SoK: Security and Privacy in Implantable Medical Devices and Body Area Networks

Michael Rushanan, Aviel D. Rubin  
Computer Science  
Johns Hopkins University  
Baltimore, MD, USA  
michar1@cs.jhu.edu, rubin@cs.jhu.edu

Denis Foo Kune, Colleen M. Swanson  
Computer Science and Engineering  
University of Michigan  
Ann Arbor, MI, USA  
foo@eecs.umich.edu, cmswnsn@umich.edu

Abstract—Balancing security, privacy, safety, and utility is a necessity in the health care domain, in which implantable medical devices (IMDs) and body area networks (BANs) have made it possible to continuously and automatically manage and treat a number of health conditions. In this work, we survey publications aimed at improving security and privacy in IMDs and health-related BANs, providing clear definitions and a comprehensive overview of the problem space. We analyze common themes, categorize relevant results, and identify trends and directions for future research. We present a visual illustration of this analysis that shows the progression of IMD/BAN research and highlights emerging threats. We identify three broad research categories aimed at ensuring the security and privacy of the telemetry interface, software, and sensor interface layers and discuss challenges researchers face with respect to ensuring reproducibility of results. We find that while the security of the telemetry interface has received much attention in academia, the threat of software exploitation and the sensor interface layer deserve further attention. In addition, we observe that while the use of physiological values as a source of entropy for cryptographic keys holds some promise, a more rigorous assessment of the security and practicality of these schemes is required.

I. INTRODUCTION

The integration of computing devices and health care has changed the landscape of modern medicine. Implantable medical devices (IMDs), or medical devices embedded inside the human body, have made it possible to continuously and automatically manage a number of health conditions, ranging from cardiac arrhythmia to Parkinson's disease. Body area networks (BANs), wireless networks of wearable computing devices, enable remote monitoring of a patient’s health status.

In 2001, the estimated number of patients in the United States with an IMD exceeded 25 million [1]; reports from 2005 estimate the number of patients with insulin pumps at 245,000 [2], [3]. IMDs have become pervasive, spurred by the increased energy efficiency and low cost of embedded systems, making it possible to provide real-time monitoring and treatment of patients [4]. Low power system optimizations [5], ultra-low-power wireless connectivity [6], and the development of numerous lightweight communication protocols (e.g., on-demand MAC) [7]–[9] have helped make small-scale sense-actuate systems like IMDs and BANs a reality. Through sensors, these systems can collect a range of physiological values (e.g., heart rate, blood pressure, oxygen saturation, temperature, or neural activity) and can provide appropriate actuation or treatment (e.g., regulate heart rate or halt tremors). On-board radios enable wireless data transfer (or wireless medical telemetry [10]) for monitoring and configuration without sacrificing patient mobility or requiring surgical procedures to physically access the devices.

The need for security and privacy of medical devices has received increasing attention in both the media and the academic community over the last few years—a perhaps telling example is the recent revelation that Vice President Dick Cheney had the wireless telemetry interface on his implanted pacemaker disabled [11]. In the academic community, the seminal work by Halperin et al. [12], which introduces a class of wireless threats against a commercial implantable cardiac defibrillator (ICD), has been followed by numerous papers researching techniques to improve the security and privacy of medical devices.

Even though the likelihood of targeted adversarial attacks on IMDs and BANs may be debatable, the consequences of an insecure system can be severe. Indeed, Fu and Blum [13] observe that while the hacking of medical devices is a “red herring”, poor security design can result in real vulnerabilities. For example, the existence of malware on networked medical devices can result in unreliable data or actuation, impacting both the integrity and availability of the systems in question. Any private data on the system may be exposed, leading to a breach of confidentiality.

Although traditionally there has been little incentive for medical device manufacturers to incorporate security and privacy mechanisms for fear of inhibiting regulatory approval [14], the FDA has recently called for manufacturers to address cybersecurity issues relevant to medical devices for the entire life cycle of the device, from the initial design phase through deployment and end-of-life [15]. Although these calls are in the form of draft guidelines for ensuring appropriate medical device security, there is evidence that the FDA means to use these guidelines as grounds for rejection of premarket medical device submissions [16].

Ensuring security and privacy in the context of safety-critical systems like IMDs, however, is more nuanced
than in the traditional computer science setting. As Halperin et al. [17] observe, the security and privacy goals of IMDs may at times conflict with the safety and utility of these devices. For example, eavesdropping on communications between an IMD and its programmer may reveal a sensitive medical condition, or querying an IMD with an unauthenticated programmer may allow clandestine tracking, both of which compromise the privacy of the affected patient. Unauthenticated communication can lead to denial of service attacks, in which legitimate communication is prevented from reaching the device or the device’s battery is needlessly depleted [12], as well as replay and injection attacks, in which potentially dangerous commands sent to the device can alter the patient’s therapy [12], [18], [19]. On the other hand, using traditional cryptographic mechanisms to ensure secure communication and storage of data can compromise the safety of the patient. If the patient needs treatment outside of his normal health care context (e.g., at the emergency room), it is necessary for health care professionals to have the ability to identify and access the IMD in order to diagnose and treat the patient.

Balancing security, privacy, safety, and utility is a necessity in the health care domain [14]. Multiple academic disciplines (e.g., embedded systems, computer security, and medicine) have independently explored the IMD/BAN problem space. We go beyond related work [17], [19], [20] by providing a comprehensive overview of security and privacy trends and emerging threats, in order to facilitate uptake by research groups and industry.

Moreover, we provide a more formal adversarial model and classification of threats than the work of Halperin et al. [17] and Zhang et al. [20]. By identifying and analyzing popular research trends in this space, we observe that current work may be roughly subdivided into three classes: the security of the wireless telemetry, detection and prevention of software vulnerabilities, and the security of the hardware architecture and sensor interface. Our categorization allows us to easily trace the evolution of IMD/BAN research, connect current work to related notions from the field of RFID security and privacy, and identify emerging threats in this space.

We identify challenges computer science researchers face in examining the security and privacy of medical devices, including the lack of reproducibility of research results. Access to medical devices is a common problem that limits researchers’ ability to validate prior results; food-grade meat as a phantom also complicates reproducibility due to its inaccurate approximation of a human body [8], [21]. In addition, we provide clear definitions of IMDs and BANs and describe the relevant communications standards, including clarifying the term medical device, which is strictly defined by the FDA. The distinction between a medical device and a device used in the context of health (e.g., FitBit, a popular tool to track physical activity) is a common source of confusion.

In the IMD/BAN space, we need to achieve trustworthy communication, trustworthy software, and trustworthy hardware and sensor interfaces. While the security of the wireless telemetry interface has received much attention in academia, both the threat of software exploits in medical devices and the security and privacy of the sensor interface are areas of research that deserve further attention. Subtle eavesdropping and injection attacks on sensor inputs, such as the work by Foo Kune et al. [22] on cardiac implantable electrical devices (CIEDs), which include pacemakers and defibrillators, and Bagade et al. [23] on compromising the privacy of physiological inputs to key generation mechanisms, are a promising avenue of future work.

II. PAPER ORGANIZATION

We provide relevant definitions and background information on IMDs and BANs in Section III and outline security and privacy goals and our adversarial model in Section IV. In Section V, we give a breakdown of the state of the art in IMD/BAN research and analyze current trends. We then discuss research challenges specific to the IMD/BAN domain and identify emerging threats in Section VI. We give concluding remarks in Section VII.

III. BACKGROUND AND DEFINITIONS

Advances in embedded systems [24] and wireless sensor networks (WSNs) [25] have made modern IMDs and BANs possible. Current embedded systems trade computing performance and memory resources for energy efficiency and lower costs. Wireless sensor networks link both homogeneous and heterogeneous autonomous devices. WSNs have been used for health care monitoring via the introduction of both wearable and implanted sensor networks [5], [26], giving rise to modern healthcare-related BANs.

A. Implantable Medical Devices and Body Area Networks

The U.S. FDA has a broad, albeit relatively strict, definition of medical devices, which range from tongue depressors to MRI machines. The U.S. Federal Food Drug & Cosmetic Act [27, Section 201(h)] defines a medical device as an instrument, apparatus, machine, or other similar article which is a) officially recognized by national registries; b) intended for use in the diagnosis, cure, or prevention of a disease; and c) intended to affect the structure or function of the body. We emphasize that in order for a device to qualify as a medical device, it must undergo substantial review by the FDA before being released on the commercial market; we use this definition of medical device in this paper. The FDA also has significant global influence through arrangements with numerous foreign government organizations [28]; therefore devices, standards, and protocols used in the U.S. are likely to be of interest to other countries as well.

The U.S. Federal Communications Commission (FCC) defines wireless medical telemetry in FCC 00-211 [29, Section 3B] and FCC 47 CFR 95.401 [10] as the measurement
Cochlear implant
Cardiac defibrillator
Insulin pump
Gastric simulator

Neuro-stimulator

Figure 1. Example IMDs and ICD/Programmer communication.

Note: The text contains an illustration of various IMDs and the communication protocol between an IMD and its programmer.

1) Implantable medical devices: We define an implantable medical device (IMD) as one which is surgically placed inside of a patient’s body. Figure 1 provides examples of IMDs and an IMD programmer (or simply, programmer), and shows the high-level communication protocol of an ICD. The programmer in this context is an external device with an interface (usually a radio frequency (RF) transceiver) for communicating wirelessly with an IMD and relaying data to a device used by clinicians or other health care providers. An IMD system supports:

- Analog front end, the signal conditioning circuitry for application-specific sensing and actuation;
- Memory and storage, for storing personal health information and sensed data;
- Microprocessor, for executing device-specific software;
- Telemetry interface, often radio-based, for transmitting data between the device and a programmer or other sensor/actuator on the patient; and
- Power management, for monitoring and managing battery use for increased longevity.

IMDs are resource-constrained, requiring reduced size, weight, low peak power and low duty cycle. Past research uses resource-constrained hardware platforms such as an 8-bit Atmel-AVR and a 16-bit TI MSP430 [30] to model IMD configurations. The TI MSP430F1611 consumes energy at approximately 0.72 nJ per clock cycle. Typical IMDs are designed to last 90 months on a single battery with 0.5 A h to 2 A h of battery life [31]. These requirements minimize the impact of invasive surgeries to replace depleted implants.

Furthermore, modern IMDs rely on low-power radio communication and network connectivity to provide a remote-monitoring system [14]. The FCC has allocated the 401 MHz to 406 MHz band for Medical Devices (MedRadio) [32], sometimes called the Medical Implant Communication Service (MICS) band. This band is currently used for IMD wireless telemetry.

The MICS band allows for reasonable signal propagation through the human body without interfering with other devices. Additionally, it allows for a greater distance between the patient and external transceiver, unlike previous IMDs (e.g., a pacemaker transmitting at 175 kHz, which required a proximity within 5 cm [9]).

2) Body area networks: We define a body area network (BAN) as a wireless network of heterogeneous computing devices that are wearable. This network enables continuous remote monitoring of patient physiological values in the medical setting. In this work, we are mainly concerned with BANs as they relate to IMDs.

BANs typically include three types of devices: sensors, actuators, and a sink. In Figure 2, sensors are placed at various locations on the body, support multiple network topologies, and forward sensed data to a more computationally powerful device (e.g., a smartphone). Although related to wireless sensor networks, BANs exhibit some notable differences [33] with respect to wearability (e.g., size and power), battery availability, and transmission (i.e., the human body is a lossy medium). Moreover, reliability requirements may be stricter than in a typical wireless sensor network, depending on how safety-critical the application.

As we are most interested in BANs as they relate to IMDs, we only give a brief overview of the communication standards for clinical environments [34]. The ISO/IEEE 11073 [35] standard spans the entire BAN communication stack, while Health Level 7 (HL7) [36], Integrating the Health Enterprise (IHE) [37] and the recent ASTM F2761 (MDPnP) [38] standard only describe the
application layer. While at least some security mechanisms are mentioned in these standards, most are optional, presumably to ensure interoperability. Foo Kune et al. [34] find that by enabling these security mechanisms in combination with known security protocols, a vast majority of security requirements could be satisfied. The Association for the Advancement of Medical Instrumentation (AAMI) is working on TIR-57, a draft guidance document to start standardizing secure Information Technology (IT) practices for clinical environments.

IV. SECURITY AND PRIVACY IN IMDs AND BANs

In this section, we first review security and privacy goals for IMDs and BANs. We then present our adversarial model and discuss security threats.

A. Security and Privacy Goals

We recognize the following security goals for IMDs and BANs, building on the models provided by Halperin et al. [17], Burleson et al. [14], and Zhang et al. [20]. These properties should hold throughout the entire life cycle of the IMD/BAN devices, including appropriate disposal of explanted devices.

- **Confidentiality**: Data, device information, and device system structures should be accessible only to authorized entities (i.e., appropriate entities) and these entities should be authenticated (i.e., the identity of entities communicating with devices should be verifiable). The system should also satisfy data origin authentication (i.e., the source of any received data should be verifiable). In particular, data should be kept confidential both in storage and while in transmission.

- **Integrity**: Data, device information, and device system structures should not be modifiable by unauthorized entities.

- **Availability**: Data, device information, and device systems should be accessible when requested by authorized entities.

IMDs and BANs should also satisfy the following privacy goals; we include criteria from Halperin et al. [17], Denning et al. [39], and Kumar et al. [40] for completeness. Although these goals bear some overlap with confidentiality, we include the full list in order to allow for a more comprehensive treatment of privacy (apart from security) in the context of IMDs and BANs.

- **Device-existence privacy**: Unauthorized entities should not be able to determine that a patient has an IMD/BAN.

- **Device-type privacy**: If device-existence privacy is not possible, unauthorized entities should not be able to determine what type of IMD/BAN is in use.

- **Specific-device ID privacy**: Unauthorized entities should not be able to determine the unique ID of an IMD/BAN sensor.

- **Measurement and log privacy**: Unauthorized entities should not be able to determine private telemetry or access stored data about the patient. The system design phase should include a privacy assessment to determine appropriate policies with respect to data access.

- **Bearer privacy**: Unauthorized entities should not be able to exploit IMD/BAN properties to identify the patient.

- **Tracking**: Unauthorized entities should not be able to leverage the physical layer (e.g., by monitoring analog sensors or matching a radio fingerprint [41]–[43]) to track or locate a patient.

B. Adversarial Model

Following the standard approach in computer security literature, adversaries may be distinguished based on their goals, capabilities, and relationship to the system in question. We have the following classification criteria.

1) An adversary is either active or passive:

- Passive adversaries are able to eavesdrop on all communication channels in the network, including side channels, or unintentional communication channels.

- Active adversaries are able to read, modify, and inject data over the communication channel.

2) An adversary is either an external or internal entity with respect to the system. That is, an adversary may either be an outsider or an insider with a legitimate system role (e.g., manufacturer employees, patient, physician, or hospital administrator).

3) An adversary may be either a single entity or a member of a coordinated group of entities.

4) An adversary may be sophisticated, relying on specialized, custom equipment, or unsophisticated, relying only on readily available commercial equipment.

All system components of IMDs and BANs may be used as attack surfaces, or points of potential weakness, by an adversary (e.g., any existing sensors, actuators, communication networks, or external programming devices). In addition, the adversary may have the following targets and goals with respect to the specified target.

1) The patient: The adversary may wish to obtain private information concerning the patient (e.g., whereabouts, diagnosis, or blackmail-worthy material), or cause physical or psychological harm to the patient.

2) The device or system manufacturer: The adversary may wish to engage in corporate espionage or fraud.

3) System resources: The adversary may wish to utilize system resources and may be unaware of the type of device or network compromised. That is, the adversary does not knowingly target an IMD/BAN.

---

1 At the time of this writing, a public version of the AAMI TIR-37 draft was not yet available.
C. Threats

We classify IMD and BAN security and privacy threats found in the literature into the following categories:

- **The telemetry interface**, which is typically wireless. Threats include a passive adversary who eavesdrops on wireless communications and an active adversary who attempts to jam, replay, modify, forge, or drop wireless communications.
- **Software threats**, which consider an adversary that can alter the logic of the system (e.g., through software vulnerabilities) to affect expected operation.
- **Hardware and sensor interface threats**. An adversary may have knowledge of the internal hardware architecture or analog sensors and may use that knowledge to attack the system. Specifically, sensor threats stem from the implicit trust that the system places on those sensor inputs, under the assumption that physical contact with the sensor is necessary to alter the signal. An active attacker, however, may introduce remote interference to sensing in order to affect actuation.

These categories inform our analysis of security and privacy research trends in Section V.

V. MEDICAL DEVICE SECURITY AND PRIVACY TRENDS

We follow the broad categorization of IMD and BAN security and privacy threats given in Section IV-C in order to analyze research trends in the literature. That is, we group research according to the relevant attack surface: the telemetry interface, software, and hardware/sensor inputs. We give an explicit categorization of relevant research with respect to security threats and goals in Table I. Due to the large amount of work on the wireless telemetry threats, we separate the wireless threats into subclasses. An overview of current research, grouped thematically and by publication year, is given in Figure 3.

As Figure 3 indicates, the vast majority of results in the literature focus on threats to the telemetry interface, while a limited number of papers consider software threats. Since very few papers deal with threats to the sensor interface, we defer discussion of this emerging threat to Section VI-C.

A. Securing the Wireless Telemetry Interface

Halperin et al. [12] introduce a class of wireless threats against a commercial ICD; since then, attacks on the telemetry interface of IMDS have received a large amount of attention [18], [77], [78]. At the physical layer, Halperin et al. [12], targeting an ICD, and Li et al. [18], targeting an insulin pump system, develop passive and active attacks against their respective device using an off-the-shelf software defined radio (SDR) platform. In the devices and programmers analyzed, the communication links do not use an authenticated channel and transmit unencrypted data without freshness checks, thereby allowing eavesdropping, replay [12], and injection attacks [18].

Unsurprisingly, many authentication techniques have been proposed to secure the wireless telemetry of IMDS and BANs, including the use of biometrics, distance-bounding authentication, out-of-band authentication, external devices, and anomaly detection. We explore each of these areas individually below.

1) Biometrics: Popular techniques for key generation and key agreement in IMDS/BANs include the use of biometrics, or physiological values (PVs) [44], [46], [55], [56], [59], [63], [67], [69]–[71], [73]. Electrocardiograms (ECGs) are a common choice as a source of key material in these protocols, although other PVs such as heart rate, blood glucose, blood pressure, and temperature have been proposed [70].

The choice to use ECGs is motivated by a well-cited paper by Poon et al. [71], which asserts that the time between heartbeats, or interpulse interval (IPI), has a high level of randomness. IPI has the additional benefit that it can be measured anywhere on the body and many IMDS in use today can measure IPI without modification.

A typical approach to PV-based key agreement between an IMD and programmer, for example, involves both devices taking a measurement of the chosen PV. This measured PV is used to generate a cryptographic key that is agreed upon by both devices, which is then used to establish an authenticated channel. The basic assumption is that physical contact (or at least physical proximity) with the patient is required in order to precisely measure the chosen PV.

Security analyses of these protocols have been mostly ad hoc in nature, however, and in general more comprehensive assessments are required. For example, Rostami et al. [19] demonstrate simple, but damaging attacks against OPFKA [46] and IMDGuard [61], which we discuss in Section V-A4.

Chang et al. [51] also explore the use of IPI, drawing attention to the issue of noise in real-world measurements. Later work by Rostami et al. [44] presents a more robust IPI-based authentication protocol, which unlike previous work, takes into account both the impact of measurement noise and provide a more rigorous security analysis. We discuss the subtleties and potential difficulties of using IPI as part of a key agreement protocol in more detail in Section VI-B and Section VI-C.

2) Distance-Bounding Protocols: Distance bounding [79] is a technique that establishes physical distance between two entities by timing the delay of sent and received transmissions. This distance bound can be computed over various signals such as RF or ultrasonic sound (which is an acoustic signal above 20 kHz). A number of IMDS/BAN access control and authentication protocols use distance bounding [18], [47], [51], [53], [64]. However, distance bounding by itself provides for only weak authentication, in which physical proximity between devices is established but identity and authorization are not, thereby requiring the use of additional authentication techniques.
Figure 3. Trends in Security and Privacy Research on IMDs/BANs.
Table I
IMD AND BAN SECURITY AND PRIVACY THREATS AND DEFENSES

<table>
<thead>
<tr>
<th>Threat</th>
<th>Attacks</th>
<th>Confidentiality</th>
<th>Integrity</th>
<th>Availability</th>
<th>Privacy</th>
<th>Safety</th>
<th>Defenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wireless eavesdropping</td>
<td>[12], [18], [49]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>[12], [18], [44], [45], [46], [47], [50], [51], [53], [55], [58], [59], [61], [63], [64], [67]–[75]</td>
</tr>
<tr>
<td>Wireless modification</td>
<td>[12], [18], [19]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>[12], [18], [44], [45], [47], [50], [51], [53], [55], [58], [59], [61], [63], [64], [67]–[75]</td>
</tr>
<tr>
<td>Wireless replay</td>
<td>[12], [18]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>[12], [18], [44], [45], [47], [50], [51], [53], [55], [58], [59], [61], [63], [64], [67]–[75]</td>
</tr>
<tr>
<td>Wireless jamming</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[61], [68]</td>
</tr>
<tr>
<td>Analog sensor injection</td>
<td>[22]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[22]</td>
</tr>
<tr>
<td>Battery depletion</td>
<td>[12]</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>[12], [58], [62], [68]</td>
</tr>
<tr>
<td>Protocol Design Flaws</td>
<td>[12], [18], [19], [23], [49], [60]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>Software Flaws</td>
<td>[76]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>[57], [76]</td>
</tr>
<tr>
<td>Side channels</td>
<td>[23], [54], [60]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>[54]</td>
</tr>
</tbody>
</table>

A typical distance-bounding protocol between a programmer and IMD, for example, involves the programmer proving to the IMD that it is physically close (e.g., within 3 cm). Rasmussen et al. [64] use ultrasonic sound signals to compute the distance bound of a programmer and IMD, since it is impossible for an attacker to send audio data that propagates faster than the speed of sound. Shi et al. [47], [53] use received signal strength (RSS) variation to differentiate BAN devices on the same body from external signals (i.e., attacker transmissions). This technique relies on the observation that the RSS variation between two BAN devices on the same body is more stable than the RSS between an on-body device and an external device. Jurik et al. [56] make use of ECG signals to establish the continued proximity of an authenticated mobile device to a user.

Distance bounds are also computed over body-coupled communication (BCC). BCC uses the human body as a transmission medium, requiring physical proximity to the patient in order to communicate. Li et al. [18] introduce wireless attacks against BCC and find that both passive and active attacks are mitigated for distances greater than 0.5 m. Chang et al. [51] inject artificial signals through the patient’s body to authenticate BAN devices on the same body. These signals, however, only achieve an estimated 0.469 to 5.429 bits per hour, making this technique impractical.

In the related field of RFID, system implementations have inaccurately assumed distance-bounding guarantees as a result of short read ranges (e.g., 10 cm). Kfir et al. [80] introduce a relay attack in which two coordinated adversaries fool an RFID reader into believing that the RFID tag is nearby. Relay attacks can be mitigated with context-aware communication [81], a method which requires the user to perform an uncommon, but easily repeatable movement in order to be authenticated. The applicability of this defense to IMDs is debatable, however, because a patient may not be able to authenticate in the event of a medical emergency.

Cremers et al. [82] provide a classification of distance-bounding attacks that assumes weak authentication, suggesting additional evaluation is required before such protocols are used in the medical setting; the adversarial capabilities necessary to launch these attacks are included in our model. Cremers et al. use the terminology verifier and prover to describe the participants in distance-bounding protocols; the verifier establishes physical proximity to the prover. The attacks consider various adversarial capabilities for falsifying physical proximity to the prover. Specifically, the adversary may modify transmissions between a verifier and prover. He may introduce his own dishonest prover, or he may collude with other dishonest entities. Lastly, he may also exploit honest provers (e.g., by first allowing the prover to establish physical proximity, then jamming subsequent prover transmissions and authenticating in the prover’s stead).

3) Out-of-Band (OOB) Authentication: OOB techniques make use of auxiliary channels, such as audio, visual, and tactile, that are outside the established data communication channel [12], [39], [72], [83]. Using auxiliary channels for authentication obviates the need for trusted third parties and key pre-distribution schemes. A common assumption in these schemes is that the chosen out-of-band channel is resistant to eavesdropping attacks.

Halperin et al. [12] propose an OOB authentication scheme that uses a low-frequency audio channel. The basic idea is that the IMD uses a zero-power RFID device to generate a random key and transmit it over the audio channel. The patient is alerted when a key exchange occurs through vibrations produced by a piezo element connected
to the RFID device. The programmer, at a distance of no more than 0.6 m to 0.9 m [60], listens for the key and then establishes a secure authenticated channel with the IMD.

Halevi et al. [60] examine a passive adversary with the ability to deploy (or otherwise make use of) a general-purpose microphone (e.g., PC microphone) in the vicinity of the IMD/programmer communication. Halevi et al. show that although the measured piezo sound accuracy varies with distance, the average key retrieval correctness at 0.9 m, computed for multiple supervised methods, is as high as 99.88%. This contradicts Halperin et al.’s [12] earlier experimental result, which indicates the audio channel is resistant to eavesdropping.

Alternatively, Denning et al. [39] and Li et al. [72] opt for visual OOB authentication. Denning et al. propose the use of ultra-violet or visible tattoos to record permanent IMD keys. This mechanism allows emergency authentication, but does not allow for key revocation and may suffer from usability concerns [39]. Li et al. [72] require the users to visually inspect simultaneous LED blinking patterns in order to achieve authentication in BANs. The usability of this scheme is unclear and it is unlikely to be appropriate for emergency scenarios, so its applicability to IMDS is limited.

4) External Wearable Devices: A unique approach to securing IMD/BAN telemetry makes use of external devices worn by the patient. The basic idea is that this external device mediates communication with the IMD, thereby providing both confidentiality for transmitted data and protection against unauthorized communication. One concern with the use of such devices is their acceptability to the patient, however. Denning et al. [39] treat this issue in some detail and study the usability of several possible authentication methods, including external devices and password tattoos.

Denning et al. [68] propose an external device, called the cloaker, that proxies authorized communication to the IMD. If the cloaker is absent, the IMD communicates openly (e.g., in case of a medical emergency, the cloaker fails open). A malicious programmer can exploit this fail-open behavior by selectively jamming the cloaker or otherwise convincing the IMD of the cloaker’s absence, so Denning et al. suggest additional mitigation techniques to prevent such an attacker from communicating with the IMD.

Gollakota et al. [58] and Xu et al. [61] use friendly jamming to protect IMD communication, which uses jamming constructively to prevent unauthorized communication. IMDGuard [61] employs an external wearable device, called the Guardian, to enable access control and confidential data transmissions. The Guardian first authenticates the programmer and then uses an ECG-based key agreement mechanism to authenticate itself to the IMD. Temporary keys can then be issued to allow a secure channel between the programmer and the IMD. In the event that an attacker jams the messages from the Guardian device to the IMD, the Guardian initiates an active defense by jamming all IMD transmissions. However, IMDGuard has the disadvantage of requiring modifications to the IMD itself (which is difficult in practice with respect to already-deployed devices) and the suggested ECG-based key agreement scheme suffers from security flaws. Rostami et al. [19] show a simple man-in-the-middle attack that reduces the effective key length from 129 bits to 86 bits. This attack takes advantage of a protocol flaw in the second round of reconciliation (in which the two parties verify they know the same key), which can be spoofed to reveal one bit per block.

The shield [58] works by listening for and jamming all IMD transmissions and unauthorized commands. Given the shield’s proximity and jamming power, the assumption is that only the shield can cancel out its own jamming signal and decode IMD transmissions. This design mitigates both passive and active wireless attacks, but the security of the system relies on the assumption that an attacker whose distance from the IMD is greater than the distance between the IMD and the shield will be unable to recover IMD transmissions, even if the attacker is equipped with multiple input and multiple output (MIMO)-systems and directional antennas. Tippenhauer et al. [49] challenge this assumption, however, and show that MIMO-based attacks are possible in the presence of an adversary with two receiving antennas from distances of up to 3 m.

5) Anomaly Detection: Anomaly detection attempts to automatically identify resource depletion and malicious communication, as well as distinguish between safety and security events [45], [50], [62]. This is generally achieved by observing patterns over time, such as physiological changes or IMD access patterns (e.g., programmer commands, date, or location).

Hei et al. [62] obtain and use normal IMD access patterns as training data for their supervised learning-based scheme. The resultant classification is used to identify anomalous IMD access in real time. That is, Hei et al.’s method tries to detect abnormal access attempts and block such authentication from proceeding, before any expensive computations take place. In this way, the IMD is protected against denial of service attacks that deplete the system’s resources. This scheme is designed for non-emergency settings, however, and Hei et al. recommend that either the IMD automatically detect emergency conditions and fail open, or that hospitals have access to a master device key. The feasibility and security provided by these two approaches is not considered.

Another anomaly detection approach makes use of audits; Henry et al.’s scheme [45] observes correlated physiological changes when an insulin bolus is administered by tracking acoustic bowel sounds. These observations are recorded as an audit log for retroactive verifiability of intended system execution. While useful, a limitation of passive anomaly detection is that such schemes do not provide medical device integrity, and so need to be used in conjunction with another mechanism that protects communications.
At the physical layer, wireless transmissions from an attacker are likely to deviate in physical characteristics from legitimate programmer transmissions. Zhang et al. [50] propose a medical security monitor, MedMon, which is an external device that detects anomalous transmissions by examining physical characteristics of the transmitted signal; such characteristics include received signal strength, time of arrival, differential time of arrival, and angle of arrival. When an anomalous transmission is detected, MedMon can initiate either a passive defense (e.g., by alerting the patient) or an active defense (e.g., by blocking the transmissions from reaching the medical device).

The characteristics of the device used for anomaly detection (and any associated audit logs) have important implications for the overall security of the system. Suggested anomaly detection implementations make use of dedicated devices, such as analog sensor systems [45], or extend the functionality of personal devices, such as smartphones [50], [62]. Offloading heavy computation to another device like a smartphone might improve the IMD’s battery life, but significantly increases the attack surface, as malware on mobile devices is common [84]. In addition, regulatory barriers for medical devices may make this approach difficult.

B. Software Threats

Software running on medical devices spans a wide range of complexity. An increasing number of medical devices are reliant on digital circuits controlled by software, rather than analog circuits. Faris [85] notes that in 2006, a major milestone was crossed when over half of deployed medical devices contained software. So far there has been a lack of detailed analysis of IMD software. However, there have been efforts to verify proper functionality by simulating an artificial heart to interface with cardiac pacemakers [57], [86]. Although these testing methods are not directly tailored to security, the tests reduce software bugs and may therefore reduce possible software vulnerabilities.

Devices communicating over a BAN, in addition to their application code, have to include a telemetry interface that increases both the amount of code and the number of possible bugs. It is not surprising, then, that software is one of the main reasons for FDA recalls of computer-related issues [48]. Sandler et al. [87] report that in 2010, the FDA issued 23 recalls of defective devices, six of which were likely caused by software defects. Alemzadeh et al. [48] report that the percentage of computer-related recalls between 2006 and 2011 was between 30% to 40%. In this study, software defects are found to be the cause of 33% of computer-related class I recalls (reasonable chance of patient harm), 66% of class II recalls (temporary or reversible adverse effects), and 75% of class III recalls (non-compliant, but unlikely to cause harm).

Bugs in medical devices have been a cause of over 500 recalls reported between 2009 and 2011 by the FDA [52]. While there exists no method to extrapolate from the reported bugs to those existing in deployed devices, the number reported is most likely only a lower bound. Fu reports that failures in medical device software often result from a failure to apply known system engineering techniques [88], indicating that the problem is partially solvable today.

Moreover, the presence of a telemetry interface on the device may expose software bugs to a remote attacker. Evidence of the brittleness of software implementations is apparent when investigating security vulnerabilities, including those in proprietary firmware. Hanna et al. [76] perform the first public software security analysis of an automatic external defibrillator (AED). By reverse engineering the device, the authors successfully target three software packages responsible for programming device parameters, collecting post-cardiac device data, and updating the AED. The authors locate four vulnerabilities, one of which enables arbitrary code execution on the device.

The need for secure coding practices for safety-critical devices is clear. However, closed source for medical devices makes it challenging to run a static analyzer on the source code, let alone obtain the firmware. With proprietary protocols and the special MICS band used on the wireless telemetry interface, traditional fuzzing tools such as Peach Fuzzer [89] have not developed modules appropriate for testing medical devices.

A related security vulnerability is the existence of malware on medical devices. Regardless of whether the intent of the attacker is to compromise a medical device, malware can significantly impact the performance and reliability of safety-critical devices such as IMDs [13].

VI. RESEARCH CHALLENGES AND EMERGING THREATS

In this section, we identify and address challenges computer science researchers face in examining the security and privacy of medical devices and discuss promising areas for future work. In particular, we discuss common problems, identifying partial solutions and highlighting areas where further work is needed. A particularly difficult issue is the lack of reproducibility of research results in this field; given the safety-critical nature of IMDs and some BANs, it is critical that proposed attacks and defenses be thoroughly and independently evaluated in order to accurately assess risk of the attack and efficacy of the defense. A second area of concern, which we discussed briefly in Section V-A, is the use of physiological values to secure IMDs/BANs. The evaluations in the literature are limited in scope, partially because of the lack of availability of appropriate data sets for use by researchers and partially because the focus has been on protocol design rather than on a rigorous assessment of the use of biometrics for cryptographic key establishment.

We first address issues related to reproducibility in Section VI-A, before moving to a discussion of the use of physiological values in Section VI-B.
A. Reproducibility challenges

Lack of access to devices is a common problem; access to medical devices is either non-existent or limited to older, end-of-life models that have been received from patients, relatives, or physicians. The ICD that Halperin et al. [12] study, for example, is a model introduced to the market five years earlier. Without access to the devices themselves, researchers are necessarily limited in their ability to analyze potential attacks and defenses; often device hardware configurations are not public knowledge. Research results from groups that have managed to acquire and study particular IMDs are not likely to be validated by others, if only because of lack of equipment. While there have been some efforts to provide access to medical devices [90], direct access to devices from manufacturers by the security research community appears to be limited at present.

A second issue in computer security and privacy experiments on medical devices is the use of food-grade meat as a phantom, or human tissue simulator [12], [49], [58]. As Clark and Fu [21] observe, this method does not lead to reproducible experiments, possibly due to the introduction of uncontrolled variables that can affect the impedance of the tissue or propagation of signals in the phantom. Instead, researchers should use a calibrated saline solution at 1.8 g/L at 21 °C [91, Table 10, p. 30] with electrodes to inject the appropriate simulated physiological signals. The complete design is described in the ANSI/AAMI PC69:2007 standard [91, Annex G]; this is the accepted standard for electromagnetic compatibility of medical devices by researchers, device manufacturers, and regulators.

B. Physiological values as an entropy source

As mentioned in Section V-A1, the use of physiological values as a building block for security and privacy mechanisms is widespread in the literature. In particular, much research relies on the use of ECGs for security and privacy mechanisms. ECG measurements have been suggested for use in authentication [44], key establishment [55], [61], [71], and proximity detection [56] protocols (i.e., determining if one or more devices are in physical contact with the same body). Several systems have devices generate a shared secret key by reading the ECG signal through physical contact with the same person [23], [46], [55], [59], [61], [67], [74].

Most of these ECG-based mechanisms rely on the reported randomness of the IPI, or the amount of time between individual heartbeats [44], [61]; Rostami et al. [19], [44] suggest that sufficient entropy may be extracted from the least significant bits of properly quantized IPIs. There are some inconsistencies in the literature with respect to the quality of randomness it is possible to extract [64], [66], [70], however, and in studying this issue, researchers have been limited by a lack of sufficient real-world data. In particular, it is important to understand the impact of confounding factors such as health and age on the amount of entropy in IPI, in order to ensure that appropriate protocol parameters are chosen for entropy extraction.

In addition, Chang et al. [51] draw attention to the fact that the feasibility of these schemes relies on the ability of two devices to measure (and agree on) IPI in the presence of noise. Therefore, realizing such schemes may be more difficult using real-world data, rather than data collected in controlled environments (as measured by physicians with advanced medical equipment). Chang et al.’s results are indicative that measurement noise must be taken into account; later work by Rostami et al. [44] address this concern by taking into account and optimizing for these error rates.

Most evaluations have relied on an aggregation of heart rate databases from the MIT PhysioNet portal [92], which provides access to a large number of waveforms (collected by clinicians) ranging from healthy sinus rhythms to irregular heartbeat rhythms, or arrhythmias. Many suggested protocols are evaluated using either unspecified databases [23], [46], [55], [61], [67], [74] or arrhythmia databases [44], [59], [75], [93]. To extract random bits for a given record, the mean and standard deviation of the record are used to first quantize the bits, with a subset of the least significant bits treated as random. For example, Rostami et al. [44] quantize the IPI data into 8-bit representations and take the four least significant bits as random; the amount of entropy is estimated empirically using the classical definition of Shannon entropy (i.e., average entropy). A statistical battery of tests is then applied to the extracted bits—typically the (basic) subset of the NIST test suite [94] appropriate for the amount of data available.

Following the state of the art [95], [96], the assessment of a true random number generator (TRNG) for cryptographic purposes requires: a) an assessment of the quality of the entropy source itself (and a justification that the physical process being measured is random); b) an analysis of the efficiency and robustness of the extraction method (and the impact of the extraction method on the statistical properties of the TRNG); and c) cryptanalysis in the suggested use case (e.g., if an adversary can observe the entropy source or has an advantage in guessing future bits, this is not good for cryptographic use).

In particular, statistical analysis of the output of a TRNG, such as testing the output using the NIST test suites, is not sufficient to determine suitability for use in key agreement. The statistical properties of the physical phenomena need to be well-understood; properly quantizing the data and extracting bits that are close to uniform requires an accurate characterization of the distribution. For example, in the case of IPI, if the suggested methods for bit extraction do not ensure that the distribution characteristics used at time of authentication are accurate, the resulting bits may exhibit bias. We discuss the issue of observability of the IPI entropy source in more detail in the next section.
C. Emerging threats: sensors, remote attacks, and privacy

The traditional assumption with respect to IMDs and BANs is that many physiological signals stay within a patient’s body, limiting the exfiltration of data and the possibility for signal injection attacks. Recent studies, however, show that both are possible.

To date, the design constraints of IMDs have carefully dealt with the possibility of accidental electromagnetic interference, but do not consider the possibility of an active attacker. Recent work by Foo Kune et al. [22] shows that intentional interference at a CIED sensor interface is possible. By injecting a signal that mimics a cardiac waveform, Foo Kune et al. show that it is theoretically possible to alter the therapy delivered by the CIED, although the current range of this attack is very limited (on the order of a few centimeters). Reliance on sensor readings to achieve accurate and timely actuation, combined with increasingly sophisticated attacks, highlights the need to carefully consider adversarial capabilities and how best to achieve trustworthy systems.

Similarly, if the assumption that certain physiological signals stay within the human body is incorrect, both the security and privacy of schemes may be affected. For example, the use of physiological values as a source of entropy in key agreement schemes relies heavily on the assumption that it is not feasible for an adversary to observe the given biometric. A standard assumption in current literature is that the adversary cannot make physical contact with the target patient. In this sense, protocols that make use of physiological values to generate a shared key can be viewed as body-coupled communication protocols, whereby the key is transmitted via the human body. Although the assumption that an adversary does not have physical contact has merit in practice, we remark that this adversarial model neglects subtle classes of attacks by people known to the victim; ideally, new technologies should not enable “perfect crime” scenarios, even for the most sophisticated of attackers. As more and more people become active participants in (potentially insecure) BANs, moreover, it may be possible for a person close to the victim (i.e., with physical contact) to inadvertently aid a remote attacker (e.g., by leaking patient biometrics or performing signal injection attacks on sensors/wireless telemetry).

Remote attackers are also a concern today, especially with respect to observing physiological values assumed to be secret. Rostami et al. [44] and Chang et al. [51] both recognize the need to consider remote sensing of IPI. Rostami et al. attempt to extract IPI from video footage of the target, following work by Poh et al. [97] on the correlation between color fluctuations and IPI. Although Rostami et al. fail to replicate these results, other recent work in this area [98], [99] indicates that such attacks deserve further attention.

As a final remark, recent results in Bagade et al. [23] show that the ECG data of one person may be observable from another person’s physiological signals, if the two are in physical contact. That is, if two individuals touch, the ECG of one person is coupled to the EEG of the other person. We conclude that while the use of ECG (and other physiological values) as a security mechanism appears to hold some promise, cryptanalysis and entropy assessments need to be undertaken more rigorously.

A related area of research is the study of neurostimulators, which are IMDs designed to send electrical pulses to the nervous system, including the brain. These devices are used to treat conditions such as epilepsy, Parkinson’s, and obsessive compulsive disorder, with ongoing human trials exploring their efficacy in treating severe depression. Very little computer security and privacy research has been completed on these devices, and as the technology progresses, the need for further work in this area becomes more pressing. Denning et al. [65] give a brief overview of potential security and privacy implications with respect to neurostimulators, but concrete results in this area are lacking. A related question is explored by Martinovic et al. [54]: the authors’ side channel attacks in the context of brain-computer interfaces (BCIs), which measure and respond dynamically to a user’s brain activities, thereby allowing communication without words or gestures. Although the study is preliminary in nature, Martinovic et al.’s results support the hypothesis that personal information, such as passwords and whether or not a particular person is known to the target, may unintentionally leak through BCI use.

VII. Concluding Remarks

In this paper, we have given a cohesive narrative of security and privacy research in IMDs and BANs, analyzing current and emerging research trends: namely the security of the IMD/BAN telemetry and sensor interfaces and the need for trustworthy software. Our analysis in Section V-A shows that much attention has been paid to securing the telemetry interface and many useful approaches have been developed.

We have identified several areas for future work, such as the need for a more rigorous assessment of the use of physiological values as a source of entropy for cryptographic keys. As mentioned in Section V-B, the increasing complexity of software in IMDs and the history of FDA software-related recalls highlights the need for future work ensuring the trustworthiness of IMD and BAN software.

Finally, as discussed in Section VI-C, the possibility of EMI attacks on the sensor interface and eavesdropping on physiological signals formerly thought to be private is indicative of the need for a more nuanced approach to security and privacy research for medical devices. Computing devices that interface with the brain are becoming more advanced and more popular, both in the entertainment (in the form of BCI-integrated gaming) and health care industries (in
the form of neurostimulators). The ability to record and analyze brainwaves in real time using implanted computing devices that alter the brain's functionality has far-reaching implications for security and privacy, moving well beyond the traditional treatment of these topics in computer security.

ACKNOWLEDGMENTS

We thank our shepherd Srdjan Čapkun, Daniel Holcomb, Joel Van Der Woude, Amir Rahmati and the anonymous reviewers for their helpful comments. This work was supported by STARMent, a Semiconductor Research Corporation program, sponsored by MARCO and DARPA, the Dept. of HHS (SHARPS) under award number 90TR0003-01, and the NSF under award number CNS-1329737, 1330142.

REFERENCES


