Distributed Analytics and Information Sciences (DAIS)

International Technology Alliance (ITA)

Initial Program Plan (IPP)
Summary of Accomplishments

21st September 2016 to 15th January 2018

Agreement W911NF-16-3-0001

During the Initial Program Plan (IPP) period of September 2016 to January 2018, several scientific advances in areas related to distributed analytics were made by researchers of the alliance. During this period, the alliance researchers published 94 articles in peer-reviewed conferences and journals, of which 85 (90%) were collaboratively authored across multiple organizations. About a third (32%) of these papers had government co-authors. Furthermore, alliance members organized various conferences, workshops and symposia related to the topic of distributed analytics. Some of the selected technical advances made during the IPP period are enumerated below:

A team of researchers from Yale University, IBM U.S. and University College London created new distributed control architectures to bring SDN (software defined networks) to mobile networks. While SDN is typically considered a centralized approach and mobile ad-hoc networks rely on use of distributed control mechanisms, the team showed a hybrid design that takes the best of the two paradigms: (i) global network view and control programmability of SDN and (ii) robustness of distributed protocols. This can provide efficient distributed control in tactical coalition environments. The pros and cons of the proposed methods were studied and highlighted in the SDN prototype by use of off-the-shelf mobile devices[1].

Researchers at Imperial College, IBM U.S., Yale University and U.S. Naval Research Lab investigated and acquired a new understanding of how the SDC (software defined coalition) performance is related to inter-domain synchronization levels and network structural properties. Specifically, they established analytical methods and performance bounds for various degrees of synchronization information shared among SDC domains, the accuracy of which was confirmed by extensive experiments based on both real and synthetic networks. These results serve as a fundamental guidance for SDC protocol designs or upgrades for significance performance improvement in future coalition networks.[2]

A team of researchers at University of Massachusetts, Amherst; Imperial College and U.S. Army Research Lab investigated the performance tradeoffs among the three factors of communication bandwidth consumption, in-network processing (computation) and data caching in coalition networks. New distributed algorithms and techniques were developed to achieve the optimal system performance tradeoffs in terms of energy efficiency, latency and cache hit probability, while delivering the specified quality of information. These results reveal that the energy efficiency under the optimization framework considering communications, computation and caching improves by as much as 88% when compared to any of the two factors alone in the future SDC infrastructure.[3]

A team of researchers from Purdue University, Defence Science and Technology Laboratory and IBM U.S. proposed a new paradigm for letting devices in a coalition environment generate their

[1] K. Poularakis, Q. Qin, E. Nahum, M. Rio, and L. Tassiulas. "Bringing SDN to the Mobile Edge." IEEE Smart World Congress DAIS Workshop, San Francisco, August 2017. http://sl.dais-ita.org/science-library/paper/doc-1392
[2] https://dais-ita.org/system/files/1712.04161.pdf
[3] F. Zafari, J. Li, K. Leung, D. Towsley and A. Swami, Optimal Energy Tradeoff among Communication, Computation and Caching with QoI-Guarantee, Submitted to ICDCS 2018. https://dais-ita.org/node/2162

own policies, created an architecture for the same, and showed that it works for security and access control in coalition tactical environments[4]. A keynote talk on generative policies was given at ACM SACMAT 2017 Conference.

DAIS ITA researchers organized a workshop (PADG) on generative policies for managing big data in federated environments in Boston in December 2017.[5] The successful workshop (PADG) should be the first in a series of several workshops on this topic.

A team of researchers from Purdue University, Defence Science and Technology Laboratory and IBM U.S. came up with the concept of quality of policies, and proposed analysis techniques that could assess the quality of policies using provenance techniques that collect data about actions executed by users in the context of business workflows.[6]

Researchers from University College London, U.S. Army Research Labs and IBM U.S. designed a novel algorithm to distribute the data required for distributed analytics on machines in a bandwidth limited network, without violating the bandwidth and storage constraints, while ensuring that the data transfer deadlines are met. These algorithms can improve efficiency and responsiveness of data analysis during coalition operations.[7]

Researchers from IBM U.S., Imperial College, U.S. Army Research Lab, and Pennsylvania State University designed and analyzed an algorithm to perform the gradient descent in distributed machine learning without nodes having to share all their instances with each other. The method is based on (relatively infrequent) computations of averages of vectors stored on nodes, which is adaptively controlled for obtaining optimal learning performance. Researchers from University College London and IBM U.S. then designed a method to efficiently perform these averages thus significantly decreasing the time required to convergence for the gradient descent. These algorithms allow different agents in the coalition to help each other learn similar tasks without having to share all their training data with each other.[8]

Researchers from IBM U.S., U.S. Army Research Lab, Raytheon BBN Technologies, Pennsylvania State University, IBM U.K. and Defence Science and Technology Laboratory studied the problem of flexibly, dynamically, and adaptively moving, positioning, and instantiating computing tasks and data in federated, distributed edge systems. They did this by modelling code and data

---

[4] E. Bertino, S. Calo, M. Touma, D. Verma, C. Williams, and B. Rivera. "A cognitive policy framework for next-generation distributed federated systems: concepts and research directions." In IEEE 37th International Conference on Distributed Computing Systems (ICDCS), 2017, pp. 1876-1886.

[5] https://dais-ita.org/padg2017

[6] E. Bertino, A. Jamal, S. Calo, C. Makaya, M. Touma, D. Verma and C. Williams, Provenance-based Analytics Services for Access Control Policies, IEEE Services Conference, 2017. http://sl.dais-ita.org/science-library/paper/doc-1234

[7] T. Tuor, S. Wang, T. Salonidis, B. Ko and K. Leung, Distributed Machine Learning at Resource-Limited Edge Nodes, Demo at INFOCOM 2018.

[8] S. Wang, T. Tuor, T. Salonidis, K. Leung, C. Makaya, T. He and K. Chan, "When Edge Meets Learning: Adaptive Control for Resource-Constrained Distributed Machine Learning", Proc of IEEE InfoCom 2018. https://dais-ita.org/node/854

fragments as dynamic particles on the vertices of a graph, using physics-style potential functions to give forces on the particles: moving them apart or closer together as necessary to ensure that relevant code and data for an application can come together quickly for execution. These algorithms can improve the efficiency and responsiveness of data analysis during coalition operations. [9]

Researchers from Cardiff University, Defence Science and Technology Laboratory and Pennsylvania State University created an innovative approach to model the evolution of human groups on a generative basis, to anticipate their possible mutation, which enables exploration of tensions associated with affiliation to multiple groups and the influence on inclusion and exclusion of individuals. The insights could be useful in analyzing situations in disturbed locations.[10]

Researchers from Cardiff University, Yale University and Defence Science and Technology Laboratory invented a new mathematical model for the mutation of social groups, with a focus on group fracture, where individuals leave groups of which they are members.[11] These models incorporate a selection of social and psychological theory which describes these phenomena as a way to understand their interplay, and describe the trade-offs and challenges.

Researchers from Pennsylvania State University and University of Massachusetts, Amherst developed a new approach to analyzing social network behavior. Using motifs[12], as originally proposed in biology, it is possible to analyze interactions in human networks that are too complex for traditional structural social network analysis. These interactions are represented as critical substructures which give insights into latent structure and bridging. The concept is being explored further in the context of reciprocity with researchers at Cardiff University.

Researchers from Pennsylvania State University, IBM U.S. and the University of Southampton were the first to study the problem of optimal provisioning of resource-intensive delay-sensitive analytics services from the edge of a network by considering service placement and request scheduling jointly under both additive (communications, computation) and sub-additive (storage) constraints[13]. In the homogeneous case, where all nodes and all services have the same characteristics, the team developed a constant-factor approximation algorithm; for the general case efficient heuristics were developed. In both cases, the performance achieved was near optimal in terms of number of served requests.

---

[9] B Ko, B Kraczek, T Salonidis, P Basu, K Chan, T La Porta, A Martens and J. Lambert, Physics-inspired Models for Agile Code and Data in Federated Edges, IEEE Smart World Congress DAIS Workshop, San Francisco, CA, August 2017. http://sl.dais-ita.org/science-library/paper/doc-1411

[10] R. Whitaker, D. Felmlee, D. Verma, A. Preece, G. Williams, "From Evolution to Revolution: Understanding Mutability in Large and Disruptive Human Groups" Proc. SPIE DSS Fusion, Anaheim, CA, April 2017.

[11] S. Eshghi, G. Williams, G. Colombo, L. Turner, D. Rand, R. Whitaker and L. Tassiulas, "Stability and fracture of social groups", Allerton Conference Allerton Conference on Communication, Control, and Computing, October 2017,

[12] https://dais-ita.org/node/2146

[13] T. He, H. Khamfroush, S. Wang, T. La Porta and S. Stein, "It's Hard to be Social: Joint Service Placement and Request Scheduling for Social Edge Applications", submitted to ICDCS 2018, https://dais-ita.org/node/1443

Researchers from Cardiff University, IBM UK and IBM U.S. pioneered the use of a structured associative memory models called vector symbolic architectures, for representing and orchestrating complex decentralized workflows. This approach allowed use of Artificial Intelligence technologies to enable rapid self-organization and composition of complex services in a distributed manner[14].

Researchers from Cardiff University, IBM UK and U.S. Army Research Lab created a new approach to exploit human-machine coalitions for situational understanding based on Subjective Bayesian Networks, i.e., Bayesian Networks with imprecise probabilities[15]. The method learns effectively from sparse data and, coupled with conversational interfaces, can be used in real-world coalition decision making.

The use of machine learning techniques for real-world operations is usually constrained by lack of good training data. Researchers from University of California, Los Angeles and IBM U.S. defined an approach to use deep learning techniques to generate property-preserving synthetic sensor data[16], which can significantly ease the challenges involved in such training data acquisition.

During the IPP the DAIS ITA program contributed 94 peer reviewed external publications, including 6 journal papers and 4 best paper/poster awards. These along with the 55 internal publications were collectively written by 170 authors from 35 organizations within and beyond the DAIS ITA alliance, with a strong collaborative profile for most of the publications. In addition to this there are about 35 publications "in press" (submitted or accepted but not yet published and therefore not yet in science library). For additional details visit the DAIS ITA Science Library which contains all currently published publications, authors and organizations involved http://sl.dais-ita.org/science-library.

[14] G. Bent, G. De Mel, S. Rallapalli, C. Simpkin and I, Taylor, "Decentralized Microservice Workflows for Coalition Environments", IEEE Smart World Congress DAIS Workshop, San Francisco, CA, August 2017.
[15] D. Braines, A. Thomas, L. Kaplan, M. Sensoy, M. Ivanovska, A. Preece and F. Cerutti, "Human-in-the-Loop Situational Understanding via Subjective Bayesian Networks", International Workshop on Graph Structures for Knowledge Representation and Reasoning, Melborune, Australia, Aug 2017, http://sl.dais-ita.org/science-library/paper/doc-1453
[16] M. Alzantot, S. Chakraborty and M. Srivastava, "SenseGen: A Deep Learning Architecture for Synthetic Sensor Data Generation", First IEEE International Workshop on Behavioral Implications of Contextual Analytics, Kona, Hawaii, March 2017.