Watermarking Based Sensor Attack Detection in Home Automation Systems

Henri Ruotsalainen
Institute of IT Security Research
St. Pölten University of Applied Sciences
St. Pölten, Austria
henri.ruotsalainen@fhstp.ac.at

Albert Treytl
Dept. for Integrated Sensorsystems
Danube University Krems
Wr. Neustadt, Austria
albert.treytl@donau-uni.ac.at

Thilo Sauter
Dept. for Integrated Sensorsystems
Danube University Krems
Wr. Neustadt, Austria
and *TU Wien*, Vienna, Austria
thilo.sauter@donau-uni.ac.at

2021 IEEE 26th International Conference on Emerging Technologies and Factory Automation (ETFA) | 978-1-7281-2989-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ETFA45728.2021.9613634

***Abstract*—We present a watermarking scheme for sensor data, which accurately detects sensor voltage manipulation attacks. Due to a low complexity implementation running locally on a microcontroller, the proposed technique is well suited for home automation systems but also IoT systems where light-weight sensor nodes with restricted communication capabilities are commonly applied. The watermarking technique is based on modulating the sensor supply voltage, which embeds the watermark signal into the sensor output voltage. Hence, invasive voltage tampering attempts including ADC grounding and voltage injection attacks can be efficiently detected. According to the performed characterization results, our method is able to distinguish normal operation and an attack condition with approx. 98% accuracy for the entire ADC input voltage range. In a real-world demonstration involving a thermistor based temperature sensor all attack attempts were successfully detected.**

***Index Terms*—Security, Home automation, Attacks, Watermarking**

I. Introduction

Authentication of light-weight IoT devices is considered the fundamental security feature regardless of the application. In home automation systems (HAS) an authentication mechanism is often embedded in the communication protocol, which defines a secure data exchange between the devices and the central gateway/hub. In a common case, a secret key is exchanged as a device joins the network for the first time. This key is later used to authenticate the device and provide confidentiality via data encryption [1]. While such a standard IoT security feature is an effective way to increase the user's privacy and thwart various network attacks, this basic authentication method fails to verify the authenticity of other interfaces on a HAS device such as the (often analog) sensor inputs. This fact exposes new attack vectors in HAS networks, which involve direct manipulations of a sensor device on the physical level. This way, an attacker is no longer required to gain access to the gateway or backend server of HAS to manipulate, e.g., a heating or air condition system. Instead, sensor manipulations as simple as decreasing resistance or output voltage of a sensor might have catastrophic consequences as the control loops in HAS cannot distinguish a malicious modification from a normal sensor operation due to a lack of low level integrity checks.

In state-of-the-art studies regarding sensor attacks various ways of sensor manipulations have been presented. In [2] a sensor saturation attack was demonstrated, where a medical infusion pump control was influenced via an external infrared light source. As the infrared sensor is driven to its non-linear region, the legitimate signaling gets suppressed, which leads to malformed sensor readings. Thus, the sensor saturation is one kind of a Denial-of-Service attack in a cyber physical system. A related but somewhat more sophisticated sensor attack vector was presented in [3]. Here the authors were able to manipulate MEMS gyroscope data by acoustic signal injections. As the micro mechanical structures of a MEMS gyroscope react also to sound pressure changes, an attacker is able to craft audio signals which closely mimic the legitimate gyroscope sensor signals. In a similar vein the authors of [4] injected signals to a MEMS microphone with an amplitude modulated laser light source. In comparison to the audio based signal injection which needs to be mounted locally, the laser-based manipulation attack can be launched from a distance of approx. 100 m. As a MEMS microphone is an essential part of many voice assistants, such long range manipulation attacks are critical in HAS, where, e.g., heating can be directly controlled via voice commands.

II. Related Work

With the rising interest on sensor attack vectors, work has also been devoted to novel protection techniques in various fronts. Firstly, in [5] the authors propose a physical challenge-response mechanism to detect spoofing attacks. In their system, a special kind of duty cycling of the active sensor actuator (unknown to the attacker) is the key to reveal the attack attempts. This idea was enriched in [6] to protect passive sensors from spoofing attacks. Further detection of out-of-band sensor attacks involve sensor fusion techniques [7], filtering before analog-to-digital conversion [8] and better shielding [9]. In automatic control systems literature, sensor watermarking techniques have been proposed to detect manipulation attempts in networked control systems [10]–[12]. For instance, the authors of [12] utilized additive sensor watermarking to authenticate sensors in a linear control system.

This work is co-financed by the Gesellschaft für Forschungsförderung Niederösterreich within the FTI Call Digitalization 2018. The authors are responsible for the content of this publication.

By superimposing a watermark signal on top of an input signal, the authors could reliably detect replay attacks. This idea was extended to detect physical rerouting attacks in [10] with a multiplicative SISO filter based approach. Further, in [11] a statistical watermarking method was designed for a parallel detection of network and sensor level attacks. While these works deliver comprehensive theoretical and numerical analyses, they are optimized for control systems and often lack real-world experimental validation.

Despite the advances in the defense mechanisms, there is still plenty of room for further hardening steps against more powerful attacker models. For instance, in [5] only a non-invasive attacker model was assumed, which does not take into account sensor modification attacks. Additionally, as remarked in [13] the experimental demonstrations of the protection methods are often limited in scope.

III. Contributions

With the above mentioned issues in sensor-level security in mind, we propose a novel watermarking concept in this paper to detect invasive sensor attacks in HAS. The main contribution of the paper is a supply voltage watermarking scheme for low complexity sensor devices. Albeit the least-significant-bit (LSB) watermarking method itself is a popular way to protect, e.g., digital time series data, to our best knowledge it has not been widely applied to protect HAS sensor data in analog domain. In the digital domain watermarks have been applied to non-media data [14] either in-band, e.g., LSB-watermarking [15], or out of band, e.g., encoded in the inter packet interval [16]. We also provide a measurement based validation. We demonstrate the applicability of our scheme with a real-world implementation based on a ultra-low power microcontroller equipped with a temperature sensor. With several experiments in different environmental settings (indoors/outdoors), we reveal the parameter settings, which deliver the highest performance and reliability in terms of sensor attack detection.

The rest of the paper is organized as follows: Section IV presents the system model including state-of-the-art sensor attacks. The implementation of the watermarking scheme and the evaluation results of our scheme are given in Sections V and VI, respectively. Further in Section VII we compare our method to the state-of-the-art and give a brief overview on the impact to HAS security. Finally, Section VIII concludes the paper with planned steps for future work.

IV. Background

A. HAS Network and Device Technology

A typical HAS network, as illustrated in Figure 1 connects several HAS devices via a central HAS gateway towards a backend server. For the device to gateway communication wireless protocols such as Zigbee, WLAN or Bluetooth are often deployed. Alternatively, wired solutions such as KNX or BACNet protocols are preferred especially when a high communication reliability is of importance. The HAS user has typically local access to the gateway to pair new devices

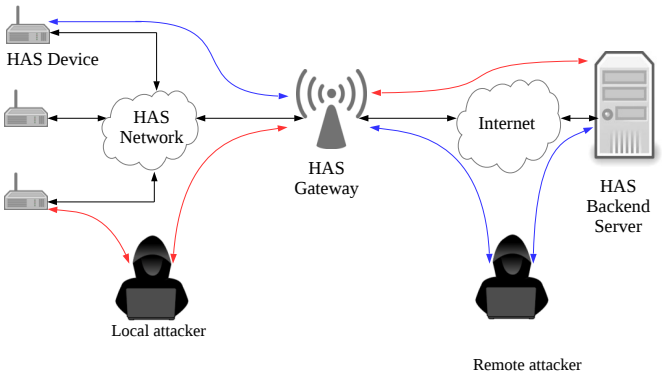


Fig. 1: HAS network with local and remote attackers depicted.

TABLE I: HAS Sensor Technology

Sensor Technology	Example Application
Infra-red Sensor	Motion Detector
Ionization	Smoke Detector
Thermistor	Temperature Sensor
Capacitive	Moisture Detector
Photodiode	Brightness Sensing

with the network and to monitor their current states. The gateway also takes care of the control actions of the HAS actuator devices, like air conditioning or heating, based on the collected sensor data. Finally, the sensor readings and the device configurations can be visualized on the back-end server.

The orchestration of the automation control necessitates versatile sensing of many kinds of environmental parameters, e.g., the moisture levels inside the walls or the temperature/brightness/humidity of the rooms. Hence, several different kinds of sensors, with examples listed in Table I, need to be deployed to achieve an accurate sensing. Common to all of these sensors is that they either change their resistance, which is then converted to a voltage using a bridge circuit or voltage divider, or produce an analog output voltage. The supply voltage influences the sensor output in most cases, too, and will be used for applying the watermark directly to the sensor's analog output.

B. Sensor Attacks

The attacks against sensor devices can be divided into two categories: invasive and non-invasive attacks (figure 2). The former category contains attacks which require a physical modification to the sensor, whereas attacks in the latter category can be executed without physical tampering of the device.

Invasive sensor attacks involve on the one hand voltage saturation attacks, where the attacker drives the ADC input voltage to a level beyond the normal operating conditions of the sensor. The sensor is pinned in saturation. With such modification the sensor will no longer respond to the changes of the measured quantity. The second type of invasive attacks aims to modify the sensor values by a voltage injection, e.g., via an external DAC to the sensor ADC input. Such setting

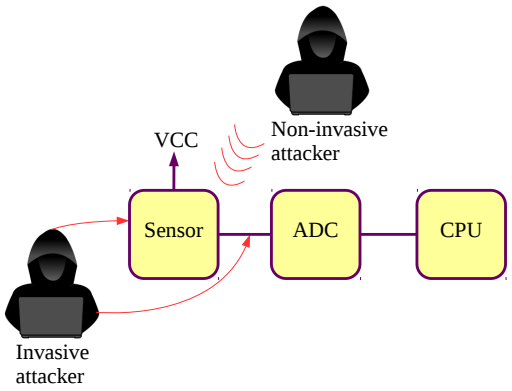


Fig. 2: A typical sensor device with locations of the invasive and the non-invasive attacker.

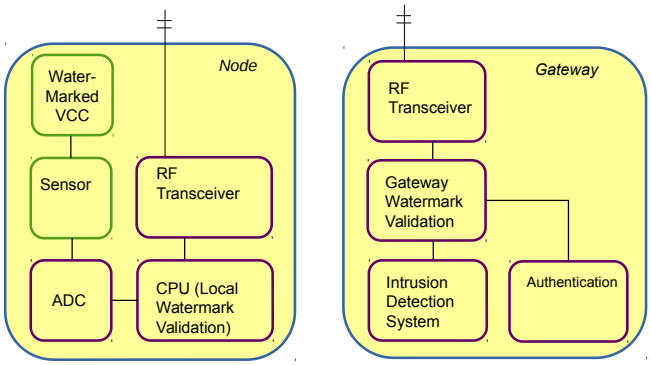


Fig. 3: Intrusion detection and authentication in HAS Network based on sensor watermarking

allows for replay attacks of the original sensor values or spoofing attacks with crafted sensor signals.

Non-invasive attacks cover on the one hand remote sensor manipulation attacks. For example, as demonstrated in [4], the properties of a MEMS microphone can be manipulated by a laser light source, which leads to vulnerabilities, e.g., in voice assistants. On the other hand, malicious signals can also be infiltrated non-invasively. The so-called out-of-band signal injection attacks are based on the fact that the sensors react to signals which are outside the application frequency range, as well as to fault injection via side-channels (e.g., via induction). Some examples of such attacks include injection of acceleration signals in MEMS sensors [3], ultra-sound command injections in voice assistants [17] and injection of AC signals via induction [18].

C. System Model

Figure 3 illustrates how our proposed intrusion detection system based on sensor watermarking is embedded in a HAS. The fundamental component in such a system is a sensor supply voltage source which applies a watermark signal (AC

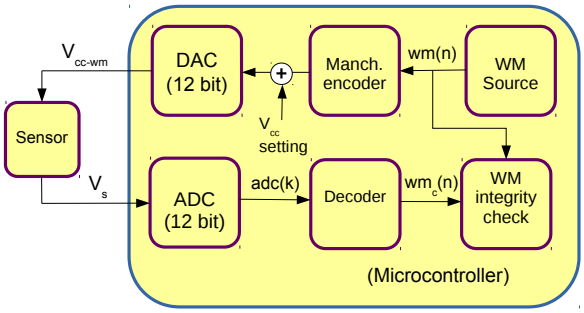


Fig. 4: Sensor watermarking implementation on a micro-controller including watermark (wm) generation, Manchester en/decoding and integrity check

component) superimposed on the sensor's supply voltage. Consequently the ADC input voltage is also perturbed by the watermark signal and hence, the embedded watermark can be firstly extracted out of the ADC readings and secondly validated against the original watermark. The watermark validation can be located on the HAS node, the gateway, or both. Our scheme supports the following security services:

1) Local integrity validation: The watermark is only recovered and validated on the sensor node, indicating the manipulation of the sensor value. As the watermark validation takes place on the HAS node, the communication overhead between the node and the gateway can be kept minimal. In case the watermark is invalid due to, e.g., a tampering attempt, a payload is sent towards the HAS gateway via a secured communication link, which ultimately triggers action to inform the user on the intrusion.
2) Gateway-based integrity validation and authentication: Since the watermark is embedded in the measurement data, it is transmitted without overhead to the gateway and can be used to authenticate the sending node as well as the integrity of the measurement. The watermark protects both manipulation of the analog sensor value as well as the transmission over the network. The latter enables retrofitting of systems originally providing no security. Thus, identification of rogue devices becomes possible.

In this paper we want to focus on the first application, although most findings also apply to the second use-case.

V. Watermarking Scheme

A. Sensor Watermarking Implementation

A watermarking implementation on an embedded sensor node able to detect various sensor signal manipulation attacks is illustrated in Figure 4. The underlying watermarking technique is a variant of the so called least-significant-bit (LSB) method, where the watermark itself is embedded by modulating the LSB bit of a binary word [19]. Since LSB

watermarking can be implemented with low amount of computational power, the method is well suited for HAS sensors.

In order to apply such a watermark to a sensor signal, the following steps are necessary:

1) A fixed length binary stream wm is generated on a device (here represented by a microcontroller or alternatively a sealed watermark generator). Afterwards, the binary stream is encoded with Manchester II code and added on top of a digital value, which controls the operating voltage of the sensor.
2) A digital-to-analog converter (DAC) generates the watermarked operating voltage for the sensor.
3) The watermarked sensor voltage signal ($\mathrm{V_s}$) is captured by an analog-to-digital converter (ADC). The steps 2. and 3. are repeated unit all the watermark bits contained in the ADC readings $\mathrm{adc}(k)$ are captured.
4) The captured binary stream $\mathrm{wm_c}$ is extracted from the digitized watermarked signal by

$$\mathrm{wm}_c(n) = \begin{cases} 0 & \mathrm{adc}(2n+1) - \mathrm{adc}(2n) \leq 0 \\ 1 & \mathrm{adc}(2n+1) - \mathrm{adc}(2n) > 0 \end{cases}. \quad (1)$$

5) A sum of bit errors β is calculated between wm and wm_c as

$$\beta = \sum_{n=0}^{\mathrm{N}-1} |\mathrm{wm}(n) - \mathrm{wm}_c(n)|, \quad (2)$$

where N denotes the length of the watermark wm. Subsequently, the bit error rate (BER) can be easily expressed as $\mathrm{BER} = \beta/N$.

Potential tampering efforts are detected by comparing the calculated bit errors with a predetermined bit error tolerance bound τ_{wm}. Thus, β exceeding a given limit indicates a sensor manipulation attack, which either can result in discarding the sensor value or is communicated, e.g., via the wireless link towards the home automation gateway/cloud server for intrusion detection. It shall be noted, that in the above model noise (thermal and/or switching based) and distortion (DAC and ADC) have an influence on τ_{wm}, since they determine the bit errors during the normal operation mode. To avoid an excessive amount of false positives, we demonstrate a systematic determination of τ_{wm} for a real device in the Section VI.

B. Sensor Reading Quality and Bias-Free Sensing

Since the sensor voltage source is modulated with a noise-like watermark signal, sensor output voltage and consequently sensor readings will inherently be affected by the watermarking.

In general we assume that the sensor signal $\mathrm{V_s}$ contains noise to be able to apply a LSB method. The LSB(s) containing the noise is replaced by the watermark. This allows the undisturbed reading of individual sensor values without a deterioration of the measured signal, as long as $\mathrm{V_{wm}}$ is not exceeding the noise floor of the original sensor signal.

Nevertheless, the different characteristic of noise and the watermark signal has to be considered if some post processing is applied, e.g., when a final sensor value is calculated from the mean of multiple readings of $\mathrm{adc(n)}$, which is the common way to suppress, e.g., ADC noise. In this case, the deviation from the true sensor value becomes dependent on the mean of the generated binary wm bit stream.

To avoid such a bias the properties of the watermark signal can be chosen to match the natural noise, but can result in a strong reduction of security. We propose an alternative way to average the ADC readings. Instead of summing up every $\mathrm{adc(k)}$ value, we select only the $\mathrm{adc(k)}$ readings during which the condition $\mathrm{adc(k)} - \mathrm{adc(k-1)} > 0$ holds. In other words, the sensor value average is built with ADC readings for which the watermark signal voltage is zero, and thus the voltage at the ADC stems only from the sensor itself.

C. Attacker Model and Security Boundaries

Among the large spectrum of sensor attacks, our watermarking scheme is designed to successfully detect the attacks involving direct voltage manipulation of the sensor ADC input voltage. In the following we define the attacker capabilities, the utilized attack methods and the attacker's level of access to a sensor device relevant for our evaluation.

Firstly, we define two attack scenarios as follows:

- Grounding attack: The attacker physically modifies, e.g., the printed circuit board of the sensor so that the sensor output voltage measured by the ADC voltage input is constantly tied to the zero potential. As such the sensor reports extremely low values towards the HAS gateway. In the same way the attacker can pin the sensor output voltage to the supply voltage.
- Voltage injection: The attacker substitutes the connection between the physical sensor and the ADC by, e.g., an external voltage source. By this way, the attacker gains full control over the physical sensor, which allows for sensor spoofing attacks. In this paper, the sensor spoofing attacks are restricted to considering only DC voltage injection attacks.

We further assume that the attacker cannot access the CPU or the memory in which the secrets (binary watermark data) are stored/processed. In addition, the above sensor attacks are performed non-synchronously to the sensor events without any side information on the watermarking itself. This security boundary applies to the local watermark validation investigated here, where the credentials are stored locally.

In principle, we can also relax these assumption when the watermark generation is provided by a dedicated circuit as depicted in Figure 3 and validated on the gateway. In this case we only need to assume that the attacker is not able to tamper the source of the watermarked sensor supply voltage.

VI. Experimental Results

A. Experimental Setup

The proposed sensor watermarking concept was realized on the ST Microsystems development board B-L072Z-LRWAN1 [20] which features a Cortex M0+ microcontroller equipped with a 12-bit DAC and a 12-bit ADC. The internal reference

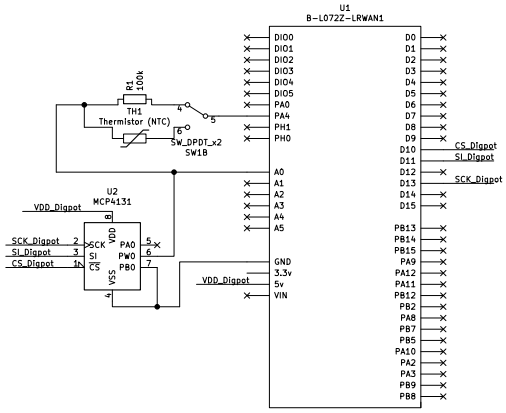


Fig. 5: Schematic of the test-bed utilized for sensor attack emulation and testing of watermark based intrusion detection.

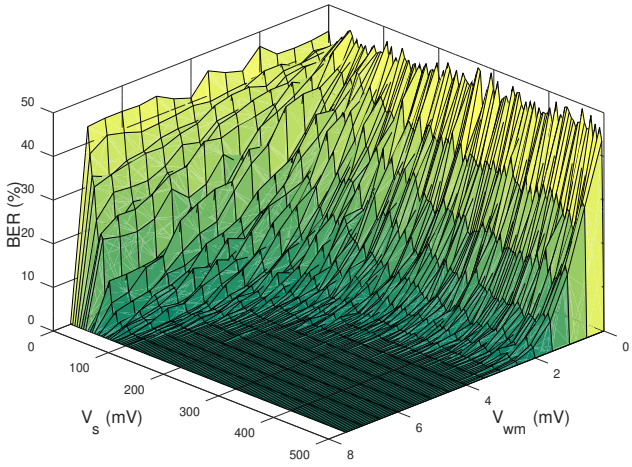


Fig. 6: BER behaviour of the measurement setup for various V_s and V_{vm} settings.

voltage for the two converters is equal to $1.224\mathrm{V}$. In addition, a $100\mathrm{k}\Omega$ digital potentiometer MCP4131 and a $10\mathrm{k}\Omega$ NTC thermistor were connected to the development board to allow for a flexible characterization of the test-bed and also to enable an automated execution of the sensor attacks. Figure 5 illustrates the interconnections between the components. The watermarked supply voltage is generated by the on-board DAC (pin PA4), which is connected to the digital potentiometer via the resistor (evaluation mode for our parameter studies) or thermistor (measurement mode providing real measurements). Finally, the digital potentiometer is on the one hand used to emulate a resistive sensor for test-bed characterization purposes in the evaluation mode. On the other hand, a setting for 0Ω connects the ADC input (pin A0) to ground, which represents the grounding attack in our setup to be used in measurement mode.

The test-bed allows for the sweeping of the sensor output voltage $\mathrm{V_s}$ by configuring the digital potentiometer. The sensor supply voltage ($\mathrm{V_{cc-wm}}$) as well as the watermark signal amplitude ($\mathrm{V_{wm}}$) are controlled by the DAC. In our experiments $\mathrm{V_{cc-wm}} = 1\mathrm{V}$ and the watermark signal is generated by switching the DAC output between $\mathrm{V_{cc-wm}}$ and $\mathrm{V_{cc-wm}} + \mathrm{V_{wm}}$.

B. Characterization of Test-Bed

In order to reliably classify the two events, namely the normal operation and the operation under an attack, a comprehensive characterization of the test-bed was conducted. Identifying the combination of parameters which lead to a high BER allows us firstly to determine the sensor signal ($\mathrm{V_s}$) range for which our scheme is effective. Secondly, the characterization reveals acceptable settings for the amplitude of the generated watermark voltage ($\mathrm{V_{wm}}$), which introduces noise on the actual sensor signal.

In the best case, the amplitude of $\mathrm{V_{wm}}$ shall on the one hand be large enough to deliver significantly different β values which allow for a successful classification, and on the other hand be low enough to keep the noise component within acceptable bounds for the measured sensor value. With the two adjustable parameters $\mathrm{V_{wm}}$ and $\mathrm{V_s}$, the BER values were evaluated for $0 \leq \mathrm{V_{wm}} \leq 7.5\mathrm{mV}$ and $0 \leq \mathrm{V_s} \leq 500\mathrm{mV}$. The BER for each parameter setting was calculated out of a watermark wm consisting of 512 randomly distributed bits.

The results, as illustrated in Figure 6, depict the BER behaviour in normal operation. Under the two extreme conditions, i.e., zero resistance equal to fixing $\mathrm{V_s} = 0$ and/or $\mathrm{V_{wm}} = 0$, the watermark cannot be encoded in the signal $\mathrm{V_s}$. Consequently the BER becomes close to 50%, since the watermark bits are evenly distributed in the watermark but only one $\mathrm{wm_c}(n)$ value can be retrieved. By increasing $\mathrm{V_{wm}}$ in steps of $1/2048V$ (minimum step size of the 12 bit DAC), the BER decreases steadily until a plateau of about 0% is reached. Between $0 \leq \mathrm{V_{wm}} \leq 3\mathrm{mV}$ the increased BER can be explained by the noise in the ADC, which results in erroneous $\mathrm{wm_c}$ bits. On the other hand, in the region where $0 \leq \mathrm{V_s} \leq 50\mathrm{mV}$, the dominating sources for bit errors are the non-linearity (and noise) connected to both DAC and ADC.

The effect of the modulated sensor supply voltage to the sensor signal quality was investigated with an NTC thermistor. As mentioned in Section V, the implemented watermarking involves amplitude modulation, which causes a bias to the sensed temperature values. This effect was validated by comparing measured temperature values calculated by averaging ADC readings without and with watermarking. The results given in Figure 7 illustrate the bias for various $\mathrm{V_{wm}}$ values. Since the mean of the watermark signal is non-zero, the signal difference in an extreme of $\mathrm{V_{wm}} = 100\mathrm{mV}$ can be up to $25\mathrm{mV}$ in $\mathrm{V_s}$ and equals a temperature difference of up to $2°\mathrm{C}$. Therefore, for the smaller $\mathrm{V_{wm}}$ the bias of ca.

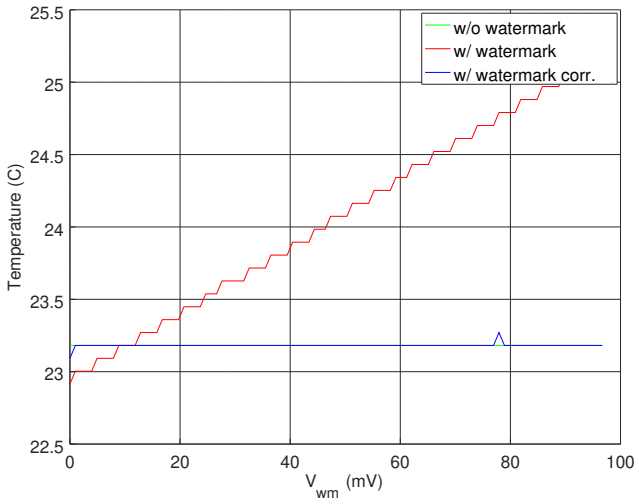


Fig. 7: Measured temperature without watermark and ordinary averaging (green), with watermark and ordinary averaging (red) and with watermark and selective averaging (blue) for various $\mathrm{V_{wm}}$ settings.

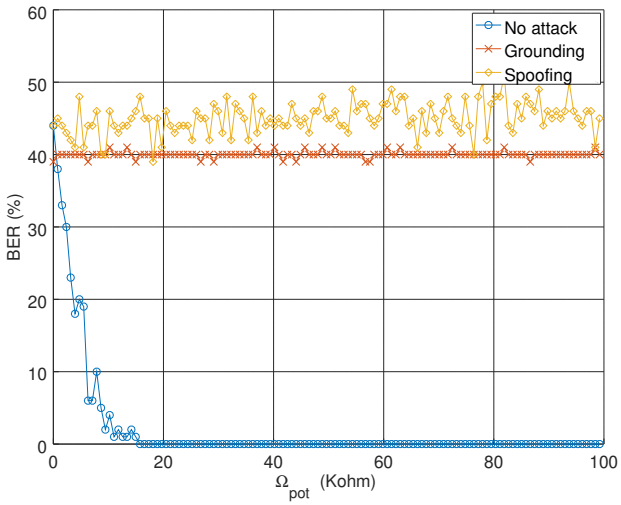


Fig. 8: BER behavior of the measurement setup during normal operation, grounding attack, and voltage injection attack conditions.

0.5°C might already be tolerable for temperature sensing in HAS. Nevertheless, by applying the correction by the selective averaging, the measured temperature values nearly overlap with the original measurements without watermarking. Similar effects of watermark induced shifts have been observed in [16].

C. Sensor Attack Detection

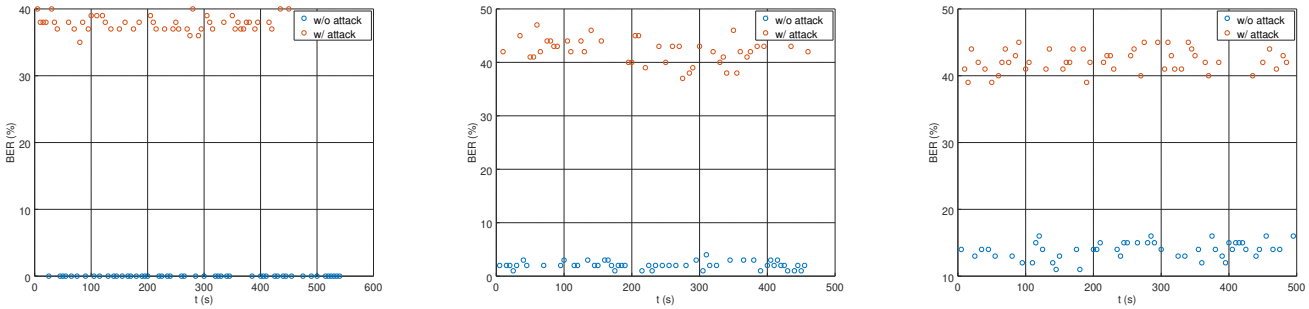
The main goal of our watermarking scheme is to detect ADC voltage manipulation attempts. Thus, a set of measurements without attack, with the grounding attack and with the voltage injection attack were performed. For each scenario, data points were collected for various resistance values of the digital potentiometer, representing the physical sensor. The BER values illustrated in Figure 8 indicate a large difference in attained BER between the normal behaviour and the behaviour under an attack. Only for small resistance values the gap between the states is smaller (see section VI-B and Figure 6 on small values of $\mathrm{V_s}$). Since both attacks erase the embedded watermark signal from the ADC input completely, the BER increases to nearly 50% for each resistance setting. Thus, fixing the tolerance bound heuristically to $\tau_{\mathrm{wm}} = 167$ (at watermark length $N = 512$) allows us to correctly detect sensor attacks in 126 resistance parameter settings out of 128 in total, which yields a detection rate of 98.4%.

Next, the efficiency of the watermark scheme was evaluated with the NTC thermistor in different real-world operating conditions. For the evaluation we considered three scenarios: 1) indoors, 24 °C 2) outdoors, 7 °C and 3) freezer, -15 °C. In each case, 200 measurements were made by randomly applying the grounding attack to the sensor with a time interval of $5\mathrm{s}$ and $\mathrm{V_{wm}} = 7.3\mathrm{mV}$. The BER results are given in Figure 9 showing noise and voltage range effects on the attack detection. It can be seen that the noise for each state increases but also the gap between the states decreases with decreasing temperature. While the sensor attacks here can still be fully distinguished, the lower temperature values cause an increase in NTC resistance, which in turn steers the ADC input voltage in our setup towards $0\mathrm{V}$. In this voltage regime, noise and nonlinear effects become dominant, which decreases the quality of the captured $\mathrm{wm_c}$. Thus, at the boundaries of the NTC temperature range, the detection accuracy decreases due to false positives.

VII. Discussion

The state of the art literature mentions several ways to detect sensor attacks and therefore, in the following, we will briefly discuss the differences between the previous work and our proposed sensor watermarking scheme. In [21] a challenge-response mechanism based on a physical unclonable function (PUF) is built using MEMS relays (SensorPUF). In such a system, tampering attempts to the physical quantity, to the challenge or to the response will invalidate the authentication and consequently also the sensor reading. The main difference between the SensorPUF and our scheme is the applicability to different sensor technologies. Whereas the PUF relies on supply voltage variations in specific sensor/MEMS structures, our scheme can be implemented more flexibly on different passive and active sensors not providing these characteristics.

A related concept, which is somewhat closer to the proposed method is presented in [5]. In this work special signaling strategies were constructed for sensor-actuator systems which are able to expose eavesdropping and spoofing attacks. Although the authors of this paper were able to present many



(a) (b) (c)

Fig. 9: BER behaviour of a NTC temperature sensor with and without the grounding attack in (a) 24 °C (indoors) (b) 7 °C (outdoors) and (c) -15 °C (freezer)

use-cases with a high detection accuracy, the work failed to consider the invasive attacker model. The presented solution can counterfeit against invasive attacks.

A further sensor authentication technique is given in [22], where the unique noise characteristics of a sensor are used to construct fingerprints. In comparison to our method, the fingerprinting necessitates additional complexity to extract and to remotely store the fingerprints. In the local watermark based detection such efforts are not strictly necessary and thus our method is better suited for large-scale deployments of resource limited devices.

Another design which is conceptually related to ours is the so called physical watermarking scheme implemented in automatic control systems [12]. There the general principle states that an additional watermark signal laid on top of a control input can reveal network and/or sensor attacks. Various watermarking schemes to control systems have been given, which are intended to detect e.g. replay [12], network [10] and sensor [11] attacks. A common factor to those techniques is that they apply mostly to discrete linear time-invariant state-space models and thus require the entire control loop for their operation. In comparison to our method, which is able to detect an intrusion with only a few samples, the state-space models require much larger sample sets to converge to a correct detection decision. Moreover, detectors in physical watermarking schemes require solving of moderately complex optimization programs which render them computationally intensive.

Next we discuss the impact of the presented sensor attack detection to HAS security. To our best knowledge, most works on authentication in HAS are concentrated on the cryptographic methods, aiming to verify authenticity of the device communication [1]. In addition, the effects of the sensor attacks in HAS environments have not been considered in detail. With our scheme, the trust in the sensor level can be improved as potentially every sensor reading can be authenticated. This is particularly important in HAS, where the sensors play a pivotal role as a part of the system, e.g., of an air-conditioning or heating systems. As the designed watermarking scheme can be implemented on ultra low power microcontrollers with almost no additional components, the deployment to most HAS end devices shall be feasible. One of the potential security application of our watermarking scheme is an early detection of physical tampering attempts to prevent catastrophic events, e.g., damaged equipment. Such protection mechanism shall be achievable by coupling the sensor attack detection to the central HAS gateway, which notifies the user and triggers suitable control actions to protect the critical HAS assets.

Concerning the sensor data quality degradation due to supply voltage watermarking, the bias free sensing as given in Section V B shall apply regardless of the sensor type. As long as the sensor voltage sampling is performed for the unmodulated parts of the supply voltage settings, i.e. during binary watermark value zero, the captured value is unbiased. In case such selective sampling scheme cannot be implemented, e.g. due to more complex way of watermark embedding, additional error correction steps might be required. Nevertheless, as our proposed scheme is of a closed-loop type, characterization for detrimental effects is reasonable as the excitation (watermarked supply voltage signal) and the response (watermarked sensor voltage) are both available. Thus, a linearization step which maps a watermarked sensor value to a non watermarked one might enable a way to mitigate for offsets in sensor data caused by the watermarking itself.

VIII. Conclusions

We presented a novel sensor watermarking scheme which is able to reliably detect invasive sensors attacks. The method was implemented on a microcontroller platform which demonstrates its applicability in ultra low power HAS devices. With a measurement based evaluation, we were able to determine the attack detection performance. According to our evaluation results, a normal operation and an attack condition can be distinguished with approx. 98% accuracy. Furthermore, we identified a way to perform bias free sensor measurements while utilizing the watermarking in analog domain. In follow-up works, we will focus on extending the concept for

further types of sensors (e.g., capacitive sensors) and more advanced attacker models. On the security side, we plan to find more general rules for the trade-off between watermark amplitude V_{wm} and amplitude of the sensor signal V_s as well as the watermark length responsible for the strength of the protection, when an attacker actively tries to insert a manipulated watermark signal.

REFERENCES

[1] P. Kumar, A. Gurtov, J. Iinatti, M. Ylianttila, and M. Sain, "Lightweight and secure session-key establishment scheme in smart home environments," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 254–264, 2016.

[2] Y. Park, Y. Son, H. Shin, D. Kim, and Y. Kim, "This ain't your dose: Sensor spoofing attack on medical infusion pump," in *10th USENIX Workshop on Offensive Technologies (WOOT 16)*. Austin, TX: USENIX Association, Aug. 2016. [Online]. Available: https://www.usenix.org/conference/woot16/workshop-program/presentation/park

[3] T. Trippel, O. Weisse, W. Xu, P. Honeyman, and K. Fu, "Walnut: Waging doubt on the integrity of mems accelerometers with acoustic injection attacks," in *2017 IEEE European Symposium on Security and Privacy (EuroS P)*, 2017, pp. 3–18.

[4] T. Sugawara, B. Cyr, S. Rampazzi, D. Genkin, and K. Fu, "Light commands: Laser-based audio injection attacks on voice-controllable systems," in *29th USENIX Security Symposium (USENIX Security 20)*, 2020.

[5] Y. Shoukry, P. Martin, Y. Yona, S. Diggavi, and M. Srivastava, "Pycra: Physical challenge-response authentication for active sensors under spoofing attacks," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1004–1015. [Online]. Available: https://doi.org/10.1145/2810103.2813679

[6] Y. Zhang and K. Rasmussen, "Detection of electromagnetic interference attacks on sensor systems," in *2020 IEEE Symposium on Security and Privacy (SP)*, 2020, pp. 203–216.

[7] K. S. Tharayil, B. Farshteindiker, S. Eyal, N. Hasidim, R. Hershkovitz, S. Houri, I. Yoffe, M. Oren, and Y. Oren, "Sensor defense in-software (sdi): Practical software based detection of spoofing attacks on position sensors," *Engineering Applications of Artificial Intelligence*, vol. 95, p. 103904, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197620302402

[8] C. Kasmi and J. Lopes Esteves, "Iemi threats for information security: Remote command injection on modern smartphones," *IEEE Transactions on Electromagnetic Compatibility*, vol. 57, no. 6, pp. 1752–1755, 2015.

[9] D. F. Kune, J. Backes, S. S. Clark, D. Kramer, M. Reynolds, K. Fu, Y. Kim, and W. Xu, "Ghost talk: Mitigating emi signal injection attacks against analog sensors," in *2013 IEEE Symposium on Security and Privacy*, 2013, pp. 145–159.

[10] R. M. Ferrari and A. M. Teixeira, "Detection and isolation of routing attacks through sensor watermarking," in *2017 American Control Conference (ACC)*, 2017, pp. 5436–5442.

[11] P. Hespanhol, M. Porter, R. Vasudevan, and A. Aswani, "Statistical watermarking for networked control systems," in *2018 Annual American Control Conference (ACC)*, 2018, pp. 5467–5472.

[12] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 93–109, 2015.

[13] I. Giechaskiel and K. Rasmussen, "Taxonomy and challenges of out-of-band signal injection attacks and defenses," *IEEE Communications Surveys Tutorials*, vol. 22, no. 1, pp. 645–670, 2020.

[14] A. Soltani Panah, R. Van Schyndel, T. Sellis, and E. Bertino, "On the properties of non-media digital watermarking: A review of state of the art techniques," *IEEE Access*, vol. 4, pp. 2670–2704, 2016.

[15] I. Kamel and H. Juma, "A lightweight data integrity scheme for sensor networks," *Sensors (Basel)*, vol. 11, no. 4, pp. 4118–4136, 2011.

[16] T. Bigler, A. Treytl, and T. Sauter, "Side-channel watermarking for lorawan using robust inter-packet timing," in *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, vol. 1, 2020, pp. 1367–1370.

[17] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu, "Dolphinattack: Inaudible voice commands," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 103–117. [Online]. Available: https://doi.org/10.1145/3133956.3134052

[18] J. Selvaraj, G. Y. Dayanıklı, N. P. Gaunkar, D. Ware, R. M. Gerdes, and M. Mina, "Electromagnetic induction attacks against embedded systems," in *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, ser. ASIACCS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 499–510. [Online]. Available: https://doi.org/10.1145/3196494.3196556

[19] A. Bamatraf, R. Ibrahim, and M. N. B. M. Salleh, "Digital watermarking algorithm using lsb," in *2010 International Conference on Computer Applications and Industrial Electronics*, 2010, pp. 155–159.

[20] *STM32L0 Discovery kit LoRa, Sigfox, low-power wireless*, ST Microsystems, 3 2021.

[21] J. Tang, R. Karri, and J. Rajendran, "Securing pressure measurements using sensorpufs," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2016, pp. 1330–1333.

[22] C. M. Ahmed, J. Zhou, and A. P. Mathur, "Noise matters: Using sensor and process noise fingerprint to detect stealthy cyber attacks and authenticate sensors in cps," in *Proceedings of the 34th Annual Computer Security Applications Conference*, ser. ACSAC '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 566–581. [Online]. Available: https://doi.org/10.1145/3274694.3274748