International Technology Alliance in Distributed Analytics & Information Sciences

Initial Program Plan

Applicable Period: Sept 21, 2016-January 15, 2018

Table of Contents

Introduction

DAIS-ITA (International Technology Alliance in Distributed Analytics and Information Sciences) is a collaborative partnership between the U.S. Army and the UK Ministry of Defence which brings together researchers from U.S. Army Research Laboratories and UK Defence Science & Technology Laboratory to work alongside a consortium of universities and industrial research laboratories in U.S. and UK. The goal of the alliance is to foster collaborative fundamental research in both nations that will to enable secure dynamic semantically aware distributed analytics for situational understanding in coalition operations. The members of the alliance seek to break down barriers, build relationships, develop mutual understanding and work in partnership to develop technology for the U.S. and UK military.

The consortium is led by IBM, which has major research and development operations in both nations. U.S. members of the consortium are University of California at Los Angeles, University of Massachusetts at Amherst, Pennsylvania State University, Purdue University, Stanford University, Yale University and Raytheon BBN Technologies. UK members of the consortium are Cardiff University, Imperial College London, University of Southampton, University College London, Airbus Group and BAE Systems.

DAIS-ITA consists of three components: the Basic Research Component and two Technology Transition Components, one each for U.S. or UK-led efforts. The Basic Research Component provides for fundamental research, the results of which will be in the public domain. The Technology Transition Components will provide for the application of the fundamental-research results to military, security and commercial applications to foster the best technologies for future defense and security needs.

This document describes the initial program plan (IPP) for the DAIS-ITA Basic Research Component, and provides an overview of the research work to be undertaken from September 21, 2016 to January 20, 2018.

The scope of basic research in the program spans two technical areas: Dynamic Secure Coalition Information Infrastructures (TA-1) and Coalition Distributed Analytics and Situational Understanding (TA-2). TA-1 will perform fundamental underpinning research for enabling distributed, dynamic, secure coalition communication/information infrastructures that support distributed analytics to derive situational understanding. Coalition operations at the tactical edge encounter severe resource constraints and rapid changes in the environment. The research in TA-1 seeks to develop techniques for dynamic, self-configuring services that build services "on-demand," taking into account changing mission needs, context and resource constraints, while seeking to protect coalition information and assets. TA-2 will explore the principles underlying distributed analytics and situational understanding, taking into account the fact that coalition operations involve complex multi-actor situations, have information with a high degree of complexity, needs to be processed in a time-sensitive manner at a high tempo, and are required to align itself with human needs and capabilities.

The outputs of the basic research component of the program will advance the state-of-the-art, develop fundamental knowledge, and provide generalizable results. This fundamental science will be manifested in scientific publications in peer reviewed conferences and journals, books covering subjects in scope of the program, as well as trained researchers. Experimental validation of the research is critical and any experimentation software will be made available across the Alliance (ideally as open source) and may be integrated into an experimental framework to enable wide-scale experiments to validate inter-disciplinary research.

The initial program plan consists of six projects, each of which address issues that cut across both technical areas. From an organizational perspective, the first three projects address more issues in TA-1, while the last three projects address more issues in TA-2. The six projects are:

- P1: *Software Defined Coalitions*: will explore the principles by which different elements across a coalition could be composed via control plane interactions to form a virtualized larger element.
- P2: *Generative Policy Models for Coalitions*: will investigate approaches for policy based management in a coalition environment with sufficient autonomy to its constituent elements.
- P3: *Agile Composition for Coalition Environments*: will explore new architectures in which analytics code and data of various types (ISR, HUMINT etc) are mobile and composed together optimally.
- P4: *Evolution of Complex Adaptive Human Systems*: explores the properties of external groups relevant to a coalition operation, and how such external groups evolve over time.

- P5: *Instinctive Analytics in a Coalition Environment*: investigates approaches for data and services can be matched together to create the analytics services that are autonomous and optimal.
- P6: *Anticipatory Situational Understanding for Coalitions*: explores new analytics algorithms for proactive situation understanding that can enable create intelligent advisors for human in the loop systems.

After describing the overall research vision, this IPP describes each of the projects in more detail.

Research Vision

With the explosion in low cost phones, wearables and the Internet of Things, most coalition operations will take place in an environment with a diverse set of small elements capable of computation, storage and communication. We propose leveraging the various devices available across the coalition members to create a system with distributed collaborative and cooperative capabilities. This interconnected system will provide an infrastructure for performing analytics required for coalition operations. It will leverage all the services offered by a wired backend infrastructure (e.g. a backend cloud system, data center or available cellular network infrastructure) but it will not be critically dependent on a continuous connectivity to the backend.

We envision a future where the interconnected system operates seamlessly across networks and systems belonging to different organizations (i.e. coalition members or sub-groups within a single coalition member). This system is frequently charged with performing tasks that require creating dynamic groups on a short notice. Such dynamic groups may be short-lived (days or hours), but could also last for a longer period (months). Differences in the pedigree of disparate systems belonging to different organizations necessitate the development of approaches that work with partial visibility, partial trust, and cultural differences, while simultaneously dealing with the challenges of a dynamically changing situation in which power, computation and connectivity may be severely constrained.

We want the ability to create an intelligent interconnected system, i.e. a system that can analyze the situation on the ground in real-time, anticipate the situation likely to happen in the future, and determine whether the situation requires human involvement. If the situation does not require human involvement, the system would undertake the most appropriate automatic action to the situation. When the situation needs human involvement, the system will recommend alternative courses of actions, along with their pros and cons. We refer to this capability that coordinates different elements, with opportunistic assistance from a fixed infrastructure with interrupted connectivity, as the *distributed coalition intelligence*.

2,5 and 10 Year Goals

The goal of our basic research is to discover and formulate the scientific principles that enable the physical realization of *distributed coalition intelligence* at the conclusion of our 10-year research agenda. This physical realization will require the transition of our basic research into the appropriate systems and solution development. We use the metaphor of a *distributed brain* to describe the end-vision. Just as the human brain is made of two parts, a left hemisphere and a right hemisphere, the distributed coalition intelligence will be an aggregation of several smaller sub-brains, each sub-brain belonging to a coalition member. All of the sub-brains work in a coordinated manner to perform analytics, and leverage the assets and knowledge available across the entire system. Just like the left hemisphere and right hemisphere of the human brain react differently to different stimuli, we expect different sub-brains to react differently in any situation, but the overall distributed system coordinates the different reactions in a seamless manner as needed. A pictorial representation of the concept is shown in Figure RP-1.

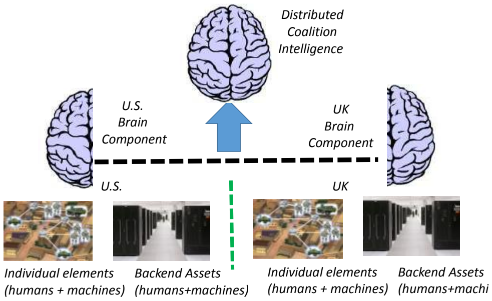

Distributed Coalition Intelligence
U.S. Brain Component
UK Brain Component
U.S.
UK
Individual elements (humans + machines)
Backend Assets (humans+machines)
Individual elements (humans + machines)
Backend Assets (humans+machi

Figure RP-1

To attain the 10-year goal outlined above, we need to understand the fundamental principles underlying some of the key properties of the distributed coalition intelligence when applied to analytics. Our 5-year goal is to understand those principles underlying those properties.

In order to achieve our strategic vision, we must get an insight into the following properties by the end of 5-years.

❖ *Composability*: How do we compose smaller elements into a larger aggregate that works like a seamless whole? What are the principles that link the attributes of a component to the larger whole, and how can we compose components belonging to different organizations with partial visibility and control in an environment with limited resources?

- *Interactivity*: How do different computing elements and people interact with each other, both with other members of the groups and to external stimulus from the environment? How should we model and understand the interactions between different elements and information sources? How do different sub-brains work together as a larger aggregate brain under?
- *Optimality*: How can elements work together to obtain the optimal results in an environment with constrained resources? How can analytics be performed so that optimal performance is obtained automatically, instead of requiring complex manual optimization?
- *Autonomy*: How can elements work together in a proactive manner understanding future situations sufficiently well to operate with a degree of autonomous behavior? How can a system determine that autonomous operation is inappropriate and human intervention is needed? How can different elements simplify the cognitive burden involved to best assist humans in the loop when intervention is needed?

Understanding the principles behind these four attributes will allow us to attain significant capabilities for military defense as articulated in the UK MoD Technology Roadmap and in the U.S. third offset strategy.

Our six projects are defined so that the insights we obtain from them can be combined to help us understand the underpinnings of the four attributes. Our current view on how the different projects can be linked together to obtain the understanding of the four properties at the 5-year point is shown in Figure RP-2. Specifically, we plan on combining the results from projects P1, P2, P3 and P4 to understand the principles underlying composability, the results from projects P2, P4 and P6 to understand the principles that explain interactions among groups, the results from projects P1, P3 and P5 to understand how to self-optimize a system under limited resources, and the results from projects P2, P5 and P6 to understand the principles of autonomy. The results from the 5 years will be combined to get insights into principles governing the distributed coalition intelligence, which will enable us to physically realize such a system in 10 years by combining these scientific insights with appropriate systems building efforts.

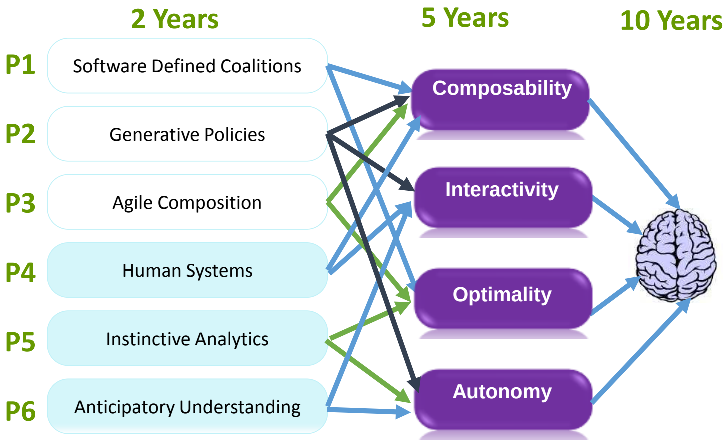


Figure RP-2

Project P1: Software Defined Coalitions (SDC)

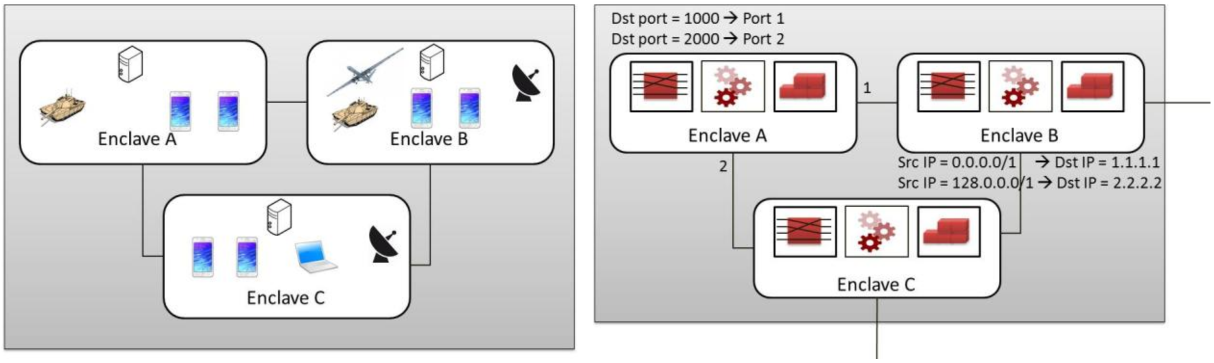
Project Champion: Kin Leung, Imperial College	
Primary Research Staff	**Collaborators**
Christopher Dearlove (BAE)	Ananthram Swami (ARL)
Dave Conway-Jones (IBM-UK)	Christopher Williams (Dstl)
Don Towsley (UMass)	Kevin Chan (ARL)
Erich Nahum (IBM-US)	Leandros Tassiulas (Yale)
Franck Le (IBM-US)	Liang Ma (IBM-US)
Kelvin Marcus (ARL)	Prithwish Basu (Raytheon – BBN)
Kin Leung (Imperial)	
Miguel Rio (UCL)	
Paul Yu (ARL)	
Richard Yang (Yale)	
Saikat Guha (Raytheon – BBN)	

To attain the vision of a distributed brain, we need to understand how the many different elements in a coalition can be composed together to act as a unified whole, despite resource (e.g., bandwidth, power) constraints, disruptions and partial visibility of partner assets. The focus of project P1 is to discover the fundamental principles and techniques by which we can obtain such a composed infrastructure, which we call a Software Defined Coalition (SDC), since it extends Software Defined Networking (SDN)'s concept of separating the control and data plane to all coalition resources, including storage and computation, enabling a new level of agility and dynamism.

Tactical coalition networks have two unique challenges related to composability. First, control can be *distributed across coalition members* with different policies, priorities, resources, and levels of trust among each other. Coalition members require secure access to resources across coalition boundaries in order to interoperate and achieve desired mission goals. Second, the level of *dynamism* is very high, due to the variety and velocity of mobile resources (troops, vehicles, etc.), network partitioning, and incomplete information, among other factors. In particular, network *fragmentation and re-integration* can be a common occurrence and must be a central focus of any approach for composing resources. Coalition operations frequently require the creation of dynamic communities of interests to conduct operations. This requires establishing functional networks across organizations within minutes, if not seconds. This is in stark contrast to agreements among commercial fixed networks that happen at a time scale of days or months.

In order to address these challenges, we will research the concept of software defined coalitions. A Software Defined Coalition (SDC) is a dynamic virtualized aggregate of individual elements that can be created, dissolved and reconfigured in a timeframe of seconds to minutes. Each of the individual elements in the aggregate is a highly configurable and dynamic collection of storage, compute and network resources. The core concept of SDN is to separate control aspects of the network from the data plane flow, enabling centralized management, rapid configuration and adjustment of individual elements to achieve higher-level goals. SDC applies the same concept of logical control plane separation to storage and computational resources, obtaining a control plane architecture that can compose diverse elements together in a dynamic manner.

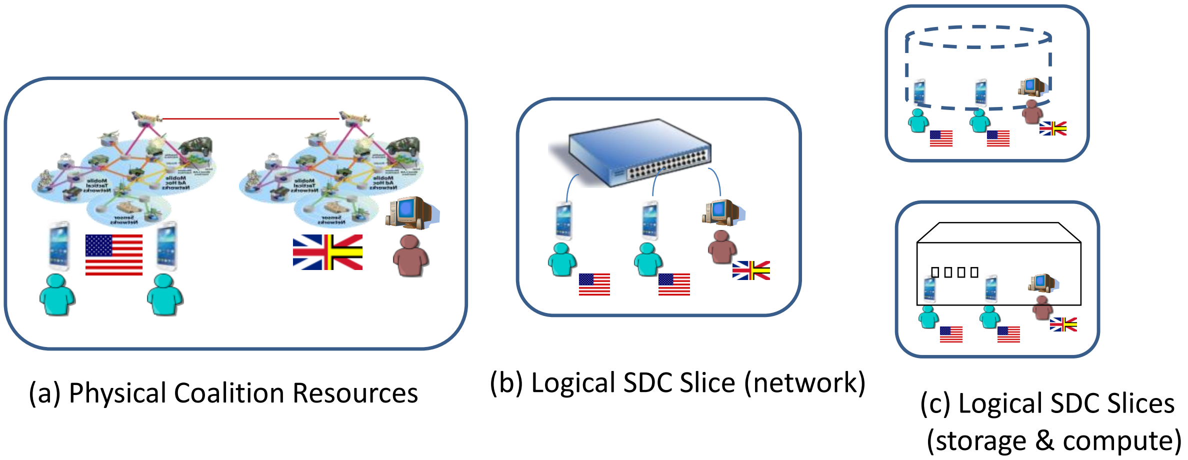
The notion of SDC is based on the abstraction of an *enclave*. An enclave is a set of *resources* (e.g., compute nodes, link bandwidth, storage, power, radios) that are controlled by a *single logical* controller. This logical controller may consist of multiple physical controllers, but all of them collectively act as a single unit in a coordinated manner (e.g., backup or failover controllers are physically redundant but logically single). An enclave can belong to a coalition partner (e.g., U.S., UK) or to a small unit of a coalition partner (e.g., U.S. Army working with U.S. Marines for a joint mission). The controller of an enclave is responsible for managing its resources. An enclave can *fragment* into two when the network is partitioned and *(re-)integrate* when communication is restored or introduced.



(a) An SDC with 3 Enclaves (b) A logical, abstracted view

Figure P1-1: SDC and Enclaves

A coalition partner may have multiple enclaves under its control (e.g., a battalion, a squad). To form an SDC, enclaves under different coalition partners will have to peer with and expose some set of resources to their partners. To enable enclaves to coordinate with each other, and at the same time to protect the security of individual enclaves, we propose that each enclave appear logically as an abstract aggregate of computation, connectivity and storage capacities available to a peering enclave, where the abstraction is computed based on both the needs and security policies of the peering enclave. The difference between physical and logical views is shown in Figure P1-1.


(a) Physical Coalition Resources

(b) Logical SDC Slice (network)

(c) Logical SDC Slices (storage & compute)

Figure P1-2. SDC Slice Concept

Within this overall system, the controllers in the environment communicate and coordinate with each other to create *SDC Slices*. An *SDC slice* is a collection of resources provided by one or more enclaves to work together to provide virtual communications, computation and storage resources for coalition members. As an example, consider an environment with two enclaves, one operated by the U.S. and another operated by the UK. A dynamic coalition community of interest is formed involving two U.S. war-fighters using their hand-held devices with a UK analyst working off his laptop. An SDC slice creates a virtual connection spanning the coalition and enables the entire communication infrastructure to look like a single virtual network that maintains connectivity between the three involved machines. From the perspective of the members in the dynamic coalition group, this connectivity is provided for their exclusive use with appropriate security policies and an acceptable quality of service. The concept of SDC analogously extends to storage and computation in that the different machines belonging to the war-fighters can logically appear to be a single logical storage system (or database) or as a single logical *computing system* that can be used for a specific mission function. Figure P1-2 shows a logical SDC slice along with the associated physical resources that support that view.

Our goal is to discover the principles and techniques that can help to form SDC slices not only for networking (a virtual network connecting members of the dynamic community of interest) but also for storage (a virtual storage or database system shared by the members), for computation (a virtual computer cluster shared by the members).

It is important to emphasize that SDC is not just a straightforward combination of the software defined network, storage and computation, which have been investigated individually in single-domain environments. It is so because there exist complicated tradeoffs among communications bandwidth, computation and memory usage to accommodate the specific coalition needs of the infrastructures and services under consideration. For example, availability of copies of data and analytics in local cache memory can significantly reduce the amount of communication bandwidth consumption, while computation can compress a huge volume of data into a small amount, which consumes less bandwidth for transmission and occupies less memory. The notion of SDC enables the infrastructure owners and users to exploit such tradeoffs among the three types of resources in dynamic coalition environments. We plan to focus on the investigation of such tradeoffs from the infrastructure and service perspective.

Any novel networking approach for the SDC raises new security concerns, and these will be addressed in P1. SDCs raise issues both within and between enclaves, to ensure that information is exchanged confidentially and with integrity between the properly authenticated participants. The data plane will be secured dynamically using Network Function Virtualization (NFV) techniques, which are complementary to SDN. Securing the control plane is even more vital, given the need to authenticate legitimate coalition partners, and the potentially more serious consequences of losing command and control. Solutions will be considered that include conventional server-based solutions as well as those based on modern cryptographic techniques, such as identity and attribute based encryption. Security will thus be a concern cutting across all aspects of P1.

In the IPP, two complementary activities will be undertaken concurrently in order to develop the initial design principles for the SDC. These activities will lay the foundation for the subsequent BPPs to fully investigate the underlying design and use of control plane architectures addressing the requirements outlined above.

- *Theoretical Foundations of Software Defined Coalitions*: We will determine the fundamental performance limits and security characteristics associated with the concept of software defined coalitions.

- *Dynamic and Self-Optimized Control Plane for Software Defined Coalitions:* We will investigate the design principles for the control plane to support the dynamic requirements of software defined coalitions, and optimize their structure and behavior automatically.

Task 1.1: Theoretical Foundations of Software Defined Coalitions

Primary Research Staff	Collaborators
Don Towsley (UMass)	Ananthram Swami (ARL)
Erich Nahum (IBM-US)	Kevin Chan (ARL)
Kin Leung (Imperial)	Leandros Tassiulas (Yale)
Paul Yu (ARL)	Liang Ma (IBM-US)
Saikat Guha (Raytheon – BBN)	Prithwish Basu (Raytheon – BBN)
	Richard Yang (Yale)

The selection of an appropriate SDC architecture, its configuration, allocation of SDC slices to coalition missions, and the ability to handle dynamics typical of military scenarios are extremely challenging to analyze. This task focuses on modeling and analysis issues associated with SDC. Our initial focus will be on the three issues of: (1) Infrastructure fragmentation and re-grouping, (2) maintaining quality of service in the presence of fragmentation, (3) distributed allocation of infrastructure resources to meet coalition mission demands.

Fragmentation and re-grouping occur due to mobility and fluctuations in the communication channel. To investigate performance limits for finite, fragmented infrastructures, we need to understand how network global structural/functional properties (e.g., connectivity, fragmentation, network diameter, controllability, robustness, etc.) relate to various network configurations, mobility patterns, and (potentially) environmental variations in heterogeneous (coalition) network settings, and, more importantly, how such complicated relations translate into fundamental conditions and laws that govern network transition rules. We will apply principles from *percolation theory* to this challenging problem. Percolation theory is well suited to the study of fragmentation and transitions from fragmented networks to connected networks. This will provide performance limits and associated conditions that will enable SDC controllers to proactively use fragmented resources and decide when fragmentation is useful for furthering a mission. SDC must be equipped with capability to use mobile assets (e.g., UAVs, vehicles, robots and drones) to provide connectivity among fragmented infrastructures on the control and data planes in order to maintain expected quality of service. Furthermore, it must be equipped with the capability of fragmenting the resources when this will enhance the mission. Using percolation theory, we will determine rules for when we should move resources for the purpose of enhancing connectivity and performance or for introducing mission specific (forced) fragmentation. Resources including communication, compute and storage must be allocated to SDC slices in light of competing mission demands. Optimal resource allocation must be performed in a distributed manner and must consider the possible fragmentations and temporary connectivity by mobile assets. We will extend existing *distributed optimization theory* and investigate the multi-objective optimization framework to serve as the theoretical foundation for such resource allocation in the SDC. We elaborate on these aspects below.

Percolation theory, controlled mobility, and software defined coalitions: Percolation theory has been used to model and investigate network performance limits, but simplifying assumptions are needed to make the model tractable. These include infinite networks, random node mobility, and simplified network topologies in order to obtain tractable results such as asymptotic performance limits. Percolation theory is well equipped to study spatial and temporal properties of network fragmentation when the network lies in the "subcritical" connected regime; that is, when link or node probabilities lie below a critical percolation threshold. We propose to use percolation theory to explore the following fundamental problems from the perspective of different performance metrics:

(a) *Connectivity*. We explore what the critical percolation threshold, such as density and/or average number of neighbors per node, is so that the connectivity of the entire network can be maintained (or with small probability of being fragmented). We will investigate how different SDC architectures (corresponding to different shape models per node) affect this threshold between a fragmented and a connected state of the network. The conditions on the threshold of the connectivity can be relaxed when we consider local connectivity. In particular, when global connectivity cannot be guaranteed, we need to get insights into how clusters (connected sub-networks) with different sizes are distributed (small probability of the formation of sub-networks with only a few nodes is preferred), and understand how it is determined by percolation thresholds.

(b) *Controlled Mobility:* We would first investigate the impact of factors such as parameter regimes, network architecture, density of mobile assets, and latency-throughput trade-offs on the connected-fragmented threshold for SDCs. This analysis can provide insights into how to position data and analytics given where nodes currently are, without allowing controlled mobility, and by allowing controlled mobility. This will allow us to validate how effective controlled mobility can be within such environments.

(c) *Controllability*. Even if the network is fragmented, we still require the network to be controllable under the SDC architecture. Therefore, an intuitive way to characterize controllability is that each cluster must contain at least one SDN controller. Thus, we investigate under what conditions such controllability in dynamic network environment can be guaranteed.

(d) *Robustness*. One way of characterizing robustness in the SDC infrastructure is using the vertex-connectivity in each cluster. Specifically, the vertex-connectivity directly determines the number of independent paths (i.e., internally vertex-disjoint paths) between any node pair in a cluster. Therefore, high vertex-connectivity in a cluster implies that the SDC controller has more flexibility in configuring data flows between nodes and is more robust against possible node/link failures/interruptions. Hence, the relationship between the vertex-connectivity and the percolation threshold is also of paramount importance in this task. Our goal is therefore to understand the robustness thresholds for an SDC architecture, develop a framework in which one can optimize the SDC architecture to obtain the best thresholds, and finally provide useful inputs, in terms of parameter regimes and network architecture, to the distributed-algorithm development part of our project. On the other hand, one may need to operate an SDC beyond a second "robustness-threshold," which is deeper into the "supercritical" connected regime, in order to facilitate handling distributed analytics workloads at a compute throughput above the threshold driven by the mission objectives. While such robustness properties have received attention for pure network settings, adaptation to SDC slices has not been done.

For all the above fundamental problems, we will start with infinite size networks as percolation theory gives accurate results for infinite systems, which nevertheless will then shed light on approaches in finite systems. Specifically, by leveraging results in infinite networks, we study an approximate characterization of the achievable properties of a connected finite SDC. We will explore ways to apply percolation theory to make accurate predictions of conditions under which finite SDCs are connected and endowed with good performance properties. We will determine bounds on accuracy and on developing accurate approximations.

The significance of the results that we will derive using percolation theory is that they will not only provide fundamental understanding of the network behaviors in the coalition SDC infrastructure, but also serve as building blocks to further integrate more complicated issues that exist in the network. For instance, the network might be rich in mobile nodes. To understand how node mobility may affect network performance limits, we propose to address this issue by mapping node mobility models (e.g., group mobility models) to node appearance probabilities at particular locations in the network, and modeling link fluctuations to a specific distribution of their associated shapes in the continuum percolation model. In this way, the afore-developed methods can be reapplied to address such mobility issues. In addition, for security reasons, SDC controllers may also be enabled as mobile nodes. Such SDC mobility can again be integrated into our percolation-theory-based framework, although heterogeneous node models may be required.

An important feature of our proposed SDC framework is that dynamics are experienced not just by nodes and network links alone, they are also experienced by *data* and *analytics* objects. Although the dynamics of data/analytics objects are typically *controlled*, e.g., as a result of optimizing caching, it is interesting to also consider simple random dynamics for load balancing, e.g., deflect a computation or data to a neighboring node in an SDC slice when the current node becomes overloaded, or perform random replication under fragmentation. We will initially characterize percolation thresholds under such higher order random dynamics in an SDC slice and subsequently consider proactive and reactive models of data/analytics dynamics relative to network dynamics.

Leveraging percolation theory and capacity results for infinite networks, we propose to also study the impact of controlled dynamicity on the capacity of SDCs. Mobility can improve network performance. For example, uncontrolled mobility increases per node throughput capacity in MANETs from O($\sqrt{n}$) to O(1) at the cost of unbounded end-to-end latencies[1]. We conjecture that "controlled mobility" preserves the O(1) throughput scaling, while allowing bounded latency. We will investigate new scaling laws for "data" throughput in an SDC slice with a given distribution of controller locations, assuming various models of dynamism of the network and the aggregate workload. We will focus on system metrics over time intervals relevant to the coalition mission under study.

Resource allocation in software defined coalitions: Distributed utility maximization is a well-established technique to achieve resource allocation. However, the technique to date focuses on allocating networking resources such as communication bandwidth and transmission power. We will develop a new unified framework that accounts for computing and storage as well as communication resources. Our key insight regarding the inclusion of storage is to associate each content (e.g., data file) or aggregate with an expiration timer and to set the different timer values for different contents so as to maximize an aggregate utility measure subject to storage capacity constraints. Individual utilities reflect the importance of different contents, missions, and desired tradeoffs between storage and network traffic. The proposed framework will also be a building block on which Project 3 can design content-based networks where data is cached at multiple routers. Our research here will focus on the proper definitions of such utilities, and how to balance between the need of associating utilities with contents, missions, and users. We plan to consider not only typical network performance parameters associated with resource allocation, but also overheads

[1] Grossglauser, M., & Tse, D. N. (2002). Mobility increases the capacity of ad hoc wireless networks. Networking, IEEE/ACM Transactions on, 10(4), 477-486.

associated with the allocation decisions and security/trust factors for sharing of data and analytics. In addition, the optimization formulation must account for diverse and possibly contradictory coalition needs. The solution to the formulation is expected to align and harmonize the potentially conflicting mission objectives. Our research will focus on a new theoretical foundation for the development of distributed algorithms to solve this class of problems.

We will also extend distributed optimization theory to trade off performance between computation, communications, and storage for data-analytic applications. For example, data processing can take place at different elements and intermediate results and data are forwarded from element to element across the network infrastructure as processing continues until completion. We have shown[2] that the formulation to achieve the optimal desirable tradeoff between communication and computation is a non-convex optimization problem for which existing solution techniques do not generally apply. Our initial results reveal that distributed optimization algorithms are capable of generating the optimal resource allocation and tradeoff for several problem formulations. In addition, we have developed distributed optimization algorithms that account for storage[3]. To allocate resources to missions with conflicting needs, we also plan to extend our investigation from non-convex optimization with a single utility to a multi-objective optimization (MOO) framework where the contradictory objectives cannot be scalarized into one utility function. Our research will focus on the following five aspects: (a) Developing appropriate resource allocation formulations through appropriate definitions of utility functions and constraints to capture performance tradeoffs among communications, storage, and computation for given missions. The new formulations will consider not only network performance such as delay, capacity and efficiency, but also constraints such as security/trust concerns of sharing of data and analytic, as well as new definitions of utilities to align objectives of different missions. (This aspect of the work focusing on infrastructure is coupled with Task 3.1 with the service perspective.) (b) Considering the possible infrastructure fragmentations and temporary connectivity by mobile assets as factors in the resource allocation formulation to quantify performance improvement and resilience of the infrastructure with augmented connectivity. (c) Extending distributed optimization theory[4] to ensure distributed solutions to the above non-convex problem formulations. (d) Extending existing MOO frameworks[5,6] and developing solution techniques, distributed algorithms where possible, to allocate infrastructure resources in coalition environments, (e) Using the new theory to develop efficient distributed algorithms for resource allocation or control in the SDC including fragmented infrastructures. The complexity and overheads of the newly proposed distributed algorithms will also be investigated and minimized.

As part of the above resource allocation research threads, we will develop and quantify metrics for evaluating the heuristic algorithms for infrastructure settings where optimal performance is not guaranteed. These metrics need to provide assessment of not just network performance but aspects that capture the essence of interoperability, tactical coalition operations and heterogeneous networking. Metrics will be independent of specific system architectures or solutions but take into account diverse and possibly contradictory coalition needs, and cover aspects such as overheads, complexity, responsiveness, 'optimality' of control, security/trust, alignment with mission objectives, network awareness and resilience.

We will evaluate the abilities of proposed algorithms to adapt to dynamic conditions using simple Markov models that allow control of the timescales governing infrastructure and workload changes. Ultimately, we will merge the resulting frameworks to design and analyze distributed optimization algorithms accounting for all SDC resources. These new distributed algorithms will be used to solve corresponding problems such as analytic/service placement in Project 3.

[2] S. Nazemi Gelyan, K.K. Leung and A. Swami, “QoI-aware Tradeoff Between Communication and Computation in Wireless Ad-hoc Networks,” IEEE PIMRC, Valencia, Spain, September 2016.

[3] M. Dehghan, L. Massoulie, D. Towsley, D. Menasche, Y.C. Tay. “A Utility Optimization Approach to Network Cache Design,” Proceedings of Infocom 2016, April 2016.

[4] S. Nazemi, K. Leung, and A. Swami, “A Distributed, Energy-Efficient and QoI-Aware Framework for In-Network Processing. ” IEEE PIMRC Conference, Washington, D.C., Sept 2014.

[5] Z.-Q. Luo and S. Zhang, “Dynamic Spectrum Management: Complexity and Duality,” IEEE Journal of Selected Topics in Signal Processing, Vol. 2, No. 1, pp. 57-73, 2008.

[6] E. Bjornson and E. Jorwcieck, “Optimal Resource Allocation in Coordinated Multi-Cell Systems,” Froundations and Trends in Communications and Information Theory, Vol. 9, No. 2-3, pp. 113-381, 2013.

Another type of resource-allocation technique at finer time scales is the scheduling mechanisms responsible for sharing a specific resource (e.g., a processor) among competing demands. Using the aforementioned algorithms, an SDC slice may have several processors at geographically dispersed locations. We plan to model the interactions and optimize the performance of the distributed resource allocation with schedulers across multiple time scales.

This research task will lead to a fundamental understanding of dynamism in SDC infrastructures, and justify the necessity of employing mobile assets (e.g., UAVs, vehicles, robots and drones) for performance improvement. If mobile assets are indeed required to provide connectivity among fragmented infrastructures, will provide inputs for minimizing the usage of mobile assets while satisfying specific network-wise performance requirement. Last, the task will lead to an understanding of how to trade off different types of resources in an optimal distributed manner. Such fundamental understanding can serve as the basis for the development of efficient operation and configuration of the control plane to be investigated in Task 1.2.

Task Milestones:

Date	Description
Q1	− Establish initial distributed utility optimization (DUO) formulation and derive distributed algorithms to achieve the optimal tradeoff between communications and computation. − Establish latency-vs.-throughput tradeoffs of distributed computation in an SDC in the supercritical connected regime and an associated robust-operation threshold on connectedness.
Q2	− Incorporate storage into the DUO formulation with network performance parameters (e.g., responsiveness, control optimality) and the consideration of overheads and complexity of heterogeneous network environments. Joint investigation on this with Task 3.1 will be carried out. − Incorporate storage into the robust-operation-threshold calculation and evaluate how storage affects the percolation thresholds.
Q3	− Enhance DUO formulation and establish MOO formulation with suitable utility functions to align different mission objectives, and resource and security constraints in SDC. − Incorporate resource and security constraints in SDC into the threshold calculation, and incorporate active control within SDC slices.
Q4	− Incorporate mobility of resources, and evaluate temporal connectivity thresholds in the presence of storage and compute resources.
Q5	− Development and validation of distributed solutions for the enhanced DUO or MOO formulation of resource allocation in the SDC. − Incorporate finite-size networks and realistic network topologies in the percolation-based fundamental thresholds and compare these thresholds with performance of algorithmic resource allocation evaluated by the DUO/MOO formulation.

Task 1.2: Dynamic Control Plane for Software Defined Coalitions

Primary Research Staff	Collaborators
Christopher Dearlove (BAE)	Christopher Williams (Dstl)
Dave Conway-Jones (IBM-UK)	Don Towsley (UMass)
Franck Le (IBM-US)	Erich Nahum (IBM-US)
Kelvin Marcus (ARL)	Kin Leung (Imperial)
Miguel Rio (UCL)	Leandros Tassiulas (Yale)
Richard Yang (Yale)	

A key challenge of SDC is that the level of *dynamism* is very high, due to the variety and velocity of mobile resources (e.g., troops, vehicles, etc.), network partitioning and re-aggregation, incomplete information, online mission changes, among other factors. We will conduct a systematic investigation of the control plane for SDCs, considering the following:

- Control-plane architectures and key architecture components.
- Control plane programming with high-level abstractions, build-in verifications, measurements and security capabilities.

Control plane architectures: The unique features of tactical networks, in particular their dynamicity, make it inefficient to design SDC controllers in a blind manner following the architectures of SDN control in conventional networks or cloud orchestration in data center networks. For example, the fundamental separation of control from data in the SDN paradigm can be obtained by having a separate node as the controller entity that maintains the state of the network. In a conventional network with low dynamicity, this can be done with relatively little overhead or availability penalty. In a MANET environment, this approach may not work well, for example, when the controller can be weakly connected or disconnected from those that it controls.

A first item of this research task is to systematically identify and evaluate the potential architectures for a control plane to enable SDC slices on demand across multiple enclaves. One such architecture consists of a distributed, peering based SDC control plane sitting logically above the layer of peering enclave controllers. Its functionalities will be embedded and distributed physically across the enclaves, including new distributed resource-allocation algorithms that we will develop (in conjunction with Task 1.1). An alternative is to arrange SDC controllers in a hierarchical manner (e.g., a tree with two or more levels) with enclave controllers at the bottom of the hierarchy, a higher-level controller coordinating those controllers that sit below it in the tree, and cross-control links from higher-level controllers to lower-level controllers to increase availability and robustness due to dynamism. Yet an alternative architecture is that a distributed computing based protocol and a logically centralized control system run in parallel, with the decisions by the centralized system having higher priority, achieving a type of opportunistic SDC optimization. Many other architectures are also possible. To avoid degenerating into simple architecture enumerations, this research will conduct a systematic, principled investigation of both existing and new control-plane architectures, and perform both qualitative and quantitative assessments on their benefits and tradeoffs. Specifically, this research will include a *principle-based architecture design technique*[7], which the research staff have applied successfully in designing the IETF standard Application-Layer Traffic Optimization framework, based on primal-dual optimization for problem decomposition to induce a modular architecture. To guide the architecture design, this research will introduce a complete set of metrics to reflect the level of performance, security, and trustworthiness of enclaves (including third-party enclaves) that can be used to design and evaluations of possible SDC architectures achieving sharing of resources within each enclave among coalition partners.

Going beyond high-level architecture design and analysis, whose main goal is to achieve high-level problem decomposition through modular abstractions, this research will investigate fundamental infrastructure optimization mechanisms on the control plane. Specifically, instead of considering dynamicity as a challenge to be overcome, we propose a research paradigm in which dynamicity can be used as an advantageous feature whenever possible. We identify multiple types of controlled dynamicity, including controlled mobile data, controlled mobile analytics (code), controlled placement of controllers, controlled placement of assets to enhance the controller-resource channel, and controlled mobility to enhance the resource plane. Consider SDC settings with the presence of controllable mobile assets such as UAVs, robots, vehicles, etc. These controllable assets can be deployed to *proactively* prevent fragmentation or *reactively* heal fragmentation, in the settings to enhance the communication resource of the resource plane. In such a setting, the controller directs available mobile assets to position themselves in a manner to provide and maintain connectivity and preserve network topology within some constraints. As another example, a controller in an enclave can require mobile assets to provide at least two disjoint routes to each resource they can reach. The mobile assets then position themselves within those constraints (i.e., at least two routes). If one route is broken, the mobile assets respond and establish a new one *before* connectivity is lost. When an enclave satisfies the constraints, it does not need to update its state with the controller. However, when the

[7] H. Xie, Y. Yang, A. Krishnamurthy, Y. Liu, and A. Silberschatz “P4P: Provider Portal for Applications,”. Proceedings of ACM SIGCOMM 2008, pp. 351-362

constraints cannot be satisfied, the controller is informed, collects the current state of the enclave, and then takes appropriate actions, including modifying the instructions to the mobile assets. Suitably controlled mobile assets therefore help to reduce the overhead of SDC by reducing the frequency of state update. Given the benefits illustrated by the preceding examples, we propose the concept of *controlled dynamicity as a key architecture element* for our SDC design. Our investigation of controlled dynamicity will leverage insights from our existing work on controlling mobile assets to enable real-time[8] or delay-tolerant communications[9] in the face of network partition. The focus of Task 1.2 will be on controlled dynamicity as an architecture component, with emphasis on controlled placement of controllers, controlled placement of assets to enhance the controller-resource channel, and controlled mobility to enhance the resource plane. Task 1.2 will leverage the theoretical foundation to be established in Task 1.1, in particular on percolation with controlled mobility. This task will also collaborate with P3, which investigates controlled mobile data, controlled mobile analytics (code). Together, they target reap the whole benefit of controlled dynamicity.

Recognizing that dynamicity is not always controllable, this research will also understand and reduce the impacts of uncontrolled dynamcity. When forced fragmentation happens, fundamental tradeoff will need to be made, due to fundamental challenges such as the CAP Theorem in distributed computing. Different missions may allow different levels of relaxation of consistency. For example, one mission may require that only controller is elected, say using the Paxos algorithm, to enforce strong consistency. In an example scenario of forced fragmentation where two controllers used to control the same enclave but the two controllers are partitioned, relaxed consistency may allow both enclaves to proceed independently to favor availability. With both controlled dynamicity optimization and better understanding of the impacts of undesirable dynamcity, this research targets to fully reap the benefits of logically centralized control to enable SDC and at the same time understand its limitations.

East-west interfaces as modular, decomposition abstractions supporting coalition composition and dynamism: Established paradigms for SDN focus on controllers within a single administrative domain. In the context of coalition operations, SDC controllers need to be extended to communicate information and state across different enclaves. Hence, we introduce another key architecture element: *communications and coordination abstractions for inter-enclave control*. The state-of-art of inter-domain coordination in the Internet is BGP, which sends path vectors (e.g., results of local computation) around. However, this advertisement model is limited to network connectivity only and is request-oblivious, i.e., the sender of the BGP updates does not depend on what the receiver wants. We believe that greater efficiency in control and coordination can be gained by adopting a need-to-know only communication paradigm. For example, in previous work[10], we explored a coordination paradigm where the different controllers talk to each other only when needed. We will investigate additional information propagation approaches, including *receiver declared query*, where the details of the declaration are controlled by the receiver. As opposed to the current BGP model where information is advertised indiscriminately, the need-to-know only communication principle allows a more efficient exchange of state information.

Control plane programming with high-level specification, build-in verifications, measurements and robustness capabilities: The need to handle dynamicity and coalition, either controlled or un-controlled, can add substantial complexity in the operation of the SDC control plane. This research will introduce multiple capabilities spanning the control plane operation life cycle, from specification to verification and to measurements, in order to substantially simplify the task.

To simplify the specification of the SDC control plane without losing generality, we propose to develop Turing-complete, higher-level control plane programming abstractions as an architectural element to simplify the formation of SDC slices. Elevating SDC programming from lower-level “assembly” primitives to higher level programming, we plan to develop both compilers and runtime approaches to realize the higher level programming abstractions, enabling more complex tasks to be automated. We have introduced basic forms of such higher-level

[8] P. Basu and J. Redi, “Movement control algorithms for realization of fault-tolerant ad hoc robot networks,” IEEE Network, vol 18, no 4, 2004, pp. 36-44.

[9] T. He, K. Lee and A. Wami, “Movement Dispatch-and-search: dynamic multi-ferry control in partitioned mobile networks, ” ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2011.

[10] K. Gao, X. Wang, and Y.R. Yang. Routing-State Abstraction Based on Declarative Equivalence. IETF Draft draft-gao-routing-state-abstraction. July 2015.

abstractions in the context of wired SDN environments[11]. We will build upon that paradigm to support storage and computation resources, and enable repositioning of controllable assets, to enable both proactive creation of control plane segments and rapid reaction (e.g., fast repair) and control of dynamicity. Aggregating multiple instances of resources into a virtualized aggregate using load balancing is an approach to provide availability, failover, and scalability to SDC slices. We propose to enable a higher-level abstraction of a load balancing service for distributed analytics, which can then be mapped to the lower level primitives that could be network-based (e.g., OpenFlow) or session-based (e.g., HAProxy) based on the underlying environment. In a coalition setting, one partner may follow a network-based paradigm while the other follows session-based paradigm. We will leverage our previous work on application-aware load balancing[12] to design the new analytics-focused service.

To analyze and verify a control plane specification, we will explore a novel formal analysis approach based on *routing algebras*. Traditionally, three methods have been applied to analyze network algorithms and protocols: (1) Model checking relies on a formal system model that is exhaustively checked against desired properties. This technique often suffers from state space explosion; (2) Theorem proverbs help users construct formal proofs. They require and rely on heavy human intervention to guide the search for proofs; (3) Routing algebras rely on algebraic structures to model routing protocols[13]. At a high level, routing algebras model network protocols and algorithms using algebraic structures (e.g., semi-rings), properties (e.g., associatively) and operations to derive sufficient and necessary conditions for desirable properties. They have been shown to be able to model the rich policy control of contemporary routing protocols[14]. These abstractions have allowed researchers to reason about network protocols, and identify fundamental properties (e.g., strict monotonicity for convergence, isotonicity for optimality of paths, etc.) for different families of routing protocol[15]. In recent work, we extended this theoretical framework to model and reason about the connectivity of heterogeneous domains, and derived new routing mechanisms with provable properties[16]. However, that work still assumes that each domain relies on traditional routing protocols, and did not consider software defined network capabilities. For example, we will explore how routing algebras can be extended to model the semantics and capabilities of SDC – e.g., questions such as, can information exchanged between coalition partners be abstracted as a signature with a partial order? Can the policies applied to information exchanged between coalition partners be represented as an algebraic operation? – and then use them to reason about the correctness and other properties of the required algorithms and protocols.

Complementing formal analysis, extending our previous work[17], and leveraging the flexibility of SDC, we propose *shadow (i.e., what if) slices* as a novel, inherent, pre-deployment testing capability of SDC. A shadow slice runs in parallel with production slices, and can carry either production workload or test workload. The main purpose of a shadow slice is not to carry production workload, but for operational tasks such as pre-deployment validation (e.g., validate no black holes before deployment) and debugging (e.g., execute at diluted timescale). In adversarial environments, shadow slices also enable SDC to frequently transition between different configurations to mitigate the effect of denial of service attacks.

[11] A. Voellmy, J. Wang, Y. R. Yang, B. Ford, and P. Hudak. Maple: Simplifying SDN Programming using Algorithmic Policies. In Proceedings of the ACM SIGCOMM 2013 Conference, pages 87–98. ACM, 2013.

[12] H. Jiang, A. Iyengar, E. Nahum, W. Segmuller, A. Tantawi, and C. Wright. "Design, implementation, and performance of a load balancer for SIP server clusters." IEEE/ACM Transactions on Networking (TON) 20, no. 4 (2012): 1190-1202.

[13] M. Gondran, and M. Minoux. Graphs, dioids and semirings: new models and algorithms. Vol. 41. Springer Science & Business Media, 2008.

[14] J. Sobrinho, "Network routing with path vector protocols: Theory and applications." ACM conference on Applications, technologies, architectures, and protocols for computer communications, 2003, pp. 49-60.

[15] T. Griffin, and J. Sobrinho. "Metarouting." ACM SIGCOMM Computer Communication Review 35, no. 4 (2005): 1-12.

[16] F. Le, G. Xie, and H. Zhang, "Theory and new primitives for safely connecting routing protocol instances," ACM 2011, vol. 41, no. 4.

[17] A. Alimi, Y. Wang and Y.R. Yang. Shadow Configurations as a Network Management Primitive.In Proceedings of SIGCOMM'08.August 2008

To manage path dynamics, we will leverage our experience in multi-path transport[18] to enable multi-path communication that meets coalition security policies. For example, imagine an end host communicating with a peer over a trusted network using MPTCP when a new untrusted path is made available (e.g., through a third-party cellular network). To take advantage of this resource while maintaining security, programmatic dynamicity satisfies policies by establishing a secure tunnel via NFV over the untrusted network and provides multi-path connectivity.

Task Milestones:

Date	Description
Q1	- Design and implement Turing complete, dynamic SDC control plane programming of basic networking resources with automatic dependency tracking and re-execution (from incremental programming to lambda programming) to hand basic dynamism of networking resources in an SDC.
Q2	- Study and define a taxonomy of existing SDx control architectures, including not only SDN control architectures but also control architectures of other resources, for example in the context of cloud computing (e.g., Openstack messaging architecture); survey of known fundamental control primitives (e.g., Paxos for state synchronization), fundamental limits such as the CAP Theorem and their relaxation potentials to better understand fragmentation/reassemble challenges in SDC. As part of the taxonomy, determine security advantages/disadvantages of different approaches. As part of the taxonomy, pay particular attention to dynamism (fragmentation/reassembly), and consider varying metrics. - Extend dynamic SDC control plane programming of networking resources from L2-L4 to L2-L7, to support in-network security and forwarding (SDN -> SDN+NFV).
Q3	- Initial design of SDC control architectures in the context of SDC with one partner. The design and analysis needs to extend from one controller to a group of controllers. Define weakly or disconnected controllers and understand the tradeoffs, in particular consistency and availability tradeoffs in challenging settings including forced fragmentation (limited resources and content availability). - Defining and determining metrics for evaluation, considering factors such as overheads, complexity, responsibility, resiliency, security, network awareness, etc. - Determine suitability of research facilities for potential research evaluation. - Extend dynamic SDC control plane programming from protocol specific to protocol oblivious, to include both wired networks and wireless networks (SDN + SDR, with extension to P5 or others as an optional possibility).
Q4	- SDC control-resource plane capacity optimization: controlled control-resource channel optimization through controlled mobility and in-band signaling. - Extend dynamic SDC control plane programming beyond networking to allow SDC slices with networking and computation resource demands (SDN+NFV -> SDC).
Q5	- SDC control-resource plan optimization extension, including controlled controller placement. - Initial SDC control architecture with initial inter-controller (east-west) design supporting basic infrastructure information such as topology hiding. - Extend dynamic SDC control plane programming to allow SDC slices with networking, computation and storage resources. - Proactive setup and maintain of an SDC slice for relatively stable environment.

[18] Y. Lim, Y. Chen, E. Nahum, D. Towsley, and K. Lee. "Cross-layer path management in multi-path transport protocol for mobile devices." In IEEE INFOCOM 2014, pp. 1815-1823.

Linkages

Project 2 focuses on the design of a novel policy management framework that can support generative policy models for controlling access to information in highly dynamic coalition environments. Generative policy models can provide an efficient mechanism for dynamic control plane architectures. Franck Le (IBM-US) will work with Project 2 team to explore the strong linkages between the two projects.

Project 3 considers the problem of supporting agile composition of code and data in a coalition environment. The algorithms developed in task 1.1 will be used in analyzing performance limits of services in project 3. Project linkage will be provided by Kin Leung (Imperial) who is participating in both projects.

Project 5 provides a way to dynamically compose the analytics from the appropriate services and will provide insights into the development of a self-organizing declarative platform for analytics. Linkage between activities between Project 1 and Project 5 will be provided by Leandros Tassiulas (Yale).

Project 6 provides requirements for the workload that Project 1 needs to support through its concept of software defined coalitions. Miguel Rio (UCL) Simon Julier (UCL) will collaborate with Simon Julier (UCL) of Project 6 to maintain linkage between these projects.

Project P2: Generative Policy Models for Coalitions

Project Champion: Elisa Bertino, Purdue University	
Primary Research Staff	**Collaborators**
Alan Cullen (BAE)	Christopher Gibson (IBM-UK)
Bill Williams (BAE)	Franck Le (IBM-US)
Brian Rivera (ARL)	Geeth de Mel (IBM-UK)
Christopher Williams (Dstl)	Maroun Touma (IBM-US)
Diane Felmlee (PSU)	Patrick Dantressangle (IBM-UK)
Dinesh Verma (IBM-US)	Ramya Raghavendra (IBM-US)
Elisa Bertino (Purdue)	Sean Lane (Purdue)
Emil Lupu (Imperial)	Stephen Pipes (IBM-UK)
Rachel Bellamy (IBM-US)	Supriyo Chakraborty (IBM-US)
Saritha Arunkumar (IBM-UK)	
Seraphin Calo (IBM-US)	

Different parts of a coalition are governed by their own sets of policies defined as directives used to guide their actions. The vision of a distributed coalition intelligence requires a dynamic, secure and resilient information infrastructure that needs to conform to the policies of each coalition member. The appropriate policy based management framework will help to attain key attributes such as autonomous operation, composing systems together, and controlling interaction among elements.

Policy technologies have been used successfully in management of IT systems and networks, but prevalent approaches tend to rely on rule-based systems that rely on centralized services. Coalition environments are highly dynamic, distributed, and heterogeneous, frequently without access to a centralized infrastructure. Although advances have been made for policy enforcement in coalition operations[19], many challenges remain in addressing the high degree of dynamism and mobility encountered in coalition operations.

In coalition environments, policy issues need to address characteristics of the humans involved in the missions as well as the computer systems involved. Current policy approaches are computer-centric and do not take such considerations into account. New policy models that can adequately capture human aspects, both cultural and sociological are needed.

The current state of the art in policy management infrastructures focuses on automated enforcement of directives. However, in a dynamic coalition environment, blind enforcement of predefined policies may prevent the delivery of a critical piece of information from a coalition partner that may be important for mission effectiveness. Policy infrastructures for coalitions must provide for the ability to trade-off mission effectiveness against policy relaxation, and support policy adjustment and negotiation to maximize mission effectiveness while minimizing risk. Current infrastructures for policy fail to live up to this challenge. In order to advance the current state of the art, we propose to undertake two research tasks.

[19] Calo, S. B., Karat, C. M., Karat, J., Lobo, J., Craven, R., Lupu, E., ... & Bandara, A. (2010). Policy Technologies for Security Management in Coalition Networks. Network Science for Military Coalition Operations: Information Exchange and Interaction, 146.

- *Generative Policy Algorithms and Systems*: We will research new policy architectures in which elements can generate their policies under a loose set of guidance from a central coalition commander, and investigate algorithms that ensure consistency and coherence in the operation of such a system.

- *Generative Policy based Security and Resource Management:* We will apply the generative policy model to create new models for security management in coalition operations, and enable the concept of resource that can contain their policies within themselves.

Figure P2-1 shows an example of how our main policy framework entities could be integrated within a coalition network.

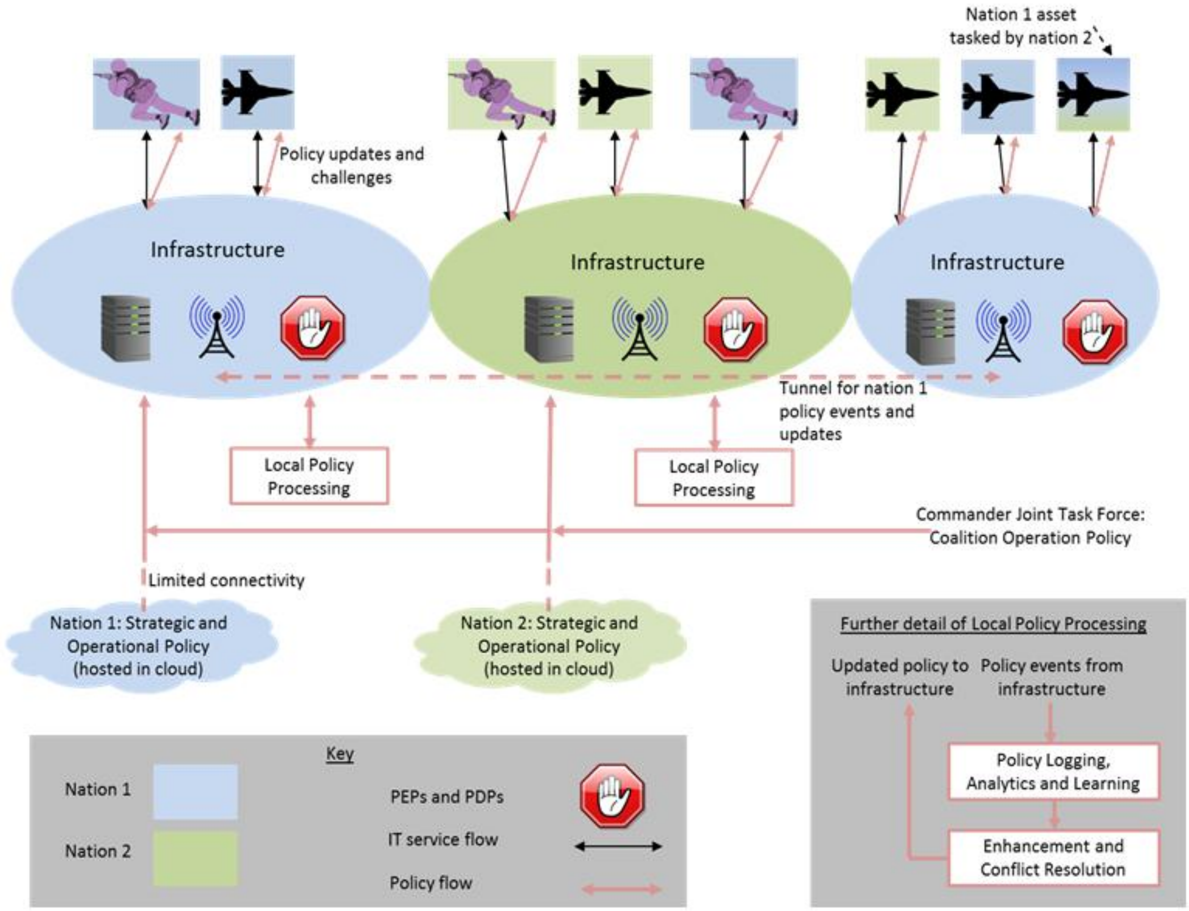

Nation 1 asset
tasked by nation 2
Policy updates and challenges
Infrastructure
Infrastructure
Infrastructure
Tunnel for nation 1 policy events and updates
Local Policy Processing
Local Policy Processing
Commander Joint Task Force: Coalition Operation Policy
Limited connectivity
Nation 1: Strategic and Operational Policy (hosted in cloud)
Nation 2: Strategic and Operational Policy (hosted in cloud)
Further detail of Local Policy Processing
Updated policy to infrastructure
Policy events from infrastructure
Policy Logging, Analytics and Learning
Enhancement and Conflict Resolution
Key
Nation 1
Nation 2
PEPs and PDPs
IT service flow
Policy flow

Figure P2-1. Example of Applying Policies in Coalition Networks

Task 2.1: Generative Policy Algorithms and Systems

Primary Research Staff	Collaborators
Alan Cullen (BAE)	Patrick Dantressangle (IBM-UK)
Bill Williams (BAE)	Ramya Raghavendra (IBM-US)
Brian Rivera (ARL)	Sean Lane (Purdue)
Christopher Williams (Dstl)	Stephen Pipes (IBM-UK)
Diane Felmlee (PSU)	Supriyo Chakraborty (IBM-US)

Primary Research Staff	Collaborators
Elisa Bertino (Purdue)	
Emil Lupu (Imperial)	
Rachel Bellamy (IBM-US)	
Saritha Arunkumar (IBM-UK)	
Seraphin Calo (IBM-US)	

The work in task 2.1 is organized into a number of activities. The first activity is the definition of a policy lifecycle. The second activity is the definition of a corresponding framework. As such a framework has to be based on a formal model of policies, we will investigate suitable formal models for generative policies. These two activities are complemented by approaches for policy analytics, which cover many different aspects of policy management, including policy conflicts and policy impacts. In what follows we elaborate on these activities.

Policy Life Cycle

The policy based management framework must accommodate the lifecycle for coalition operations as shown in Figure P2-2. The three stages shown in the figure form the basis for the formalization of the policies used in the governance of coalition missions. The roles of humans differ across these stages. In the *strategic planning* stage, high-level commanders negotiate and develop common operating principles applicable to coalition operations in general. This would determine policies independent of any specific coalition mission. The *operation planning* stage occurs when the specific mission (e.g., to plan to rescue a hostage, launch a humanitarian effort to save civilians due to a dam breach, or initiate an immunization drive to prevent the outbreak of a disease) is known. During operation planning, decisions are made regarding which assets are to be used and policies governing the operation of these assets are defined. During the *operation* stage, assets are actively involved, and policies are used to manage the deployment and activities of assets and people. Policy based decision-making needs to be highly automated, while involving humans in the loop as needed. Decisions in the strategic planning stage are made in days; decisions in the operations planning stage are made in hours; and, decisions in the federated network operations stage must be made in seconds or less. These time scales must be taken into account in the development of policy analysis algorithms for the various stages.

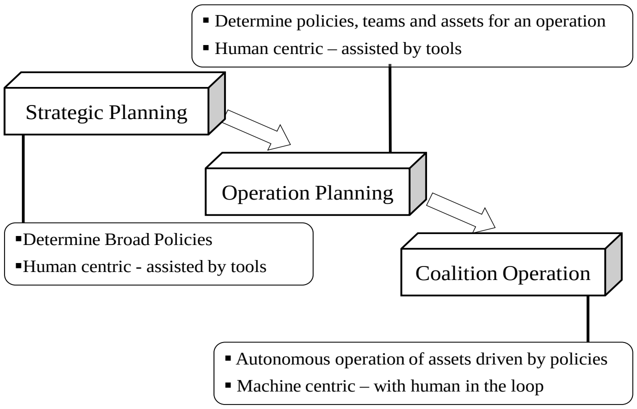

▪ Determine policies, teams and assets for an operation
▪ Human centric – assisted by tools
Strategic Planning
Operation Planning
▪Determine Broad Policies
▪Human centric - assisted by tools
Coalition Operation
▪ Autonomous operation of assets driven by policies
▪ Machine centric – with human in the loop

Figure P2-2

Policies will go through an iterative refinement as they progress through the different stages; the policies at each stage being constrained by the policies at next stage. Policies also go from less specific to more specific, and the computational infrastructures available at each of the life-cycle stages differ across life-cycle stages.

A critical challenge in the policy lifecycle is the dynamic evolution of policies – which we refer to as *dynamic policies*. Policies may have to evolve because of changes in situations, in contexts, or in goals. Also one may find that the current set of policies do not cover all situations of interest (*policy incompleteness*), that actions taken by agents in a given context violate policies (*policy inconsistency with respect to agent behavior*), or that certain policies never apply to any situations of interest (*policy irrelevance*). All these findings represent indications that policy evolution is required.

The models for how policies will be used would also differ. During strategic and operation plan, we can envision commanders jointly making decisions. They can use policy-based approaches implemented as software running on their handhelds/laptops with possible connectivity to backend cloud services.

During the coalition operation stage, access to a backend cloud infrastructure may not be available, but assets may be operated according to the hierarchical structure shown in Figure P2-3. A central entity (e.g., at a base camp or coalition joint task force headquarters) supervises each coalition operation, which in turn consists of multiple operational groups. An operational group may consist of war-fighters from one or more coalition members. Policies at the strategic stage and/or operations planning stage could determine the membership of operational groups. The war-fighter in each operational group is controlling a variety of assets.

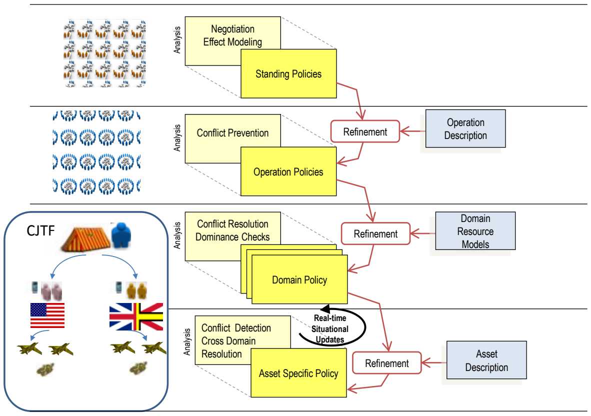


Figure P2-3

We have summarized the successive refinement of policies and the different environments in which they need to operate in Figure P2-3. The standing policies are made during the strategic planning stage and they are developed using negotiations among commanders. At that stage, we need to analyze the impact of the agreed policies on the effectiveness of operations and their conformance with national policies. The operation policies are determined by a set of military planners deciding how to perform a specific mission. At this stage, we need to determine which aspects (or domains) of operation need to be controlled by policies and what those policies ought to be. Once the mission is underway, we assume a dynamic Community of Interest (CoI) formed across coalition members, where a Coalition Joint Task Force (CJTF) commander leads the mission, comprised of U.S. and UK members, with each group controlling assets such as UAVs, mules and computing systems.

Generative Policy Framework and Policy Formal Model

To deal with the complexities of the coalition environment and the variety of context that may be encountered on the ground, we need a new policy framework for federated network operations. Instead of the current model of having the CJTF send a static set of policy rules to humans and assets in the team, we propose a new framework where each element in the system is able to generate its own sets of policy rules as needed, according to the events being seen and its local context.

In this new framework, each element is sent a *policy generation grammar* which would allow a richer mechanism for specifying policies. We call this framework *generative* since the coalition authority will only distribute policy grammars, and the network elements would generate their own policies. Within this generative framework, we will create new policy information models and the grammars that would support the dynamic creation and modification of policies including humans in the loop when needed.

Referring back to Figure P2-3, the standing policies formed during strategic planning are negotiated and described in natural language. These can be converted directly into a standard policy language such as CIM-SPL by a supporting policy negotiation tool, and can capture the goals and requirements that must be followed in conducting joint operations. During the operations' planning stage, when policies for specific missions are determined, we assume that policies are described in a form that enables some freedom for individual elements, e.g. a Context Free Grammar (CFG) or another suitable mechanism. We will develop the appropriate information model for policies description, and we also provide a syntax, for example based on XML. This allows us to have a specific formalism to use for use cases to be developed in Task 2.2. However alternative approaches will be identified. Mission specific policies capture the range of information and actions that are expected to be relevant in the coalition mission, and may also be determined using analysis tools. These models support the specification of policies that are aligned with coalition expectations. Their joint specification establishes the need to investigate automated support for Human-in-the-Loop (HIL) policy negotiation and the characteristics of the tooling needed to support it.

During federated network operations, the policies and policy models are deployed to the distributed coalition elements. The CJTF commander sends policy models to war-fighters involved in the mission, who can use handhelds and other devices available to them to process and use them. In particular, these abstract policy models are refined by the war-fighters into more concrete policies (to be used by assets like UAVs and mules) taking into account the local environment and the operational context. These are the set of policy generators that provide an enabling mechanism for each element of the distributed system to determine its own policies.

We consider that there are separate coalition members' coordination sites, the operational sites supporting member teams, and the IT, robotic and sensor equipment supporting each individual war-fighter. The number of different **policy generators** and their relationships need to be determined and an overall architecture developed to manage the control flow. The policy framework needs to model the hierarchical levels of a policy infrastructure, and capture **relationships** among the different abstraction layers as well as the dynamic interactions among coalition members.

Analysis is needed at each level of refinement, and any potential conflicts resolved. At the lowest layer where executable policies that pertain to a particular resource are determined, policy analysis can be done in a conventional manner using existing techniques. However, at the strategic planning and operation planning stages, the impact of human behavior and how humans will react to different situations need to be taken into account. We plan to incorporate human-centric analysis of policies at these stages, as discussed below in *Policy Impact Analytics*.

There is also a linkage between the lower levels of policy enforcement and the higher levels. With policies being defined for different resources and within different domains at various levels of the system architecture, their specification and maintenance becomes a problem. We will investigate ways that the system itself could **learn the policies** that are appropriate to achieving the goals of a dynamic, distributed coalition.

The consequences of policy enforcement are monitored through logging and analytics. Different metrics for effectiveness would pertain to different domains (security, operations, etc.), and learning mechanisms can be employed at local and coalition member levels to identify the most effective policies. This information could also be used to resolve policy conflicts. Policies are then updated or changed based on operational experience and/or human intervention. There would need to be local overrides for emergency situations or when network segmentation affects policies that require inputs from other nodes that have become unreachable.

The policy generators would revert to the generators set by each coalition member when the CoI disbands. The **policy lifecycle** for each individual organization would be similar to that of the coalition except that the control flow is internal to the organization.

During the first project year, based on the policy lifecycle and policy framework, we will define an architecture that will include the key architectural components and associated services. Requirements concerning the security of the architectures will be identified as well as attacks that specifically exploit the generative policy approach.

Policy Analytics

Our policy lifecycle and framework require two types of policy analytics: (a) *operation policy analytics* that evaluate policies for different assets during the operation of the federated network, determine conflicts and strategies for solving conflicts, and determine the need for policy evolution and, (b) *policy impact analytics* that evaluate the impact of policies defined during strategic planning independent of missions, or of specific policies defined during operation planning for a mission. The former deals with issues of distributed autonomous management of policies in a dynamically changing environment requiring rapid decision making, while the latter deals with aspects that are of longer duration and need to anticipate the needs of future operations. During the first project year, we plan to identify all different types of policy analyses that are needed and that will result in a set of corresponding policy analytics services.

These analysis algorithms check if a set of policy generation grammars result in a consistent set of actions, and the impact they may have on the overall operation of the system. Such analysis techniques for previous models of policies have been developed by our team members in the past[20].

Maintaining overall policy consistency in a system based on generative policies involving humans in the loop represents a significant challenge because: the freedom introduced in policy formulation makes it more difficult to

[20] Lupu, E.C. and Sloman, M., 1999. Conflicts in policy-based distributed systems management. IEEE Transactions on software engineering, 25(6), pp.852-869.

analyze policies for conflicts in advance; the local autonomy in policy formulation may introduce un-anticipated conflicts (locally or globally) that arise only in specific run-time contexts; and, involving humans in policy formulation and policy decisions also introduces an element of human inconsistency.

To address these challenges we will use a combination of strategies that combines: propagating constraints on the policy generation from the global level to the local one but also from the local level to dependent contexts, including compliance checking of policies with propagated constraints; using federation techniques in order to combine and check for the consistency of policies across local environments; using techniques such as argumentation to reason with conflicting policies and understanding how decisions are affected by the conflicts; establishing "safety" boundaries for policy decisions in order to avoid combinations of policies that lead to unsafe outcomes; and, monitoring at run-time for the occurrence of policy conflicts, attempting to resolve them and referring the conflict to human decision where necessary.

We will develop policy analysis techniques associated with the information models that would provide the means for assessing key characteristics of the sets of policies being employed within a particular system (correctness, coverage, conflict detection, etc.) and evaluate their impact in given tactical scenarios. Context free grammars are not closed under many operations. We will thus explore new analysis algorithms. We will also develop policy applications, e.g., schemes that can automatically infer rules governing the behavior of intermediaries in a network flow, or innovative use-cases of integrated data policy systems.

We will also develop data-intensive approaches based on logged data concerning agent requests, and agent actions to identify indicators for policy evolution. One example is to keep track of which policies have been checked before a certain action has been executed by an agent, and determine whether the action is compliant with the policies. If not, this indicates that policies are inconsistent. Supporting such types of analytics require representing relevant actions, such as agent request, policy evaluation, agent action, as temporal sequences and then mining these sequences to identify the indicators of interest.

Policy Deconflicting: One particularly important class of analytics concerns the resolution of policy conflicts. In a distributed and possibly fragmented coalition environment, where coalition elements are highly autonomous and yet share resources, conflicts among different coalition elements are likely to happen. Conflicts can be of different types and arise for multiple reasons. Some arise because multiple policies recommend inconsistent decisions in regards to the same request, others may arise because of conflicts on the concurrent use of resources, and others can still arise because some circumstances arising at run-time have not been foreseen in the policy specification. For example, if multiple coalition elements share a data repository, and each element has its own access control policies concerning the access to the repository, a conflict may arise if multiple policies are applicable to the same request and some policies allow access, whereas others deny access. In this case, we talk of *conflicts arising because of multiple applicable policies concerning a single request*. A different situation is represented by conflicting requests for the same resource by different parties. An example is of two coalition elements concurrently requiring the use of a drone with specialized infrared equipment for high priority missions with a real-time deadline. In such cases, whereas each single request would be allowed if taken in isolation, it is clear that both requests cannot be satisfied at the same time and need to be serialized, if at all possible. In this case, we talk of *conflicts arising because of multiple conflicting requests and the lack of policies concerning concurrent resource usage.* More complex situations can occur where the different types of conflicts can arise for a given set of shared resources and concurrent requests for resource use. Policy deconflicting is critical for dynamic policies as before evolving a policy (or set of policies) one must assess whether new conflicts would be introduced.

A comprehensive and articulated conflict management process is thus required. It is important that such a process be organized around the following conflict management lifecycle: (1) Conflict prevention – to anticipate conflicts as much as possible, to investigate restrictions on policy generators to prevent conflicts, to reconcile conflicting policies, and to introduce policies for conflict resolution; (2) Dynamic conflict resolution – since not all possible conflicts may be identified in advance and/or it may not always be possible to reconcile possible policies in advance, one must be prepared to dynamically resolve conflicts at "run-time", in some cases resorting to the intervention of humans; (3) Post-conflict assessment and learning – once a conflict is dynamically resolved, it is important to gather data concerning the way the conflict was resolved and data related to the conflicting requests, including request context information. Such data can then be used for future conflict resolution and lead to enhanced conflict prevention. Such activity is particularly crucial when humans are involved in the conflict resolution (step 2) since this would allow the conflict management system "to learn" from humans.

Supporting such a conflict management process requires addressing several research challenges, including:

Static policy analysis techniques for conflict identification. An example is represented by techniques for static analyses of XACML policies based on the use of multi-terminal binary decision trees, in turn based on model checking. Such techniques, implemented as part of the EXAM environment by members of the project P2 team[21], allow policy analysts to statically determine for which access requests two XACML policies conflict and for which they do not. The results of the policy conflict analysis are presented in EXAM using a bi-dimensional grid. The use of such techniques requires however major extensions to deal with policy languages different from XACML and to support the analysis of conflicts arising because of multiple conflicting requests for concurrent resource usage.

Techniques for policy reconciliation. Based on results from the static policy analysis, the policies may need to be revised in order to eliminate conflicts. For example, in the case of access control policies, the analysts may decide that when two policies – say P1 and P2 – conflict, P1 has precedence, or they may decide to restrict or expand the applicability of the policies. In order to support policy reconciliation a possible approach is to define an algebra of policy operations. An example of such an algebra is the one developed for XACML[22]. In addition to the development of an algebra suitable for a wider range of policy languages than XACML, important theoretical questions concerning such an algebra have to be addressed, such as completeness, e.g., whether the algebra allows one to express all possible policy reconciliation strategies, and whether the algebra is minimal. In addition, tools need to be developed supporting the high-level specification of algebraic expressions and the automatic generation of restructured policies, expressed in the policy language of interest.

Techniques and strategies supporting dynamic conflict resolution. In general there are various strategies for dealing with conflicts at run-time. One simple approach is to put in place a default conflict resolution strategy; for example, in the case of access control policies, a simple default strategy is one by which if one policy grants access and the other denies access, the policy denying access prevails. Another possible strategy is to look for past occurrences of the same conflict and apply the same conflict resolution decision taken in the past. Finally, another strategy is to ask a (group of) human user(s) to solve the conflict. Notice that such strategies are often complementary. For example, one may first look for past occurrences of the same conflict and automatically apply the same decision. If there are no such occurrences, one may apply the default conflict resolution policy. If no such default exists, the system should prompt a human for conflict solving. In other cases, even though the system may be able to automatically solve the conflicts, the involvement of humans may still be required to validate the decisions. As part of project P2 we will identify various strategies for dynamic conflict resolution, including possible conflict resolution steps and the optimal sequence of resolution step applications.

A critical issue when involving humans in conflict resolution is which information is to be presented to humans in order to facilitate their decision taking. Various strategies are possible, including providing humans with: examples of similar decision situations, recommendations with explanations, and examples of decision consequences. How to visualize such information is a critical issue. Also which strategy to apply and how to prioritize the various strategies may depend on whether the conflict resolution decision must be taken by a single individual or multiple individuals. To address such issues, we will investigate human factors concerning policy conflict resolution. It is also important to notice that groundbreaking work on behavioral economics and prospect theory by Kahnenam (recipient of the 2002 Nobel Prize in Economics) has shown that when making decisions humans have certain limitations and are subject to certain biases[23]. Therefore it is important that strategies to be investigated for conflict resolution processes involving humans be informed by such previous work as well as more recent work on bounded rationality[24] in order to determine the best information that should be presented to users.

Techniques for assessing decision consequences and learning from past decisions. We need tools to acquire and log relevant information concerning the consequences of past decisions and to “quantify” such consequences.

[21] Lin, D., Rao, P., Bertino, E., Li, N. and Lobo, J., 2010. EXAM: a comprehensive environment for the analysis of access control policies. International Journal of Information Security, 9(4), pp.253-273.

[22] Rao, P., Lin, D., Bertino, E., Li, N. and Lobo, J., 2011. Fine-grained integration of access control policies. Computers & Security, 30(2), pp.91-107.

[23] Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. science, 185(4157), 1124-1131.

[24] Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. Psychological review, 103(4), 650

Analytic techniques need to be developed that allow one to mine data concerning past decisions. Data mining techniques may need to incorporate information about biases and circumstances under which decisions were taken, and continuously learn, which requires the use of active learning techniques. Results from such assessment and learning must be used to automatically revise policies and conflict resolution. As part of project P2 we will develop learning strategies suitable for dynamic and uncertain environments and develop tools for automatic policy evolution and versioning based on machine learning results.

Policy Impact Analytics: Policy impact analysis develops algorithms that could help in determining the impact of strategic and mission policies on the effectiveness of the operation and using insights from the current ongoing operations to improve those policies. In order to do so, models for analyzing human behavior in response to specific policies need to be developed.

Individuals and human groups will be called upon to choose among, and enact policies, and therefore our multidisciplinary approach will be informed by research on human behavior. Here we utilize social science perspectives that outline the ways in which people: 1) form decisions; and, 2) work together in coalitions and groups. On a broad level, social science theories of human interaction often emphasize the costs and rewards associated with outcomes of interactions.

Social exchange theory[25] is a particularly comprehensive perspective of this type, with roots in economics, sociology, and psychology. Its distinct advantages include its generalizability and the fact that it can be readily operationalized. Social exchange theory has been applied to a wide range of human behaviors[26], although not yet to coalition operations. Variants of social exchange theory include rational choice, behavioral economics, and game theory. Here we will extend the tenets of social exchange theory to coalition decision-making, basing our arguments on previous models applied to interchanges in social relationships [27,28].

The basic tenets of social exchange theory, as applied to group decision-making, include:

- People attempt to maximize the outcomes associated with the decisions they make, with outcomes defined as rewards minus the costs associated with a particular decision
- Satisfaction with a decision is a positive function of outcomes and a negative function of the actor(s)' comparison level (i.e., a standard expectation for outcomes based on past experience or protocol)
- Commitment to a decision, or a line of action, is a positive function of satisfaction and investments (i.e., "sunk costs," or nontransferable material and psychological resources already spent), and a negative function of alternatives (i.e., other available options for payoffs)

The main propositions are summarized as follows:

$$\textit{Outcomes} = \textit{Reward} - \textit{Costs} \qquad (1)$$

$$\textit{Satisfaction} = \textit{Outcomes} - \textit{Comparison Level} \qquad (2)$$

$$\textit{Commitment} = \textit{Satisfaction} - \textit{Alternatives} + \textit{Investments} \qquad (3)$$

In other words, a coalition will decide to pursue a decision, or a line of action, when the anticipated benefits exceed the costs, when the expected outcome is greater than the standard (e.g., protocol), when the alternative options are worse, and the investments into that line of action are high. Note that operationalization of these concepts is context specific. Particular situations and policies will determine the criteria and the manner in which the concepts of Rewards, Costs, and so on, are operationalized. The negotiations of mission policies and strategic policies provide context that we will model and to which we will apply elements of social exchange theory.

[25] Emerson, R. M. (1972). Exchange theory, Part I: A psychological basis for social exchange. Sociological theories in progress, 2, 38-57.

[26] Felmlee, D. H. (2001). No couple is an island: A social network perspective on dyadic stability. Social Forces, 79(4), 1259-1287.

[27] Thibaut, J. W., & Kelley, H. H. (1959). The social psychology of groups.

[28] Arriaga, X. B., & Rusbult, C. E. (1998). Standing in my partner's shoes: Partner perspective taking and reactions to accommodative dilemmas. Personality and Social Psychology Bulletin, 24(9), 927-948.

Another crucial aspect of policy impact analysis is the ability to determine how humans will make policy related decisions in a specific context, and then use them to create policy rules that machines can use to implement in various situations. These algorithms can be used in the post operation analysis model.

Our research agenda will thus address the following major issues: the support of human decision processes, including human influences in policy specification, employing human decisions for conflict resolution, and the manner in which policies should be applied to the control of human assets. The tasks involved include:

1) *How to support user decision making (both single user and collaborative) about policies and conflicts.* Complementary elements of a possible approach would involve: providing **relevant** meta-data, providing **examples** of similar decision situations, providing **recommendations with explanations,** and providing examples of **decision consequences.**
2) *How to learn from human decision-making (both single user decision and collaborative decision).* Complementary elements of a possible approach involve: recording decisions by users and decision contexts/situations/missions and building models from these; recording outcomes of decisions about conflicts and about application of policies; developing/using human behavior models; and, identifying the circumstances under which a decision made by a human would differ from the one taken by the system.
3) *How to decide when a human needs to be involved in a conflict resolution.* In general, if there is sufficient information captured in the system, human intervention will not be required. Such information could include system defaults, prioritization of policies, or maintenance of history (past occurrences of the same conflict and how they were resolved). However, the decisions led to by such information may themselves lead to an uncertain outcome. There is a need to extend models to allow the *confidence* in a conflict resolution to be evaluated at the same time as the decision. Human intervention would then be solicited when the information available in the system does not lead to a sufficient degree of confidence in the decision made. In some cases, even though the system may be able to automatically solve the conflicts, it may also solicit human oversight because of the critical effects of the decision.
4) *How to apply policies to humans.* This will require: understanding the effects of policies on human behavior and decision making; understanding how to best deal with human non- or late compliance with the policy without impairing usability; allowing in the policy formulation for the necessarily stochastic characteristics of human behavior and for graceful degradation; understanding human biases in making decisions across different groups and how policies/decisions are affected by these biases; and, understanding how to explain/present policies to humans.

Task 2.1 - IPP Activities

During the first project year, the following activities will be carried out:

1) Definition of Policy Lifecycle for Coalitions and Design of the Corresponding Architecture

A clear understanding of policy lifecycle for coalitions is the crucial building block on which the policy framework and software architecture can be built. In this activity we will refine the policy lifecycle shown in Figure P2-2 to include more detailed phases and iterations. Figure P2-4 shows a refined lifecycle as a state transition diagram. Notice that several states can be refined into several sub-states. A notable example is the Policy Rules Enforcement State that can be refined into the following sub-states: (a) identification of applicable policy rules; (b) conflict resolution. In turn the conflict resolution state can be refined into several sub-states according to the conflict resolution strategies outlined in the previous discussion of Task 2.1. Developing a comprehensive policy lifecycle for coalitions requires refining the states into sub-states, identifying additional transitions, and events/circumstances triggering transitions. An example is identifying the circumstances under which the need to evolve the rules of a policy would lead to modification of the policy generation grammar. Also notice that given an initial policy generation grammar, the policy refinement activity can be concurrently executed by multiple agents as we can expect that the same grammar will be distributed to multiple members of the coalition that concurrently refine the grammar.

The policy lifecycle developed as first step of Task 2.1 (part of the first milestone for this project) will be further refined by taking into account requirements obtained from the scenarios developed in Task 2.2, from the analysis of the needs of the Autonomous Operational Units (OAUs) part of Task 2.2, and from the other projects through the linkage activities. An example of very relevant application scenario for the policy lifecycle and architecture is represented by the support for secure data and unambiguous name resolution from Project P3. In particular, such resolution as well as data security policies (such as policies concerning digital signature and encryption for data) can be encoded and enforced by our policy framework.

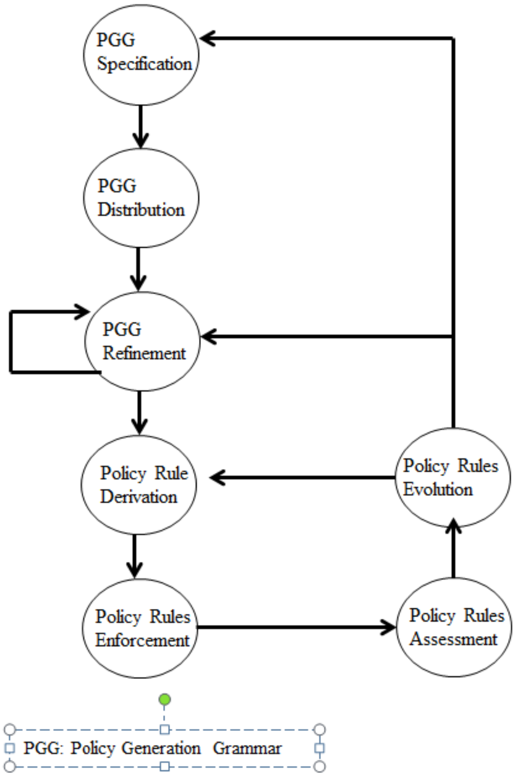


Figure P2-4

Based on the lifecycle developed during this activity, we will identify several characteristics/features of the lifecycle including:

a) input required for each state;

b) for each state, whether there is human involvement and, if so, the nature of the human involvement (e.g. providing input, validation, inspecting policies), and what is required to better support humans (for example human-readable policies);

c) software tools needed to support the activities to be executed at each state.

Based on the detailed policy lifecycle, and especially based on the identified software tools, a policy architecture will be designed.

2) <u>Formal Model and Syntax for Generative Policies</u>

The development of a formal model for generative policies and the use of this model for distributed coalitions requires the investigation of several theoretical aspects and the development of its formal foundational definitions. In this activity we will conduct such investigations and lay the foundations for the policy model used in our framework. A first key question that will be investigated is related to the policy generation grammar refinement as different components of the policy generation grammar can be refined in order to generate the actual policy rules. For example, the initial policy generation grammar may not include the alphabet of the terminal symbols (e.g. the actual values that may appear in the elements of the rules generated by the grammar), and thus each party refining the policy generation grammar must instantiate the grammar with its own alphabet. A second key question is related to the definition of a notion of action compliance with respect to the rules of a policy. An initial possible definition is that an action complies with a policy if the action can be derived from the policy grammar, where the policy grammar is the grammar expressing the policy rules. However this definition needs to be validated. We need to develop similar definitions formalizing the notion of the applicability of a policy rules to an action and the notion of conflicts among policies. We notice here that conflicts between two or more policies can be total, if the policies conflict for each action to which they are applicable, and partial, if the policies conflict only on a subset of actions to which they are applicable. Finally we also need to develop a notion of policy complexity and "compactness" as one of the foundations for the policy metrics[29]. Another important question is to formulate a notion of semantic correctness for the policy rules generated from the policy grammar.

3) Policy analytics services and policy conflict management

The design of a set of policy analytics services will be one of the fundamental and novel contributions of project P2 as our goal is to provide a wide spectrum of analytics supporting the design, refinement, enforcement, assessment and evolution of policies in distributed coalition. Our work in this activity will focus on identifying for each phase in the policy lifecycle (see Figure 2-2), the type of analytics that is required. For example, for the policy rule assessment phase, we need analytics providing indications whether the policy needs to evolve.

In addition, in this activity we will analyze work done in psychology concerning how humans navigate in conflicts and identify results and insights relevant to our context. As one novel aspect of our conflict management is to understand how humans solve conflicts and whether we can build a system that "learns" from humans, we will analyze research done in psychology focusing on how many decisions one has to observe from individuals in order to determine which their decision metrics are. Insights from such work will be taken into account as requirements towards the design of an automatic, or semi-automatic, conflict management system.

4) Impact analysis strategies and algorithms

In this activity, we will identify a set of metrics according to which the impact can be assessed. Understanding which metrics are best applied to a generative policy based system is a challenge within itself, and one that we will address during the first year. The difficulty arises because in such a system, operational agility is favored over pre-planning and autonomy is obtained at the price of increased coordination costs, albeit also inducing reduced complexity in the policy specification. But neither agility, nor autonomy are easily measurable in a generic manner, especially when the baseline is difficult to characterize.

A potential baseline for comparison could be provided by current practice in Joint Coalition Taskforce operations which involve a significant amount of pre-planning. Operational needs that fall outside the pre-planned behavior are either not met or give rise to exceptions that need to be dealt with by re-configuring the system manually. The success of a policy-based system could thus be measured in terms of a reduced number of exceptions and reduced time to the re-configuration. True advantages would further be revealed if it were possible to estimate the impact of operational needs that are not met in a timely fashion. We need to investigate to what extent this baseline can be established, but a qualitative reduction of the number of use-cases that require manual reconfiguration would already be an indication of good performance.

We will compare generative policy-based systems against non-generative ones to evaluate the reduction in complexity of a generative specification (e.g. reduced number of rules) and measure the additional overhead induced by the coordination required between autonomous policy generators (e.g. computational load, time and number of messages). However, we need to investigate how to best measure the gains obtained from increased autonomy and agility in response to changing circumstances.

Applying generative policies to human subjects as well as devices is expected to improve compliance with mission plans, whilst not inhibiting behavior and initiative in unforeseen circumstances. Whilst these gains can be illustrated qualitatively in set use-cases, how to best measure the impact of this flexibility and that the right trade-off between flexibility and security has been achieved will be investigated.

Measures of improvement or performance in specific applications of generative policy-based systems e.g., network or firewall reconfiguration will be considered alongside the more generic metrics and comparisons described above. Within such specific contexts absolute measures of performance such as computational overhead

[29] Elisa Bertino, Barbara Catania, Elena Ferrari, Paolo Perlasca: A logical framework for reasoning about access control models. ACM Trans. Inf. Syst. Secur. 6(1): 71-127 (2003)

and response time, which are traditionally associated with any software application will be made to ensure that they satisfy application requirements.

In addition to identifying the metrics, we will identify the methodologies and algorithms according to which each metric can be evaluated (such as quantitative vs. qualitative). For metrics whose evaluation requires observing human subjects, we will design the experiments that would be required.

Task Milestones:

Date	Description
Q1	− Architecture for the management of generative policies for coalition environments
Q2	− Formal model and syntax for generative policies
Q3	− Policy analytics services
Q4	− Policy conflict management algorithms
Q5	− Impact analysis strategies and algorithms

Task 2.2: Generative Policy based Security and Resource Management

Primary Research Staff	Collaborators
Alan Cullen (BAE)	Christopher Gibson (IBM-UK)
Bill Williams (BAE)	Franck Le (IBM-US)
Brian Rivera (ARL)	Geeth de Mel, (IBM-UK)
Christopher Williams (Dstl)	Maroun Touma (IBM-US)
Dinesh Verma (IBM-US)	Ramya Raghavendra (IBM-US)
Elisa Bertino (Purdue)	
Emil Lupu (Imperial)	

To validate the policy framework and assess the usefulness of our analytics, we need to apply them to different coalition **domains** and evaluate how different policies might interact. It is thus critical that we identify and characterize classes of policies relevant for current and next-generation coalitions. Characterizing interactions among classes of policies is also critical. Performance policies, availability policies, and security policies, for example, can have conflicting consequences when applied to the same set of resources. Capturing these indirect interactions, and resolving the resulting management issues requires investigation, as does the ability to reason about and prioritize conflicting requirements based on the situational context. Techniques for argumentation reasoning or programming with priorities may be applicable for this purpose.

The activities work in Task 2.2 is organized into a number of activities. The first activity focuses on creating a vision for next-generation coalitions, such as coalitions involving robots, intelligent machines, drones, sensors, and actuators, possibly organized into autonomous systems. Based on this vision, categories of policies will be identified. In parallel, however, as part of task T2.2 we will also develop two concrete applications of the policy model and framework: the first will focus on policies for coalition firewalls, the second will focus on an autonomous system scenario. When dealing with next-generation coalitions, characterized by autonomous systems, it is critical that all entities within these systems be self-describing and tightly coupled with policies.

Next-generation Coalitions

In order for our research to be effective for coalition operations, we need to develop generative policies for operations and scenarios which a U.S. UK coalition is likely to encounter in 10-15 years in the future. Targeting that environment provides for adequate time for our research to be useful in the field given the typical time-scale for technology adoption within the militaries of both nations. Towards that goal, one of our first task is to envision the type of coalition operations that are likely to typical in the future. While the specifics would remain uncertain, we can safely assert that future coalition operations will see a large number of unmanned vehicles, both aerial drones as well as ground mules, to assist the soldiers in their operation. Furthermore, dynamic communities of interest will be established among coalition partners which will require the use of assets from two or more countries. In such types of operations, we would need to have policy based control of how these assets can be shared.

Policies could be defined on which mules/drones can be shared for different types of dynamic communities, how new autonomous elements can be automatically included in the group, e.g. when a set of mules in a community find another mule stranded in the middle during their operation, how much information is shared among them, where the mules/drones are allowed to go etc. The goal will be to have comprehensive set of policies which would allow such coalitions to self-govern themselves and dynamically adapt to changes to the mission goals and contexts.

In this use-case, we envision creating policy generators that are provided to the coalitions on how the assets can be used. When a dynamic community of interest is formed, a second level of policy generators are created which are provided to autonomous mules and drones in the community. The mules and drones then use those policies to control the communications, the level of control which they allow to the other coalition partners. One specific goal that we want to illustrate in this use-case is that the mules/drones are able to relax policy constraints and allow enhanced access to a coalition partner in those cases when doing so will be able to better meet the mission objectives of a dynamic community of interest.

Coalition Firewall Policies

As a specific and contained domain to which we can apply our generative policy models, we will consider the generation and enforcement of firewall policies for dynamic coalitions. The members of the coalition that join a particular mission would agree on a set of mission policies. This would involve the specification of high-level network-wide security policies, specified as policy generation grammars, and the generation of new firewall configurations, obtained as policy rules from the policy generation grammars, to bring the system to a policy-compliant state. The policies of the different coalition members would be individually generated to achieve the overall mission security policies according to the policy models formulated in Task 2.1. The mission policy would specify whether or not a specific set of users (roles) should have access to a certain set of services. The specific elements associated with the firewalls protecting the different services would deduce low-level configurations which satisfy the high-level policy, or raise an alarm when the network security policy cannot be satisfied. The latter case would lead to the invocation of analytics to determine how the inconsistency can be resolved, and may lead to a modification of the higher level policy.

Solutions for Self-Describing Resources

A promising direction for the management of complex distributed environments is to make the major elements of the system **self-describing**. We want to create an architecture where policy mechanisms are tightly coupled with the computing elements. Notice that information is critical in such environments in that: (a) the information itself is the critical resource to be protected; (b) the information describes some critical resource. Therefore in such integrated architectures, we can create new models for information assurance, e.g., policies can capture information as data sources are processed in different components of the system, providing a new model for traceability of information, and allowing better provenance on information flows. In order to build this capability, we would use techniques like **functional encryption**. Different keys would then provide different results when used against the same data object enabling not only the control of access to the data by end users but also the use of the data in compositional data aggregation/analytics services. Generative policies at a higher level would determine and automatically generate the right number and set of encryption keys.

Computing and knowledge resources could also be virtualized as autonomous elements. At a high level, the characteristics of the elements of the system (hardware, software, datasets, etc.) would be captured in **Virtual Objects** (VOs) that would represent them within the system. The VOs would track the status of the assets they represent, and would be able to communicate with them and control them to the extent that the actual asset can be externally controlled. **Autonomous Operational Units** (AOUs) would represent collections of hardware and software assets that jointly are capable of performing an operation or process. The AOUs include a management system with cognitive capabilities, and maintain rules, process flows and policies associated with their operational domain. They can be goal driven, discovering VOs and other AOUs that can provide necessary capabilities for carrying out their responsibilities. They can predict the expected behavior of physical assets, and would incorporate learning mechanisms for improving their effectiveness and performance.

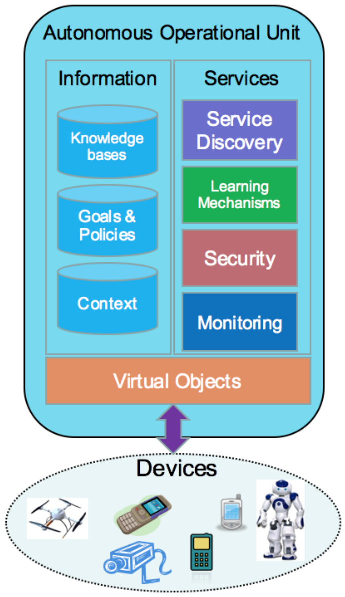


Figure P2-5

The AOUs would maintain relationships with other AOUs, so that their collection forms a (logically) distributed collaborative system. These collections would establish patterns of behavior, provide mechanisms for continuously optimizing the functioning of their associated physical systems, and would enforce complex policies and workflows associated with multiple interacting organizations, e.g., coalitions. Such architectures relate to earlier work on Self-Managed Cells (SMC)[30] and Multi-Agent Systems[31] (MAS).

A specific case of self-describing resources is when the resource is an analytics service as defined in Project P5. We would apply the concept of self-describing resources to services, and show that this enables a more appropriate service architecture for coalition operation. This work will be done in collaboration with P5 team.

Task 2.2 – IPP Activities

During the first project year, the following research activities will be carried out.

1) <u>A vision of next-generation coalitions</u>

In this activity we will start shaping our vision outlined above. We will consider a number of scenarios and identify a number of relevant categories of policies. It is important to mention that our vision will evolve along with our research in the program.

2) <u>Coalition Firewall Policies</u>

In this activity, we will consider a detailed firewall policy system and one side analyze the system to determine whether it is adequate for coalitions or whether extensions may be needed. On the other side, we will develop a specification of the policy grammar for such firewall and derive the firewall policy rules using the policy model developed in Task 2.1. Results from this activity will be critical for the initial assessment of the policy framework and for driving modifications/extensions to the framework for the subsequent years of the project.

3) *<u>Navigation Policies for Drone Swarms</u>*

Autonomous systems represent a challenging domain in which to test our policy framework. They are also very relevant in the context of our vision for next-generation coalitions. Therefore testing our policy framework on such systems is a critical activity. In this activity we will consider, as initial example of autonomous system, the case of navigation systems for swarms of small drones having to carry out collaborative missions (for example for data

[30] N Dulay, E Lupu, M Sloman, J Sventek, N Badr, S Heeps, Self-managed cells for ubiquitous systems

[31] Wooldridge, Michael (2002). An Introduction to MultiAgent Systems. John Wiley & Sons

gathering or formation of ad-hoc communication networks for emergency management). Recent work[32,33,34] has been laying concepts and architectures underpinning the Internet of Drones (IoD) that share characteristics from three different large-scale networks, namely air-traffic control network, cellular network, and Internet. A critical element in such architectures is represented by the navigation system that governs the use of airspace by drones. An airspace can be organized into three main elements: airways playing a similar role to the roads; intersections formed by at least two airways; and nodes which are the points of interest reachable through an alternating sequence of airways and intersections. Each of these three has a geometric shape. Movement of drones inside the airways and intersections is regulated according to policies. An example of a policy is that drones must move only in the designated direction(s) of an airway or intersection.

The airspace is partitioned into zones; each zone consists of a set of airways, intersections and nodes and can be represented as graph. Vertices and edges in such graph are described by attributes; examples of data encoded by these attributes are the minimum performance required from any drone that wishes to travel along the particular element, such as drone range limitations, landing restrictions, and other physical constraints. Attributes may also contain more detailed information about a particular element; for example, the attributes associated with a node representing a park can have a map of the park which a drone could use upon entry to the node. A portion of airspace is either public or private. For private elements, *access* policies for drones can be specified as part of the attributes associated with the elements. However policies concerning exceptions must be also be provisioned allowing for example the traversal of private elements in the case of emergency. In addition, even for public airspace portions one may want to enforce access policies; for example only drones whose owners are certified can be admitted in densely populated areas. We envision that a policy generation grammar for access policies will specify the components of such policies; for example it will specify that such an access policy must include as elements: a specification of drones allowed under the policy; a specification of the parties allowed to fly the drones; a specification of the public spaces covered by the policy. Different parties can then insatiate such policy grammar according to their own local knowledge, constraints, and situations. For example which "spaces" are to be considered public may be different across different geographical regions.

In addition to policies concerning drone navigations, other policies are related to collaborative missions of drone swarms. For example, suppose that a swarm of drones need to collect data from a certain area within a certain time (for example to collect aerial images following a chemical spill in a river). However because of congestions of drone airspace, only one drone can move forward through a certain zone. So the swarms must autonomously decide a strategy for selecting the drone that moves forward (such decision may depend on the drone autonomy, type of equipment carried by the drone and so forth). Notice that such decisions can be based on policies as well as on optimization techniques. It is important to notice here that policies used by the drone swarms to take decisions are dependent also on the navigation policies. For example, the drone swarm may select a drone that then is not allowed to navigate in a zone that has to be traversed in order to reach the destination. In the first year of the project, we will design and develop several such scenarios to test our policy model and framework and identify further requirements. Our initial drone-based scenarios can then be extended to include other types of vehicles, robots, sensors, and actuators.

4) *End-to-end solutions for self-describing resources and services*

We will identify the technologies needed to construct end-to-end solutions for self-describing resources and services. The specific motivating instances that will be the basis of our research studies will be those that are involved in the other activities in Task 2.2, viz., Next Generation Coalitions, Coalition Firewalls, Drone Swarms, etc. Some of the key technologies for distributed, collaborative systems that will be considered include Multi-Agent Systems (MAS), Dynamic Distributed Federated Databases (DDFD) and Distributed State Machines (DSM). The principal aspects that need to be addressed include: discovery – how do Autonomous Operational Units (AOUs) locate each other; recognition – how are AOUs with the requisite capabilities identified; community establishment –

[32] M. Gharibi, R. Boutaba and S. Waslander: Internet of Drones. CoRR abs/1601.01289 (2016)

[33] S. Devasia and A. Lee, "A scalable low-cost-uav traffic network (unet)," arXiv:1601.01952v2, 2016.

[34] R. Hall: An Internet of Drones. IEEE Internet Computing 20(3): 68-73 (2016)

how are teams of agents formed and maintained; and, policy identification – what sets of policies capture the operational requirements of the collaborative systems

Task Milestones:

Date	Description
Q1	− A vision of next-generation coalitions
Q2	− Application of generative policies applied to firewalls for coalition operations
Q3	− Application of generative policies applied to autonomous systems
Q4	− Definition of end-to-end solutions for self-describing resources and services
Q5	− An application of self-describing resources and services to implement a specific analytics scenario for coalition operations

Linkages

Project P2 investigates generative models for policies, which play a key role in coalition operations and thus have a linkage with all the other projects in the program. While the precise linkages will become clearer as the projects evolve, we have identified team members who will be responsible for linkage of Project P2 activities with the activities in the other projects.

Project	Liaison Responsible
P1	Franck Le (IBM-US)
P3	Christopher Gibson(IBM-UK)
P4	Diane Felmlee (PSU)
P5	Geeth de Mel (IBM-UK)
P6	Supriyo Chakraborty (IBM-US)

During the first year, the linkage activities will focus on:

1) identify relevant categories of policies arising in the other projects;
2) identify initial requirements from the other project concerning: (a) the specification of policies; (b) the enforcement of policies; (c) the evolution of policies; (d) the human involvement in policies.

Results from the above linkage activities will lead to revision and extension of the policy framework developed by Task 2.1 and application of the framework to specific scenarios in Task 2.2.

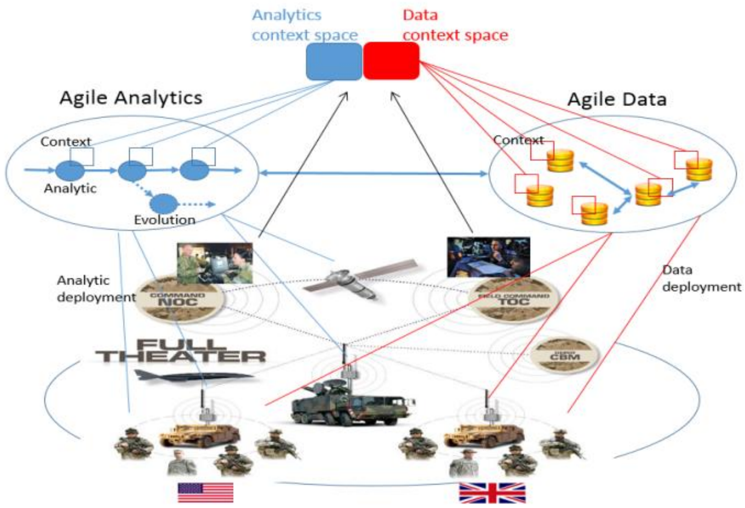
Project P3: Agile Composition for Coalition Environments

Project Champion: Lixia Zhang, University of California, LA	
Primary Research Staff	**Collaborators**
Bongjun Ko (IBM-US)	Ananthram Swami (ARL)
Christopher Gibson (IBM-UK)	Andreas Martens (IBM-UK)
Kevin Chan (ARL)	Christopher Williams (Dstl)
Kin Leung (Imperial)	Don Towsley (UMass)
Liang Ma (IBM-US)	Pablo Bermell-Garcia (Airbus)
Lixia Zhang (UCLA)	Shiqiang Wang (IBM-US)
Mark Herbster (UCL)	Theodoros Salonidis (IBM-US)
Michael McDowell (Airbus)	Ting He (PSU)
Prithwish Basu (Raytheon – BBN)	
Thomas La Porta (PSU)	

The vision of distributed coalition intelligence requires that actionable insights be derived through various analytics tasks using services and resources available from different partners and networks that comprise the coalition infrastructure. At the same time, constrained infrastructure with limited bandwidth, power and mobility introduces several fundamental challenges in composing services dynamically. Frequent disruptions (e.g., as due to mobility and the need for operational agility) render pre-planned analytic service deployments infeasible.

Data sources for analytics services are often distributed across the network and the data can be large in size. Finding relevant data sources is particularly difficult in coalition networks where such sources may be spread across coalition partners and subject to access policy constraints. There is a pressing need for a unified computing and networking platform that can provide an *efficient*, *scalable*, *reliable*, and *secure* means to dynamically compose, position, and execute analytics services in coalition environments. In order to enable new principles to create agile composition of services, we propose the radical concept that analytics and data should both be treated equally in a novel architecture for tactical coalition operations where both analytics and data possess the same attributes of replication, division, movement and composition. We refer to data and analytics possessing those attributes as *agile data* and *agile analytics*. When data and analytics can both move, they will result in a virtual information flow– a distributed environment in which data and analytics can find each other, with each of them moving to the other or an intermediate infrastructure component as needed. Our approach is to provide names that can represent structures and semantics of both analytics and data, and develop extensions to the Content Based Network (CBN) paradigm to enable data and analytics to discover each other and execute in an efficient manner. This fundamentally new architecture enables finding information, analytics and supporting resources and composing them in a dynamically changing network *jointly and optimally*.

The new agile analytics and data paradigm provides a fundamentally different approach to composing services. In current research, services are considered as static objects that are physically tied to a node. The analytics services considered in this project consist of smaller services components, which are mapped to the underlying tactical infrastructures. This aspect introduces high complexity to the analytic service composition problem because, in addition to composing a service at the logical level, the physical network locations of the tactical network where the services and data may move need to be taken into account. In addition, composition of services and thus execution of the analytics applications become constrained by the need to transmit information among those smaller services, and the composed services get disrupted if network connectivity is disrupted, a phenomenon that happens all too often at the tactical theater in coalition operations. On the other hand, if services can migrate freely to the location that is best suited for the right combination of data and analytics, new types of services can be readily



composed by moving the data and analytics elements to the best location eliminating the strong dependency on network connectivity by the composed service.

In addition to discovering the algorithms and architectural mechanisms that allow us to build this new architecture, we want to understand the performance limits of this new architecture and understand how to build in self-optimization as an implicit aspect of this architecture.

In this IPP, we will develop principles, theories, and fundamental architectures of this concept in the following two inter-dependent tasks:

- *Task 3.1: Fundamental Limits and Models for Agile Code and Agile Data*: We will determine the fundamental performance limits and security characteristics associated with the concept of agile code and agile data.

- *Task 3.2: Adaptive Mechanisms for Agile Code and Agile Data in Coalition*: We will investigate novel architectures and new algorithms that can leverage agile code and data to dynamically adapt analytics processing in a tactical coalition environment as its configuration changes.

Task 3.1: Fundamental Limits and Models for Agile Code and Agile Data

Primary Research Staff	Collaborators
Bongjun Ko (IBM-US)	Ananthram Swami (ARL)
Kevin Chan (ARL)	Andreas Martens (IBM-UK)
Kin Leung (Imperial)	Christopher Williams (Dstl)
Prithwish Basu (Raytheon – BBN)	Don Towsley (UMass)
Thomas La Porta (PSU)	Shiqiang Wang (IBM-US)
	Ting He (PSU)

Within this task, we will undertake two concurrent activities exploring alternative approaches for modeling a system with agile code and agile data. In the first activity, we model the system using the analog of potential energy from Physics, while in the second activity, we model the system using algorithms and approaches that will be developed in project P1.The second activity provides a common linkage between this project and P1.

Potential Metrics for Composition of Data and Analytics

The new proposed architecture allows analytic services and data to migrate freely to the most suitable location for the required combination of data and analytics, given the network state and workload of the tactical coalition network. In order for the composition to happen seamlessly, we need models and algorithms to characterize and determine which analytics tasks should flow to data and when, and which and should flow to analytics tasks when data, i.e., to determine the right positioning of data and analytics. The right answer will depend on the nature of the analytics and the data that is available to be processed by the analytics, as well as the resources (CPU, network, storage) on which the analytics functions are to be executed. Our initial work will investigate fundamental models that capture the core functionality of a large class of analytics applications.

In order to determine the positioning of data and analytics, we propose two related concepts – potential Φ and distance ρ. This task will focus on developing abstract definitions of Φ and ρ, and algorithms to determine the optimal locations of the data and services.

The distance, ρ, is an abstraction, which characterizes the difference between a desired service and an available service, or the degree of match between a service and data. The distance between a desired service and an available service is based on a semantic difference of the result a service can provide and what is requested. For example, the distance between a request for a weather forecast and a climate report is larger than the same request and the daily average for that location, which is larger than a real-time forecasts. Distance can also be applied to multiple services that must be composed into a larger service. Joint work will be carried out with Project 5 to determine a precise definition of the distance. At a high level, a distance metric between a data entity and an analytics entity is determined by several factors, such as the compatibility between the two entities, the fitness of the data to the desired result (i.e., the "mission-fitness"), the amount of computation that is involved in the analytics at any location, etc. The definition of distance provides a way for any entity to measure how "far" another entity is in a logical manner. This distance metric although initially constructed by the researchers, will be refined by the data and the services through metric learning[35,36] using approaches developed in task 3.2.

We consider two types of entities that can migrate – services that must be composed to form a larger service may migrate to be close to or be co-located with each other, or data may migrate to be closer to a service that requires it. If the entities can be co-located, then their operation becomes robust against network connectivity with the exception of connectivity to the consumer of the service. The definition of the potential Φ associates with a location or entity, and the tendency to migrate towards (or away from) that location or entity, i.e., agile entities move from a location with a higher potential to a location with a lower potential. In particular, the potential Φ is related to the system performance such as network load, local storage utilization, and latency of processing; it can be seen as the "cost" of operating the system under the current condition. The value of Φ also depends on aspects of system performance relating to interoperability including heterogeneous networking and tactical coalition operations. For example, the potential may be higher at coalition boundaries where the network connection becomes more disruptive; the potential may also be higher when one coalition partner's task is running on the compute device of another coalition partner (due to possible security risks). Algorithms for making migration decisions will be based on these potential and distance metrics. We will focus on a solution that utilizes a limited capability of mobility prediction, where migration occurs pro-actively when possible and reactively when necessary. We will also explore the use of "controlled mobility" assets provided by the SDC slice infrastructure described in Project 1. The placement and migration algorithms based on the potential and distance metrics are also highly relevant to the self-optimized use of resources in the defined SDC infrastructure slice in Project 1. We will jointly carry out the investigation with Project 1 to maximize the synergy between the two projects, where the focus in Project 3 here is on the data and analytics for services, and Project 1 will provide the fitness measures of the underlying resources in the SDC infrastructure.

Our initial research will first define mathematical "gravity laws", which are based on force vectors[37]. Such laws will differ from Newtonian gravity laws; we borrow similar concepts and apply them to networked systems for the dynamic configuration of analytics task and data placement. The sum of force vectors results in the potential acting on the entity. These vectors can represent attractive forces or repulsive forces. In addition, we define an impedance factor. Attractive forces can be exerted by users that consume the service (a service will tend to move towards its consumer), among services that must collaborate, as well as among services that require a certain type of data (we may also model data as having a force). The strength of the force will be a function of the distance, ρ, between the service and the desired service, and the data and the desired data, as described above. The smaller the distance between what is desired and what is offered, the stronger the attractive force will be. In this way, suitably matched services and data will tend to be co-located. The sum of all attractive forces at a service will also be related

[35] Crammer, K., & Gentile, C. (2013). Multiclass classification with bandit feedback using adaptive regularization. Machine learning, 90(3), 347-383.

[36] Basu, P., Krishnan, R., & Brown, D. W. (2008, November). Persistent delivery with deferred binding to descriptively named destinations. In Military Communications Conference, 2008. MILCOM 2008. IEEE (pp. 1-8).

[37] Wang, G., Cao, G., & La Porta, T. (2006). "Movement-assisted sensor deployment," Mobile Computing, IEEE Transactions on, 5(6), 640-652

to the amount of service demands. It will be stronger if a service is in high demand. Note that, in practice, some randomization on the attractive forces may need to be introduced in order to avoid deadlock or livelock situations resulting from more than one, equally attractive forces or the uncertainty of the physical metrics representing the forces in dynamic, coalition networks.

Repulsive forces will be exerted by similar services, servers that have a low residual capacity, or parts of the network that are congested. These forces will tend to spread similar services across the network, load balance both storage of data and processing across nodes, and distribute traffic load. The total of forces of the vectors from all the users requesting a service, all services requesting data, and current network state (position of data, services and traffic load) will move the services and data to some stable and sustainable points (Like in the case of attractive forces, however, one needs to be careful not to fall into deadlock or livelock points that can result from repulsive forces exerted by multiple services). In addition, strong forces on an object (analytics or data) from multiple directions can result in the object being split or replicated; a chunk of data can be partitioned into multiple pieces, each of which can be processed by multiple analytics tasks, and an analytics task can be replicated and executed at the same time to process distributed set of data sources. In such a case, it will be appropriate to apply repulsive forces to the clones of the object in order to achieve to achieve a well-balanced distribution of the cloned objects throughout the network.

In addition to attractive and repulsive forces, we will model the impedance to movement, which will retard movement or limit how far elements may move. The impedance will be a function of the degree of difficulty to migrate the element, which will be composed of factors such as network connectivity, path length, network congestion, overhead involved in the migration and preferences in the coalition policy. Incorporating the impedance factor will allow an element to tend to move along the path with lower impedance, while hard impedance, e.g., a roadblock, may result from a very restrictive policy present in the coalition environment.

Our hypothesis is that defining the right potential and learning the right distance metric in any distributed environment provides a common framework to understand and analyze a variety of problems associated with the agile analytics architecture. The framework will describe the relationship and interactions among the different forces and the potential and distance metrics. In this framework, the dynamics of data and analytics entities can be handled naturally since it would warp the potential field causing new entities to be discovered and moved appropriately until equilibrium is achieved. This naturally leads to a control mechanism for dynamic analytics and data placement/migration based on time-varying demands and system conditions. In addition to migration, the findings in this activity can be further extended to include options of replicating and deleting services, and we will explore the possibility and ways of such extension. We will also explore pro-active migration in which analytic services and data may move even in the absence of service requests. This may be beneficial to allow analytics that typically collaborate, or new data that may be used by a service, to migrate when network load is low in advance of anticipated requests. Of particular interest is how these energy models, the corresponding potential fields, and the system metrics can capture the constraints across coalition boundaries, which can quite possibly introduce non-linearity in the models due to the coalition policies and network heterogeneity. Hence, it is important to develop a deeper understanding of the characteristics of these metrics and validate their suitability in coalition scenarios.

Furthermore, it is worth noting that the notion of potential and distance can be readily incorporated as parts of the utility definitions in determining movement, replication and split of data and analytics. As a result, the distributed utility maximization techniques to be devised on Project 1 will be applicable to the CBN agile architecture. In addition, it is likely that in coalition environments, only partial (and perhaps corrupted) information is available to support the optimization. It is also possible that the derived potential and distance functions have an unfavorable shape for performing distributed optimization in some part of the system. Joint work will be carried out with Project 1 to study how to address these challenges.

A key research challenge here is to efficiently manage the information related to the temporally evolving attractive, repulsive, and impedance potential fields as well as the aforementioned distance metrics in the agile code and agile data architecture to be studied in Task 3.2. In this activity, we will explore novel methods for intelligently aggregating namespaces, wherever feasible, to make the process of discovery between analytics and data objects more efficient.

To validate our approach to developing the potential metrics, distance measure, and their relevance in mapping to physical metrics, we aim to apply these measures to the mechanisms for the placement of the analytics and data to be studied in Task 3.2. The fitness and appropriateness of our metric definitions will be assessed and characterized in terms of how concise and effective our metrics are when it comes to making the decision of moving and placing the analytics and data objects in the optimization tasks, compared to traditional approaches of using raw physical metrics. In particular, we expect the combination of physical distance and semantic distance to play an important role for mission-fit placement and composition of services and data. For this, we will jointly work with Project 5 and Task 3.2 to assess the efficacy and effectiveness of such combination through data-driven analysis and simulation of the placement tasks.

Fundamental Limits of Analytics Capacity and Performance

It is important to understand the capacity limits of this new CBN-enabled agile architecture, when it will perform better than traditional analytics architectures, and when it will not. In a traditional architecture, system capacity corresponds to network throughput. However, for the new agile data and agile analytics paradigm, the simple notion of capacity is no longer adequate. Toward that end, we propose the notion of Analytics Capacity that characterizes the maximal task processing rates that can be supported considering all possible ways of optimizing service composition and positioning under multi-dimensional resource (e.g., compute, network, and storage) and coalition specific constraints.

Capacity of the agile analytics architecture is not the amount of data that can be transferred, but the volume of useful analytics results that can be generated in a given unit of time. To compute this measure, we need to introduce the characteristics of the analytics operations and application workloads into the capacity equation and examine which configurations of agile networked systems have the capacity to run a given application split in the agile manner. Furthermore, the analytics capacity can be increased by leveraging the multiplicity of common intermediate components across different applications to eliminate redundant computation results (and even intermediate “transient states”) and by trading off the overall utility for usage of resources, including those used for moving the analytics codes and data, for executing the analytics tasks and updating the states of their execution over distributed entities in the network. Correspondingly, our capacity model will consider these attributes of the new architecture. In addition, the optimal service composition and positioning will be determined by maximizing the appropriately defined utility via the distributed optimization techniques to be developed on Project 1. Moreover, CBN-based approaches will exploit caching of named data, analytics functions and intermediate results, and thus could potentially improve the analytics capacity significantly compared to the case where no CBN facilities are available.

Further, the impact of the overhead, in terms of extra bandwidth, processing and storage consumption, resulting from movement, replication and processing agile data and analytics on the analytics capacity of the new architecture will need to be considered. Note that since the underlying resource availability on the SDC slice (as reported by techniques to be developed on Project 1) is expected to vary dynamically over time, the re-mapping of logical analytics and data objects to these physical resources will also result in signaling overhead and the definition of Analytics Capacity will need to incorporate such overhead into consideration. Metrics comparing analytics capacity to physical capacity of the network (e.g., measured using metrics like max-flow) can provide an estimate of the overhead associated with the approach.

Our approach towards finding the Analytics Capacity will start with defining a generic mathematical formulation that captures the characteristics of a large class of analytics applications. This formulation will incorporate the relationship among the amount of available data, how the data is distributed (at different compute entities, among different coalition partners), the network communication overhead, the computational demand of analytics tasks and workflows, and the accuracy of analytics result. This relationship can be built based on the algorithms and performance guarantees for solving distributed empirical risk minimization (ERM) problems[38] which is the key building block of many analytics applications. We will extend known results by building new theoretical models that include cases where the data is non-uniformly distributed across different computing entities. Performance bounds on the relationship between data distribution and analytics accuracy will be derived. This will then be integrated with the communication and task processing capacity in distributed systems, where the latter can be a “max-flow, min-cut” type capacity metric, to conclude a general metric for analytics capacity. These metrics

[38] Zhang, Y., Xiao, L. (2015). DiSCO: Distributed optimization for self-concordant empirical loss. in Proceedings of the 32th International Conference on Machine Learning, pp. 362-370, 2015.

would consider not just the input data locations, but would also need to incorporate data/code replication as well as controlled mobility phenomena. We will incorporate constraints related to heterogeneity and tactical coalition operations in the capacity formulation.

The performance limits and the optimality of the approaches for self-organizing composition resulting from the concepts of potential and kinetic energy as described earlier also need to be analyzed. While the energy based approaches provide an innovative heuristic that can be used in a practical architecture, we need to determine how close these heuristics can come to the optimal organization of code and analytics under some set of simplifying assumptions that allow analytical modeling. We note that the concepts of potential function and potential minimizing algorithms for dynamic analytic service composition are similar in spirit to the concepts of Lyapunov function and Lyapunov drift minimization used in the mathematical techniques of stochastic optimization, which enables the design of capacity-achieving algorithms for the agile architecture. Our research will explore this connection and establish the performance bounds of the corresponding approaches.

Task Milestones:

Date	Description
Q1	− Study the key elements that need to be included in the potential and distance metrics and analytics capacity formulation, by considering analytics applications in distributed coalition operation scenarios.
Q2	− Initial definition of distance (measure between perceived desired service and available service; joint with P5.1) − Initial analytical model for capturing the relationship between analytics accuracy and data availability
Q3	− Initial definition of potential (tendency to migrate towards a point) − Enhanced analytical model on analytics accuracy under different data distributions and availability
Q4	− Framework that captures the relationship among distance, potential, and different types of forces − Incorporate communication and task processing capacity in the model, propose initial notion of analytics capacity
Q5	− Algorithm that uses distance, potential, and forces for migrating services, including use of distributed optimization framework − Propose analytics capacity considering all system elements, including constraints related to heterogeneity and tactical coalition operations.

Task 3.2: Adaptive Mechanisms for Agile Code and Agile Data in Coalition

Primary Research Staff	Collaborators
Bongjun Ko (IBM-US)	Ananthram Swami (ARL)
Christopher Gibson (IBM-UK)	Christopher Williams (Dstl)
Kevin Chan (ARL)	Kin Leung (Imperial)
Liang Ma (IBM-US)	Pablo Bermell-Garcia (Airbus)
Lixia Zhang (UCLA)	Prithwish Basu (Raytheon – BBN)
Mark Herbster (UCL)	Theodoros Salonidis (IBM-US)
Michael McDowell (Airbus)	Thomas La Porta (PSU)

In order to understand the principles for agile code and agile data, we will undertake two concurrent activities in the task. The first activity is to understand how the emerging paradigm of content based networking can be used to create adaptation mechanisms, and the second activity tries to find performance limits for such adaptation by drawing an analog from the concept of kinetic energy in Physics.

Content-based Discovery, Distribution, and Execution of Codes and Data in Coalition

Content-Based networking (CBN), also known as Information-Centric networking (ICN), is a paradigm in which data delivery in the network is done based on the *name* of each piece of the desired data rather than its location as indicated by the IP address of the data host[39]. A unique novelty of CBN is that *both* content and analytics can be treated as named data pieces, this enables us to use the named data networking model not just for data, but also for analytics, making them both independent of location. Therefore, we can identify data as well as analytics by names, and move them to appropriate locations for analysis or execution, respectively. Because of the decoupling between data/analytics from their locations, the CBN paradigm can be adopted for a variety of applications. In particular, CBN enables distributed discovery by sending queries with the desired data (or analytics), and enables delivery and reuse of distributed datasets of relevance to particular analytics tasks. Similarly, CBN can also facilitate the networking substrate to support a variety of new primitives besides simple forwarding, e.g., replication, split, combination, and movement of both data and functional tasks at execution time to enable distributed analytics.

The power of CBN lies on routing and forwarding based on content/analytics names, realizing our aforementioned vision of potential-theoretic distributed service composition. The problem of distributed composition, positioning, and management of data can be significantly simplified by the adoption of a CBN approach. For example, in order to enable semantic discovery of data for analytics, one can systematically develop naming conventions to encode the content semantics into names, so that the network forwarding decisions can take into account a variety of semantic and structural properties of data, including transient data that are results of partially-completed computation. Such encodings of the semantic properties of data into “routable” names becomes especially useful when dealing with unstructured, ephemeral data such as audio, video, and other time-series data generated by various sensors, such that they can also be adequately discovered, located, and consumed within CBN frameworks, along with other data sets that are named after traditional naming conventions. The key challenges here are (a) how to automatically discover and describe such semantic properties of data, and (b) how to design a name space in which such expressive description can be adequately encoded into names that can be handled efficiently in CBN context. For the first issue of semantic data description, we will collaborate with Project 5, in particular with the task investigating self-description of the data and analytics, while this task will focus on the name space design issues for such semantically described data.

In tactical coalition environments, a critical factor that enables a CBN to fetch data and analytics from anywhere is the mandatory requirement of securing every single piece of data: every piece of data must be cryptographically signed, and encrypted as necessary. This requires a well established certificate system for data signing, verification, and key distribution. Furthermore, the coalition of different administrative network domains requires proper resolution between different CBN namespaces, or even between a CBN domain and non-CBN domain, so that named objects in one domain can be seamlessly translated into those in another. Such resolution and security requirements can be governed by the policies of the coalition, on which we will collaborate with Project P2.

In the context of moving both the data and analytics, CBN methods can also be used to discover not only richly described data but also other analytics entities (i.e., the instances of analytics tasks) based on their richly described roles in the computation task, e.g., Intentional Naming[40]. Note that such semantic roles of the analytics entities can be effectively described by the corresponding meta-description of the data they consume (input data) and produce (output data). Complex analytics tasks can be further expressed in terms of logical dependency graphs consisting of not only nodes denoting subtasks with named attributes, but also those denoting named data objects

[39] Meisel, M., Pappas, V., & Zhang, L. (2010, September). Ad hoc networking via named data. In Proceedings of the fifth ACM international workshop on Mobility in the evolving internet architecture (pp. 3-8).

[40] Basu, P., Krishnan, R., & Brown, D. W. (2008, November). Persistent delivery with deferred binding to descriptively named destinations. In Military Communications Conference, 2008. MILCOM 2008. IEEE (pp. 1-8).

produced or consumed by these subtasks. Such description of the analytics tasks shall be encoded into names to facilitate distributed routing and forwarding of data and analytics within CBN framework. Once the names of data and analytics are determined, they can be discovered and routed by one another via CBN's routing methods. The routing decision shall be based on (a) the distance metrics between data and analytics and between analytics sub-tasks, which is investigated in task 3.1 (in collaboration with Project 5), (b) the availability and 'fitness' of the resources to execute the analytics on, which can be obtained via control-plane of the physical SDC slices investigated in Project 1, and (c) the algorithms that determine the movement of the analytics code and data.

Through proper naming, CBN can also enable network management to obtain efficient operations for coalition networks. We can define different pieces of information available for network management in a name space that reflects their availability. By using intelligent replication models[41], we can create a virtual information or analytics overlay that can provide a stable way to retrieve network management data from replicated copies of the information, even if the primary source of management data is not available. This allows us to reduce the network management problem to be a special case of the general agile analytics architecture defined above. A particular application of this approach is to use CBN framework for manage distributed resources (CPU, memory, and network) in Software-Defined Coalition (SDC) slices which will be investigated in Project 1. We will seek for opportunities to apply CBN architecture in distributed SDC control-plane mechanisms through collaboration with P1 researchers.

In this IPP, we will focus our efforts on laying the foundation for realizing the concepts of agile code and agile data by developing a new CBN architecture that enables the movement of the analytics and data in a distributed fashion in tactical coalition environments. As part of the architecture development activity, we aim to specify the architecture in more detail, by (i) **designing the name space for data and analytics functions**, (ii) **develop methods for cross-domain data security and name space resolution**, (iii) **define the right primitives to realize the concepts of agile code and agile data**, and (iv) **validate its performance by means of simulation as well as modeling and analysis**. We plan to design the architecture in a modular way so that different variations for adapting this architecture by using the concepts of potential and kinetic energy can be supported seamlessly.

Our vision for agile code and agile data is that this concept can effectively and efficiently address the dynamic and austere nature of tactical coalition environments. In essence, treating both the data and analytics tasks as movable, replicable, and divisible components as needed would allow for dynamic scheduling and resource allocation for them in a much finer granularity under the dynamics of the physical systems than existing approaches, and CBN-based architecture will provide the necessary platform that facilitates such adaptive mechanisms. The challenge for the CBN framework, however, is to develop CBN architectures that will be able to cope with system dynamics resulting from rapid changes in the locations of network nodes, resource availability, and the location of the data and analytics. In particular, since keeping track of these changes may incur excessive communication overheads (and hence energy and bandwidth resources on mobile devices at the edge), one needs to carefully consider the tradeoff between the ability to distributed their accurate states and the cost associated with it. In the IPP, we will explore this tradeoff space by investigating how the current CBN approaches can handle these dynamics and identifying new approaches based on the metrics definitions that will be established in this IPP.

Another interesting aspect in the design of the name spaces for CBN is that a semantically aware CBN framework presents a rich space of tradeoffs among expressiveness of names, information leakage that may be caused by expressive semantics, and scalability of distributed service composition. As such, it is important to develop theoretical models for these tradeoffs and also analytic and algorithmic methods to control information leakage across coalitions in a tunable manner to enable "need to know" data sharing, as well as novel namespaces that can express data and analytics functions across a broad range of the spectrum, from a detailed description of functions to aggregate/obfuscated descriptions. Here, scalability at a desired level of information leakage and expressiveness will be achieved by eliminating the flow of redundant data by enabling appropriate data transformations at the right places in the network by means of intelligent analytics elements. The aforementioned activities in the IPP will establish the foundational models and system architecture that help us attain these long term visions.

[41] Ko, B. J., & Rubenstein, D. (2005). Distributed self-stabilizing placement of replicated resources in emerging networks. IEEE/ACM Transactions on Networking (TON), 13(3), 476-487

Adaptive Scheduling and Resource Allocation through Kinetic Energy

Coalition networks are highly dynamic and frequently run into issues such as network fragmentation, congestion, lack of computing resources, energy depletion, data protection, etc. In our proposed architecture of agile analytics and data, analytics and data can be moved and replicated as needed. Similarly, data can be partitioned into different subsets and the same analytics run on different subsets of the data and the results combined together. All of these possibilities correspond to the same analytics function, which will yield the same functional result that can be realized by many different variations. Each variation of an analytics function is a different instantiation of the same logical operation, but with a different possible set of tasks and sub-tasks (e.g., permutation of order, split-and-merge, replication, aggregation, etc.). Each of these variations will be a viable option for composing the analytics result. The challenge will be to determine the variation of the analytics function that optimizes system performance metrics such as throughput, delay and/or quality of the analytics result, given the current physical network state in a dynamic manner.

Just like the potential energy and distance metrics developed in Task 3.1., which enable moving analytics and data towards each other, we introduce the concept of *kinetic energy* to capture the different variations that can be used with an analytics function. The kinetic energy of the system is a metric of application behavior during system operation, which is derived from various resource usage profiles that will be used in a given configuration of the analytic system, (i.e., a function of Key Performance Indicator (KPI) resource utilization metrics such as CPU, memory, disk, network bandwidth and energy expenditure). A kinetic energy metric thus captures various aspects of resource usage during system operation. A higher kinetic energy corresponds to higher quantity or higher speed of resource usage. It may also correspond to a higher momentum in terms of overhead that will be required during application reconfiguration. Different variations of the analytics function, each of which results in the same functional output, will correspond to different levels of kinetic energy since the variations use different resources and result in different measured KPIs.

In systems that obey Newtonian laws of physics, kinetic energy and potential energy are conserved and can be converted into each other. However, in the world of computation and analytics, we do not expect such a conservation law to hold in general. The total energy of the system, which includes its kinetic energy and potential energy, however need to be minimized, even if the conservation laws do not hold. We propose a scheduling framework, which estimates appropriate total energy of the system in different analytic application variations and optimizes system performance by selecting an optimal analytic function variation and configuration. The task of the scheduler is to continuously monitor the current kinetic energy and potential energy and distance metric of the current configuration of data and analytics application. The scheduler will also estimate the kinetic energy that corresponds to the different variations of the same analytics function. We will develop estimation and resource allocation algorithms for the scheduler, both assuming a hypothetical centralized scheduler with perfect information, and subsequently more realistic scheduler implementations that operate with incomplete state and in a decentralized manner.

A first challenge will be to define kinetic energy definitions suitable for the analytic tasks encountered in coalition environments. Our first area of investigation will be identifying the nature and requirements of analytic applications encountered in coalition network environments. Subsequently, we will seek to define the right kinetic energy functions, along with the appropriate distance and potential functions developed in task 3.1, based on the structure of the analytics under consideration. We will experiment with different definitions of the kinetic energy in terms of different KPIs and topologies of analytics tasks/sub-tasks, and determine the efficacy of this approach to automatically reconfigure a system to its best possible organization. In the end, we strive to develop a model of the kinetic energy function that can capture multiple system metrics into a scalar value (or at least into a lower dimensional space), such that it can be efficiently and effectively utilized in the system optimization for placing analytics and data.

A second challenge will be the design of mechanisms that accurately model application behavior. By building a model of application behavior with respect to a set of KPIs such as CPU utilization, memory utilization, network bandwidth utilization, network location etc., we will seek to predict how that application and its variations will behave when placed on the network with either known, or predicted environment parameters. The models built can be discrete and/or continuous depending on the requirements of the system, in terms of how accurate that application profiling needs to be. Both model types can be learnt by example and using machine learning approaches. At many points in the network this is a relatively well-constrained problem with known issues, but towards the network edge that kind of certainty diminishes, since, in a dynamic network setting not all KPIs are expected to be known at all times. Thus the models will need to predict some unknown KPIs based on KPIs that are measured in “similar” circumstances previously observed, for which various machine learning models (unsupervised, supervise, or semi-supervised) can be applied to establish the most proper measure of such similarity. These models should allow efficient, dynamic updates to account for new and previously unseen configurations.

Toward the edge of the network in a live operational environment, it becomes increasingly difficult to determine the analytic application’s behavior through direct/indirect querying of resources due to incomplete knowledge of the measured network state. Under such conditions, we will develop heuristic mechanisms that yield approximate near-optimal solutions. Such heuristics will seek to provide near optimal estimation of the kinetic and/or potential energy present in the system. Since the application that we wish to estimate kinetic energy from is non-stationary with respect to the network, the resources and the environment that it operates in, any models, feature extractions, tuning parameters or learned near-optimal behavior will need reassessing and retraining if it is to provide an accurate means to model application behavior. Heuristics based upon incremental learning where the heuristic is updated when a large enough change is seen within the system that it is going to estimate from, are likely to provide both speed and accuracy benefits over other more static approaches. This ensures that the learned model remains relevant to the system, allowing for a more accurate and timely estimation. The challenge is that the system that is being modeled is highly dynamic across many disparate variables. In addition to environmental conditions, dynamism exists due to strategic changes being made to optimize the network infrastructure. Taking input from the other elements of this project that are constantly assessing the network for its analytics capacity and adjudging when changes should be made will also be taken into account as a driver to remodeling the system in case of failure and thus enabling timely and accurate estimations. Effectively managing a machine learning approach across such a changing environment, while still producing timely and relevant outputs, will be a challenge addressed by this project. Where possible, active learning can be implemented to improve on the learning phase in terms of accuracy and time spent collecting and learning from data. To improve upon the regular implementation of this method we will initiate it through supervised learning, but when in operation will become an unsupervised learning approach. The supervised learning will establish the boundaries of expected operation of the system, which the unsupervised approach will be constrained by. The unsupervised approach will then constantly assess the network and determine when to retrain and to what level of detail and updating the model when appropriate. This approach will require a much more agile network model to be developed that can work in unison with its ever changing environment that it is a model of. The output will be a rapid means for estimating application behavior accurately and in a timely manner.

The estimated application behavior profile, summarized through a metric such as kinetic energy can be used by various resource allocation mechanisms that aim to optimize system performance. We will investigate how to combine application behavior metrics such as kinetic energy with potential energy and distance metrics developed in Task 3.1., toward incorporation in resource allocation mechanisms that optimize for energy metrics. For example, when the scheduler finds a variation where the reduction in the kinetic energy of the system is higher than the difference in potential energy that results from a reconfiguration of the system (which is a measure of the effort required to move between the different configurations), the system can move over to the new variation. Such a decision can be formally captured and optimized using a reinforcement learning approach. Another way of combining energy metrics is to define aggregate energy metrics, such as one that captures the trade-off in resource usage (kinetic energy metric) and reliability through replication of compute and data resources (potential metric). Thus, in addition to reactive resource allocation mechanisms, accurate modeling of application behavior will enable the development of proactive resource allocation mechanisms. For example, an application variation with minimum predicted aggregated energy metric reflecting reliability can be selected and deployed in order to minimize the risk of application unavailability when demanded. Finally, we will also explore the relationship of such internal energy metrics with external metrics such as throughput, delay, reliability or quality of computation result using analytic, simulation and machine learning approaches.

Task Milestones:

Date	Description
Q1	− Investigate the requirements for realizing the concepts of agile data and agile code based on CBN, and find the linkages to other projects and tasks. − Investigate key elements and requirements for developing learning-based data and analytics placement at tactical coalition edges.

Q2	− Analyze key elements in name space design and distance measures for named objects in CBN coalition. − Establish models and metrics for kinetic energy for characterizing application behavior under different variations (manifested by replication, splitting and reordering of application components) and measurements of system KPIs such as CPU utilization, memory, network bandwidth and energy consumption.
Q3	− Develop name space models for data and analytics functions that are amenable for cross-domain CBN-based discovery and routing in coalition environments. − Establish lightweight and accurate estimation approaches for modeling application behavior profiles based on measured KPIs using machine learning techniques.
Q4	− Develop the essential set of primitives in CBN for moving, splitting, and replicating the data and analytics based on name resolution and potential energy metrics. − Develop incremental near-optimal application behavior profiling approaches that operate under uncertainty on system KPI and network state using online machine learning techniques.
Q5	− Analyze properties (stability and convergence) of the CBN architecture for agile code and agile analytics in dynamic coalition network through modeling and experiments. − Develop models that combine application behavior metrics, such as kinetic energy with potential energy, and distance metrics toward incorporation in resource allocation and optimization mechanisms. − Establish the relationship of energy metrics with external metrics such as throughput, delay, reliability or quality of computation result through analytic, simulation or machine learning approaches.

Linkages

Project P1 is investigating several algorithms that will provide insights into the fundamental limits of software defined coalitions. These algorithms can be useful for applying in analyzing fundamental limits of analytics in the context of agile code and agile data. The linkage between these algorithms, and their applications to problems within the scope of the program will be provided by Kin Leung (Imperial), who is working on both of the projects.

There is obvious linkage of P3 activities to Project P2 which is looking at policy based management. Linkage between these two projects will be provided by Christopher Gibson (IBM-UK).

Project P3 have several synergies with Project P5, which is exploring approaches for self-organization in services. Linkages between P3 and P5 will be provided by common researchers working across the two projects, including Bongjun Ko (IBM-US) and Thomas La Porta (PSU). P3 team will work with task 2 of Project 5 which addresses resource assignment for instantiating services, and re-provisioning based on changing network state. More specifically, we will provide them with the means to use our metrics in driving decisions of re-provisioning.

Project P6 is looking at approaches for anticipatory situational understanding. Mark Herbster (UCL) will collaborate with Simon Julier (UCL) of Project P6 to ensure that both teams are appraised of progress in different projects. Further linkages will be provided by Pablo Bemell-Garcia (Airbus) who is part of both projects.

Project P4: Evolution of Complex Adaptive Human Systems

Project Champion: Diane Felmlee, Pennsylvania State University	
Primary Research Staff	**Collaborators**
David Mott (IBM-UK)	Alun Preece (Cardiff)
Dean Clarke (Airbus)	Cheryl Giammanco (ARL)
Diane Felmlee (PSU)	Dinesh Verma (IBM-US)
Jared Somerville (Airbus)	Gavin Pearson (Dstl)
Leandros Tassiulas (Yale)	Grace-Rose Williams (Dstl)
Nicholas Christakis (Yale)	Sanaz Yeganefard (IBM-UK)
Nick Jennings (Imperial)	Tien Pham (ARL)
Nirmit Desai (IBM-US)	Troy Kelley (ARL)
Rachel Bellamy (IBM-US)	
Roger Whitaker (Cardiff)	
Sebastian Stein (Southampton)	
Soheil Eshghi (Yale)	

The purpose of Project 4 is to undertake basic research in how complex adaptive human systems evolve in conditions relevant to coalition operations and how they can be proactively influenced and exploited. This requires exploring the properties of groups of humans, and the principles defining how such groups would react to external stimuli and interactions with other collectives. There is a substantial interdisciplinary literature concerning complex human systems spanning sociology, anthropology, psychology, biology, computing and economics. However, these inter-disciplinary activities need to be re-examined in terms of relevance to coalition operations.

One area which has been relatively unexplored in the realm of complex adaptive human systems is the modeling and analysis of external groups. Within coalition operations, the groups belonging to the coalitions can be characterized as internal groups, and there has been a substantial body of work examining the interaction between internal groups of a coalition, some examples being concepts of shared understanding in coalitions[42], cognitive modeling of coalitions using ACT-R cognitive architectures[43] , cultural network analysis for cooperating groups[44], or modeling cultural differences between U.S. and UK coalitions[45]. In contrast to the understanding of such internal groups, the understanding of external groups has not been examined in any significant manner. Our focus for

[42] P. R. Smart, T. Huynh, D. Mott, K. Sycara, D. Braines, M. Strub, W. Sieck, and N. Shadbolt, “Towards an Understanding of Shared Understanding in Military Coalition Contexts”, in Proc. Annual Conference of the International Technology Alliance, 2009.

[43] P. R. Smart, and K. Sycara, “Cognitive Social Simulation and Collective Sensemaking: An Approach Using the ACT-R Cognitive Architecture”, in Proc. International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE), 2014

[44] W. Sieck, “Cultural network analysis: Method and application”, Advances in cross-cultural decision making, pp. 260–269, 2010.

[45] S. Poteet, P. Xue, J. Patel, A. Kao, C. Giammanco, and I. Whiteley, "Linguistic sources of coalition miscommunication”, in NATO RTO HFM-142 Symposium on Adaptability in Coalition Teamwork, 2008.

research in P4 will be on modeling of external groups, such as civilian populations, non government organizations, insurgents and other groups which are not likely to be accessible or cooperating with coalition members, and who may not be amenable to direct measurements or observation.

Current work on understanding the properties of complex adaptive human systems relies primarily on accurate measurements of human behaviors. Such measurements can be obtained passively by reports of field entities (sensors or humans) or proactively by probing the potentially misbehaved groups. However, it is challenging to measure human systems in a dynamic environment since these signals are likely to be noisy and incomplete. Moreover, many external groups may be partially beyond the network sensing range. Hence, for such cases, proactive measurements are not feasible, and passive measurements (relying on sensors and humans) are not reliable.

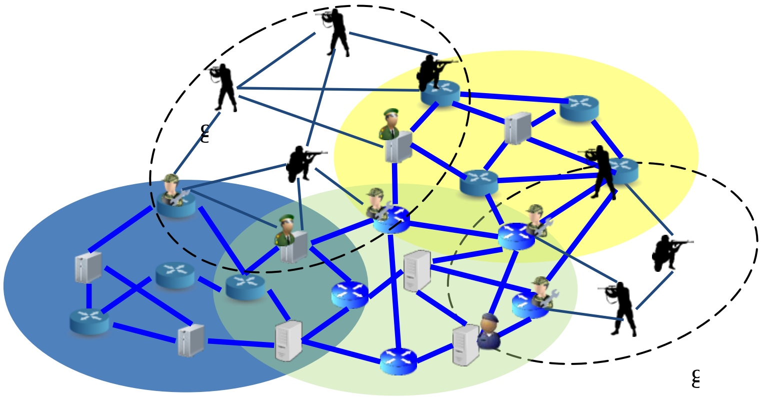


Figure P4-1

As shown in figure P4-1, although human entities in Groups A and B form a connected social network, only a few of them may be directly accessible to the coalition network infrastructure and different elements of the coalition are only able to partially observe or influence these groups. Even when human behaviors are not directly measurable, their impact may still be indirectly observed or measured and the causality can then be inferred. Based on such observations, our goal is to infer the fine-grained human behaviors, such as mobility and cohesion, within each external group, thereby predicting their future behaviors and thus influencing individuals (especially the ones at the network edge) to exhibit preferred behaviors. To this end, we propose a hybrid mobile and social analytics approach based on both proactive measurements and passive inference for the fundamental understanding of external complex human systems in coalition based scenarios.

In order to advance the current state of the art, we propose to undertake two research tasks.

- *Generative and Predictive Models for external group behavior*: We will propose and evaluate new models that can characterize the behavior of external groups, using techniques such as mathematical modeling, simulations and indirect measurements that can be obtained on such external groups.

- *Cognitive architectures for modeling characteristics of external groups:* We will explore how factors such as emotion can affect reasoning capabilities of group members by developing models within cognitive architectures that can provide further characterizations of group mutability and the effect of intervention strategies

In order to make coordinated progress on each of the two tasks, one activity which we will take in the IPP period is the definition of a conceptual model, i.e. an information model or ontology for concepts underlying external groups. The conceptual model provides an enumeration of the key concepts required for analyzing and understanding external groups. These concepts would include a definition of the types of entities in the group, their characteristics, and relations between entities, which would characterize the various groups, behaviors, social and "authority" structures, communication structures, emotional states. The conceptual model will also include representations of ideas that define the context in which mutability occurs, such as culture, religion, "ideas", reasoning (whether "good" and "bad"), geography, world events, roles, motivations of key players in these events. The conceptual model will provide both the syntax as well as the semantics which would allow the development of finer grain models in both of the tasks. The first task will use the model to understand the characteristics and properties of external groups using mathematical models and computer simulations. The second task will use the model to create a meta-heuristic framework, which captures the relationships between requirements for analysis (such as the need to predict how certain behaviors or dimensions of a group will change over time) to the appropriate models and systems that are available to provide causal explanations for such changes and suggest suitable interventions.

Task 4.1: Generative and Predictive Models for external group behavior

Primary Research Staff	Collaborators
Diane Felmlee (PSU)	Alun Preece (Cardiff)
Leandros Tassiulas (Yale)	Cheryl Giammanco (ARL)
Nick Jennings (Imperial)	Dinesh Verma (IBM-US)
Nirmit Desai (IBM-US)	Gavin Pearson (Dstl)
Rachel Bellamy (IBM-US)	Sanaz Yeganefard (IBM-UK)
Roger Whitaker (Cardiff)	Tien Pham (ARL)
Sebastian Stein (Southampton)	Troy Kelley (ARL)
Soheil Eshghi (Yale)	Grace Rose-Williams (Dstl)
	David Mott (IBM-UK)

Understanding how complex adaptive human systems behave and evolve under a range of assumptions is an essential first step in this task which will allow us to develop a new theoretical underpinning of system behavior for scenarios and context relevant for coalition operations. Specifically we will determine the salient features that characterize particular types of adaptive human systems and how these might be detected. To achieve this, together with Task 4.2, we will develop a definition and comprehensive ontology (which we call the conceptual model of group mutability) of different kinds of adaptive human systems, considering dimensions such as: group structure; levels of movement; intra- and inter-group connectivity; independence; density; ostracism; clustering within the larger groupings; presence within multiple networks, and exposure to environmental influences; dominant key actors etc. From this ontology, and combined with the knowledge of key features characterizing and driving dynamics, we will go beyond the existing complexity literature to model and predict adaption of collective human behavior using a combination of mathematical modeling (e.g., game, chaos and complexity theories) and agent based modeling. We will use these theories to inform which parameters must be measured in order to detect and understand the types of behavior that are being exhibited in external groups.

Although complicated human behaviors are complex unconstrained social activities that are challenging to capture using mathematical models, recent work on spatial-temporal networks[46] demonstrates that in large social networks with complex human interactions, simple behavior models can still characterize the main features of the complex behavior. The key observation is that different human behavior leaves unique footprints/signatures within the complex system that can be measured and exploited.

These observations imply that human behaviors can be inferred based on externally observable effects that can be directly measured using coalition network sensors (either passively or actively). We propose to develop a new set of models addressing different categories of possible human behaviors. For these models, each of them corresponds to a unique set of human behaviors and social parameters. The benefit of this approach is that for transient or out-of-scope human behaviors (possible anomalies) that are not directly measurable, their impact on the group can still be measured or reported by some group human entities who have direct interactions with the network infrastructures. Furthermore, identifying which behavior model is being followed is also complex and often counter intuitive. As an example, specific human groups being monitored may exhibit dynamics in coordinated mobility as shown in[47]. Such human dynamics may be interpreted in two ways: (i) group topology changes over time due to coordinated planned operations, or (ii) individual knowledge of group members evolves over time by the spread of

[46] Williams, M. J., Whitaker, R. M., & Allen, S. M. (2012). Decentralised detection of periodic encounter communities in opportunistic networks. *Ad Hoc Networks*, *10*(8), 1544-1556.

[47] Webberley, W. M., Allen, S. M., & Whitaker, R. M. (2015). Retweeting beyond expectation: Inferring interestingness in Twitter. *Computer Communications*.

information, which in turn depends on opportunities for dissemination, resulting in apparent coordinated behavior[48]. Additional behaviors, drivers and response patterns within groups can provide as the basis for underlying descriptors of group evolution and adaptation. We would model and analyze the behaviors of such groups using techniques from social exchange theory, social comparison theory, relevance theory and resource mobilization theory.

In our research, we will postulate new models that can establish causality, and provide a computational methodology to predict human action based on the shared goals of the collective. The origins of this pursuit date back to the international arms race[49], where for example according to a number of scholars, differential equation modeling played a part in quelling the arms race between the United States and the Soviet Union during the Cold War.

Alongside social agent based approaches through which detailed behaviors can be embedded and modeled, we will build on systems of dynamic, differential equation models to theorize about the interaction between actors in scalable coalition situations, where each actor represents the actions of one coalition, or external faction. As an example, two actors can be modeled by a system of coupled ordinary differential equations as follows[50]:

$$dx/dt = a_1(x^*-x) + b_1(y - x)$$

$$dy/dt = a_2(y^* - y) + b_2(x - y),$$

where x and y denote the behavior of actor 1 and actor 2 at any time t, respectively, a and b are model parameters, and x* and y* represent external standards, goals, or equilibria to which each actor, in the absence of the other actor, will tend to approach over time. This coupled, dynamic model specifies that the instantaneous rate of change in each actor's behavior is a function of two components: i) the difference between an actor's behavior and that actor's own standards, or goals, and ii) the difference between the behaviors of both actors. This type of model can be used to predict situations under which the behavior of two sets of actors, or coalitions, will interact in ways that generate over time a) stable, convergent outcomes (e.g., develop compromises), versus those that result in b) divergent, unstable paths (e.g., joint, conflicting behavior), or finally, those that produce c) repeating, cyclical, oscillating outcomes (e.g., oscillations between periods of escalation and de-escalation of cooperation and/or conflict). However, such a simple model cannot fully capture the tactical operations in challenging environments. Therefore, we will investigate multi-dimensional nonlinear systems to predict chaotic, dynamic patterns. Equally when the interactions are more complex and non-linear, modeling by differential equations may be impeded. Our dual strategy of using computational agent based approaches for computational modeling of collective interactions provides useful complementarily.

According to some social scientists[51], the formation of groups among individuals can be viewed as a way to maximize some measurable and definite outcome. Human systems come into being because the participants want to maximize their chances of achieving some external goal, and they have come to the rational conclusion that cooperating with other people is the best way to do this. In these cases, the evolution of coalition behavior can be modeled as a utility maximization problem, which can then be solved under specific assumptions. The application of utility maximization techniques could be used to provide a causal model for interactions in specific situations, e.g. when modeling behavior of terrorist organizations interacting with non-governmental organizations in a coalition environment.

Other social science theories[52] postulate homogamy as one of driving reasons for complex human systems to arise, where people choose to collaborate with others not because they view it as a means to an end, but as an end in

[48] Allen, S. M., Chorley, M. J., Colombo, G. B., & Whitaker, R. M. (2012). Opportunistic social dissemination of micro-blogs. *Ad Hoc Networks*, *10*(8), 1570-1585.

[49] Richardson, L. F. (1960). *Arms and insecurity: A mathematical study of the causes and origins of war*. Boxwood Press.

[50] Felmlee, D. H., & Greenberg, D. F. (1999). A dynamic systems model of dyadic interaction. *The journal of mathematical sociology*, *23*(3), 155-180.

[51] Walsh, D. T. (2006). A structural approach to the study of intra-organizational coalitions.

[52] Axelrod, R. M. (1970). Conflict of interest: A theory of divergent goals with applications to politics. Markham Pub. Co.

itself. Many people simply enjoy interacting with like-minded others. In this case, coalition members would tend to undertake behaviors which increase their interaction with other entities with similar attributes. We propose to model homogamy using mathematical models, which can then be analyzed jointly with utility maximization techniques to give a causal prediction of human systems in coalition settings, e.g. the dynamics driving non-governmental organizations interacting with fundamentalist groups in an environment like Iraq or Afghanistan.

Another approach for the causal model is to formulate the human system as an optimization problem by identifying each collective's internal properties. We propose to create new objectives that combine characteristics of agent-based models, statistical models and mathematical models to represent the collectives and interaction between these collectives. Interactions among collectives can be modeled as structured and constrained interaction among individuals in the collective. By imposing topological and graph-based constraints on how collectives interact with their individual members, we would be able to analyze the behavior of interacting collectives by solving this optimization problem

IPP Activities

In the first year of the program, the team's first goal is to establish a common goal and vocabulary that can act as a means for effective communication between the team members, many of whom are working together for the first time. Leveraging the collaboration meetings planned in December 2016 and January 2017, we would clarify the context, goals, scenarios of research, and exchange reviews about the various types of social theories, mathematical and agent modeling techniques that can be used to model external groups. This would then lead to selection of the suitable preferred approach for each aspect of external groups that we will study.

Our next goal is to define a conceptual model for the representation of different attributes associated with external groups. The conceptual model will be defined in conjunction with all researchers in both tasks of the project, and will define a common vocabulary that captures the key concepts required to model external groups. Further details are given under Task 4.2, which shares this activity. Subsequently, each of the tasks may end up extending and refining the model in different directions, based on the type of modeling that is performed.

External groups can mutate due to many factors, and such mutations can take the form of polarization among the groups, increase in terror incidents, civil unrest or political upheavals. Our next task is to identify such factors that can cause mutations like polarization in the external groups. We would model and study how emotions spread within external groups, using techniques such as epidemiological models[53] to understand the spread of ideas, memes and polarizing emotions.

We would also consider specific scenarios for external groups, e.g. modeling terrorist groups in the theater of coalition operations, and consider approaches that can model interactions among these external groups. Our goal will be to find the properties of the external groups that would remain unchanged (immutable) under different evolutionary approaches. Some insights into the behavior of external groups can be obtained from analyzing the information posted from current regions of conflict on social media such as Twitter or Facebook. We would analyze such online information and explore what insights into the behavior of external groups can be obtained from this analysis. Coalition members may want to influence the evolutions of external groups, and they need to develop appropriate intervention strategies for influencing the evolution of external groups. We would develop an understanding of the impact of different intervention strategies using various modeling techniques..

Accordingly, we are defining the following milestones for this task.

Task Milestones:

Date	Description
Q1	− Clarification of context, research goals, key concepts (including types of mathematical modeling and agent based modeling), agreed scenarios for research, initial review of

[53] Hethcote, H. W. (1989). Three basic epidemiological models. In Applied mathematical ecology (pp. 119-144). Springer Berlin Heidelberg.

Date	Description
	mathematical and agent modeling techniques and tools, leading to selection of preferred approaches
Q2	− Creation of a "group mutability conceptual model", a conceptual model for describing group evolution and mutability (shared milestone with task 2). The conceptual model will provide a language for key concepts such as groups, behaviors, social and "authority" structures, communication structures, emotional states, the context in which mutability occurs, culture, religion, "ideas", reasoning ("good" and "bad"), geography, world events and the roles, motivation of key players in these events, etc.
Q3	− Identify factors that cause mutations in external groups, and model the evolution of characteristic (e.g. emotions, ideas and memes) in external groups. Refinement or extension
Q4	− Apply the models for external group evolution to specific scenarios, e.g. occurrence of terrorist incidents in theater of coalition operations. Identify properties for evolution of the external groups, analyzing group-wise sampling/observation strategies to deduce the status of an external group.
Q5	− Analysis of online behavior of external groups, and models that extract insights about evolution of external groups from these models, giving insight into the potential for effective interventions.

Task 4.2: Cognitive architectures for complex human-agent collectives

Primary Research Staff	Collaborators
David Mott (IBM-UK)	Cheryl Giammanco (ARL)
Jared Somerville (Airbus)	Gavin Pearson (Dstl)
Nicholas Christakis (Yale)	Grace-Rose Williams (Dstl)
Nick Jennings (Imperial)	Leandros Tassiulas (Yale)
Sebastian Stein (Southampton)	Sanaz Yeganefard (IBM-UK)
Dean Clarke (Airbus)	Tien Pham (ARL)
	Troy Kelley (ARL)

Our overall goal in the project is analyzing the behavior of mutability of external collectives is to gain an understanding of how, in a coalition context, the coalition network could be used to monitor, understand and affect the emerging behavior of external collectives. We would like to have architectures that enable us to:

i) develop methods to assist a non-specialist user to obtain a situational understanding of such emerging behavior;

ii) investigate how we might indicate to the user the impact of actively probing behavior changes in response to coordinated coalition actions; and

iii) ensure predictive models are used effectively to proactively influence collective behavior.

Task 4.2 supports these goals by developing models to capture and predict the behavior of external collectives and enable situational understanding of such behavior. Whereas Task 4.1 is focusing on the development of mathematical and agent based simulation models, Task 4.2 explores the use of Cognitive Architectures[54,55,56] that allow a more detailed model of human behavior to be constructed, based upon social and psychological principles that can focus on more cognitive aspects influencing human behavior and collective mutability. It also seeks to understand how we identify actions (external stimuli), such as interventions at different points in collectives, which can be taken to disrupt/encourage different types of behavior for a range of scenarios.

Understanding how complex adaptive human systems behave and evolve under a range of assumptions is an essential first step, allowing us to develop a new theoretical underpinning of such behaviors for relevant scenarios and context. To this end the project will develop a definition and comprehensive ontology (which we call the **conceptual model of collective mutability**) of different kinds of adaptive human systems, considering dimensions such as: collective structure; levels of movement; intra- and inter-collective connectivity; independence; density; ostracism; clustering within the larger groupings; presence within multiple networks, and exposure to environmental influences; dominant key actors etc. This will support the determination of the salient features that characterize particular types of adaptive human systems and how these might be detected and will inform methods to model and predict adaption of collective human behavior by mathematical modeling and agent based modeling (Task 4.1) and Cognitive Architectures (Task 4.2). These models will allow us to investigate how such behavior can be proactively influenced for a well-defined range of scenarios. In addition, when seeking to apply probes to determine collective behavior in circumstances where human behaviors are not directly measurable, such a model may assist the inference of fine-grained human behaviors within each external collective from indirect measurements.

We aim to develop Causal Models instead of Correlative Models. Existing models for predicting human behaviors are predominantly based on correlations between observed behaviors, as opposed to an underlying model that manifests behavior with a view to inferring causality. The conceptual model of collective mutability will be augmented with a causal model and this will inform, and be informed by, the mathematical modeling and agent based modeling in task 4.1 and the cognitive modeling in task 4.2. Such models may assist the design of suitable interventions when undesirable collective behavior is encountered.

Human behavior is dependent on factors such as whether the humans are alone, part of a small collective or part of a large crowd where they may act as a complex adaptive system. Their behavior also depends on their past training, the culture and organizational collective structure to which they belong.. Understanding human behavior in the context of these organizational and other external factors may be informed by knowledge of how human problem solving and reasoning (both good and bad) occurs when working in a collective. There are significant ‘hindrances’ and “analytical pitfalls” to “intuitive” human thinking and perception that can cause illogical reasoning that potentially leads to undesirable behavior by a collective[57],[58] . Such hindrances include confirmation bias, tendencies towards negative conformity, groupthink, and polarization. We will explore more detailed cognitive models based upon the use of cognitive architectures to test hypotheses about the effects of factors such as emotion on cognition (and vice versa) [59, 60, 61, 62] and in the causation of desirable and undesirable behaviors [63]. We thus aim to represent

[54] Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S. , Lebiere, C. and Qin, Y. (2004) An integrated theory of the mind, Psychological Review, vol. 111, no. 4, pp. 1036–1060

[55] Wray, R. E, and Jones, R.M. An Introduction to Soar as an Agent Architecture, https://www.researchgate.net/publication/228398068_An_introduction_to_Soar_as_an_agent_architecture

[56] Sun, R. 2004. The CLARION cognitive architecture: Extending cognitive modeling to social simulation, http://www.cogsci.rpi.edu/~rsun/sun.clarion2005.pdf

[57] Halpern, D. F. (1999) "Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker" New directions for teaching and learning, 1999(80), 69-74.

[58] Mott, D. , Shemanski, D. R. , Giammanco, C. , Braines, D. Collaborative human-machine analysis using a Controlled Natural Language (2015). SPIE Next Generation Analyst III (2015).

[59] Gratch, J., and Marsella, S. (2004) A Domain-independent Framework for Modeling Emotion, Journal of Cognitive Systems Research, Volume 5, Issue 4, 2004, Pages 269-306

[60] Keltner D., and Horberg, E., J.,Emotion-Cognition Interactions http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.471.3155&rep=rep1&type=pdf

interactions among collectives via multi-level cognitive models. For smaller collectives, detailed individual models using cognitive architectures like ACT-R (within task 4.2) can be combined with graph models defining connections to obtain a model of collective behavior. Theoretical analysis can then be used to derive the fundamental principles governing such collectives. Larger collectives will be modeled (within task 4.1) by a simplified representation that can abstract out the essential characteristics of the individual or small collectives model and be used to run larger scale simulations to preserve detail and create multiple scales of resolution.

IPP Activities

In the first year of the program, the team's first goal (along with Task 4.1) is to establish a common goal and vocabulary that can act as a means for effective communication between the team members, many of whom are working together for the first time. Leveraging the collaboration meetings planned in December 2016 and January 2017, we would clarify the context, goals, scenarios of research, and exchange reviews about the various types of social theories, mathematical, agent and cognitive architecture-based modeling techniques that can be used to model external collectives. Task 4.2 will then focus on the use of Cognitive Architectures for modeling human members of external collectives

Our next goal is to define a "collective mutability conceptual model" for the representation of knowledge about the mutability of collectives and the effects of interventions. Such a model will provide a language for key concepts (in terms of types of entity, their characteristics, and relations between entities), such as collectives, behaviors, social and "authority" structures, communication structures, emotional states, as well as concepts defining the context in which mutability occurs, such as culture, religion, "ideas", reasoning ("good" and "bad"), geography, world events and the roles, motivations of key players in these events. The conceptual model will provide both the language in which knowledge can be expressed[64], and a formal underpinning that could be used to build **inferential** and **causal models** connecting behavior to inputs. Such a causal model could potentially be used by systems that need to identify features of collectives, and to infer future behavior and the effects of interventions. The conceptual model (including relationships of causality) is likely to be more comprehensive that the specific behavioral models that underpin the work within Task 4.1, but will inform the development of these specific models, and the results found in Task 4.1 may inform the validation and further development of the conceptual model. Thus the conceptual model will be defined in conjunction with all researchers in both tasks of the project, and will define a common vocabulary that captures the key concepts required to model external collectives. Subsequently, each of the tasks may end up extending and refining the model in different directions, based on the type of modeling that is performed.

Coalition members may want to influence the evolutions of external collectives, and they need to develop appropriate intervention strategies for influencing the evolution of external collectives. We would develop an understanding of the impact of different intervention strategies using all of the types of modeling techniques to be employed in task 4.1 and 4.2. Limited information availability and difference seeking utilities in agents (agents valuing difference in their actions with their neighbors) can lead to sub-optimal actions (bad collective reasoning). The release of information is a possible tool to avoid polarization and wrong reasoning among agents with such utilities. Such approaches can provide a way to model intervention strategies and their impact.

[61] Lin J, Marc Spraragen M, Blythe J, and Zyda, M, EmoCog: Computational Integration of Emotion and Cognitive Architecture, Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference

[62] Marinier, R. P, and Laird, J. E., Toward a Comprehensive Computational Model of Emotions and Feelings, Proceedings of the Sixth International Conference on Cognitive Modeling, 172-177. Mahwah, NJ: Lawrence Earlbaum.

[63] Smart, P. R. , Tang, Y. , Stone, P. , Sycara, K. , Bennati, S. , Lebiere, C. , Mott, D. , Braines, D. , Powell, G. (2014) Socially-distributed cognition and cognitive architectures: towards an ACT-R-based cognitive social simulation capability In, Annual Fall Meeting of the International Technology Alliance, Cardiff, UK and http://eprints.soton.ac.uk/367239/1/ITAAFM%20Team%20Cognitionv8.pdf

[64] Mott, D. (2016) The ITA Controlled English report: http://nis-ita.org/science-library/paper/doc-3165

Mutability of collectives and the effects of interventions may also be researched by construction of more detailed "behavioral models" based upon Cognitive Architectures. A cognitive architecture provides a framework of functional modules that may be composed into a "**cognitive model**" that behaves given "sensory" inputs and background contextual information. The functional modules are based upon psychological theories of cognition, such as memory, attention, and arousal; thus resulting behavior may potentially be considered to be psychologically valid. A cognitive model has the potential to represent a greater level of detail about the cognitive workings of the agents than the low-level models of task 4.1, covering more subtle and complex aspects of collective mutability, and this model should be based upon social science theory such as those explored in Task 4.1 as well as ideas on the spread of emotions and disrupting cascades[65]. One such possibility is the modeling of the effects of emotion (or even aspects of brain function) on "bad reasoning": for example how people may revert to potentially incorrect, non-reflective reasoning under emotional stress. The modeling of such "bad reasoning" may provide insights into how people can be misdirected and radicalized, and how interventions may (or may not) affect such behaviors.

The conceptual model defined jointly with task 4.1 can be expanded by defining a **meta-heuristic framework**, which models the relationships between requirements for analysis (such as the need to predict how certain behaviors or dimensions of a collective will change over time) to the appropriate models and systems that are available to provide causal explanations for such changes and suggest suitable interventions.

The use of "probes" to understand the behavior of external collectives is the concern of both tasks, and such sampling and probing can be done proactively, or reactively in response to some event in an external collective. Caution needs to be exercised in such probing since the external collective may react in a different manner to such observations and probing by the coalition members, and the use of cognitive architectures may assist the modeling of such undesirable detection as the result of cognition or reasoning.

Task Milestones:

Date	Description
Q1	− Clarification of context, research goals, key concepts (including "Bad reasoning/critical thinking/emotion"), and scope of mutability modeling, agreed scenarios for research, selection of modeling representation language
Q2	− Creation of a "collective mutability conceptual model", a conceptual model for describing collective evolution and mutability (shared milestone with task 1). The conceptual model will provide a language for key concepts such as collectives, behaviors, social and "authority" structures, communication structures, emotional states, the context in which mutability occurs, culture, religion, "ideas", reasoning ("good" and "bad"), geography, world events and the roles, motivation of key players in these events, etc.
Q3	− Create a detailed "cognitive model" of the behavior of individuals within mutable collectives, using a cognitive architecture within a multi-agent framework to explore the effects of emotion on "bad reasoning" and for the evaluation of intervention strategies. Model the spread of bad reasoning using approaches like cascade analysis and spread of extreme emotions
Q4	− Analysis of pro-active and reactive sampling strategies to characterize optimal approaches for the state detection of an external collective, with application to specific scenarios.
Q5	− Extension of the collective mutability model to capture causal explanations for collective evolution and mutability. Initial construction of a meta-heuristic framework based on these causal explanations, that can assist the user in suggesting interventions and in determining which modeling techniques are most suitable for the user’s requirements

[65] Soheil Eshghi, Leandros Tassiulas , Nicholas Christakis, Ideas for work based on P4 – Evolution of Complex Adaptive Human Systems

Linkages

Project P4 investigates evolution of external groups in coalition environment, which has synergies with research activities being undertaken in projects P2, P5 and P6. While the precise linkages will become clearer as the projects evolve, we have identified team members who will be responsible for linkage of Project P4 activities with the activities in the other projects.

Project	Liaison Responsible
P2	Diane Felmlee, PSU
P5	Sebastien Stein, Southampton
P6	Coordination between Roger Whitaker, Cardiff and Alun Preece, Cardiff

During the first year, the linkage activities will focus on understanding the role of external human groups in driving the requirements for the other projects, and for understanding how technical advances envisioned by the other projects can have an impact on the evolution of human groups.

Project P5: Instinctive Analytics in a Coalition Environment

Project Champion: Thomas La Porta, Pennsylvania State University	
Primary Research Staff	**Collaborators**
Bongjun Ko (IBM-US)	Alun Preece (Cardiff)
Graham Bent (IBM-UK)	Flavio Bergamaschi (IBM-UK)
Heesung Kwon(ARL)	Geeth de Mel(IBM-UK)
Ian Taylor (Cardiff)	Ramya Raghavendra (IBM-US)
Jorge J Ortiz (IBM-US)	Swati Rallapalli (IBM-US)
Lance Kaplan (ARL)	
Leandros Tassiulas (Yale)	
Percy Liang (Stanford)	
Peter Waggett (IBM-UK)	
Sebastian Stein (Southampton)	
Thomas La Porta (PSU)	

The success of future military coalition operations—be they combat or humanitarian—will increasingly depend on the coalition's ability to share data and data processing services (e.g., aggregation, summarization, fusion) at the network edge. This is mainly because, future coalition networks will be composed of a set of heterogeneous assets—exposing their capabilities through micro-service architectures—that come together in an ad hoc manner to fulfill coalition needs, thus realizing a fully distributed information processing eco-system across the coalition boundaries.

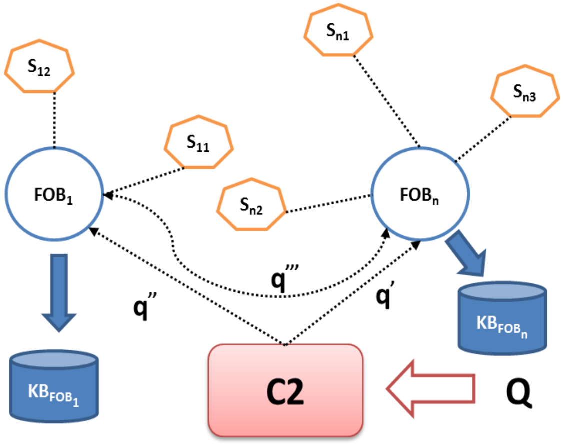
However, one needs to have the means to bring these services together in a seamless and timely manner to support decision making in this new operational context; traditional approaches in which knowledge about services are centralized to match against user requirements will no longer scale in this setting. Thus, in this project we propose a system in which users specify their needs in a declarative manner, and the system infers required services (or compositions) by automatically discovering (or composing) them with respect to the declared user needs. We propose mechanisms in which services are discovered by features that are learnt over time so that the knowledge about these distributed services is derived in an (semi) automated manner. Furthermore, we will make our approach context sensitive (i.e., relevant to user needs) while being also sensitive to the uncertainties in the operating environment. Ultimately—with all these components together—we envision a brain-inspired computing paradigm to automatically match distributed coalition services to requirements so that the full potential of the future military networks may be realized.

We believe the analytic services available in a coalition environment should be automatically discovered and matched to support the tasks at hand without requiring any complex input from the user. We call such an infrastructure instinctive—an infrastructure that reacts automatically to address the analytics task at hand. In order to meet this objective, this project will undertake research into theorems, frameworks and mechanisms required to create instinctive analytics infrastructures. This will require bringing together multiple strands of research from the fields of machine learning, knowledge representation, optimization, and systems engineering [66,67]. Within the

[66] Johnson, M. P., Rowaihy, H., Pizzocaro, D., Bar-Noy, A., Chalmers, S., La Porta, T. F., & Preece, A. (2010). Sensor-mission assignment in constrained environments. Parallel and Distributed Systems, IEEE Transactions on, 21(11), 1692-1705

coalition environment, we view the problem—i.e., how to best find services to interpret data in a distributed resource constrained environment subjected to policy—as a distributed optimization problem.

Let us consider the scenario depicted by Figure P5-1; the command and control (C2) obtains local (or global) situational understanding by querying information from forward operating bases (FOB) which in-turn get localized information from associated services (S_{ij}). These services could be providing any information collection, information processing or information fusion functions. Furthermore, FOBs and S_{ij}s may have communication links with varying quality to obtain localized situational awareness information from other resources, thereby improving their own situational understanding. In this complex web of interactions, some key questions to be answered are: (1) how to formally specify requirements in distributed coalition environments; (2) how to discover shared data, services, and their features, automatically to satisfy requirements; (3) how to create, expand, and use shared domain models automatically; (4) what strategies and mechanisms can be used to efficiently offload services to the network edge and across coalition boundaries by taking into account their computational properties and costs; (5) how constraints of the operating environment and service fit can be used to optimally allocate and adjust services; and (6) how the uncertainty in the operating environment—both in terms of data and user requirements—affects feature discovery and service matching for requirements.



Fiaure P5-1: Distributed Analvtic Services

These questions identify the need to perform a continuous set of distributed choices about how services should be employed (or composed) to satisfy the demands of multiple tasks; we aim to develop theoretical models to answer these questions and to design new approaches to self-describe, self-discover, and self-assemble both data and services in support of coalition mission goals. In doing so, we will address the topic of *Optimizing the Matching of Services to Tasks*. While performing the matching, we will also investigate issues that may arise due to the partial views of available services from uncontrolled sources (i.e., it may be hard to estimate the relevance of such services to the tasks at hand) and the trust in available services (e.g., services from uncontrolled sources might not fully meet the advertised functionality).

Project 5 provides a way to dynamically compose the analytics from the appropriate services and, together with Project 3, will provide insights into the development of a self-organizing declarative platform for analytics. This platform will underpin the analytics layer for the system that will perform information fusion for anticipatory situational understanding, addressed in Project 6. The strong linkage between Projects 5 and 6 will be led and coordinated by Alun Preece (Cardiff) and Ramya Raghavendra (IBM-US) to ensure that the necessary set of analytic services required for P6 are developed.

In order to advance the current state of the art, we propose to undertake two research tasks.

➢ *Intelligent distributed analytic compositions*: We will explore approaches that can use semantic technologies to form and orchestrate services, and enable the attainment of a self-* architecture.

➢ *Anticipatory anticipations of coalition analytics:* We will explore approaches and algorithms for optimally instantiating service instances.

These two inter-linked tasks are described further below.

[67] Preece, A., Gomez, M., de Mel, G., Vasconcelos, W., Sleeman, D., Colley, S., ... & La Porta, T. (2008, April). Matching sensors to missions using a knowledge-based approach. In SPIE Defense and Security Symposium (pp. 698109-698109). International Society for Optics and Photonics.

Task 5.1: Intelligent distributed analytic compositions

Contributors	Collaborators
Heesung Kwon (ARL)	Alun Preece (Cardiff)
Ian Taylor (Cardiff)	Flavio Bergamaschi (IBM-UK)
Jorge J Ortiz (IBM-US)	Geeth de Mel (IBM-UK)
Lance Kaplan (ARL)	
Percy Liang (Stanford)	
Peter Waggett (IBM-UK)	
Thomas La Porta (PSU)	

In this task we have two related activities:

- ***Self-Describing Services****:* We plan to explore new paradigms by which we can make services self-describing—as opposed to existing manual approaches—in order to enable seamless composition. We propose novel learning mechanisms in which consumer-service interactions are observed to infer the capabilities of available services in an automated manner. Furthermore, we will account for the various types of uncertainty—uncertainty of user needs due to ambiguity in the service request, uncertainty in the data on which the analytics is being performed due to either compromised or uncontrolled data sources, uncertainty due to the availability (e.g., service failure or unexpected unavailability) of a service, uncertainty because of limited views caused by coalition environments, and uncertainty in the exact analytics processing—found in the environment and model them so that we can associate confidence values to the extracted features based on the context.
- ***Self-Discovering Services****:* We propose novel service discovery mechanisms wherein goals are declared and the system automatically finds the desired services—be they atomic or composite—on its own by matching the inferred features of services to the declared goals. However, the transient nature of services at the tactical edge makes the discovery of real-time availability of services a difficult task, i.e., just because a service was discovered at time *t* does not mean it is available at time *t+1*. We will develop novel decentralized service discovery algorithms to effectively match the underlying network protocols, filter by service descriptions and subscription preferences, and provide means to better manage the dynamic nature of both the application and the network. Furthermore, the discovery needs to be resilient to the uncertainties in both the requirements and service descriptions.

Self-Describing Services:

In order to provide a bird's-eye-view of the operating environment, especially in distributed coalition environments, data from multiple heterogeneous sources need to be analyzed using coalition-wide shared services. Currently, much of this analysis is pre-planned with the data required to perform the analyses being collected from a variety of sensors and sources and processed in predefined stove-piped systems with the results being disseminated to the users. To provide greater flexibility, micro-service architectures in which new applications are created from available component services[68,69] are being more widely adopted. While micro-services provide greater flexibility, micro-service composition techniques face significant challenges when they need to be combined into complex services. This is exacerbated when they are required to operate in areas where connectivity is no longer guaranteed.

[68] Namiot, D., & Sneps-Sneppe, M. (2014). On micro-services architecture. *International Journal of Open Information Technologies*, *2*(9), 24-27.

[69] Bermell-Garcia, P., Verhagen, W. J., Astwood, S., Krishnamurthy, K., Johnson, J. L., Ruiz, D., ... & Curran, R. (2012). A framework for management of Knowledge-Based Engineering applications as software services: Enabling personalization and codification. Advanced Engineering Informatics, 26(2), 219-230.

Furthermore, the challenge is increased in the coalition context in which no single party may have a complete view of the network.

One approach to address the above challenge is to create data and service modules that are self-describing, and that can self-discover and self-assemble to meet the required goals while adhering to policies. Such an approach would enable the system to discover data-to-services and services-to-services relationships so that effective service negotiation is made possible in disconnected, dynamic environments. To achieve this goal of self-description, we propose to investigate techniques to automatically learn concepts/properties of different domains, create and extend ontologies with respect to new knowledge, and have mechanisms to dynamically expand knowledge bases with new information. The domain models we need for these analyses will be represented in a manner that is machine processable. We will collaborate with P3 on the semantic data description aspect of this work.

Existing techniques in self-description, discovery, and assembly require service providers and consumers to agree on a set of definitions that describe the service and how to consume them—i.e., application programmers must manually compose these descriptions and actively publish them for consumers. We propose to examine ways to automate this process by boosting existing descriptions through observational characterization. In order to characterize services we will look at feature-extraction techniques that extract characteristic behaviors from passively collected traces of communication channels between service consumers and services. For example, we can collect time-series data from communication channels between cooperating entities, feed these into a neural network, extract the feature-vector from the hidden layers, and feed these to a clustering algorithm; similar transfer-learning techniques have been used in image processing[70].

When services are used in isolation or in compositions to satisfy user needs, it is paramount that constraint attached with those services are respected. Typical in systems, we use policies to facilitate such. However, in dynamic environment such as the once we focused in this program, traditional policy frameworks fail to meet requirements such as security and access sufficiently—e.g., users may use services differently in different contexts, thus requiring different usage and access policies for the same service or composition; thus the challenge is to generate appropriate polices in context based on the capability/usage patterns of services with respect to the requirements. P2 aims to investigate on a new policy paradigm in which policies could automatically be generated, activated (or deactivated) in context, thus we would collaborate with P2 team to apply the concept of generative policies with respect to self-describing resources and services, and show that this enables a more appropriate service architecture for coalition operation.

Furthermore, to characterize data at the edge, we will explore more efficient techniques for latent process modeling and change-point detection in time-series data. The former will generate topic labels that are used to summarize data sets so that matching can take place efficiently, and the latter will approximate a probabilistic mixture model that can be used to compare service behaviors. We will design new machine learning techniques that work optimally at the edge where resources are constrained, since current techniques will not scale in resource-constrained environments. More specifically, we will examine online/stochastic-learning techniques under such constraints and explore how learning techniques such as Gibbs sampling could be tailored to approximate inference at the tactical edge[71]. We note that our results will be inherently probabilistic and such uncertainties must be captured explicitly for the benefit of upper layers.

To account for existing disparate coalition labeling/description of services we propose to investigate and augment the techniques based on Latent Semantic Analysis (LSA), context dependent Latent Dirichlet Allocation (LDA), and word embeddings to existing service descriptions and work flows. We further propose to investigate how the resulting similarity scores can be used to search for candidate matching services. The work on self-describing services should also improve the capability to resolve the coalition labeling/description through the development of a common automatically generated human/machine understandable language for service description.

[70] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?. In Advances in Neural Information Processing Systems (pp. 3320-3328).

[71] Steinhardt, J., & Liang, P. (2014). Filtering with abstract particles. In Proceedings of the 31st International Conference on Machine Learning (ICML-14) (pp. 727-735).

Self-Discovering Services:

Once we obtain such domain models, we can utilize concepts from declarative networking to effectively map the "*what"* of requirements to the "*how"* of the services by considering the logical relationships among them. In declarative networking[72], the goals of missions are declared and the system automatically infers how to take actions like routing or access control. Motivated by this notion, we propose to develop the concept of declarative analytics. In declarative analytics, an analyst declares the type of results they want to obtain, and the system automatically infers how to meet the desired goals by performing different subsets of the analysis in a distributed manner, thus promising a new way to address challenges. By representing the features of data and services with respect to ontologies we are contextualizing them so that we can select appropriate resources given a particular context of operation.

The declarative approaches, such as the ones mentioned above, offer a part of the solution; we recognize such logical and conceptual frameworks may not scale well in large networks. We therefore propose to investigate techniques for performing distributed analytics based on the self-organizing properties of networks. Such techniques have proven to be effective for self-organizing distributed databases[73], where fitness functions are used to determine the emerging network structure. We propose to investigate the use of similar techniques to perform self-organization and assembly of distributed services.

Our intuition is that such self-discovering systems will provide new ways of solving the problem of matching tasks to services in a dynamically changing environment. Instead of adhering to static planning to solve the matching problem, each task running on a node can perform a local search to find out which of the services in its neighborhood are available or otherwise. However, finding available services is a hard problem to be solved for the tactical edge due to the dynamism and the temporal constraints. To solve this problem, first, we propose a distributed search protocol, which is capable of dynamically adapting its modes of operation according to the current mobility constraints and application profiles. For example, one-to-many applications (i.e., one service serves many concurrent users) may be best optimized using a proactive service announcement approach, whereas many-to-few applications may be better optimized using a more reactive scheme to provide the most up to date information at minimal bandwidth cost. This efficient approach will be capable of providing a real-time catalogue of recently available services along with annotations. We then employ a distributed allocation protocol which discovers real instances of the services, and mediates competition for these services among missions; this mediation may be done at the services themselves by determining their own fit for a mission and the cost to declare themselves for a certain mission.

Moreover, to assist in discovery, we will define new notions of distance that encompass the idea of mission-fit (e.g., a better fit could be *closer*) and cost (e.g., network distance or demand on services). This work will be done jointly with Project 3. In Project 3, the work focuses on a distance measure as it relates to the physical location of the processing and data, the state of the network, and the ability to move data and processing closer; in this project, we focus on the notion of mission-fit distance. We note that the mission-fit distance is different from the uncertainty associated with data, processing, and user requirements as the former affects the resource allocation while the later affects the value of the result.

To validate our approach on knowledge base expansion, we will compare our automated feature extraction mechanisms against labeled datasets for different types of micro-services. We can characterize typical micro-service behavior by composing services by hand and comparing our manual labels/descriptions with automatic labels/descriptions; the goal here is to approximate manual labeling. If successful, we could also use the approach to characterize and correct manually constructed—or inferred—descriptions. Furthermore, we plan to use well-tested Mechanical Turk approaches to validate the sensibility of the automatically generated (or inferred) annotations.

[72] Loo, B. T., Condie, T., Garofalakis, M., Gay, D. E., Hellerstein, J. M., Maniatis, P., ... & Stoica, I. (2006, June). Declarative networking: language, execution and optimization. In Proceedings of the 2006 ACM SIGMOD international conference on Management of data (pp. 97-108).

[73] Bent, G., Dantressangle, P., Stone, P., Vyvyan, D., & Mowshowitz, A. (2009, September). Experimental evaluation of the performance and scalability of a dynamic distributed federated database. In Proc. 3rd Ann. Conf. International Technology Alliance.

Moreover, we plan to simulate a dynamic service eco-system so that we can stress test our automated annotation approach.

Task Milestones:

Date	Description
Q1	− Use NLP and semantic analysis and other techniques to model query intent with respect to relevance and domain-specific information
Q2	− Create an initial service scenario with subject matter experts (joint with 5.2); Definition of distance metric with P3; Build support for time-based partitioning and temporal joins, both of which are novel for distributed processing of time series data
Q3	− Apply word embeddings and topic modeling work with respect to coalition vocabularies, self-description services and application of generative policy models; Use semantic analysis and NLP techniques to match labeled time series with user-specified queries; Build novel statistical mining techniques (e.g. frequent location mining, common trajectories, next location prediction etc.) to enable us to perform anomaly detection in a variety of use cases
Q4	− Techniques to use/augment semantic distance measures to match requirements to inferred service descriptions; Discover the topic/domain of the set of services and data based on descriptions and data values and labels; Build learning-based mining techniques and platforms that are required for P6 activities
Q5	− Application of neuromorphic computing into the self-organizing data collection architectures; Workshop/conference papers related to project topics (1) word embeddings and topic modeling for service description creation; (2) matching of use requirements, possibly in natural language, to service descriptions (3) self-describing services driven generative policies (4) Self organizing data collection architectures for neuromorphic computing, etc.; Match data to services using service descriptions and data values and associated labels

Task 5.2: Anticipatory instantiation of coalition analytics

Primary Research Staff	Collaborators
Bongjun Ko (IBM-US)	Flavio Bergamaschi (IBM-UK)
Graham Bent (IBM-UK)	Geeth de Mel (IBM-UK)
Heesung Kwon (ARL)	Ramya Raghavendra (IBM-US)
Lance Kaplan (ARL)	Swati Rallapalli (IBM-US)
Leandros Tassiulas (Yale)	
Percy Liang (Stanford)	
Sebastian Stein (Southampton)	
Thomas La Porta (PSU)	

In this task we have three related activities:

- ***Self-Allocating Services****:* We will introduce a rigorous framework for allocating services and resources to meet coalitional goals that accounts for different levels of *specificity*. This latter term captures the fact that declared goals can be generic or more specific with respect to requirements. The proposed framework will allow the modeling and assessment of crucial mission tradeoffs such as the information required to specify coalitional goals, the necessary time and resources for delivering on goals, uncertainty in both requirements and services, and so forth. We propose novel methods for the systematic description of goals with different levels of specificity, which will be employed accordingly in the resource matching and service allocation (or composition). We will also provide estimates of confidence in the allocation based on the uncertainty in the operating environment.

- ***Self-Provisioning Services:*** In dynamic environments it is important to have strategies to optimally reallocate or re-provision service plans. This is necessary, for example, when services unexpectedly fail, when more suitable services become available during plan execution, or when new exploitation opportunities of existing services are identified. In this regard, we propose distributed and online allocation mechanisms in which a system automatically reallocates and matches services as the network state evolves. To do so, we will consider various optimization techniques such as online resource matching methods which require no new information, and probabilistic reasoning about service performance. Furthermore, we propose frameworks to determine distributed service requirements by leveraging the contextual information from the underlying network structure of a coalition (e.g., membership, subgroups) so as to increase the efficacy of the repurposing.
- ***Non Von Neumann Analytics:*** We propose to investigate the potential for performing distributed analytics based on brain-inspired computational models. We will explore the potential of neuromorphic processing to be extended to a distributed network setting, as a radically new approach for future coalition operations.

Self-Allocating Services:

We propose novel service allocation protocols to automatically instantiate service plans to best meet the requirements of users; we consider the fact that the dynamism associated with the operational environment may dictate the pool of available service instances which may not fully meet the specified requirements or the coalition members may not have complete visibility of the data and services offered by others. Thus, it is hard to accurately quantify the relevance of available services to the task at hand. Furthermore, uncontrolled sources of data and services mean that we may not trust each of the services equally to comply with the advertised functionality. Ensuring robustness in the face of these challenges is a critical requirement of our project; we believe the important features that will enable us to meet these challenges include 1) redundancy—i.e., the allocation should account for redundancy in order to ensure task completion and to enhance the credibility of results; and 2) quick response—i.e., the allocation should be able to quickly adapt when new information becomes available so as to match the most relevant of the available resources to the task at hand. Moreover, since service allocation based on self-describing data and services is analogous to distributed task dispatching frameworks on logical data in the big data paradigm, we plan to borrow the concepts from such work[74] to instantiate service plans effectively and efficiently in distributed dynamic environments, especially by considering their scalable and compositional language models.

We will also deal with the uncertainty in data and service descriptions. There are many important properties to be considered for such a protocol: cost (e.g., complexity, bandwidth), regulations (e.g., a certain coalition partner only provides a percentage of a service utility but it is readily available), temporal constraints (e.g., a service that gives 100% utility is available in an hour’s time but a service with 80% utility is available now) to name a few. We aim to capture such properties in simple but useful utility functions and develop strategies to select best fits based on the mission requirements (e.g., accuracy vs. quick response) so that the utility is maximized in the given context. We will also develop resource allocation algorithms that account for uncertainty. These algorithms will allow users to tune requirements of confidence intervals of the quality of results when making allocation decisions. Users may be risk seeking, encouraging the system to take on tasks that may provide high quality results with low probability, or they may be more risk-averse. Our work will initially provide resource allocation of information sources in environments with stochastic demands and utility. Such uncertainty and fuzziness also call for dynamic service composition, which creates services on-demand based on the operational requirements, resource availability and constraints. In order to ensure robust and timely service allocation, and execution even in the presence of highly uncertain services, we will also build decision theory to incorporate redundancy and dynamic re-provisioning policies into complex service workflows.

[74] Srivatsa, M., Ganti, R., Borbash, S., & Agrawal, D. (2014, March). A BSP approach to composite network analysis. In IEEE PERCOM Workshop, 2014 (pp. 407-412).

Finally, fuzzy allocation not only has to take into account the local objective to allocate most fitting analytic services to the tasks but also try to optimize for global objective—i.e., how to allocate services such that global utility functions are maximized. With the goal of optimizing the global objective functions, we will explore easy tuning of weights on the different tasks such that some highly critical tasks receive higher priority in the global optimization.

Self-Provisioning Services:

Once services are allocated, they operate over time periods during which the mission service needs may change—e.g., new missions may emerge, other missions may end, while the network state may also have changed. Given the need to operate without interruption, we propose to leverage self-describing services to allow reconfiguration in real-time in a fully distributed manner. As conditions change, the *cost* of certain services will change, perhaps based on the demand or their location in a network. The importance of these services to missions may also change as missions emerge, or are changed by users. We propose to formalize the usefulness of the services to missions developed in Task 5.1 into mathematical utility functions to incorporate the cost to each mission. Much like frameworks such as network utility maximization[75] (NUM), we hope to prove that we can reach optimal allocations of services, continuously, in a distributed manner. Unlike the normal NUM formulation, however, we will deal with evolving service descriptions, and the ability of services to declare a value to a user, and to rely on local and shared results of ongoing analytics. The algorithms need to deal with a tradeoff between the disruption to an existing task and the delay of a new task, and policies developed in context with Project 2 would need to be used to determine the dynamic behavior in this case.

In addition, we will design online policies for service deployment and repurposing, which satisfy the distributed analytics requirements of the present, and respects other long-term performance and operation coalition goals. The optimality criteria may vary from maximum performance to minimum resource consumption so as to sustain and prolong coalition operation. These objectives can be determined dynamically, based on the task importance, the mission's impact, and so forth. One major breakthrough of the proposed policies is that they will provide the best possible outcome without the need for any prior information about the service availability and mission requirements. This is expected to revolutionize the way coalitions deploy, exploit, and repurpose such distributed services. Our approach will build on related methodologies and results that will be developed in the context of Project P3.

There are few fundamental issues that we need to address in order to make our *self describing, discovering, allocating, and provisioning* approaches be truly distributed. For example, communication between the edge and the FOBs is expensive. As such, it cannot be assumed that we can collect information centrally—even at FOB level—to infer features of services. Thus, we must be able to push computation to the edge and tailor existing algorithms to minimize resource consumption. We will examine the use of distributed learning techniques under the given resource constraints at the edge, including limited memory and bandwidth. In order to maximize the information content that is sent from the edge to a FOB, we will explore means to execute unsupervised (or semi-supervised) clustering techniques, adapted to such environments. Existing algorithms mostly assume unbounded resources and are designed for very large-scale systems and datasets. We look to approximate the effectiveness of these techniques in resource-constrained environments through inference on sparse inputs combined with algorithms that minimize resource consumption – such as minimizing the communication cost.

An important aspect of *self describing, discovering, allocating, and provisioning* services to users' needs is the fact that these activities may occur across organizational boundaries. While this flexibility and sharing of services is a key strength of the instinctive analytics infrastructure we propose, it also raises new research challenges. In particular, each coalition member will have its own mission priorities or policies that may conflict with those of other members, and this could lead to strategic behavior that harms the overall coalition objectives. For example, a partner may be reluctant to offer critical services that it may potentially need in the future or, during mediation, may misrepresent the cost and mission-fit of its services, in order to favor requests from within its own organization. To mitigate such strategic behaviors, we will build on the frameworks of game theory and social semantics [108,109], and we will propose novel service mediation mechanisms that incentivize coalition members to

[75] Neely, M.J., 2010. Stochastic network optimization with application to communication and queueing systems. Synthesis Lectures on Communication Networks, 3(1), pp.1-211.

reveal the cost, mission-fit, and availability of their services truthfully, in order to best meet the coalition's overall objectives.

Another key part of our research is to address the uncertainty in analytic processes. We will initially focus on the uncertainty of the data; at this level uncertainty may be known if data sources are under control of the users or may have an estimation if data is being provided by a coalition partner. We will then extend the work to include uncertainty in analytics processing; such uncertainties, for example, are caused by the processing services being provided by coalition partners. A key aspect of this type of uncertainty is that the analytic services may be prone to unexpected failures or delays in processing, which have to be anticipated and mitigated by the service consumer. We will also address the uncertainty in the availability of services by improving the service discovery processes to provide more responsive real-time discovery information to the subsystems. Moreover, we will explore uncertainty in user requests. This may be caused by vagueness in user requests, and will be mitigated by augmenting the knowledge of domain properties. Finally, in order to provide a more comprehensive analysis, the proposed methods will account for different levels of specification, or specificity, for the coalition goals. The more specific a goal is, the easier it is to match to resources and services. On the other hand, specificity increases the communication and computation overheads for describing the goals and may also reserve scarce resources.

Non Von Neumann Analytics:

A high-risk high-reward goal for Project 5 is to investigate disruptive approaches to distributed situational understanding and the required associated services. Specifically, we will consider a new approach that is based on a non Von Neumann neuromorphic computing model[76]. The demands of future tactical environments will require new types of technologies to overcome the requirements imposed by limited communication bandwidth and power. The trend in traditional Von Neumann computing is towards computing devices that scale at best linearly with computational complexity whereas non Von Neumann computing models scale logarithmically with computational complexity and use significantly less. A number of current research programs, such as DARPA's Systems of Neuromorphic Adaptive Plastic Scalable Electronics (SyNAPSE) program and the European Commission Human Brain Project, have begun the development of a new generation of brain-inspired neurosynaptic computational architectures that are compact and consume little power. To support these developments, a new generation of neuromorphic computing models is being developed, in which both data and associated processing are distributed naturally within the network of processing elements. Such models have been shown to be applicable to a range of signal processing and analytics tasks.

These developments naturally lead to the question of whether these brain-inspired computing models can be extended to the service composition problem in which both data and associated processing are distributed naturally within a network based on a non Von Neumann processing paradigm. Such an approach will potentially suggest a radically new approach to network processing in which the data and analytics are represented in fundamentally different ways to today's processing architectures and one in which the realization of the *distributed brain* may become significantly more like a real brain. We propose to investigate how such new computing models can be applied across a coalition network and the types of inter-process communication that are required to support such a model. Whilst we recognize that this proposal is of high technical risk, the potential benefits are likely to be of great significance and will help to inform the development of future network operations.

To validate the service allocation and re-provisioning, we plan to perform theoretical analysis, and simulation of matching and service allocation algorithms to determine bounds with respect to optimal assignments. This will also help us in evaluating the different definitions of distance and allocation algorithms. Furthermore, we plan to perform comparison of matching decisions and descriptions made by automated algorithms and those made manually to determine quality of learning algorithms. In order to validate the scalability of the proposed research, we plan to use agent-based simulations. Additionally, we also plan to investigate the termination, correctness, and completeness properties of the mechanisms used in these simulations and experiments by means of critical argument and mathematical modeling.

[76] Arthur, J. V., Merolla, P., Akopyan, F., Alvarez, R., Cassidy, A., Chandra, S., ... & Modha, D. S. (2012, June). Building block of a programmable neuromorphic substrate: A digital neurosynaptic core. In IEEE Neural Networks (IJCNN), 2012

Task Milestones:

Date	Description
Q1	− Collate relevant non-military data sets and develop requirements from these (SH, IBM-US, Yale, ARL); Determine with 5.1 how candidate services and data pools will be formatted (PSU, Cardiff, Yale); Explore potential collaboration with P1 for using percolation theory to: (i) Support resilient connectivity for distributed service placement; (ii) Ensure resilience when mapping neuromorphic circuits onto a distributed network architecture (IBM-UK, PSU, ARL)
Q2	− Develop military scenarios relevant to the government with subject matter experts (SH, IBM-US, IBM-UK, Yale, ARL); Determine level of uncertainty in system (PSU, SH, IBM-UK, ARL)
Q3	− Determine methods for meeting redundancy requirements for uncertainly and availability (PSU, IBM-US, Cardiff); Initial results from the exploration of techniques for identifying structural isomorphism between trained neuromorphic circuits and corresponding neuromorphic micro-services (e.g. corelets, threshold linear networks) with known functionality (IBM-UK, ARL).
Q4	− Define utility functions for services (SH, PSU, Yale); Self-organizing data collection architectures for Non Von Neumann architectures (IBM-UK, ARL)
Q5	− Define workflow models for re-provisioning (SH, PSU); Define and evaluate algorithms for resources allocation based on utility and redundancy (PSU, Cardiff, SH, Yale); Initial results from the exploration of mapping of neuromorphic circuits onto distributed network architectures in terms of performance and resilience (IBM-UK, ARL).

Linkages

Project P5 has linkages with various other projects in the program. The following researchers will act as liaisons to explore linkages between P5 and the other projects.

Project	Liaison Responsible
P1	Leandros Tassiulas (Yale)
P2	Geeth de Mel (IBM-UK)
P3	Bongjun Ko (IBM-US) and Thomas La Porta (PSU)
P4	Sebastien Stein (Southampton)
P6	Alun Preece (Cardiff) and Ramya Raghavendra (IBM-US)

During the first year, the linkage activities will focus on understanding how the algorithms and approaches being explored in P5 could leverage the algorithms being done in the other projects, and how the two projects could join forces to demonstrate use-cases of the P5 algorithms and technologies.

Project P6: Anticipatory Situational Understanding for Coalitions

Project Champion: Alun Preece, Cardiff	
Primary Research Staff	**Collaborators**
Alun Preece (Cardiff)	Alistair Nottle (Airbus)
Chris Willis (BAE)	Erin Zaroukian (ARL)
Dave Braines (IBM-UK)	Federico Cerutti (Cardiff)
Heesung Kwon (ARL)	Gavin Pearson (Dstl)
Mani Srivastava (UCLA)	Gavin Powell (Airbus)
Pablo Bermell-Garcia (Airbus)	Jon Bakdash (ARL)
Ramya Raghavendra (IBM-US)	Lance Kaplan (ARL)
Simon Julier (UCL)	Mudhakar Srivatsa (IBM-US)
Supriyo Chakraborty (IBM-US)	Richard Tomsett (IBM-UK)
	Ross Lund (Dstl)
	Simon Bray (Dstl)
	Tien Pham (ARL)

Anticipatory situational understanding (ASU) is fundamental to support decision-making and autonomy by all agents within the coalition. Situational understanding is the "*... product of applying analysis and judgment to the unit's situation awareness to determine the relationships of the factors present and form logical conclusions concerning threats to the force or mission accomplishment, opportunities for mission accomplishment, and gaps in information*"[77]. Therefore, each agent in the coalition must be able to form *an awareness of its environment*, perform inferences on that awareness *to identify concepts and relationships*, *predict* how the environment will evolve, and *assess* how its actions can shape this future evolution. Project 6 will explore the algorithms and techniques that will be used to develop ASU. It will investigate how this can be carried out in a distributed coalition with heterogeneous agents, each of which has its own role, has access to its own sources of information (which can be hard or soft), has its own computational resources, needs to make its own decisions and can undertake different actions. The key scientific challenges lie in how the different levels of representation and reasoning interact with one another to allow the flow of information, uncertainty and reasoning between the different levels. To meet these challenges, Project 6 will develop new representations of the environment, integrating knowledge-based reasoning with machine learning, investigate new methods for human-machine collaboration, and identify new inference algorithms for multi-level distributed fusion and estimation.

Our key innovations are centered around new approaches to support transfer of information between different levels of abstraction in a fluid manner. Traditional approaches to ASU have adopted a model that representations live in a strict hierarchy in which low-level inference must be completed before high-level inference can be carried out. We believe that information processing agents should be able to translate, support, and move between different levels of abstraction from the state of individual targets to entire high-level summaries of an operation. Key to achieving this goal are dynamic models for individual agents and the coalition as a whole that will learn and encode (1) *hierarchical information representations* spanning the different levels of abstraction, including knowledge of entities and relationships in the world which are tailored to each agent's needs, (2) *memories* of past and current

[77] http://www.globalsecurity.org/military/library/report/call/call_01-18_ch6.htm

states in the world together with probabilistic dependencies among variables that capture the likelihood of future states, and (3) *sharing mechanisms* which make it possible for agents to share both their memories and the meta-information which describes the representation used to encode those entities. We will use these models to develop algorithms and procedures that can (a) *estimate* the current state of the world by fusing uncertain structured and unstructured data from sources ranging from machine sensors of varying modalities (e.g., acoustic, camera, pressure, location) to human agents and other social sensors (e.g., Twitter, Facebook, Instagram); (b) *predict* future states of the world by projecting sequences of actions and their consequences in time; and (c) *reason* about the feasibility and implications of the chosen actions in terms of driving the world to the desired state, maximizing the operational efficiency of coalition decision-making.

Project 6 will develop the scientific underpinnings of an end-to-end robust and adaptive information fusion pipeline for the processing of collected hard and soft data – including traditional signal processing as well as analysis of soft sources such as text and open source intelligence – through to ASU. Our work will focus on the creation of theory-based dynamic models and mechanisms to integrate multiple fusion capabilities in a distributed decision-support context, spanning multiple levels of abstraction from the state of individual entities to high-level situation overviews. The research aims to produce approaches that can operate in heterogeneous environments typical in coalition operations, comprising both machine and human agents with different responsibilities and capabilities.

For operational efficiency within the dynamic coalition environment, a key focus of our approach is the ability for the pipeline to rapidly reconfigure itself to support situational understanding. This requires an awareness of the current situation together with the ability to predict what it could become, and it is closely linked to work in Project 3 in which prediction capabilities are required to inform proactive configuration of network services and other resources. We seek to attain this goal while maintaining transparency to human users by incorporating knowledge of the state of information components within the system, including key metadata that enables the system to respond and predict how the operational context will evolve.

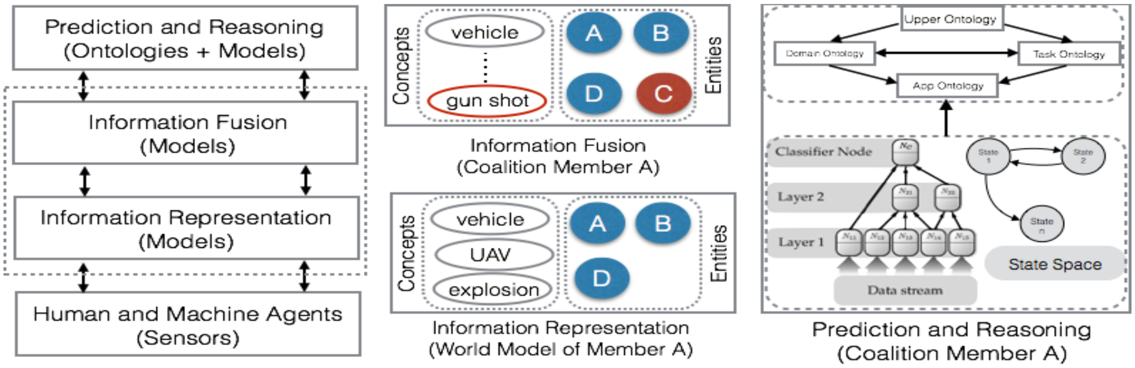


Figure P6-1. (left) Different stages of the end-to-end information fusion pipeline; (center) concepts and entities are modeled by each member; information fusion is used to combine models obtained from other members of the coalition and estimate the current state of the world; (right) models and knowledge bases (including ontologies) are used to predict future states and reason about them.

The goal is to allow for distributed ASU within a coalition network. To illustrate the various technical challenges we consider a coalition with four members A, B, C, and D where each member seeks to attain situational understanding, consisting of an awareness of its immediate situation and a prediction of how this will evolve. A conceptual representation of an end-to-end information fusion pipeline maintained at each member is shown in Figure P6-1(a). The lower layer consists of a collection of human and machine agents, under a particular member, acting as sensory sources providing structured and unstructured data. The information representation layer will use the incoming data streams to learn concepts and model entities together with their relationships at multiple levels of semantic granularity. The history of past observations will also be implicitly encoded in these models. The information fusion layer will have algorithms and techniques, developed to perform fusion over concepts and entities (represented by their models) obtained from other coalition members, and it will estimate the current state of the world. The prediction and reasoning layer will use the estimated current state, together with the state space of the models to predict the future state. It will also use knowledge bases – including formal ontologies – to reason about the future state. Initially, the pipeline will rely on human agents to provide expert knowledge for reasoning, but the goal of our research is to build mature models and algorithms that can over a period of time reduce the human intervention and attain greater autonomy but without entirely replacing human involvement and oversight. There will be a bi-directional exchange of information occurring between the different layers. In the forward path, the inferences at the lower layer will act as input for the next higher layer. In the feedback path, information will be used to adjust the model and algorithm parameters and possibly actuate the sensors differently.

In order to advance the current state of the art, we propose to undertake two research tasks.

- *Human/agent knowledge fusion*: We will explore approaches that can obtain situation awareness by combining the knowledge of humans along with agent based systems.
- *Deep learning for multi-layer situational understanding:* We will explore the use of deep learning approaches to obtain situational understanding in complex multi-layer environments.

Task 6.1: Human/agent knowledge fusion

Primary Research Staff	Collaborators
Pablo Bermell-Garcia (Airbus)	Jon Bakdash (ARL)
Dave Braines (IBM-UK)	Federico Cerutti (Cardiff)
Simon Julier (UCL)	Ross Lund (Dstl)
Alun Preece (Cardiff)	Gavin Pearson (Dstl)
Ramya Raghavendra (IBM-US)	Tien Pham (ARL)
Mani Srivastava (UCLA)	Erin Zaroukian (ARL)

This task focuses on multiple information processing nodes (human and machine agents) working together for anticipatory situational understanding (ASU), as distinct from Task 6.2 that focuses on the fusion architecture of a single node. We seek to develop a novel distributed approach for ASU and evaluate it along multiple dimensions, e.g., question type (open or closed), primary data type (hard or soft), and mission context (tactical or operational).

Supporting situational understanding requires a system to facilitate making inferences about states of the world; this means that the system must, at least at some level, operate in terms of human-understandable concepts and relationships. Humans are sources of information as well as recipients of it, across a range of fusion levels from low to high, so the machine parts of a distributed ASU system must be capable of working with human-friendly information and knowledge representations. This is even more so in a dynamic coalition network requiring the ASU system to assist humans in the selection of actions that should lead to desirable future outcomes (i.e., to be proactive) while also being auditable and transparent to its users.

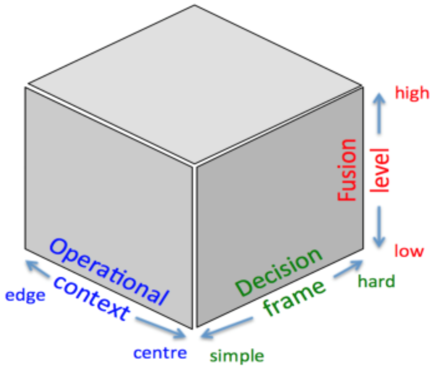


Figure P6-2. Multi-dimensional context for ASU

An agent’s decision-making context has multiple dimensions, some of which are shown in Figure P6-2. The *operational context* in the coalition network (from edge to center) constrains an agent’s view of the network, their connectivity, their availability, and the types of resources and structures they have access to. *Decision frames* reflect coalition decision problems of varying complexities, based on factors such as the agent’s goals, knowledge, values, and environment. *Fusion levels* refer to the syntactic granularity and structure of coalition information available to an agent from low-level (signal, unstructured) to high-level (semantics, structured). All of these dimensions need to be represented to some extent in the ASU system, to enable sharing of context between coalition agents (human and machine).

To achieve integrated ASU across a broad range of abstraction levels in Task 6.1 we will address the research question of *how to link human-understandable representations of concepts and relationships to patterns recognized in sensed data?* Considering deep learning methods such as Hierarchical Temporal Memory (HTM) as a starting point, these algorithms enable the formation of concepts in terms of (learned) patterns at multiple levels in a hierarchy. At the higher levels, these patterns will often correspond to *percepts* that could be named (e.g., *car*, *SUV*, *crowd*). Through supervised learning, the HTM could associate actual names (*referents*) with the percepts so that, a particular set of patterns could become associated with the corresponding name. We envisage this being the point at which the HTM system would link to a human-understandable knowledge-representation system for representation and reasoning about the perceived situation elements. Percepts, through their referents, can be mapped to concepts in the knowledge representation, and can be organized in schemas, allowing the making of inferences necessary for situation understanding and communication between the machine and human parts of the ASU system.

Human-machine ASU is a kind of human-computer collaboration and needs to take account of trade-offs between human and machine capabilities. As a simple example, consider a scenario where aerial imagery shows two vehicles stopped at a particular location. Machine processing of the imagery data may be able to identify the objects in the scene (vehicles, people, features of the location, etc) but a much harder problem for the machine would be to determine that a meeting is taking place. Associating this data with the concept *meeting* would be much easier for a human, especially when the target individuals may be trying to avoid revealing that they are interacting (requiring a significant amount of tacit knowledge). We want to make it easy for human agents in such cases to simply *tell* the system particular pieces of information rather than relying on machine learning to detect it. This facility is also important when available data may be too sparse for an agent to learn percepts robustly, or the learning process may be too slow compared with simply *telling* the agent something. Many relevant aspects of the situation will not be physical in nature, for example, the cultural attributes of external organizations and groups, and these will be harder still for a machine to identify through training and to factor them into the decision-support context.

Our starting point for the proposed research in Task 6.1 will be to build on prior work in controlled natural language (CNL) knowledge representations that are both machine-processable and human-understandable which we have previously applied in an information fusion and situational understanding context. CNL supports traditional knowledge representation models, in a way that is more transparent to users and domain experts, reducing knowledge engineering effort and providing social semantics as well as formal semantics. In the near term, we hypothesize that *an integrated ASU system in which "deep learning" and CNL-based knowledge representation and reasoning work in conjunction via a percept-referent mechanism provides improved performance on fusion problems across a range of abstraction levels compared to non-integrated approaches.*

We would seek to explore how such an integrated ASU system could handle both "learned" and "told" input, and to recognize, reason about, and explicate patterns. Going further, we would aim to explore how the integrated system can perform the kinds of activities traditionally associated with a combination of ontological and machine learning approaches to improve ASU, specifically: using deep learning to acquire, extend, or align knowledge bases; and using knowledge bases as background information in learning. A fundamental challenge for the research is to understand how the predictive capabilities of an approach like HTM can work in conjunction with more traditional predictive models commonly applied to structured and semi-structured data. We envisage the CNL-based knowledge representation component of our proposed system being useful here also, in framing information obtained from multiple disparate coalition sources. Using this approach also positions us well to advance related capabilities in "selection and tailoring of knowledge-based reasoning components", again drawing on the human-friendly but machine-readable representations to enable the human-machine hybrid team to share information in a seamless manner.

Task Milestones:

Date	Description
Q1	− Establish ASU baseline for comparison - the hypothesis is that our Human/Agent Knowledge Fusion (HAKF) approach will outperform the baseline. Establish research method and ways of working. Establish requirements for a representative set of ASU problems / test cases / datasets.
Q2	− [Jointly with Task 6.2] Collect ASU problems, test cases, datasets etc. Contribute these to the DAIS Experimental Framework and make them open source where possible.
Q3	− Define initial HAKF model or choice of candidate models for testing on ASU problems.

Date	Description
Q4	− Write roadmap paper on HAKF for DAIS Annual Meeting defining principles, baseline, hypotheses, and a research agenda for the first 5 years of DAIS.
Q5	− Perform initial evaluation of Q3 HAKF model(s) on Q2 ASU cases/datasets. Contribute reproducible results to DAIS Experimental Framework and make open source where possible. Extend Q4 HAKF roadmap paper with results and submit for publication..

Task 6.2: Deep learning for multi-layer situational understanding

Primary Research Staff	Collaborators
Pablo Bermell-Garica (Airbus)	Lance Kaplan (ARL)
Supriyo Chakraborty (IBM-US)	Ross Lund (Dstl)
Simon Julier (UCL)	Alistair Nottle (Airbus)
Heesung Kwon (ARL)	Gavin Pearson (Dstl)
Alun Preece (Cardiff)	Tien Pham (ARL)
Mani Srivastava (UCLA)	Mudhakar Srivatsa (IBM-US)
Chris Willis (BAE)	Richard Tomsett (IBM-UK)

This task focuses on the fusion architecture of a single (machine) information processing node for anticipatory situational understanding (ASU), as distinct from Task 6.1 that focuses on multiple nodes (human and machine agents) working together. We seek to develop a novel combination of deep learning (subsymbolic) and symbolic reasoning (for communication and rationalization), and evaluate the performance of our proposed approach in terms of, e.g., an ability to achieve temporal understanding (an essential requirement for anticipation) across multiple layers) and an ability to explain/rationalize “deep” processing".

To perform information fusion at varying levels of semantic granularity, the choice of the *models used to represent the world, and the information agents* is extremely important. This distributed collection and processing of information in a coalition setting also has similarities to the manner in which the neocortex in the human brain does processing of the various sensory inputs for decision-making and intelligence. Neuroscientists observe that the processing units in the neocortex are independent of the sensing modality and the signals that are fed into them. For example the same learning structures are responsible for handling both acoustic signals (from our ears) and visual signals (from our eyes). When we encounter a new phenomenon or learn a new concept in the real world, the existing structures in the brain are trained to learn the pattern associated with the new concept and then identify its occurrence in an incoming signal. Motivated by the above we identify the following properties that are necessary for any chosen model to operate in a dynamic coalition setting:

- **Heterogeneity**: The selected models should be able to handle different types of information sources (human and machine) together with semantic models.
- **Composability:** The models must operate in a distributed setting, and be resilient to changes in coalition membership (i.e., joining and leaving of coalition members, information sources).
- **Hierarchical Learning:** Should be able to automatically learn concepts from input data streams through iterative generalization to higher-level semantic abstractions. The learning should support both feedback and feed-forward message passing and have the ability to incorporate prior information at any level of the hierarchy in the form of knowledge-bases available across the coalition.
- **Prediction Capabilities:** The model should provide ability to project forward in time and explore possible sequences of future states, under constraints of timeliness of the chosen action and resources available to a coalition.

There are several current “deep learning” methods that satisfy subsets of the above properties and are gaining success in pattern recognition and classification problems, for example, recognizing objects in images. These methods apply hierarchical learning to automatically identify high-level abstractions in data using complex structures composed of non-linear transformations of data. HTM is one such promising general-purpose deep learning technique that can operate in both supervised and unsupervised modes. It is designed to imitate the human neocortex and is naturally distributed. Compared to other deep-learning methods, HTM integrates the notion of time more tightly into the learning process and patterns are obtained by pooling data over both space and time (using spatial and temporal pooling algorithms). Furthermore, by learning sequences of patterns occurring in close proximity in time, HTM is able to predict future states and explain the effect of an external action, which is important to attaining ASU. Organized as a hierarchy of regions, a HTM provides a uniform learning mechanism independent of the spatio-temporal data type used as an input, by internally converting each signal into a sparse-representation that is then used to learn patterns both in space and time. HTM also maintains a history of prior observations in the form of states and their transitions (see Figure P6-1(c)).

Currently, the learning in HTMs occurs in both feed-forward and feedback directions through message exchanges between the top and bottom layers. HTMs do not allow additional training inputs to be applied to the in-between layers independently, i.e., input signals are fed only to the lower and topmost layers of the HTM. We propose that augmenting the learning algorithm to allow semantic inputs from knowledge bases may increase the flexibility and composability of systems such as HTM. Finally, HTM is still in its formative stage, and while there is empirical evidence regarding its ability to perform certain tasks the theoretical underpinnings are not all understood. Thus, one key research question for Task 6.2 is: *can a "deep learning" method like HTM be successfully applied to higher levels of abstraction in terms of situational understanding?* For example, progressing from recognizing objects such as SUVs to recognizing that certain kinds of SUV movements are characteristic of smuggling and, going further, that particular kinds of smuggling may be linked to insurgent activity? Such an approach would be the basis for our hypothesis that the fusion of deep learning approaches with agile knowledge representation and reasoning could form the basis for a powerful and flexible system to support ASU in a coalition context.

We envisage a combination of recognition and reasoning being important in dealing with the complete set of abstraction levels, with "deep learning" methods continuing to improve at the lower levels, and potentially move upwards, and reasoning approaches being employed at the higher levels, where inferences and predictions need to be explainable and open to close inspection by humans. State of the art deep learning approaches tend to yield models that are not transparent, leading to significant brittleness. Therefore a second key research question for Task 6.2 is: *how can "deep learning" be combined with reasoning approaches to cover the complete set of abstraction levels for ASU?*

Task Milestones:

Date	Description
Q1	− Establish ASU baseline for comparison - the hypothesis is that our Deep Learning for Multi-Layer Situational Understanding (DLMLSU) approach will outperform the baseline. Establish research method and ways of working. Establish requirements for a representative set of ASU problems / test cases / datasets.
Q2	− [Jointly with Task 6.1] Collect ASU problems, test cases, datasets etc. Contribute these to the DAIS Experimental Framework and make them open source where possible.
Q3	− Define initial DLMLSU model or choice of candidate models for testing on ASU problems.
Q4	− Define the roadmap for DLMLSU defining principles, baseline, hypotheses and open problems.
Q5	− Perform initial evaluation of Q3 DLMLSU model(s) on Q2 cases/datasets. Contribute reproducible results to DAIS Experimental Framework and make open source where possible. Extend Q4 DLMLSU roadmap paper with results and submit for publication.

Linkages

Project 1 addresses the problem of creating software defined coalitions: how to compose a coalition of a set of nodes, respecting resources and data management objectives. Task 6.1 will directly complement this project by providing a specification of what information must be exchanged between the agents to achieve ASU. As such, it provides a set of constraints on the network which must be respected. Simon Julier (UCL) will collaborate with Miguel Rio (UCL) of Project 1 to ensure that both teams are appraised of progress in different projects.

Project 2 focuses on the design of a novel policy management framework that can support generative policy models for controlling access to information in highly dynamic coalition environments. The resulting high degree of autonomy, due to each member generating their own policies, would require the framework to provide mechanisms that ensure consistency, coherence, and flexible enforcement of the policies. Furthermore, these policies will also govern the type and fidelity of the information being shared between coalition members, which in turn will effect both the quality of the models learned and their use for performing ASU. Alun Preece (Cardiff) will work with Project 2 team via Supriyo Chakraborty (IBM-US) to explore the strong linkages between the two projects.

Project 3 considers the problem of supporting agile composition in a coalition environment. In this paradigm, data and computational capabilities move together through a network, supporting dynamic and distributed computation. Tasks 6.1 and 6.2 will directly inform the design of this architecture. The concepts from Task 6.1 will be used to explore what data needs to be mobile and can move across the network. Task 6.2 defines the analytics which are required for ASU. Simon Julier (UCL) will collaborate with Mark Herbster (UCL) of Project 3 to ensure that both teams are appraised of progress in different projects.

Project 4 focuses on the development of models for understanding the activities of groups external to the coalition, including physical crowds and online social networks. Elements of these models will inform ASU of situations involving those kinds of groups and activities. Alun Preece (Cardiff) will collaborate with the Project 4 team via Roger Whitaker to ensure that this linkage is exploitable to the benefit of both projects.

Project 5 provides a way to dynamically compose the analytics from the appropriate services and, together with Project 3, will provide insights into the development of a self-organizing declarative platform for analytics. This platform will underpin the analytics layer for the system that will perform information fusion for anticipatory situational understanding in Project 6. The strong linkage between Projects 5 and 6 will be led and coordinated by Alun Preece (Cardiff) and Ramya Raghavendra (IBM-US) to ensure that the necessary set of analytic services required for Project 6 are developed.