SpringerBriefs in Computer Science

Linan Huang · Quanyan Zhu

Cognitive Security
A System-Scientific Approach

Springer

SpringerBriefs in Computer Science

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

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 Linan Huang
Brooklyn, NY, USA Quanyan Zhu
December 2022

Contents

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