Idaho Incubation Fund Program

Final Report Form

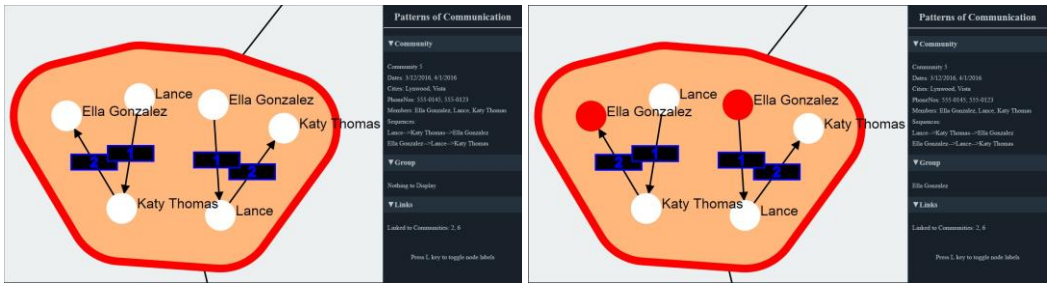
OSP Proposal No.	**7855**
Name:	**Gaby Dagher**
Name of Institution:	**Boise State University**
Project Title:	Malicious Community Extractor (MACE): A Robust Toolkit for Unmasking Criminal Networks

Information to be reported in your final report is as follows:

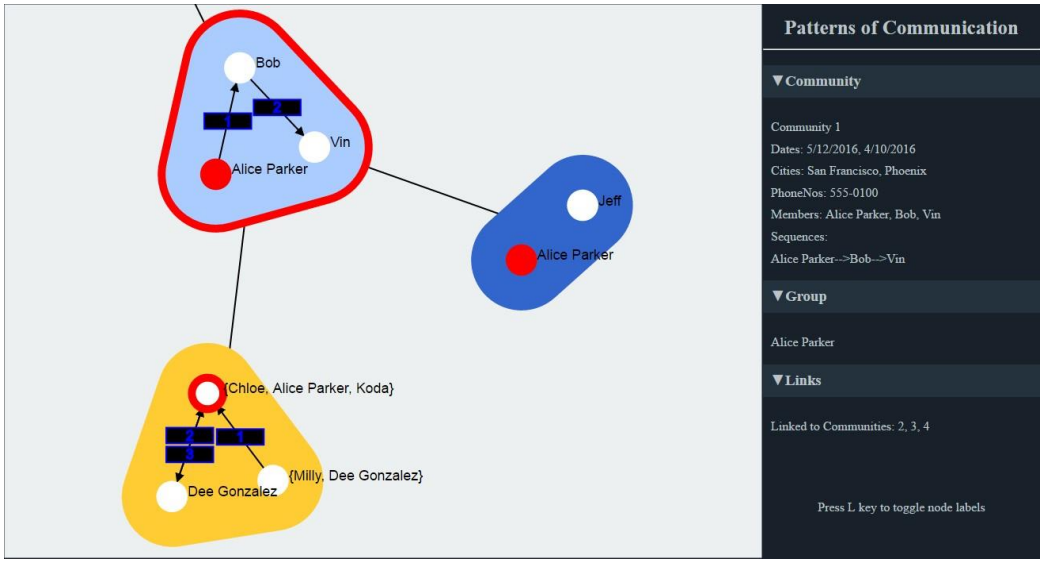
1. Provide a summary of overall project accomplishments to include goals/milestones met, any barriers encountered, and how the barriers were overcome:

   To automate criminal network detection from large data sets, we built MAlicious Community Extractor (MACE) toolkit. It is designed using a novel sequential pattern mining algorithm that efficiently identifies potential criminal networks while providing an interactive interface that graphically displays communication patterns and relationships within and between the criminal networks. More specifically:

   - To extract criminal communities and identify relations between members and communities, we propose two algorithms that focus on an accurate extraction process from any type of document. Our solution allows investigators to easily sort through the documents found in a suspect's computer. With extraction of criminal communities, it's important to know how the criminal networks interact, and our solution helps the investigator see how separate communities can be related to each other.
   - Extracting active communities which are communities that are found to interact frequently throughout the entire document set. Each active community can contain one or more order of interactions which allows an investigator to observe how the active community interacts. This can be important because the investigators observations can find which member interacts the most, which member starts the communication, and how the communities interact.
   - With the extraction of communities in a conglomerate of information, it's important to display the information in an efficient, effective manner. Our solution outputs the criminal network analysis in an organized display that allows the investigator to clearly see prevalent communities, active communities, community relations, patterns of communication, and extracted information about these communities such as places, phone numbers, or topics.



These figures are close ups of a single selected active community. The links show the pattern of communication between the members, and the numbers show the order of the communication with the information about the community on the right which shows the information extracted from the documents, and the sequence pattern. The figure on the left is different from the community on the right because a member of the community, Ella Gonzalez, is selected in the figure on the right. Because it is selected, under the group tag in that figure, Ella Gonzalez's name can be seen, signifying that she is the member selected.



Three communities can be related to each other through direct and indirect relations. The first community contains the same individual, Bob, as the second community which shows a direct relation. The first community also has a direct link to the third community because they both contain Alice Parker as a member. Through this, it can be seen that the second community and the third community have an indirect relation because they are both are directly related to the first community.

#	PROJECT TASKS	Q1	Q2	Q3	Q4
	Milestones:		A	B	C
1	User-Centered Interface	X	X		
2	Quality Assurance + Benchmarking		X	X	
3	Beta Phase Testing			X	X

Specific outcomes:

- We completed developing the interface for MACE that is optimized for how investigators intend to use the MACE toolkit.

- We completed the execution of the quality assurance plan.

- We completed the Beta-Testing of the MACE toolkit to evaluate usability, ensure functionality and to validate accuracy.

2. Describe the current state of the technology and related product/service:

   Currently, the development of the toolkit and its interface has been completed. The tool can handle small to medium size of document set. However, we discovered based on our tests that the tool is not efficient enough with respect to large document set. We are currently analyzing our code to determine the reason and come up with a solution to the issue to make the tool more efficient and usable in practice.

3. List the number of faculty and student participants as a result of funding:

   # of faculty: 1

   # of students: 8 (4 graduate and 4 undergraduate)

4. What are the potential economic benefits:

   Any law enforcement agency interested in detecting criminal networks from a suspect's digital media could employ our toolkit, and for all types of crimes. In addition, cybersecurity companies can also utilize the proposed technology to extract relevant information about on-line malicious attacks and breaches carried out against a terminal machine or server. Several cyber forensic tools are commercially available, including Forensic Toolkit®, EnCase®, CAINE and The Sleuth Kit®. However, unlike the MACE toolkit, there is currently no theoretical approach nor commercial tool that allows the investigator to visualize the criminal networks and their pattern of communications in documents captured on a suspect's computer. As a result, we hope that law enforcement agencies as well as cybersecurity companies will be eager to add MACE to their cyber forensics laboratories.

5. Description future plans for project continuation or expansion:

   Once the efficiency issue has been addressed, we will start marketing our toolkit to cybersecurity companies, and to law enforcement agencies at the local, state, and federal levels. Dependent upon the results from our rigorous market analysis and planning, we intend to effectuate the commercialization of the MACE toolkit by creating a start-up company or directly out-licensing the technology to an established company for dissemination.

1. We will work with the Office of Technology Transfer (OTT) in a two-pronged approach for the most effective and appropriate licensing strategy: (1) creating a startup around this technology or (2) out-licensing to an existing entity.

2. We will complete the internal process for Entrepreneurial Conflicts of Interest for startup formation while concurrently identifying potential licensees.

3. In preparing the licensing strategy, we will analyze each licensing opportunity individually in a manner that reflects the business needs and values of BSU.

6. Please provide a final expenditure report (attached) and include any comments here:

FINAL EXPENDITURE REPORT

A. FACULTY AND STAFF		
Name/Title	$ Amount Requested	Actual $ Spent
Dr. Gaby Dagher	13,403.00	13,403.00 *
B. VISITING PROFESSORS		
Name/Title	$ Amount Requested	Actual $ Spent
N/A		
C. POST DOCTORAL ASSOCIATES/OTHER PROFESSIONALS		
Name/Title	$ Amount Requested	Actual $ Spent
N/A		
D. GRADUATE/UNDERGRADUATE STUDENTS		
Name/Title	$ Amount Requested	Actual $ Spent
Anthony Harris	34,696.61 (students)	2,040.00
Hannah Johnson		634.40
Tanya Khatri		910.00
Manish Kumar		14,101.60 *
Cybil Lesbyn		2,054.00
Danyal Mohammadi		5,120.00
James Souder		572.00
Yi Xie		9,420.00 *
E. FRINGE BENEFITS		
Rate of Fringe (%)	$ Amount Requested	Actual $ Spent

22 % Dr. Gaby Dagher	3,077.89	2,922.77 *
6.5% Students	2,270.00	2,269.73 *
PERSONNEL SUBTOTAL:	**$53,447.50**	**53,447.50**
F. EQUIPMENT: (List each item with a cost in excess of $1000)		
Item/Description	$ Amount Requested	Actual $ Spent
1.		
2.		
EQUIPMENT SUBTOTAL:		$0.00
G. TRAVEL		
Description	$ Amount Requested	Actual $ Spent
1.		
2.		
TRAVEL SUBTOTAL:		
H. PARTICIPANT SUPPORT COSTS:		
Description	$ Amount Requested	Actual $ Spent
1.		
2.		
3		
PARTICIPANT SUPPORT COSTS SUBTOTAL:		
I. OTHER DIRECT COSTS:		
Description	$ Amount Requested	Actual $ Spent
1.Computer for student employee	1,475.50	1,475.50
2 Student fees for Manish Kumar	4,377.00	4,377.00
3.		
OTHER DIRECT COSTS SUBTOTAL:	5,852.50	5,852.50
TOTAL COSTS (Add Subtotals):	59,300.00	59,300.00
TOTAL AMOUNT REQUESTED:		59,300.00
TOTAL AMOUNT SPENT:		59,300.00

* Based on encumbered Payroll and Fringe that will post on 07/13/18. All payroll and fringe is for work performed through end date of grant – 06/30/18.

7. List invention disclosures, patent, copyright and PVP applications filed,

technology licenses/options signed, start-up businesses created, and industry involvement:

We have almost completed a journal paper based on the on the current functionalities of the MACE toolkit to be published in a prestigious journal (e.g. Elsevier, Springer). We will also look for the possibility of patenting the algorithm we use to extract malicious communities.

8. Any other pertinent information:
   --