# Security and Privacy of Biomedical Cyber-Physical Systems

Tamara Bonaci

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Howard J. Chizeck, Chair

M. Ryan Calo

Blake Hannaford

Tadayoshi Kohno

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

**Abstract**

**Security and Privacy of Biomedical Cyber-Physical Systems**

Tamara Bonaci

Chair of the Supervisory Committee:
Professor Howard J. Chizeck
Department of Electrical Engineering

Advances in cyber-physical systems (CPS), machine learning, big data techniques, and in cloud computing having been enabling ever more data to be collected about systems and their users, in search for unique features and interesting patterns. This, in turn, has been giving rise to the *personalization trend*, an approach where a cyber-physical system uses observed features and patterns in order to better adopt to users' needs, abilities, and preferences. Examples of personalized technologies are many, from buildings learning about inhabitants' daily routines and preferences [13], to music, video and shopping recommendation systems [19, 14, 1].

The *personalization trend* is expected to be particularly important for biomedical cyber-physical systems, where data about patients, and/or medical practitioners is expected to allow systems to better adapt to medical needs. Yet, this trend is not without risks. Any time data about users and systems is recorded, processed, and possibly stored for future analysis, *security and privacy risks* arise. Misusing the collected data gives rise to threats ranging from compromising or breaking systems to shaming, manipulating or even physically harming users. Moreover, in biomedical CPS, some biosignals or data about genetic material may contain not only the current information about patients, but may allow predictions to be made about patients' future, or their relatives.

Security and privacy issues related to personalized CPS are thus front and center, and this dissertations focuses on those arising in biomedical cyber-physical systems. In doing so,

we start from human components of such systems, and propose that *users' idiosyncrasies, in the way users interact with systems, may expose these systems to potential security and privacy risks. At the same time, however, users' unique traits can be used to increase the systems' security, privacy and usability properties.*

To investigate the stated hypothesis, this dissertation focuses on three questions: (1) how do (how could) biomedical cyber-physical systems use users' idiosyncrasies, (2) what security and privacy vulnerabilities may arise from users' unique traits, and (3) how can users' idiosyncrasies be leveraged to increase systems' security and privacy? The question about possible vulnerabilities is answered by analyzing properties of *brain-computer interfaces*, an example of emerging neural engineering technology. The last question is answered in the context of the next generation teleoperated robotic systems, focusing specifically on surgical robots.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## GLOSSARY

AES: The Advanced Encryption Standard

ANN: Artificial Neural Network

ARP: The Address Resolution Protocol

BCI: Brain-Computer Interface

CL-DBS: Closed Loop Deep Brain Stimulator

CPS: Cyber-Physical System

CPU: Central Processing Unit

DBS: Deep Brain Stimulator

DDS: DoS Defense System

DDOS: Distributed Denial-of-Service Attack

DOF: Degree of Freedom

DOS: Denial-of-Service attack

DWT: Discrete Wavelet Transform

EBM: Evidence Based Medicine

ECG: Electrocardiogram

EDR: Electrodermal Response

EEG: Electroencephalogram

EMG: Electromyogram

ERD: Event Related Desynchronization

ERN: Error-Related Negativity, One Possible Event Related Potential Component

ERP: Event Related Potential

ERS: Event Related Synchronization

E-STOP: Emergency Stop Button Used in Teleoperated Robotic Systems

FDA: US Food and Drug Administration

FLS: Fundamentals of Laparoscopic Surgery

GUI: Graphical User Interface

HCI: Human-Computer Interaction

HIP: Haptic Interaction Point

HIPAA: US Federal Health Insurance Portability and Accountability Act

HMM: Hidden Markov Model

IDS: Intrusion Detection System

IPS: Intrusion Prevention System

ITP: Interoperable Telesurgery Protocol

LQR: Linear Quadratic Regulator

MAC: Multiple Access Channel

MAC-GF: Multiple Access Channel with Generalized Feedback

MITM: Man-in-the-Middle Attack

N400: N400, One Possible Event Related Potential Component

NSF: National Science Foundation

PCA: Principal Component Analysis

PII: Private Identifiable Information

P300: P300, One Possible Event Related Potential Component

P600: P600, One Possible Event Related Potential Component

ROS: Robot Operating System

SCP: Slow Cortical Potential

SMR: Sensorimotor Rhythm

SSVEP: Steady State Evoked Visual Potential

SVM: Support Vector Machine

UAV: Unmanned Aerial Vehicle

UDP: User Datagram Protocol

VQ: Vector Quantization

VSD: Value Sensitive Design

WSN: Wireless Sensor Network

WTC: World Trade Center

## ACKNOWLEDGMENTS

This dissertation is the culmination of several years of work, but also guidance, help and support from many wonderful people who have made my years as a graduate student possible, and mostly enjoyable, and without whom I would not be approaching the finish line (and the coveted purple kiss).

First, I would like to thank my advisor and my mentor, Howard Chizeck, for all of his help, support, guidance, patience, and encouragement. Howard has been an amazing mentor, in all aspects of an academic career, but also in many aspects of life. He has been my champion, but also a safety net whenever I needed it, and I have greatly benefited from that. I am looking forward to our next endeavor together. To serendipity.

I would also like to thank other members of my PhD committee, for all of their guidance, support and patience: to Ryan Calo and Batya Friedman, for opening a whole new world to me, and providing me with new tools to work with, to Blake Hannaford, Tadayoshi Kohno, and Jeff Ojeman, for their great questions and comments that have so many time significantly enhanced my work, and to Franziska Roesner, for all of her support and great advices on my career path.

Thank you also to my other collaborators: Aaron Alva, Tim Brown, Nguyen Le my Chau, Jeffrey Herron, Kevin Huang, Fethya Mohamed Ibrahim, Tom Lendvay, Andrew Lewis, Tyler Libey, Charles Matlack, Brian Mogen, Patrick Moore, Xiyu Ouyang, Margaret Thompson, Hannah Werbel, Junjie Yan, and Tariq Yusuf for all their help, suggestions and advices during my dissertation work.

Conducting experiments with human subjects requires a great deal of time and coordination, and I am extremely grateful for all the help I have had on both of my experimental studies. Many thanks to Charles Matlack for teaching me about experimental design process, and for all of his help and support with my very first IRB application; to Jeffrey

Herron, for all the help with the development of both sets of experiments, and for all the programming he has done; to Tyler Libey and Brian Mogen, for making the Flappy Whale game available to us; to Hawkeye King, for the help with the design of teleoperation security experiments; to Danying Hu, Kevin Huang and Andrew Lewis, for all of their help with Raven II setup and troubleshooting, as well as to Junjie Yan, for all of his help conducting teleoperation security experiments.

I have been fortunate to be a part of a friendly, supportive and vibrant lab, and would like to thank current and past members, for all of their help, support, insights, humor and friendship: Nava Aghdasi, Paul Bartell, Tim Brown, Lei Cheng, David Caballero, Nicklas Gusafsson, Andrew Haddock, Mohammad Haghighipanah, Jeffrey Herron, Brady Houston, Danying Hu, Kevin Huang, Iris Jiang, Sina Nia Kosari, Andrew Lewis, Jack Lindsay, Charles Matlack, Muneaki Miyasaka, Sharon Newman, Fredrik Ryden, Astrini Sie, Margaret Thompson, Lee White, and Junjie Yan.

Thank you also to many other CSNE, EE and TPL people for all of their great advices throughout my graduate experience, as well as on my career path: Jeff Bilmes, Sam Burden, Eric Chudler, Bryan Crockett, Robert Bruce Darling, Maryam Fazel, Karen Fisher, Steven Graham, Mary Guiden, Deborah Harper, Lise Johnson, Brenda Larson, Alex Llapitan, Pim Lustig, Emily McReynolds, Brian Otis, Josh Patrick, Hal Perkins, Raj Rao, Eve Riskin, Rad Roberts, and John Sahr.

During my time at UW, I have gotten a great deal of support from other graduate students. I am grateful to Amittai Axelrod, Colby Boyer, Daniela D'Auria, Shelly Jang, Mike Katell, Julie Medero, Karthik Mohan, Temitope Oluwafemi, Jason Silver, Jessica Tran, and Soumya Vasisth for their support and encouragement. I am especially in debt to Nicole Nichols and Nicole Thomas for their great advices, help, support and encouragement.

Being a full-time student while raising a child has made me very much aware of how precious the time is, and incredibly grateful to my family's support network. Thank you to my parents, Sanja and Miljenko Vurin, and my parents-in-law, Ivanka and Juraj Bonaci, for taking care of my son when I got back to work after my son was born, and when there

were only a few people in the whole universe whom my husband and I trusted enough to leave our son with. Thank you to our friends Anita and Josip Medved, as well as to the staff at the Bright Horizons, especially to Elvira Knowles and Kristen Farnam.

I am and forever will be grateful for the support of my family. Thank you to my wonderful son Daniel, who does not know a life without Mom going to school, but who has been the best, most supportive toddler one can imagine. Thank you to my daughter-to-be, for her patience during the time she has spent stuck in an uncomfortable position while I was wrapping up my dissertation. Finally, thank you to my amazing husband Davor, who has been my greatest support, my biggest advocate, my rock and my safe haven, and without whom I would have never even consider a PhD. Through this whole journey, Davor has picked up more slack, and has showered Daniel and me with more love, affection and support than I would have ever thought possible. Davor, thank you for everything!

My graduate career and work described in this dissertation were supported in part by the National Science Foundation Grant CNS-1329751, the National Science Foundation Award EEC-1028725, and by the gift from the University of Washington Tech Policy Lab.

## DEDICATION

To my husband Davor, my son Daniel and my daughter-to-be.

I love you “to infinity and beyond” :)

# Chapter 1

# INTRODUCTION

Advances in cyber-physical systems (CPS), machine learning and big data techniques, as well as in cloud computing having been enabling ever more data to be collected about systems and their users, in search for unique features and interesting patterns. This, in turn, has been giving rise to the *personalization trend*, an approach where a cyber-physical system leverages observed features and patterns in order to better adapt to users' needs, abilities, and preferences. Examples of personalized technologies are many, with many more to come, from buildings learning about inhabitants' daily routines and preferences [13], to music, video and shopping recommendation systems [19, 14, 1].

The *personalization trend* is expected to be particularly important for biomedical cyber-physical systems, where data about patients, and/or medical practitioners is expected to allow systems to better adapt to medical needs. The development of personalized biomedical CPS is expected to facilitate:

- More efficient and effective delivery of medical treatments,
- Safer and faster execution of medical procedures, with less negative outcomes and
- Increase in patients' ability and willingness to accept their medical devices.

Personalized CPS, are not, however, without risks. Any time data about users and systems is recorded, transmitted, processed, and possibly stored for future analysis, *security and privacy risks* arise. Misusing the collected data gives rise to threats ranging from compromising or breaking systems to shaming, manipulating or even harming users. Moreover, in biomedical CPS, some biosignals or data about genetic material may contain not only the current information, but may allow predictions to be made about patients' future, or their relatives.

This dissertation focuses on security and privacy risks that have recently arisen, or may arise with *biomedical cyber-physical technologies*. In doing so, we take the following stance:

- Security and privacy represent important issues for emerging cyber-physical technologies, and should be taken into account during the design and development of these technologies.
- Even though, in many cases, no real concerns or threats with these technologies have been reported yet, we should not wait for a problem to occur, as the consequences may be too severe.
- When analyzing and addressing potential security and privacy issues surrounding cyber-physical systems, we can, and often times should leverage the existing knowledge, available techniques and mitigation strategies from the cyber security community. Yet, there exists two fundamental differences between cyber and cyber-physical system. Those differences are the *physical component* and the *human component* of a cyber-physical system.

**Physical component of a CPS:** Knowledge about physical constraints of a cyber-physical system, such as, for example, knowledge about a system's dynamics, possible physical constrains, and ultimately about the laws of physics that a system needs to obey, may provide constraints and limitations on the system as a whole. At the same time, this knowledge provides additional information that can be leveraged to enhance the security and privacy of the system.

**Human component of a CPS:** Human users (operators, patients) have a unique way of interacting with a cyber-physical system. For example, users' biosignals are increasingly being used to personalize biomedical systems, and those signal have been shown to be user-specific [150, 119]. Similarly, it has recently been shown that users have a unique way of interacting with their touch-based devices, in terms of forces and torques applied to those devices [39, 238]. These users' idiosyncrasies and unique features may expose cyber-physical systems to potential security and privacy risks. At the same time, however, users' unique traits and ways of interacting with a system can be used to increase the system's security, privacy and usability properties.

This dissertation investigates the *human component* of a biomedical cyber-physical system and, in doing so, it focuses on three guiding questions:

1. How do (how could) biomedical cyber-physical systems use users' idiosyncrasies to potentially improve a system's performance?
2. What security and privacy vulnerabilities may arise from users' unique traits and system interaction idiosyncrasies?
3. How can users' idiosyncrasies be leveraged to increase systems' security and privacy?

These question are answered in the context of two emerging biomedical technologies, *brain-computer interfaces* and *teleoperated robotic systems.*

### *1.1 Brain-Computer Interfaces*

Brain-computer interfaces (BCIs) represent an augmentative communication and control technology which enables direct communication between the brain and the external environment through the use of different electro-physiological signals. The initial motivation for the development of BCIs came from the medical field, where these devices have been used to provide basic communication capabilities to people suffering from neuromuscular disorders [231]. In recent years, BCI-enabled communication has had a surge in popularity in non-medical applications, such as advertising and marketing, fiction and gaming, for example, [17, 12, 15, 16]).

The expansion in capabilities and application space opens up, however, a host of questions related to potential inappropriate use and misuse of BCIs. Leveraging recent neuroscientific results, for example [203, 115], BCIs can be misused to extract private information about users' memories, prejudices, religious and political beliefs, and possible neurophysiological disorders. In order to improve privacy and security properties of emerging BCI technologies, this dissertation seeks to answer the following research questions:

(i) How could BCIs be misused to extract users' private information?

(ii) Focusing on non-invasive BCIs, which components of the electroencephalogram (EEG) are most suitable for private information extraction?

(iii) How do we quantify the amount of exposed information?

(iv) How do we prevent and mitigate the identified privacy threats?

To investigate the extent to which is extraction of private information possible with BCIs, in Chapter 3 we design and execute a series of experiments with human subjects. We present users with a variety of visual stimuli, and analyze their responses to the presented stimuli. We focus on the Event Related Potential (ERP), neural responses associated with specific sensory, cognitive and motor events [141], and investigate the feasibility of different ERP components, such as P300, N400, P600 and ERN (Error Related Negativity), for private information extraction.

To quantify the amount of exposed private information, in Chapter 4 we map the considered privacy attack into the communication-theoretic setting, and show that it can be modeled as a two-user multiple access channel with generalized feedback (MAC-GF) [56]. We then propose the success probability and the equivocation rate as metrics quantifying the attack impact, and show the prevention of the considered privacy threat to be similar to information hiding in communication [46].

Based on our hypothesis that electroencephalograms can be decomposed in real time into characteristic components, which provide sufficient information about a user's conscious and intended messages, in Chapter 5 we propose an approach to prevent the identified privacy threats, referred to as the BCI Anonymizer. The BCI Anonymizer is an interface between the EEG electrodes (BCI sensors) and BCI applications. It takes raw electro-physiological signals as inputs, and decomposes them into a collection of characteristic components. Upon request, instead of the complete signal, the BCI Anonymizer provides an application with a subset of requested components [60].

### *1.2 Teleoperated Robotic Systems*

Teleoperated robotic systems, which allow human operators to control remote robots through a communication network, are envisioned to soon be used in combat zones, in space and underwater missions, as well as in areas of natural disasters and underdeveloped areas. In such circumstances, robots' portability becomes important, as well as their ability to operate with limited power resources, in challenging environments where basic infrastructure may not exist. Additionally, in those areas, operator-robot communication over available networks may be targeted by attackers, exposing the whole system to cyber security threats.

The potentially open and uncontrollable nature of the communication medium may allow attackers to jam, disrupt, or take over the communication between an operator and a robot.

To render teleoperated robotic systems secure against possible attacks, Chapter 7 focuses on two questions:

(a) How would an attacker compromise a teleoperated robotic system?

(b) What the applications of such a cyber security attack might be?

We identify possible attacks, and based on their impact on teleoperators, classify them into intention manipulation, intention modification and hijacking attacks [47].

In Chapter 8, we then analyze the feasibility of the identified attacks, and using the next generation teleoperated surgical platform, the *Raven II*, demonstrate that an attacker can currently control a wide range of a robot's functions, manipulate a robot's feedback, and even completely override an operator's inputs. Moreover, we show that it is possible to abuse the robot's existing emergency stop mechanism to execute efficient (one packet) denial-of-service attacks [47]. Through a series of experiments involving human subjects [47, 50], we then assess the actual impact level these attacks have on operators. Our experiments are based on established robotic surgery tasks, and we quantify the impact using the following metrics: the overall procedure (trial) time, the subjective assessment of difficulty, and the Fitts' index of difficulty.

We introduce Fitts' law as a novel way of quantifying the impact of cyber security attacks on cyber-physical systems. It is a formal method of characterizing subjects' performance in terms of the duration of point-to-point reaching movements, and in this dissertation, we use it to:

(A) Quantify the increase in difficulty of a teleoperated robotic procedure when a system is under attack,

(B) Establish impact equivalence between different attacks, and

(C) Predict the impact of other possible attacks.

Using Fitts' law, we further show that an attack does not impact all components/subtasks of a teleoperated robotic procedure equally. Typically, different attacks have a less prominent

impact on the "free movement" component than on the "fine motor (homing)" component of a procedure [50].

In Chapter 9, we propose resorting to a multi-layer approach to security and privacy in order to enhance safety, security, privacy and reliability of teleoperated robotic systems. We first propose a few simple changes to the communication protocol used for teleoperation, which would make teleoperated procedures resilient to some of the identified attacks [47]. Based on our hypothesis that every teleoperator has a unique way of interacting with a robot and with the environment, we then propose two new approaches to reliably identify and authenticate human operators (*haptic passwords*) and to quickly detect potential attacks on teleoperated robotic surgery (*operator signatures*) [61]. The operator signatures approach leverages the available information about a surgeon's haptic device, a robot's end effectors, and the exchanged messages, in order to detect malicious activities, but also any potential robot's and surgeon's anomalies during a procedure.

Due to their emerging nature, security- and privacy-focused research around brain-computer interfaces and teleoperated robotic systems has involved several common steps:

(i) Understanding the systems' properties, and identifying possible security and privacy threats,

(ii) Evaluating the identified attacks theoretically and experimentally,

(iii - a) Developing prevention and mitigation tools based on traditional security methods, and

(iii - b) Developing security and privacy tools that leverage cyber-physical properties of the systems.

In addition to developing engineering approaches against possible attacks (steps (iii - a) and (iii - b)), in Chapters 6 and 9 we further advocate for the need for an interdisciplinary effort, involving neuroscientists, roboticists, engineers, computers scientists, legal scholars and ethicists, to develop appropriate standards, policies, regulations and laws, guiding the proper use of emerging biomedical cyber-physical systems, especially in the security and privacy context.

### *1.3 Thesis Contributions*

Focusing on two emerging biomedical cyber-physical systems, this dissertation makes fundamental contributions towards understanding and defining the differences between cyber and cyber-physical systems in the security and privacy context. It enhances the basic knowledge of possible security and privacy threats against emerging *personalized cyber-physical system*. Further, leveraging the unique properties of cyber-physical systems, it proposes novel mitigation and prevention strategies against the identified attacks. These mitigation and prevention strategies are based on the uniqueness of human component within a cyber-physical system. Specific contributions of this work are:

- **Theoretical and experimental analysis of privacy attacks against BCIs:** Chapters 3 and 4 present experimental and theoretical results of the feasibility of extraction of private information using BCIs. Experimental analysis is conducted through a series of experiments involving human subjects, where, as a part of the experiments, subjects were presented with a variety of visual stimuli, and their responses to stimuli are analyzed. To quantify the amount of exposed private information, we propose the success probability and the equivocation rate as metrics, and we show the prevention of the considered privacy threat to be similar to information hiding in communication [46].

- **BCI Anonymizer - the first engineering approach to preventing BCI privacy attacks:** Based on our hypothesis that electroencephalograms can be decomposed in real time into characteristic components, which provide sufficient information about a user's conscious and intended messages, in Chapter 5, we propose an approach to prevent the identified privacy threats. This approach introduces an interface between the BCI sensors and BCI applications, referred to as the BCI Anonymizer. It takes raw electro-physiological signals as inputs, and decomposes them into a collection of characteristic components. Upon request, instead of the complete signal, the BCI Anonymizer provides an application with a subset of requested components [60].

- **Broader impact of the BCI Anonymizer:** In Chapter 6, we advocate that the development of prevention tools against privacy attacks on brain-computer interfaces should be an interdisciplinary effort, involving neuroscientists, neural engineers, ethicists, as well as legal, security and privacy experts. We focus the following issues that may arise: (a) assurance that third-party BCI applications will not be manipulating the BCI Anonymizer, (b) users' level of control over the accessed resources, (c) possible burden on users to understand what information is being accessed, and publicly shared when they use their BCIs, and (d) BCI manufacturers' willingness to add the BCI Anonymizer as a new component to their system. To address these issues, we propose a novel privacy addendum [91] to the traditional agreement between BCI manufacturers, BCI application developers and BCI users as a first step towards enhancing users' privacy.

- **The first taxonomy and experimental analysis of cyber security attacks against teleoperated robots:** In Chapters 7 and 8, we identify possible cyber security attacks against teleoperated robotic systems, and classify them based on the impact they have on a human operator. For each of these classes, we assess the level of the actual impact of an attack on a teleoperated procedure through a series of experiments involving human participants. Our experiments are based on established robotic surgery tasks.

- **Haptic Passwords - novel biometric-based approach to identification and authentication in teleoperated robotic systems:** In Chapter 9, we propose a novel biometric technology, based on human operator's interaction with a haptic device, which provides an operator with a sense of touch. Our technique uses wavelet-based analysis to extract operators' unique haptic interaction features. The extracted features are then used as inputs to a neural network which performs operators' identification and authentication. Our experimental results show that the proposed haptic-based authentication system has a high identification accuracy, and that it is resistant to forgery attacks.

- **Operator Signatures - the novel monitoring and detection technique for teleoperated robotic systems:** In Chapter 9, we also introduce the concept of operator signatures, a new approach to monitor, analyze, and validate performance of human operators. This approach is based on the assumption that each operator interacts with a remote robot in a unique way, thus generating a unique biometric (signature), which can be extracted and used for further validation.

- **The first attempt at regulatory guidance towards enhanced security for teleoperated robotic systems:** In Chapter 10, we discusses preliminary legal and policy applications of operator signatures for teleoperated robotic surgery. We focus on the three main task operator signatures seek to address: identification, authentication, and evaluation. We also discuss possible legal benefits from operator signatures. In particular, we discuss how operator signatures can refine the standard of care for robotic surgical procedures, and how they may, possibly for the first time, provide objective empirical evidence of an individual operator's actions during robotic surgery.

Chapter 2

# PRIVACY CONCERNS RELATED TO BRAIN-COMPUTER INTERFACES

## *2.1 Brain-Computer Interface*

A BCI is a communication system between the brain and the external environment. In this system, messages between an individual and an external world do not pass through the brain's normal pathways of peripheral nerves and muscles. Instead, messages are typically encoded in electro-physiological signals, such as electroencephalograms (EEG), signals directly measuring electrical potentials produced by neural synaptic activities [231, 230, 232].

The initial motivation for the development of BCIs came from the growing recognition of the needs of people with disabilities, and of potential benefits that BCIs might offer. The first BCI was developed in the 1970s [231]. Since then, many research programs have focused on the development of BCIs, for assistance, augmentation and repair of cognitive and sensorimotor capabilities of people with severe neuromuscular disorders, such as spinal cord injuries or amyotrophic lateral sclerosis.

In recent years, however, BCIs have had a surge in popularity in fiction, gaming, entertainment and marketing. There are currently several consumer-grade BCIs (e.g., Emotive System [8], NeuroSky [17], and g-tec Medical Engineering [9]) offering relatively low-cost EEG-based BCIs and software development kits to support and facilitate expansion of BCI-enabled applications. The supported applications can broadly be classified into: (i) *accessibility tools*, such as mind-controlled mouse and keyboard, (ii) *hands-free arcade games*, such as Pong [3], and (iii) *"serious games"*, i.e., games with purpose other than pure entertainment, such as attention and memory training [240].

BCIs are also beginning to be used for gaming and entertainment. In 2011, the Necomimi, an entertainment BCI shaped like cat ears was chosen as one of the top 50 inventions of the year [12], and in May 2013, the first neurogaming conference was held, gathering more

than involved 50 companies [16]. In the same month, Samsung, in collaboration with the University of Texas, presented an attempt at controlling mobile devices via BCIs [26]. Recently researchers from Taiwan proposed a method of predicting success of an online game by analyzing a user's electromyographic (EMG) signals (i.e., electrical signals produced by a user's skeletal muscles) over the first 45 minutes of the game [59].

In addition to gaming and entertainment industry, in recent years, market research companies have also shown an increased interest in BCI-enabled technologies. For example, in 2008. the Nielsen Company, a marketing company studying consumers and their behavior, has acquired Neuro-Focus, a neurological research company. As stated in the press release, the goal of the acquisition was the development of neural engineering technologies aimed at "helping clients understand their consumers and their viewers in more depth, detail, and accuracy than ever before possible" [15].

This expansion in the application space and user population is expected to bring a new wave of energy to BCI-enabled technologies. In particular, the BCIs are expected to become simpler to use, requiring less training time and user effort, while enabling faster and more accurate translation of a user's intended information [230].

### *2.1.1 Components of a BCI*

A BCI is a system used to translate electrophysiological signals, reflecting activity of central nervous system, into a user's intended messages that act on the external world [231]. From an engineering perspective, it is a *communication system*, consisting of inputs (user's neural activity), outputs (external world commands), and components translating inputs to outputs, known as *signal acquisition* and *signal processing*. A high-level block diagram of a typical BCI is depicted in Figure 2.1.

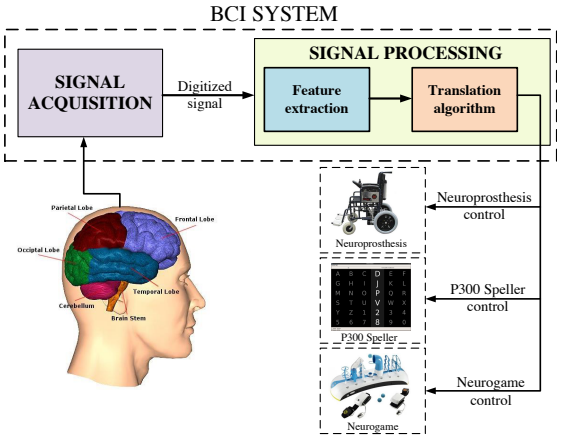



**Figure 2.1:** High-level block diagram of a typical brain-computer interface.

Based on the recording location, BCIs can be divided into:

(i) Invasive,

(ii) Partially invasive, and

(iii) Non-invasive systems.

Invasive BCIs are directly implanted into the brain during a surgery. They enable the highest quality measurements of neural activity. Partially invasive BCIs are implanted inside the skull, typically on top of the brain. They provide signals of lower noise and higher selectivity than non-invasive BCIs, which record neural signals from the scalp. Most non-invasive BCIs are based on electroencephalography (EEG). While known to be susceptible to noise and signal distortion, EEG signals are easy to measure. In addition, EEG-based BCIs have relatively low cost and low risk, which makes them the most widely used BCI devices [230].

Signal processing component of a BCI typically consists of two parts: *feature extraction* and *translation (decoding) algorithms*. BCIs may use neural signals' features in the time domain (such as amplitudes of event related potentials) or in the frequency domain (such as $\mu-$ or $\beta-$ rhythm amplitudes). Often used feature extraction methods are spatial filtering, voltage amplitude measurement and spectral analysis. In this dissertation, we focus on time domain signal features, in particular on the Event Related Potentials (ERPs).

Translation algorithms take abstract feature vectors, reflecting specific aspects of a user's current EEG signals, and transform those vectors into application-specific commands. Depending on an application, many different translation algorithms are being used in BCIs, yet, as pointed out in [231], any effective translation algorithm should be able to adapt to:

- Individual user's signal features,
- Spontaneous variations in recorded signal quality, and
- Adaptive capacities of the brain (neural plasticity).

### *2.2 Event Related Potential*

A important class of currently used non-invasive BCIs uses Event Related Potential (ERP) as a feature of interest. ERP are defined as neural responses associated with specific sensory, cognitive, and motor events [141]. An ERP response typically consist of either positive

or negative voltage peak(s), related to the "higher" brain processes, involving memory, attention or expectation. These patterns of signal changes after specific stimuli can be extracted from the overall EEG by *averaging*.

### *2.2.1 ERP Techniques in Neural Engineering*

The first ERP recordings from awake human subject were obtained in the mid 1930s [141]. The modern era of ERP research began, however, several decades later (in the mid 1960s), with the discovery of the first cognitive ERP component, referred to as the Contingent Negative Variation (CNV). The next major advance was the discovery of the P300 ERP component in the 1965. Since then, a large number of ERP components have been discovered and used in neuroscience. These components can broadly be classified into [141]: (a) visual sensory responses, (b) auditory sensory responses, (c) somatosensory, olfactory and gustatory responses, (d) language-related ERP components, (e) error detection, and (f) response-related ERP components.

Among these classes, more prominent components, which are being investigated as a part of this project, are: P300, N400, P600, and error-detection ERN (error-related negativity), thought to reflect the following [141]:

- **P300** - cognitive processes involving stimulus evaluation or categorization (the most often used ERP response);
- **N400** - reaction to a meaningful or potentially meaningful stimulus, including words, pictures, sounds, smells or faces (language-related ERP component);
- **P600** - a reaction to hearing or reading a grammatical error or other syntactic anomaly (language-related ERP component);
- **ERN** - processes occurring after an error is committed in multiple-choice tasks, even if a person is not explicitly aware of that error (error detection ERP component).

Over the last two decades, a large body of neuroscientific research has investigated and reported how different ERP components can be used to infer information about a user's intent, cognitive and behavioral processes, as well as about his/her affective and emotional states. The P300 ERP component, whose example is shown in Figure 2.2, is typically

observed over the parietal cortex as a positive peak at about 300 milliseconds after the stimulus[1]. This component is typically elicited as a response to infrequent or particularly significant auditory, visual or somatosensory stimuli, when interspread with frequent or routine stimuli. This component is one of the most widely used, and one of the important advantages of the P300-based BCIs is the fact that the P300 is typical, or naive, response to a desired choice, thus requiring no initial user training.

The best known application of the P300 response is a spelling application, the P300 Speller, proposed and developed by Farwell and Donchin in 1988 [83]. More recently, the P300 response has been used to recognize a subject's name in a random sequence of personal names [203], to discriminate familiar from unfamiliar faces [150], and for lie detection [28].

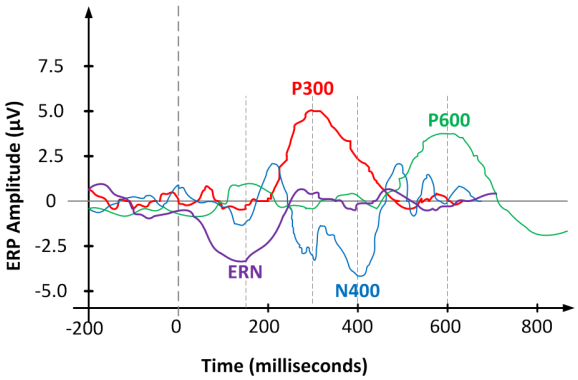



**Figure 2.2:** Examples of Event Related Potential (ERP) components of an EEG recording. Red curve denotes a typical P300 response, blue curve a N400 response, green curve a P600, and purple curve a typical ERN response.

Another well-investigated component is the N400 response, associated with semantic processing, and typically observed as a negative peak at about 400 milliseconds after the stimulus. The N400 has recently been used to infer what a person was thinking about, after he/she was primed on a specific set of words [224]. This ERP response has also been linked to the concept of *priming*, an observed improvement in performance in perceptual and cognitive tasks, caused by previous, related experience. Similarly, the P600 component has been used to make an inference about a user's sexual preferences, and estimates of a user's anxiety level, derived from his/her EEG signals, has been used to draw conclusions about that user's religious beliefs [115].

[1]The component's name typically reflects the time of the peak and whether the peak is positive (P) or negative (N).

### *2.2.2 Advantages and Disadvantages of ERP Techniques*

ERPs have several important advantages over other behavioral and physiological measures, that make them especially suitable as features of choice in BCIs [141]:

(A1) There are no fundamental restrictions on the amount of ERP data that can be collected from a user, since ERP responses are recorded using non-invasive EEG technique.

(A2) ERP responses have a *temporal resolution* of 1 millisecond (under optimal conditions, even better than 1ms).

(A3) ERP responses can provide a continuous measure of processing between a stimulus and a response, thus making it possible to determine which processing stages are affected by a specific experiment.

(A4) ERP responses can provide an online measure of processing of stimuli, even when there is no behavioral response.

On the other hand, BCI responses are known to have fairly poor spatial resolution. In addition, these signals have a relatively small amplitude, thus typically requiring a large number of trials in order to measure them accurately. Due to their essentially binary nature (i.e., the response to a specific stimulus is either elicited, or it is lacking), the ERP responses are very good for tasks such as spelling and speech synthesis, but they exhibit a poor performance in motor control tasks, such as cursor movements.

## *2.3 Legal and Ethical Considerations in Neural Engineering*

With an increasing number of neural engineering applications, in particular BCIs and neural imaging, researchers have recognized the need to address emerging ethical and legal questions arising from their use [82, 120, 113, 114, 218, 65]. In 2003, Jonsen introduced *neuroethics* as "a discipline that aligns the exploration and discovery of neurobiological knowledge with human value system". It was then recognized that neuroethics will have to address questions related to [114]:

(a) Incidental findings,

(b) Surrogate and biomarkers of diseases, and

(c) Commercialization of cognitive neuroscience.

In 2005, The Committee on Science and Law considered possible legal implications of neural engineering [218]. An emphasis was put on privacy implications of neural imaging, in particular on the use of neural imaging in non-medical research. The committee recognized *neuromarketing*, defined as the field of marketing research that studies consumers' sensorimotor, cognitive, and affective response to marketing stimuli [137] and *brain fingerprinting*, defined as a technique that purports to determine the truth by detecting information stored in the brain [218], as emerging non-medical areas using neural imaging data. The committee observed important similarities between genetic and brain data, in the following: (1) "both genetic and brain data hold out the promise of prediction (not only disease, but also behavior)," and (2) " both types of information expose unique and personal, and to a large extent, uncontrollable aspects of a person that previously were unobservable" [218]. Based on these observations, the committee proposed exploring and leveraging for neuroethics those medical, ethical and legal rules already set forth in genetic research.

More recently, in [82], Farahany observed that modern neuroscience and neural engineering pose a new set of legal challenges to the existing Self-Incrimination doctrine of the Fifth Amendment, which states that "no person shall be compelled to prove a charge from his own mouth, but a person may be compelled to provide real or physical evidence" [82]. The author presented several examples, showing how is modern neuroscience expected to facilitate evidence collection during criminal investigation. The presented examples strongly indicate that the traditional border between testimonial and physical evidence becomes blurry when applied to the evidences collected by neural engineering techniques. To address the problem, Farahany proposed that an evidence can arise in four different ways:

(E1) From the identifying characteristics inherent to individuals,

(E2) Automatically, without conscious processing,

(E3) Through memorialized photographs, papers, and memories, or

(E4) Through responses uttered silently or aloud.

The author then proposed the new spectrum of *identifying, automatic, memorialized and uttered* evidence should be used to appropriately adapt the Self-Incrimination doctrine.

## *2.4 Security and Privacy Considerations in Neural Engineering*

### *2.4.1 The Use of Neural Data for Identification and Authentication*

Based on the observation that neural signals of each individual are unique and can therefore be used for biometrics [150], many researchers have recognized potential benefits of using neural data for user *identification*, defined as the identity selection out of a set of identities, and *authentication*, defined as verification that the claimed identity is valid (authentication) [150, 192, 183, 182]. EEG signals have shown to be particularly useful for these application.

In [192], the authors proposed a method using $\alpha$-rhythm for identification. They reported correct classification scores in the range of 72% to 84%. Further, in [196], the authors proposed an EEG-based identification methods, that uses data collected only from the two frontal electrodes. In [194], the authors present an overview of biometric identification methods based on EEG, electrocardiogram (ECG) and the skin conductance signals, also known as electrodermal response (EDR).

In [150], the authors proposed an authentication methods based on Gaussian Mixture Models and maximum *a posteriori* model adaptation. The authors further investigated the practicality of different mental tasks for authentication, and showed that some task are more appropriate than others. Finally, [40] proposed that neural data can be used to prevent coercion attacks (also known as rubber hose cryptanalysis), where a user is forced to reveal cryptographic secrets known to him/her. The proposed approach is based on the idea of *implicit learning*. Instead of asking a user to consciously memorize a secret and use it for identification and authentication, in the proposed approach a user is identified and authenticated based on specific patterns he/she learned and can use without ever being aware that he/she knows these patterns.

### *2.4.2 Neurosecurity. Privacy and Security Challenges in Neuroscience*

Advances in neural engineering are expected to continue driving improvements in both medical and non-medical neural applications. In particular, it is expected that the number of implanted neural devices will continue increasing in the future [88]. In 2009, Denning

et al. [76] recognized that "the use of standard engineering practices, medical trials, and neuroethical evaluations during the design process can create systems that are safe and that follow ethical guidelines; unfortunately, none of these disciplines currently ensure that neural devices are robust against *adversarial entities* trying to exploit these devices to alter, block, or eavesdrop on neural signals". The authors identified potential security threats that can be mounted against implanted neural devices, and introduced the term "neurosecurity" as "the protection of the confidentiality, integrity, and availability of neural devices from malicious parties with the goal of preserving the safety of a person's neural mechanisms, neural computation, and free will" [76].

## *2.5 Problem: Brain Malware in BCIs*

Probably the most concerning BCI-related example was showcased at the 2012 USENIX Security Symposium [102, 32]. The authors presented "brain spyware", the first malicious software designed to infer users' private information using a BCI [153]. They used a commercially-available BCI to present a user with visual stimuli and record his/her EEG neural signals. Focusing on the P300 response, the authors analyzed the recorded signals in order to detect a user's: (a) chosen single digit, (b) banking information, (c) month of birth, (d) location of residence, and (e) if a user recognized the presented set of faces.

While the authors [153] focused only on the P300 response, it is not hard to imagine brain spyware applications being developed to extract private information about users' memories, prejudices and beliefs, but also about their possible neuro-physiological disorders. Currently, there does not seem to exist a way to resist these attacks. Moreover, recent results [140] show that attempts at willful deception can themselves be detected from an individual's neural signals. Going a step further, the same authors [140] show that non-invasive brain stimulators, emitting imperceptible DC electrical currents, can be used to make a user's responses noticeably slower when attempting to lie.

### *2.5.1 Scope of the Brain Malware Problem*

Information extracted using brain "spying" applications, such as the presented brain spyware, might be of interest to multiple parties, and one can easily imagine the following

examples of concerning BCIs use:

**Example 1:** As exemplified in Farahaney's work [82], an access to an individual's memories and emotional responses might be used by police enforcement and government agencies during criminal investigation, as well as for crime and terrorism prevention.

**Example 2:** BCI-recorded neural signals may be used in a variety of entertainment and relaxation applications. A person's emotional response and satisfaction/annoyment level may, for example, be used to provide better (more accurate) music and/or movie recommendations. Similarly, information about a person's activity and anxiety levels may be used to tailor a more personalized training routine or a relaxation session.

**Example 3:** Personal information, extracted from neural signals, could also be used for targeted advertisement, where in addition to (or instead of) information about a person's activities on the Internet, an advertiser/retailer would have a real-time access to a person's level of interest, satisfaction, or frustration with the presented material.

**Example 4:** On the other end of the spectrum, however, the extracted information about a person's memories, prejudices, beliefs or possible disorders could be used to manipulate a person or coerce her/him into doing something unpleasant or potentially illegal.

**Example 5:** Finally, the extracted neural information could also be used to cause physical or emotional pain to a person. Examples of such actions have already been noted in the security literature. Denning et al. [76], presented the case of individuals who placed flashing animations on epilepsy support webpages, eliciting seizures in some patients with photosensitive epilepsy.

Summing up the presented examples, it is clear that the impact of exploiting or mishandling BCI-enabled technologies may be severe. So the question arises: is it in the public interest that other entities, private organizations, or even the government have an unrestricted access to all the private information that can be extracted from a person's neural activity? It appears there are currently no regulations preventing the access to the extracted information, outside of medical HIPAA privacy and security rules [10]. Thus, privacy and

security concerns arising from (mis)use of BCIs represent an important issue that deserves immediate attention and careful consideration.

### *2.5.2 Brain Malware Attack Model*

As a first step towards preventing brain malware, in this dissertation, we consider an attacker who uses non-invasive BCI devices, mostly intended for consumer use, to extract private information about users. Manufacturers of non-invasive EEG-based BCIs currently often distribute software development kits with their products, as well as technical support [153]. Their intention is to promote application development, but such "open- development" platforms may compromise users' privacy and security, since there is currently no review process, standards and guidelines in place to protect users, nor technical protection to restrict inappropriate or malicious BCI use.

As depicted in Figure 2.1, a typical BCI system consists of three main components: an acquisition system, an application, and a signal processing system, and the existing BCI "open development" platforms typically grant every application developer a full control over all of these components. We therefore assume that an attacker has an access to all of these resources, and analyze how can an attacker uses these resources to develop malicious applications.

#### *Attacker Type*

Based on the way an attacker uses the BCI technology, we distinguish between two types of attackers. The first type of an attacker extracts users' private information by hijacking the legitimate components of a BCI. Such an attacker exploits for malicious purposes those feature extraction and decoding algorithms that are intended for the legitimate BCI applications. The second type of an attacker extracts users' private information by adding or replacing the legitimate BCI components. Such an attacker implements additional feature extraction and decoding algorithms, and either replaces or supplements the existing BCI components with additional malicious code.

*Methods of Extracting Private Information*

We further focus on scenarios where an attacker interacts with users by presenting them with specific sets of stimuli, and then records their responses to the presented stimuli. In the current literature, there are several well-established methods of presenting stimuli to users:

(S1) **Oddball paradigm** - a technique where users are asked to react to specific stimuli, referred to as *target stimuli*, hidden as rare occurrences in a sequence of more common, non-target stimuli [111].

(S2) **Guilty knowledge test** - a technique based on the hypothesis that a familiar stimulus evokes a different response when viewed in the context of similar, but unfamiliar items [233].

(S3) **Priming** - a technique that uses an implicit memory effect where one stimulus may have an influence on a person's response to a later stimulus [224].

We assume that an attacker can use any of these methods to facilitate extraction of private information. In addition, an attacker can present malicious stimuli in an overt (conscious) fashion, as well as in a subliminal (unconscious) way, with subliminal stimulation defined as the process of affecting people by visual or audio stimuli of which they are completely unaware [34]. Overview of subliminal stimulation is given in Chapter 3, but generally speaking, ways to achieve unawareness typically include reducing either a stimulus intensity or its duration below the required level of conscious awareness.

## *2.6 Summary*

Privacy and security threats arising from the use of BCI-enabled technologies may not pose a critical concern at this moment, given a fairly limited deployment of BCIs outside of research and medical communities. We believe, however, that the right time to address potential issues is now, and propose that methods to prevent and mitigate BCI-enabled privacy and security threats should be developed in the early design phase, and embedded throughout the entire life of the technology.

We view the development of these prevention and mitigation tools as an interdisciplinary effort, involving neuroscientists, neural engineers, ethicists, as well as legal, privacy and security experts. In this Chapter, we take the initial step towards facilitating the necessary interdisciplinary discussion by analyzing the current BCI technology, and by proposing the first model of a brain malware attacker.

### *2.7 Acknowledgment*

This work is supported by Award Number EEC-1028725 from the National Science Foundation, and by the University of Washington Tech Policy Lab. The content of this Chapter is solely responsibility of the authors, and does not necessarily represent the official views of the National Science Foundation.

We thank Jeffrey Herron and Charles Matlack, and Professors M. Ryan Calo, Howard J. Chizeck, Tadayoshi Kohno, and Jeffrey G. Ojemann for helpful suggestions, and ideas that have inspired and enhanced this work.

# Chapter 3

# EXPERIMENTAL ANALYSIS OF BRAIN MALWARE

In this Chapter, we experimentally investigate how feasible it is to extract private and sensitive information about BCI users, based on their recorded neural signals. In doing so, we use commercial-grade non-invasive BCI devices, and we focus on EEG signals, in particular on the Event Related Potential (ERPs) components of an EEG. We investigate several different ERP components – P300, N400, P600, and ERN, but our main focus is on the P300 component, and its feasibility as a brain malware signal of interests, when it is being used in an oddball paradigm.

We specifically focus on the feasibility of private and sensitive data extraction when stimuli are being presented to a user in a subliminal fashion, i.e., in a manner where a user is not consciously aware of the presented stimulus, but (s)he may still be reacting to it. We see this type of brain malware as a most malicious one, but also the most likely one. In order to collect sufficient data from a user, an attacker will likely want to conceal his/her presence, and the fact that there might be anything out-of-the-ordinary with the used BCI application, and, due to their nature, subliminal stimuli might be just the stimuli type to enable that. We thus continue this Chapter by providing a brief overview of the relevant research on the use of subliminal stimuli in neuroscience.

## *3.1 Preliminaries: Subliminal Stimuli in Neuroscience*

*Subliminal stimulation* is defined as the process of affecting people by visual or audio stimuli of which they are completely unaware [34]. Ways to achieve unawareness typically include reducing a stimulus' *intensity* or *duration* below the required level of conscious awareness. Some authors suggest that there exist two types of unconscious perception:

- **Subjective**, where a user cannot report the stimuli (but may be able to detect it),
- **Objective**, where a user cannot detect the stimuli [173].

Most authors agree, however, that the required level for visual subliminal stimuli is below 100 milliseconds [188].

Since its inception in the 1957, perhaps no issue in neuroscience (and the psychology of persuasion) has drawn more controversy than subliminal stimulation. Many subliminal alarmists hypothesized "that what we do not see and what we do not know can hurt us" [51]. Over the next few decades, there were several highly publicized examples of subliminal stimulation, even though there were no objective methods to measure the stimuli effect, nor were the presented results replicable. In 1957, the marketing firm reported a dramatic increase in soda and popcorn sales upon subliminally presenting movie-goers with messages "Drink Coca-Cola" and "Eat popcorn" [51]. In 1970, researchers subliminally presented subjects with pictures of octagons, and the exit questionnaire responses indicated the subjects clearly preferred the flashed octagons over the other geometric shapes they have not seen.

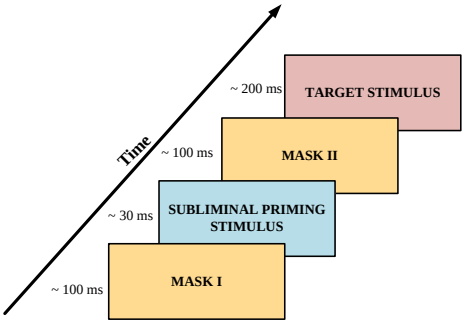



**Figure 3.1:** A graphical representation of subliminal (masked) priming. During one trial of the experiment (approximately 450 ms), a user is presented with the first mask, whose duration is typically about 100ms, followed by a *subliminal (masked) stimulus*, whose duration is typically about 30ms. The masked stimulus is followed by the second mask, whose duration is again about 100ms. Finally, the second mask is followed by a visible (conscious) *target* stimuli.

More recently, with advances in neural engineering (and neural imaging, in particular), interest in subliminal stimulation research has reignited. Researchers became particularly interested in the idea of *priming*, the process of perceiving an object in order to improve accuracy and speed of its recognition in subsequent encounters [35], and its extension, *masked (subliminal) priming* [188, 35, 75, 74, 99, 72, 73, 71]. In masked priming, a user is presented with visual stimuli that are flashed so briefly they cannot be consciously observed (referred to as *subliminal (masked) stimuli*), yet they can facilitate the subsequent processing of related visible (conscious) stimuli, referred to as *target stimuli*. A graphical representation of masked priming is depicted in Figure 3.1.

In [35], the authors reported a 17% improvement in the ability to name an object upon

masked priming. Authors of [75] reported masked (subliminal) stimuli have a measurable influence on both electrical and haemodynamic measures of brain activity. In the subsequent work [72], the authors showed that the priming effect with subliminal numbers has observable effects on both semantic and motor levels of brain activity. These results provided a clear evidence that semantic-level processing of masked primes is indeed possible. In the same work, the authors, however, showed that subliminal primes have a very short effect, typically lasting less than 500 milliseconds. Finally, in [73] the authors showed that subliminal priming can be made crossmodal, for example, from visual priming stimuli to auditory target stimuli.

### *3.2 Experimental Logistics and Subjects Demographics*

Relying on the presented results with subliminal stimulation, in our experimental study we seek to answer if subliminal stimuli, used in an oddball paradigm, can be used to infer information about users based on their P300 responses to the presented stimuli. All experimental data presented in this Chapter is collected as a part of the study "Privacy and Security by Design in Brain-Computer Interfaces", approved by the University of Washington Institutional Review Board (approval # 45241 - B). We collected the data from nine subjects, eight of which are undergraduate and graduate students from the University of Washington. More information about subjects' demographics can be found in Appendix A, table A.1. Out of the collected data sets, however, we only used seven for experimental analysis, since two data sets were not complete.

We acknowledge that in our experiments, we resorted to, and were approved to use *deception* towards our subjects. Due to the nature of subliminal stimulation, we did not inform the subjects before the experiment that they would be presented with subliminal sequences of images, words or numbers while they were participating in the experiment[1]. This withholding of complete information from subjects was necessary since any notion of subliminal stimuli would have impacted a subject's alertness and concentration, and would have therefore negatively impact the results of the study. In addition, there was no

[1]Subjects were, however, generally aware that they are participating in the experimental study whose purpose was to enhance privacy of BCI-enabled technologies

other alternative to quantifying the difference in extraction capabilities when subliminal and conscious stimuli were used.

In all experiments involving subliminal stimuli, however, we put strict limits on the kind of information we try to extract, and we focus only on *personally relevant information*, which we define as information that a subject has already voluntarily and overtly provided before any experiment even started. Examples of personally relevant information that we use in this study include:

(I1) Two **numbers** that have a special meaning to a subject, and that a subject is likely to immediately recognize (for example, a part of his/her phone number, year of birth, license plate number).

(I2) Two **names** that have a high significance and/or a special meaning to a subject (for example, a subject's mom's first name, his/her sibling's first name, his/her significant other's first name, or the name of his/her pet).

(I3) Two **places (geographical locations)** that have a special meaning to a subject (for example, a subject's hometown, his/her favorite vacation place, or the city (s)he would very much like to visit).

### *3.3 Experimental Setup*

This experimental study consisted of several phases.

**Experimental Questionnaire:** At the beginning of an experimental session, a subject is introduced to the study, and then provided with a de-identified (coded) questionnaire, asking for personally relevant information (I1)–(I3). The used questionnaire is provided in Appendix A. The provided personally relevant information is then used to generate several user-specific BCI tasks.

**Recording Equipment Set Up:** Once user-specific BCI tasks were generated, a subject is fitted with a soft, stretchy EEG cap with holes in specific locations corresponding to the electrode placement sites. Seven EEG electrodes are fitted into these holes so that they sit near a subject's scalp, and a conductive gel is squirted into the electrode housing to improve the electrical connection. A reference electrode is clipped to the subject's earlobe, and one EMG electrode is placed on a forearm of the subject's choosing.

**Experimental BCI Tasks:** In order to collect experimental data used for further analysis, the subject is asked to play 10 instances of a BCI game, the *Flappy Whale*. Details of the game are presented in subsection 3.3.2. During each of the instances of the Flappy Whale game, the subject is presented with subliminal stimuli. Subliminal stimuli are chosen and presented in the same order to all of the subjects:

(S1) Logos of international fast food chains,

(S2) Photos of famous and anonymous persons,

(S3) Logos of car makes,

(S4) Logos of local and national coffee shops,

(S5) Black-and-white names, including the two names relevant to the subject,

(S6) Logos of national and international banks,

(S7) Black-and-white numbers, including the two numbers provided by the subject,

(S8) Logos of national sports leagues,

(S9) Black-and-white places, including the two places of significant meaning to the subject,

(S10) Logos of national coffee shops and one relevant name provided by the subject.

**Training Data Collection:** Once all experimental data is collected, we ask the subject to participate in several simplified instances of the Flappy Whale game, where (s)he is presented with overt visual stimuli. The presented stimuli include:

(a) (**P300 Training**) Overt numbers from 0–9, where the subject is beforehand asked to focus on one specific number,

(b) (**N400 Training**) Overt words spelling out two sentences, that differ only in the last word: *I drink coffee with milk and sugar* and *I drink coffee with milk and socks*, and

(c) (**P600 Training**) Four overt sentences, consisting of similar words, but with differing meanings: *(i) The cat from the mice fled ran across the room.*; *(ii) The cat that fled from the mice ran across the room.*; *(iii) The mice that from the cat fled ran across the room.*; and *(iv) The mice that fled from the cat ran across the room.*

**Conclusion and Exit Questionnaire:** Once all experimental and training data is collected, we help the subject remove the recording equipment, and clean any remaining conductive gel. The experimental session is concluded by asking the subject to fill out an exit questionnaire (example provided in Appendix A). The whole experimental session takes on average 2.5 hours.

### *3.3.1 Experimental Questions and Approach*

The overarching question in this experimental study is whether private and sensitive information about a BCI user can be extracted using subliminal stimuli, presented via the oddball paradigm, and by analyzing the user's P300 response to those stimuli. During the experimental design, however, this question has given rise to several related questions:

(Q1) Does **the type** of the presented stimuli (e.g., names vs. photos of famous and anonymous people) have an effect of the feasibility to extract private and sensitive information about a BCI users? If yes, which stimuli type best enables private information extraction?

(Q2) Does **the duration** of a subliminal stimulus on the screen have an effect on the feasibility of private information extraction?

(Q3) Does **the position** of a subliminal stimulus on the screen have an effect on the feasibility of private information extraction?

(Q4) Does **priming** have any effect on the feasibility of private information extraction? If so, is *implicit priming* sufficient to improve extraction feasibility?

(Q5) Could different ERP components (e.g., N400, P600) of an EEG signal also be used to extract private and sensitive information about a BCI user in a subliminal fashion?

(Q6) How aware and alert are users to potential unusual events during a BCI game?

In order to address question (Q1), we design an experimental session to consist of ten different BCI games per user. In every game, different type of subliminal stimuli is used. During the design phase of this experiment, we came to a conclusion that the answer to (Q2) is *yes*. Moreover, we observed that subjects were able to consciously observe, and even recognize the fast food and the sport leagues logos if those logos were presented on

the screen for any amount of time longer than 10 milliseconds. In order to address that, and to make the presented stimuli subliminal for the majority of subjects, throughout this study, we present all instance of subliminal stimuli on the screen for only 7 milliseconds. This duration was determined empirically, and it was a result of technical limitations of the screen used in the study. Due to its maximal refresh rate of 144Hz, any stimuli duration shorter than 7ms was not achievable. Moreover, during the design phase of the experiment, we observed a close relation between questions (Q2) and (Q3). In order to address question (Q3) further, in different BCI games, subliminal stimuli were placed at different positions on the screen, as presented in Figure 3.2:

(P1) Next to the score information,

(P2) In the white cloud above the whale,

(P3) In the woods behind the whale, or

(P4) On top of the whale.

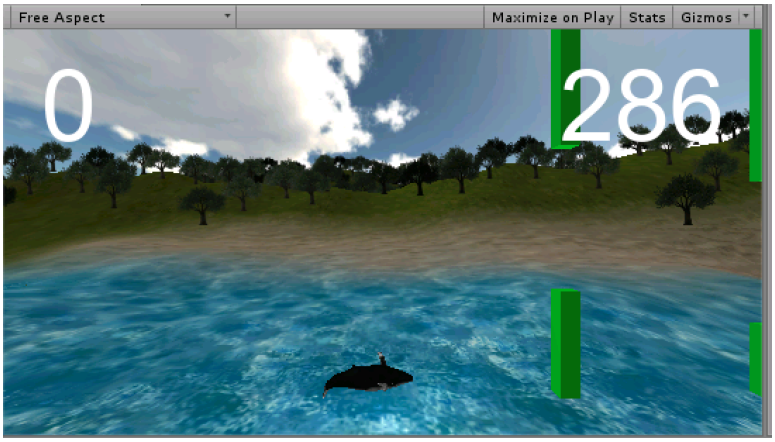



**Figure 3.2:** Screen shot from the BCI game Flappy Whale, used as a part of the brain malware experimental study. The position of the whale in the game is controlled through an EMG signal recorded from a subject's forearm.

### *3.3.2 Flappy Whale - Experimental Brain Malware with Subliminal Stimuli*

A screen shot from the *Flappy Whale* game is depicted in Figure 3.2. In this BCI game, the position of the whale on the screen is controlled through an EMG signal, recorded from a subject's left or right forearm. The goal of the game is to move the whale through the

arriving tubes, without touching them. Every time the whale moves through the tubes without touching them, positive points are assigned to a player, depending on the game level. If, however, the whale touches the tubes, depending on the game level, certain number of points is lost. As the game progress, and a player becomes more proficient, the game becomes harder, as the tubes start arriving at the whale faster, and the distance between them decreases.

The Flappy Whale game was originally developed as a part of the Center for Sensorimotor Neural Engineering 2014 Tech Sandbox competition [4], and it was repurposed for this experimental study. Two main changes we made were:

(C1) The addition of subliminal stimuli within the game, and

(C2) The addition of logging capabilities about events occurring during the game.

The Flappy Whale game is adapted into a subliminal brain malware by adding five different visual stimuli in each game. Depending on their type, the stimuli are placed in different locations on the screen, but duration of a single stimulus is 7 milliseconds, corresponding to one frame of the game. Each stimulus is repeated exactly ten times during one five-minute game, with inter-stimulus interval randomly chosen between 3 and 7 seconds.

### *3.4 Data Collection and Preprocessing*

Throughout this experimental study, we used the g-tec bioamplifier to record subjects electrophysiological signals (EEG and EMG). Additionally, during every five-minute instance of the Flappy Whale game, we record information about presented stimuli, as well as about a subject's actions. Overall, for every game instance, experimental data set consists of the following information:

(D1) **EEG signal**: Based on the recommendations by Krusienski et al. [133], EEG signals are recorded from seven EEG electrodes placed in locations Cz, P3, Pz, P4, Po7, Po8 and Oz, in accordance with the International 10-20 electrode placement system [109], as depicted in Figure 3.3.

(D2) **EMG signal:** EMG signal, used to control the position of the whale on the screen, was recorded from a subject's forearm. Both EEG and EMG signals were recorded with sampling rate of 256Hz.

(D3) **Information about subliminal stimuli:** For every presented stimulus, we record its index, the time it first occurs on the screen, the time it first disappears from the screen, as well as its position on the screen (relative $x$ and $y$ coordinates).

(D4) **Information about the whale:** Every time the position or the velocity of the whale changes, information about the whale's $x$ and $y$ coordinates, and its velocities in $x$ and $y$ directions is recorded.

(D5) **Information about the tubes:** Every time the position of the tubes changes, information about the game difficulty level, and about $x$ and $y$ coordinates of the middle point between the tubes is recorded.

(D6) **Information about whale's flaps:** Every time an EMG signal crosses the prespecifid threshold, and a flap command is sent to the whale, information about the occurrence of that event, and $x$ and $y$ coordinates of the whale are recorded.

Recorded data (D1)–(D6) is imported and synchronized in Matlab using the *Brain Hacking toolbox*, and accompanying Graphical User Interface (GUI), developed specifically for this experimental study. A screen shoot of the Brain Hacking GUI is depicted in Figure 3.4.

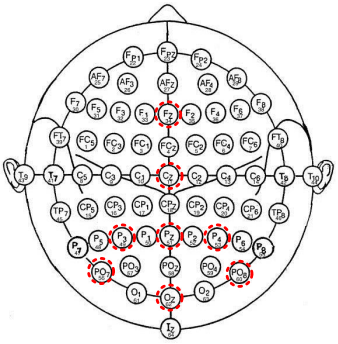


**Figure 3.3:** Graphical representation of EEG electrodes' positions during experimental study. Electrodes naming is based on the International 10-20 electrode placement system.

After synchronizing EEG and EMG data with the presented subliminal stimuli, electrophysiological data is preprocessed, in order to remove potential noise and movement artifacts. We start the denoising process by first epoching the data, i.e., windowing the data such that every EEG/EMG time window consists of 300 milliseconds of the signal before a stimulus, and 700 milliseconds of the signal after the stimulus. An example of the recorded EEG data from a single channel before and after epoching is presented in Figures 3.5 and 3.6.

Once epoched, electrphysiological data is preprocessed by removing the baseline, i.e., by removing the mean value of the signal before the presented stimulus, and filtered with a low-pass and a high-pass filter.

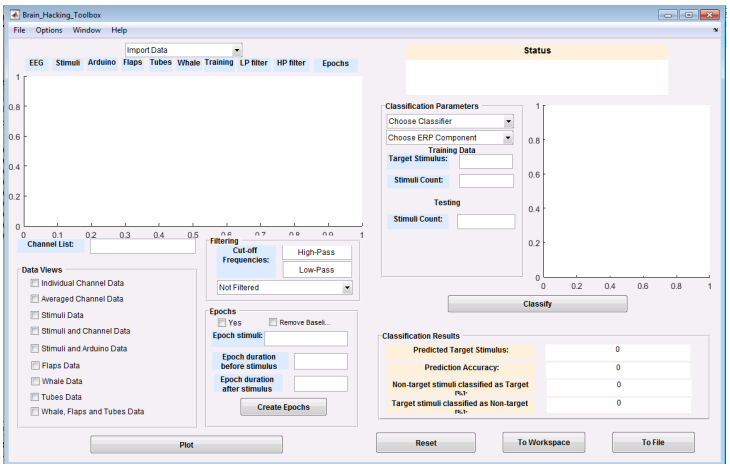



**Figure 3.4:** Screen shot of the Brain Hacking toolbox GUI, developed specifically for this experimental study. The toolbox allows experimental data to be imported in Matlab, different data sources to be synchronized in time, and preprocessed for further analysis.

## *3.5 Data Analysis and Feasibility of Subliminal Brain Malware*

Lastly, in order to determine the feasibility of subliminal stimuli for extraction of BCI users' private and sensitive information, we resort to binary classification, and search for a subject's preferred (target) stimulus. We use the data collected during the training phase to train the classifier, and the rest of the data, collected during the first ten instances of the Flappy Whale game for analysis of subliminal brain malware feasibility. Relying on recent results by Krusienski et al. [133], we use a simple classifier, the Support Vector Machine (SVM), to determine what a subject's preferences are towards the presented stimuli.

### *3.5.1 Preliminaries: Support Vector Machine*

Determining the presence or the absence of an ERP component in an EEG signal can be understood as a *binary classification problem*, and the discriminant function of such a problem can be seen as having a decision hyperplane defined as [133]:

$$wf(x) + b = 0(1) \tag{3.1}$$

where $x$ represents the feature vector, $f(x)$ a transformation function, $w$ the vector of classification weights and $b$ the bias term. For linear classification methods, function $f(x)$ is simply an identity transformation, $f(x) := x$, whereas for nonlinearly separable problems, function $f(x)$ typically represents a kernel transformation that maps features into a higher

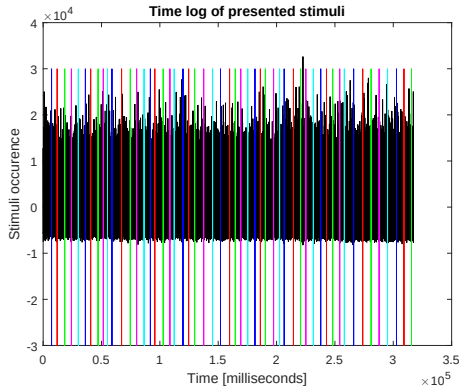



**Figure 3.5:** An example of the recorded single-channel experimental data before epoching.

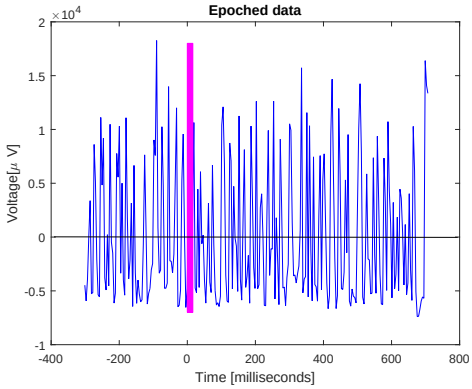



**Figure 3.6:** An example of the recorded single-channel experimental data after epoching. Electrophysiological data is epoched (divided into windowed data) such that each window consists of 300 milliseconds of the signal before the stimulus, and 700 milliseconds of the signal after the stimulus.

dimensional space in the attempt to create a linearly separable set.

The Support Vector Machine (SVM) classification method is one of the most accurate and most often used classification methods in ERP research [149]. The main idea behind this non-probabilistic binary classification method is to find the separating hyperplane between two classes so that the distance between the hyperplane and the closest set of points from both classes is maximized. Stated differently, in this supervised learning algorithm, our goal is to maximize the margin between the two classes, and this goal can be expressed as the following optimization problem [225]:

$$\begin{aligned} &\text{minimize} \|w\| \\ &\text{subject to } y(w \cdot f(x) + b) \geq 1 \end{aligned} \tag{3.2}$$

where $y \in [\pm 1]$ represents class labels.

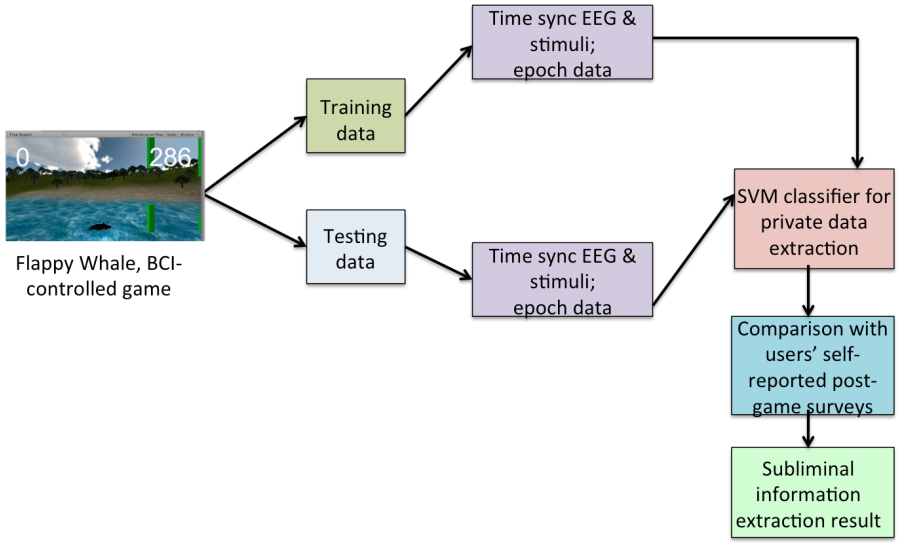



**Figure 3.7:** Simplified block diagram of data analysis sequence. Upon data preprocessing, epoched data is used as input to the SVM classifier, and the output of the classifier is compared with subjects' self-reported preferences.

### *3.5.2 Feasibility of Subliminal Information Extraction*

A simplified block diagram of the whole data analysis sequence is depicted in Figure 3.7. Once classification results are obtained for each of the ten Flappy Whale games, we compare a subject's estimated preferences towards the presented stimuli with those self-reported in the experimental questionnaire. Our preliminary (and limited results, based on experimental data sets from only seven subjects) indicate that there is an overlap between the estimated (classified) and self-reported preferences significantly larger than chance. This result support the hypothesis that extraction of private and sensitive information about BCI users may be possible through subliminal stimulation.

## *3.6 Subjects' Awareness and Alertness*

In order to confirm that the presented stimuli were indeed subliminal for the majority of our subjects, at the end of the experimental session, we provide subjects with an exit questionnaire (example shown in Appendix A). Due to a deceptive nature of the study, however, we are not able to explicitly ask the subjects whether or not they have noticed any stimuli (logos, pictures, or words) during experimental games. Instead, we resort to asking

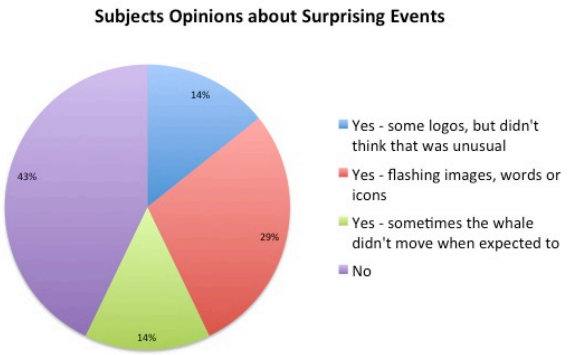



**Figure 3.8:** Experimental results indicating the level of subjects' awareness of the presented stimuli. Based on these preliminary results, it seems that almost a half of BCI users notices presented stimuli, even if they presented on the screen for only 7ms.

a relatively open-ended question: *Did you notice anything surprising or unusual during the experiment?* and a follow-up: *If yes, could you please explain below.*

Our preliminary results (based on a limited data set of only seven subjects) are presented in Figure 3.8. Three of seven people did not notice any of the presented stimuli, whereas three did notice some of the presented stimuli (even though stimuli were only presented on the screen for seven milliseconds). Our assumption is that the remaining subject (one out of seven) found the question too suggestive, and decided that there must have been something, but (s)he missed it.

The fact that almost half of the subjects did actually notice some of the presented stimuli could be taken, at the first glance, as an encouraging sign that the problem of subliminal brain malware may not be as severe as initially thought. Unfortunately, our earlier question (*How do feel about the experiment, in which information was extracted from your brain signals, now that you have participated?*) seems to indicate otherwise, as depicted in Figures 3.9 and 3.10. It appears that before the experimental session begins, half of the subjects, knowing that the purpose of the study is the enhancement of BCI privacy, do not feel completely comfortable about participating. After the experiment, the majority of the subjects feels comfortable. This majority includes those subjects who reported that they did observe something unusual during the experiment. These limited results seem to indicate that even those BCI users who may consciously notice some of malicious subliminal

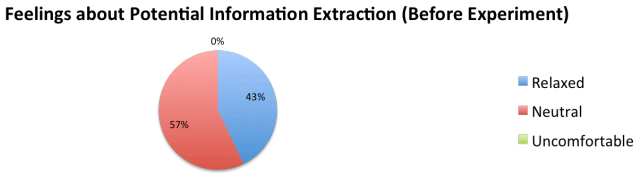



**Figure 3.9:** Experimental results indicating the level of subjects' alertness about the experiment, before the experiment started.

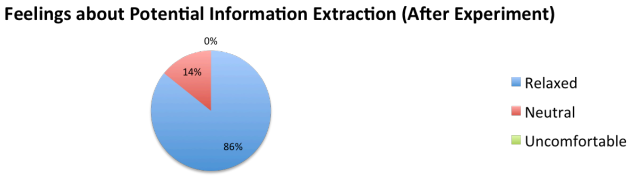



**Figure 3.10:** Experimental results indicating the level of subjects' alertness about the experiment, before the experiment finished.

stimuli may not feel alarmed by their presence.

## *3.7 Summary*

In this Chapter, we experimentally investigate the feasibility of extraction of BCI users' private and sensitive information, based on their recorded electrophysiological signals. We specifically focus on the feasibility of private data extraction when stimuli are presented to a user in a subliminal fashion, i.e., in a manner where a user is not consciously aware of the presented stimulus, but (s)he may still be reacting to it. We further investigate the level of users awareness and alertness to potential nefarious events within their BCI applications.

Our experimental results indicate that subliminal brain malware, where seemingly benign BCI games are used to expose private and sensitive information about a user, based on his/her unconscious and uncontrollable responses to stimuli that (s)he is not aware of, are possible, thus posing a considerable privacy risk for BCI users. Moreover, our preliminary

results about users' alertness and awareness indicate that even when/if a user does notice something out-of-the-ordinary within the game, (s)he does not seem to be alerted by an unexpected/unusual event.

Brain malware, and in particular subliminal brain malware, is a developing concern, given recent deployment of BCI technologies outside of research and medical purposes. We believe that the right time to address these issues is now, and propose that methods to prevent and mitigate BCI-enabled privacy threats should be developed in the early design phase, and embedded throughout the entire life of the technology. We view the development of these prevention and mitigation tools as an interdisciplinary effort, involving neuroscientists, neural engineers, ethicists, as well as legal, privacy and security experts.

### *3.8 Acknowledgment*

This work is supported by Award Number EEC-1028725 from the National Science Foundation, and by the University of Washington Tech Policy Lab. The content of this Chapter is solely responsibility of the authors, and does not necessarily represent the official views of the National Science Foundation.

We thank Jeffrey Herron, Lise Johnson, Tyler Libey, Charles Matlack, Brian Mogen, Patrick Moore, Hannah Werbel and Professor M. Ryan Calo, Howard J. Chizeck and Eric Chudler for suggestions, and ideas, as well as for all the help with the acquisition of the BCI device used in the experiments (Dr. Lise Johnson and Dr. Eric Chudler), for the help with experimental design and IRB approval (Jeffrey Herron and Charles Matlack), with experimental setup (Jeffrey Herron, Brian Mogen and Tyler Libey), with the development of tools for experimental analysis (Jeffrey Herron and Hannah Werbel), as well as for the help with trial runs (Jeffrey Herron and Patrick Moore).

# Chapter 4

# INFORMATION-THEORETIC ANALYSIS OF BRAIN MALWARE

In this Chapter, we cast the brain malware problem into a well-developed information-theoretic framework, which allows us to ask the following four questions:

1. How do we objectively measure an attacker's efficiency in achieving his/her goals?
2. How do we measure the efficiency of any deployed prevention/mitigation mechanism against brain malware?
3. Does there exist an optimal sequence of stimuli an attacker should present to a user in order to minimize the time and the effort needed to infer information of interest about the user?
4. Does there exist an adaptive (online) way of pruning out a subset of stimuli based on the observed user's responses?

We seek to answer the first two questions by casting the brain malware problem into the communication framework, which further allows us to analyze the interaction between an attacker and a potential defense mechanism using game theoretic tools.

## *4.1 Casting Brain Malware Problem into the Communications Framework*

We address the question of objectively measuring an attacker's efficiency, and an efficiency of a possible prevention/mitigation system by casting the brain malware problem for non-invasive BCIs into a well-developed information-theoretic framework. In order to do so, we first review how a BCI, operating in a benign environment, can be modeled as communication channel. A summary of notation used in this Chapter is given in Table 4.1.

In [180], the authors modeled non-invasive EEG-based BCIs as a communication channel, based on assumptions that:

(i) The main purpose of BCIs is communication with the environment, and

(ii) The environment will wait until a user's intention is clear before taking any action.

**Table 4.1:** Summary of notation from Chapter 4.

| Symbol | Definition |
|---|---|
| $\mathcal{W}_i$ | Finite alphabets of source messages $W_1, W_2$ |
| $W_1$ | Source messsage, modeling a user's intent |
| $W_2$ | Source messsage, modeling a user's private information |
| $\mathcal{X}_i$ | Finite alphabets of input messages $X_1, X_2$ |
| $X_1$ | User's intent encoded in neural activty |
| $X_2$ | User's private information, encoded in neural activty |
| $\mathcal{Y}$ | Finite alphabets of output message $Y$ |
| $Y$ | User's intent and private information, as contained in a user's recorded neural signals |
| $\mathcal{Z}_i$ | Finite alphabets of feedback messages $Z_1, Z_2$ |
| $Z_1$ | Graphical display feedback to a user |
| $Z_2$ | Feedback message about user's private information |
| $\hat{W}_1$ | Decoded user's intent |
| $\hat{W}_2$ | Decoded user's private information |
| $\mathcal{P}_e$ | Error probability |
| $\mathcal{P}_{success}$ | Probability of a successful brain malware attack |
| $\mathcal{P}_{chance}$ | Probability of successfully decoding a user's intent by chance |
| $\Delta$ | Equivovation rate, a measure of an attacker's inability to correctly decode a user's private information |

### *4.1.1 Intent as a Discrete Random Process*

The authors of [180] assumed that a user's intent can be described as a sequence $W = (W_1, W_2, \dots)$, where each $W_i$ is an element of a finite alphabet, $\mathcal{W} = \{w_1, \dots, w_m\}$. To capture the fact that a user's intent may (and will) differ, depending on a situation, they assumed that an intent is generated by a random process $\mathbb{P}_W(w)$. Thus, for any $n$, the conditional probability of $n$-length intent can be found as [180]:

$$\mathbb{P}_{W^n}(w^n) := \prod_{i=1}^{n} \mathbb{P}_{W^i|W^{i-1}}(w_i|w^{i-1}) \tag{4.1}$$

### *4.1.2 Coding/Decoding Process as a Discrete Memoryless Channel*

In BCIs, classifiers are usually used to determine the intended messages based on recorded neural signals (for example, the intended left- or right-hand motor imagery, or the spelled letter), with some probability of making errors and incorrectly decoding messages. This allows us to think of coding/decoding process as a *discrete memoryless channel*.

For example, the authors of [180] modeled motor-imagery-based BCIs as *binary symmetric channels (BSC)* [66]. The $k$-th input to the channel is a random variable $X_k \in \{0, 1\}$, representing a user's motor imagery, where $x_k = 0$ corresponds to the left hand movement and $x_k = 1$ to the right. The $k$-th output of the channel is a random variable $Y_k \in \{0, 1\}$, representing a user's decoded intent. In this channel, $x_k = y_k$ represents a correct decoding (inference), and $x_k \neq y_k$ a case when a decoding error occurs. The probability of an error, $\mathcal{P}_e$, can be computed as [180]:

$$\begin{aligned} \mathcal{P}_e &:= \mathbb{P}_{Y_k|Y^{k-1},X^k}(y_k|y^{k-1}, x^k) = \mathbb{P}_{Y^k|X^k}(y_k|x_k) \\ &= \begin{cases} 1-\epsilon, & \text{if } y_k = x_k, \\ \epsilon, & \text{else} \end{cases} \end{aligned} \tag{4.2}$$

where $\epsilon$ denotes a parameter which can be learned (inferred) from BCI training data.

### *4.1.3 Graphical Display as Noiseless Feedback*

A vast majority of non-invasive BCIs operate by providing visual feedback through a graphical display. Such a feedback contains information about the previous channel outputs,

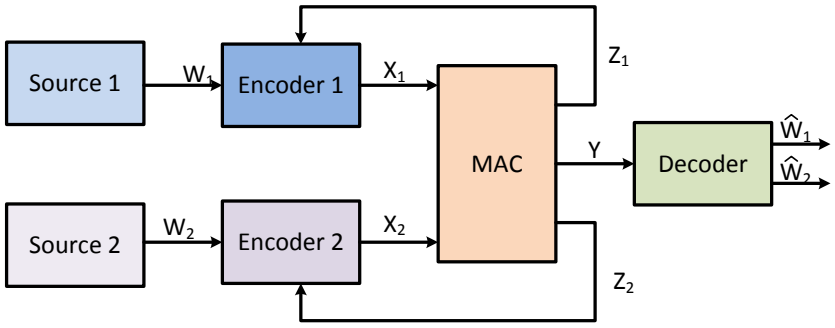



**Figure 4.1:** A block diagram of a two-user Multiple Access Channel with Generalized Feedback (MAC-GF). Signals $W_1$ and $W_2$ represent source messages to be transmitted to the receiver. $X_1$ and $X_2$ are input messages, processed by encoders, and $Y$ is an output message. Signals $Z_1$ and $Z_2$ represent feedback messages to encoders and $\hat{W}_1, \hat{W}_2$ the decoded messages on the receiver side.

$y_1, y_2, \ldots, y_k$, which is then used to choose the next channel input, $x_{k+1}$.

By assuming that a graphical display conveys information about the decoded output without noise and errors, the authors of [180] modeled it as a causal and noiseless feedback channel. Thus, combining ideas described above, an EEG-based BCI can be modeled as a *binary symmetric channel with noiseless feedback* [66].

## *4.2 Preliminary: Multiple Access Channel with Generalized Feedback*

A Multiple Access Channel with Generalized Feedback (MAC-GF) is a communication channel where two or more sources transmit information to a single destination, and each source observes a different feedback message. More details about this type of a communication channel can be found in seminal papers by Carleial [56], Zhen et al. [242], and Ozarow [181], but a simple block diagram of a two-user MAC-GF is depicted in Figure 4.1. We can mathematically represent such a discrete memoryless MAC-GFs as:

$$(\mathcal{X}_1 \times \mathcal{X}_2, \mathbb{P}(y, z_1, z_2|x_1, x_2), \mathcal{Y} \times \mathcal{Z}_1 \times \mathcal{Z}_2) \tag{4.3}$$

where $\mathcal{X}_1$ and $\mathcal{X}_2$ represent finite input alphabets, $\mathcal{Y}$ a finite output alphabet, and $\mathcal{Z}_1$ and $\mathcal{Z}_2$ finite feedback alphabets.

Value $\mathbb{P}(y, z_1, z_2|x_1, x_2)$ represents the conditional probability that output and feedback messages are equal to $(y, z_1, z_2) \in \{\mathcal{Y} \times \mathcal{Z}_1 \times \mathcal{Z}_2\}$ given input messages $(x_1, x_2) \in \{\mathcal{X}_2, \mathcal{X}_2\}$,

and it can be computed as [56]:

$$\begin{aligned} & \mathbb{P}(y^n, z_1^n, z_2^n | x_1^n, x_2^n) \\ = & \prod_{i=1}^{n} \mathbb{P}(y_i, z_{1i}, z_{2i} | x_{1i}, x_{2i}, y^{i-1}, z_1^{i-1}, z_2^{i-1}) \\ = & \prod_{i=1}^{n} \mathbb{P}(y_n, z_{1i}, z_{2i} | x_{1i}, x_{2i}) \end{aligned} \tag{4.4}$$

A typical code for discrete memoryless MAC-GFs consists of:

- Two independent message sources, generating random messages $W_1 \in \{1, 2, \ldots, M_1\}$ and $W_2 \in \{1, 2, \ldots, M_2\}$.
- Two encoding functions:

$$\begin{aligned} x_{1n} &= f_{1n}(W_1, Z_1^{n-1}), \\ x_{2n} &= f_{2n}(W_2, Z_2^{n-1}), n = 1, 2, \ldots, N \end{aligned} \tag{4.5}$$

  where $x_{1n}, x_{2n}$ denote codewords of input messages $X_1, X_2$; $W_1, W_2$ source messages; $Z_1^{n-1}, Z_2^{n-1}$ feedback messages, and $N$ the length of input messages $X_1, X_2$.

- One decoding function:

$$(\hat{W}_1, \hat{W}_2) = g(Y^N) \tag{4.6}$$

  where $\hat{W}_1, \hat{W}_2$ denote decoded source messages and $Y^N$ output message.

For independent and uniformly distributed source messages $W_1, W_2$, error probability, $\mathcal{P}_e$ can be defined as [242]:

$$\mathcal{P}_e := \mathbb{P}[(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)] \tag{4.7}$$

The capacity region of a general MAC-GF is not known. However, several coding schemes have been proposed in the literature (e.g., [56, 242]), and achievable regions for those schemes were derived.

### *4.3 Modeling Brain Malware as MAC-GF*

We start by modeling a user's intent as a discrete random sequence $W_1$ over finite alphabet $\mathcal{W}_1$. Similarly, we model one specific instance of a user's private information (for example, a

user's credit card PIN) as a discrete random sequence $W_2$ over finite alphabet $\mathcal{W}_2$. Building on section 4.1, we assume both intent $W_1$ and instance of a user's private information $W_2$ are generated by discrete random processes, as described by equation (4.1).

We next assume that intent and private information are both encoded into a user's neural activity, which can be modeled as input messages $X_1^N$ and $X_2^N$, consisting of $N$ codewords $x_{1i}$ and $x_{2i}$, $i = 1, \ldots, N$:

$$\begin{aligned} x_{1n} &= f_{1n}(W_1, Z_1^{n-1}) \\ x_{2n} &= f_{2n}(W_2, Z_2^{n-1}), \ n = 1, 2, \ldots, N \end{aligned} \tag{4.8}$$

Feedback message $Z_1$ represents feedback provided to a user about his/her intent. We distinguish three cases:

- We can have an open-loop BCI, where no feedback about the decoded intent is provided. This case can be modeled as a special case of a MAC-GF, namely the discrete memoryless MAC with one-sided different generalized feedback [242].
- Next, we can have a closed-loop BCI, where feedback about the decoded intent is assumed to be perfect (causal and noiseless).
- Finally, we can also have a closed-loop BCI where feedback about user's intent is affected by noise.

In all of the cases, feedback message $Z_2$ represents an attacker abusing the system, to represent different stimuli to users and to make inferences about their private information.

In a BCI, a user's neural activity is recorded using an electrode array. Therefore, in a signal acquisition component of a BCI, both intent and private information are observed as single output signal $Y$. In a signal processing component of an attacked BCI, signal $Y$ is translated into a decoded intent, $\hat{W}_1$ and a decoded private information, $\hat{W}_2$.

**Note 1:** In case no attack is mounted against a BCI, the described model easily reduces to a discrete memoryless channel, described in section 4.1.

**Example 1:** Let's consider a user, trying to spell a word `blue`. Let's assume his credit card PIN is `1234`. In this case, values of random variables $W_1$ and $W_2$, representing source messages are:

$$W_1 = \texttt{blue} \text{ and } W_2 = \texttt{1234}$$

These source messages get encoded into inputs $X_1$ and $X_2$:

$$X_1 = \{x_{11}, x_{12}, x_{13}, x_{14}\} \text{ and } X_2 = \{x_{21}, x_{22}, x_{23}, x_{24}\}$$

In an EEG-based BCI, these input messages are recorded from a user's skull as an EEG signal $Y = \{y_1, y_2, \ldots, y_8\}$, sampled with some sampling rate $f_s$ and discretized with some quantization resolution, $r_{A/D}$.

Using feature extractor and decoding algorithms, an attacker can decode output signal $Y$ into a user's intent and an instance of private information:

$$\begin{aligned} \hat{W}_1 &= \text{anything} \\ \hat{W}_2 &= \texttt{1234} \text{ or } \hat{W}_2 = \texttt{9876} \end{aligned}$$

### *4.3.1 Modeling Attacker's Goals*

In brain malware, an attacker's goal is to correctly infer a user's private information. Thus, an attack is considered *successful* if an attacker can correctly decode a user's private information:

$$\mathbb{P}[\text{Attack successful}] := \mathbb{P}[\hat{W}_2 = W_2] \tag{4.9}$$

While an attacker, in general, does not care about correctly decoding users' intents, he might be willing to hide presence within a BCI, in order to avoid alarming users, and to keep them engaged with a malicious application for as long as possible. To maintain this *stealthiness*, an attacker will want to correctly decode users' intents at least as successfully as a benign application would. Keeping this observation in mind, we redefine a *successful attack* as follows.

**Definition 1** *A brain malware attack is considered successful if an attacker can correctly decode a user's private information, while at the same time hiding his/her presence within the system. An attacker is considered hidden if (s)he can decode a user's intents equally as good as a benign BCI application would. Thus, success probability of a brain malware attack*

*is equal to:*

$$\begin{aligned} \mathcal{P}_{success} &:= \mathbb{P}[\hat{W}_1 = W_1, \hat{W}_2 = W_2] \\ &= \mathbb{P}[(\hat{W}_1, \hat{W}_2) = (W_1, W_2)] \\ &= 1 - \mathcal{P}_e \end{aligned} \tag{4.10}$$

An attacker's goal can now be formalized as an optimization problem:

$$\begin{aligned} &\text{maximize } \mathcal{P}_{\text{success}} \\ &\text{subject to } \mathbb{P}[\hat{W}_1 \neq W_1] > \mathcal{P}_{\text{chance}} \\ &\qquad \text{Attacker's resources} \end{aligned} \tag{4.11}$$

where $\mathbb{P}[\hat{W}_1 \neq W_1] > \mathcal{P}_{\text{chance}}$ represents a benign application's constraint on successfully decoding a user's intent, and *attacker's resources* denote all other constraints imposed on the attacker, such as computational power and energy. Value $\mathcal{P}_{\text{chance}}$ denotes the value of successfully decoding the intended message by chance. While this value depends on the source and input alphabets, as well as assumptions made about a communication channel, it represents a lower bound requirement on a decoding algorithm used by a BCI. Using equation (4.10), an attacker's goal (4.11) can be rewritten as:

$$\begin{aligned} &\text{minimize } \mathcal{P}_e \\ &\text{subject to } \mathbb{P}[\hat{W}_1 \neq W_1] > \mathcal{P}_{\text{chance}} \\ &\qquad \text{Attacker's resources} \end{aligned} \tag{4.12}$$

### *4.3.2 Modeling Interaction Between An Attacker and a Privacy-preserving BCI*

The main goal of a hypothetical BCI privacy-enhancing mechanism would be to prevent an attacker from gaining access to a user's private data. Thus, a brain malware attack on a user of a BCI with a privacy-enhancing mechanism can be modeled as an *information hiding game* between two non-cooperative players, a privacy-enhancing mechanism and an attacker. The information hiding problem has been extensively studied in recent years. For more details, see for example [190, 174].

In the BCI information hiding game, the first player tries to maximize a payoff function, while the opponent (an attacker) tries to minimize it. One possible choice for a payoff function is the *equivocation rate* [235], which we redefine as follows.

**Definition 2** *The equivocation rate,* $\Delta$*, is a measure of the degree to which an attacker is unable to correctly decode a user's private information. It is equal to the conditional entropy of a private information* $W_2$*, given output message* $Y$*:*

$$\begin{aligned} \Delta &:= H(W_2|Y) = \sum_{y\in\mathcal{Y}} \mathbb{P}[y] H(W_2|Y=y) \\ &= -\sum_{y\in\mathcal{Y}} \sum_{w_2\in\mathcal{W}_2} \mathbb{P}[w_2, y] \log\left(\mathbb{P}[w_2|y]\right) \end{aligned} \tag{4.13}$$

The information-hiding game can now be formalized as the following optimization problems:

$$\begin{aligned} \textit{BCI Anonyimizer:} \quad & \text{maximize } \Delta \\ & \text{subject to system resources} \\ \text{Attacker:} \quad & \text{minimize } \Delta \\ & \text{subject to attacker's resources} \end{aligned} \tag{4.14}$$

### *4.4 Summary*

In this Chapter, we cast the brain malware problem against EEG-based BCIs into an information theoretic framework, in order to evaluate an attacker's private information efficiency, and a prevention/mitigation mechanism's efficiency in preventing brain malware. In doing so, we model brain malware attacks as a Multiple Access communication Channel with Generalized Feedback (MAC-GF). We further introduce the equivocation rate, defined as a measure of the degree to which an attacker is unable to correctly decode a user's private information as a metric to evaluate interaction between a an attacker and a prevention/mitigation mechanism.

### *4.5 Acknowledgment*

This work is supported by Award Number EEC-1028725 from the National Science Foundation, and by the University of Washington Tech Policy Lab. The content of this Chapter

is solely responsibility of the authors, and does not necessarily represent the official views of the National Science Foundation.

We thank Jeffrey Herron, Charles Matlack, and Fredrik Ryden and Professors Howard J. Chizeck and Jeff A. Bilmes for helpful suggestions, ideas, and inspiration that have helped in writing of this chapter.

# Chapter 5

# BCI ANONYMIZER, AN ENGINEERING APPROACH AGAINST BRAIN MALWARE

## *5.1 Similarity Between BCI and Smartphone Industry*

We start this chapter by revisiting the block diagram of a typical BCI, depicted in Figure 2.1. If such a system is compromised, then an attacker, who controls an application and a signal processing component of the system, may be able to: (i) receive the whole recorded neural data and (ii) run a classification/decoding algorithm on it, in order to extract a user's private data.

Let's now compare the case of a compromised BCI to the case of a compromised smartphone, where an attacker tries to gain access to a user's *private identifiable information* (PII), defined as any information that can be used on its own or with other information to identify, contact, or locate a single person, or to identify an individual in context [98].

In the smartphone industry, such attacks on users' privacy are typically prevented by limiting access to phones' operating systems and users' PII. In other words, an application has an access only to a limited subset of PIIs, operating system states and functionalities (for examples of current prevention and mitigation strategies in smartphone industry, please see e.g., [62, 244, 79, 152]).

## *5.2 BCI Anonymizer: The Main Idea*

We observe that BCI-recorded neural signals have the same role as smartphone PII data. Thus, an access to users' neural signals, recorded using BCIs, should be limited. We therefore propose the following approach for BCI-enabled technologies: **in order to protect a user's neural data, a new component, referred to as a *BCI Anonymizer*, should be added to a BCI system. The BCI Anonymizer should be a secured and trusted component that takes raw neural signals and decomposes them to specific com-**

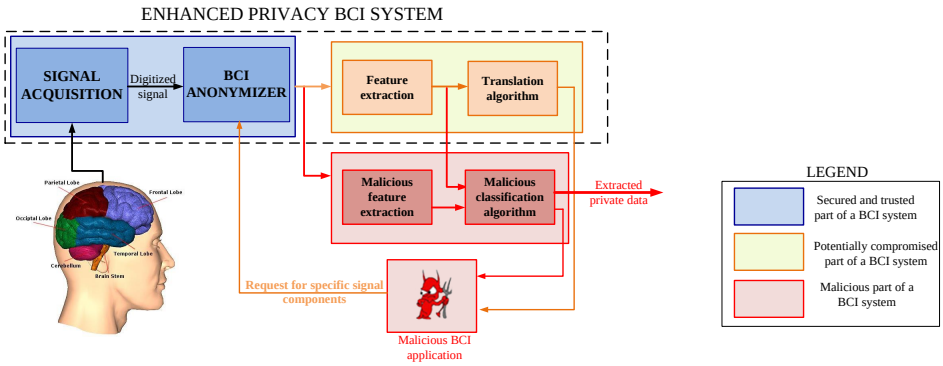



**Figure 5.1:** A high level block diagram of a BCI with the BCI Anonymizer. An attacker may access and control feature extraction and translation algorithms (denoted as orange blocks in the diagram), or may add an additional set of algorithms for private data extraction (denoted red in the diagram), but an attacker cannot access states and functionality of signal acquisition and BCI Anonymizer components (denoted blue in the diagram).

**ponents. Upon request, instead of the complete recorded neural signal, the BCI Anonymizer provides a BCI application with a subset of requested signal components.**

A block diagram of a BCI system with the proposed BCI Anonymizer component is shown in Figure 5.1. This approach, and the BCI Anonymizer rely on several important assumptions. Those are:

(1) The BCI applications will only ask for those components of the recorded neural signal that they require in order to function properly.

(2) The BCI applications will not be able to misuse and manipulate those components they have been given access to for nefarious purposes.

(3) The BCI Anonymizer is a secured and trusted component that decomposes the whole recorded neural signal, and never gives access to it to any BCI application.

(4) The BCI Anonymizer will be able to decompose the recorded signal in a reliable and fast enough fashion.

These assumptions raise several critical engineering, scientific and industrial challenges that will have to be addressed before any further development and wider adaptation of the proposed BCI Anonymizer. In the rest of this chapter, we tackle upon the following three of these challenges:

(a) Why could (why does) the proposed recorded signal decomposition approach work, and why should decomposing the signal be enough to prevent privacy attacks?

(b) Who, when and how should be in charge of vetting BCI applications to make sure they are only requesting necessary and benign components of recorded neural signals?

(c) Why real time signal decomposition?

### *5.3 Why Does the Proposed Signal Decomposition Approach Work?*

Most of the existing, as well as those BCI applications emerging in near future, can broadly be divided into *medical* and *non-medical*. Medical BCI applications are typically grouped into the following main categories [147]:

(M1) Communication,

(M2) Environment control,

(M3) Locomotion, and

(M4) Neurorehabilitation,

and non-medical application into the following seven categories [223]:

(NM1) Device control,

(NM2) User state monitoring,

(NM3) Evaluation (neuroergonomics and neuromarketing),

(NM4) Training and education,

(NM5) Gaming and entertainment,

(NM6) Cognitive improvements, and

(NM7) Safety and security.

Most EEG-based BCI applications, both medical and non-medical, rely on one of the following EEG components/bands as a feature of interest:

(C1) **Slow cortical potentials (SCPs)**, defined as slow voltage changes in the cortex area (among the lowest frequency features of EEG). Typically, negative SCPs shifts represent cortical activation associated with movement, while positive SCP shifts accompany reduced cortical activation [147].

(C2) **Sensorimotor rhythms (SMRs)**, defined as the changes in the $\mu$ rhythm (8-12 Hz) or the $\beta$ rhythm (18-26 Hz), typically recorded over sensorimotor cortex. Changes in $\mu$ and $\beta$ rhythms are often referred to as Event-Related Desynchronization (ERD) and Event-Related Synchronization (ERS), and they are typically associated with movement, sensation and motor imagery [147].

(C3) **Steady State Visual Evoked Potential (SSVEPs)**, defined as potentials generated by exciting the retina using visual stimuli modulated at certain frequencies [175].

(C4) **Event Related Potential (ERPs)**, defined as responses to an infrequent or a particularly meaningful stimulus within a stream of frequent standard stimuli [147].

Based on the used component, BCI control and communication strategy differ very significantly, in terms of:

(D1) The expected amount of training time required in order to successfully use an application,

(D2) The expected amount of user effort, and

(D3) The expected amount of time a user will be able to hold attention on an application.

Thus, simply by labeling an application as one using component $C_i$, certain expectation on the way it should operate are already set. Moreover, neither of the listed components does not seem suitable for"dual use", i.e., a scenario where a component is at the same time used for a valid purpose, and for a privacy attack. For example, a malicious P300 speller does not seem feasible, since it is unlikely that a user will at the same be able to respond to flashed letters and to a potential malicious stimulus. Thus, in summary, there are three main reasons making the proposed signal decomposition feasible:

(R1) Most BCI applications use *only* one among possible SCPs, SMRs, SSVEPs and ERPs EEG components.

(R2) By labeling an application as one using a specific EEG components, certain expectation about it are already set, and

(R3) None of the most often used EEG components allows for a dual use at the same time.

## *5.4 Why Real Time Decomposition?*

Upon closer consideration, one can observe that both Martinovic et al. [153] and we, in our experimental analysis, have actually used a *batch* processing method to extract private information about a user. In both sets of experiments, a user was presented with a random sequence of stimuli, his/her responses to the whole sequence of presented stimuli were recorded, and then the recorded signals were processed and used to infer information about a user. The same is true in many other ERP-based BCI applications, such as P300 Speller [83].

It therefore seems feasible that the same approach, to first record neural signals and batch-decompose them into components before giving them to a BCI application, may also be used to mitigate some of the privacy attacks on BCI-enabled communication. The batch approach does not, however, protect the stored neural data. The real-time approach, on the other hand, mitigates privacy attacks that might occur during:

(i) Storage,

(ii) Transmission, or

(iii) Data manipulation by a BCI system,

since the complete signal is never stored or transmitted. Thus, while more challenging, the real-time signal decomposition is a critical step in ensuring privacy and security of BCI-enabled technologies.

### *5.4.1 Possible Signal Processing Approaches*

In developing the appropriate real-time signal decomposition algorithm for the BCI Anonymizer, we propose that two families of signal processing algorithms might be useful: (a) the wavelet transforms and (b) Empirical Mode Decomposition.

#### *Wavelet Transform*

The wavelet transform [68, 52] has been found suitable in a wide range of bioengineering applications, including capturing and localization of transient features of an EEG signal [29].

The wavelet transform uses a set of *basis functions*, based on a *mother wavelet*, to decompose the original signal into shifted and scaled versions of the mother wavelet. A mother wavelet is a waveform of an effectively limited duration and has an average value of zero. A convenient feature of wavelet transform is the fact that a set of so-called "complementary" wavelets can decompose the original signal without gaps or overlaps, so that the applied decomposition is mathematically reversible. On the other hand, as pointed out by [160], the wavelet transform, similar to the Fourier transform, has problems with uncertainty principle, causing issues with time/frequency resolution tradeoff. For these reasons, it may not be suitable for all BCI applications.

*Empirical Mode Decomposition*

The Empirical Mode Decomposition (EMD) [110] is a signal decomposition method well suited for nonlinear and non-stationary data signal processing. This method is a data driven method, which makes it adaptive and quite flexible in nature. It decomposes a time series signal into multiple zero-mean amplitude modulated/frequency modulated oscillatory function, known as *"intrinsic mode functions" (IMFs)*, that take on a similar role as basis functions in the wavelet transform.

One concern of EMD use is a lack of formal guarantees of convergence in the process, due the *ad hoc* nature of the approach. This concern was resolved by a modification of the way that the underlying support functions are obtained [106]. EMD methods have recently been used in a variety of biological and medical applications, for example, for EEG signal preprocessing [70, 184], separation and removal of ocular artifacts from the EEG signal [168], and epileptic seizure detection [217].

Recently Sweeney et al. [214] proposed an extension of the original EMD method, the Ensemble EMD with Canonical Correlation Analysis (CCA), and applied it to efficiently remove movement artifact from EEG signals.

## *5.5 How to Go about BCI Applications Vetting?*

In order to make sure that BCI applications are only requesting those components of the recorded neural signal that they need to operate, and that those signal components are

being used for valid (benign) purposes, BCI industry will have to set up a mechanism to analyze and validate available applications. There are two main vetting approaches that can be taken:

(A1) Centralized approach, and

(A2) Distributed approach.

In a centralized approach, there exist a centralized entity (a private company, a federal agency, or an industrial society) that scrutinizes every application, and only allows those applications deemed as appropriate to become a part of the app store. On the other hand, in a distributed approach, every application can freely be added to the app store, where anyone can freely use it, analyze it, and report findings about it for everyone else to see. Both approaches have been taken in the smartphone industry, and both approaches have experienced some critiques, from being deliberately too slow to vet a valid and benign application, in order to get an upper hand over competition, to being too slow to remove malicious content from the app store, once it has been legitimately reported as malicious.

One potential approach for the BCI industry might be a combination of a centralized and a distributed approach, the so-called *hybrid approach*, where multiple BCI manufacturers, federal agencies, user advocates, as well as other interested parties would participate in the creation of application privacy and security requirements and vetting rules. Once requirements and rules are created, however, the vetting process itself should be made as fast, automatized and as transparent as possible.

### *5.6 Summary*

In this Chapter, we present the concept of the BCI Anonymizer, an engineering approach to enhance privacy of BCI users. The basic idea of the BCI Anonymizer is to pre-process neural signals, before they are stored and transmitted, in order to remove all information except specific intended BCI commands. Unintended information leakage is prevented by never transmitting and never storing raw neural signals and any signal components that are not explicitly needed for the purpose of BCI communication and control.

The BCI Anonymizer should be a part of a BCI, and never of any external network or computational platform. It thus acts as a secured and trusted and trusted subsystem that takes raw neural signals and decomposes them to specific components. Upon request, instead of the complete recorded neural signal, the BCI Anonymizer provides a BCI application only with a needed subset of requested signal components.

### *5.7 Acknowledgment*

This work is supported by Award Number EEC-1028725 from the National Science Foundation, and by the University of Washington Tech Policy Lab. The content of this Chapter is solely responsibility of the authors, and does not necessarily represent the official views of the National Science Foundation.

We thank Jeffrey Herron, Charles Matlack and Professor Howard J. Chizeck for suggestions, ideas, and help with the concept of the BCI Anonymizer.

Chapter 6

# SOCIETAL IMPACT OF BRAIN MALWARE AND BCI ANONYMIZER

Drawing from an observation that a wide range of mobile information technologies already exist, and that BCI technology can be considered the extreme example of cybernetic integration, we begin this Chapter by briefly assessing the state of practices and policies, currently governing the relationship between information, communication technologies, and society.

In doing so, we turn to the smartphone industry once again, and draw another link between smartphones and BCI technologies. It is now generally a standard for smartphone operating systems to provide users with granular control of access permissions to their private identifiable information. This is typically done by giving users options to grant or deny access to specific, atomic resources on the device (e.g. data storage or sensor stream), and relying on users to make choices informed by their knowledge of the information content of those resources.

Unfortunately, this approach has had its own set of drawbacks. To begin with, in many instances, the operating system does not exercise sufficient control over all communication channels to enforce security policies. A recent example of this is third-party embedded advertising software broadcasting users' PIIs over insecure communication channels [25]. Second, some implementations do not give users complete control of accessed resources at the desired level of granularity. That is, requests for resource access are bundled so that a user may have to choose between compromising a resource exposing private information, and not using an application at all. For example, the use of location services on Android phones was offered only on condition of allowing continuous location tracking [23]. Finally, the most challenging problems may be that users are expected to fully understand what information is encoded in the data resources they grant or deny access to, and they cannot

filter/hide private information in a resource while simultaneously making it available to an application.

Our belief is that for brain-computer technologies, given their early development and deployment stage, a combination of technological approaches, tech policies and standards should be capable of resolving and avoiding many of the insufficiencies of the existing security and privacy practices for more ubiquitous consumer products. In order to develop such approaches and policies, however, in this Chapter, we focus on a few fundamental questions that will have to be resolved before BCIs come to more a widespread use:

1. How do we provide assurance to BCI users that the applications they are using are not trying to manipulate the system, in order to access their sensitive and private information?
2. How do we grant users a sufficient level of access control over their resources?
3. How do we relieve users from an unnecessary burden of having to understand what information may be encoded in data resources they are granting or denying access to?
4. How do we incentivize BCI manufacturers to incorporate a privacy-enhancing component into their systems?

In order to tackle these questions, we turn to Value Sensitive Design, a systematic approach to accounting for human values in technology design [89]. More specifically, we draw inspiration for this Chapter from the recent VSD project, proposing the development of a *privacy addendum* for open-source software licenses [91].

### *6.1 Value Sensitive Design*

Value Sensitive Design (VSD) is a theoretically grounded approach that allows for a principled and comprehensive way of accounting for human values throughout the design of technology. It leverages a unifying and iterative tripartite methodology, consisting of *conceptual, empirical, and technical investigations*. *Conceptual investigation* focuses on key stakeholders and values that they hold important. *Empirical investigation* engages with actual or potential stakeholders, with the goal of evaluating potential success or failure of the

investigated technology. In doing so, empirical investigation leverages a wide range of quantitative and qualitative methods used in social sciences, including observations, interviews, surveys, experimental manipulations, and document collection, as well as measurements of users' behavior and physiology. Finally, *technical investigation* focuses on features, architecture, and infrastructure of the analyzed technology [89].

VSD has been applied to a large variety of technology design challenges, including human-computer interaction [90], human-robotic interactions [122], and computer-supported cooperative work [164]. In this Chapter, we focus only on the conceptual investigation, and more specifically on *direct and indirect stakeholders* investigation.

*Direct and Indirect Stakeholders:* Earlier VSD work [90] has shown the need to consider two groups of people that may potentially be affected by a technology, referred to as direct and indirect stakeholders. Direct stakeholders are defined as those people who directly interact with technology, and indirect stakeholders as the people whose data or presence may be implicated by the technology, even though they are not directly interacting with it.

### *6.1.1 Value Sensitive Design and BCI Technology*

As a part of the conceptual investigation of the existing and emerging BCIs, we identify ten distinct groups of stakeholders, seven of which we consider to be direct stakeholders:

(D1) BCI manufacturers

(D2) BCI application developers,

(D3) People using BCIs for augmentation,

(D4) People using BCIs for medical reasons,

(D5) Medical practitioners,

(D6) Patients' caregivers, and

(D7) BCI attackers.

Additional groups of stakeholders are (I1) privacy merchants, (I2) marketers and advertisers, and (I3) family members not directly involved in patients' care, and these groups we consider to be indirect stakeholders.

Disregarding the values considered to be important by BCI attackers, we identify the following values to be important to one or more groups of stakeholders:

(V1) Reliability,

(V2) Ease of use,

(V3) Privacy,

(V4) Independence,

(V5) Autonomy, and

(V6) Ease of access to users' private information.

We recognized that all direct stakeholders (except for attackers) care about values reliability and privacy. Similarly, we anticipate that the majority of direct stakeholders, except maybe medical practitioners and patients' caregivers, will value ease of use. We further expect that people using BCIs for medical reasons will deeply value independence, while their caregivers are likely to care about autonomy. Marketers and advertisers, as well as privacy merchants, on the other hand, are expected to care most about the ease of access to users' private information.

In answering questions (1)–(4), however, our focus is on the following categories of direct stakeholders: (D1) BCI manufacturers, (D3) and (D4) people using BCIs for medical reasons or for augmentation, and (D7) BCI attackers, and we mostly focus on the value of privacy. Inspired by [91], we analyze these stakeholders and the value of privacy through two specific aspects: informed consent and threat analysis.

### *6.2 Informed Consent*

Informed consent is a construct (typically, either written or oral) that provides human users with a level of protection while participating in, or being involved in a specific event. Its intent is to provide users with sufficient knowledge about the event, and with a choice to decide about their participation [80]. As stated in [91], "...when implemented well, informed consent creates conditions by which end users are positioned to protect themselves and their privacy as they want through selective participation."

A model of informed consent typically consists of six components: *disclosure, comprehension, voluntariness, competence, agreement and minimal distraction*, defined as follows [91]:

- **Disclosure** - the act of providing appropriate and accurate information to the intended audience.
- **Comprehension** - the intended audience's ability to understand what has been disclosed.
- **Agreement** - the intended audience's ability to agree or decline to participate in the event, once informed about it.
- **Competence** - the intended audience's mental, emotional and physical capability to give informed consent.
- **Voluntariness** - the intended audience's ability to give informed consent without coercion or undue manipulation.
- **Minimal Distraction** - the requirement to have a streamlined "informing" and "consenting" process, that do not unduly distract intended audience from their goals.

### *6.2.1 Informed Consent and BCI Technologies*

Similar to [91], to examine those behaviors we wish to require or recommend for BCI manufacturers and application developers, we examine the current and the emerging BCI technologies through the lenses of a model of informed consent. Going component by component, we observe the following:

**Disclosure:** We anticipate that BCI users (both people using BCIs for augmentation and as a medical device) will want to know the answers to the same questions listed in [91]:

(Q1) What information does the BCI collect?

(Q2) Who all will have access to the collected information?

(Q3) How long will the information be retained?

(Q4) Which purposes will collected information be used for?

(Q5) Will a user's identity be protected and how?

**Comprehension:** While it is indeed impossible to guarantee that a BCI user has a sufficient level of understanding of the presented information, it it important to require from BCI application developers to disclose all information about all signals and their components that will be recorded and used. This disclosure should be made at a specific level of reading and comprehension (for example, at 12-grade reading level), and as potentially restrictive as this requirement may initially seem, it is rather necessary in order to avoid overloading users with complex engineering and neuroscientific terms and expression, which would effectively cancel out the purpose of informed consent.

**Agreement:** Devising an appropriate agreement for BCI technologies, we encounter problems already listed in [91], namely the problem of visible ongoing opportunity to revoke the consent at any point in time, and the problem of requesting the withdrawal of all data collected earlier.

**Competence and Voluntariness:** The questions of competence and voluntariness become especially interesting for BCI applications, especially for people using BCIs for medical reasons. In many instances, this group of stakeholders may represent a vulnerable group of users, and a special attention should be devoted to them, to make sure no unreasonable promises are being made to them, in return for their private and sensitive data. In addition, information that this group of users is even using the device may be sensitive in and of its own, thus requiring the information about the type of a BCI user to be protected as well.

Finally, some types of BCIs devices may not be intended for chronic use, and potential removal of the device after a certain period of time may result in a negative setback for users, especially those using the device for medical reasons. That information should be conveyed to all users before they agree/disagree to the use of a device. More importantly, this information may have an effect on the ongoing nature of an informed consent, raising the question of when should a user be considered competent to make a decision about his/her BCI device.

**Minimal Distraction:** In order to implement this component of the informed consent model, we propose to follow the same practice as outlined in [91].

### *6.3 Threat Analysis for BCI Technologies*

The model of informed consent protects BCI users' privacy by giving users control over their participation and interaction with a system. In order to gain a full understanding of potential impact BCIs may have on users' privacy, however, analyzing and understanding potential threats that may arise from the use of these devices is equally important.

Drawing again from [91], we divide possible threats into those that arise as **side-product of a technical design**, and those that arise as a **consequence of an intentional and malicious action**. Combining results from Chapters 2–3 with findings in [91], we broadly group possible threats against BCIs into the following categories:

**Threat 1:** Disclosure to Unauthorized Parties,

**Threat 2:** Unauthorized Use of Individual Data,

**Threat 3:** Unauthorized Request (Search) for Individual Data,

**Threat 4:** Unauthorized Use of Aggregated Data,

**Threat 5:** Unauthorized Fusion of BCI Data with Unexpected External Data

**Threat 1: Disclosure to Unauthorized Parties:** As stated in [91], an unauthorized disclosure of information can affect user's reputation, as well as their right "to be left alone". That is particularly so with BCI users, where neither users using the system for a medical reason, nor those using it for augmentation may be willing to share with other people the information that they are doing that. The users may feel protective about the fact that they are using the system, as well as the purpose that they are using the it for.

**Threat 2: Unauthorized Use of Individual Information:** This class of threats addresses the problem of *secondary use of intended BCI information* [91]. For example, possessing information about the fact that a user is currently using a BCI device, and about his/her level of performance with a device may put a user in an unfavorable position.

**Threat 3: Unauthorized Request (Search) for Individual Data:** This threat is a novel threat, specific to BCI systems. It is introduced to account for threats such as brain malware, where a BCI application developer manipulates a user into thinking that application is being used for one purpose, when at the same time, (s)he introduces new

components into the application (such as subliminal stimuli) in order to maliciously extract private and sensitive information about a user.

**Threat 4: Unauthorized Use of Aggregated Data:** As stated in [91], this class of threats addresses those issues that arise from unauthorized use of BCI users' aggregated data. It is often times assumed that data aggregation and de-identification increases a user's level of privacy. That, however, is often not the case, and it especially may not be the case with users' neural signal, which have been shown to be unique to a user.

**Threat 5: Unauthorized Fusion of BCI Data with Unexpected External Data** This threat addresses an issue of fusion of a user's BCI data with other, potentially publicly available data, in order to make additional and advanced inferences about a user.

### *6.4 Privacy Agreement for BCI Technologies*

Combining results of informed consent and threat analyses for BCI technologies, we observe that *Privacy Addendum 1*, developed in [91] represents a meaningful first attempt of addressing questions (1)–(4) for BCIs. One possible way to make brain malware, and any similar attempts at unwillful extraction of private information harder to execute is to make it mandatory for BCI manufacturers to include the *Privacy addendum* every time they are directly interacting with BCI application developers, but also every time the device and all of its accompanying software and services are being leased or sold, possibly even to end users.

Based on the observations from informed consent and treat analyses of BCIs, modifications will, however, have to be made to the Addendum 2, originally referred to *Location Aware Privacy Principles*. We propose to use Addendum 2 as the basis for a novel Addendum 3, which we refer as the *BCI Privacy Principles*. The new Addendum accounts for the identified differences between BCIs and location-aware systems, namely the difference between competence and voluntariness components of the informed consent model, and the new threat class.

The updated text of the Addendum 1 from work [91], and the Addendum 3 are given below:

**Addendum 1: Privacy Addendum**

This addendum contains the additional terms applicable to development and distribution of a work (Work) containing all or a portion of the Program or that is otherwise derived from the Program.

I. You agree that:

(a) Your compliance with this Addendum is a material condition of your license to the Program.

(b) You will include in any follow-on licenses you make with other developers for building other applications or services for the Program and your Work the same terms and conditions as this license addendum provides (and you will bind any firm that acquires your firm to the same terms and conditions as this license provides).

(c) Your development, use, and/or distribution of a Work constitutes an enforceable public commitment to comply with the provisions of this addendum.

(d) Any collection, use, disclosure, and/or storage of private identifiable information about end-users will be undertaken in accordance with the **BCI Privacy Principles** (attached).

(e) End users are entitled to enforce the terms of this Addendum and the BCI Privacy Principles as third party beneficiaries of the Agreement or as otherwise permitted under applicable law.

(f) Any violation of the terms of this Addendum may constitute an unfair and/or deceptive trade practice in violation of state and federal consumer protection law.

II. You further agree that all distributed Works will: Clearly, conspicuously, and verifiability (a) warn end users that the Work may disclose their private and sensitive information to third parties; and (b) obligate you to comply with the BCI Privacy

Principles (set forth in Addendum 3, attached) by, without limitation, causing the following text (with sections in parentheses modified accordingly) to appear when the Work is first installed and at reasonable intervals thereafter: *When you use this application, (Software Name) (or other malicious software which takes advantage of (Software Name)) may cause your BCI device to communicate your private and sensitive information arising from your collected biosignals - to application providers and/or third parties online*

*Please be aware of your circumstances and your safety and use appropriate caution when using (Software Name). (Developer/Distributor) adheres to the BCI Privacy Principles (attached). These principles require us to get your prior informed consent for any collection, use, disclosure and/or storage of your private and sensitive information. Please review the (Software Name) privacy policy (include URL)*

III. You also agree that to the extent practicable, your implementation of the BCI Privacy Principles will be consistent with the best practices set forth from time to time at the Intel Research/University of Washington Privacy Best Practices website.

**Addendum 3: BCI Privacy Principles**

1. End users will be informed, in a manner reasonably designed to provide actual notice and prior to any collection, use, retention, or disclosure of personally identifiable information, of the following:

   (a) What personal information, biosignals, and their components will be collected;

   (b) How that personal information, biosignals and their components will be used;

   (c) To whom that personal information, biosignals and their components will be disclosed, how the recipient will be able to use that data, whether the recipient in turn will be able to transfer the data and whether the recipient is obligated to comply with these Principles.

(d) How long (or how often) the personal information, biosignals and their components will be disclosed (e.g. at the time of initial connection only, or periodically during the use of the software)

2. To the extent reasonably practicable under the circumstances, end users will be given conspicuous notice of the opportunity to prohibit the proposed collection, use, retention, and/or disclosure of their private identifiable information in whole or in part, and a reasonably simple mechanism for taking advantage of such opportunity. Where practicable, a Work will provide an easy method by which an end user can prohibit such collection, use, retention, and/or disclosure, on a case by case basis, at their discretion.

3. Personally identifiable information will be deleted regularly when it is no longer needed by the end-user or for the correct functioning of the Work.

4. Licensees will implement administrative, technical, and/or other safeguards appropriate in light of the sensitivity of the data to protect private identifiable information from unauthorized access, use, disclosure, or damage.

### *6.5 Summary*

In this Chapter, we focus on four fundamental questions that will have to be resolved before BCIs with enhanced privacy properties come to more a widespread use. Those involve: (A) how do we provide assurance to BCI users that applications they are using are not trying to manipulate the system, in order to access their sensitive and private information; (B) how do we grant users a sufficient level of access control over their resources; (C) how do we relieve users from an unnecessary burden of having to understand what information may be encoded in data resources they are granting or denying access to; (D) how do we incentivize BCI manufacturers to incorporate a BCI Anonymizer component withing their systems?

In answering these questions, we turn to the Value Sensitive Design, systematic approach to accounting for human values in technology design [89]. We build upon the recent VSD project that has proposed the first privacy addendum for open-source software licenses [91].

We analyze the existing BCI technologies in the VSD context, and show how the proposed privacy addendum applies to privacy-enhanced BCIs.

### *6.6 Acknowledgment*

This work is supported by Award Number EEC-1028725 from the National Science Foundation, and by the University of Washington Tech Policy Lab. The content of this Chapter is solely responsibility of the authors, and does not necessarily represent the official views of the National Science Foundation.

We thank Professors M. Ryan Calo, Howard J. Chizeck and Batya Friedman for helpful suggestions, ideas and inspiration, leading to this chapter.

Chapter 7

# CYBER SECURITY ATTACKS ON TELEOPERATED ROBOTIC PROCEDURES

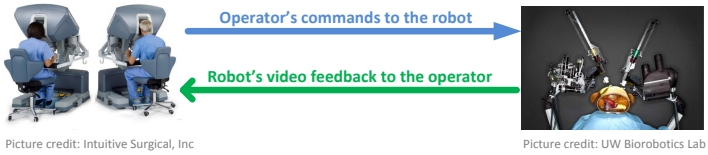
## *7.1 Teleoperated Robotic Systems*

A teleoperated robotic systems can be defined as a system where a human operator uses a robot to interact with far away, inaccessible or dangerous environments at a distance [208]. A typical teleoperated robot consists of one or more robotic arms and end effectors controlled by a human operator. The robot is often referred to as a *manipulator* or a *slave*, and a human operator as a *master*.

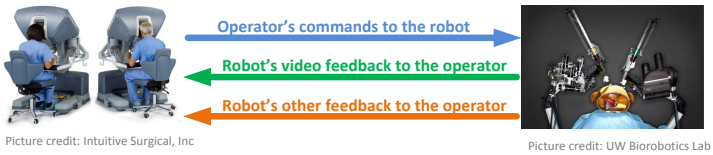
All teleoperated robotic systems use visual feedback via cameras and monitors to show an operator what a robot is doing. Depending on other types of information that may be exchanged between a robot and a human operator, teleoperated robotic systems can be either *unilateral* or *bilateral*, as depicted in Figure 7.1. In *unilateral systems*, position or force commands are sent one way from a human operator to a robot, whereas in *bilateral systems*, in addition to operator's commands being sent to a robot, information about robot's motion, force or haptic (tactile) information is sent back to the operator.

Teleoperated robotic technology dates back to the 1940-s, when such systems were used for the first time to handle radioactive material. The first teleoperated systems were simply long grasping tools. These were soon replaced by more complex mechanical system, which were eventually replaced by more evolved electro-mechanical systems, and then by human-controlled robots. The nuclear industry has to date remained the main producer and consumer of teleoperated robots [127]. For example, teleoperated robots were sent into the Fukushima Daiichi power plant to assess the ongoing reactors meltdown after the 2011 tsunami disaster.

Today, teleoperated robots are in a wide spread use in many applications, including combat zones, disaster relief efforts, handling of explosive or dangerous material, as well as




(a) Unilateral teleoperated robotic system




(b) Bilateral teleoperated robotic system

**Figure 7.1:** Simplified block diagrams of (a) unilateral and (b) bilateral teleoperated robotic systems. In unilateral systems, position or force commands are sent from an operator to a robot, and in bilateral systems, information about robot's motion, force or haptic information is sent back to the operator. In both unilateral and bilateral systems, visual feedback is sent back from a robot to an operator.

space and underwater explorations. For example, mobile rescue robots were used as a part of the clean up and relief effort in the World Trade Center (WTC) disaster [176], and in the 2012 Costa Concordia ship wreak catastrophe [18].

Benefits of teleoperated robotic systems are many. An ability to interact with robots at a distance allows help and relief to be delivered where it would otherwise be too dangerous and inappropriate for human operators to act in person, or at places where skilled human personnel may simply be lacking. Size and motion scaling of a human operator's intended actions allows for robotic procedures to be performed in otherwise inaccessible areas, as well as in very small or extremely large areas.

To date, robotic surgery remains one especially successful example or teleoperated robotic technology. For example, so-called *micromanipulators* are being routinely inserted through a patient's abdominal wall, in order to allow a surgeon to operate on the human body internally, without large incisions, thus resulting in faster and less-complicated recovery [145].

## *7.2 Telerobotic Surgery*

The use of robots in surgery dates to 1985, when *Puma 560* industrial robot was used for needle placement in brain biopsy [124]. In 1988, the *Probot*, developed by Imperial

College, was used to perform prostate surgery [124]. The first actual teleoperated surgical manipulator, where a surgeon indirectly controlled a manipulator using a computer, was developed in the 1990s by the Stanford Research Institute (SRI) [241]. Further developments in teleoperated surgical procedures were advanced by three commercial systems: the *Aesop* and the *Zeus*, developed by Computer Motion, and the *da Vinci*, by Intuitive Surgical [124].

Given that in a teleoperated surgery, a surgeon does not directly control robot's end effectors, but rather, operates through a surgical console (surgical computer interface), an obvious next step for telerobotic surgery was a distance extension between a surgeon and a robot. In September 2001, Dr. Marescaux used the *Zeus* to perform the first transatlantic teleoperated robotic surgery, operating from New York City on a patient in Strasbourg [151]. In 2005, the *da Vinci system* was used to perform the first transcontinental telesurgery, between a surgeon in Sunnyvale and patients in Cincinnati and Denver [241].

In 2014, surgical robotics market was estimated at $3.2 billion [22], and there are several FDA-approved, routinely used surgical robots today. For example, *Neuromate*, developed by Renishaw, is a stereotactic surgical robot, used for brain surgery and pre-operative surgical planning [20]. *RIO* (StykerCorp.) and *ROBODOC* (Curexo Technology) are orthopedic robots, used for unicompartmental and total knee arthroplasty [21], and *Cybernknife* (Accuray) is a robot used for non-invasive tumor removal. It operates by precisely delivering a high dose of radiation to a tumor area, while limiting the exposure of the surrounding tissue to radiation [5].

The best known, and most widely used surgical robot today is the *da Vinci*, developed by Intuitive Surgical [6]. To date, more than 2000 *da Vinci* systems have been sold worldwide, and more than two million patients have been treated using it. It is a unilateral master-slave teleoperation system, used in minimally invasive surgical procedures, where a surgeon controls surgical instruments, inserted into a patient's abdomen. The abdomen is inflated with carbon dioxide, offering enough space for a surgeon to see and operate [127].

### *7.2.1 Next Generation Teleoperated Surgical Robots*

In the next five years, surgical robotics market is expected to reach $20 billion, mostly due to the introduction of the next generation devices, systems, and instruments to operating rooms [22]. These next generation systems are envisioned to provide immediate medical relief in under-developed rural areas, areas of natural and man-made disasters, as well as in combat zones [105].

There are currently several active research efforts, developing the next generation surgical manipulators and surgical consoles. One example is the *Ibis IV*, pneumatically actuated minimally invasive surgical manipulator [216], and a master system based on a delta motion platform [215] (Tokyo Institute of Technology). Another example is an upper-limb exoskeleton master station [187], allowing for whole-arm motion to be scaled down for surgical tasks (UC Santa Cruz). For emergency response and battlefield applications, robot's portability becomes an additional important requirement, and two portable surgical manipulators under current research and development. Those are the *M7* (SRI International) [130] and the *Raven II* (Applied Dexterity) [2].

### *7.2.2 Raven II - The Next Generation Surgical Robotics Research Platform*

The *Raven II*, shown in Figure 7.2, is a research platform used for investigation of advanced robotic-assisted surgery techniques [198, 104]. It is the first technology that supports both software development and experimental testing for surgical robotics. It was developed at the University of Washington with NSF support, and is now manufactured and distributed by Applied Dexterity [2]. It is currently used as a research tool by 15 institution in the U.S., Canada, France, United Kingdom, Denmark, Israel and South Korea.

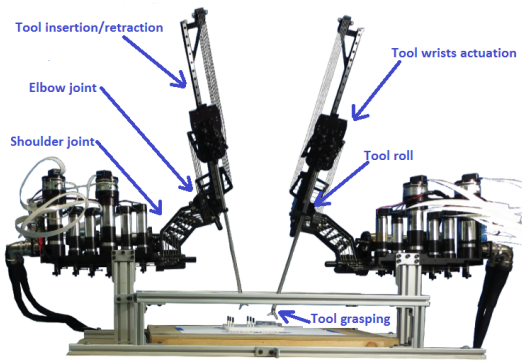



**Figure 7.2:** The *Raven II* next generation teleoperated robotic surgery research platform. The system consists of two 7 degrees-of-freedom surgical manipulators. The motion axes of the robot are: shoulder joint, elbow joint, tool insertion/retraction, tool roll, tool grasping, tool wrist 1 actuation and tool 2 wrist actuation. Picture credit: Applied Dexterity, Inc.

The *Raven II* consists of two 7-degrees-of-freedom (DOF) surgical manipulators, divided into three main subsystems: the static base that holds all seven actuators, the spherical mechanism that positions the tool, and the tool interface. The motion axes of the robot are: shoulder joint (rotational), elbow joint (rotational), tool insertion/retraction (linear), tool roll (rotational), tool grasping (rotational), tool wrist 1 actuation (rotational), and tool wrist 2 actuation (rotational). DC motors mounted to the base actuate all motion axes. The motors of the first three axes have power-off brakes to prevent tool motion in the event of a power failure. Each manipulator has a total (moving plus non-moving) mass of approximately 10 kg, which includes motors, gear heads and brakes. A tool interface allows quick changing of tools, and transmits motion to the tool rotation, grasp and wrist axes. The links and control system support a 3-axis wrist.

The *Raven II* software is based on open technologies, including Linux and the Robot Operating System (ROS) [193]. The low-level control system includes real-time Linux processes (modified by the RT-Prempt Config kernel patch), running at a deterministic rate of 1000 Hz. Key functions running inside the 1000 Hz servo-loop are:

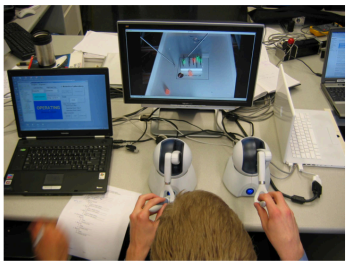


**Figure 7.3:** An example surgical control console, used with the *Raven II*. The console consists of three main parts: surgical GUI, shown on the laptop screen, surgical video transmission on the LCD monitor, and two Omni haptic devices [Picture credit: Sankaranarayanan et al. [205].]

- Coordinate transformations,
- Forward and inverse kinematics,
- Gravity compensation, and
- Joint-level closed-loop feedback control.

The link between the control software and the motor controllers is a USB 2.0 interface board, designed with eight channels of high-resolution 16-bit digital-to-analog conversion for control signal output to each joint controller, and eight 24-bit quadrature encoder readers. The board can perform a read/write cycle for all 8 channels in 125 microseconds. The two *Raven II* arms are controlled by a single PC with two USB 2.0 boards [104].

An example surgical console for the *Raven II* is shown in Figure 7.3. Control inputs and robot feedback, which includes video and haptic information, are transmitted using a communication standard for surgical teleoperation, the Interoperable Telesurgery Protocol (ITP) [130]. The ITP allows communication between heterogeneous surgical consoles (masters) and robots (slaves), regardless of their individual hardware and software. In this protocol, messages from a surgeon and a robot are exchanged using the User Datagram Protocol (UDP) [191], a minimal-latency transport layer protocol, providing unreliable service. Possible negative effects of the UDP's unreliability, which may include out-of-order arrivals, packet duplications and losses, are reduced by transmitting a surgeon's inputs in small increments, based on the assumption that surgical tool motions are continuous.

A surgeon's control message consists of the following fields [130]:

- Position increments,
- Orientation increments,
- Indicator variable "button state" defining actuation of end effectors,
- "Surgeon mode" variable, used to coordinate indexing between a master and a slave,
- Sequence number, and
- Checksum.

### *7.2.3 Extreme Environments Experiments*

Some of the envisioned applications for the next generation teleoperated robotic systems assume operating environment significantly different than current environments. One such example may be a combat zone. Such an extreme environment imposes specific constraints on the next generation teleoperated robotic systems. For example, manipulators may have to operate lacking a basic infrastructure, with limited power resources, in humid, as well as in hot, desert-like climates [241, 198]. Despite harsh conditions, safety of an environment and any personnel in the vicinity of a robot, as well as of the robot itself must be maintained and guaranteed through the entire procedure.

In recent years, several teleoperated robotic experiments were conducted in extreme environments, using the *Raven* and the *M7* systems [105, 129, 142, 128]. In the Hap\SMRT

field experiment [142], the *Raven* was deployed in the Mojave desert. It was controlled across the Internet, with the final link being a UAV-enabled wireless network, where the UAV flew in a pattern around a MASH tent. In NEEMO 12 (mission 16), remote telesurgery was tested in an underwater habitation module in Florida [143]. In these experiments, the following network factors were recognized as critical to system's performance [144]:

- Communication latency,
- Jitters,
- Packet delays and out-of-order arrivals,
- Packet losses, and
- Devices failures.

### *7.3 Problem: Cyber Security Threats Against Teleoperted Robotic Systems*

In addition to stochastic but benign network patterns, operator-robot communication over publicly available networks expose teleoperated robotic procedures to problems most likely not present in current settings. Due to the open and uncontrollable nature of communication networks, it becomes easy for malicious entities to jam, disrupt, or take over the communication between a robot and an operator.

Recent examples, such as Stuxnet worm, specifically designed to target programmable logic controllers, and blamed for ruining a significant part of Iran's nuclear centrifuges [81], exemplify possible issues when a cyber-physical system is targeted explicitly. To date, security has not been a concern for telereobotic robotic systems. Yet researchers have recognized that a variety of possible cyber security vulnerabilities against such systems may be possible [136]. While a few approaches, focusing mainly on private communication [222], and on the ability to verify the code on a robot's side [64], have recently been proposed, there is currently little understanding of what the actual risks are.

This lack of understanding of the actual risks is a function of two factors. At the moment, it is not known:

1. How easy would it be for an attacker to compromise a teleoperated surgery system?
2. What might be the applications of such cyber security attacks?

Not being able to answer these questions makes it hard to understand what the challenges to improving cyber security of teleoperated robotic procedures are, much less to address them. In this dissertation, we seek to answer the above questions through an empirical analysis of one teleoperated surgical system, the *Raven II*. Our work is experimental, and we seek to provide an informed understanding of risks and defenses. We start by analyzing the attacker model.

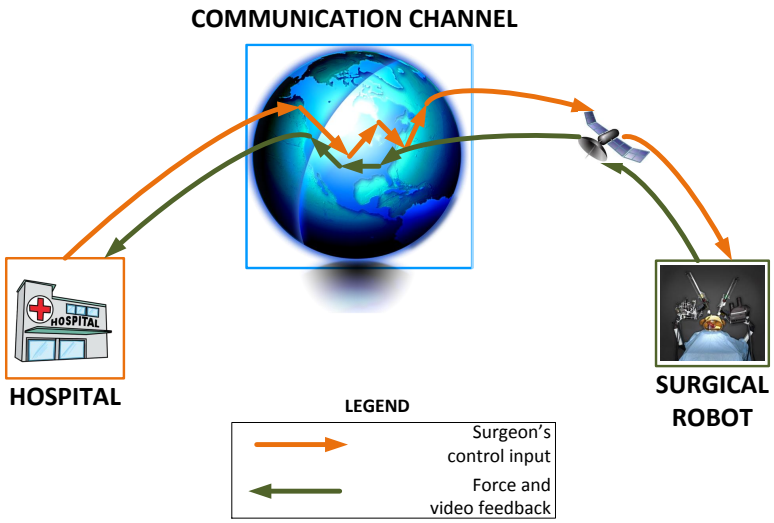



**Figure 7.4:** Visualization of a typical setup for a teleoperated robotic procedure. Orange color indicates an operator's control messages, and green color a robot's feedback messages.

### *7.3.1 Attacker Model*

Many applications of teleoperated robotic systems are envisioned to be extreme conditions applications, such as search and rescue missions, where robots will be expected to operate in low-power and harsh conditions, over a potentially lossy communication channel. The last communication link may possibly be a wireless link to a drone or a satellite, providing connection to a trusted facility, such as a hospital with an established infrastructure, as depicted in Figure 7.4. In such operating conditions, we recognize two attack vectors are feasible [47]:

(a) **Endpoint compromise**, where either an operator's control console or a robot can be compromised, and/or

(b) **Communication-based attacks**, where an attacker may intercept the existing network traffic, inject new malicious traffic, or both.

Endpoint compromises are less interesting since physical access to either side will likely be strictly monitored. Communication-based attacks thus represent a more feasible way to compromise the system. Moreover, due to their abundance and variability, mitigating these attacks is likely to be intellectually challenging, making this the most difficult part of the system to protect. The most likely point of attack appears to be between the network up-link and a teleoperated robot. Since communication will likely be wireless, on-the-field attackers will be able to disrupt the link or manipulate traffic contents. In the rest of this dissertation, we thus focus on disruption and manipulation attacks against teleoperation communication links.

### *7.3.2 Attack Classification*

Based on their impact on human operators, we classify possible attacks on teleoperated robotic systems into three categories:

(A) Intention modification,

(B) Intention manipulation, and

(C) Hijacking attacks.

**Intention modification attacks** occur when an attacker directly impacts an operator's intended actions by modifying his/her messages while packets are in-flight, and an operator has no control over them. These attacks are relatively easy to observe when executed correctly, through e.g., unusual robot movements, robot becoming randomly engaged or disengaged, or unusual delays in movements.

**Intention manipulation attacks** occur when an attacker only modifies feedback messages (e.g., video or haptic feedback), originating from a robot. An operator's messages (and his/her intent) are assumed to be valid. These attacks can prove to be more difficult to mount, simply because of the amount of data that a robot transmits, but if executed correctly, these attacks may be harder to detect and prevent, since they are quite subtle.

Since feedback is assumed to be valid, an operator's (valid) actions may unintentionally become harmful to the environment, or even people in the vicinity of the robot.

In **hijacking attacks**, a malicious attacker causes a robot to completely ignore the intentions of an operator, and to instead perform some other, potentially harmful actions. Some possible attacks includes both temporary and permanent takeovers of a robot, and depending on the actions executed by the robot after being hijacked, these attacks can be either very discreet or very noticeable.

In addition, we consider the role an attacker needs to assume within the system in order to mount an attack, and with respect to that, we classify attackers into two groups:

(i) **Network observer**, and

(ii) **Network intermediary**.

A **network observer** initially eavesdrops on information exchange between an operator and a robot, and based on the collected information, starts inserting false messages into the network, while still allowing both benign parties to communicate directly. A **network intermediary** (i.e., a man-in-the-middle (MitM) attacker) assumes a role of an intermediary between a robot and an operator, thus completely preventing the benign parties from communicating directly. In a real-life attack scenarios, this can be done using methods such as ARP poisoning [27].

For a while now, security experts (e.g., [125, 31]) have been suggesting that the design of security and privacy algorithms for complex systems, including cyber-physical systems, should be an integral part of the system design. A proactive approach towards security and privacy would prevent many security problems from occurring. Moreover, leveraging knowledge and experience from cyber security community, we may be able to predict and avoid many vulnerabilities during the design phase. With teleoperated robotic systems, we have a rare opportunity to design systems and their appropriate security and privacy mechanisms in parallel, and that is exactly the approach advocated and taken in this dissertation.

## *7.4 Related Work in the Security of Cyber-Physical Systems*

Security of cyber-physical systems has been a rapidly growing research area, and in recent years, researchers have been focusing on the topics of monitoring and estimation (e.g., [167, 54]), networked control systems verification (e.g., [220, 84]), as well as robust communication, consensus and distributed computation (e.g., [212, 213, 185]). We give a brief overview of recent security results for the CPS most relevant to teleoperated robotic systems: networked control, automotive, medical systems.

### *7.4.1 Security of Networked Control Systems*

It has been shown that some of the attack classes against networked control systems, wireless sensor networks, and multi-agent systems can be mitigated by relying on a system's dynamics (see, e.g., [31, 55, 166]). In [57, 167, 166], the authors assumed that a system's dynamics are linear, and showed that a simple optimal controller and a Kalman filter can be used to guarantee the desired probability of detecting attacks, such as replay, false data injection and integrity attacks, given a certain model. In [31], the authors considered a networked control system with linear dynamics under a denial-of-service (DoS) attack. They proposed that a semi-definite programming approach can be used to find a causal feedback controller that ensures that the given networked system operates properly (i.e., that ensures that the system's objective function is minimized) while maintaining the system's security and power constraints. The approach proposed in that paper can be applied to teleoperated robotic systems where the linear system dynamics assumption is not satisfied.

### *7.4.2 Security of Automotive Systems*

Automobiles are becoming highly computerized and increasingly "connected", as well as semi-autonomous and autonomous. Recent research has shown that although automotive computer standards indeed describe mechanisms to improve security, these mechanisms are not universally implemented on all the computers in modern cars [132, 58]. In [132], through an analysis of the security properties of all the critical computerized components of a car, the authors found that an attacker connected to the vehicle's internal computer

networks can affect the state of all the analyzed components. In [58], the authors provided a experimental study of an external attack surface on a modern automobile, and they discussed structural characteristics of the automotive ecosystem that make addressing the identified vulnerabilities challenging. Some of the observed challenges can be easily avoided in teleoperated robotic systems, given a relatively early design phase these systems are currently at.

### *7.4.3 Security of Medical Systems*

Security and privacy issues related to telemedical applications were first recognized in the mid-1990s [148, 234]. After the establishment of the Health Insurance Portability and Accountability Act (HIPAA) [10], patients privacy became a primary concern, and researchers focused on the confidentiality of transmitted and stored patient data. More recently, it has been observed that many modern implantable medical devices, including pacemakers and implantable cardioverter-defibrillators, are vulnerable to a variety of attacks, which allow attackers to wirelessly obtain private patient information and change device settings in ways that can directly impact patient health [95, 101].

### *7.4.4 Security of Teleoperated Robotic Systems*

Recently, security concerns related to telerobotic surgery systems have emerged, with a focus on system verification [64], communication reliability [221, 222], and private and authenticated communication [136]. More recently, several research groups considered the impact of cyber security attacks on search and rescue robotic systems [45, 226]. In [45], we proposed the first taxonomy of possible attacks against search and rescue robotic systems, and in [226], the authors experimental investigated possible physical indicators of cyber security attacks against rescue robots.

## ***7.5 Requirements on Teleoperated Robotic Systems***

The use of communication networks to control manipulators at a distance enables us to abstract teleoperated robotic procedures to the information exchange between an operator

and a manipulator. A teleoperated robotic system can thus be seen as an *information system*, depicted in Figure 7.4. While the core functionality of this information system remains the same, services can now be delivered across large distances. These remote environments may, however, put specific constraints and performance requirements on teleoperated robotic systems.

### *7.5.1 Performance Requirements on Teleoperated Robotic Systems*

Teleoperated robot may have to operate in areas lacking basic infrastructure, often with limited access to power and communication resources. Despite severe operating conditions, safety of an environment and any personnel in the vicinity of the robot, as well as of the robot itself, remains priority, and in order to guarantee it, the following performance requirements have been identified as necessary [130, 142, 219]:

(R1) Reliable delivery of control inputs to a robot, and force and video feedback to an operator.

(R2) Stable operation under large communication latencies, and in the presence of jitter and packet losses.

(R3) Robustness to communication delays, device failures and unexpected events.

(R4) Ability to complete ongoing teleoperated tasks in spite of catastrophic events, such as communication blackout.

### *7.5.2 Cyber Security Requirements on Teleoperated Robotic Systems*

The outlined performance requirements are related to the unpredictable, but benign nature of communication networks. However, in order to develop safe and reliable next generation teleoperated robotic systems, able to conform with the stated performance requirements, it is necessary to ensure that these systems are cyber secure and privacy preserving. In [136, 78, 239], the authors recognized that confidentiality, data integrity and authorization are important in guaranteeing privacy and safety of teleoperated surgery patients. We extend this list, and recognize the following traditional objectives of information security [163] as necessary security objectives for teleoperated robotic systems:

(S1) **Privacy/confidentiality**, defined as the ability to keep information secret from all but authorized operators and remote robots.

(S2) **Data integrity**, recognized as the property ensuring that messages to and from an operator have not been corrupted by communication errors or unauthorized entities on their way to a destination.

(S3) **Identity authentication**, representing the ability to confirm unique identities of an operator and a robot.

(S4) **Message authentication**, defined as the ability to undeniably confirm message origin. Closely related to this objective is signature.

(S5) **Signature**, denoting the ability to undeniably tie control commands to an operator, and feedback data to a robot. This objective will be important when multiple operators collaboratively control a remote robot.

(S6) **Authorization**, representing the ability to ensure that only authorized operators can control robots.

(S7) **Access control**, denoted as the mean to ensure that specific control inputs are restricted to privileged operators. This objective may be especially important for teleoperated robotic systems used in education and training, since it is expected trainees will not have the same access rights as certified, experienced operators.

(S8) **Time-stamping**, defining the ability to record the times when control and feedback information are created.

(S9) **Receipt**, denoted as an acknowledgment that information has been received.

(S10) **Confirmation**, representing an acknowledgment that services have been provided. In telerobotic surgery systems, receipt and confirmation objectives are inherently enabled through force and video feedback.

(S11) **Revocation**, defined as the ability to revoke authorization at any point of time. This objective is important for those manipulators intended to be used for only a limited number of times.

When mounting an attack against teleoperated robotic systems, an adversary may exploit a specific system property, and disrupts an ongoing procedure by invalidating any of security objectives (S1)-(S11).

### *7.6 Summary*

A teleoperated robotic systems is as a system where a human operator uses a robot to interact with far away, inaccessible or dangerous environments at a distance [208]. Today, teleoperated robots are being used in a variety of applications, including combat zones, disaster relief efforts, handling of explosive or dangerous material, as well as space and underwater explorations, and their benefits are many.

Many predict that robots today are in a stage similar to that of personal computers in the 1970s, and that the application of robotics will only continue to grow, reaching the number of a couple of billions in a few decades [92]. In the future, teleoperated robots will likely be expected to combine the existing publicly available networks with temporary *ad-hoc* and satellite networks to send video, audio and other sensory data to remote operators [142]. Such teleoperated systems will be used to provide relief in under-developed areas, areas of natural and human-caused disasters, as well as in combat zones [105]. The question arises, however: what if teleoperated robotic systems are compromised?

To date, security has not been a primary concern for teleoperated robotic systems. Yet the problem has recently been recognized [136, 221, 222]. At the moment, however, there is little understanding of what the actual cyber security threats against teleoperated robotics are, and what the impact and implications of these attacks might be.

In this chapter, we analyze properties, constraints and performance requirements on teleoperated robotics. We then identify two attacks vectors as feasible for teleoperated robotic systems, endpoint compromise and communication-based attacks. Based on the observation that communication-based attacks represent a more feasible way to compromise the system, we focus on this attack vector.

Focusing on the impact that communication-based attacks can have on human operators, we then classify possible attacks into three categories: intention modification, intention manipulation and hijacking attacks. In the following chapter, we experimentally evaluate

the actual impact of these attacks, as well as the effort that an attacker would have to invest in mounting these attacks.

### *7.7 Acknowledgement*

This work is supported by the National Science Foundation, Grant # CNS-1329751. Any opinions, findings, and conclusions or recommendations expressed in this Chapter do not necessarily reflect the views of the funding agencies.

We thank Jeffrey Herron, Andrew Lewis, Junjie Yan, Tariq Yusuf, and Professors Howard J. Chizeck, Blake Hannaford and Tadayoshi Kohno for the help in understanding differences, requirements, as well as possible deficiencies and attacks that make security of teleoperated robotic systems different than security of other cyber and cyber-physical systems.

# Chapter 8

# EXPERIMENTAL ANALYSIS OF CYBER SECURITY ATTACKS ON TELEOPERATED ROBOTS

In this chapter, we systematically analyze cyber security attacks against teleoperated robotic systems. We focus on four broad classes of possible attacks:

1. **Intent modification attacks**, defined in Chapter 7 as those attacks where an attacker directly impacts an operator's intended actions by modifying his/her messages while packets are in-flight, and an operator has no control over them.
2. **Hijacking attacks**, i.e., those attacks where an attacker causes a robot to completely ignore the intentions of an operator, and to instead perform some other, potentially harmful actions.
3. **Denial-of-service attacks**, defined as all attacks that make a system or a communication resource unavailable to their intended and legitimate users [112].
4. **Delay attacks**, defined as those attacks where messages between an operator and a remote robot are intentionally and maliciously delayed with the goal of negatively impacting a teleoperated procedure.

For each of the analyzed classes of attacks, we seek to provide an informed understanding of risks and impacts, based on an evaluation of a real technology, the *Raven II*, an advanced teleoperated robotic surgery system. For each attack class, we demonstrate one or more practical examples of attacks, and assess the level of impact on a surgical procedure through a series of human subjects' experiments [47, 50]. Our experiments are based on established robotic surgery tasks, and we quantify the impact using the following metrics: the overall procedure (trial) time, the subjective assessment of difficulty, and the Fitts' index of difficulty.

We introduce Fitts law as a novel way of quantifying the impact of cyber security attacks on cyber-physical systems. It is a formal method of characterizing subjects' performance in

terms of the duration of point-to-point reaching movements, and we propose it can be used to:

(A) Quantify the increase in difficulty of a teleoperated robotic procedure when a system is under attack,

(B) Establish impact equivalence between different attacks, and

(C) Predict the impact of other possible attacks.

## *8.1 Experimental Logistics and Subjects Demographics*

All experimental data used in this Chapter is collected as a part of the study "Analysis of the Impact Adversarial Attacks Have on Teleoperated Procedures", approved by the University of Washington Institutional Review Board (approval # 46946 - EB). All of our subjects are undergraduate and graduate students from the University of Washington, ranging in age from 19 to 30 years. More information about them can be found in Appendix B, in tables B.2–B.4. We can observe that there were 41 experimental sessions in total, out of which 32 were conducted with unique subjects. That is because several subjects have participated in multiple experimental sessions.

We acknowledge that a non-medical student's behavior may differ from a surgeon's behavior, but that is acceptable, and an established experimentation method in surgical robotics, since work [144] has shown that both surgical and non-surgical subjects, upon gaining proficiency, achieve similar results in simple surgical robotic tasks, such as those used in our experiments.

## *8.2 Intent Modification Attacks*

### *8.2.1 Experimental Setup*

We consider three subgroups of intent modification attacks:

(i) Reordering,

(ii) Packet loss, and

(iii) Content modification.

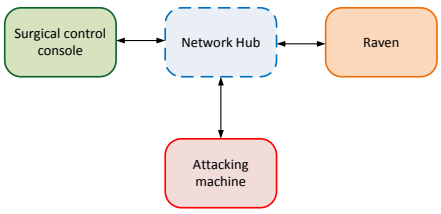



**Figure 8.1:** Experimental setup for intent modification attacks: the attacking machine, *Eve* is running Kali Linux, and all attack implementations are written in Python with the Scapy framework.

We mount these attacks by establishing communication between the surgical control console and the *Raven II* through a network hub, as shown in Figure 8.1. This allows us to connect an external computer to the same subnetwork, and use it to observe and modify communication between a surgeon and a robot. While connecting an attacking computer to the same subnetwork used for communication between a subject and a robot certainly represents a simplification compared to the path an attacker would have to actually undertake, this simplification allows us to abstract a communication between a human operator and a teleoperated robotic systems into two layers: a layer common to all communication systems, and an layer specific to teleoperated robotic systems. By employing the mentioned simplification, we choose to focus on the layer specific to teleperated robotic systems, instead of focusing on known and well-established methods of penetrating communication networks (for examples, see e.g., [125, 85, 157]).

Our attacking computer is running Kali Linux, and all of the analyzed attacks are implemented in Python, using the Scapy framework [37].

### *8.2.2 Experimental Description*

To experimentally investigate impact of intention modification attacks on a teleoperated procedure, we ask our human participants to execute the *Fundamentals of Laparoscopic Surgery (FLS) block transfer task*, a standard test used to train and test surgeons [143], where a subject uses robot's graspers to move six rubber blocks, one at the time, from the left side of the FLS pegborad to the right, and then back to the left side. When moving from left to right, a block is picked up from the peg with the left hand, transferred in the air to the right hand, and then placed on the right peg. Hands are reversed when moving

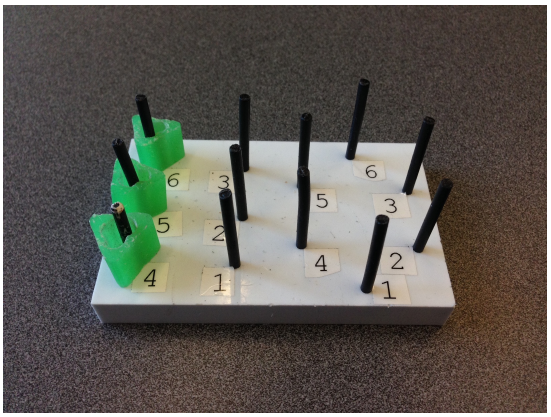



**Figure 8.2:** A pegboard used in intent modification experiments. Each participant was asked to move a rubber block from one of the positions 4, 5, 6 on the left-hand side to one of the positions 2, 3, 6 on the right-hand side.

from right to left. One trial consists of moving all blocks from left to right and then back from right to left, totaling in twelve transfers.

Due to the nature of our investigation, where we focus on the impact of attacks, rather than on subjects' proficiency, we make three simplifications to the standard FLS block transfer task:

(S1) Instead of six rubber blocks, we only use three.

(S2) The subjects are asked to move pegs only from left to right, and the right-to-left movement is not required.

(S3) The subjects are allowed to pick up blocks with a grasper of their choosing, and they are not required to transfer blocks in the air from one hand in the other.

With these simplifications, our trial consists of moving all blocks only from left to right, totaling in three transfers. The pegborad we use is shown in Figure 8.2. In all of the mounted attacks, subjects are aware that attacks might be mounted while they are executing the task, but they do not know when they are being targeted, nor which attack is being mounted against them.

### *8.2.3 Surgeon's Intent Reordering*

Intent reordering is a simple zero-knowledge attack where, instead of forwarding a surgeon's packet to the *Raven II*, we add it to a queue on the attacking machine, referred as *Eve*, that pops items out in a random order once it reaches the maximum length. As a result, all surgeon's messages are delivered to the *Raven II* with a negligible delay (caused by the time spent in the queue). The *Raven II*, however, does not implement all of the received messages. It skips those messages with sequence numbers received out of order, and the effect of skipped messages is a jerky motion of robot's arms, immediately observable by experiment participants.

### *8.2.4 Surgeon's Intent Loss*

Intent loss is another zero-knowledge attack, where we randomly drop individual surgeon's packets or groups of packets. As a result of packets being dropped, the *Raven's* motion

becomes delayed and jerky. We investigate what are the largest tolerable *packets dropping rates*, $\eta$, that still result in a reasonably compliant robot. To do so, we wrote a Python script that sweeps the space of allowed dropping rates $\eta$, $0 \leq \eta \leq 1$.

For individual packet drops, we observe that $\eta \geq 0.55$ generally makes the robot operable, but difficult to use, because grasping becomes challenging. When $\eta$ increases to 0.9, the robot becomes almost unusable, in particular when the required movements are small and precise. For group packet drops, we consider groups consisting of 100 packets (10% of packets transmitted every second), and we find that packet dropping rate $\eta \leq 0.2$ results in a generally operable robot, but the robot's movements are still jerky.

### *8.2.5 Surgeon's Intent Modification*

Leveraging knowledge about the structure of a surgeon's packets, as defined by the Interoperable Telesurgery Protocol [130], we modify surgeon's packets on-the-fly before forwarding them to the *Raven II* through our malicious proxy, *Eve*. Some of the attacks we consider:

(A1) Changing the commanded changes in position,

(A2) Changing the commanded changes in rotation,

(A3) Inverting the grasping states of robotic arms,

(A4) Inverting a combination of the above attacks to fully invert left and right robotic arm, and

(A5) Randomly scaling the commanded changes in position and rotation.

Most of these attacks have a noticeable impact on the *Raven II* immediately upon launch. In particular, if an attack involves any changes to grasping state, even a modification of a single packet has a profound impact on the FLS block transfer task. Unsurprisingly, the least noticeable case is the attack affecting the positions of robot's arms, as long as the modified changes are within the allowed region. Once the modified changes require too large or too fast changes in the positions of robotic arms, thus effectively requiring too high currents, the robot's safety mechanism clips the currents, resulting in a noticeably slower robot motion.

To investigate the human subjects' *assessment of task difficulty*, after every FLS task, we ask the subjects to evaluate the difficulties of the following specific task components:

($D_1$) Reaching each of the blocks,

($D_2$) Grabbing the blocks,

($D_3$) Moving between the pick-up and the put-down locations, and

($D_4$) Performing the task as a whole.

The reported difficulties range from 0 (easy) to 7 (hard), and we analyze data from five subjects. For each intent modification attack scenario and each subject, we sum up the four self-reported difficulty, thus obtaining a single number as a representation of the perceived difficulty of an attack. We refer to this number as the *attack difficulty score*, $D$ [50]:

$$D := D_1 + D_2 + D_3 + D_4 \tag{8.1}$$

The obtained results are presented in Table 8.1, where $NA$ denotes the case when no attack is mounted, and $A$ the case when all instance A1–A5 of intent modification attacks are mounted during one FLS block transfer task.

**Table 8.1:** Subjective assessment of difficulties when intent modification attacks are mounted.

| Subject | $D_1$, NA | $D_1$, A | $D_2$, NA | $D_2$, A | $D_3$, NA | $D_3$, A | $D_4$, NA | $D_4$, A | $D$, NA | $D$, A |
|---|---|---|---|---|---|---|---|---|---|---|
| Subject 1 | 5 | 6 | 7 | 7 | 5 | 6 | 7 | 7 | 24 | 26 |
| Subject 2 | 3 | 2 | 4 | 5 | 2 | 4 | 2 | 3 | 11 | 14 |
| Subject 3 | 2 | 1 | 5 | 3 | 1 | 2 | 5 | 5 | 13 | 11 |
| Subject 4 | 4 | 5 | 3 | 7 | 1 | - | 4 | 6 | 12 | 18* |
| Subject 5 | 5 | 6 | 3 | 5 | 2 | 5 | 4 | 6 | 14 | 20 |

The case of Subject 3 is quite peculiar, since it seems that that subject does not notice the attack. Based on the reported results, we conclude that other subjects do notice attacks occurring. However, all of them carry the task to completion. Moreover, they are able to adjust to the attacks within 1-1.5 seconds time period (even in the case of complete inversions of robot's arms). Performing a random combination of the described attacks does,

however, result in several typical errors, such as dropping the block, moving the robotic arms outside of the allowed workspace, or triggering E-stop, in order to avoid undesirable robot's movements.

## *8.3 Hijacking Attacks*

In hijacking attacks, an attacker assumes a role of a *network observer*, who can eavesdrop on packets between a surgeon and a robot, without modifying them. After sufficient reconnaissance, an attacker then injects new, malicious packets into the network, in order to impact the surgical procedure. In our case, the reconnaissance phase requires only capturing the current packet's *sequence number*, at which point we are able to take over control of the robot. We consider two types of hijacking attacks, namely:

(H1) Sequence number leading attack, and

(H2) Force reset.

### *8.3.1 Sequence Number Leading Attack*

Leveraging again prior knowledge about the structure of a surgeon's packet [130], we conduct the following sequence number leading attack: we first read a single surgeon's packet, and extract its sequence number, $seqNum$. We then add a random offset, $rand$, to a new malicious sequence number, $\tilde{seqNum}$

$$\tilde{seqNum} = seqNum + rand \tag{8.2}$$

where the only requirement is that the offset needs to be less than 1000. We compose a new (empty) surgeon packet with the new sequence number, $\tilde{seqNum}$ and send it to the *Raven II*. At this point, we take over the control of the robot, since the robot attributes the large jump in sequence numbers to packet drops, and as long as the system does not lose more than a second worth of data, the operation continues. From this point on (until the sequence number wraps back to the beginning), our malicious leading packets are implemented and the surgeon's packets are ignored due to the difference in the sequence numbers, and we effectively take control over the teleoperated procedure.

### *8.3.2 Force Reset Attack*

A interesting extension to the described sequence number leading attack, where we abuse the way packet drops are handled and sequence numbers are processed, is the force reset attack, where were abuse the robot's inherent safety mechanism, preventing the robot's arms from moving too fast or moving outside of the allowed area. Every time the *Raven's* arms are commanded to move too fast, or go to an unsafe position, the robot's software imposes a system-wide halt, referred to as software emergency stop (E-stop). This is to protect both electrical and mechanical components of the robot, as well as to ensure an extra level of safety for patients, and human operators standing near the robot.

By sending a leading packet to the robot, where at least one of the changes in position or rotation is too large, and would cause the *Raven II* to either go too fast or to go to a forbidden region, we are able to E-stop the robot. Moreover, by repeatedly sending a malicious leading packet as the one just described, we are able to easily stop the robot from ever being properly reset, thus effectively making a surgical procedure impossible.

## *8.4 Denial-of-Service Attacks*

### *8.4.1 Experimental Setup*

Similar to intent modification attacks, to experimentally investigate impact of DoS attacks on teleoperated surgical procedures, we establish communication between the surgical control console and the robot through a network hub, in order to connect our attacking computer, *Eve* on the same subnetwork. The attacking computer is running Windows 7 SP3, and all of the analyzed attacks are implemented in C#.

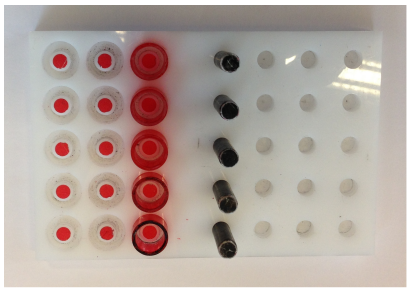


**Figure 8.3:** A board used in telerobotic Fitts' tasks. A subject uses a robot to move the given cylindrical blocks from wells to the pegs. The depicted board represents the "thick" board, "close" configuration scenario.

### *8.4.2 Experimental Description*

To experimentally investigate the impact of DoS attacks on teleoperated robotic procedure, we ask our human subjects to participate in the *telerobotic Fitts' task*. This task was first proposed in [127] as a way of measuring motor control performance of human operators using a teleoperated robot. In this task, a subject uses a robot to move a plastic cylindrical block from a well on the right hand side of the board to a peg on the left hand side of the board, as shown in Figure 8.3. One experimental trial consists of moving five blocks.

In our implementation of the telerobotic Fitts' task, we consider two types of pegs:

(P1) "Thick", and

(P2) "Thin" peg,

with widths respectively equal to 8.00 and 4.60 millimeters

For both types of pegs, we consider two board configurations, defined as functions of the *movement amplitude*, i.e. the center-to-center distance between a block pick-up location and a target peg:

(B1) "Close" board configuration, and

(B2) "Middle" board configuration,

with center-to-center distances, respectively equal to 31.20 and 69.80 millimeters. These center-to-center distances respectively correspond to (B1) the distance between the closest row of pegs and the closest row of wells (31.20mm) and (B2) the second closest row of pegs and of wells (69.80mm)).

For both types of pegs and both board configurations, we consider three attack scenarios:

(AD1) Benign case, when no attack is mounted,

(AD2) Intermediate DoS, and

(AD3) Severe DoS.

To mount intermediate and severe DoS attacks, our attacking machine is set to inject fake packets into the subnetwork, in accordance with transmission of legitimate surgeon and

robot messages, with the goal of creating network layer congestion (network layer attacks). The DoS severity is controlled by the number of malicious threads instantiated on the attacking machine. Each malicious thread generates fake packets of the length 256B with frequency of 1000Hz, and sends them both to the *Raven II* and to the surgical console.

For the intermediate DoS attack, we instantiate 80 malicious threads and for the severe DoS attack 150. These numbers of threads are obtained through an empirical analysis, where the goal was to find the minimal number of threads such that the impact of the attack is noticeable (intermediate DoS), and the maximum number of threads such that the robot is still usable (i.e., the robot is not E-stopping due to too many dropped packets, or too high current levels) (severe DoS).

Combining the considered peg types, board configurations and DoS attack severities, each subject is asked to execute 12 different trials, and the trials are organized in the order presented in Table 8.2, where the order is not know to the participants. This order of trials is specifically chosen to cancel out any inadvertent effect of subjects' learning. In addition, before starting the defined experimental sequence, the subjects are given ample time to learn how to use the system and to gain proficiency with it.

### *8.4.3 Preliminaries: Fitts' Law and One-Part Models of Pointing Performance*

Fitts' law is an empirical model, developed in 1954 as a way to model human performance in pointing tasks. It is often used in human-computer interaction (HCI) research to characterize subjects' performance during simple movement tasks under different speed-accuracy conditions, and it has been applied to physical pointing underwater [126], in near-zero gravity [86], as well as with microscopic targets [135].

Fitts' law characterizes subjects performance in terms of the duration of point-to-point reaching movements, and defines the *movement time*, $T$, as a function of movement distance to the target, $D$, and the width of the target, $W$ [127, 155]:

$$T = a + b \log_2 \left( \frac{D}{W} \right) = a + b(ID) \tag{8.3}$$

where intercept parameter $a$ represents the non-movement time needed to start and stop (finish) the task, and slope parameter $b$ is taken to represent an inverse of the inherent speed

**Table 8.2:** Order of experimental trials for denial-of-service (DoS) attacks.

| **Trial** | **Peg type** | **Configuration** | **Attack type** |
|---|---|---|---|
| 1 | thin | close | no DoS |
| 2 | thin | close | intermediate DoS |
| 3 | thin | close | severe DoS |
| 4 | thin | middle | intermediate DoS |
| 5 | thin | middle | severe DoS |
| 6 | thin | middle | no DoS |
| 7 | thick | close | intermediate DoS |
| 8 | thick | close | no DoS |
| 9 | thick | close | severe DoS |
| 10 | thick | middle | severe DoS |
| 11 | thick | middle | intermediate DoS |
| 12 | thick | middle | no DoS |

of the device, i.e.:

$$b \propto \frac{1}{speed}$$

Parameter $ID$ represents the *Fitts' index of difficulty*, defined as a function of target distance, $D$ and target width, $W$. Summary of notation for this Chapter is provided in Table 8.8.

In [146], the original Fitts' index of difficulty from equation (8.3), $ID$, was redefined as:

$$ID_{Shannon} := \log_2\left(\frac{D+W}{W}\right) = \log_2\left(1+\frac{D}{W}\right) \tag{8.4}$$

Formulation (8.4) is typically referred to as the *Shannon formulation*, due to its similarity with the Shannon's channel capacity theorem, characterizing the channel capacity as a

function of signal and noise powers [155]. This analogy between Fitts' law and the channel capacity theorem is further extended by typically expressing the Fitts' index of difficulty in *bits*.

For an analyzed task, indices of difficulties, $IDs$, are determined in the task design phase (computed from the chosen target distances and widths), and the movement time is measured for every task trial. The metric of interest is the *index of performance*, $IP$, which quantifies how movement times change with task difficulty. It has two competing definitions; *the direct division component*:

$$IP_1 := \frac{ID}{T} \tag{8.5}$$

and *the version derived from linear regression*, using equation (8.3):

$$IP_2 := \frac{1}{b} \tag{8.6}$$

Fitts' formulation, given by equation (8.3), and Shannon formulation, represented by equation (8.4), are often referred to as *one-part models of pointing performance*, because they depend only on the ratio of movement distance to the target, $D$, and the width of the target, $W$, but not on their absolute values [209].

### *8.4.4 Preliminaries: Two-Part Models of Pointing Performance*

**Two-part models of pointing performance** are defined as those models that allow for separable contributions of $D$ and $W$ to movement time, $T$, where *separable* means that values of interest are individual values of $D$ and $W$, rather than just their ratio, and thus the contributions of $D$ and $W$ to movement time $T$ can be weighted with their own constants [209].

The first two-part model of pointing performance was proposed by Welford in 1971, as an alternative explanation for those Fitts'-like tasks where the produced data did not follow the Fitts' model [228]:

$$T = a + b_1 \log_2(D) - b_2 \log_2(W) \tag{8.7}$$

Recently, Shoemaker et al. [209] introduced a variation of Welford's two-part model,

referred to as the *Shannon-Welford formulation*:

$$T = a + b_1 \log_2(D + W) - b_2 \log_2(W) \tag{8.8}$$

As pointed out by Shoemaker, formulation (8.8) combines aspects of the one-part Shannon model (8.4) and the two-part Welford model (8.7) in such a way that the mapping of noise and signal from information-theoretic origins to movement distance to the target and target width are the same as with the Shannon model, but signal and noise are broken into independent terms [209].

### *8.4.5 Fitt's Law Analysis*

*Fitts' Indices of Difficulty*

To analyze task difficulties under DoS attacks, we apply Fitts' law to the experimental data obtained from six subjects. The Fitts' indices of difficulty are computed using distances $D_1$ and $D_2$ respectively equal to 31.20mm for the "close" and 69.80mm for the "middle" board configuration. Widths $W_1$ and $W_2$ are computed as the difference between the block diameter and the peg diameters. For "thin" and "thick" pegs, we obtain:

$$\begin{aligned} W_1 &= w_r - w_{p,1} = 12.8 - 4.6 = 8.2\text{mm} \\ W_2 &= w_r - w_{p,2} = 12.8 - 8 = 4.8\text{mm} \end{aligned}$$

Combining all board configurations and all peg widths, from equation (8.4) we obtain four different indices of difficulty, as presented in Table 8.3.

**Table 8.3:** Fitts' indices of difficulty, $ID$, for two pegborad configurations (close and middle) and two peg types (thin and thick).

| **Configuration** / **Peg type** | close, 31.20mm | middle, 69.80mm |
|---|---|---|
| thin, 8.2mm | 2.3 | 3.3 |
| thick, 4.8mm | 2.9 | 3.9 |

*Data Preprocessing*

In order to compute indices of performance, $IP$, for 12 experimental trials, we combine data collected from six subjects, and computed the mean, $\mu$ and the standard deviation, $\sigma$, of movement times. We then discard outliers, defined as those where the movement time was greater than $\mu + 0.5\sigma$, or less than $\mu - \sigma$. Most outliers occur during the first experimental trial, which seem to imply that subjects are still adjusting to the setup (even though every subject was successful in achieving the prescribed level of proficiency before they continued with the experimental trials).

*Fitts' Indices of Performance*

After discarding the outliers, we combine the trials corresponding to the same attack scenario, and through linear regression find intercept and slope parameters, $a$ and $b$ for each attack scenarios. The obtained results are presented in Table 8.4.

**Table 8.4:** Parameters of the Fitts' model, $a$ and $b$, and Fitts' indices of performance, $IP$, for three considered denial-of-service (DoS) scenarios (no attack, intermediate attack, severe attack).

| **Attack scenarios** | **Slope** $b$ | **Intercept** $a$ | $IP_2$ [bits/s] |
|---|---|---|---|
| No Dos | 0.11 | 3.74 | 9.21 |
| Intermediate DoS | 0.12 | 5.56 | 8.57 |
| Severe Dos | 1.11 | 2.75 | 0.90 |

We observe that DoS attacks have a significant impact on the overall task, since the index of performance decreases under both attack scenarios, especially under the severe DoS case, where it decrease by more than a factor of ten.

### *8.4.6 The Telerobotic Fitts' Law*

Fitts' law was originally developed for *simple movement tasks under different speed-accuracy conditions*. This assumption, however, may not be satisfied for telerobotic Fitts' tasks[1],

[1]As well as for many other applications of Fitts' law.

where in addition to a movement task, a subject needs to perform additional *fine motor tasks*, which involve picking up a cylindrical block or putting that block down on the appropriate peg.

The impact of these additional fine motor tasks on the overall task is assumed to be represented by the intercept parameter $a$ of the original one-part Fitts' model (equation (8.3)). When the Fitts' index of difficulty, $ID$, is equal to 0, no movement should be necessary to execute the tasks, since $ID = 0$ implies that the distance to the target is $D = 0$. Yet, in many such cases, measured movement time is typically strictly positive, $T > 0$, and the model (8.3) implies that the time is governed purely by the intercept parameter.

This observation about *pure movement* and *fine motor* parts of the overall task has an impact on the Fitts' index of performance, and index values (8.5) and (8.6) may significantly differ, depending on the task. However, the observation that the overall task can be divided into different parts implies that those indices are not competing - they are simply conveying different information.

Based on the assumption that telerobotic Fitts' tasks can be divided into two parts, *pure movement* and *fine motor* tasks, we propose a new metric, referred to as the *ratio between pure movement and fine motor tasks*, $\zeta$, and define it as:

$$\zeta := \frac{IP_2 - IP_1}{IP_2} = 1 - \frac{IP_1}{IP_2} = \frac{T - b \cdot ID}{T} = \frac{a}{T} \tag{8.9}$$

where parameters $IP_1$ and $IP_2$ represent indices of performance, defined by equations (8.5) and (8.6). Parameter $\zeta$ takes on values from the range $[0, 1]$, where value $\zeta = 0$ implies that the considered task consists only of pure movement tasks, and value $\zeta = 1$ that the considered task consists only of fine motor tasks, since the index of difficulty, $ID = 0$.

In many telerobotic Fitts' tasks, parameter $\zeta$ will be close to 0, implying that the fine motor tasks are only a small component of the overall task, and the larger the value of $\zeta$ is (for the same overall task setup), the more prominent the fine motor task is in the overall task.

The defined metric $\zeta$ is especially important when evaluating the impact of cyber security attacks on telerobotic systems, where the same attack may not affect each part of the

**Table 8.5:** Duration of experimental movement times (in seconds), Fitts indices of performance $IP_1$ and $IP_2$, Fitts' model parameters $a$ and $b$, and ratio between pure movement and fine motor tasks, $\zeta$, for all subjects over twelve experimental trials.

| **Exp. trial** | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $ID$ | $a$ | $b$ | $IP_1$ | $IP_2$ | $\zeta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Thin, C, NA | 6.78 | 3.58 | 5.05 | 4.12 | 8.12 | 3.96 | 2.91 | 3.74 | 0.11 | 0.57 | 9.21 | 0.94 |
| Thin, C, IA | 8.53 | 4.54 | 6.46 | 4.29 | 8.43 | 6.29 | 2.91 | 5.56 | 0.12 | 0.39 | 8.57 | 0.96 |
| Thin, C, SA | 10.44 | 4.65 | 5.52 | 4.32 | 9.92 | 6.75 | 2.91 | 2.75 | 1.11 | 0.43 | 0.91 | 0.52 |
| Thin, M, NA | 7.58 | 2.80 | 4.64 | 2.99 | 7.56 | 4.33 | 3.96 | 3.74 | 0.11 | 0.79 | 9.21 | 0.91 |
| Thin, M, IA | 11.37 | 5.71 | 6.33 | 4.31 | 9.44 | 7.13 | 3.96 | 5.56 | 0.12 | 0.55 | 8.57 | 0.94 |
| Thin, M, SA | 13.52 | 5.90 | 6.93 | 4.81 | 8.94 | 6.22 | 3.96 | 2.75 | 1.11 | 0.51 | 0.91 | 0.43 |
| Thick, C, NA | 6.90 | 3.84 | 3.56 | 3.65 | 4.94 | 4.23 | 2.27 | 3.74 | 0.11 | 0.72 | 9.21 | 0.92 |
| Thick, C, IA | 5.41 | 5.73 | 5.83 | 4.01 | 7.32 | 6.96 | 2.27 | 5.56 | 0.12 | 0.49 | 8.57 | 0.94 |
| Thick, C, SA | 5.82 | 4.84 | 6.93 | 4.37 | 8.49 | 6.13 | 2.27 | 2.75 | 1.11 | 0.49 | 0.91 | 0.46 |
| Thick, M, NA | 4.28 | 3.93 | 4.06 | 3.38 | 5.78 | 5.21 | 3.25 | 3.74 | 0.11 | 0.95 | 9.21 | 0.89 |
| Thick, M, IA | 6.39 | 4.80 | 5.64 | 5.64 | 7.05 | 7.34 | 3.25 | 5.56 | 0.12 | 0.66 | 8.57 | 0.92 |
| Thick, M, SA | 9.09 | 4.98 | 6.17 | 6.28 | 8.95 | 8.59 | 3.25 | 2.75 | 1.11 | 0.56 | 0.91 | 0.39 |

considered task equally.

The fitted Fitts' model (equation (8.3)) for each of our twelve trials, and each of the six subjects are presented in Table 8.5, where $S_i$ denotes subjet $i$, $C$ the close pegboard configuration, $M$ the middle pegboard configuration, $NA$ the case when no DoS attack is mounted, $IA$ the case of intermediate DoS attack, and $SA$ the case of severe DoS attack.

We observe that when no DoS attack is mounted, the major part of the trial corresponds to the fine motor tasks (picking up the cylindrical block and positioning it down on the appropriate peg). This observation is also true for intermediate DoS attacks, where parameter $\zeta$ tends to be larger than 0.9 for all considered cases ($\zeta = 0.96, 0.94, 0.94, 0.92$). However, this observation is not true for the cases of severe DoS. Under severe DoS attacks, parameter $\zeta$ decreases to the value of 0.5, indicating that, under attack, the movement task

becomes more prominent than in benign case. This observation intuitively makes sense, since under attack, overshoots and undershoots are expected to happen more often that in benign scenarios.

### *8.4.7 Subjective Assessment Analysis*

For each of the 12 experimental tasks, we again ask subjects to evaluate the difficulties of: $D_1$ - reaching each of the blocks, $D_2$ - grabbing the blocks, $D_3$ - moving between the pick-up and the put-down locations and $D_4$ - performing the task as a whole, where the allowed difficulties range from 0 (easy) to 7 (hard). As before, we sum up the four reported difficulties to obtain the *attack difficulty score* [47]. The obtained attack difficulty scores are presented in Table 8.6, where $S_i$ again denotes subject $i$, $D$ the attack difficulty score, $C$ the close pegboard configuration, $M$ the middle, $NA$ the case of no attack, $IA$ the case of intermediate DoS attack, and $SA$ the case of severe DoS.

**Table 8.6:** Subjective assessment of difficulties for denial-of-service (DoS) attacks.

| **Exp. trial** | $D$, $S_1$ | $D$, $S_2$ | $D$, $S_3$ | $D$, $S_4$ | $D$, $S_5$ | $D$, $S_6$ |
|---|---|---|---|---|---|---|
| Thin, C, NA | 10 | 9 | 4 | 6 | 13 | 0 |
| Thin, C, IA | 15 | 12 | 11 | 10 | 16 | 4 |
| Thin, C, SA | 22 | 16 | 15 | 12 | 20 | 6 |
| Thin, M, NA | 5 | 4 | 5 | 11 | 11 | 4 |
| Thin, M, IA | 19 | 19 | 7 | 13 | 17 | 6 |
| Thin, M, SA | 28 | 17 | 10 | 14 | 13 | 5 |
| Thick, C, NA | 7 | 21 | 15 | 14 | 9 | 8 |
| Thick, C, IA | 9 | 18 | 16 | 13 | 20 | 8 |
| Thick, C, SA | 7 | 20 | 12 | 14 | 21 | 10 |
| Thick, M, NA | 12 | 4 | 14 | 11 | 13 | 4 |
| Thick, M, IA | 18 | 16 | 14 | 11 | 17 | 10 |
| Thick, M, SA | 25 | 19 | 15 | 14 | 17 | 11 |

We observe that subjects always rate all attack scenarios as being more difficult than no-attack scenarios. Moreover, the subjects almost always rate the "severe" DoS cases as being more difficult than "no DoS" or "intermediate DoS" cases. The only exceptions seem to be two cases where we start with the "intermediate" attack severity. In those two example, several users both seem to switch the difficulty of "intermediate" and "severe" DoS case.

An interesting effect can be observed when averaging the attack difficulty scores and the trial times over all users, and comparing them for all twelve trials. The results are summarized in Table 8.7, where $S_i$ denotes subject $i$, $\bar{T}$ the average trial time, and $\bar{D}$ the average attack difficulty score. The same results are graphically presented Figures 8.4 and 8.5. For the case where no attack is mounted against the system, the average trial times for different pegboard configurations and different peg widths are decreasing (5.27, 4.99, 4.52, 4.44s), which seems to indicate there is a *learning effect* present. However, the averaged DoS difficulty scores increases with time (7, 6.67, 12.33, 9.67). This may indicate that the subjects are getting fatigued or annoyed (even though their performance is improving). A similar, but a less prominent trend can be observed with the "intermediate" attack as well, where the average trial times remain approximately constant, yet the DoS difficulty score is increasing (11.33, 13.5, 14, 14.33).

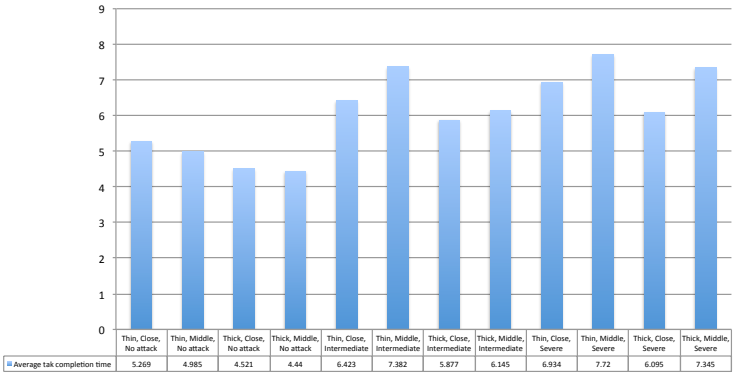



**Figure 8.4:** Average trial times for DoS attacks averaged over all users. We observe that task completion times decrease as subjects progress through experimental trials.

**Table 8.7:** Duration of experimental movement times (in seconds), average trial times and average attack difficulty, $D$, averaged across all subjects.

| **Exp. trial** | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $\bar{T}$ | $\bar{D}$ |
|---|---|---|---|---|---|---|---|---|
| Thin, C, NA | 6.78 | 3.58 | 5.054 | 4.12 | 8.12 | 3.96 | 5.27 | 7 |
| Thin, C, IA | 8.53 | 4.54 | 6.46 | 4.29 | 8.43 | 6.29 | 6.42 | 11.33 |
| Thin, C, SA | 10.44 | 4.65 | 5.52 | 4.32 | 9.92 | 6.75 | 6.93 | 15.17 |
| Thin, M, NA | 7.58 | 2.80 | 4.64 | 2.99 | 7.56 | 4.33 | 4.99 | 6.67 |
| Thin, M, IA | 11.37 | 5.71 | 6.33 | 4.31 | 9.44 | 7.13 | 7.38 | 13.5 |
| Thin, M, SA | 13.52 | 5.89 | 6.93 | 4.81 | 8.94 | 6.22 | 7.72 | 14.5 |
| Thick, C, NA | 6.9 | 3.84 | 3.56 | 3.65 | 4.94 | 4.23 | 4.52 | 12.33 |
| Thick, C, IA | 5.41 | 5.73 | 5.83 | 4.01 | 7.32 | 6.96 | 5.88 | 14 |
| Thick, C, SA | 5.82 | 4.84 | 6.93 | 4.37 | 8.49 | 6.13 | 6.09 | 14 |
| Thick, M, NA | 4.28 | 3.93 | 4.06 | 3.38 | 5.78 | 5.21 | 4.44 | 9.67 |
| Thick, M, IA | 6.39 | 4.80 | 5.64 | 5.64 | 7.05 | 7.34 | 6.15 | 14.33 |
| Thick, M, SA | 9.09 | 4.98 | 6.17 | 6.28 | 8.95 | 8.59 | 7.35 | 16.83 |

## *8.5 Implications of Analyzed Attacks*

Attacks analyzed in this Chapter pose not only technical challenges, but also considerable risks to remote robots, environment and any personnel in the vicinity of robots, as well as to human operators and teleoperated robotic industry as a whole. For example, a compromised teleoperated surgical robot in the midst of even a routine surgical procedure could potentially be used to inflict considerable internal wounds to a patient. Moreover, any extra procedure time, caused by a compromised system, may have severe consequences on a procedure outcome, as well as a patient's recovery. Finally, compromised data and video streams could pose a risk to patient privacy. A surgeon's actions, haptic feedback and a robot's video stream may all contain private and protected patient-related information. For instance, images in the video stream may contain patient identifying features or may expose

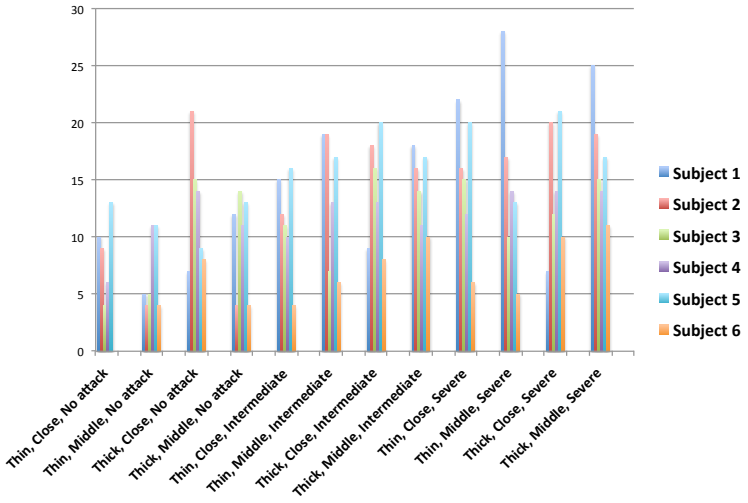



**Figure 8.5:** Average attack difficulty scores for DoS attacks. We observe that attack difficulty scores increase as subjects progress through experimental trials.

portions of the body that the patient would prefer to keep private.

For surgeons, the possibility of surgery systems being compromised complicates the issue of legal responsibility for their actions during procedures. A surgeon will not typically have direct access to a robot and will only operate based on the exchanged information. In a compromised system, for example, haptic feedback may be modified to cause a surgeon to harm a patient. If one could claim that it was reasonable to expect that a surgeon should have noticed that haptic feedback was modified, than the resulting malpractice lawsuit might be strengthened.

Teleoperation security threats may have further implications on surgical robots themselves, since mounted attacks may cause robots to break, or to damage other nearby equipment in the operating room. Finally, any security holes in teleoperated systems present an existential threat to the field of surgical robotics as a whole. Even if attacks are rare, any harm caused by a surgical robot could undermine the public faith in these systems. From a patient perspective, all the advantages in recovery or success rate that come from teleoperated surgery may not be worth the risk of having a potentially hijacked machine operate on them.

## *8.6 Summary*

In this Chapter, we experimentally analyze four broad classes of possible attacks against teleoperated robotic systems: intent modification attacks, hijacking attacks, denial-of-service

attacks, and delay attacks. We specifically focus on denial-of-service and delay attacks, based on the observation that these attacks cannot be mitigated using available cryptographic solutions.

All of our experiments are conducted using *Raven II*, an advanced teleoperated robotic surgery system. For each attack class, we demonstrate one or more practical examples of attacks, and assess the level of impact on a surgical procedure through a series of human subjects' experiments [47, 50]. Our experiments are based on established robotic surgery tasks, and we quantify the impact using the following metrics: the overall procedure (trial) time, the subjective assessment of difficulty, and the Fitts' index of difficulty.

Our DoS experimental results indicate that human operators exhibit learning effects across the given sequence of experimental trials, implying that operators are capable of adapting to unfavorable network conditions. This observation, while positive for system defenders, does not imply that cyber security attacks against teleoperated robotic systems are not a problem. On the contrary, it urges us to quickly develop efficient prevention and mitigation methods, while indicating that in disastrous scenarios, where communication networks may inadvertently be clogged or even DoS-ed, teleoperated robotic systems will remain functional and capable of providing the necessary services.

### *8.7 Acknowledgment*

This work is supported by the National Science Foundation, Grant # CNS-1329751. Any opinions, findings, and conclusions or recommendations expressed in this Chapter do not necessarily reflect the views of the funding agencies.

We thank Lei Cheng, Danying Hu, Kevin Huang and Andrew Lewis, for their help with *Raven II*, Nguyen Le my Chau, Fethya Mohamed Ibrahim, and Xiyu Ouyang for their help with the execution of the experiments, and Jeffrey Herron, Junjie Yan, and Tariq Yusuf, and Professors Howard J. Chizeck, Blake Hannaford and Tadayoshi Kohno, for all of their help with experimental design, setup, execution, as well as with experimental data analysis.

**Table 8.8:** Summary of notation from Chapter 8.

| Symbol | Definition |
|---|---|
| $\eta$ | Packet dropping rates |
| $D_1$ | Assessment of difficulty of reaching a block |
| $D_2$ | Assessment of difficulty of grabbing a block |
| $D_3$ | Assessment of difficulty of moving between pick-up and put-down |
| $D_4$ | Assessment of difficulty of performing the task as a whole |
| $D$ | Attack difficulty score |
| $seqNum$ | Sequence number of the captured packet |
| $rand$ | Random offset |
| $\tilde{seqNum}$ | Sequence number of a malicious packet |
| $T$ | Movement time |
| $D$ | Movement distance to the target |
| $W$ | Width of the target |
| $a$ | Non-movement time needed to start and stop the task |
| $b$ | Inverse of the inherent speed of the device |
| $ID$ | Fitts' index of difficulty |
| $ID_{Shannon}$ | Shannon index of difficulty |
| $IP$ | Index of performance |
| $IP_1$ | Index of performance, direct division formulation |
| $IP_2$ | Index of performance, linear regression formulation |
| $b_1$ | Weight for the movement distance to the target |
| $b_2$ | Weight for the width of the target |
| $\mu$ | Mean of the experimental movement times |
| $\omega$ | Standard deviation of the experimental movement times |
| $\zeta$ | Ratio between pure movement and fine motor tasks |

Chapter 9

# PREVENTING CYBER SECURITY ATTACKS ON TELEOPERATED ROBOTS

## *9.1 Multi-Level Approach to Security of Teleoperated Robotic Systems*

While the attacks that we analyzed in Chapter 8 primarily target specific vulnerabilities of the *Raven II*, the identified exploits will have to be addressed for any teleoperated robotic system. There are several important steps robotics and security community can take to make the next generation teleoperated robotic systems more safe, secure and privacy preserving:

1. Implement updates, changes and enhancement to the used software,
2. Use available hardware for security enhancement,
3. Leverage the existing cyber and cyber-physical security methods, and
4. Develop new security techniques specific to teleoperated robotic systems.

We see the implementation of steps 1–4 as a *multi-layered approach to security and privacy of teleoperated robots*, where the goal of every layer is to provide an additional level of security and privacy to all information exchanged between an operator and a remote robot.

### *9.1.1 Implementation of Software Updates, Changes and Enhancements*

As with many other cyber and cyber-physical systems, maintaining high quality of the used software, and keeping that software (proprietary, as well as open-source) up-to-date will be an important first step toward enhanced security and privacy of teleoperated robotic systems.

**Security update to ITP:** One example piece of software where a security update would certainly be beneficial is the Interoperable Telesurgery Protocol (ITP) [130]. Some of the attacks demonstrated in Chapter 8 could have been easily prevented had the packets' sequence number processing been implemented differently (for example, sequence number

leading attack would not be possible), and had the protocol had checksum checking implemented (e.g., operator intent modification attack would not be possible). Thus, *sequence number processing* and **checksum checking** are the minimal changes needed to increase security. Moreover, implementing these changes will not impact system performance at all.

**Packets processing rate:** Another observed feature that can be turned into a security vulnerability is the fact that teleoperated robots (including *Raven II*) typically execute command packets as soon as they are received. This means that if a burst of packets is received, a robot may start moving very fast and in jerky motions. This may increase the wear on a robot's joints and motors, but more importantly, it may pose danger to humans and the environment in the vicinity of the robot. Moreover, at the moment, an attacker can deliberately cause control commands to be received in bursts. To protect against so-called *burst attacks*, we propose *limiting a robot's processing rate* to a value sufficiently large to never be reached in benign scenarios, but low enough to protect the robot from harm due to a flood of commands.

### *9.1.2 Hardware Solutions*

Many teleoperated robotic systems already implement a variety of hardware and mechanical solutions in order to enhance system safety, to prevent potential damage to a robot, as well as to protect humans and the environment in the vicinity of the robot from possible injuries/damage. One such example are **bounding boxes for robotic arms' joints**. Leveraging this safety-intended tool for security would in many cases be easy to implement, and beneficial for the whole systems. For example, using bounding boxes for robotic arms' joints as a hardware security tool, would prevent robotic arms from ever moving too fast or in jerky motions due to malicious commands sent to the robot.

### *9.1.3 Leveraging the Existing Cyber Security Techniques*

#### *Encryption and Authentication*

Several attacks that we demonstrated in Chapter 8 were successful due to the fact that valid packets were accepted by the robot from any source. For the *Raven II*, this was almost

certainly an oversight, and it is easy to fix. However, we need to consider the larger problem of how to protect against a more sophisticated *packet spoofing attacks* that also spoof source IP and port information. One straightforward answer is to *encrypt all data streams* between the two endpoints rendering all but the man-in-the middle attacks impossible. An advantage that teleoperated robotic systems have over many other communication systems is that there is likely dedicated staff at one end of the system at least. This means that there exists an out-of-band communication method, such as texting or talking on the phone, to exchange a private piece of information that can be used to authenticate data streams. By encrypting and authenticating data streams between an operator's console and a robot, an attacker's ability to initiate intention modification, manipulation, or hijacking attacks becomes severely hampered.

**Cost of Encryption:** In order to investigate the cost of encrypting all data exchanged between an operator and the *Raven II*, that is, all data in the network, but not the side out-of-band communications, we used an intermediary computer with Intel Core2 Quad CPU processor running at 2.5GHz, to execute cryptographic tasks on. We acknowledge that the results obtained through this analysis do not necessarily represent the exact results we would have observed had we encrypted all packets closer to an operator and to the *Raven II*, but for this analysis we only wanted to measure the added overhead of cryptographic operations. We used the Advanced Encryption Standard (AES) encryption method [67], and considered three different key lengths:

- 128-bits,
- 192-bits, and
- 256-bits.

For all key lengths, no noticeable increase in CPU usage was observed, compared to the baseline case where the intermediary computer only received packet and forwarded them further. However, we observed an increase in memory usage, with the average increase of 3000KB. This increase value will likely be acceptable for the majority of teleoperated robotic systems. As expected, we did not observe a significant memory usage difference between

different key lengths. Thus, the use of encryption and authentication has low cost and high benefits to telerobotic surgery, mitigating many analyzed attacks.

*Preventing Denial-of-Service Attacks*

Denial-of-service attacks, evaluated in Chapter 8, were successful in disrupting teleoperated robotic procedures mainly because no mitigation mechanism against it is currently in place. In the security community, prevention and mitigation of DoS attacks typically relies on a combination of malicious activity detection, network traffic monitoring and malicious traffic blocking [139]. The existing approaches can generally be divided into preventive and reactive methods [165]. Some of the existing methods against DoS attacks are:

(i) Intrusion Prevention Systems (IPSs) [206],

(ii) DoS Defense System (DDS),

(iii) Blackholing [186],

(iv) Pipes cleaning [30], and

(v) Channel surfing [236].

Intrusion Prevention Systems (ISPs)-based approaches are a type of preventive methods. They require a known attack signature in order to be able to stop the attack. These attack signatures typically use packets' content and network behavior as features of interest. The problem with DoS attacks, however, is that it is relatively easy for an attacker to flood the communication channel with legitimate packets. For example, an attacker can simply capture one legitimate message between a robot and an operator, copy it multiple times and overflow the network with it. Similarly, in a distributed DoS setting, it may be very hard to distinguish between malicious and legitimate network behavior, since the attack task is spread over a large number of computers.

Blackholing [186] is a reactive DoS mitigation strategy, where all the traffic from an attacking network entity is being rerouted to a non-existent server, typically referred to as the "black hole". The problem with this approach, however, is that in rerouting all the traffic from an attacking network entity, we may end up rerouting the legitimate traffic as well,

thus effectively completely preventing communication between a robot and an operator. In order for this approach to be effective, there would have to exist a way to quickly and efficiently distinguish between a valid and a malicious traffic, and only reroute the malicious traffic.

Pipe cleaning [30] is another reactive DoS mitigation method. In it, all traffic is passed through a so-called "scrubbing center" where all packets are inspected and only legitimate ones are forwarded. The problem with this approach is the fact that many teleoperated robotic systems require (near) real time operation and communication, and this approach may negatively impact that requirement.

Analyzing the existing DoS mitigation strategies exposes challenges unique to the security of teleoperated systems, namely tensions between real-time operation and security [50]. It therefore may be hard to find one out-of-the box approach to prevent DoS attacks against teleoperated robotic systems, and successfully use it in a variety of telerobotic scenarios. A more feasible approach may be to try to combine some of the existing proactive and reactive approaches.

*Tensions Between Security and Performance of Teleoperated Robotic Systems*

Throughout the experimental analysis presented in Chapter 8, we observed another feature specific to the security of teleoperated robotic systems. Namely, in many different instances, and with varying severity, we observed *tensions* between a system's security and privacy and other desired system properties. A few examples of those tensions are as follows:

**Tension Between Real-Time Operation and Security:** In order to ensure fast enough operation, many teleoperated robotic systems resort to using datagram protocols. It is typically assumed that operators motions and commands are continuous, and that transmission rates between an operator and a robot are sufficiently high, so that occasional benign packet losses have negligible effects on the overall procedure. Yet, in a hostile setting, an attacker with sufficient knowledge of a system may abuse the protocol, and specifically drop certain packets in order to cause maximal damage (harm), while being cautious about his/her own resources. Since datagram-based protocols are likely to remain the preferred choice for tele-

operated robotic systems, an appropriate strategy to mitigate this type of threats will have to be found.

**Tensions between Safety and Security**: Many teleoperated robotic systems use an Emergency stop (E-stop) button. In a benign case, an E-stop is a mechanism designed to improve safety of near-by equipment and operators, and of a robot. Our experiments have shown, however, that the existence of E-stop may actually lead to decreased safety and security of a robot and of people in the case of a compromised system. An attacker with sufficient knowledge of the system may easily abuse E-stop to render a robot unusable. For example, by occasionally sending leading packets, where at least one of the changes is sufficiently large, an attacker may cause the system to be permanently E-stopped. The challenge thus arises to reconcile the benefits of E-stop in the benign case with its possible negative consequences in the adversarial setting.

**Tension between Fast Feedback and Privacy:** Many of the demonstrated attacks may be mitigated by encrypting and authenticating all communication between an operator and a robot. Yet, due to sheer quantity of video data from a robot, and the real-time operation requirement, encrypting the entire feedback channel may not be feasible. In this case a trade-off between the real-time feedback and possible privacy requirements/expectations in the remote environment may arise. Based on the experimental results, however, we propose that authenticating all packets should be the minimum required feature for any teleoperated robotic system operating over a public network, so as to assure that packets from any other sources are never accepted as real.

### *9.1.4 Solutions Specific to Teleoperated Robotic Systems*

Observed tensions between security and privacy, and other desired system properties may render many existing security techniques infeasible for teleoperated robotic systems, possibly requiring new security approaches to be developed. Many of the new methods may rely on the unique component of teleoperated robotic systems, namely the *human component* of a system.

Human users (operators, patients) have a unique way of interacting with a cyber-physical

system. For example, users' biosignals are increasingly being used to personalize biomedical systems, and those signal have been shown to be user-specific, and unique to users [150, 119]. Similarly, it has recently been shown that users have a unique way of interacting with haptic devices, in terms of forces and torques applied on a haptic tool [39, 238]. These users' idiosyncrasies and unique features may expose cyber-physical systems to potential security and privacy risks. At the same time, however, users' unique traits and ways of interacting with a system can be used to increase the system's security, privacy and usability properties.

In the rest of this chapter, we investigate the ways in which operators' idiosyncrasies can be used to increase and enhance security and privacy properties of a teleoperated robotic system. We focus on two specific examples:

(a) Haptic passwords [238], and

(b) Operator signatures [61].

*Haptic Passwords*

Haptic passwords are a new biometric method to identify and authenticate operators of a teleoperated robotic system (or users of touch screens, stylus-based tablets, or smart phones). This biometric method is based on our hypothesis that ways in which users haptically interact with devices provides unique user-dependent features, and that we can use those unique features for user identification and authentication.

*Operator Signatures*

Leveraging again our hypothesis that each operator interacts with a remote robot in a unique way, thus generating a unique biometric, we introduce the concept of operator signatures as a method to monitor an operator's and a robot's actions in real time, in order to enable:

(A) Continuous identification and authentication of an operator,

(B) Real time monitoring and validation of an operator's and a robot's actions, and

(C) Enhanced operator training.

*Relationship Between Haptic Passwords and Operator Signatures*

Although haptic passwords and operator signatures rely on the same hypothesis, that human operators have unique ways of interacting with their operator consoles, with robots, as well as with remote environments, these two concepts fulfill different security roles in a teleoperated robotic system. A simplified block diagram of a teleoperated robotic system with haptic passwords and operator signatures-based monitoring components is shown in Figure 9.1.

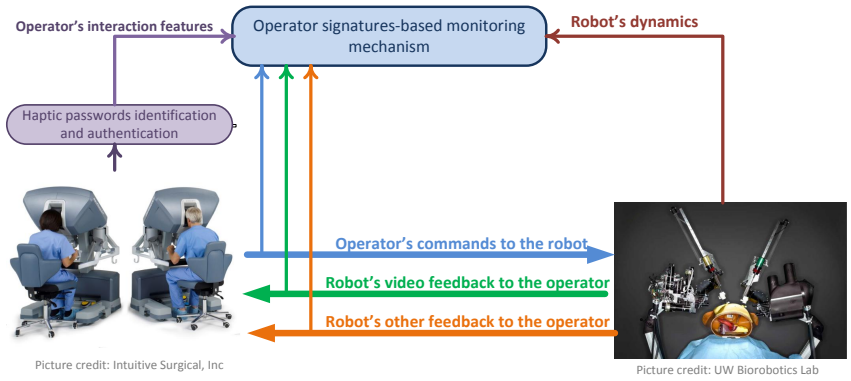



**Figure 9.1:** Block diagram of a teleoperated robotic systems with implemented haptic passwords and operator signatures systems. Haptic passwords represent a local identification and authentication component of a system, that operates based solely on the information available on an operator's side of a system. Operator signatures-based monitoring component is a global system component, operating continuously throughout the whole procedure, based on information from both sides of the system, as well as the exchanged messages.

Haptic password represent a *local* identification and authentication component of a system. It operates based only on the information available on an operator's side of a system (local information), and it is activated at the beginning of a teleoperated procedure, and possibly in some discrete steps after that, to re-authenticate the operator.

Operator signature-based monitoring component is a *global* system component, operating continuously throughout a whole teleoperated procedure. It performs continuous identification, authentication, and validation of an operator, as well as real time monitoring of both an operator and a robot. This monitoring mechanism bases its decisions and actions on the following information:

- Messages exchanged between a robot and an operator,

- o Information available about an operator from his/her console (haptic password, console interaction features, etc.), and
- o Information available about a robot.

### *9.2 Haptic Passwords*

Many existing cyber and cyber-physical systems already rely on the use of passwords for identification and authentication of human users, and the same is expected from teleoperated robotic systems, especially with the foreseen increase in their application space. Existing identification and authentication methods can broadly be classified into those using *alphanumerical passwords* and those relying on classical *biometric properties of users*, such as fingerprints, voice data, or users' iris.

Alphanumerical passwords are still the most widely used because they are relatively easy to implement, and typically have a relatively simple updating process. There are, however, drawbacks to the use of alphanumerical passwords, mostly caused by human users, who tend to:

- Use overly simplistic passwords that are easy to memorize, but also easy to break,
- Reuse their passwords across different platforms and systems, and
- Not update their passwords regularly.

Additionally, alphanumerical passwords may have limited possible password spaces, and may thus be vulnerable to dictionary and brute force search attacks [36, 178].

Biometric-based identification and authentication systems, on the other hand, remove the burden of having to memorize a password. Nonetheless, most widely used biometric-based systems, such as fingerprint or iris recognition, come with their own set of shortcomings, including:

- ⋆ Potential privacy issues that may arise from the use of these passwords [118],
- ⋆ Relatively lower accuracy rate [117], and
- ⋆ Limited ability to update these passwords.

These drawbacks create the need for new identification and authentication systems, preferably such that combine the good properties, but avoid the shortcomings of alphanumerical and biometric-based system. One possible novel system is based on *haptic interaction*. Leveraging the fact that each individual user interacts with a force feedback (haptic) device in a unique way, we propose a new type of biometric identification and authentication, referred to as a *haptic password system*. Such a system is expected to significantly increase the space of possible passwords, making dictionary and brute force attacks much harder to accomplish. In addition, unlike other biometric-based identification and authentication methods, haptic-based passwords can be updated if the need arises.

### *9.2.1 Background and Related Work*

#### *Haptic Interaction System*

A simplified block diagram of a haptic interaction system is depicted in Figure 9.2. In such a system, an operator interacts with a remote or a virtual environment through the use of a *haptic device*, which enables a *bi-directionally flow of force information* between an operator and an environment. The haptic device senses the position, velocity, applied forces and torques of the operator's hands, and uses this information to update the position of the haptic interaction point (HIP), which has a role similar to that of a mouse cursor in virtual environment. Collision detection algorithms then determine points of contact between the HIP and objects in the virtual environment, and appropriate interaction forces are computed and applied to the user's hand. This provides an extra channel for human-computer interaction, referred to as *haptic force feedback*. Haptic force feedback enhances the performance of human-computer interaction, since users tend to have a unique and distinguishing way of responding to haptic force feedback, which can be used for identification and authentication.

#### *Related Work on Behavioral Biometrics-Based Identification and Authentication*

Biometrics, defined as the use of human characteristics for identification and authentication purposes can broadly be divided into two main categories: physiological and behav-

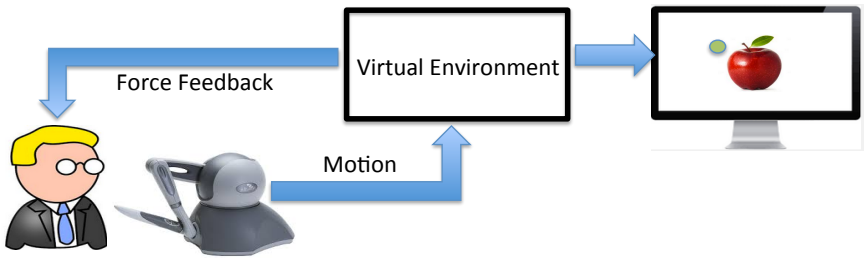



**Figure 9.2:** Block diagram of a haptic interaction system. An operator interacts with a remote or a virtual environment through the use of a haptic device, which enables a bi-directionally flow of force information between an operator and an environment.

ioral [237]. *Physiological biometrics* rely on physical attributes, such as fingerprints, hand geometry, facial features, neural and other biosignals, typically assumed to be permanent and static over the course of a person's life. *Behavioral biometrics*, on the other hands, rely on invariant (or presumed to be invariant) features and traits related to a person's behavior, such as speech, walk, and typing, for identification and authentication [87].

Early behavioral biometrics were based on keystroke dynamics or mouse movements, with the average error rates varying between 5%-15% [211, 179, 63, 170, 121]. In order to improve these relatively high error rates, various authors started proposing *multi-modal biometrics* approaches, combining physiological and behavioral features, such as voice, face and signatures [131].

Moving a step further, in recent years, researchers have been showing an increased interest towards the use of touch behavior biometrics in order to identify and authenticate users. In [69], authors proposed a password application where a user draws strokes on a touch screen as a password. The application uses pressure, coordinates, size, speed and time of a stroke to identify a user. Similarly, authors of [243] proposed that a user can be identified based on the way (s)he taps on a touch surface with one or more fingers. They suggested using acceleration, pressure, size and timing of a single tap as identification and authentication features.

### *9.2.2 Preliminaries*

#### *Discrete Wavelet Transform*

The Discrete Wavelet Transform (DWT) is a modified wavelet transform for which wavelets are discretely sampled to deal with discrete signals. The main idea of the DWT is to rep-

resent a time series as a linear combination of a set of functions generated from a mother wavelet. The weighting parameters are called wavelet coefficients. A key advantage of any wavelet transformation is that it captures both frequency and localized (in time) information. This facilitates the feature extraction process later on [97].

The DWT coefficient of signal $x$ is calculated by passing it through a series of filters generated from a mother wavelet filter. The mother wavelet filter $g$ is a low-pass filter that satisfies the standard quadrature mirror condition [207]:

$$G(z)G(z^{-1}) + G(-z)G(-z^{-1}) = 1 \tag{9.1}$$

where $G(z)$ denotes the z-transform of the filter $g$. Its complimentary high-pass filter can be obtained as:

$$H(z) = zG(-z^{-1}) \tag{9.2}$$

The mother wavelet filters are used to generate the series of filters of increasing width:

$$\begin{aligned} H_{i+1}(z) &= H(z^{2^i})G_i(z) \\ G_{i+1}(z) &= G(z^{2^i})G_i(z) \end{aligned} \tag{9.3}$$

with initial condition $G_0(z) = 1$. Filters (9.3) can equivalently be expressed in the time domain as:

$$\begin{aligned} h_{i+1}(k) &= [h]_{\uparrow 2^i} \times g_i(k) \\ g_{i+1}(k) &= [g]_{\uparrow 2^i} \times g_i(k) \end{aligned} \tag{9.4}$$

where notation $[\cdot]_{\uparrow m}$ denotes upsampling by a factor of $m$. Figure 9.3, from [238], shows a typical block diagram of the DWT process. At each level in the diagram, the signal is decomposed into low and high frequencies. The high frequency component of each level is regarded as the detail coefficient of the corresponding level.

*Artificial Neural Network*

Artificial neural networks (ANNs) are a family of statistical learning models, inspired by biological neural networks, and used to estimate or approximate functions that can depend

on a large number of generally unknown inputs. These learning models typically consist of multiple interconnected *neurons*, whose connections have tunable weights, making ANNs adaptable and capable of learning. The weights are used to amplify or de-amplify original input signals. Once weighted, the signals are added together and passed into the *activation function*, which is used to convert an input into an output.

From a hierarchical point of view, an ANN consist of an *input layer*, a *hidden layer* and an *output layer*, and there can be any number of nodes per layer. Typically, there are also multiple hidden layers to pass through before ultimately reaching the output layer. ANNs can broadly be divided into *feedforward networks*, allowing signals to pass through the layers of the network in a single direction, and *feedback networks*, where signals are allowed to pass through layers in both directions.

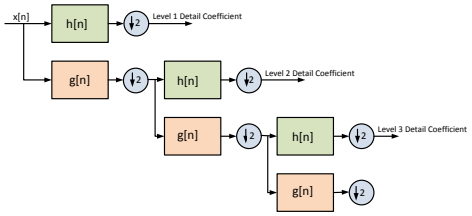



**Figure 9.3:** Block diagram of a discrete wavelet transform sub-band decomposition. [Figure credit [238].]

Learning in ANNs practically means finding an algorithm to update the neurons' weights, typically in some optimal sense. In other words, given a specific task, and a specific class of ANNs, $F$, we want to use the set of observations to find $f^* \in F$ which solves the given task in some optimal sense. This, in turns, means that we have to define a cost function $C : F \to \mathbb{R}$, such that, for the optimal solution, $f^*$, no other solution has a cost less than the cost of the optimal solution.

There exist three major learning paradigms for ANNs, each corresponding to a particular abstract learning task, *supervised learning, unsupervised learning and reinforcement learning*. For more details on ANNs, please refer to e.g., [100, 107].

### *9.2.3 Haptic Passwords System*

Our proposed haptic password system consists of three main parts: (i) data collection, (ii) feature extraction, and (iii) classification, as shown in Figure 9.4.

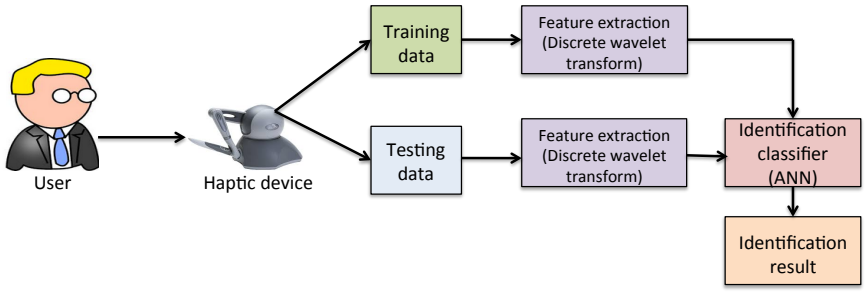



**Figure 9.4:** Block diagram of the haptic passwords system.

*Data Collection Component*

For our proposed haptic-based identification and authentication system, we collect the following data:

(D1) Position of the tip of a haptic device in virtual environment $(x, y, z)$,

(D2) Applied forces, $(f_x, f_z)$,

(D3) Orientation of the haptic device $(q_{\text{pitch}}, q_{\text{roll}}, q_{\text{yaw}})$

The state vector, $v$, is organized as:

$$v := [x, y, z, f_x, f_z, q_{\text{pitch}}, q_{\text{roll}}, q_{\mathbf{yaw}}] \tag{9.5}$$

All experimental data is recorded at a 30 Hz sampling rate, where the software starts recording data when the tip of the haptic device makes contact with the virtual paper, and stops when no more contact is detected.

*Feature Extraction Component*

Since our recorded haptic signals contain transient and localized features, we choose the DWT as an extraction method, since DWT, like all wavelet methods, capture signals frequency properties while conserving its local features. The feature vector for each experimental trial is obtained in the following steps, first described in [116]:

(S1) The position data, $(x, y, z)$, is differentiated to obtain the velocity data, $(v_x, v_y, v_z)$.

(S2) The data set of each trial is resampled to 128-point length (i.e. for each trial the data size is $8 \times 128$, where 8 is the dimension of the data). This resampling make the discrete wavelet transform process feasible.

(S3) The DWT is applied to each channel separately. The mother wavelet is the Duabechies Wavelets of order 4. For each channel, seven levels of detail coefficients, $D_1 - D_7$, are obtained.

For detail coefficients $D_1 - D_5$, the following statistical features are found to represent the time frequency distribution:

(F1) Maximum of wavelet coefficients in each level,

(F2) Minimum of wavelet coefficients in each level,

(F3) Mean of wavelet coefficients in each level, and

(F4) Standard deviation of the wavelet coefficients in each level.

The lengths of detail coefficients $D_6$ and $D_7$ are 2 and 1 respectively, and because of that, we insert them into the feature vector directly. Therefore, the feature vector for each channel, $f_i$ is given as:

$$f_i := [w_1, w_2, w_3, w_4, w_5, w_6 \equiv D_6, w_7 \equiv D_7] \tag{9.6}$$

where $w_i$ defined as:

$$w_i := [\max(D_i), \min(D_i), \text{mean}(D_i), \text{std}(D_i)] \tag{9.7}$$

The length of the feature vector for every channel is 23. Finally, the feature vector for every experimental trial is obtained by combining together feature vectors for every channel:

$$F := [f_1, f_2, \ldots, f_8] \tag{9.8}$$

*Classification Component*

Based on the obtained feature vector, $F$ an artificial neural network is implemented to complete the user classification task. Our chosen ANN consists of 184 inputs, one hidden layer with 500 neurons and number of outputs equal to the number of system users. The output $O$ is thus a vector of length $N$, where each output element can be between 0 and 1,

where a zero-value of the $i^{\text{th}}$ element indicates that it is unlikely that the collected data is generated by user $i$, while a value of 1 indicates that data is very likely generated by user $i$.

We use a **scaled conjugate gradient backpropagation** [169] supervised learning method to train the network. All training parameters use default settings. In order to obtain satisfying training results, the stop criterion is set to minimize the mean square error before validation failures reach 100 or the performance gradient is less than $1 \cdot 10^{-10}$.

### *9.2.4 Experimental Analysis*

To evaluate the proposed haptic password system, we conduct an experimental study with human experiments, where we ask out participants to interact with a virtual 3D environment via a 3 degree-of-freedom (DOF) haptic device, the SensAble PHANToM Omni [154]. Using the haptic device, the participants manipulate the virtual pen, which enables them to write on a virtual paper. The virtual paper is slightly tilted (15 degrees) towards the user, and the position of the pen is visually depicted as a red cursor and a black shadow, representing the projection of the pen tip on the virtual paper. In addition, force feedback from the interaction between the pen and the paper is rendered haptically, allowing the participants to both see and feel the virtual paper.

#### *Subjects Demographics*

The conducted experiments are a part of the study "Analysis of the Impact Adversarial Attacks Have on Teleoperated Procedures", approved by the University of Washington Institutional Review Board (approval # 46946 - EB). All of our subjects are undergraduate and graduate students from the University of Washington, ranging in age from 22 to 30 years (9 subjects in total). There are eight right-handed participants and one left-handed participant. Most of the subjects had not used a stylus haptic device prior to the experiments. More details about experimental subjects can be found in Appendix B.

*Experimental Task*

Before each experiment, subjects are asked to explore the environment and get used to the haptic device, and the sensation of its force feedback. Once they gain proficiency with the haptic device, they are asked to execute the following three tasks:

(T1) Draw an L-shaped pattern ,

(T2) Write word 'SEAHAWK' (all in uppercase), and

(T3) Sign their own name (own signature).

Subjects are given a practice period before each task type in order to gain sufficient proficiency, and to limit possible nehative learning effects. After practice, each task was repeated 10 times per user.

### *9.2.5 Experimental Results*

*Relative Password Variation*

To evaluate our proposed haptic password system, we consider *the relative password variations* for different subjects and different tasks.

**Definition 3** *The relative password variation is the variation of one subject's password relative to the distance to its most similar subject's password. The smaller the variation is, the better classification performance will be. The relative password variation can be computed as:*

$$PV_i = \frac{\sum_{j=1}^{N} \|F_{ij} - \bar{F}_i\|_2}{N} \cdot \frac{1}{\min_{j \neq i} \|\bar{F}_i - \bar{F}_j\|_2} \tag{9.9}$$

*where:*

$PV$ *- password variation,*

$N$*- number of trials,*

$F_{i,j}$ *- Feature vector of user* $i$ *in trial* $j$*, and*

$\bar{F}_i$ *- Mean feature vector of user* $i$*.*

Table 9.1 shows the relative password variation among 9 subjects for the given tasks, where $S_i$ denotes subject $i$. We notice that among the given tasks, the signature task varies the least. The main reason for this result if probably the fact that most subjects find it very easy to write their own signatures (on a paper, or on a touch surface), and the mental effort required for this task is likely lower than for the first two tasks. The intra-subject performance and execution of this task is thus more consistent. Not surprisingly, signature data generates better user authentication results than the data from the first two tasks.

**Table 9.1:** Relative password variation for the given haptic interaction tasks.

| **Task** | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | $S_8$ | $S_9$ |
|---|---|---|---|---|---|---|---|---|---|
| **L-shape** | 1.11 | 1.01 | 0.36 | 0.66 | 0.46 | 1.14 | 0.55 | 0.75 | 0.87 |
| **Seahawk** | 0.67 | 1.16 | 0.23 | 0.88 | 0.76 | 0.68 | 0.87 | 0.90 | 1.02 |
| **Signature** | 0.63 | 0.57 | 0.32 | 0.62 | 0.46 | 0.50 | 0.62 | 0.58 | 0.71 |

*User Classification and Authentication*

For haptic-based user classification and authentication, the ANN algorithm was trained using each subjects' seven trials per task, thus leaving three trials per task for evaluation. All $\binom{10}{7} = 120$ training and testing combinations were examined, and classification performance was computed by averaging results for all combinations.

Since the output of the algorithm was a 9-dimensional vector, representing the likelihood that data was generated by a particular subject, for each examined trial, the data was attributed to a specific user as follows:

$$i := \arg\max_j O(j) \tag{9.10}$$

We refer to equation (9.10) as the **classification task**.

Finally, in the **authentication task**, a user is authenticated as a valid user if the likelihood that the given data was generated by him/her is higher than some predefined

threshold, $t$:

$$O(i) > t \tag{9.11}$$

The classification performance results for the first three tasks are shown in Table 9.2. Even for the L-shape pattern, considered to be the simplest task, the classifier is able to attribute data to the correct human subject in the 95.93% (25.9 out of 27) cases. When the task became more complex, and possibly more personalized, as it is the case with the latter two tasks of writing the word 'SEAHAWK', and the subject's personal signature, the method attributes data to the correct human subject in the 98.95% (26.7 out of 27), and 100% cases, respectively.

**Table 9.2:** User classification correctness rate for the given haptic interaction tasks.

| **Task** | Correct classification rate |
|---|---|
| **L-shape** | 95.93% |
| **Seahawk** | 98.95% |
| **Signature** | 100% |

### *9.2.6 Theoretical Analysis of the Possible Space of Haptic Passwords*

In order to characterize the strength of a haptic password system, we next find the upper bound on a possible number of unique haptic passwords, and later on, the lower bound on the number of unique users that can be authenticated using such a haptic systems. In doing so, we observe that such bounds depend on several parameters of a haptic password system:

(P1) The used haptic device,

(P2) The considered measured parameters, such as position, and applied torques and forces, and

(P3) The duration of an authentication sequence (i.e., the length of available authentication data).

For example, in our experimental analysis, we used the SensAble PHANToM Omni [154] haptic device, and at every time point, we collected information about the position of the tip of a haptic device in virtual environment $(x, y, x)$, applied forces $f_x$ and $f_z$ and orientation of the haptic device $(q_{\text{pitch}}, q_{\text{roll}}, q_{\mathbf{yaw}})$, resulting in an eight-dimensional state vector $v$

$$v := [x, y, z, f_x, f_z, q_{\text{pitch}}, q_{\text{roll}}, q_{\text{yaw}}] \tag{9.12}$$

Moreover, every authentication sequence in our experimental trial consisted of 128 8-dimensional data points.

We next make the following simplifying assumptions:

(A1) Measured parameters are mutually independent,

(A2) For every parameter, every value within the possible range is equally likely to be achieved, and

(A3) Recorded 8D data points are all mutually independent,

and after consulting the haptic device specifications, we first find the possible set size for every measured dimension as:

$$|v| := [|x|, |y|, |z|, |f_x|, |f_z|, |q_{\text{pitch}}|, |q_{\text{roll}}|, |q_{\text{yaw}}|] \tag{9.13}$$

Relying on assumptions (A1) and (A2), we can find the possible set size of a single data point within an authentication sequence to be equal to:

$$|d| := \prod_{i=1}^{8} |v|_i = |x| \cdot |y| \cdot |z| \cdot |f_x| \cdot |f_z| \cdot |q_{\text{pitch}}| \cdot |q_{\text{roll}}| \cdot |q_{\mathbf{yaw}}| \tag{9.14}$$

Relying now on assumption (A3), we find the space of possible unique haptic passwords by asking how many different 8D sequences of length 128 can we drawn from a set of of possible single data points within an authentication sequence. This question is equivalent to asking how many 128-combinations with repetition are there within the set of possible single data points [138]:

$$|HP| := \left(\binom{|d|}{128}\right) = \binom{|d| + 127}{128} \tag{9.15}$$

where $|d|$ denotes the set size of the set of single data points within an authentication sequence, defined by equation (9.14), and $|HP|$ the size of the set of unique haptic passwords.

We next recall the fact that users of a haptic passwords system may (and will) exhibit variability in the way they interact with a haptic device on a same authentication task. That variability may result in different authentication sequences from the possible set of haptic passwords $HP$ corresponding to the same user. Yet, those slightly varying authentication sequences should still be attributed to the same user. Because of that, the size of the haptic set $|HP|$ represents an over-estimate of the possible unique users that can successfully use our proposed haptic system. In order to take this variability into account, we propose to find an 8-dimensional hypersphere, defining a single user's variability. The volume of that hypersphere can be computed by finding the maximum variances of every measured parameters. Finally, in order to find the number of unique users that can successfully use the proposed haptic systems, we find how many non-overlapping hypersphere, defining a single user's variability, can be fitted within the set of possible haptic passwords.

### *9.3 Operator Signatures*

In teleoperated robotic systems, an operator typically has a good understanding of a robot's expected "behavior" (dynamics and actions under normal operating conditions). A remote robot responds to an operator's commands with predictable behavior due to an underlying *mathematical model* of a robot, known both to the robot and to the operator. This mathematical model can exist in a variety of forms, and it is often used to represent and analyze the robot's dynamics and actions. More elaborate models may also take into account the dynamics of the communication network, in order to better predict delays, anomalies or communication failures.

The same depth of understanding of an operators' behavior is, however, almost never available. Operators are typically assumed to be trained, skilled and authorized to control a robot, and it is often expected that they will execute a remote procedure at their highest level of performance and attention. These assumptions may not hold valid due to a variety of reasons, many of which are out of operators' control. At the moment, however, neither a robot, nor any other part of a teleoperated robotic system has a systematic way to analyze

and/or validate operator's actions.

In order to enhance a teleoperated system's ability to monitor and validate an operator's actions, we observe that every operator has a unique way of communicating with, and controlling a remote robot. We thus propose to record features of an operator's interaction with a remote robot, and to use this information to learn parameters of a mathematical model representing an operator's actions during a remote procedure, which we refer to as *operator signature*. Once such a model is available, it can be used in hard real-time (online), in soft real-time, or offline to help with detection of possible discrepancies between an operator's expected and his/her exhibited behavior.

Access to a unique operator signature during a teleoperation procedure is expected help with three tasks:

(V1) Continuous identification and authentication of an operator (**evaluation**)

(V2) Real time monitoring and validation of an operator's and robot's actions, and (**validation**)

(V3) Enhanced operator training (**training).**

Additionally, being able to perform tasks (V1)–(V3) will allow for:

(W1) Easier detection of benign anomalies on the operator side of the system,

(W2) Enhanced security of teleoperated robotic systems, and

(W3) Improved way teleoperation control skills are being taught, trained and evaluated.

### *9.3.1 Background on Operator Signatures in Teleoperated Robotic Surgery*

The idea to record forces and torques that surgeons apply on surgical console during a robotic surgery, and to combine these data with robotic tool/tissue interaction data, collected on a robot's side, is not new. A number of authors have shown that such data can been used to assess the level of surgical skill, and distinguish between novice and expert surgeons [200, 195, 201, 199, 202, 197]. These authors defined 14 types of tool/tissue interactions and associated each interaction type with a unique surgeon's force/torque signature. Using the experimental data from 10 surgeons who performed laparoscopic cholecystomy, the authors trained a Hidden Markov Model (HMM) for each subject and each step of the

procedure. The obtained HMMs where used to analyze discrepancies between expert and novice surgeons, and a statistically significant difference between two groups of surgeons was observed. Moreover, the authors observed the major differences between skill level were observed in:

- ◇ Force/torque amplitudes,
- ◇ Types of tool/tissue interactions used, and transitions between them,
- ◇ The time spent in each tool/tissue interactions, and
- ◇ The overall procedure time.

### *9.3.2 Technical Details of Generation and Use of Operator Signatures in Robotic Surgery*

The goal of operator signatures is to gain better understanding of an operator's behavior and actions. In other words, we want to identify those operators' traits and features that will allow us to determine an operator's behavior and current state. We anticipate the following data will be useful in doing so:

(DT1) Position and velocities of operators' instruments, such as haptic devices,

(DT2) Forces and torques that operators apply to the instruments,

(DT3) Positions and velocities of robots' end effectors, and

(DT4) Messages exchanged between operators and robots.

When thinking about a teleoperated robotic surgical procedure, however, there are several components that may introduce variability into an operator signature. Those can broadly be grouped into:

(X1) Features specific to the medical condition/disease being treated.

(X2) Features defining a patient's state, such as the patient's age, gender, weight, height, blood pressure, overall well-being, the severity of the treated medical conditions, as well as other medical indications that may affect the conducted procedure.

(X3) Features defining a robot's and network's states.

(X4) Features defining a surgeon's skill level, and current mental capacity.

In a specific case it may not be possible to mutually de-correlate features (X1)–(X4). Regardless of a possible correlation, the observed features will be used to learn/infer parameters of a mathematical model, representing an operator's unique operator signature. In doing so, we recognize the following step as the minimum necessary set of steps:

(ST1) Choose a mathematical model of an operator.

(ST2) At the beginning of a teleoperated robotic procedure, record the first batch of measurable data on both sides of the system, and messages exchanged between an operator and a robot.

(ST3) Use the recorded data to identify (learn, infer) parameters of the mathematical model chosen in Step ST1.

(ST4) Based on the identified parameters of the model, determine an operator's identity.

(ST5) At each time interval, where the length of time interval is determined based on the type of a procedure:

(ST - i) Predict the expected output of the operator signature. The predicted output will typically consist of the measurable data on the operator side and the messages sent by an operator.

(ST - ii) Record data on the operator's and the robot's side, and the exchanged messages.

(ST - iii) Compare the observed (measured) data with the predicted data.

(ST - iv) If the observed and the measured data align (within the given threshold), declare that the operator is valid.

(ST - v) If there is a discrepancy between the observed and the measured data, announce that there is an anomaly in the remote procedure.

**Choosing an Appropriate Mathematical Model:** There exist a variety of mathematical models that can be used to model operator's actions. Those may include: (i) linear and nonlinear dynamical models, (ii) graphical models, including Markov random fields and Bayesian networks, and (iii) algorithmic models.

**Choosing an Appropriate Set of Measurable Data:** Similar to persons' handwritings, where many people may write a single letter in an identical way, simply looking at a single

letter (in our case, at a single feature) may not be enough to extract an operator's unique signature. Yet, just as an individual's handwriting can be identified given a large enough sample of their writing, an operator's signature will likely also require a sufficiently sized set of recorded data. There is a large number of measurable parameters that may contain information that could be considered a part of an operator's signature:

(Y1) Position, velocity, acceleration and orientation of an operator's tools,

(Y2) Position, velocity, acceleration and orientation of a robot's end effectors,

(Y3) Forces and torques applied by an operator,

(Y4) Forces and torques applied by a robot on the surrounding environment,

(Y5) Time differences between two consecutive control messages,

(Y6) Time differences between two consecutive feedback messages, and

(Y7) The overall procedure time.

**Choosing an Appropriate Model Training Method:** A variety of feature extraction and model training methods can be used to develop an operator's signature, including system identification, statistical and machine learning methods. The appropriate choice of a method will depend upon:

(Q1) The type of a remote procedure,

(Q2) The type of a chosen mathematical model, and

(Q3) The available measured data.

Any chosen method must be reliable, with low false positive and false negative results. In other words, any chosen method should guarantee the following, simplistically depicted in Figure 9.5:

(R1) A single set of measured parameters should never correspond to more than one operator's signature (**injection**).

(R2) A single set of measured parameters should always match to at least one operator's signature (**surjection**).

(R3) A set of measured parameters should always correspond to one and only operator's signature (**bijection**).

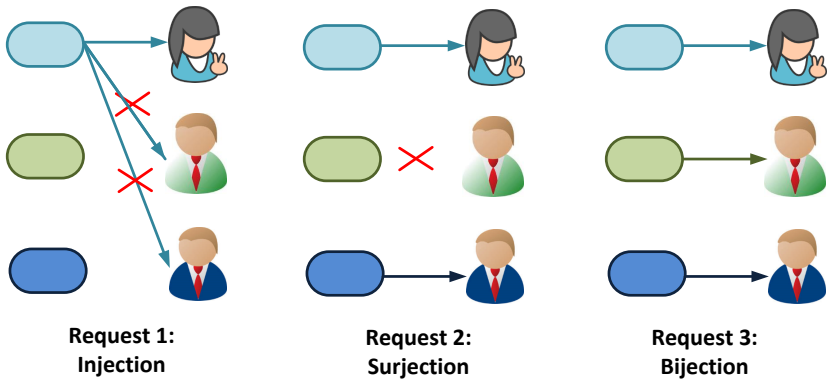



**Figure 9.5:** Requests on operator signatures. A reliable operator signature should be a bijective function (i.e., an injective and a surjective function), where every set of measured parameters always corresponds to one and only one operator's signature.

**Choosing an Appropriate Validation Time Interval:** Considering the proposed operator analysis and validation approach, it is obvious that a shorter validation time interval enables a quicker detection of possible anomalies in a remote procedure. On the other hand, a short time interval requires a faster data collection and data analysis, which may impact validation reliability. There is an inherent trade-off between the length of the validation time interval and the reliability of validation. Choosing an appropriate validation time interval will therefore depend on the type of a remote procedure and the perceived risk that the procedure may get compromised.

**Choosing an Appropriate Validation Technique and Validation Threshold:** Similar to feature extraction and model training methods, choosing an appropriate validation technique will depend on a variety of parameters, such as the type of a remote procedure, the type of the chosen mathematical model, and the available measured data. In addition to the reliability of a chosen validation method, we will also be interested in its computational overhead and efficiency.

### *9.3.3 Experimental Analysis*

As a proof of concept, we present a simple model of operator signatures trained using experimental data collected by Dr. Lee White between September 2010 and January 2012, under Department of Defense Grant W81XWH-09-1-0714: "Virtual Reality Robotic Simulation for Robotic Task Proficiency: A Randomized Prospective Trial of Pre-Operative Warm-up" [229].

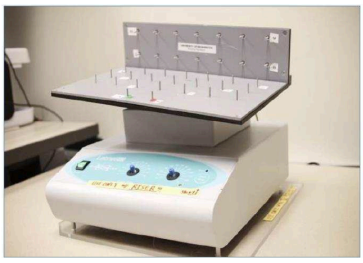


**Figure 9.6:** Rocking pegboard experimental setup. In this task, surgeons are asked to move a pair of elastomeric rings in a prescribed sequence of pegs and tool movements around a pegboard mounted on a chemistry rocker. [Picture credit [229].]

Experimental data was collected from ten surgeons, who performed three series of robotic surgery training task referred to as *rocking pegboard*, depicted in Figure 9.6. In this task, surgeons were asked to move a pair of elastomeric rings in a prescribed sequence of pegs and tool movements around a pegboard mounted on a chemistry rocker undulating at a rate of eight cycles per second [123].

The data was recorded at a frequency of 30Hz, using *SurgTrak*, a custom-developed system for recording surgical performance [229], and it consisted of position and orientation of the left and the right surgical tool, and poses of tool graspers. Using the recorded data, we compute velocities and angular velocities of the tools and the tool grasper. The computed velocities are then used for model training.

**Data Quantization:** Overall, the analyzed data is 14-dimensional, resulting in more than 240000 14-dimensional data points for every surgeon. In order to allow for easier modeling and performance validation, we cluster the recorded data into 70 distinct clusters using Vector Quantization (VQ) method (K-means algorithm) [93]. That allows us to reduce every 14-dimensional recorded data point into a one dimensional data point (cluster index). The problem with this approach, however, is the fact that the conducted dimensionality reduction inherently causes some loss of information. Thus, in order to be sure no significant information about surgeons' performance has been lost, we analyze whether two desirable properties of our experimental data are maintained once surgeons' outputs are clustered:

- **High inter-surgeons variability:** We want to make sure that we have preserved enough information about each surgeon's performance, so that no two surgeons have the same (or sufficiently similar) clustered output.
- **Low inter-trial variability:** We also want to make sure that we have preserved enough information about every surgeon, so that every surgeon's unique and distinct features are maintained over all trials.

The first requirement, high inter-surgeons variability, is equivalent to our earlier requirement (R1), which allows us to distinguish between different surgeons on the same trial. The second requirement, low inter-trial variability, is directly related to extracting every surgeon's unique features, and building his/her unique operator signature. This requirement is related to our earlier requirements (R2) and (R3).

### *9.3.4 Experimental Results*

**Inter-surgeons variability:** The results related to the inter-surgeon variability are presented in Figures 9.7–9.9, depicting the comparison between clustered outputs for surgeons 1-5. Figure 9.7 depicts the first 50 time steps during the first rocking pegboard task, Figure 9.8 the first 150 time steps during the second pegboard task, and Figure 9.9 the first 300 time steps during the third rocking pegboard task.

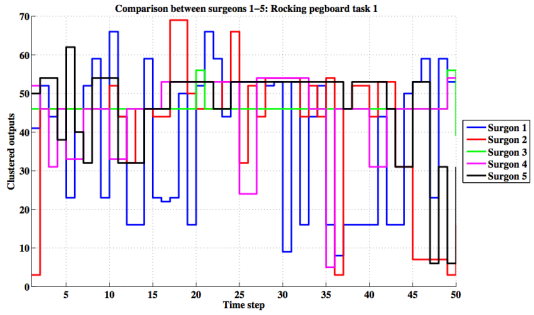



**Figure 9.7:** (a) Comparison between clustered outputs for surgeons 1–5 for the first 50 time steps during the first rocking pegboard task.

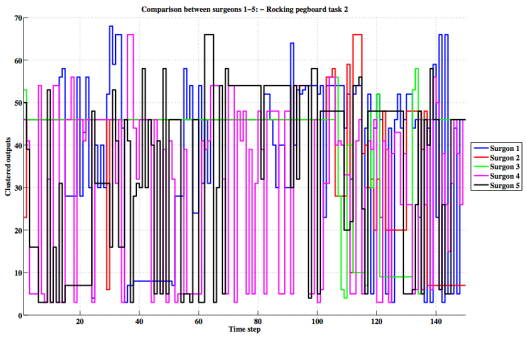



**Figure 9.8:** (b) Comparison between clustered outputs for surgeons 1–5 for the first 150 time steps during the second rocking pegboard task.

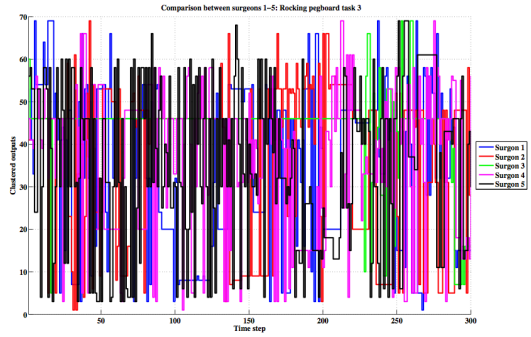



**Figure 9.9:** (c) Comparison between clustered outputs for surgeons 1–5 for the the first 300 time steps during the third rocking pegboard task. From all three figures, we observe each surgeon performs a task in a clearly distinct way.

In all figures, we observe every surgeon has a clearly distinct sequence of actions, over all trials and all three considered time intervals. Similar results are observed when comparing performance of all surgeons over longer time intervals. Thus, we conclude the chosen clustering methods allows us to maintain a high enough inter-surgeons variability.

**Inter-trial variability:** The results related to the inter-trial variability are presented in Figures 9.10–9.13, depicting inter-trial variability for two surgeons, Surgeon 2 and Surgeon 7. Figure 9.10 depicts the first 200 time steps of Surgeon 2's actions, and Figure 9.11 the first 200 time steps of Surgeon 7's actions. Similarly, Figure 9.12 depicts actions of Surgeon 2 over time interval [300-500], and Figure 9.13 actions of Surgeon 7 over the same time interval.

We observe that the variability over different trials (different executions of the same

task) is present for both surgeons. It appears, however, that the variability over different trials for the same surgeon is smaller than the variability between different surgeons. In addition, as depicted in Figures 9.11 and 9.13, the variability between different trials seems to decrease as time progresses.

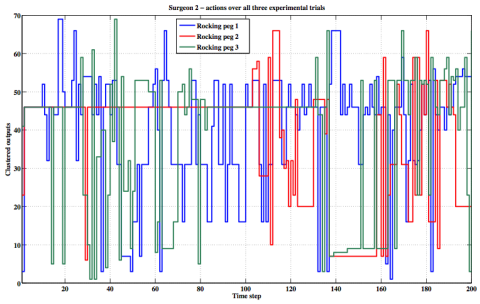



**Figure 9.10:** (a) Inter-trial variability for Surgeons 2 and 7 – 2. The first 200 time steps for Surgeon 2.

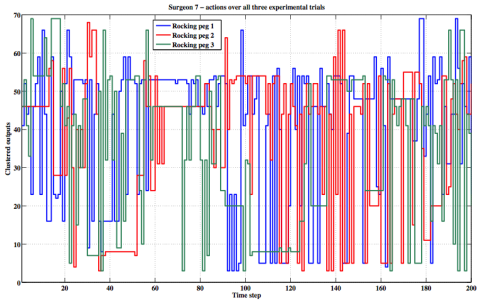



**Figure 9.11:** (b) Inter-trial variability for Surgeons 2 and 7 – 2. The first 200 time steps for Surgeon 7.

## *9.4 Summary*

Teleoperated robotic systems are complex cyber-physical systems, consisting of several components: human operators, controlling the remote manipulator, remote robots, operating in remote environments, and communication networks, connecting human operators and robots. In order to enhance security, privacy and safety of these systems, in this Chapter, we propose a *multi-layered approach to security and privacy*, consisting of at least these four

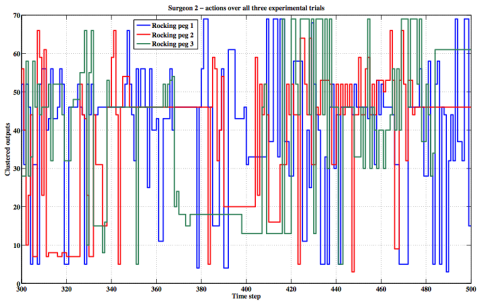



**Figure 9.12:** (c) Inter-trial variability for Surgeons 2 and 7 – 2. Observed actions of Surgeon 2 in time interval 300–500.

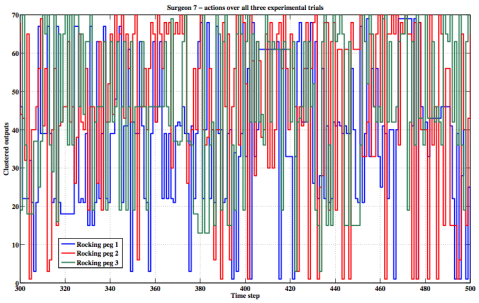



**Figure 9.13:** (d) Inter-trial variability for Surgeons 2 and 7 – 2. Observed actions of Surgeon 7 in time interval 300–500. From the presented figures, we observe that the Surgeon's actions vary across different trials (different executions of the same task), however, the variability across different trials is significantly smaller than the variability between different surgeons.

steps: (1) the implementation of updates, changes and enhancement to software used within the system, (2) the use of available hardware for security enhancement, (3) the leverage of the existing cyber and cyber-physical security methods, and (4) the development of new security techniques specific to teleoperated robotic systems. The goal of every layer (1)–(4) is to provide an additional level of security and privacy to all information exchanged between an operator and a remote manipulator.

We further observe that in many teleoperated robotic systems there may exist tensions between security and privacy, and other desired properties that the system should hold. Those tensions may render many existing security techniques infeasible for teleoperated robotic systems, possibly requiring new security approaches to be developed. Many of the

new methods may rely on the unique component of many teleoperated robotic systems, namely the *human component* of a system.

Human users (operators, patients) have a unique way of interacting with a cyber-physical system, and these users' idiosyncrasies may expose cyber-physical systems to potential security and privacy risks, but at the same time, these unique traits can also be used to increase the system's security, privacy and usability properties. In this Chapter, we propose two new security methods leveraging users' unique interaction with haptic interfaces, haptic passwords and operator signatures.

Haptic passwords are a new biometric method to identify and authenticate operators of a teleoperated robotic system, based on our hypothesis that ways in which users haptically interact with devices provides unique user-dependent features. Similarly, operator signatures leverage the same hypothesis to monitor operators actions in real time.

### *9.5 Acknowledgement*

This work is supported by the National Science Foundation, Grant # CNS-1329751. Any opinions, findings, and conclusions or recommendations expressed in this Chapter do not necessarily reflect the views of the funding agencies.

We thank Nava Aghdasi, Aaron Alva, Jeffrey Herron, Kevin Huang, Thomas Loos, Andrew Lewis, Lee White, Junjie Yan, Tariq Yusuf, and Professors M. Ryan Calo, Howard J. Chizeck, Blake Hannaford, Tadayoshi Kohno and Thomas S. Lendvay for helpful discussion, suggestion, and ideas, as well as for the help with implementation and experimentation with different prevention and mitigation strategies.

**Table 9.3:** Summary of notation from Chapter 9.

| Symbol | Definition |
|---|---|
| $x$ | Mother wavelet function |
| $G(z)$ | Z-transform of the mother wavelet function |
| $H(z)$ | Filter complimentary to the mother wavelet function |
| $F$ | Family of classes of ANN functions |
| $f^*$ | Optimal class of ANN functions with respect to cost $C$ |
| $x, y, z$ | Position of the tip of haptic device in the virtual environment |
| $f_x, f_z$ | Applied forces |
| $q_{pitch}, q_{roll}, q_{jaw}$ | Stylus orientation |
| $v$ | State vector |
| $v_x, v_y, v_z$ | Velocity data |
| $D_i$ | Detail coefficient of level $i$ |
| $f_i$ | Feature vector |
| $w_i$ | $w_i := [\max(D_i), \min(D_i), \text{mean}(D_i), \text{std}(D_i)]$ |
| $PV$ | Password variation |
| $N$ | Number of experimental trials |
| $F_{ij}$ | Feature vector of user $i$ in trial $j$ |
| $\bar{F}_i$ | Mean feature vector of user $i$ |
| $\|v\|$ | Set size for every measured dimension |
| $\|d\|$ | Set size of 8-dimensional single data points within an authentication sequence |
| $\|HP\|$ | Set size of possible unique haptic password |

Chapter 10

# BROADER IMPACT OF THE USE OF OPERATOR SIGNATURES

Based on the assumption that each operator interacts with a remote robot in a unique way, thus generating a unique biometric, in Chapter 9, we introduce the concept of operator signatures as a method to monitor an operator's and a robot's actions in real time. We anticipate that operator signatures will be useful for three main purposes:

(A) Continuous identification and authentication of an operator,

(B) Real time monitoring and validation of an operator's and robot's actions, and

(C) Enhanced operator training.

The use of operator signatures for identification, authentication and real-time monitoring is expected to be especially helpful against a number of cyber security threats, which may be mounted against teleoperated systems. In addition, operator signatures provide an efficient and reliable logging, forensic and training method. The properties that make operator signatures a viable tool for securing systems against cyber-security threats can also become a strong evidentiary tool with legal implications in the realms of liability and medical malpractice.

Yet, the use of operator signatures-based methods introduces a new set of questions that are not currently present in teleoperated robotics. We recognize the following as important questions which will have to be addressed before signature based-methods become more ubiquitous:

1. What benefits do operator signatures-based methods bring to teleoperated robotics, and what is their anticipated legal applicability?
2. What potential issues may arise from the use of operator signatures-based methods in teleoperated robotics?

3. Once operator signatures-based methods detect a failure or a malicious activity within the system, what is the set of meaningful next steps that the robotic system should take?

In the rest of this chapter, we tackle these questions in the context of medical/surgical robotics. We put a specific emphasis on the first question.

### *10.1 Brief Overview of the Recent Legal Cases in Robotic Surgery*

There is currently only one FDA approved minimally invasive surgical robotic system on the market in the US, the *da Vinci*, manufactured by Intuitive Surgical [6]. In recent years, Intuitive Surgical, surgeons, and hospitals have been parties in a number of legal actions. Thus far, Intuitive Surgical has been involved in: strict product liability [94], strict malfunction liability [227], negligence [11], breach of warranty [94], misrepresentation [41], medical malpractice [41], and Fraud Claims Act legal actions [24].

In two legal cases where Intuitive Surgical was a party, the plaintiff patient alleged Intuitive Surgical was liable as manufacturer when the robot was not used [41]. For example, in *Mracek v. Bryn Mawr Hospital* [41], a patient underwent prostate surgery that was intended to be conducted using a robot. But the "robot malfunctioned during the surgery and displayed 'error' messages" . Instead, the surgeon manually used laparoscopic equipment, and the patient later suffered a gross hematuria. The patient brought strict product liability, strict malfunction liability, negligence, and breach of warranty actions against Intuitive Surgical and the hospital. In an unprecedential opinion, the patient's argument was rejected, and it was concluded that there was "no record evidence that would permit a jury to infer [patient's injuries] were caused by the robot's alleged malfunction" [41].

Several claims related to robotic surgeons discussed the lack of training Intuitive Surgical or the hospital have provided to the operators [41]. For example, in *Silvestrini v. Intuitive Surgical, Inc.* [41], the plaintiff alleged that Intuitive Surgical was "responsible for training [hospital] staff members to use the surgical robot and that such training was 'totally lacking or woefully inept or inadequate' ". In the same case, the plaintiff also alleged that the hospital was responsible for "training its staff to use the robot".

The conclusion that can be drawn from the existing robotic surgery cases points that there is currently no consistent legal response to when something goes wrong during a robotic surgical procedure. Most of the cases thus far can be broadly grouped into those seeking legal recourse against the manufacturer, and those focusing on the sufficiency of training or certification of an operator.

## *10.2 Benefits and Legal Applicability of Operator Signatures-Based Methods in Teleoperated Surgery*

In their "2015 Top 10 Health Technology Hazards" report [7], The ECRI Institute listed both robotic surgery and cyber security. Robotic surgery was listed as a threat because of "complications due to insufficient training", and cyber security because of "insufficient protections for medical devices and systems". Operator signatures-based technologies have the potential to address both of these hazards, and to offer a new prospective into an operator's actions during a teleoperated surgery. Operator signatures can be seen not only as a new source of information about a remote procedure, but also as one of the first forms of empirical evidence for an operator's actions during a remote surgical procedure. Having this information available is, in turns, expected to facilitate and lead to:

- A growth in medical research on proper techniques,
- A clarified standard of care for sufficient surgical training,
- A clearer liability delineation, and
- Increased accountability resolution.

### *10.2.1 Growth in Medical Research on Proper Techniques*

Since their emergence in the mid-1970's, *medical malpractice* claims have been rapidly increasing in numbers and cost, and the threat of medical malpractice claims has become an integral past of the US health care costs. In 2010, the national medical malpractice cost was estimated at $55.6 billion per year, of which $17-$29 billion per year was estimated to be from preventable medical injuries [162].

One of the purposes of medical malpractice claims is to *improve* patient safety by compensating injured patients, and more practically, by *linking liable conduct to an established*

*standard of care*. This, in other words means, that in order for a plaintiff to prevail in a medical malpractice case, there must be some determination of a physician's proper standard of care. Currently, however, there is no single codification of a standard of care, and to determine the standard of care in a medical malpractice action, parties tend to look at a number of sources, including:

- Expert testimony by a qualified expert,
- Guidelines adopted by specialized medical societies and public institutions,
- Applicable case laws, and
- Applicable regulations.

Since robotic surgery can still be considered a new medical practice, in many cases, there is not yet an established standard of care. For instance, it has been noted that there currently does not exist a standard of care for patient selection for robotic surgery [189]. One way in which operator signatures-based methods are expected to help resolving this issue is by providing empirical data that will facilitate the development of new evaluation methods for robotic surgery. These evaluation methods will, in turn, facilitate the development of new guidelines, as well as new training techniques for robotic surgery.

### *10.2.2 Clarified Standard of Care for Sufficient Surgical Training*

For a long time, medical research has been very influential in determining an appropriate standard of care for particular medical situation, see for example [161]. Along these lines, operator signatures-based methods have a potential to aid and enable medical research in a few directions. First, operator signatures can be used to develop a new knowledge base for supporting *evidence-based medicine* practices. Evidence-based medicine (EBM) is defined as "the judicious use of the best current evidence about the care of the individual patient" [204], and it integrates the best external evidence with individual clinical expertise and patients' choices.

Operator signatures can aid EBM by providing a means to address "important technical aspects of surgery...that are currently inadequately studied" [134]. For EBM to be useful, however, there must exist a knowledge base that provides surgeons with external studies

they can use to evaluate their own practices. Current studies on technical surgical aspects are often burdensome because they require observers to review video of procedure [38]. Operator signatures-based methods have potential to make such studies significantly less burdensome and more reliable. Operator signatures can also aid EBM by providing surgeons with their own 'playback' of robotic surgical procedures they executed. This provides an ability for direct comparison of surgeons' actions with themselves, as well as with other surgeons in the field.

In addition, operator signatures can potentially be used to monitor and control for differences in skill in surgical robotics research. Authors of [77] recently identified the need to control for a surgeon's expertise in randomized control trials. They contend that traditional randomized control trials for surgical interventions suffer from "expertise bias" because they typically do not account for the individual surgeon's expertise. This "expertise bias" may skew results toward less technically challenging procedures.

Finally, the availability of operator signatures can enable research as to the validity of operator signatures themselves. This type of validation research will be necessary to support the use of operator signatures in legal cases, and it could aid in standard of care evaluations in medical malpractice cases.

### *10.2.3 Clearer Liability Delineation*

Operator signatures-based method will make available data that help is distinguishing between a potential robotic malfunction and a surgeon's error. This has a potential to help with distributing liability to where it may be warranted, and with increasing accountability for all involved parties. In many cases "the most difficult burden the plaintiff has in a case against a health care provider is proving that a breach of the standard of care in fact caused the plaintiff's injuries" [156]. The plaintiff must prove a number of aspects that are typically in a defendant's favor. These aspects include demonstrating causation between a physician's departure from the standard of care and the injury suffered, even though "important scientific issues of causation are unsettled" [172].

By providing data about a surgeon's actions, operator signatures make the question

whether a surgeon departed from an established standard of care significantly easier to answer. In these evaluative instances, operator signatures provide objective evidence of the procedure in question. They can demonstrate that a surgeon:

- Acted consistently with his/her past procedures of the same nature,
- Deviated from his/her previous actions, but in ways within the asserted standard of care, or
- Deviated substantially from his/her previous actions.

Additionally, the authorization properties of operator signatures could be used to track surgeons' states, such as if they are or are not fatigued.

Further, operator signatures may be useful to demonstrate that the alleged source of the plaintiff's harm was not a surgeon. For instance, the surgeon may have proceeded in ways that were consistent with his/her past acceptable operators. Yet, the plaintiff was still injured. Here, the surgeon would be able to show that his/her actions were within the standard of care, and that some other action, possibly a robotic failure, was the source of the harm. Or, the surgeon may have proceeded in ways completely inconsistent with the alleged surgeon. There, the *identification* property of operator signatures may be used to demonstrate that it was an individual other than the alleged operator who caused harm.

Finally, just as airplane black boxes provide investigators with a record of actions taken, operator signatures may also be useful as evidence of what occurred during robotic surgery. With an actual record in place, liability between actors may be more clearly delineated.

### *10.2.4 Increased Accountability Resolution*

**Robotic Failure:** To demonstrate how operator signatures-based methods increase the accountability resolution in robotic surgery, we start by considering a scenario where a surgical robot mechanically fails, but the failure is only partial, and it still allows a surgeon to continue the procedure. The surgeon is aware of the partial malfunction, and using the available knowledge, decides to take additional actions to mitigate the partial malfunction. Even with the operator's mitigating actions, the patient suffers serious injury.

In this robotic failure scenario, operator signatures may be used to delineate whether or not the surgeon had knowledge of the robotic mechanical failure. This delineation between whether or not the surgeon knew about a defect matters for products liability and malpractice. If the surgeon knew about the mechanical defect, then the manufacturer could argue that it is not liable because of the *learned-intermediary doctrine*, which immunizes a manufacturer from liability when the manufacturer provides adequate warnings to the sophisticated user [159].

Whether or not the surgeon, operating the robot, can be considered a learned intermediary is, however, an open question, since the scope and the applicability of the doctrine is not completely defined [53], and some researchers (e.g., [159]) have suggested that because surgeons cannot be presumed to have expertise in engineering, in a products liability action, whether a surgeon is a learned intermediary will require an adequate foundation.

**Fatigued Surgeon:** Let's now consider the case where a surgeon performs a remote procedure while dramatically fatigued. In this case, the robotic surgical system worked with no malfunctions. During the surgery, however, the surgeon committed an error, which resulted in serious injury to the patient.

Medical personnel's fatigue is a clearly identified issue. For example, in study [158], it was found that residents' fatigue was "prevalent, pervasive, and variable and accounted for an increased risk of medical error". Thus, detections of a surgeon's own state may be useful to determining hospital policies for fatigue and other personal impairments. For example, the detection of measurable fatigue may indicate that a surgeon is unfit to perform the operation. From an institutional level, the hospital may be able to use operator signatures to develop scheduling and policies on fatigue. Such procedures may help limit the hospital's liability, and more importantly, reduce fatigue.

**Compromised Communication Link:** We lastly consider the case where the communication link between a surgeon and a robot is compromised. An unknown attacker compromises the link and takes control of the robot, thus becoming able to take actions that can seriously injures the patient. In this case, by logging operator signatures on both sides of the network, a mismatch can be identified after the incident occurred by using the identification properties of operator signatures. This mismatch in the commands being sent by

the surgeon and the commands received by the robot could be used to decisively show that the surgeon was not in control of the robot at the time of the harm.

Once established that it was not the surgeon who caused injuries to the patient, the questions of liability would shift to other factors, such as:

(a) What vulnerability did the intervening actor use?

(b) Who was responsible for maintenance of system or network where the vulnerability was exploited?

(c) Is there contractual language involved in assigning liability for cyber security incidents?

(d) Does the FDA provide guidance or requirements for cyber security risks?

(e) Is there relevant state or Federal law that discusses liability for cyber security incidents?

### *10.3 Potential Issues Arising from the Use of Operator Signatures-Based Method in the Teleoperated Surgery*

The use of operator signatures-based methods in robotic surgery increases a potential for several issues which are not present in the current robotic surgery. We recognize *potential privacy implications on surgeons* as a most important question. Although practices of monitoring operators' actions already exist in commercial aerial and ground transportation industries, see for example [103, 33, 96], operator signatures-based methods may be seen as privacy intrusive, and may have negative impacts on surgeons' actions.

These methods are anticipated to be useful in identification and authentication of surgeons, as well as for real time monitoring and validation of surgeon's actions throughout the remote procedure, thus providing data about a surgeon's:

(D1) Expertise and skill level,

(D2) Performance during the ongoing procedure,

(D3) Potential (both harmless and harmful) deviations from the surgeon's earlier executions of a similar procedure,

(D4) Potential deviations from other (equally skilled) surgeons' executions of a similar procedure,

(D5) Potential omissions, risky actions and errors, and

(D6) Mental state and capacity during the procedure.

This data could be used by multiple parties, for various purposes. For example, the operators signatures data could be used by:

- Surgeons, for self-evaluation and training,
- Patients, as a part of treatment and further diagnosis,
- Other surgeons, for learning and training,
- Manufacturers, to better adapt a robot to surgeons' needs
- Hospitals, for scheduling and evaluation,
- Professional medical societies and federal agencies, for surgeons' evaluation and certification,
- Health insurances, for insurance and billing purposes,
- Courts, as potential proofs against surgeons, hospitals, and manufacturers, and
- Malicious entities, aiming to harm surgeons' reputation and integrity.

From the listed examples, we see that there are many potential benign applications of operator signatures data, but also many that surgeons may have a hard time agreeing with. Moreover, we anticipate that an unregulated access to operator signatures data may make surgeons:

(S1) Less willing to perform robotic surgery, even when doing so is beneficial for a patient.

(S2) Prone to taking more conservative and less risky actions, even when riskier actions are warranted, and would actually be beneficial.

(S3) More nervous and prone to take defensive steps in order to justify their actions.

These negative consequences of the use of operator signatures data clearly indicates that there is a strong need to *regulate who, when (under which circumstances) and for how long should have access to the data*, and *how such data should be used*.

Treating operator signatures data as other biometric or biosignal data, especially as a person's genetic or neural data may provide the first avenue for determination of rules of proper use of this data.

### *10.4 Perspective on the Next Step Once Operator Signatures-Based Alarm Has Been Raised*

Once operator signatures-based mechanism detects an anomaly during a teleoperated robotic procedure, it should react, allowing the system to prevent potential mechanical failures, human errors or malicious activities from causing further damage to the system. There are several potential ways the system could go about that:

(O1) Completely stop the ongoing teleoperated procedure to the nearest safe stopping point,

(O2) Completely stop the ongoing teleoperated procedure, and restart it from some safe starting point,

(O3) Completely stop the ongoing teleoperated procedure, and move to the manual operating mode,'

(O4) Disconnect a robot and an operator, and switch the robot to a (semi-)autonomous operating mode.

In many cases, options O1–O3 may not be viable for teleoperated robotic surgery, leaving the last option as the only available. Yet, the option of switching to a (semi)-autonomous operating mode opens up a variety of technical, as well as legal and regulatory questions. Some important technical challenges in enabling a teleoperated robot to switch to an (semi)-autonomous operating mode include:

(C1) Defining the set of system states and dangerous surgeons' commands that, when detected, allow a robot to block them and switch to autonomous operating mode,

(C2) Defining the allowed set of operations a robot is allowed to perform in the emergency autonomous mode, end

(C3) Defining the rules to communicate system states that a robot will use when starting autonomous operation.

### *10.5 Summary*

Operator signatures-based techniques are expected to enable more efficient and reliable identification, authentication and evaluation of operators of teleoperated robotic systems,

based on the assumption that every operator interacts with, and controls a remote robot in a unique way. Additionally, operator signatures-based method are expected to enable robust and reliable detection of possible anomalies during a teleoperated procedure, as well as to have a wide application in teaching and training phase of different teleoperated procedures.

In this chapter, we address legal, ethical and societal questions that are likely to arise with the use of operator signature-based methods as a part of teleoperated robotic systems. We focus on the following three questions: (i) what are the benefits of using the concept of operator signatures in teleoperated robotic systems, (ii) what potential issues may arise from the use of operator signatures in teleoperated system and (iii) what is the legal and ethical perspective on the expected next steps within a teleoperated robotic system, once signature-based alarm has been raised. We put the specific emphasis on the first question, which we answer in the context of the next generation surgical robotic systems.

We show that operator signatures-based techniques can refine the standard of care for robotic surgical procedures, and that they can be used to better determine the minimum level of training of medical personnel using surgical robots. We further also show that operator signatures-based mechanisms may provide objective empirical evidence of an individual operator's actions during robotic surgery, thus allowing for a clearer delineation of operators' liabilities, and well as a more reliable accountability resolution.

### *10.6 Acknowledgement*

This work is supported by the National Science Foundation, Grant # CNS-1329751. Any opinions, findings, and conclusions or recommendations expressed in this Chapter do not necessarily reflect the views of the funding agencies.

We thank Aaron Alva, Jeffrey Herron, Professors M. Ryan Calo, Howard J. Chizeck, and Margot Kaminski, as well as the participants of the 4th Annual Conference on Robotics, Law and Policy (WeRobot 2015) for the helpful discussion, suggestions, and ideas.

Chapter 11

# RELEVANT OTHER PROJECTS, REMAINING QUESTIONS AND FUTURE WORK

In addition to security and privacy investigations of brain-computer interfaces and tele-operated robotic systems, I was involved with a few other projects: (i) security of wireless sensor networks, (ii) jamming attack analysis and (iii) ethical and legal considerations of emerging closed-loop deep brain stimulators. This chapter gives a brief overview of these projects.

## *11.1 Security of Wireless Sensor Networks (WSNs)*

In this project, our focus was on persistent attacks that can skew network structure or allow an attacker to gain knowledge about internal-states of the network. One such attack is the *node capture and cloning*, where an attacker physically compromises (captures) network sensors (nodes) and extracts the data specific to them. The attacker then uses compromised nodes to mount a variety of secondary attacks such as jamming, false data injection or selective data collusion. In addition, the attacker may fabricate functionally equivalent replicas of captured nodes, referred to as *clones*, and deploy them back into the WSN, in order to gain control over the network or tamper with network operation.

As a part of this project, we proposed and developed a method to mitigate node capture attacks, which does not depend on the number of compromised and cloned nodes present in the network, thus guaranteeing graceful network degradation under attack. This involved: (a) developing a mathematical model of the node capture and cloning attack, (b) analyzing the existing clone detection methods and developing a new optimization framework to facilitate efficient detection, and (c) modeling and analyzing interaction between the network manager and a persistent attacker in a game-theoretic setup.

### *11.1.1 Threat Model of Node Capture and Cloning Attack*

To develop a mathematical model of the node capture and cloning attack, we mapped the given network security problem into a control-theoretic problem. We then showed that the actions of an attacker and the corresponding network response can be modeled as the following dynamical model [43]:

$$\begin{aligned} \dot{x} &= \lambda[M - x(t)] - \mu x(t) + \omega_n \\ y(t) &= \gamma x(t) + \omega_\alpha(t) \end{aligned} \tag{11.1}$$

where $x(t)$ denotes the number of compromised network nodes, $M$ the number of nodes an attacker needs to compromise to get an access to all cryptographic secrets, $\lambda$ the average rate at which an attacker captures nodes, $\omega_n$ the number of deployed clones, $\omega_\alpha$ the number of valid nodes whose all cryptographic secrets are compromised, $\gamma$ the detection rate of clones and $\mu$ the minimum revocation rate required to guarantee secure connectivity of all valid nodes.

Using dynamical model (11.1), we then mapped the problem of network response into the optimal control problem:

$$\bar{u}(t) = -\mu x(t) + u_{ff} \tag{11.2}$$

where $\bar{u}(t)$ represents the total number of nodes to be revoked from the network at time $t$, $-\mu x(t)$ the number of detected nodes that should be revoked, and $u_{ff}$ the number of nodes that do not have to be revoked, but whose cryptographic secrets should be refreshed at time $t$. We further showed that the revocation rate $\mu$ can be found as the solution (optimal control gain) of the following Linear Quadratic Regulator (LQR) problem:

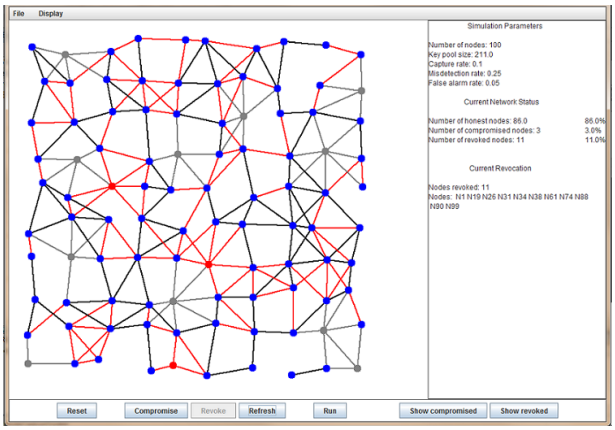



**Figure 11.1:** A snapshot of the developed simulator, used for analysis of network's dynamics, stability and performance under the node capture and cloning attack. Blue vertices represent valid network nodes, red compromised and cloned nodes, and grey revoked nodes. Black edges represent valid, protected communication links, red compromised links, and grey links whose cryptographic secrets are revoked.

$$J = \int_0^\infty [x(\tau)^T Q x(\tau) + u(\tau)^T R u(\tau)] d\tau \tag{11.3}$$

where cost components $Q$ and $R$ can be completely characterized in terms of network parameters. Details of network parameters mapping and costs $Q$ and $R$ can be found in [43, 44]. Using the developed control theoretic framework (11.1)–(11.3), we developed a simple network simulator, which allows one to analyze dynamics, stability and operational point of a WSN under node capture and cloning attack, and to jointly evaluate the performance of applied monitoring, detection and revocation methods. A snapshot of the simulator is depicted in Figure 11.1.

### *11.1.2 Convex Optimization Framework for Clone Detection in WSNs*

As a part of this project, we analyzed the existing methods for detection of cloned and compromised nodes. Based on the conducted analysis, we observed that WSNs experience three negative effects under the attack. First, the undetected malicious nodes may significantly affect WSN's operation. Second, all clone detection methods impose an additional communication and storage overhead on the network. Finally, in some clone detection methods, a subset of valid nodes may falsely be identified as cloned and revoked. As a result, these false revocations may significantly degrade network's performance.

Based on this observation, we developed an analytical (convex optimization) framework for evaluation of clone detection methods, which allows quantitative analysis of the impact of the attack and the corresponding network's response. Moreover, the developed framework provides a systematic way for a network manager to choose the appropriate parameters of a detection method, such that the WSN performance is optimal with respect to identified negative effects. The developed framework considers four costs experienced by the targeted WSN, which we defined as [49]:

1. *Communication Cost* $C_1$ – the average number of bits a node has to transmit during one iteration of detection. This cost is proportional to the average number of transmissions required by each node during detection.
2. *Storage Cost* $C_2$ – the average number of bits a node needs to store during one iteration of detection. This cost is proportional to the number of messages stored by an individual node.

3. *Cost of Undetected Compromised Nodes* $C_3$ – the impact a set of compromised nodes has on the performance of the WSN. This cost is defined as the total number of bits injected or corrupted by all compromised nodes.
4. *Cost of False Revocation* $C_4$ – the reduction in the available bandwidth (in bits), due to the fact that a node falsely identified as compromised is revoked and cannot transmit messages any more. This cost is applicable to those detection methods that allow false alarms, and it is proportional to the number of falsely identified nodes per one iteration of clone detection.

Using costs $C_1 - C_4$, we showed that an inherent tradeoff between an additional communication and storage overhead imposed by the detection methods, and the effectiveness of that method can be characterized by minimizing the convex combination of costs, leading to the following optimization problem [49, 48]:

$$\min C = \beta_1 C_1 + \beta_2 C_2 + \beta_3 C_3 + \beta_4 C_4 \tag{11.4}$$

where $\beta_i, \ \ i \in [1, 4]$ represent the weights given to the specific costs. The weights are nonnegative real numbers, satisfying $\sum_{i=1}^{4} \beta_i = 1$. Using optimization framework (11.4), we analyzed several clone detection methods, and for each of them, proposed an efficient optimal parameter selection algorithm. Detailed derivation of parameter selection algorithms can be found in [49].

### *11.1.3 Game-Theoretic Framework for Node Capture and Cloning Attacks*

As a final step in designing efficient WSNs, secure against the node capture and cloning attacks, we observed that there are two aspects to the attack, namely: (1) the impact of the attack on the network performance, and (2) the interaction between the attacker and the network manager over the lifespan of the network. To incorporate both of these aspects, we showed that the node capture and cloning attacks can be analyzed in a game-theoretic framework [42]. We modeled the attacker-network manager interaction as simultaneous, noncooperative, two-player games, where an attacker, choosing the appropriate capture strategy and capture rate, represents one player, and the network manager, choosing the appropriate detection and revocation methods and corresponding rates, the other player.

We considered two models of the attack: a deterministic dynamical model (11.1), and a more general stochastic (Markovian) model [42]. The Markovian model represents an attacker whose capture actions appear random to the network manager. For such an attacker, the capture process can be modeled as a Poisson process $X(t)$ with capture rate $\lambda$, where $X(t)$ represents the number of captured nodes present in the network at time $t$ [42]:

$$\mathbb{P}[X(t) = k] = e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \tag{11.5}$$

Similarly, we showed that for every node, revocation time $r_i$ can be assumed to have exponential distribution with the revocation rate equal to the probability of detection, $\mathbb{P}_d$:

$$\mathbb{P}[r_i \leq T] = 1 - e^{-\mathbb{P}_d T},\ T \geq 0 \tag{11.6}$$

We then showed that taken together, assumptions about the capture and revocation processes imply that the node capture and cloning attack on a WSN can be modeled as an M/M/N/N queue, where the capture process maps into the arrival process, the number of captured nodes into the number of customers in the system, and the revocation process into the service process.

For deterministic model (11.1), we developed three games, all of which have quadratic utility for the network, whereas the attacker's utility depends on the assumptions about his abilities. For stochastic model (11.5), (11.6), we developed a game with convex utility functions. For each game, we proved the existence of a pure strategy Nash Equilibrium and presented an efficient way of solving the game. Detailed derivation can be found in [42].

### *11.2 Analysis of Control Channel Jamming Attack on WSNs*

In addition to node capture and cloning attack, we studied the *control channel jamming attack* against wireless networks. In wireless networks, control channels are essential for timely dissemination of network management information. Thus, by simply jamming management messages, an attacker can efficiently disrupt network services. The problem of preventing such an attacker becomes increasingly challenging in the presence of *compromised insiders*. We formulated the control channel jamming attack as a control-theoretic problem, and developed a linearized dynamical model of a wireless network under this attack. Using

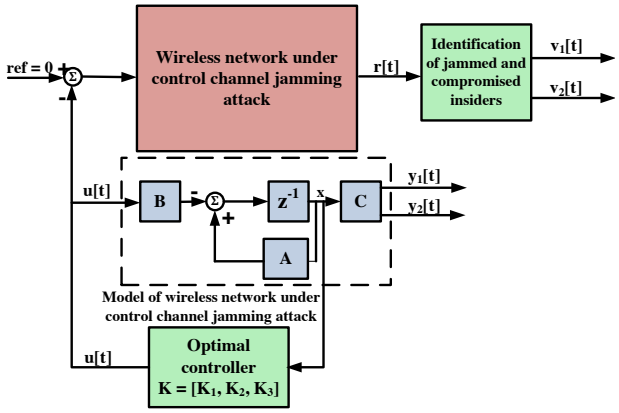



**Figure 11.2:** High-level block diagram of a wireless network under control channel jamming attack: $u[t]$ represents the number of users to be removed from the network, $r[t]$ the users feedback information, $v_1[t]$ the measured probability that users are jammed, $v_2[t]$ the measured probability that users are compromised, $y_1[t]$ the number of jammed users and $y_2[t]$ the number of compromised insiders.

linear quadratic optimal control theory, we developed a method for a network to efficiently respond to the attack. We also showed how feedback information about jammed frequency channels and the current positions of jammed network users can be used to estimate the set of compromised insiders. A high-level block diagram of a wireless network under the control channel jamming attack is depicted in Figure 11.2.

### *11.3 Legal Considerations of Emerging Closed-loop Deep Brain Stimulators*

Deep brain stimulator (DBS), also known as "brain pacemaker" is an accepted and clinically effective form of neuroprosthetic treatment, with a user basis of more than 100 000 implanted patients. DBS are being userd for a variety of common and debilitating neurological disorders, such as Parkinson's disease and essential tremor [171].

This device typically consists of a control and power unit implanted in a user's chest cavity, connected to a set of trans-cranial electrodes which extend into a user's brain. Current DBS implementations, however, have no sensors, thus lacking the capacity to determine whether a user is currently experiencing pathological symptoms. Such *open loop* systems must be continuously active while a user is conscious in order for the system to grant any therapeutic benefits. This continual stimulation overexposes users to potential side effects of DBS, which may include physical sensations such as tingling, burning, and proprioceptive distortions. Additionaly, some users have also been reporting neuropsychiatric side effects, including cognitive and speech disfunction, impulsivity, and changes in self-image. Con-

stant stimulation also limits the implanted system lifespan, as the DBS control unit must be surgically replaced once its battery is exhausted.

Recent research efforts suggest that more advanced implementations of DBS, where stimulation is selectively delivered depending on the state of the patient, may be viable. Proof of concept for a *closed-loop deep brain stimulator* (CL-DBS), capable of detecting the onset of a user's symptoms via co-implanted sensors, then delivering stimulation proportional to the duration and severity of the pathological brain activity, has recently been presented. Such a device offers several important advantages: the minimization of adverse side effects, the capability to tune implant response to symptom severity, and extended system lifespan [108].

While the engineering challenges surrounding closed-loop DBS are still the subject of ongoing research and development, in this project we investigated several societal, legal and ethical questions related to the potential use of these emerging technologies. The driving force behind this project was the investigation whether giving users volitional control over a CL-DBS system is ethically and legally permissible. Our findings indicate that granting a patient (user) a volitional control over his/her CL-DBS is not only permissible, but is in fact advantageous when compared to the alternative of making the system's operation entirely automatic. From an ethical perspective, volitional control maintains the integrity of the self by allowing the user to view the technology as restoring, preserving, or enhancing ones abilities without the fear of losing control over one's own humanity. This preservation of self-integrity carries into the legal realm, where giving users control of the system keeps responsibility for the consequences of its use in human hands. Our analysis framed these issues within the context of tort liability, since the tort law framework provided numerous points of comparison and commonality with existing technologies. Further, being predicated on the concept of individual responsibility, tort theory allowed us to draw directly on the the philosophical issues at stake.

### *11.4 Remaining Questions and Future Work*

Work presented in this dissertation focuses on security and privacy risks that may arise with *personalized cyber-physical systems*. These are the systems able to adapt to their

users' needs, abilities and preferences by collecting and analyzing data about themselves and their users, in search for unique features and patterns which are then used for system personalization.

Even though, in many cases, no real concerns or threats have been reported yet, many of emerging personalized CPS will likely be used in safety-critical application where possible security and privacy compromises may have severe and long-lasting negative effects. Because of that, as well as the early developmental stage of these technologies, this dissertation advocates that security and privacy for personalized CPS should be considered as a part of the design.

In doing so, this dissertation focuses on two emerging biomedical cyber-physical systems, brain-computer interfaces and teleoperated robotic systems. We start from the human component of the cyber-physical system, and propose that *users' idiosyncrasies, in the way users interact with a system, expose these systems to potential security and privacy risks. At the same time, however, users' unique traits can be used to increase the system's security, privacy and usability properties.*

Major parts of this dissertation focus on: (i) understanding the risk of private information extraction when users' electro-physiological signals are used as information sources (ii) prevention of malicious private information extraction, (iii) understanding cyber security risks against teleoperated robotic systems, and (iv) prevention of cyber attacks against teleoperated robots.

### *11.4.1 Understanding the Risks of Private Information Extraction from Users' Electro-physiological Signals*

In Chapters 3 and 4, we present experimental and theoretical analysis of the feasibility of private information extraction when non-invasive EEG-based BCIs are being used. Experimental analysis is conducted through a series of experiments involving human subjects, where, as a part of the experiments, subjects were presented with a variety of visual stimuli, both conscious and subliminal. To analyze the obtained experimental data, we focus on the Event Related Potential (ERP), neural responses associated with specific sensory, cognitive

and motor events [141], and investigate the feasibility of different ERP components, such as P300, N400, P600 and ERN (Error Related Negativity), for private information extraction.

Possible steps to extend this work include:

- Investigation and evaluation of the feasibility of other EEG components for private information extraction
- Investigation of the feasibility of other electro-physiological signals as sources of private information.
- Investigation of correlation between different electro-physiological and biosignals. More broadly, investigation of possible ways to perform sensor fusion between different electro-physiological and biosignals in order to improve and enhance biomedical and bio-engineered systems.
- Experimental and theoretical investigation of novel methods of presenting stimuli to users in order to efficiently and reliably elicit their responses.

### *11.4.2 Prevention of Malicious Private Information Extraction*

Based on our hypothesis that electroencephalograms can be decomposed in real time into characteristic components, which provide sufficient information about a user's conscious and intended messages, in Chapter 5, we propose an approach to prevent the identified privacy threats, referred to as the BCI Anonymizer. It is an interface between the BCI sensors and BCI applications, taking raw electro-physiological signals as inputs, and decomposing them into a collection of characteristic components.

Possible next steps to extend this work include:

- Making algorithms used for filtering of different components of EEG robust and fast enough, in order to make the BCI Anonymizer a real time BCI component.
- Implementing the BCI Anonymizer in hardware, as an actual part of a BCI, and evaluating its performance.
- Investigation of the BCI Anonymizer's applicability to other bio-amplifiers.
- Investigation of other possible methods to prevent private information extraction in BCIs.

### *11.4.3 Understanding Cyber Security Risks Against Teleoperated Robots*

In Chapters 7 and 8, we identify possible cyber security attacks against teleoperated robotic systems, and classify them based on the impact they have on a human operator. For each of these classes, we assess the level of the actual impact of an attack on a teleoperated procedure through a series of experiments involving human participants. Our experiments are based on established robotic surgery tasks, and we quantify the impact using Fitts' law. We show that an attack does not impact all components of a teleoperated robotic procedure equally. Typically, different attacks have a less prominent impact on the "free movement" component than on the "fine motor (homing)" component of a procedure [50].

Possible next steps to extend this work include:

- Gaining better understanding of subtasks of a teleoperation procedure.
- Theoretical and experimental analysis of the impact of various cyber attacks on different teleoperation subtasks.

### *11.4.4 Prevention of Cyber Attacks Against Teleoperated Robots*

Based on our hypothesis that every teleoperator has a unique way of interacting with a robot and with the environment, referred to as *operator signature*, in Chapter 9 we propose new approaches to identify and authenticate a teleoperator, and to quickly detect potential attacks on teleoperated robotic surgery [61]. These approaches leverage the available information about a surgeon's haptic device, a robot's end effectors, and the exchanged messages, in order to detect malicious activities, but also any potential robot's and surgeon's anomalies during a procedure.

Possible next steps to extend this work include:

- Investigation of haptic passwords, and their applicability to other "touch" based devices, such as tablets, smart phones, and touch screen terminals.
- Implementation and experimental evaluation of operator signatures-based monitoring and detection system, in terms of its speed, accuracy and robustness.
- Investigation and development of protocols and mechanisms to switch to safe and secure operating mode once an anomaly or malicious activity has been detected. These

mechanisms may involve switching to a temporary autonomous operating mode.

# Chapter 12

# CONCLUSION

People are increasingly embracing the trend of technology personalization, where different systems that they use in everyday life adapt to their needs, abilities and preferences. This adaptation comes as a result of a system's ability to collect data about itself and its users, and analyze it in search for unique features and patterns which are then used for system personalization. Examples of personalized cyber-physical technologies are many, with many more to come, from buildings learning about inhabitants' daily routines and preferences [13], to music, video and shopping recommendation systems [19, 14, 1].

The personalization trend is expected to be particularly important for biomedical cyber-physical systems, where data about patients, and/or medical practitioners is expected to allow systems to better adapt to medical needs. Yet, this trend is not without risks. Any time data about users and systems is recorded, processed, and possibly stored for future analysis, security and privacy risks arise. Misusing the collected data gives rise to threats ranging from compromising or breaking systems to shaming, manipulating or physically harming users.

This dissertation focuses on security and privacy risks that have recently arisen, or may arise with personalized cyber-physical technologies. We approach the issue by focusing on the human component of a cyber-physical system, and propose that users' idiosyncrasies, in the way users interact with a system, expose these systems to potential security and privacy risks. At the same time, however, users' unique traits can be used to increase the system's security, privacy and usability properties.

Drawing from the stated hypothesis, this dissertation focuses on three questions: (1) how do (how could) biomedical cyber-physical systems use users' idiosyncrasies, (2) what security and privacy vulnerabilities may arise due to the use of users' unique traits, and (3) how can users' idiosyncrasies be leveraged to increase systems' security and privacy?

These question are answered in the context of two emerging biomedical technologies, brain-computer interfaces and teleoperated robotic systems.

In Chapters 2 – 6, potential privacy risks related to non-invasive brain-computer interfaces are investigated and addressed. We present experimental and theoretical results of the feasibility of extraction of private information using BCIs. Experimental analysis is conducted through a series of experiments involving human subjects, where, as a part of the experiments, subjects were presented with a variety of visual stimuli, and their responses to stimuli are analyzed. Based on the hypothesis that EEGs can be decomposed in real time into characteristic components, which provide sufficient information about a user's conscious and intended messages, we then propose an approach to prevent the identified privacy threats, referred to as the BCI Anonymizer.

In Chapters 7 – 10, security risks that may arise with the next generation teleoperated robotic system are identified and addressed. We classify possible attacks based on the impact they have on a human operator, and for each class, assess the level of the actual impact of an attack through a series of experiments involving human participants. Based on our hypothesis that every teleoperator has a unique way of interacting with a robot and with the environment, referred to as operator signature, we then propose a new approach to quickly detect potential attacks on teleoperated robotic surgery [61]. This approach leverages the available information about a surgeon's haptic device, a robot's end effectors, and the exchanged messages, in order to detect malicious activities, but also any potential robot's and surgeon's anomalies during a procedure.

This dissertation enhances the basic knowledge of personalized biomedical cyber-physical systems. It relies on, and at the same time makes fundamental contributions to security and privacy, robotics, and neural engineering research areas. By leveraging the unique properties of cyber-physical systems, it proposes novel mitigation and prevention strategies, based on the uniqueness of a human component within a cyber-physical system. The obtained results are broader than the two investigated applications, and they are expected to be relevant and applicable to a wide range of emerging cyber-physical technologies.

# BIBLIOGRAPHY

[1] Amazon (last accessed: May 7, 2015).

[2] Applied Dexterity (last accessed: May 18, 2015).

[3] Brain Bats: Mind-controlled Pong (last accessed: June 1, 2015).

[4] The center for sensorimotor neural engineering tech sandbox (last accessed: August 7, 2015).

[5] Cyberknife (last accessed: May 18, 2015).

[6] da Vinci Surgery (last accessed: May 18, 2015).

[7] ECRI Institute: 2015 Top 10 Health Technology Hazards (last accessed: May 21, 2015).

[8] Emotiv Systems (last accessed: May 5, 2015).

[9] g-tec Medical Engineering (last accessed: June 1, 2015).

[10] The Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy, Security and Breach Notification Rules (last accessed: May 19, 2015).

[11] LexisNexis Legal Newsroom Litigation: Da Vinci Surgical Robot Maker Reserves $67M to Settle Product Liability Claims (last accessed: May 21, 2015).

[12] Necomimi Brainwave Cat Ears (last accessed: August 10, 2015).

[13] Nest Labs (last accessed: May 7, 2015).

[14] Netflix (last accessed: May 7, 2015).

[15] NeuroFocus (last accesed: August 10, 2015).

[16] NeuroGaming 2013 Conference and Expo (last accessed: May 10, 2013).

[17] NeuroSky (last accessed: June 1, 2015).

[18] The New York Times: Italian Crews Struggle to Secure Cruise Ship (last accessed: May 7, 2015).

[19] Pandora Internet Radio (last accessed: May 7, 2015).

[20] Renishaw (last accessed: May 18, 2015).

[21] Robotic Surgery in Arthroplasty (last accessed: May 18, 2015).

[22] Surgical Robots Market 2015-2021 - Shares, Strategies, and Forecasts for the $20 Billion Market (last accessed: May 18, 2015).

[23] TechHive: Your Android Phone is Tracking You (last accessed: June 2, 2015).

[24] U.S. ex rel. Antoon v. Cleveland Clinic Found. (Summary) (last accessed: May 21, 2015).

[25] The Wall Street Journal: Mobile-App Makers Face U.S. Privacy Investigation (last accessed: June 2, 2015).

[26] Samsung Demos a Tablet Controlled by Your Brain (last accessed: June 1, 2015), 2013.

[27] C. L. Abad and R. I. Bonilla. An Analysis on the Schemes for Detecting and Preventing ARP Cache Poisoning Attacks. In *the Proceedings of the 27th International Conference on Distributed Computing Systems Workshops*, pages 60–60. IEEE, 2007.

[28] V. Abootalebi, M. H. Moradi, and M. A. Khalilzadeh. A New Approach for EEG Feature Extraction in P300-based Lie Detection. *Computer Methods and Programs in Biomedicine*, 94(1):48–57, 2009.

[29] H. Adeli, Z. Zhou, and N. Dadmehr. Analysis of EEG Records in an Epileptic Patient Using Wavelet Transform. *Journal of Neuroscience Methods*, 123(1):69, 2003.

[30] S. Agarwal, T. Dawson, and C. Tryfonas. DDoS Mitigation via Regional Cleaning Centers. Technical report, Sprint ATL Research Report RR04-ATL-013177, 2003.

[31] S. Amin, A. A. Cárdenas, and S. S. Sastry. Safe and Secure Networked Control Systems Under Denial-of-Service Attacks. In *Hybrid Systems: Computation and Control*, pages 31–45. Springer, 2009.

[32] S. Anthony. Hackers Backdoor the Human Brain, Successfully Extract Sensitive Data, August 2012.

[33] P. H. Baas, S. G. Charlton, and G. T. Bastin. Survey of New Zealand Truck Driver Fatigue and Fitness for Duty. *Transportation Research Part F: Traffic Psychology and Behaviour*, 3(4):185–193, 2000.

[34] R. B. Baldwin. Kinetic Art: On the Use of Subliminal Stimulation of Visual Perception. *Leonardo*, pages 1–5, 1974.

[35] M. Bar and I. Biederman. Subliminal Visual Priming. *Psychological Science*, 9(6):464–468, 1998.

[36] S. M. Bellovin and M. Merritt. Encrypted Key Exchange: Password-Based Protocols Secure Against Dictionary Attacks. In *the Proceedings of IEEE Computer Society Symposium on Research in Security and Privacy*, pages 72–84. IEEE, 1992.

[37] P. Biondi. Scapy Documentation, 2010.

[38] J. D. Birkmeyer, J. F. Finks, A. O'Reilly, M. Oerline, A. M. Carlin, A. R. Nunn, J. Dimick, M. Banerjee, and N. J. O. Birkmeyer. Surgical Skill and Complication Rates after Bariatric Surgery. *New England Journal of Medicine*, 369(15):1434–1442.

[39] C. Bo, L. Zhang, and X.-Y. Li. SilentSense: Silent User Identification via Dynamics of Touch and Movement Behavioral Biometrics. *arXiv preprint arXiv:1309.0073*, 2013.

[40] H. Bojinov, D. Sanchez, P. Reber, D. Boneh, and P. Lincoln. Neuroscience Meets Cryptography: Designing Crypto Primitives Secure Against Rubber Hose Attacks. In *the Proceedings of the 21$^{st}$ USENIX Security Symposium*. USENIX, 2012.

[41] T. Bonaci, A. Alva, J. Herron, R. Calo, and H. J. Chizeck. I Did It My Way: On Law And Operator Signatures for Teleoperated Robots. In *the Proceesings of the 4$^{th}$ Annual Conference on Robotics, Law and Policy*.

[42] T. Bonaci and L. Bushnell. Node Capture Games: A Game Theoretic Approach to Modeling and Mitigating Node Capture Attacks. In *the Proceedings of the 2$^{nd}$ Conference on Decision and Game Theory for Security,*, pages 44–55, 2011.

[43] T. Bonaci, L. Bushnell, and R. Poovendran. Node Capture Attacks in Wireless Sensor Networks: A System Theoretic Approach. In *the Proceesings of the 49$^{th}$ IEEE Conference on Decision and Control*, pages 6765–6772, 2010.

[44] T. Bonaci, L. Bushnell, and R. Poovendran. Probabilistic Analysis of Covering and Compromise in a Node Capture Attack. Techical Report #001, University of Washington, Network Security Lab (NSL), Seattle, WA, 2010.

[45] T. Bonaci and H. J. Chizeck. On Potential Security Threats Against Rescue Robotic Systems. In *the Proceedings of the 2012 IEEE International Symposium on Safety, Security, and Rescue Robotics*, pages 1–2. IEEE, 2012.

[46] T. Bonaci, J. Herron, C. Matlack, and H. J. Chizeck. Securing the Exocortex: A Twenty-first Century Cybernetics Challenge. In *the Proceedings of the 2014 IEEE Conference on Norbert Wiener in the 21st Century*, pages 1–8. IEEE, 2014.

[47] T. Bonaci, J. Herron, T. Yusuf, J. Yan, T. Kohno, and H. J. Chizeck. To Make a Robot Secure: An Experimental Analysis of Cyber Security Threats Against Teleoperated Surgical Robots. *arXiv preprint arXiv:1504.04339*, 2015.

[48] T. Bonaci, P. Lee, L. Bushnell, and R. Poovendran. Distributed Clone Detection in Wireless Sensor Networks: An Optimization Approach. In *the Proceedings of the 2$^{nd}$ IEEE International Workshop on Data Security and Privacy in Wireless Networks,*, pages 1–6, 2011.

[49] T. Bonaci, P. Lee, L. Bushnell, and R. Poovendran. A Convex Optimization Approach for Clone Detection in Wireless Sensor Networks. *Pervasive and Mobile Computing*, 2012.

[50] T. Bonaci, J. Yan, J. Herron, T. Kohno, and H. J. Chizeck. Experimental Analysis of Denial-of-Service Attacks on Teleoperated Robotic Systems. In *the Proceedings of the 6th ACM/IEEE International Conference on Cyber-Physical Systems*, 2015.

[51] L. A. Brannon and T. C. Brock. The Subliminal Persuasion Controversy: Reality, Enduring Fable, and Polonius's Weasel. *Persuasion: Psychological Insights and Perspectives*, 1994.

[52] C. S. Burrus, R. A. Gopinath, H. Guo, J. E. Odegard, and I. W. Selesnick. *Introduction to Wavelets and Wavelet Transforms: A Primer*, volume 23. Prentice Hall Upper Saddle River, 1998.

[53] S. Calabro. Breaking the Shield of the Learned Intermediary Doctrine: Placing the Blame Where it Belongs. *Cardozo Law Review*, 25:2241, 2003.

[54] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry. Attacks Against Process Control Systems: Risk Assessment, Detection, and Response. In *the Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, pages 355–366. ACM, 2011.

[55] A. A. Cárdenas, S. Amin, and S. Sastry. Research Challenges for the Security of Control Systems. In *the Proceedings of the 3rd Conference on Hot Topics in Security*, 2008.

[56] A. Carleial. Multiple-Access Channels with Different Generalized Feedback Signals. *IEEE Transactions on Information Theory*, 28(6):841–850, 1982.

[57] R. Chabukswar, Y. Mo, and B. Sinopoli. Detecting Integrity Attacks on SCADA Systems. In *the Proceedings of the 18th IFAC World Congress*, pages 11239–11244, 2011.

[58] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner, T. Kohno, et al. Comprehensive Experimental Analyses of Automotive Attack Surfaces. In *the Proceddings of the 2011 USENIX Security Symposium*, 2011.

[59] Y.-T. Chiu. Mind Reading to Predict the Success of Online Games, February 2013.

[60] H. J. Chizeck and T. Bonaci. Brain-Computer Interface Anonymizer, February 6 2014. US Patent App. 14/174,818.

[61] H. J. Chizeck, T. Bonaci, and T. Lendvay. Enhanced Security and Safety in Telerobotic Systems, July 3 2013. US Patent App. 13/935,436.

[62] B.-G. Chun and P. Maniatis. Augmented Smartphone Applications through Clone Cloud Execution. In *the Proceedings of the 12$^{th}$ Conference on Hot Topics in Operating Systems*. USENIX Association, 2009.

[63] N. L. Clarke and S. M. Furnell. Advanced User Authentication for Mobile Devices. *Computers & Security*, 26(2):109–119, 2007.

[64] K. Coble, W. Wang, B. Chu, and Z. Li. Secure Software Attestation for Military Telesurgical Robot Systems. In *the Proceedings of the Military Communications Conference*, pages 965–970. IEEE, 2010.

[65] J. Contreras-Vidal. Ethical Considerations Behind Brain-Computer Interface Research, December 2012.

[66] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.

[67] J. Daemen and V. Rijmen. *The Design of Rijndael: AES-the Advanced Encryption Standard*. Springer Science & Business Media, 2002.

[68] I. Daubechies. The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005, 1990.

[69] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann. Touch me Once and I Know It's You!: Implicit Authentication Based on Touch Screen Patterns. In *the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 987–996. ACM, 2012.

[70] Z. De-xiang, W. Xiao-Pei, and G. Xiao-jing. The EEG Signal Preprocessing Based on Empirical Mode Decomposition. In *the Proceedings of the 2nd International Conference on Bioinformatics and Biomedical Engineering*, pages 2131–2134. IEEE, 2008.

[71] S. Dehaene, J.-P. Changeux, L. Naccache, J. Sackur, and C. Sergent. Conscious, Preconscious, and Subliminal Processing: A Testable Taxonomy. *Trends in Cognitive Sciences*, 10(5):204–211, 2006.

[72] S. Dehaene, N. Kanwisher, and J. Duncan. The Neural Bases of Subliminal Priming. *Functional Neuroimaging of Visual Cognition (Attention and Performance Series, 20)*, 2004.

[73] S. Dehaene and L. Naccache. Can One Suppress Subliminal Words? *Neuron*, 52(3):397–399, 2006.

[74] S. Dehaene, L. Naccache, L. Cohen, D. Le Bihan, J.-F. Mangin, J.-B. Poline, and D. Rivière. Cerebral Mechanisms of Word Masking and Unconscious Repetition Priming. *Nature Neuroscience*, 4(7):752–758, 2001.

[75] S. Dehaene, L. Naccache, G. Le Clec'H, E. Koechlin, M. Mueller, G. Dehaene-Lambertz, P.-F. van de Moortele, and D. Le Bihan. Imaging Unconscious Semantic Priming. *Nature*, 395(6702):597–600, 1998.

[76] T. Denning, Y. Matsuoka, and T. Kohno. Neurosecurity: Security and Privacy for Neural Devices. *Neurosurgical Focus*, 27(1):1–4, 2009.

[77] P. J. Devereaux, M. Bhandari, M. Clarke, V. M. Montori, D. J. Cook, S. Yusuf, D. L. Sackett, C. S. Cinà, S. D. Walter, B. Haynes, et al. Need for Expertise Based Randomised Controlled Trials. *British Medical Journal*, 330(7482):88, 2005.

[78] N. Dowler and C. J. Hall. Safety Issues in Telesurgery-Summary. 1995.

[79] W. Enck. Defending Users Against Smartphone Apps: Techniques and Future Directions. In *Information Systems Security*, pages 49–70. Springer, 2011.

[80] R. R. Faden, T. L. Beauchamp, and N. M. King. A History and Theory of Informed Consent. 1986.

[81] N. Falliere, L. O. Murchu, and E. Chien. W32. Stuxnet Dossier. *White paper, Symantec Corp., Security Response*, 5, 2011.

[82] N. Farahany. Incriminating Thoughts. *Stanford Law Review*, 64:11–17, 2011.

[83] L. A. Farwell and E. Donchin. Talking off the Top of Your Head: Toward a Mental Prosthesis utilizing Event-Related Brain Potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–523, 1988.

[84] H. Fawzi, P. Tabuada, and S. Diggavi. Secure Estimation and Control for Cyber-Physical Systems under Adversarial Attacks. *IEEE Transactions on Automatic Control*.

[85] B. A. Forouzan. *Cryptography & Network Security*. McGraw-Hill, Inc., 2007.

[86] B. Fowler, S. Meehan, and A. Singhal. Perceptual-Motor Performance and Associated Kinematics in Space. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(6):879–892, 2008.

[87] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication. *IEEE Transactions on Information Forensics and Security*, 8(1):136–148, 2013.

[88] M. Friedlander. Neural Implants Come of Age, June 2012.

[89] B. Friedman. Value Sensitive Design. *Interactions*, 3(6):16–23, 1996.

[90] B. Friedman, P. H. Kahn Jr., A. Borning, and A. Huldtgren. Value Sensitive Design and Information Systems. In *Early Engagement and New Technologies: Opening up the Laboratory*, pages 55–95. Springer, 2013.

[91] B. Friedman, I. Smith, P. H. Kahn Jr., S. Consolvo, and J. Selawski. Development of a Privacy Addendum for Open Source Licenses: Value Sensitive Design in Industry. In *the Proceedings of the 2006 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 194–211, 2006.

[92] W. Gates. A Robot in Every Home. *Scientific American*, 296(1):58–65, 2007.

[93] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Springer Science & Business Media, 1992.

[94] M. Goldberg. Robotic Arm Went Crazy-The Problem of Establishing Liability in a Monopolized Field. *Rutgers Computer & Tech. LJ*, 38:225, 2012.

[95] S. Gollakota, H. Hassanieh, B. Ransford, D. Katabi, and K. Fu. They Can Hear Your Heartbeats: Non-invasive Security for Implantable Medical Devices. *ACM SIGCOMM Computer Communication Review*, 41(4):2–13, 2011.

[96] R. Grace, V. E. Byrne, D. M. Bierman, J.-M. Legrand, D. Gricourt, B. K. Davis, J. J. Staszewski, and B. Carnahan. A Drowsy Driver Detection System for Heavy Vehicles. In *the Proceedings of the AIAA/IEEE/SAE Digital Avionics Systems Conference*, volume 2, pages I36–1. IEEE, 1998.

[97] A. Graps. An Introduction to Wavelets. *IEEE Computational Science & Engineering*, 2(2):50–61, 1995.

[98] S. Greene. *Security Program and Policies: Principles and Practices*. Pearson Education, 2014.

[99] A. G. Greenwald, R. L. Abrams, L. Naccache, and S. Dehaene. Long-Term Semantic Memory Versus Contextual Memory in Unconscious Number Processing. *Journal of Experimental Psychology-Learning Memory and Cognition*, 29(2):235–246, 2003.

[100] K. Gurney. *An Introduction to Neural Networks*. CRC press, 1997.

[101] D. Halperin, T. S. Heydt-Benjamin, B. Ransford, S. S. Clark, B. Defend, W. Morgan, K. Fu, T. Kohno, and W. H. Maisel. Pacemakers and Implantable Cardiac Defibrillators: Software Radio Attacks and Xero-power Defenses. In *the Proceedings of the 2008 IEEE Symposium on Security and Privacy*, pages 129–142. IEEE, 2008.

[102] N. Halverson. Brain Hackers Pluck Your Private Data, August 2012.

[103] T. C. Hankins and G. F. Wilson. A Comparison of Heart Rate, Eye Activity, EEG and Subjective Measures of Pilot Mental Workload During Flight. *Aviation, Space, and Environmental Medicine*, 69(4):360–367, 1998.

[104] B. Hannaford, J. Rosen, D. W. Friedman, H. H. King, P. Roan, L. Cheng, D. Glozman, J. Ma, S. N. Kosari, and L. White. Raven-II: An Open Platform for Surgical Robotics Research. *IEEE Transactions on Biomedical Engineering*, 60(4):954–959, 2013.

[105] B. M. Harnett, C. R. Doarn, J. Rosen, B. Hannaford, and T. J. Broderick. Evaluation of Unmanned Airborne Vehicles and Mobile Robotic Telesurgery in an Extreme Environment. *Telemedicine and e-Health*, 14(6):539–544, 2008.

[106] S .D. Hawley, L. E. Atlas, and H. J. Chizeck. Some Properties of an Empirical Mode Type Signal Decomposition Algorithm. *IEEE Signal Processing Letters*, 17(1):24–27, 2010.

[107] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999.

[108] J. Herron and H. J. Chizeck. Prototype Closed-loop Deep Brain Stimulation Systems Inspired by Norbert Wiener. In *the Proceedings of the 2014 IEEE Conference on Norbert Wiener in the 21st Century*, pages 1–6. IEEE, 2014.

[109] R. W. Homan, J. Herman, and P. Purdy. Cerebral Location of International 10–20 System Electrode Placement. *Electroencephalography and Clinical Neurophysiology*, 66(4):376–382, 1987.

[110] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis. In *the Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, volume 454, pages 903–995. The Royal Society, 1998.

[111] S. A. Huettel and G. McCarthy. What is Odd in the Oddball Task? Prefrontal Cortex is Activated by Dynamic Changes in Response Strategy. *Neuropsychologia*, 42(3):379–386, 2004.

[112] A. Hussain, J. Heidemann, and C. Papadopoulos. A Framework for Classifying Denial of Service Attacks. In *the Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pages 99–110. ACM, 2003.

[113] J. Illes, M. P. Kirschen, J. D. E. Gabrieli, et al. From Neuroimaging to Neuroethics. *Nature Neuroscience*, 6(3):205–205, 2003.

[114] J. Illes and E. Racine. Imaging or Imagining? A Neuroethics Challenge Informed by Genetics. *The American Journal of Bioethics*, 5(2):5–18, 2005.

[115] M. Inzlicht, I. McGregor, J. B. Hirsh, and K. Nash. Neural Markers of Religious Conviction. *Psychological Science*, 20(3):385–392, 2009.

[116] P. Jahankhani, V. Kodogiannis, and K. Revett. EEG Signal Classification Using Wavelet Feature Extraction and Neural Networks. In *the Proceedings of the IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing*, pages 120–124. IEEE, 2006.

[117] A. K. Jain, A. Ross, and S. Pankanti. Biometrics: A Tool for Information Security. *IEEE Transactions on Information Forensics and Security,*, 1(2):125–143, 2006.

[118] A. K. Jain, A. Ross, and S. Prabhakar. An Introduction to Biometric Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, 2004.

[119] B. Johnson, T. Maillart, and J. Chuang. My Thoughts Are Not Your Thoughts. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 1329–1338. ACM, 2014.

[120] A. R. Jonsen. *The Birth of Bioethics*. Oxford University Press, USA, 2003.

[121] R. Joyce and G. Gupta. Identity Authentication Based on Keystroke Latencies. *Communications of the ACM*, 33(2):168–176, 1990.

[122] P. H. Kahn Jr, H. Ishiguro, B. Friedman, T. Kanda, N. G. Freier, R. L. Severson, and J. Miller. What is a Human?: Toward Psychological Benchmarks in the Field of Human–Robot Interaction. *Interaction Studies*, 8(3):363–390, 2007.

[123] K. Kahol, R. M. Satava, J. Ferrara, and M. L. Smith. Effect of Short-term Pretrial Practice on Surgical Proficiency in Simulated Environments: A Randomized Trial of the "Preoperative Warm-up" Effect. *Journal of the American College of Surgeons*, 208(2):255–268, 2009.

[124] S. Kalan, S. Chauhan, R. F. Coelho, M. A. Orvieto, I. R. Camacho, K. J. Palmer, and V. R. Patel. History of Robotic Surgery. *Journal of Robotic Surgery*, 4(3):141–147, 2010.

[125] C. Kaufman, R. Perlman, and M. Speciner. *Network Security: Private Communication in a Public World*. Prentice Hall Press, 2002.

[126] R. Kerr. Movement Time in an Underwater Environment. *Journal of Motor Behavior*, 5(3):175–178, 1973.

[127] H. H. King. *Human-Machine Collaborative Telerobotics: Computer Assistance for Manually Controlled Telesurgery and Teleoperation*. PhD thesis, 2014.

[128] H. H. King, B. Hannaford, J. Kammerl, and E. Steinbach. Establishing Multimodal Telepresence Sessions Using the Session Initiation Protocol (SIP) and Advanced Haptic Codecs. In *Haptics Symposium, 2010 IEEE*, pages 321–325. IEEE, 2010.

[129] H. H. King, B. Hannaford, K.-W. Kwok, G.-Z. Yang, P. Griffiths, A. Okamura, I. Farkhatdinov, J.-H. Ryu, G. Sankaranarayanan, V. Arikatla, et al. Plugfest 2009: Global Interoperability in Telerobotics and Telemedicine. In *the Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1733–1738. IEEE, 2010.

[130] H. H. King, K. Tadano, R. Donlin, D. Friedman, M. J. H. Lum, V. Asch, C. Wang, K. Kawashima, and B. Hannaford. Preliminary Protocol for Interoperable Telesurgery. In *the Proceedings of the International Conference on Advanced Robotics*, pages 1–6. IEEE, 2009.

[131] J. Koreman, A. C. Morris, S. Wu, D.and Jassim, H. Sellahewa, J. Ehlers, G. Chollet, G. Aversano, H. Bredin, S. Garcia-Salicetti, et al. Multi-Modal Biometric Authentication on the SecurePhone PDA. 2006.

[132] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, et al. Experimental Security Analysis of a Modern Automobile. In *the Proceedings of the 2010 IEEE Symposium on Security and Privacy*, pages 447–462. IEEE, 2010.

[133] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw. A Comparison of Classification Techniques for the P300 Speller. *Journal of Neural Engineering*, 3(4):299, 2006.

[134] M. R. Kwaan and G. B. Melton. Evidence-Based Medicine in Surgical Education. *Clinics in Colon and Rectal Surgery*, 25(3):151, 2012.

[135] G. D. Langolf, D. B. Chaffin, and J. A. Foulke. An Investigation of Fitts' Law Using a Wide Range of Movement Amplitudes. *Journal of Motor Behavior*, 8(2):113–128, 1976.

[136] G. S. Lee and B. Thuraisingham. Cyberphysical Systems Security Applied to Telesurgical Robotics. *Computer Standards & Interfaces*, 34(1):225–229, 2012.

[137] N. Lee, A. J. Broderick, and L. Chamberlain. What is "Neuromarketing"? a Discussion and Agenda for Future Research. *International Journal of Psychophysiology*, 63(2):199–204, 2007.

[138] A. Leon-Garcia. Probability and Random Processes. *Addison Wesley*, 3:126–127, 1989.

[139] G. Loukas and G. Öke. Protection Against Denial of Service Attacks: A Survey. *The Computer Journal*, page bxp078, 2009.

[140] B. Luber, C. Fisher, P. S. Appelbaum, M. Ploesser, and S. H. Lisanby. Non-invasive Brain Stimulation in the Detection of Deception: Scientific Challenges and Ethical Consequences. *Behavioral Sciences & the Law*, 27(2):191–208, 2009.

[141] S. J. Luck. *An Introduction to the Event-Related Potential Technique*. MIT press, 2014.

[142] M. J. H. Lum, D. C. W. Friedman, H. H. King, T. Broderick, M. N. Sinanan, J. Rosen, and B. Hannaford. Field Operation of a Surgical Robot via Airborne Wireless Radio Link. In *the Proceedings of the IEEE International Conference on Field and Service Robotics*, 2007.

[143] M. J. H. Lum, D. C. W. Friedman, G. Sankaranarayanan, H. H. King, A. Wright, M. Sinanan, T. Lendvay, J. Rosen, and B. Hannaford. Objective Assessment of Telesurgical Robot Systems: Telerobotic FLS. *Studies in Health Technology and Informatics*, 132:263, 2008.

[144] M. J. H. Lum, J. Rosen, T. S. Lendvay, M. N. Sinanan, and B. Hannaford. Effect of Time Delay on Telesurgical Performance. In *the Proceedings of the IEEE International Conference on Robotics and Automation*, pages 4246–4252. IEEE, 2009.

[145] M. J. H. Lum, D. Trimble, J. Rosen, K. Fodero, H. H. King, G. Sankaranarayanan, J. Dosher, R. Leuschke, B. Martin-Anderson, M. N. Sinanan, et al. Multidisciplinary Approach for Developing a New Minimally Invasive Surgical Robotic System. In *the Proceedings of the First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics*, pages 841–846. IEEE, 2006.

[146] I. S. MacKenzie. Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human-Computer Interaction*, 7(1):91–139, 1992.

[147] J. N. Mak and J. R. Wolpaw. Clinical Applications of Brain-Computer Interfaces: Current State and Future Prospects. *IEEE Reviews in Biomedical Engineering*, 2:187–199, 2009.

[148] L. Makris, N. Argiriou, and M. G. Strintzis. Network and Data Security Design for Telemedicine Applications. *Informatics for Health and Social Care*, 22(2):133–142, 1997.

[149] V. Manyakov, N, N. Chumerin, A. Combaz, and M. M. Van Hulle. Comparison of Classification Methods for P300 Brain-Computer Interface on Disabled Subjects. *Computational Intelligence and Neuroscience*, 2011:2, 2011.

[150] S. Marcel and J. del R Millán. Person Authentication Using Brainwaves (EEG) and Maximum A Posteriori Model Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):743–752, 2007.

[151] J. Marescaux, J. Leroy, M. Gagner, F. Rubino, D. Mutter, M. Vix, S. E. Butner, and M. K. Smith. Transatlantic Robot-assisted Telesurgery. *Nature*, 413(6854):379–380, 2001.

[152] C. Marforio, A. Francillon, and S. Capkun. *Application Collusion Attack on the Permission-based Security Model and its Implications for Modern Smartphone Systems*. Department of Computer Science, ETH Zurich, 2011.

[153] I. Martinovic, D. Davies, M. Frank, D. Perito, T. Ros, and D. Song. On the Feasibility of Side-Channel Attacks with Brain-Computer Interfaces. In *the Proceedings of the 21$^{st}$ USENIX Security Symposium*. USENIX, 2012.

[154] T. H. Massie and J. K. Salisbury. The Phantom Haptic Interface: A Device for Probing Virtual Objects. In *Proceedings of the ASME Winter Annual Meeting, Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, volume 55, pages 295–300. Chicago, IL, 1994.

[155] C. B. Matlack. *Adaptation for Brain-Computer Interfaces*. PhD thesis, 2014.

[156] F. McClellan. *Medical Malpractice: Law, Tactics, and Ethics*. Temple University Press, 1994.

[157] S. McClure, J. Scambray, G. Kurtz, and Kurtz. Hacking Exposed: Network Security Secrets and Solutions. 2009.

[158] F. McCormick, J. Kadzielski, C. P. Landrigan, B. Evans, J. H. Herndon, and H. E. Rubash. Surgeon Fatigue: S Prospective Analysis of the Incidence, Risk, and Intervals of Predicted Fatigue-related Impairment in Residents. *Archives of Surgery*, 147(5):430–435, 2012.

[159] T. R. McLean. Cybersurgery–An Argument for Enterprise Liability. *Journal of Legal Medicine*, 23(2):167–210, 2002.

[160] Z. Mehboob. *Information Quantification for Spike Trains and Field Potentials*. PhD thesis, University of Manchester, 2011.

[161] M. Mehlman. Professional Power and the Standard of Care in Medicine. 2013.

[162] M. M. Mello, A. Chandra, A. A. Gawande, and D. M. Studdert. National Costs of the Medical Liability System. *Health Affairs*, 29(9):1569–1577, 2010.

[163] A. J. Menezes, P. C. Van Oorschot, and S. A. Vanstone. *Handbook of Applied Cryptography*. CRC press, 1996.

[164] J. K. Miller, B. Friedman, G. Jancke, and B. Gill. Value Tensions in Design: The Value Sensitive Design, Development, and Appropriation of a Corporation's Groupware System. In *the Proceedings of the 2007 International ACM Conference on Supporting Group Work*, pages 281–290. ACM, 2007.

[165] J. Mirkovic and P. Reiher. A Taxonomy of DDoS Attacks and DDoS Defense Mechanisms. *ACM SIGCOMM Computer Communication Review*, 34(2):39–53, 2004.

[166] Y. Mo and B. Sinopoli. Secure Control Against Replay Attacks. In *the Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing, Allerton*, pages 911–918. IEEE, 2009.

[167] Y. Mo and B. Sinopoli. False Data Injection Attacks in Control Systems. In *Preprints of the 1st Workshop on Secure Control Systems*, pages 1–6, 2010.

[168] M. K. Molla, T. Tanaka, T. M. Rutkowski, and A. Cichocki. Separation of EOG Artifacts from EEG Signals Using Bivariate EMD. In *the Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing*, pages 562–565. IEEE, 2010.

[169] M. F. Møller. A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning. *Neural Networks*, 6(4):525–533, 1993.

[170] F. Monrose, M. K. Reiter, and S. Wetzel. Password Hardening Based on Keystroke Dynamics. *International Journal of Information Security*, 1(2):69–83, 2002.

[171] T. Moore, T. Brown, J. Herron, M. Thompson, T. Bonaci, S. Goering, and H. J. Chizeck. Personal Responsibility in the Age of User-Controlled Neuroprosthetics. In *the Proceedings of the $4^{th}$ Annual Conference on Robotics, Law and Policy*, 2015.

[172] T. A. Moore. *Medical Malpractice: Discovery and Trial*. Number 1. Practising Law Institute, 2002.

[173] T. E. Moore. Subliminal Perception: Facts and Fallacies. *Skeptical Inquirer*, 16(3):273–281, 1992.

[174] P. Moulin and J. A. O'Sullivan. Information-theoretic Analysis of Information Hiding. *IEEE Transactions on Information Theory*, 49(3):563–593, 2003.

[175] G. R. Müller-Putz, R. Scherer, C. Brauneis, and G. Pfurtscheller. Steady-State Visual Evoked Potential (SSVEP)-Based Communication: Impact of Harmonic Frequency Components. *Journal of Neural Engineering*, 2(4):123, 2005.

[176] R. R. Murphy. Human-Robot Interaction in Rescue Robotics. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 34(2):138–153, 2004.

[177] K. Nagatani, S. Kiribayashi, Y. Okada, K. Otake, K. Yoshida, S. Tadokoro, T. Nishimura, T. Yoshida, E. Koyanagi, M. Fukushima, et al. Emergency Response to the Nuclear Accident at the Fukushima Daiichi Nuclear Power Plants Using Mobile Rescue Robots. *Journal of Field Robotics*, 30(1):44–63, 2013.

[178] A. Narayanan and V. Shmatikov. Fast Dictionary Attacks on Passwords using Time-Space Tradeoff. In *Proceedings of the 12th ACM conference on Computer and Communications Security*, pages 364–372. ACM, 2005.

[179] M. S. Obaidat and B. Sadoun. Keystroke Dynamics Based Authentication. In *Biometrics*, pages 213–229. Springer, 1996.

[180] C. Omar, A. Akce, M. Johnson, T. Bretl, R. Ma, E. Maclin, M. McCormick, and T. P. Coleman. A Feedback Information-theoretic Approach to the Design of Brain-Computer Interfaces. *International Journal of Human-Computer Interaction*, 27(1):5–23, 2010.

[181] L. H. Ozarow. The Capacity of the White Gaussian Multiple Access Channel with Feedback. *IEEE Transactions on Information Theory*, 30(4):623–629, 1984.

[182] R. Palaniappan and K. V. R. Ravi. A New Method to Identify Individuals Using Signals From the Brain. In *the Proceedings of the $4^{th}$ Joint Conference on Information, Communications and Signal Processing.*, volume 3, pages 1442–1445, 2003.

[183] R. B. Paranjape, J. Mahovsky, L. Benedicenti, and Z. Koles. The Electroencephalogram as a Biometric. In *the Proceedings of the Canadian Conference on Electrical and Computer Engineering*, volume 2, pages 1363–1366, 2001.

[184] C. Park, D. Looney, P. Kidmose, M. Ungstrup, and D. P. Mandic. Time-frequency Analysis of EEG Asymmetry Using Bivariate Empirical Mode Decomposition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19(4):366–373, 2011.

[185] F. Pasqualetti, A. Bicchi, and F. Bullo. On the Security of Linear Consensus Networks. In *the Proceedings of the 48th IEEE Conference on Decision and Control*, pages 4894–4901. IEEE, 2009.

[186] C. Patrikakis, M. Masikos, and O. Zouraraki. Distributed Denial of Service Attacks. *The Internet Protocol Journal*, 7(4):13–35, 2004.

[187] J. C. Perry and J. Rosen. Design of a 7 Degree-of-freedom Upper-limb Powered Exoskeleton. In *the Proceedings of the 1st IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics*, pages 805–810. IEEE, 2006.

[188] M. Pessiglione, L. Schmidt, B. Draganski, R. Kalisch, H. Lau, R. J. Dolan, and C. D. Frith. How the Brain Translates Money into Force: a Neuroimaging Study of Subliminal Motivation. *Science*, 316(5826):904–906, 2007.

[189] P. G. Peters Jr. The Quiet Demise of Deference to Custom: Malpractice Law at the Millenium. *Washington & Lee Law Review*, 57:163, 2000.

[190] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn. Information Hiding-A Survey. *Proceedings of the IEEE*, 87(7):1062–1078, 1999.

[191] J. Postel. User Datagram Protocol. *ISI*, 1980.

[192] M. Poulos, M. Rangoussi, V. Chrissikopoulos, and A. Evangelou. Person Identification Based on Parametric Processing of the EEG. In *the Proceedings of the 66th IEEE International Conference on Electronics, Circuits and Systems*, volume 1, pages 283–286, 1999.

[193] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng. ROS: An Open-Source Robot Operating System. In *the Proceedings of ICRA Workshop on Open Source Software*, volume 3, page 5, 2009.

[194] K. Revett and S. T. de Magalhães. Cognitive Biometrics: Challenges for the Future. In *Global Security, Safety, and Sustainability*, pages 79–86. Springer, 2010.

[195] C. Richards, J. Rosen, B. Hannaford, C. Pellegrini, and M. Sinanan. Skills Evaluation in Minimally Invasive Surgery Using Force/Torque Signatures. *Surgical Endoscopy*, 14(9):791–798, 2000.

[196] A. Riera, A. Soria-Frisch, M. Caparrini, C. Grau, and G. Ruffini. Unobtrusive Biometric System Based on Electroencephalogram Analysis. *EURASIP Journal on Advances in Signal Processing*, 2008, 2007.

[197] J. Rosen, L. Chang, J. D. Brown, B. Hannaford, M. Sinanan, and R. Satava. Minimally Invasive Surgery Task Decomposition-Etymology of Endoscopic Suturing. *Studies in Health Technology and Informatics*, pages 295–301, 2003.

[198] J. Rosen and B. Hannaford. Doc at a Aistance. *IEEE Spectrum*, 43(10):34–39, 2006.

[199] J. Rosen, B. Hannaford, C. Richards, and M. Sinanan. Markov Modeling of Minimally Invasive Surgery Based on Tool/Tissue Interaction and Force/Torque Signatures for Evaluating Surgical Skills. *IEEE Transactions on Biomedical Engineering,*, 48(5):579–591, 2001.

[200] J. Rosen, C. Richards, B. Hannaford, and M. Sinanan. Hidden Markov Models of Minimally Invasive Surgery. *Studies in Health Technology and Informatics*, pages 279–285, 2000.

[201] J. Rosen, M. Solazzo, B. Hannaford, and M. Sinanan. Objective Laparoscopic Skills Assessments of Surgical Residents Using Hidden Markov Models Based on Haptic Information and Tool/Tissue Interactions. *Studies in Health Technology and Informatics*, pages 417–423, 2001.

[202] J. Rosen, M. Solazzo, B. Hannaford, and M. Sinanan. Task Decomposition of Laparoscopic Surgery for Objective Evaluation of Surgical Residents' Learning Curve Using Hidden Markov Model. *Computer Aided Surgery*, 7(1):49–61, 2002.

[203] J. P. Rosenfeld, J. R. Biroschak, and J. J. Furedy. P300-based Detection of Concealed Autobiographical Versus Incidentally Acquired Information in Target and Non-target Paradigms. *International Journal of Psychophysiology*, 60(3):251–259, 2006.

[204] D. L. Sackett, W. Rosenberg, J. A. Gray, R. B. Haynes, and W. S. Richardson. Evidence Based Medicine: What It Is and What It Isn't. *British Medical Journal*, 312(7023):71–72, 1996.

[205] G. Sankaranarayanan, H. H. King, S.-Y. Ko, M. J. H. Lum, D. C. W. Friedman, J. Rosen, and B. Hannaford. Portable Surgery Master Station for Mobile Robotic Telesurgery. In *the Proceedings of the 1st international conference on Robot communication and coordination*, page 28. IEEE Press, 2007.

[206] K. Scarfone and P. Mell. Guide to Intrusion Detection and Prevention Systems (IDPS). *NIST Special Publication*, 800(2007):94, 2007.

[207] M. Shensa. The Discrete Wavelet Transform: Wedding the a Trous and Mallat Algorithms. *IEEE Transactions on Signal Processing*, 40(10):2464–2482, 1992.

[208] T. B. Sheridan. *Telerobotics, Automation, and Human Supervisory Control*. MIT press, 1992.

[209] G. Shoemaker, T. Tsukitani, Y. Kitamura, and K. S. Booth. Two-Part Models Capture the Impact of Gain on Pointing Performance. *ACM Transactions on Computer-Human Interaction*, 19(4):28, 2012.

[210] R. S. Siegler. The Twenty Questions Game as a Form of Problem Solving. *Child Development*, pages 395–403, 1977.

[211] R. Spillane. Keyboard Apparatus for Personal Identification. *IBM Technical Disclosure Bulletin*, 17(3346):3346, 1975.

[212] S. Sundaram and C. N. Hadjicostis. Distributed Function Calculation via Linear Iterations in the Presence of Malicious Agentspart I: Attacking the Network. In *the Proceedings of the American Control Conference*. IEEE, 2008.

[213] S. Sundaram and C. N. Hadjicostis. Distributed Function Calculation via Linear Iterations in the Presence of Malicious Agentspart II: Overcoming Malicious Behavior. In *the Proceedings of the American Control Conference*, pages 1356–1361. IEEE, 2008.

[214] K. Sweeney, S. McLoone, and T. Ward. The Use of Ensemble Empirical Mode Decomposition with Canonical Correlation Analysis as a Novel Artifact Removal Technique. *IEEE Transactions on Biomedical Engineering*, 60(1), 2013.

[215] K. Tadano and K. Kawashima. Development of a Master Slave System with Force Sensing Using Pneumatic Servo System for Laparoscopic Surgery. In *the Proceedings of the IEEE International Conference on Robotics and Automation*, pages 947–952. IEEE, 2007.

[216] K. Tadano and K. Kawashima. Development of a Pneumatically Driven Forceps Manipulator Ibis IV. In *the Proceedings of the IEEE International Joint Conference ICCAS-SICE*, pages 3815–3818. IEEE, 2009.

[217] A. K. Tafreshi, A. M. Nasrabadi, and A. H. Omidvarnia. Epileptic Seizure Detection Using Empirical Mode Decomposition. In *the Proceedings of the IEEE International Symposium on Signal Processing and Information Technology*, pages 238–242. IEEE, 2008.

[218] The Committee on Science and Law. Are Your Thoughts Your Own?: "Neuroprivacy" and the Legal Implications of Brain Imaging, 2005.

[219] J. M. Thompson, M. P. Ottensmeyer, and T. B. Sheridan. Human Factors in Telesurgery: Effects of Time Delay and Asynchrony in Video and Control Feedback with Local Manipulative Assistance. *Telemedicine Journal*, 5(2):129–137, 1999.

[220] A. Tiwari, B. Dutertre, D. Jovanović, T. de Candia, P. D. Lincoln, J. Rushby, D. Sadigh, and S. Seshia. Safety Envelope for Security. In *the Proceedings of the 3rd International Conference on High Confidence Networked Systems*, pages 85–94. ACM, 2014.

[221] M. E. Tozal, Y. Wang, E. Al-Shaer, K. Sarac, B. Thuraisingham, and B.-T. Chu. On Secure and Resilient Telesurgery Communications Over Unreliable Networks. In *the Proceedings of the 2011 IEEE Conference on Computer Communications Workshops*, pages 714–719. IEEE, 2011.

[222] M. E. Tozal, Y. Wang, E. Al-Shaer, K. Sarac, B. Thuraisingham, and B.-T. Chu. Adaptive Information Coding for Secure and Reliable Wireless Telesurgery Communications. *Mobile Networks and Applications*, 18(5):697–711, 2013.

[223] J. B. F. Van Erp, F. Lotte, and M. Tangermann. Brain-Computer Interfaces: Beyond Medical Applications. *Computer*, (4):26–34, 2012.

[224] M. van Vliet, C. Mühl, B. Reuderink, and M. Poel. Guessing What's on Your Mind: Using the N400 in Brain Computer Interfaces. *Brain Informatics*, pages 180–191, 2010.

[225] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Science & Business Media, 2000.

[226] T. Vuong, A. Filippoupolitis, G. Loukas, and D. Gan. Physical Indicators of Cyber Attacks Against a Rescue Robot. In *the Proceedings of the 2014 IEEE International Conference on Pervasive Computing and Communications*, pages 338–343. IEEE, 2014.

[227] J. F. Weaver. *Robots are People Too: How Siri, Google Car, and Artificial Intelligence Will Force Us to Change Our Laws*. ABC-CLIO, 2013.

[228] A. T. Welford. Fundamentals of Skill. 1968.

[229] L. W. White. *Quantitative Objective Assessment of Preoperative Warm-up for Robotic Surgery*. PhD thesis, University of Washington, 2013.

[230] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan. A Review of the First International Meeting. *IEEE Transactions on Rehabilitation Engineering*, 8(2):164–173, 2000.

[231] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. Brain-Computer Interfaces for Communication and Control. *Clinical Neurophysiology*, 113(6):767–791, 2002.

[232] J. R. Wolpaw and E. W. Wolpaw. *Brain-Computer Interfaces: Principles and Practice*. OUP USA, 2012.

[233] P. R. Wolpe, K. R. Foster, and D. D. Langleben. Emerging Neurotechnologies for Lie-detection: Promises and Perils. *The American Journal of Bioethics*, 10(10):40–48, 2010.

[234] F. Wozak, T. Schabetsberger, and E. Ammmenwerth. End-to-end Security in Telemedical Networks - A Practical Guideline. *International Journal of Medical Informatics*, 76(5):484–490, 2007.

[235] A. D. Wyner. The Wire-Tap Channel. *Bell System Technical Journal*, 54(8):1355–1387, 1975.

[236] W. Xu, K. Ma, W. Trappe, and Y. Zhang. Jamming Sensor Networks: Attack and Defense Strategies. *IEEE Networks*, 20(3):41–47, 2006.

[237] R. V. Yampolskiy and V. Govindaraju. Behavioural Biometrics: A Survey and Classification. *International Journal of Biometrics*, 1(1):81–113, 2008.

[238] J. Yan, K. Huang, T. Bonaci, and H. J. Chizeck. Haptic Passwords. In *in the Proceesings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2015.

[239] Y. Yang, Z. Wang, F. Bao, and R. H. Deng. Secure the Image-based Simulated Telesurgery System. In *the Proceedings of the 2003 International Symposium on Circuits and Systems*, volume 2, pages II–596. IEEE, 2003.

[240] M.-S. Yoh, J. Kwon, and S. Kim. A BCI Game in the Form of Interactive Fairy Tale. In *the Proceedings of the 12$^{th}$ ACM International Conference Adjunct Papers on Ubiquitous Computing*, pages 389–390. ACM, 2010.

[241] A. C. Yoo, G. R. Gilbert, and T. J. Broderick. Military Robotic Combat Casualty Extraction and Care. In *Surgical Robotics*, pages 13–32. Springer, 2011.

[242] C.-M. Zeng, F. Kuhlmann, and A. Buzo. Achievability Proof of Some Multiuser Channel Coding Theorems Using Backward Decoding. *IEEE Transactions on Information Theory*, 35(6):1160–1165, 1989.

[243] N. Zheng, K. Bai, H. Huang, and H. Wang. You are How You Touch: User Verification on Smartphones via Tapping Behaviors. In *the Proceedings of the IEEE 22nd International Conference on Network Protocols*, pages 221–232. IEEE, 2014.

[244] Y. Zhou, X. Zhang, X. Jiang, and V. W. Freeh. Taming Information-stealing Smartphone Applications (on Android). In *Trust and Trustworthy Computing*, pages 93–107. Springer, 2011.

# Appendix A

# BRAIN MALWARE EXPERIMENTS

**Table A.1:** Additional information about subjects participating in experimental study "Brain-Computer Interface (BCI) Security and Privacy.

| **Subject's ID** | **Subject's Age** | **Subject's Gender** |
|---|---|---|
| S1 | 44 | Male |
| S2 | 32 | Male |
| S3 | 25 | Male |
| S4 | 26 | Male |
| S5 | 19 | Female |
| S6 | 19 | Female |
| S7 | 22 | Female |
| S8 | 19 | Male |
| S9 | 26 | Male |

**UNIVERSITY OF WASHINGTON**
**EXPERIMENTAL QUESTIONNAIRE**

**Brain-Computer Interface (BCI) Security and Privacy**

**Experiment number: _____**

**Date: _____**

**The purpose of this questionnaire is to protect your privacy. In each of the following questions, we ask you to please choose one or more items that a personal importance to you. By not revealing any details as to how you made the selection, your selections will appear random to us, thus preventing any possible privacy implications.**

1. **Numbers:**

   Please choose ***two numbers*** that have a special meaning to you, and that you would immediately recognize (for example, a part of your phone number, your year of birth, your license plate number). You may include slashes, dashes, or other punctuation that would normally be written with these numbers.

   First number: _____

   Second number: _____

2. **Names:**

   Please choose ***two first names*** that have a high significance and/or a special meaning to you (for example, your mom's first name, your sibling's first name, your significant other's first name, the name of your pet).

   Name 1: _____

   Name 2: _____

3. **Places:**

   Please choose the names of ***two places (geographical locations)*** that have a special meaning to you (for example, your hometown, your favorite vacation place, or the city you'd very much like to visit).

Place 1: _____

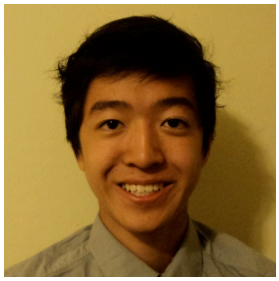
Place 2:_____

**4. People:**

Please, check all the people that you recognize on the pictures below. For those that you recognize, please write down their names and occupations.

_____ _____ _____

_____ _____ _____



_____ _____ _____

_____ _____ _____

**UNIVERSITY OF WASHINGTON**
**EXIT QUESTIONNAIRE**

**Brain-Computer Interface (BCI) Security and Privacy**

**Experiment number: _____**

**Date: _____**

**The purpose of this questionnaire is to helps us assess your experience of the experiment. Please note that we <u>will not</u> use your answers to this questionnaire to try to extract any private or personal information about you.**

1. **How do feel about the experiment, in which information was extracted from your brain signals, now that you have participated?**

   (a) **RELAXED** – i.e., I feel comfortable and relaxed about having this information extracted.

   (b) **NEUTRAL** –i.e., I do not have any particular feelings about my participation.

   (c) **UNCOMFORTABLE** – i.e., I feel a bit uncomfortable/concerned about having this information extracted.

2. **Did you notice anything that surprised you during the experiment?**

   (a) YES.

   (b) NO.

3. **If your answer to question 2 is YES, please explain below:**

4. **Would you be willing to participate in other experiments as a part of this study?**

   (a) YES.

   (b) NO.

**5. May we contact you about participating in future studies?**

(a) YES.

(b) NO.

# Appendix B

# TELEOPERATION SECURITY EXPERIMENTS

**Table B.1:** Table of experiments conducted as a part of study "Analysis of the Impact Adversarial Attacks Have on Teleoperated Procedures".

| **Experiment #** | **Attack or Defense** | **Exp. Task** | **# of Subjects** |
|---|---|---|---|
| Exp. 1 | Delay attacks | FLS task | 14 |
| Exp. 2 | Intent modification attacks | FLS task | 6 |
| Exp.3 | Denial-of-service attacks | Fitts' task | 11 |
| Exp. 4 | Delay attacks | Fitts' task | 10 |
| Exp. 5 | Haptic passwords | Haptic interaction task | 9 |

**Table B.2:** Additional information about subjects participating in experimental study "Analysis of the Impact Adversarial Attacks Have on Teleoperated Procedures" - Experiment 1.

| **Exp. #** | **Subject's ID** | **Subject's Age** | **Subject's Gender** | **Subject's Handedness** |
|---|---|---|---|---|
| Exp.1 | S1 | 19 | Female | Right |
| Exp.1 | S2 | 24 | Male | Right |
| Exp.1 | S3 | 26 | Female | Right |
| Exp.1 | S4 | 21 | Female | Right |
| Exp.1 | S5 | 25 | Male | Right |
| Exp.1 | S6 | 25 | Male | Right |
| Exp.1 | S7 | 20 | Male | Right |
| Exp.1 | S8 | 20 | Female | Right |
| Exp.1 | S9 | 19 | - | - |
| Exp.1 | S10 | 24 | Female | Right |
| Exp.1 | S11 | 22 | Female | Right |
| Exp.1 | S12 | 21 | Male | Right |
| Exp.1 | S13 | 22 | Female | Right |
| Exp.1 | S14 | 27 | Female | Right |

**Table B.3:** Additional information about subjects participating in experimental study "Analysis of the Impact Adversarial Attacks Have on Teleoperated Procedures" - Experiments 2 and 3.

| **Exp. #** | **Subject's ID** | **Subject's Age** | **Subject's Gender** | **Subject's Handedness** |
|---|---|---|---|---|
| Exp.2 | S15 | - | Male | - |
| Exp.2 | S16 | - | Male | - |
| Exp.2 | S17 | - | Male | - |
| Exp.2 | S18 | - | Male | - |
| Exp.2 | S19 | - | Male | - |
| Exp.2 | S20 | - | Female | - |
| Exp.3 | S21 | 21 | Female | Right |
| Exp.3 | S22 | 27 | Male | Right |
| Exp.3 | S23 | 28 | Female | Right |
| Exp.3 | S24 | 27 | Male | Right |
| Exp.3 | S25 | 23 | Male | Right |
| Exp.3 | S26 | 24 | Female | Right |
| Exp.3 | S20 | 20 | Male | Right |
| Exp.3 | S28 | 26 | Male | Right |
| Exp.3 | S29 | 30 | Male | Right |
| Exp.3 | S3 | 26 | Female | Right |
| Exp.3 | S30 | 24 | Male | Right |

**Table B.4:** Additional information about subjects participating in experimental study "Analysis of the Impact Adversarial Attacks Have on Teleoperated Procedures" - Experiments 4 and 5.

| **Exp. #** | **Subject's ID** | **Subject's Age** | **Subject's Gender** | **Subject's Handedness** |
|---|---|---|---|---|
| Exp.4 | S3 | 16 | Female | Right |
| Exp.4 | S26 | 24 | Female | Right |
| Exp.4 | S31 | 25 | Female | Right |
| Exp.4 | S25 | 23 | Male | Right |
| Exp.4 | S32 | 23 | Female | Right |
| Exp.4 | S21 | 21 | Female | Right |
| Exp.4 | S29 | 30 | Male | Right |
| Exp.4 | S8 | 20 | Female | Right |
| Exp.4 | S23 | 28 | Female | Right |
| Exp.4 | S28 | 26 | Male | Right |
| Exp.5 | S32 | 30 | Female | Left |
| Exp.5 | S14 | 27 | Female | Right |
| Exp.5 | S21 | 21 | Female | Right |
| Exp.5 | S23 | 28 | Female | Right |
| Exp.5 | S33 | 24 | Male | Right |
| Exp.5 | S24 | 27 | Male | Right |
| Exp.5 | S34 | 29 | Female | Right |
| Exp.5 | S26 | 24 | Female | Right |

**UNIVERSITY OF WASHINGTON**
**SUBJECT SURVEY**

**Analysis of the Impact Adversarial Attacks Have of Teleoperated Procedures**

**Experiment number:** _____

**Experimental order:** _____

**Trial 1:**

**1. How easy/difficult was it to reach the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**2. How easy/difficult was it to grab the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. How easy/difficult was it to move between the pick-up location to the put-down location**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**4. How easy/difficult was it to transfer the block from one hand to the other?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. Overall, how easy/difficult was it to perform tasks using this system?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

---

**Trial 2:**

**1. How easy/difficult was it to reach the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**2. How easy/difficult was it to grab the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. How easy/difficult was it to move between the pick-up location to the put-down location**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**4. How easy/difficult was it to transfer the block from one hand to the other?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. Overall, how easy/difficult was it to perform tasks using this system?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**Trial 3:**

**1. How easy/difficult was it to reach the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**2. How easy/difficult was it to grab the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. How easy/difficult was it to move between the pick-up location to the put-down location**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**4. How easy/difficult was it to transfer the block from one hand to the other?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. Overall, how easy/difficult was it to perform tasks using this system?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

---

**Trial 4:**

**1. How easy/difficult was it to reach the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**2. How easy/difficult was it to grab the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. How easy/difficult was it to move between the pick-up location to the put-down location**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**4. How easy/difficult was it to transfer the block from one hand to the other?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. Overall, how easy/difficult was it to perform tasks using this system?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**Trial 5:**

**1. How easy/difficult was it to reach the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**2. How easy/difficult was it to grab the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. How easy/difficult was it to move between the pick-up location to the put-down location**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**4. How easy/difficult was it to transfer the block from one hand to the other?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. Overall, how easy/difficult was it to perform tasks using this system?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

---

**Trial 6:**

**1. How easy/difficult was it to reach the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**2. How easy/difficult was it to grab the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. How easy/difficult was it to move between the pick-up location to the put-down location**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**4. How easy/difficult was it to transfer the block from one hand to the other?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. Overall, how easy/difficult was it to perform tasks using this system?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**Trial 7:**

**1. How easy/difficult was it to reach the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**2. How easy/difficult was it to grab the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. How easy/difficult was it to move between the pick-up location to the put-down location**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**4. How easy/difficult was it to transfer the block from one hand to the other?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. Overall, how easy/difficult was it to perform tasks using this system?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

---

**Trial 8:**

**1. How easy/difficult was it to reach the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**2. How easy/difficult was it to grab the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. How easy/difficult was it to move between the pick-up location to the put-down location**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**4. How easy/difficult was it to transfer the block from one hand to the other?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. Overall, how easy/difficult was it to perform tasks using this system?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**Trial 9:**

**1. How easy/difficult was it to reach the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**2. How easy/difficult was it to grab the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. How easy/difficult was it to move between the pick-up location to the put-down location**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**4. How easy/difficult was it to transfer the block from one hand to the other?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. Overall, how easy/difficult was it to perform tasks using this system?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

---

**Trial 10:**

**1. How easy/difficult was it to reach the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**2. How easy/difficult was it to grab the block?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. How easy/difficult was it to move between the pick-up location to the put-down location**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**4. How easy/difficult was it to transfer the block from one hand to the other?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

**3. Overall, how easy/difficult was it to perform tasks using this system?**
(easy) 0 1 2 3 4 5 6 7 (difficult)

# VITA

Tamara was born in Sisak, Croatia. She earned her B.Sc. in Electrical Engineering from the University of Zagreb, Croatia in 2008, and her M.Sc. in Electrical Engineering from the University of Washington in 2012. Between degrees, she worked as a software engineer in Vancouver, BC. Tamara's research focuses on security and privacy of biomedical cyber-physical systems. In order to prevent potential security and privacy against these emerging cyber-physical technologies, she has collaborated with an interdisciplinary team, involving bioengineers, medical practitioners, legal scholars, philosophers and artists. Tamara also enjoys teaching, and she has taught many graduate and undergraduate courses across multiple departments at the University of Washington. Finally, Tamara feels strongly about diversity, and she advocates that diversity, both in life and in engineering, is crucial for success.