"Final Report – Real Time Network Assessment "

12 July 2014

Sponsored by

Office of Naval Research

In response to: ONR BAA-10-001

Issued by Office of Naval Research

Grant No. N00014-10-1-0637

Name of Contractor: Carley Technologies, Inc. (CTI)

Principal Investigator, Project Scientist, or Engineer: Dr. Kathleen Carley

E-mail Address: kathleen.carley@carleytech.com

Business Address: 1924 Glen Mitchell Road, Sewickley PA 15143 USA

Phone Number: 412-741-2002

Effective Date of Contract: 15 April 2010

Contract Expiration Date: 14 April 2013

Reporting Period: 15 April 2010 through 14 April 20113

"The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either express or implied, of the Office of Naval Research or the U.S. Government."

"**Approved for Public Release; distribution is Unlimited.**"

20130723013

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
12-07-2013	Final Report	Apr 2010 - Apr 2013

4. TITLE AND SUBTITLE
Final Report - Real Time Network Assessment

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-10-1-0637

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Carley, Kathleen M.
Malloy, Eric
Carley, Larry, R.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Carley Technologies, Inc.
1924 Glen Mitchell Road
Sewickley, PA 15143-8871 USA

8. PERFORMING ORGANIZATION REPORT NUMBER
ONR-06374

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
Attn: Rebecca Goolsby, ONR Code 34
One Liberty Center
875 N. Randolph Street
Arlington, VA 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; distribution is Unlimited.

13. SUPPLEMENTARY NOTES
Final Report for ONR Grant N000141010637

14. ABSTRACT

Network analytic tools are valuable in a large number of situations of relevance to the Navy. These include support for littoral activity, MISO operations, HA/DR missions, intelligence and targeting activities. While network tools have advanced greatly, they are still relatively unused in real time situations. Current technology is limited in that the behavior cannot be observed and characterized in real time; and, even when it can be observed, most analysts have little guidance to help them realize what they are seeing. Analysts need the ability to assess and reason about tactics and behavior using network analytics applied to evolving and real time data for complex situations. This requires not just real time data and visual analytics but also it requires that the analysts are able to understand how to interpret network metrics, and when given temporal data, how to identify change, particularly in the face of messy and incomplete data. Thus, what is needed is both a real time assessment environment and a process for rapidly training / retraining military analysts in network analytics so that they can make timely and effective decisions in this complex environment.

15. SUBJECT TERMS
Dynamic Network Analysis, Training, Real Time, Assessment, Social Network, Human Terrain

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Kathleen M. Carley
None	None	None	UU	11	19b. TELEPHONE NUMBER *(include area code)* 412-741-2002

I. P.I.: Dr. Kathleen M Carley

Organization: Carley Technologies Inc., (CTI)

Grant Number: N00014-10-1-0637
Award Title: Real Time Network Assessment

II. Scientific and Technical Objectives

Network analytic tools are valuable in a large number of situations of relevance to the Navy. These include support for littoral activity, MISO operations, HA/DR missions, intelligence and targeting activities. While network tools have advanced greatly, they are still relatively unused in real time situations even though real time network assessment is key to information dominance. From an operational perspective the current technology is limited in that the behavior cannot be observed and characterized in real time; and, even when it can be observed, most analysts have little guidance to help them realize what they are seeing. Analysts need the ability to assess and reason about tactics and behavior using network analytics applied to evolving and real time data for complex situations. This requires not just real time data and visual analytics but also it requires that the analysts are able to understand how to interpret network metrics, and when given temporal data, how to identify change, particularly in the face of messy and incomplete data. Thus, what is needed is both a real time assessment environment and a process for rapidly training / retraining military analysts in network analytics so that they can make timely and effective decisions in this complex environment.

Our goal was to enable automated assessment of critical situations in teal time from a network analytic perspective and to reduce the training time needed to use such a capability. A real time network assessment systems, network training system, and the process for supporting the training of military personnel to utilize these techniques and correctly interpret the measures is critical to many Navy missions.

III. Approach

There are two primary tasks. Task 1 is the development of a table-top prototype of a real time data capture and analysis system. This task was not fully funded. Thus the scope was reduced to identifying necessary components for a real time system, identifying a workflow for those components, and to testing some of those components. Task 2 is the Training and development of training tools for dynamic network analysis.

A. Real Time Network Assessment Prototype

The basic goal was to develop a prototype for a real time data capture and analysis system. This was reduced to the following sub-tasks:

- Design a system and identify potential tools to integrate
- Demonstrate a simple system
- Conduct a feasibility assessment of data storage, maintenance, and integration requirements

- Test a web-based data feed
- Design a solution for multi-media that supports converting the information into networks

Our approach was to identify the basic components, devise a table top prototype, identify a workflow, take part in experiments to assess the viability of this workflow using a human-in-the-loop approach with lower volume canned data. The workflow was tested using humans to move and convert data thus only examining analysis in near real time. Tests were conducted at Carley Technologies with canned data, during ONR experiments using data on Hurricane Sandy that was given out faster than real time, and in training at EUCOM using data collected on Libya and Egypt in real time. The testing process was to run each tool in the sequence thought best for the workflow and to also observe other analysts using other workflows.

B. Training and Training Evaluation

In this case there were several basic sub-tasks, that based on funding revisions were:

- Develop basic training materials for Dynamic Network Analytics
- Train end users on a Dynamic Network Analytic system
- Evaluate training technology and training needs for a Dynamic Network Analytic system
- Evaluate needs for training material for a Real Time Network Analysis system

For Task 2, directed interviews with end users and observation of users being trained were used to assess training. Type of material, focal content, and desired content were considered. Outcomes examined included user preference, training time, ability to perform task, frustration, and attention.

Training was conducted for groups identified by the sponsor, and opportunistically for other groups in the DoD. All training activities were used as forums for collecting data on the training process, strategy, materials, and methods.

IV. Accomplishments

A. Task 1: Real Time Network Assessment Prototype

We demonstrated the feasibility of linking near real time network analytics to mashups and web-based informational sites that use crowd sourcing by working with GDIT, ASU, and CMU at ONR experiments using data from Hurricane Sandy.

A real time network assessment prototype was developed and tested at: Haiti, CMU, EUCOM and NORTHCOM. This system had only portions of the overarching design. Key features that were included in this prototype were an email, facebook page importer, importer for twitter, key entity extractor, and the full ora-netscenes for network analytics and visualization. This testing was used to identify and confirm what workflows were needed, and to identify and confirm that the list of component technologies made sense. The result was a confirmed list of component technologies for network analytics, a basic workflow, and a list of technologies for use in an over-arching system. We note that an overarching system will grow over time and the set of technologies listed should be viewed as a minimum set.

The basic technologies needed, other than the network analytics include: data capture, translation, deduplication, de-identification, fusion, sub-selection, storage, workflow management, workflow creation and discovery, network extraction (entity and link), image processing, and crowd-sourcing for error recovery. Of these, the technology most requested was improved automation for dynamically extracting meta-networks from source data. Based on the literature and interviews with end users a set of sub-technologies and the problems with those technologies that limited their applicability in a real time scenario were identified. See Table 1.

Table 1: Possible Automated Dynamic Meta-Network Extraction Technologies

Challenge	Key Relevant Technologies	Key Difficulties
Data cleaning	Automated typo fixing De-duplication Spelling corrections.	Node attributes
Entity identification	Machine learning e.g., CRF for entity extraction Statistical patterns (e.g., LDA and LSA) Pattern matching (e.g. tracking known bi-grams)	Identification of new n-grams (location names, organizations, and events) activities
Entity disambiguation (includes combining similar concepts into single node)	Stemmers Thesauri application Network position Statistical common patterns Pronoun identification Theme assessment (e.g., using LDA or LSA)	Expertise, beliefs/attitudes
Entity classification	Ontology thesauri application Part-of-speech mapping Statistical clustering with seeds Supervised machine learning techniques	Expertise, resources, beliefs/attitudes
Entity attributes	Predefined thesauri Part-of-speech mapping	All attributes
Entity adaptation	Temporal mapping Change in attribute* Part-of-speech mapping (use adverbs and adjectives)*	All attributes
Link identification	Proximity based link extraction Social logic based link extraction Syntactic parsing	Links to locations Group membership Links between expertise, resources and activities
Link attributes	Extraction frequency Inter-document consistency Valence or directionality assessment*	Valence
Link adaptation	Part-of-speech mapping (use adverbs and adjectives)*	Strength and Valence
Certainty assessment	User defined Inter-document consistency*	All
Coding choices	Logging	No issues

Additional drill down was conducted for dynamic network analytics and visualization for real time network assessment. The core new capability desired is dynamic metric. The development of these metrics will require fundamental basic research, are new metrics for incremental measurement of centralities and groups, alert functions for change, and new metrics for emergence and atrophication. New desired features, not present in tools examined (ORA-NetScenes, UCINET, Palantyr, Analyst

Notebook, Pajek) are: dynamic metrics, network movies with replay, top spot identifiers, auto-group comparison, and an analysis wizard.

Further testing using ORA-NetScenes revealed that the following features were most valuable for real time network assessment: comparative analytics, change assessment, change reports, change detection and fourier analysis, and heat maps for temporal frequencies. Features most requested, other than those listed above, were sequencing of data, movie like movement through temporal networks, inclusion of text mining in ORA-NetScenes, and importers for email, facebook and twitter. Versions of each of these were developed and included in ORA-NetScenes. These were then tested in near real time environments.

B. Task 2: Training and Training Evaluation

The majority of research conducted was on conducted basic research on how training should be done for network analytics, identifying core training needs, and developing and testing training materials. We tested features of the Real Time Network Assessment prototype system as well. We identified strengths and weaknesses in current training materials. We developed some new material. We assessed training needs. We conducted training.

Carley Technologies Inc. (CTI) aka Netanomics supported and provided training both in person and through telecons to personnel at NORTHCOM, EUCOM, PACOM, NPS, JIATF and USMA.

CTI review the literature on training and existing training programs focused on training people to use software. From this review a list of training technologies was constructed. Then through interviews and observation the relevance of each type of training technology was assessed. The results are in Table 2.

Table 2. Relevance of Types of Training Technologies

Material	Value to End User when First Learning	Useful to End User on Their Own
Tool Tips	High	High
Quick Start Guide	High	Low
Follow Along Lectures	High	Low
User Guide	Low	Medium
Wordy Interface	Medium	High
Powerpoint	Medium	Medium
Book	Low	High
Group Forum	Low	High
Learning by doing	Medium	High

CTI developed or improved many of these critical techniques. Then conducted testing to determine whether or not they a) improved the perception of the value of the training, b) decreased training time, and c) improved ability to do 3 fundamental tasks necessary in Real Time Network Analytics - identify key entities, enter data and visualize results. See Table 3.

Table 3. Impact of Types of Training Materials

Material	Implemented	Reduce Training Time	Improve Ability	Improve Perception of Utility
Tool Tips	Yes	Yes	Yes	Yes
Quick Start Guide	Yes	Yes	Neutral	Yes
Follow Along Lectures	Yes	Yes	Yes	Yes
User Guide	Used Existing	No	Yes	Yes
Wordy Interface	No			
Powerpoint	Used Existing	No	Neutral	Yes
Book	Draft	No	Neutral	Yes
Group Forum	Yes	No	Yes	No
Learning by doing	Yes	No	Yes	Yes

We found through this assessment that the type of materials most desired that did not exist in some form were follow along videos, formatted forum blogs for help, lessons done with data samples similar in kind to what they would use in practice, and a detailed book for more in-depth and theoretical training. In general, users preferred steps to be automated rather than to require training. The key example here is that when looking at a network, the end users wanted the network visual to be provided with tips that auto-identified key features of the data that an expert would look for, so that they did not have to remember what to look for. CTI assessed training material to see what worked, and to identify where time was spent during training, and how much time people wanted to spend on that type of training. The key finding was that most of the training time was being spend on routine tasks rather than on the more creative analytic tasks. This created competent users, but not informed analysts. End users wanted more time to be spent on the less routine tasks that required more analytic capability. See Table 4.

Table 4. Mismatch in Training Delivered and Training Desired.		
Feature	**Percentage Training Time**	**Desired Training Time**
Data Import	40%	10%
Visualization	20%	10%
Analysis	35%	50%
Interpretation	5%	30%

In addition CTI collected information from a subset of these personnel on what types of training material were desired. The most desired types of training material are: 1) follow along videos, 2) formatted forum blogs, 3) multiple data samples, 4) book, and 5) and data tips that identify key features in the data.

Carley Technologies Inc. (CTI) aka Netanomics supported multiple experiments – two using data from Hurricane Sandy and one using Kenyan election data. Our role was to provide training, to identify what training material worked and did not work in these contexts, and to identify the features needed by the DoD in a real-time network assessment technology. The first activity was an ONR sponsored event using data from Hurricane Sandy. In this case a CMU-Netanomics team briefed personnel from NORTHCOM, SOCOM, NDU, and others on the use of ORA-NetScenes for HA/DR events. The second activity was at EUCOM. EUCOM requested a workshop and training on social media and network analytics, using ORA-NetScenes (CMU-Netanomics), held September 10-12, 2012, at Patch Barracks in Germany. Netanomics supported the training to analysts on how to use these tools and provided training material and the COTS military version or ORA-NetScenes.

During the training the Libyan embassy was attacked. As a class exercise, the research team demonstrated how that event could be analyzed in real-time, teaching the analysts/students how to use these tools on a live, dynamic event. Both analysts and members of the ASU-CMU-Netanomics team collected and analyzed data on this event that leveraged on-going collection on crisis events in the countries of Arab Spring together with streaming data from Twitter and Lexis-Nexis-tagged news articles. This experience has been written up, was sent to a conference, has been accepted and is forthcoming:

Kathleen. M. Carley, Jürgen Pfeffer, Huan Liu, Fred Morstatter, Rebecca Goolsby, Forthcoming 2013, Near Real Time Assessment of Social Media Using Geo-Temporal Network Analytics, In Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), August 25-28 2013, Niagra Falls, CA.

V. Major Problems / Issues

Immediately after award, there was a change in DCMA that caused the DCMA contact listed on the ONR contract to be incorrect. The Company performed under the award during 2010, but was not able to receive reimbursement until 2011. Although the expenditure rate was on plan, the reimbursement rate was greatly delayed. All reimbursement requests filed through WAWF were never processed and never reached the PM. Finally, in early 2011, we were able to determine the problem and put through invoices with the correct DCMA contact. Since then, we have been able to invoice regularly on the project. The

final amount spent is slightly less than the total allocated because we were planning on a no-cost extension due to the delay in receiving funds. We received word that the NCE would not be granted late enough that we slightly underspent on the Grant Award.

VI. Technology Transfer

Attended diverse ONR sponsored experiments, and the IV2 experiment, and gave a briefing to Quantum Leap. These interactions involved: 1) Carley Technologies Inc. supporting the CMU team by providing the COTS version of ORA-NetScenes, 2) Carley Technologies Inc. developing specialized demos; and 3) Carley Technologies Inc. working with and interviewing end-users to identify needed and un-needed features in a real-time network assessment tool. End-users included representatives from NDU, NPS, USMA, PACOM, EUCOM, and NORTHCOM in the areas of intel, MISO, HA/DR, and targeting. The experience and data gathered was directly relevant to improving ORA-NetScenes and will support future transitions.

The work under this grant led to the following commercialization strategies that will be pursued in the future:

1. Knowledge gained from end-users defined a core minimum set of technologies needed in a real-time network assessment framework. This led to revisions in the ORA-NetScenes tool that will increase its utility and support sales as it opens up a new group of customers – those with dynamic data.
2. As part of this work we evaluated existing training materials and identified what training materials were desired by end-users. This information informs two different commercialization strategies: a) the sale of training as a service and, in particular, materials for training trainers, and b) increases the sale of the Ora-NetScenes and a related family of products by providing a more comprehensive system with integrated training materials.
3. Knowledge gained performing research under this award suggested the need for support technologies for de-identification of data. Future work will lead to a de-identification technology, which will be commercialized as a separate related tool.

C. Licensed or Patented Technology

None

D. Transitions to Obtain Funds

None

E. If Technology Transfer occurred without such interactions, please describe that as well.

Ideas developed under this research project led to improvements that were incorporated into the commercial Social Network Analysis Tool ORA-NetScenes.

F. Describe any future plans you have for Technology Transfer of ONR-funded R&D

We plan to transfer additional ideas developed under this research project into the commercial Social Network Analysis Tool ORA-NetScenes.

We plan to publish journal and conference papers to disseminate the knowledge developed under this research award.

VII. Foreign Collaborations and Supported Foreign Nationals

None

VIII. Productivity

C. Refereed Journal Articles

None.

D. Non-Refereed Significant Publications

None.

E. Books or Chapters

None.

F. Technical Reports

None.

G. Workshops and Conferences

Conference Proceedings:

Kathleen. M. Carley, Jürgen Pfeffer, Huan Liu, Fred Morstatter, Rebecca Goolsby, Forthcoming 2013, Near Real Time Assessment of Social Media Using Geo-Temporal Network Analytics, In Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), August 25-28 2013, Niagra Falls, CA.

Organized Sessions:

None.

Workshops and Presentations:

Carley, LR "Real Time Network Assessment," at 2013 ONR Program Review, Carnegie Mellon University, Pittsburgh, PA. May 2013.

Posters:

None.

H. Patents

None.

I. Awards/Honors

Kathleen M. Carley was appointed a Member of the NAS/NRC Committee on NGA Workforce Assessment
Kathleen M. Carley was appointed aMember of the NAS/NRC Committee on Math Digital Library
Kathleen M. Carley was appointed a Member of the NAS/NRC Committee on Massive Data Analysis
Kathleen M. Carley was Appointed to DHS HSSTAC, as a special government employee
Kathleen M. Carley gave an Invited Plenary: "Crisis Mapping: Big Data from a Dynamic Network Analytic Perspective," World Summit on Big Data and Organization Design, Paris, Fr.
Kathleen M. Carley gave an Invited Plenary: "Dynamic Network Approach to Health Surveillance," PACOM S&T 2013, Honolulu, HI

IX. Award Participants

Dr. Kathleen M. Carley (PI)

Dr. L. R. Carley

Mr. Eric Malloy