


SpringerBriefs in Computer Science

Linan Huang · Quanyan Zhu



# Cognitive Security

A System-Scientific  
Approach

# **SpringerBriefs in Computer Science**

## **Series Editors**

Stan Zdonik, Brown University, Providence, RI, USA

Shashi Shekhar, University of Minnesota, Minneapolis, MN, USA

Xindong Wu, University of Vermont, Burlington, VT, USA

Lakhmi C. Jain, University of South Australia, Adelaide, SA, Australia

David Padua, University of Illinois Urbana-Champaign, Urbana, IL, USA

Xuemin Sherman Shen, University of Waterloo, Waterloo, ON, Canada

Borko Furht, Florida Atlantic University, Boca Raton, FL, USA

V. S. Subrahmanian, University of Maryland, College Park, MD, USA

Martial Hebert, Carnegie Mellon University, Pittsburgh, PA, USA

Katsushi Ikeuchi, University of Tokyo, Tokyo, Japan

Bruno Siciliano, Università di Napoli Federico II, Napoli, Italy

Sushil Jajodia, George Mason University, Fairfax, VA, USA

Newton Lee, Institute for Education, Research and Scholarships, Los Angeles, CA, USA

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

Briefs allow authors to present their ideas and readers to absorb them with minimal time investment. Briefs will be published as part of Springer's eBook collection, with millions of users worldwide. In addition, Briefs will be available for individual print and electronic purchase. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, easy-to-use manuscript preparation and formatting guidelines, and expedited production schedules. We aim for publication 8–12 weeks after acceptance. Both solicited and unsolicited manuscripts are considered for publication in this series.

**\*\*Indexing:** This series is indexed in Scopus, Ei-Compendex, and zbMATH **\*\***

Linan Huang • Quanyan Zhu

# Cognitive Security

A System-Scientific Approach

 Springer

Linan Huang  
New York University  
Brooklyn, NY, USA

Quanyan Zhu  
New York University  
Brooklyn, NY, USA

ISSN 2191-5768

ISSN 2191-5776 (electronic)

SpringerBriefs in Computer Science

ISBN 978-3-031-30708-9

ISBN 978-3-031-30709-6 (eBook)

<https://doi.org/10.1007/978-3-031-30709-6>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

Humans are indispensable components in Cyber-Physical Systems (CPSs) due to their cognitive capacities and the ultimate goal to support rather than supersede humans. The close integration of humans, CPSs, and Artificial Intelligence (AI) creates AI-Powered Human-Cyber-Physical Systems (HCPSs) that drive the development of Industry 5.0 and revolutionize the future of work. Despite the remarkable cognitive capabilities (e.g., situation awareness, decision-making, and cooperation) of human users, operators, and administrators in designing, operating, supervising, and securing HCPSs, humans have been the weakest link in HCPS security.

Attackers are increasingly sophisticated in exploiting not only vulnerabilities in software and hardware but also human vulnerabilities to obtain initial credentials from human users through phishing, scamming, and various types of social engineering. These exploitable human vulnerabilities lie primarily in human cognitive processes such as perception, attention, memory, and mental operation. An adversary can use a variety of reactive (e.g., design deceptive phishing emails to evade users' attention) and proactive (e.g., generate excessive feints to overload human operators) methods to disrupt human cognitive processes so that humans misperceive the HCPS state and/or are misled into fallacious reasoning and incorrect decisions. The consequence can exacerbate and further lead to the compromise of cyber and physical components, as well as a system-level meltdown. It is both opportune and imperative to create socio-technical solutions to break such a cognitive kill chain, make humans resilient to cognition-based threats, and enhance the *cognitive security* in the new battle field of HCPSs.

To this end, in this book, we present a *system science foundation* that builds on and bridges the fields of psychology, neuroscience, data science, decision and game theory, and learning theory to develop transdisciplinary *socio-technical* mechanisms at the convergent human-technology frontier to mitigate cognitive threats. Based on the understanding of human cognition and multidimensional data from various biosensors, this book develops *human-centered assistive AI technologies* to improve cognitive resilience and harden cognitive security. Leveraging system-scientific approaches to cognitive security brings quantitative, modular, multi-scale, and

transferable solutions. This book goes further to create new metrics and characterize the fundamental limits of cognitive security.

The book investigates emerging cybersecurity concerns regarding human cognition and behavior and does so from a unique system perspective. It provides a self-contained introduction to the area of cognitive security and a succinct overview of essential system-scientific methods that play a central role in the modeling, analysis, and design of human-centric security solutions. The book uses reactive and proactive attention attacks as two case studies to demonstrate the system-scientific modeling and design of assistive solutions. Cognitive security is a multi-disciplinary and vibrant area of research. The chapters of this book are not meant to be comprehensive, but they are organized to offer readers an overview and several success stories that will motivate future research in the broad area and push the frontier of human-technology convergence.

The authors of this book would like to acknowledge their association with the Center for Cyber Security (CCS) at New York University (NYU) when a major part of this book was completed. We both like to take this opportunity to thank many of our colleagues and students for their input and suggestions. Special thanks go to members of the Laboratory of Agile and Resilient Complex Systems (LARX) at NYU, including Zhiheng Xu, Jeffrey Pawlick, Juntao Chen, Junaid Farooq, Rui Zhang, Tao Zhang, Yunhan Huang, Tao Li, Shutian Liu, Yuhan Zhao, Yunfei Ge, Yunian Pan, Patrick Yubeaton. In particular, Patrick has helped with proofreading of this book, and his suggestions have drastically improved the quality of the presentation of this book. We would also like to acknowledge the funding support from the National Science Foundation (NSF), the US Department of Energy (DOE), and the Army Research Office (ARO).

Brooklyn, NY, USA  
Brooklyn, NY, USA  
December 2022

Linan Huang  
Quanyan Zhu

# Contents

- 1 Introduction** ..... 1
  - 1.1 AI-Powered Human-Cyber-Physical Systems ..... 1
    - 1.1.1 Human Roles in Mission Stack ..... 2
    - 1.1.2 Incorporating AI Stack into Mission Stack ..... 3
  - 1.2 Cognitive Security in HCPS ..... 5
    - 1.2.1 Attack and Defense Stack ..... 5
    - 1.2.2 Feedforward and Feedback Cognitive Processes ..... 6
    - 1.2.3 Exploitable Cognitive Vulnerability ..... 7
    - 1.2.4 Cognitive Attack ..... 11
    - 1.2.5 Cognitive Security ..... 13
  - 1.3 System-Scientific Perspectives for Cognitive Security ..... 19
    - 1.3.1 Advantages of System-Scientific Approaches ..... 20
  - 1.4 Outline and Organization ..... 21
  - References ..... 22
- 2 System-Scientific Methods** ..... 27
  - 2.1 Decision Theory ..... 27
    - 2.1.1 Expected Utility Theory ..... 28
    - 2.1.2 Cumulative Prospect Theory ..... 30
    - 2.1.3 Notes ..... 31
  - 2.2 Game Theory ..... 32
    - 2.2.1 Basic Game Setting and Elements ..... 32
    - 2.2.2 Mixed Strategy and Nash Equilibrium ..... 33
    - 2.2.3 Notes ..... 34
  - 2.3 Learning Theory ..... 34
    - 2.3.1 Bayesian Learning ..... 34
    - 2.3.2 Reinforcement Learning ..... 35
    - 2.3.3 Notes ..... 37
  - References ..... 37



- 3 Cognitive Capacities for Designing, Operating, Supervising, and Securing Complex Systems** ..... 41
  - 3.1 Situation Awareness..... 41
    - 3.1.1 Perception of the Elements in the Environment ..... 43
    - 3.1.2 Comprehension of the Current Situation ..... 43
    - 3.1.3 Projection of Future Status ..... 43
  - 3.2 Problem Solving and Decision Making ..... 44
    - 3.2.1 Flexibility and Creativity ..... 44
    - 3.2.2 Adaptability and Learning..... 44
  - 3.3 Collaboration ..... 45
    - 3.3.1 Interaction and Communication..... 45
    - 3.3.2 Theory of Mind ..... 46
  - 3.4 Discussions ..... 47
  - References ..... 48
- 4 Review of System-Scientific Perspectives for Analysis, Exploitation, and Mitigation of Cognitive Vulnerabilities** ..... 49
  - 4.1 Innate Vulnerabilities ..... 49
    - 4.1.1 Attention Vulnerability ..... 50
    - 4.1.2 Subjective Risk Perception ..... 53
    - 4.1.3 Vulnerabilities of Reasoning and Decision Making..... 54
  - 4.2 Acquired Vulnerabilities ..... 57
    - 4.2.1 Lack of Security Information and Knowledge ..... 57
    - 4.2.2 Misaligned Incentive and Incompliance ..... 59
  - 4.3 Discussions ..... 60
  - References ..... 61
- 5 ADVERT: Defending against Reactive Attention Attacks**..... 67
  - 5.1 Introduction ..... 67
  - 5.2 Attention Enhancement Mechanism ..... 69
    - 5.2.1 Visual State Transition Model..... 70
    - 5.2.2 Feedback Visual-Aid Design ..... 71
    - 5.2.3 Evaluation of Attention Status ..... 72
    - 5.2.4 Q-Learning via Consolidated Data..... 74
  - 5.3 Phishing Prevention Mechanism ..... 75
    - 5.3.1 Metrics for Phishing Recognition ..... 75
    - 5.3.2 Efficient Hyperparameter Tuning ..... 76
  - 5.4 Case Study ..... 77
    - 5.4.1 Validation of Attention Enhancement Mechanism ..... 80
    - 5.4.2 Validation of Phishing Prevention Mechanism ..... 81
  - 5.5 Conclusions and Future Work ..... 82
  - References ..... 83
- 6 RADAMS: Defending Against Proactive Attention Attacks** ..... 85
  - 6.1 Introduction ..... 85
  - 6.2 Motivation and Features of Feints ..... 89

6.3	Model of IDoS Attacks .....	90
6.4	Model of Alert Triage.....	91
6.5	Model of Human Attention and Alert Inspection.....	92
6.6	Model of AI-Powered Enhancer.....	95
6.7	Conclusions and Future Work .....	97
	References .....	98
<b>7</b>	<b>Summary and Conclusions</b> .....	<b>101</b>
7.1	Summary and Insights .....	101
7.2	Applications in Trustworthy HCPS .....	103
7.3	Cognitive Insecurity of Adversaries.....	104
7.4	Expanding the Scope of Cognitive Security .....	106
	7.4.1 Breadth: Multi-Dimensional Cognitive Processes .....	106
	7.4.2 Abundance: Multifaceted Measures of Cognitive Factors ....	106
	7.4.3 Depth: Multi-Level Characterization of Cognitive Security ..	107
	References .....	108

# Acronyms

AI	Artificial Intelligence. 1–5, 13, 16–20, 44, 46, 51–53, 70, 90, 93, 97, 107
AoI	Area of Interest. 70, 72–74, 80
APT	Advanced Persistent Threat. 6, 62, 103, 107
BO	Bayesian Optimization. 71, 78, 80, 83
CDF	Cumulative Distribution Function. 30, 31
CPS	Cyber-Physical System. 1–5, 7, 12–14, 17, 19, 20, 42, 44, 45, 51, 56, 87, 90, 103, 104
CPT	Cumulative Prospect Theory. 28, 30, 31, 56
DDoS	Distributed Denial-of-Service. 12, 13, 92
DoS	Denial-of-Service. 87, 88
EUT	Expected Utility Theory. 28–31, 56, 57
HCPS	Human-Cyber-Physical System. 1–3, 5, 7, 11, 14–21, 27, 31, 34–36, 45, 46, 48, 51, 52, 57–62, 87, 103–105, 108–110
HMI	Human Machine Interface. 4, 12, 13, 17
HRA	Human Reliability Analysis. 14
IDoS	Informational Denial-of-Service. 8, 11–13, 16–19, 21, 55, 87–93, 95, 98–100
IDS	Intrusion Detection System. 15, 35