REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
22-08-2018	Final Report	5-Sep-2017 - 4-Jun-2018

4. TITLE AND SUBTITLE
Final Report: Large-Scale Network Inference: Detecting the Unknown and the Intermittent

5a. CONTRACT NUMBER
W911NF-17-1-0464

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Cornell University
Office of Sponsored Programs
373 Pine Tree Road
Ithaca, NY 14850 -2820

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

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

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
69159-NS.12

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Qing Zhao
UU	UU	UU	UU		19b. TELEPHONE NUMBER 607-255-6899

RPPR Final Report
as of 13-Mar-2019

Agency Code:

Proposal Number: 69159NS **Agreement Number: W911NF-17-1-0464**
INVESTIGATOR(S):

Name: Qing Zhao
Email: qz16@cornell.edu
Phone Number: 6072556899
Principal: Y

Organization: **Cornell University**
Address: Office of Sponsored Programs, Ithaca, NY 148502820
Country: USA
DUNS Number: 872612445 EIN: 150532082
Report Date: 04-Sep-2018 Date Received: 22-Aug-2018
Final Report for Period Beginning 05-Sep-2017 and Ending 04-Jun-2018
Title: Large-Scale Network Inference: Detecting the Unknown and the Intermittent
Begin Performance Period: 05-Sep-2017 **End Performance Period:** 04-Jun-2018
Report Term: 0-Other
Submitted By: Qing Zhao Email: qz16@cornell.edu
Phone: (607) 255-6899
Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 1 **STEM Participants:** 3

Major Goals: Accurate and timely detection of network anomalies or unusual activities, be it endogenous (caused by internal malfunctioning components) or exogenous (caused by hostile attacks and intrusions), is crucial to the functionality and survivability of tactical military networks. The objective of this research is to develop general design methodologies for large-scale network inference for anomaly detection. We aim to establish fundamental limits on sample complexity—in particular, the scaling behavior of sample complexity with respect to the problem size and the detection accuracy—and develop efficient algorithms that achieve or approach the fundamental limits with scalable low-complexity implementations. Our emphasis is on low-complexity deterministic strategies with implementations scalable to large networks.

Accomplishments: Please see attached.

Training Opportunities: This project provide valuable research experience and training opportunities to graduate and undergraduate students. One junior-level undergraduate and two Ph.D. students were involved in the project. One of the Ph.D. students received his degree during the reporting period.

This project also provided opportunities for international collaboration. The PI and the graduate students working on this project collaborated with Prof. Kobi Cohen and his students in Ben-Gurion University, Israel, all through the project duration.

Results Dissemination: Results have been disseminated through high-impact journals and conferences.

Honors and Awards: The PI was recently named a Marie Sklodowska-Curie Fellow of the European Commission and the Jubilee Professor of Chalmers University, Sweden.

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: Graduate Student (research assistant)
Participant: Chao Wang

Person Months Worked: 5.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)
Participant: Boshuang Huang
Person Months Worked: 5.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: PD/PI
Participant: Qing Zhao
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Undergraduate Student
Participant: Yirong Cheng
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

CONFERENCE PAPERS:

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: The 55th Annual Allerton Conference on Communication, Control, and Computing
Date Received: 07-Nov-2017 Conference Date: 03-Oct-2017 Date Published: 03-Oct-2017
Conference Location: Urbana, IL
Paper Title: Anomaly Detection under a Nonlinear System Cost Objective Function
Authors: Andrey Gurevich, Kobi Cohen, Qing Zhao
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 13-Mar-2019

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)
Date Received: 21-Aug-2018 Conference Date: 15-Apr-2018 Date Published: 15-Apr-2018
Conference Location: Calgary, Canada
Paper Title: Hierarchical Heavy Hitter Detection under Unknown Models
Authors: Sattar Vakili, Qing Zhao, Chang Liu, Chen-Nee Chuah
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)
Date Received: 21-Aug-2018 Conference Date: 15-Apr-2018 Date Published: 15-Apr-2018
Conference Location: Calgary, Canada
Paper Title: Active Anomaly Detection in Heterogeneous Processes
Authors: Boshuang Huang, Kobi Cohen, Qing Zhao
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)
Date Received: 21-Aug-2018 Conference Date: 25-Jun-2018 Date Published: 25-Jun-2018
Conference Location: Kalamata, Greece
Paper Title: Dynamic Search on a Tree with Information-Directed Random Walk
Authors: Chao Wang, Qing Zhao, Kobi Cohen
Acknowledged Federal Support: **Y**

DISSERTATIONS:

Publication Type: Thesis or Dissertation
Institution: Cornell University
Date Received: 21-Aug-2018 Completion Date: 5/11/18 9:12AM
Title: Sequential Design of Experiments for Anomaly Detection
Authors: Chao Wang
Acknowledged Federal Support: **N**

Summary of Accomplishments

This nine-month project aims to tackle the following three research issues that are central to large-scale network inference. These three tasks represent a logical progression in scope and level of difficulty. Despite the ambitious research agenda with a short performance period, we have obtained significant results and successfully completed all tasks. Below we summarize our accomplishments in each of the three tasks.

Task 1: achieving optimal sample complexity with respect to detection accuracy

As the first step, we focus on achieving optimal sample complexity with respect to detection accuracy under known models. Results and insights obtained in this step serve as basic building blocks for subsequent steps.

Our technical approach rests on the general theory of active hypothesis testing originated from Chernoff's seminal work on sequential design of experiments. When applied to the problem of anomaly detection, active hypothesis testing sequentially and adaptively determines where to search based on the current estimate on where the anomalies may reside. As a result, compared to a passive test that screens the entire search space indiscriminately, active hypothesis testing significantly reduces sample complexity for a given detection accuracy.

Under this task, we have developed low-complexity *deterministic* active inference strategies that achieve optimal scaling with the required detection accuracy. More significantly, compared with the randomized test developed in Chernoff's original theory, these deterministic policies offer significant performance gain in the finite regime and considerable reduction in computation, memory, and implementation complexity, especially when the network size is large. In particular, in [1,2], we tackle the problem in heterogeneous networks where anomalies manifest differently across a large number of network components. In [3,4], we adopt an application-oriented objective: minimizing network-level operation cost incurred by anomalous components. This objective function captures the varying degrees of criticality of different network components and allows a general nonlinear dependency of the cost on the time of being anomalous (e.g., the risk posed by an anomalous component may grow superlinearly with time due to inter-dependency across components and potential cascading effects in the network).

Task 2: Achieving optimal sample complexity with respect to the network size

In this task, we aim to develop active inference strategies that achieve optimal scaling of sample complexity in terms of the network size while preserving the optimal scaling with respect to the detection accuracy. The emphasis is on achieving a sublinear scaling with the network size.

The key to a sublinear scaling with the problem size is to exploit the hierarchical structure of the search space inherent to many applications. For example, network traffic flows can be aggregated based on IP prefix, leading to a tree-structured search space for

heavy hitter and denial-of-service (DoS) and distributed denial-of-service (DDoS) detection. In computer vision applications such as surveillance by UAVs with limited battery capacity, sequentially determining areas to zoom in or zoom out can quickly locate anomalies by avoiding giving each pixel equal attention.

Our first result is on designing nested hierarchical search strategies within the framework of quantitative group testing [5]. Using DoS detection as case studies, we show that this nested hierarchical search strategy achieves orders of magnitude of improvement over two prevailing sampling-based approaches in sample complexity, detection accuracy, and counter consumption. Our second major result is an active inference policy over a tree-structured search space [6,7]. This novel policy is asymptotically optimal with respect to the detection accuracy and order-optimal (more specifically, a logarithmic order) with respect to the network size. Furthermore, effectively localizing the data processing to small subsets of the search space, this policy a constant order in terms of computation and memory complexity as compared with the superlinear order of prevailing approaches.

Task 3: Achieving optimal sample complexity under unknown models

In this task, we consider the case when the stochastic models of both anomalous and normal components are unknown or only partially known. We tackle the challenging problem of achieving optimal sample complexity in both detection accuracy and network size under unknown models.

Our approach is to integrate online learning with active inference to tackle the challenge of unknown models. In [8], we take a parametric model where anomalous and nominal behaviors differ in the (unknown) values of certain parameters. In [9], we adopt a more general nonparametric model where each component can follow arbitrary unknown distributions that are potentially heavy-tailed. Under both formulations, we develop order-optimal policies. In addition to detailed theoretic analysis to establish the optimality of the proposed policies, we also demonstrate the performance of the proposed policy in intrusion detection applications using real network traffic traces, in particular, the DARPA intrusion detection dataset that contains 5-million network connections and 4 classes of attacks. The experiment results show orders of magnitude improvement of the proposed policy over existing methods (see [8]).

References:

[1]. Boshuang Huang, Kobi Cohen, Qing Zhao, “Active Anomaly Detection in Heterogeneous Processes,” to appear in *IEEE Transactions on Information Theory*.

[2]. Boshuang Huang, Kobi Cohen, Qing Zhao, “Active Anomaly Detection in Heterogeneous Processes,” in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, April, 2018.

[3]. Andrey Gurevich, Kobi Cohen, Qing Zhao, "Sequential Anomaly Detection under a Nonlinear System Cost," submitted to *IEEE Transactions on Signal Processing.*

[4]. Andrey Gurevich, Kobi Cohen, Qing Zhao, "Anomaly Detection under a Nonlinear System Cost Objective Function," in *Proc. of the 55th Annual Allerton Conference on Communication, Control, and Computing*, October, 2017.

[5]. Chao Wang, Qing Zhao, Chen-Nee Chuah, "Optimal Nested Test Plan for Combinatorial Quantitative Group Testing," *IEEE Transactions on Signal Processing*, vol. 66, no. 4, pp. 992 - 1006, February, 2018.

[6]. Chao Wang, Kobi Cohen, Qing Zhao, "Information-directed Random Walk for Rare Event Detection in Hierarchical Processes," submitted to *IEEE Transactions on Information Theory*.

[7]. Chao Wang, Kobi Cohen, Qing Zhao, "Dynamic Search on a Tree with Information-directed Random Walk," in *Proc. of IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC),* June, 2018.

[8]. B. Hemo, K. Cohen, Q. Zhao, "Searching for Anomalies over Composite Hypotheses," submitted to *IEEE Transactions on Signal Processing*.

[9]. Sattar Vakili, Qing Zhao, Chang Liu, Chen-Nee Chuah, "Hierarchical heavy hitter detection under unknown models," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* April, 2018.