefficient (and lower degree distribution) is most likely prone to lower risk. It is also useful to identify the "structural holes" in a network. If two neighbors of a node are not themselves neighbor than we say that there is a "structural hole" existing among them. Identification of "structural holes" could be useful to analyze the propagation of cascading risks. As part of this task we will study the usefulness of these features and also identify other new ones.

Initial Results on Task 1: Network Structure Formation

We define a funding network of an MDAP using the PEs that funded the MDAP's RDT&E efforts. PE is the code number assigned by the comptroller. Since PEs fund multiple MDAPs, programs that share a common PE monitor could be isolated. Procurement PEs were not considered for defining funding networks since the RDT&E interdependencies were the most critical to program performance. The funding network and the associated R-docs allowed us to do a detailed study of the performance of the member nodes and to understand the cascading effects the funding network of the three MDAPs named MDAP_A, MDAP_B and MDAP_C. The original names of these MDAPs have been removed to retain the confidentiality of the programs.

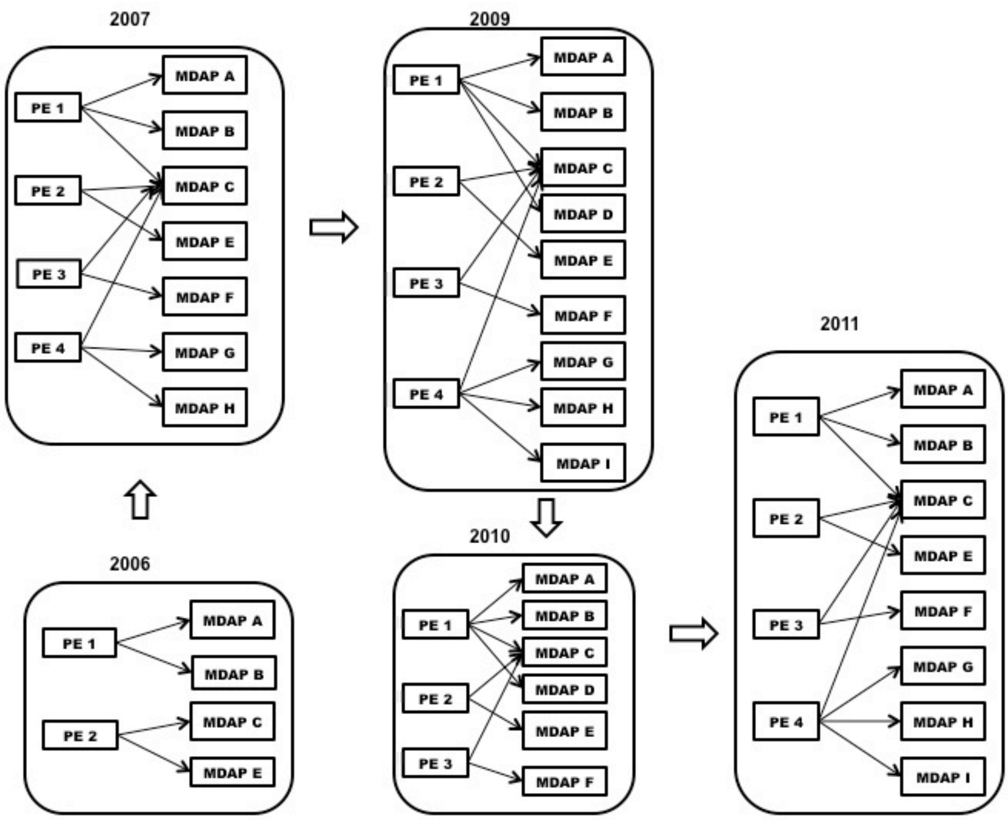


Figure 2: Evolving funding network of MDAP_A, MDAP_B and MDAP_C

Examination of the DAES reports and R-docs from years 2006-2011 related to these MDAPs shows that, MDAP_A and MDAP_B experience frequent performance breaches while MDAP_C appears to be performing as expected. We have built an evolving funding network of these three MDAPs based on the common PE accounts that they share with other MDAPs, such as MDAP_D-I. The relationship between the PE accounts and the MDAPs, extracted from the PE docs, is represented as bipartite networks. Figure 2 shows how the funding relationship of these three MDAPs and their neighbors change from 2006 to 2011. Since the PE docs for the year 2008 were unavailable, we couldn't show the funding network for that year.

From these bipartitie networks, we notice that MDAP_A and MDAP_B share only one PE account (PE 1), while MDAP_C shares multiple PE accounts (PE 1–4). It indicates that MDAP_C is prone to more inter-dependency risks.

Next we plan to measure the weight of the links between the PE account and the respective MDAPs based on the funding distribution as captured in the PE docs. This measurement can be obtained by comparing the POM and SARS data. The former describes what the PM says the program requires and the latter is what the program actually got. This comparison will give us a better understanding of the dependency of MDAPs on the associated PEs and the effect of expected and actual budget allocations on performance breaches. We will use these link weights as state features for the respective programs.

3.2 Task 2: Automated data extraction and text analysis

We plan to address the following two questions as part of this task:

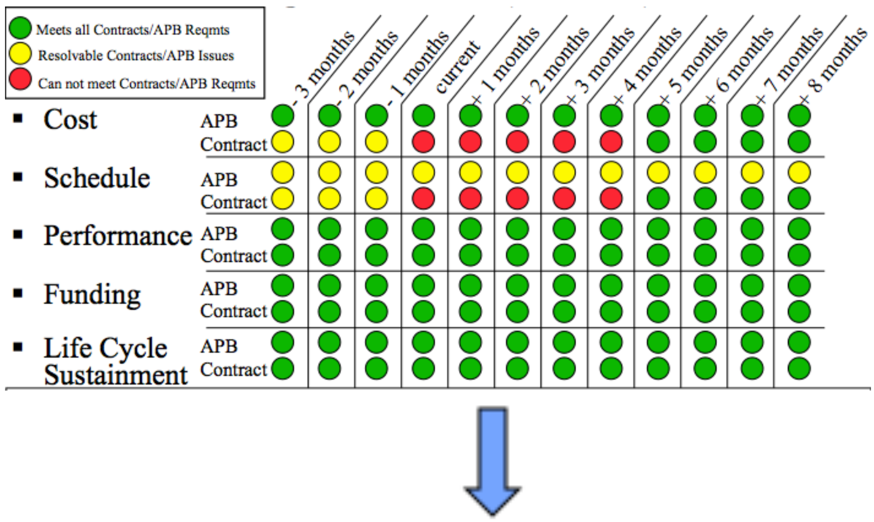
- What are the local issues that lead toward breach or near-breach situation?
- How often and why do the local mitigation efforts fail to improve the performance?
- How do we identify the non-local issues that result from the interdependencies?
- Determine the cascading effect through the network?

We plan to approach Task 2 from two perspectives: *Local perspective* where the analyses are based solely on the individual program's own data; and *Non-Local perspective* where the analyses are based on the data of MDAPs existing in the joint space of the individual program. Lessons learned from these analyses should enable the stakeholders to take appropriate measures to improve the performance of the programs. Our objective in this task is to narrow down the wealth of data present in the DAES reports in order to capture useful knowledge about the status of individual MDAPs and the MDAPs in their network. This will be achieved as follows:

3.2.1 Automatic data extraction:

The aim of this subtask is to bring the content of DAES reports, currently as Microsoft Power Point files, Adobe Acrobat PDF files and Word document, into a form suitable to further analysis. We will mainly focus on the *program status* and *issue summary*. First, bottom-up (pixel to block) image segmentation will be used in order to extract the layout of the document (O'Gorman, 1993; O'Gorman and Kasturi, 1997; Salleb and Hocini 1996). It appears from the DAES reports that the part that requires further extraction is the *program status matrix* for the following items: Cost, Schedule, Performance and Funding. The status of each of these items is given for *APB* and *Contract*. The status is a colored circle indicating three possible states: Meet all contracts (green); resolvable contracts (yellow) and cannot meet all contracts (red). The status is given for current month, past 3 months along with a forecast for the upcoming 9 months.

Once we extract the different components in the document though image segmentation using bounding boxes, nearest neighbors, linear regression (O'Gorman, 1993; Salleb and Hocini, 1996), we will translate the program status matrix into an integer-valued matrix, where green will be represented by 1, yellow by 0 and red by -1. An example of such a representation is presented in Figure 3. We will also parse DAES files to extract all the words used in the *program status* and *issue summary*. We will use Java text extraction libraries that have proven to be powerful. Hence, a report will be defined by the following components: MDAP name, Month, Year, status matrix and the extracted text from the program status description and from the issue summary.



Item	M_3	M_2	M_1	Current	P_1	P_2	P_3	P_4	P_5	…	P_8
Cost_APB	1	1	1	1	1	1	1	1	1	1	1
Cost_Contract	0	0	0	-1	-1	-1	-1	-1	1	1	1
Schedule_APB	0	0	0	0	0	0	0	0	0	0	0
Schedule_Contract	0	0	0	1	1	1	1	1	-1	-1	-1
Performance_APB	1	1	1	1	1	1	1	1	1	1	1
Performance_Contract	1	1	1	1	1	1	1	1	1	1	1
Funding_APB	1	1	1	1	1	1	1	1	1	1	1
Funding_Contract	1	1	1	1	1	1	1	1	1	1	1

Figure 3: Translation of data to integer-valued matrix

3.2.2 Advanced text analysis: Text from the program status and issue summary will be used to assess the health or status of the MDAP. We will extensively use word clouds in order to visualize the status of the MDAP as described in the corresponding text (See example of word cloud in Figure 4). While word clouds will provide a nice visualization tool that provides a general idea on the contents of the documents.

Besides visualization, we will also extract n-grams (sequence of words) and concepts or topics from the text in order to identify the issues that an MDAP is going through. For this purpose, we will use *Topic Modeling* to discover the topics discussed in the corpus. The intuition behind topic modeling is that when program managers prepare to write their monthly reports, they first have in mind a set of topics to address. They fill in the DAES using words associated with the different topics. Topic modeling identifies which words have the greatest probability of occurring together, and posits an abstract topic that conditions these probabilities. After generating the topic models for the MDAP documents, each document can be represented as a subset of the total topics, each in a proportion dependent on the content words. For instance, a report can belong to the topic “delay in schedule” and also to a topic “gap in funding”.

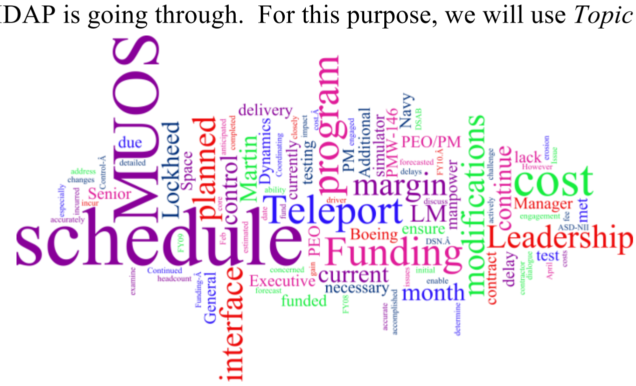


Figure 4: Example Issue Summary Word Cloud

To preprocess the documents, we will strip all the non-content words, and only keep the free text. Words and characters that are removed include section and field names, person names, punctuation, digits and stop-words. A topic model consists of a probability distribution over topics, and then for each topic, the probability of each word in the vocabulary. The parameters behind the probability distributions are treated as latent variables. By analyzing a set of observations (words in the documents), it is possible to recover the latent structure of the generative model. The particular model we use is based on Latent Dirichlet Allocation (LDA) (Blei et al., 2003) with Gibbs Sampling. For the experiment, we will use the Stanford Topic Modeling tool kit (http://nlp.stanford.edu/software/tmt/tmt-0.4/), a machine learning toolkit for natural language processing tasks.

Initial Results on Task 2:Automated Data Extraction

As discussed above, DAES reports include information of program performance in the form of text and image. Our current focus is to understand the textual descriptions in the reports. The "Issue summary" section in the report illustrates the local issues if any and possible actions to resolve them. We prepare the input file for the topic modeling tool by manually copying this information as records into a csv (comma separated value) file. Specifically, we created two input files one with set of Issues (problems encountered by MDAPs as reported in the DAES) and other with set of Actions (the tangible actions proposed by the MDAP program manager to alleviate the Issues). As described above, we preprocess the reports by stripping the non-content words, and only keep the free text. Words and characters that are removed include section and field names, person names, punctuation, digits and stop-words.

We first train a classifier to automatically identify the Issues identified in the DAES reports. Using an input file for the program MDAP_A from the previous section with few (15) records of its issues from a single year, we trained a model that will classify contents into issue related topics. The results were not informative as the data was small and so we extended the input to include issues of all the reports of MDAP_A across the years. The increased data set resulted in words like 'schedule', 'Funding', 'Launch', 'ground site' and 'cost control' to be the top words in individual topic list. Examination of the tool for consistent results is important and this technically indicates the convergence of the model. Convergence is dependent on the number of iterations the model is executed, which in turn is dependent on the data size. For a data size of 100 plus records, convergence occurred at around 800 iterations. We tested our trained model on a few (30) records of the same MDAP_A program. Test results indicate the proportion of relevance of the record to each of the topics. In Figure 5 below, we describe an example record and the proportion to which the record is relevant to the 5 topics identified by the model: Schedule, Funding, Launch, Ground site, Cost Control. As shown, **this record has a high proportion of the topic "Funding".**

Record	Schedule	Funding	Launch	Ground site	Cost Control
Funding is required to test modifications of X-interface. If this is not funded a simulator will be used for the purpose. This will incur a cost impact on the MDAP_A program.	2.68E-05	0.999836	5.12E-05	5.96E-05	2.65E-05

Figure 5: Example topic distribution for an "Issue-related" record

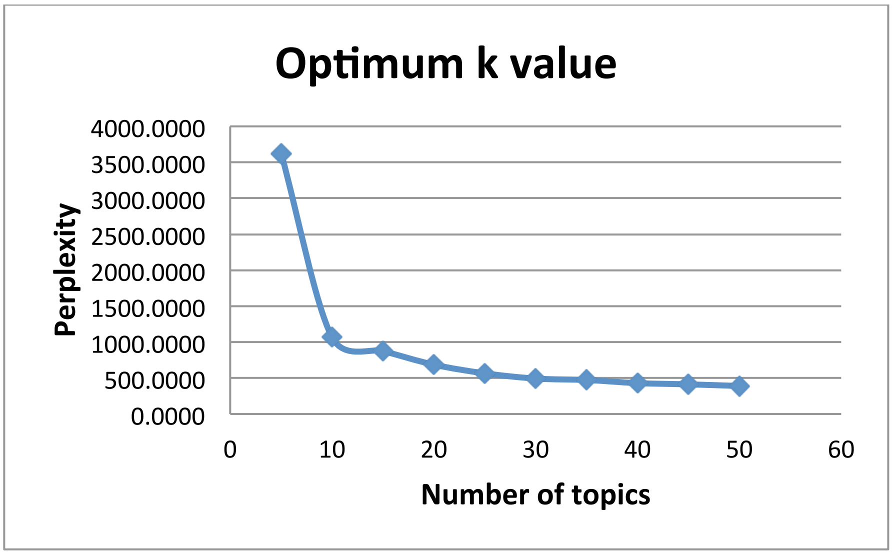


Figure 6: Determining optimal k value for "Issue" related topic

The number of topics is another important parameter for topic modeling. Initial results included 5 topics but finding optimum value for the number of topics will provide better result (Griffiths and Steyvers, 2004) in the sense that topics will be of finer granularity and hence more specific and relevant. For this we trained the model several times and recorded perplexity. Perplexity is a measure of the quality of the model learned by LDA in predicting future data from the same distribution as the data used to train the model. Lower perplexity value indicates a stable model. An experiment with the different number of topics as shown in Figure 6 signifies a k value of 20 or more is the best for our experimental data.

Our next steps in the task will involve:

- Automate the preparation of input file using PERL, a scripting language.
- Expand the input data set to include reports of all the programs across the years and train the model with this data.
- Explore parameters of the LDA model to fine-tune the results such that the top set of words in a topic list is explanatory of that topic.
- Frame a phrase by analyzing word list, for example 'Hardware issue' for better understanding and to support further analysis.
- Perform a similar topic extraction of "Action" related data.
- Scale the analysis to all MDAP programs.
- Use the extracted information to populate the Markov Decision Process in Task 4.
- Apply these topic extraction techniques to POM documents and compare it to the information in the SARS documents as discussed in Task 1.

3.3 Task 3: Local/non-local issue analysis

As part of automating the identification and analysis of local and non-local issues that lead to performance breaches, we will first evaluate the monthly mitigation forecasting for the problems from the DAES reports. We hypothesize that frequent forecasting failure along with sustaining/recurring breaches would require issue analysis. We plan to analyze the automatically extracted issues (Task 2) to reveal the presence or absence of local issues to explain the erroneous forecasting. If no significant issue can be found to explain the frequent forecasting failure, then we claim that either DAES reports do not capture the local reasons or some non-local reasons are responsible for the poor performance. We will then analyze the local issues of the neighbors in the funding network to determine if there is any non-local issue that possibly could have propagated through the network leading to performance breaches. This is work that we will pursue after we make progress on Task 2.

3.4 Task 4: Formulate a decision-theoretic model that harnesses Decentralized-Markov Decision Process (DEC-MDP) formalism:

The questions to be addressed by this task are

- What are the essential characteristics of the MDAP network that justify a DEC-MDP model?

- How to model the MDAP network as a decentralized system?
- What are the key challenges in the design of the DEC-MDP?
- What essential features should the DEC-MDP model incorporate for better predictability?

In this work, decision-making in a MDAP network is viewed as a multiagent sequential decision problem because the utility gained by each agent depends on a sequence of actions over time. Our goal is to determine the behavior of the decision-makers (agents) that best balances the risks and rewards while acting in an uncertain environment with stochastic actions.

Each agent will make its individual decisions in an environment where the state space is not fully observable, meaning, that the nodes in the network (the programs) do not exactly know in which state they are in at any particular instant because they do not have complete information about their neighbors. With the partial state information, the individual agents aim to optimize the joint reward function. This class of problems is modeled as decentralized partially observable MDP (DEC-POMDP) in literature (Bernstein et al., 2002) where at each step when an agent takes an action, a state transition occurs, and the agent receives a local observation. Following this, the environment generates a global reward that depends on the set of actions taken by all the agents. A necessary condition for stable equilibrium among agents in a multiagent system is that each agent plays a best-response to the strategy of every other agent: this is called a Nash Equilibrium. In our previous work (Cheng et al., 2012) we make the DEC-POMDP problem for a tornado tracking tractable by approximating the DEC-POMDP with a stochastic DEC-MDP model and using a factored reward function to define a Nash Equilibrium instead of the global reward function. We apply this technique to the MDAP domain. We define the reward function of this model to be composed of two different components: local reward function and global reward function. The local reward functions are dependent only on the individual agents' actions, while the global reward function depends on the action of all agents. We make this a stochastic DEC-MDP by defining a solution as a stochastic policy for each agent. A stochastic policy of an agent i is denoted by $\Pi_i(s) \in PD\ (A_i)$, where PD (A_i), is the set of probability distributions over actions A_i. Stochastic policies can cope with the uncertainty of observation and perform better than deterministic policies in a partial observable environment. We plan to apply these modeling techniques we have developed for another complex multiagent domain (tornado tracking) to the MDAP domain.

4. Conclusions & Future Work

Our multi-year research goal is to gain a deeper understanding of interdependencies among MDAPs by examining the various information sources, SARS, DAES, POMS and R-docs. This would involve establishing a statistically significant correlation between the state of MDAP network dependencies and their consequences. Our previous work in this area involved manual analysis of DAES and SARS data belonging to a small network MDAPs to determine the local and non-local issues that affect MDAP performance. As a consequence of this work, we recognized the need to analyze the data from the entire set of MDAPs in batch form to be able to build good decision models for "what-if" analysis. The volume and complexity of the data has led to our current research tasks that involve automating methods for data extraction, network analysis and decision model construction for successful and not-so-successful MDAPs. In this paper, for each task we describe our proposed work and initial results. Our hope is that as a consequence of this work, we will be able to 1) Extract the link characteristics between MDAPs; 2) Examine and compare the funding network structure characteristics of interdependent regions belonging to successful and not-so-successful MDAPs to augment our current work in "what-if" analyses; 3) Automate the data extraction and analysis process by leveraging algorithms for decision support as well as image and text analysis; 4) Continue to identify the challenges in acquiring the data from the government and program managers.

5. Acknowledgements

We are grateful to Maureen Brown and Rob Flowe for their insightful feedback on our work. We thank Maureen Brown, Graham Owen and Clinton Anderson for assisting us in the data-gathering process. We also thank undergraduate students Coreen Burke and Daniel Ball for their help with preparing the text for analysis for Task 2.

References

Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993). *Network Flows: Theory, Algorithms, and Applications.* Prentice Hall, Upper Saddle River, NJ.

Albert, R., Jeong, H., and Barabási, A.-L. (2000). Attack and error tolerance of complex networks. *Nature* 406, 378-382.

Bernstein, D. S., Givan, R., Immerman, N. and Zilberstein, S. (2002). The complexity of decentralized control of markov decision processes. *Mathematics of Operations Research,* 27(4), 819– 840.

Bertsekas, D. P. (1987). *Dynamic Programming: Deterministic and Stochastic Models.* Englewood Cliffs, NJ: Prentice-Hall. Bethke, E. (2003). *Game Development and Production*. Wordware game developer's library. Plano, Tex: Wordware Pub.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003), Latent dirichlet allocation.. *Journal of Machine Learning Research*, 3:993–1022.

Bonacich, P. F. (1987). Power and centrality: A family of measures. *Am. J. Sociol,* 92, 1170-1182.

Borgatti, S. P. (2005). Centrality and network flow. *Soc. Networks* 27, 55-71.

Brown, M. M. (2011). *Acquisition Risks in a World of Joint Capabilities*. Report prepared for Naval Postgraduate School.

Brown, M. M. & Owen, G. (2012). *Acquisition Risks in a World of Joint Capabilities: A Study of Interdependency Complexity.* Report prepared for Naval Postgraduate School.

Cheng, S., Raja, A. and Lesser, V. (2012). Multiagent meta-level control for radar coordination. To appear in *Web Intelligence and Agent Systems: An International Journal (WIAS)*, IOS Press, Netherlands.

Freeman, L. C., Borgatti, S. P., and White, D. R. (1991). Centrality in valued graphs: A measure of betweenness based on network flow. *Soc. Networks* 13, 141-154.

Flowe, R. M., Kasunic, M., & Brown, M. M. (2010). *Programmatic and Constructive Interdependence: Emerging Insights and Predictive Indicators of Development Resource Demand* (No. CMU/SEI-2010-TR-024). CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST.
Giambastiani, E. (2004). Imperatives for Transformation, *ComDef West*, March 22.
Griffiths, Thomas L.; Steyvers, Mark (April 6 2004). Finding scientific topics". *Proceedings of the National Academy of Sciences* **101** (Suppl. 1): pp 5228-5235.

Han, S. Y., Fang, Z. & DeLaurentis, D. (2012). *Acquisition Management for System-of-Systems: Requirement Evolution and Acquisition Strategy Planning*. Report prepared for Naval Postgraduate School.

Lewin, Y. (1999). Application of Complexity Theory to Organization Science. *Organization Science* 10:3(215-236).

McCallum, A. K. (2002). Mallet: A machine learning for language toolkit. Retrieved from http://mallet.cs.umass.edu

Montgomery, D. C. (2005). *Design and Analysis of Experiments* (6th Ed.). John Wiley & Sons.
Newman, M. E. J. (2011). *Networks: An Introduction*. Oxford University Press Inc., New York, USA.
O'Gorman, L. (1993). The Document Spectrum for Page Layout Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1162–1173.

O'Gorman, L. & Kasturi, R. (1997). Document Image Analysis. *IEEE Computer Society Executive Briefing*, ISBN 0-8186-7802-X Library of Congress Number 97-17283.

Raja, A., Hasan, M. R. & Brown, M. M. (2012). *Facilitating Decision Choices with Cascading Consequences in Interdependent Networks.* Report prepared for Naval Postgraduate School.

Rudin, C., Waltz, D., Anderson, R., Boulanger, A., Salleb-Aouissi, A., Chow, M., Dutta, H., Gross, P., Huang, B. & Jerome, S. (2012). Machine Learning for the New York City Power Grid. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no.2, pp.328-345.

Salleb-Aouissi, A. Vrain, C. & Nortet, C. (2007). QuantMiner: A Genetic Algorithm for Mining Quantitative Association Rules. *Proceedings of the 20th International Conference on Artificial Intelligence IJCAI* , pp. 1035-1040. Hyberadad, India.

Salleb, A. (2003). Mining Frequent Patterns for the Discovery of Association and Characterization Rules. Ph.D. thesis, Laboratoire d'Informatique Fondamentale d'Orléans Lifo, University of Orléans. France. Advisor: Professor Christel Vrain.

Salleb, A. & Hocini, H. (1996). A Page Layout Analysis System for printed documents, original title: Contribution a la conception et a la realisation d'un systeme d'analyse de documents imprimes. Engineer Degree Dissertation, Universite des Sciendes et de Technologie Houari Boumediene (USTHB) **.**

Salleb-Aouissi, A., Radeva, A., Passonneau, R., Xie, B., Khan, F., Khattak, Tomar, A., Waltz, D., McCord, M., McGurk, H. & Elhadad, N. (2011). Diving into a Large Corpus of Pediatric Notes. *ICML 2011 Learning from Unstructured Clinical Text Workshop Bellevue*, WA, USA.

Zhao, Y., Gallup, S. & MacKinnon, D. (2012). *Applications of Lexical Link Analysis Web Service for Large-Scale Automation, Validation, Discovery, Visualization, and Real-Time Program-Awareness*. Report prepared for Naval Postgraduate School.

About the Author

Anita Raja—Raja is an associate professor of software and information systems at The University of North Carolina at Charlotte, and a visiting scientist at the Center for Computational Learning Systems at Columbia University (2011–2012). She received her PhD in computer science from the University of Massachusetts at Amherst in 1998 and 2003, respectively. Raja's research focus is in the field of artificial intelligence, specifically as it relates to the study of decentralized control and reasoning in software agent systems operating in the context of uncertainty and limited computational resources. [anraja@uncc.edu]

Mohammad Rashedul Hasan --Hasan is a PhD candidate at the College of Computing and Informatics at the University of North Carolina at Charlotte. Previously, he taught in the Department of Electrical Engineering and Computer Science at North South University in Bangladesh. His research interests are in the field of Network Science, Complex Systems and Artificial Intelligence. [mhasan8@uncc.edu]

Shalini Rajanna-- Rajanna is a Master's student in the College of Computing and Informatics at University of North Carolina at Charlotte. Her research interest is in the field of text mining and intelligent systems. [srajanna@uncc.edu]

Ansaf Salleb-Aoussi-- Salled-Aouissi joined Columbia University's Center for Computational Learning Systems as an Associate Research Scientist in 2006 after a Postdoctoral Fellowship at INRIA (France). Her research interests lies in Machine Learning. She has worked on large-scale projects including the power grid. Her current research includes pattern discovery, crowdsourcing and medical informatics. Ansaf received an Engineer degree in Computer Science from the University of Science and Technology Houari Boumediene, Algeria, an M.S. and Ph.D. degrees from University of Orleans (France). [ansaf@ccls.columbia.edu]