PURE: Using Verified Remote Attestation to Obtain Proofs of Update, Reset and Erasure in Low-End Embedded Systems

Ivan De Oliveira Nunes
University of California, Irvine
ivanoliv@uci.edu

Norrathep Rattanavipanon
University of California, Irvine
nrattana@uci.edu

Karim Eldefrawy
SRI International
karim.eldefrawy@sri.com

Gene Tsudik
University of California, Irvine
gene.tsudik@uci.edu

ABSTRACT
Remote Attestation (RA) is a security service that enables a trusted verifier (Vrf) to measure current memory state of an untrusted remote prover (Prv). If correctly implemented, RA allows Vrf to remotely detect if Prv’s memory reflects a compromised state. However, RA by itself offers no means of remediating the situation once Prv is determined to be compromised. In this work we show how a secure RA architecture can be extended to enable important and useful security services for low-end embedded devices. In particular, we extend the formally verified RA architecture, VRASED, to implement provably secure software update, erasure, and system-wide resets. When (serially) composed, these features guarantee that a remote Prv has been updated to a functional and malware-free state, and was properly initialized after such process. These services are provably secure against an adversary (represented by malware) that compromises Prv and exerts full control of its software state. Our results demonstrate that such services incur minimal additional overhead (0.4% extra hardware footprint, and 100-s milliseconds to generate combined proofs of update, erasure, and reset), making them practical even for the lowest-end embedded devices, e.g., those based on MSP430 or AVR ATMega micro-controller units (MCUs). All changes introduced by our new services to VRASED trusted components are also formally verified.

1 INTRODUCTION
In recent years, the number and variety of special-purpose computing devices increased dramatically. This includes all kinds of embedded devices, cyber-physical systems (CPS), and Internet-of-Things (IoT) gadgets, that are increasingly utilized in various “smart” settings, such as homes, offices, factories, automotive systems, and public venues. Despite their many benefits, such devices unfortunately also represent natural and attractive targets for attacks, especially, remote exploits and malware infestations. As society becomes accustomed to being surrounded by, and dependent on, such devices, their security becomes extremely important.

At the low-end of the spectrum, embedded devices are designed with strict constraints on cost, physical size, and energy consumption, i.e., ultra-low-power MCUs, such as TI MSP430 [1]. It is thus unrealistic to expect such devices to have sophisticated means (akin to those available for laptops or smartphones) to prevent malware injection. Remote Attestation (RA) has recently emerged as an alternative and inexpensive mean to detect malware presence on remote low-end devices. RA allows a trusted and more powerful verifier (Vrf) to remotely measure the software state of an untrusted remote prover (Prv). As shown in Figure 1, RA is typically realized as a simple challenge-response protocol:

1. Vrf sends an attestation request with a (random) challenge (Chal) to Prv. This request might also contain a token derived from a secret that allows Prv to authenticate Vrf.
2. Prv receives the attestation request and computes an authenticated integrity check over its memory and Chal. The memory region to be attested might be either pre-defined, or explicitly specified in the request. In the latter case, authentication of Vrf in step (1) is important for overall security/privacy of Prv, since the request can specify arbitrary memory regions.
3. Prv returns the result to Vrf.
4. Vrf receives the result from Prv, and checks whether it corresponds to a valid memory state.

The authenticated integrity check can be implemented as a Message Authentication Code (MAC) over Prv’s entire memory, or a specific range thereof. However, computing a MAC requires Prv to have a unique secret key (denoted by $K$) shared with Vrf. $K$ must reside in secure storage, where it is inaccessible to any software running on Prv, except for privileged attestation code. Since the usual RA threat model assumes a fully compromised software state on Prv, secure storage implies some level of hardware support. Hybrid RA (based on hardware/software co-design) is a particularly promising approach for low-end embedded devices. It aims to provide the same security guarantees as (more expensive) hardware-based RA approaches (e.g., those based on TPMs [1], SGX [2] or

---

1http://www.ti.com/microcontrollers/msp430-ultra-low-power-mcus/applications.html

---
TrustZone [3]), while minimizing modifications to underlying hardware.

Despite major progress and many proposed RA architectures with different assumptions and guarantees [4–9], current RA approaches are limited in two major aspects: First, they assume that, if an invalid state is detected (either because of malware or errors), physical out-of-band measures must be taken to re-program and/or reset $P_{rv}$. This is a cumbersome requirement, and we show how it can be avoided by extending the RA architecture to provide secure software update, memory erasure, and proofs of reset. Second, they typically do not offer formal guarantees or any kind of provable security. We argue that, before such architectures can be adopted on a large scale, all their security services and properties should stand on solid ground. To do so, we build upon a recently proposed formally verified VRASED RA architecture. This allows us to prove security of every protocol and implementation proposed in this work. Whenever modifications to the Trusted Code Base (TCB) of the RA architecture are required, they are formally verified. This way, security is attained for both the protocol and its implementation.

Our goal is to develop an architecture that assures $\forall v f$ that a remote $P_{rv}$ has been brought to a functional malware-free state and has been properly initialized. Achieving this goal requires composing three services: (i) proof of software update to ensure $P_{rv}$’s program memory re-configuration; (ii) proof of erasure of data memory, to ensure that no data from prior executions remain, including malware that can hide itself; and (iii) proof of system reset to ensure that – after program memory update and data memory erasure – $P_{rv}$’s software state is reset and execution control flow is re-initialized. With this end-goal in mind, we propose a provably secure and formally verified architecture, called PURE: Proofs of Update, Reset, and Erasure. This work makes the following contributions:

- **New security services for low-end embedded systems**: we construct security services for proofs of remote system-wide reset, secure update and erasure by extending VRASED [9] – a formally verified hybrid RA architecture. The overhead of our design is minimal. To the best of our knowledge, PURE is the first architecture that provides proofs of system-wide reset. Secure update based on RA have been proposed for medium-end devices (capable of storing and running a secure micro-kernel) [10]. However, PURE is the first RA-based secure update service for low-end devices and also the first that is provably secure.

- **Formal verification and provable security**: We implement reset, update and erasure services by extending VRASED formally verified RA TCB. All changes to VRASED TCB are also formally verified, in order to yield end-to-end provable security. Resulting services are provably secure under the assumption of full compromise of $P_{rv}$’s software state.

- **Public-domain implementation and evaluation**: All proposed services are implemented on a real-world low-end MCU platform (TI MSP430) and deployed using FPGAs. The entire design, along with its verification, is publicly available and incorporated to VRASED open-source project [11]. Our evaluation demonstrates a low hardware footprint (0.4%), which we consider to be affordable even for low-end devices.

## 2 RELATED WORK

### Remote Attestation (RA)

Remote Attestation (RA): prior techniques fall into three categories: software-based, hardware-based, or hybrid. Security of software-based attestation [12–14] relies on strong assumptions about precise timing and/or constant communication delays, which are mostly unrealistic in the IoT ecosystem. Hardware-based methods [15–17] rely on security provided by dedicated hardware components, e.g., TPMs [1]. However, the cost of such hardware is prohibitive for low-end devices. Hybrid RA [6, 7, 9] aims to achieve security equivalent to hardware-based mechanisms, yet with lower hardware cost. It imposes minimal hardware requirements, while relying on software to minimize the complexity of the additional hardware.

### Formally Verified Security Services

In recent years, several efforts focused on formal verification of security services. In terms of cryptographic primitives, Hawblitzel et al. [18] verified implementations of SHA, HMAC, and RSA. Beringer et al. [19] verified the Open-SSL SHA-256 implementation. Bond et al. [20] verified an assembly implementation of SHA-256, Poly1305, AES and ECDSA. More recently, Zinzindohoué, et al. [21] presented HACL*, a verified cryptographic library containing the entire cryptographic API of NaCl [22]. Also, larger security-critical systems have been successfully verified, e.g., [23]. CompCert [24] is a C compiler formally verified to preserve C code semantics in generated assembly code. Klein et al. [25] designed and proved functional correctness of sel4 – the first verified general-purpose microkernel. Cabodi et al. [26, 27] took the first steps towards formalizing hybrid RA properties. Later, VRASED [9] realized a formally verified hybrid RA architecture.

## 3 BACKGROUND

We now overview relevant formal verification methods, and VRASED which serves as our main building block.

### 3.1 Formal Verification, Model Checking & LTL

Formal verification typically involves three basic steps. First, the system of interest (e.g., hardware, software, or communication protocol) is described using a formal model, e.g., a Finite State Machine (FSM). Second, properties that the model must satisfy are formally specified. Third, the system model is checked against formally specified properties to guarantee that the system satisfies them. This checking can be achieved via either Theorem Proving or Model Checking. We use the latter to verify the implementation of system modules, and the former to derive new properties from sub-properties that are proved for the modules’ implementation.

One approach to perform model checking is to specify properties as *formulæ* using Linear Temporal Logic (LTL) and system models as FSMs. Hence, a system is represented by a triple: $(S, S_0, T)$, where $S$ is a finite set of states, $S_0 \subseteq S$ is the set of possible initial states, and $T \subseteq S \times S$ is the transition relation set, which describes the set of states reachable in a single step from each state. The use of LTL to specify properties allows representation of expected system behavior over time.

We apply the model checker NuSMV [28], which can be used to verify generic HW or SW models. For digital hardware described at Register Transfer Level (RTL) – which is the case in this work –
conversion from Hardware Description Language (HDL) to NuSMV model specification is simple. Furthermore, it can be automated [29], since the standard RTL design already relies on describing hardware as an FSM.

In NuSMV, properties are specified in Linear Temporal Logic (LTL), which is particularly useful for verifying sequential systems, because it extends common logic statements with temporal clauses. In addition to propositional connectives, such as conjunction (\(\land\)), disjunction (\(\lor\)), negation (\(\neg\)), and implication (\(\rightarrow\)), LTL includes temporal connectives, thus enabling sequential reasoning. We are interested in the following temporal connectives:

- \(\Box \phi\) – \(\neg\text{next}\ \phi\): holds if \(\phi\) is true at the next system state.
- \(\Diamond \phi\) – Future \(\phi\): holds if there exists a future state where \(\phi\) is true.
- \(\mathcal{G}\phi\) – Globally \(\phi\): holds if for all future states \(\phi\) is true.
- \(\phi \lor \psi\) – Until \(\psi\): holds if there is a future state where \(\psi\) holds and \(\phi\) holds for all states prior to that.
- \(\phi \land \psi\) – Weak Until \(\psi\): holds if \(\psi\) holds for all states prior to the state when \(\psi\) holds. However, the state when \(\psi\) holds is not guaranteed to happen. In that case \(\phi\) will remain true forever.

This set of temporal connectives combined with propositional connectives (with their usual meanings) allows us to specify powerful rules. NuSMV works by checking LTL specifications against the system FSM for all reachable states.

### 3.2 VRASED Overview

VRASED [9] is a formally verified hardware/software co-design for RA targeting low-end devices. It is designed as a set of sub-modules, each guaranteeing a specific set of sub-properties. All sub-modules, both hardware and software, are individually verified. Finally, composition of all sub-modules is proved to achieve formal definitions of RA soundness and security. RA soundness guarantees that the content of AR can not be modified once RA computation starts, thus protecting against "hide-and-seek" attacks caused by roving malware [30]. RA security ensures that RA execution generates an unforgeable authenticated measurement and that the secret key \(K\) is not leaked before, during, or after attestation.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(f</td>
<td>T</td>
</tr>
<tr>
<td>TCB</td>
<td>ROM region storing SW-Att TCB = ({f</td>
</tr>
<tr>
<td>MR</td>
<td>(\text{MAC RAM}, \text{RAM region where SW-Att computation result is written, MR} = {f</td>
</tr>
<tr>
<td>AR</td>
<td>Attested Region, i.e., memory region to be attested. Can be fixed/predefined or specified in an authenticated request from ‘Yrf: AR = {f</td>
</tr>
</tbody>
</table>

Table 1: Notation for memory addresses

Figure 2 shows VRASED architecture and Table 1 summarizes the notation used throughout the rest of this paper. To achieve aforementioned goals, VRASED software (SW-Att in Figure 2) is stored in Read-Only Memory (ROM) and relies on a formally verified HMAC implementation from the HACL* cryptographic library [21]. A typical SW-Att execution is as follows:

1. Read a challenge \(\text{Chal}\) that is assumed to be stored in the memory region \(MR\).
2. Derive a one-time key from \(\text{Chal}\) and the attestation master key \(K\).
3. Generate an attestation token \(H\) by computing an HMAC over an attested memory region \(AR\) using the derived key. More formally, \(H = \text{HMAC}(K\text{DF}(K, \text{Chal}), \text{AR})\).
4. Write \(H\) into \(MR\) and return the execution to an unprivileged Software, i.e., normal applications.

VRASED hardware, as shown in Figure 2, monitors 7 MCU signals:

- \(\text{PC}\): Current Program Counter value;
- \(\text{R}_{\text{en}}\): Indicates if the MCU is reading from memory (1-bit);
- \(\text{W}_{\text{en}}\): Indicates if the MCU is writing to memory (1-bit);
- \(D_{\text{addr}}\): Address for an MCU memory access;
- \(\text{DMA}_{\text{en}}\): Indicates if Direct Memory Access (DMA) is currently enabled (1-bit);
- \(\text{DMA}_{\text{addr}}\): Memory address being accessed by DMA;
- \(\text{irq}\): Signal that indicates if an interrupt is happening (1-bit);

These signals are used to determine a one-bit reset signal output. When set to 1, it triggers an immediate system-wide MCU reset, prior to the execution of the next instruction. reset is triggered when VRASED hardware detects any violation of security properties. VRASED hardware is described in Register Transfer Level (RTL) using Finite State Machines (FSMs). Then, NuSMV Model Checker [28] is used to automatically prove that FSMs achieve claimed security sub-properties. Finally, the proof that conjunction of hardware and software sub-properties implies end-to-end soundness and security is performed using an LTL theorem prover. More formally, VRASED end-to-end security proof guarantees that no Probabilistic Polynomial-Time (PPT) adversary can win the security game in Definition 1 with more than negligible probability in the security parameter. In the rest of this paper, we show (via reductions) how PURE proofs of update and erasure can be constructed in a provably secure manner entirely based on VRASED end-to-end security guarantees; whereas we leverage specific LTL sub-properties guaranteed by VRASED hardware to design a minimal formally verified architecture for proofs of reset.
We consider an adversary, $\text{Adv}$, that controls entire software state of $P_{rv}$. $P_{rv}$ includes code and data. $\text{Adv}$ can thus modify any writable memory and read any memory (including secrets) that is not explicitly protected by VRASED access control rules. Also, $\text{Adv}$ has full access to all Direct Memory Access (DMA) controllers on $P_{rv}$.

Physical hardware attacks are out of scope of this paper: $\text{Adv}$ can not alter any hardware components to violate VRASED rules, e.g., by modifying code in ROM, inducing hardware faults, or retrieving $P_{rv}$ secrets via physical side-channels. Protection against physical attacks is orthogonal and attainable via standard tamper-resistance techniques [31]. We assume that the MCU architecture strictly adheres to, and correctly implements, all of its specifications. In particular, we assume that program counter (PC) always contains the address of the instruction being executed in a given cycle. Whenever a DMA controller attempts to access main system memory, a DMA-address signal ($\text{DMA}_{\text{addr}}$) reflects the address of the memory location being accessed and a DMA-enable bit ($\text{DMA}_{\text{en}}$) must be set. DMA can not access memory when $\text{DMA}_{\text{en}}$ is off (i.e., logical zero). The reset signal triggers an immediate (before executing the next instruction) system-wide reset. At the end of a successful reset, all registers (including PC) are set to zero before resuming normal software execution flow. Resets are handled by the MCU in hardware. Thus, reset handling routine can not be modified.

4 ADVERSARIAL MODEL & ASSUMPTIONS

We consider an adversary, $\text{Adv}$, that controls entire software state of $P_{rv}$, including code and data. $\text{Adv}$ can thus modify any writable memory and read any memory (including secrets) that is not explicitly protected by VRASED access control rules. Also, $\text{Adv}$ has full access to all Direct Memory Access (DMA) controllers on $P_{rv}$. Physical hardware attacks are out of scope of this paper: $\text{Adv}$ can not alter any hardware components to violate VRASED rules, e.g., by modifying code in ROM, inducing hardware faults, or retrieving $P_{rv}$ secrets via physical side-channels. Protection against physical attacks is orthogonal and attainable via standard tamper-resistance techniques [31]. We assume that the MCU architecture strictly adheres to, and correctly implements, all of its specifications. In particular, we assume that program counter (PC) always contains the address of the instruction being executed in a given cycle. Whenever a DMA controller attempts to access main system memory, a DMA-address signal ($\text{DMA}_{\text{addr}}$) reflects the address of the memory location being accessed and a DMA-enable bit ($\text{DMA}_{\text{en}}$) must be set. DMA can not access memory when $\text{DMA}_{\text{en}}$ is off (i.e., logical zero). The reset signal triggers an immediate (before executing the next instruction) system-wide reset. At the end of a successful reset, all registers (including PC) are set to zero before resuming normal software execution flow. Resets are handled by the MCU in hardware. Thus, reset handling routine can not be modified.

5 PROOFS OF REMOTE SYSTEM-WIDE RESET

(A POR)

A "Proof of (remote system-wide) Reset" (POR) is needed in several situations. First, POR can be used to restart software execution flow if $P_{rv}$ reaches an unexpected state. Second, if $P_{rv}$ has a secure boot module, resetting is a way to trigger secure boot code (e.g., to wipe-out privacy sensitive data memory after completion of a task). Third, after a software update, POR is needed to make sure that execution of newly installed software is initiated properly by restarting the MCU control flow execution. Without a reset, the MCU might start execution of the new code using state kept from software that was previously executing. Execution of new code might thus start at an inappropriate point, leading to an invalid control flow, unpredictable behavior, and even return-oriented programming (ROP) attacks [32]. Forcing a reset ensures that execution of the new code is properly initiated. POR syntax and security goals are formalized in Definitions 2 and 3, respectively. PURE POR scheme is presented in Construction 1. We prove that it is secure in Section 5.1 and formally verify its implementation in Section 5.2.

\textbf{Definition 2 (syntax).} POR is a tuple (Request, Reset, Verify) of algorithms:

- \text{Request}^{\text{rf} \rightarrow \text{P}_r}() : \text{algorithm initiated by } \text{Vrf} \text{ at time } t \text{ to request a proof that } P_{rv} \text{ has performed a reset at some point } t' \text{ in time, where it must hold that } t' > t. \text{ As a part of } \text{Request, } \text{Vrf} \text{ sends a challenge to } P_{rv}.

- \text{Reset}^{\text{P}_r \rightarrow \text{Vrf}}(\text{Chal}) : \text{algorithm executed by } P_{rv} \text{ to perform a reset and use Chal to provide an unforgeable proof that a reset has happened.}

- \text{Verify}^{\text{Vrf}}(H, \text{Chal}) : \text{algorithm executed by } \text{Vrf} \text{ upon receiving } H. \text{ It outputs } 1 \text{ if } H \text{ is a valid proof in response to } \text{Request. Otherwise, it outputs } 0.

\textbf{Definition 3.}

3.1 POR Security Game (POR-game): Challenger plays the following game with $\text{Adv}$:

(1) $\text{Adv}$ is given full control over $P_{rv}$ software state and oracle access to Reset calls.

(2) At time $t$, $\text{Adv}$ is presented with Chal.

(3) $\text{Adv}$ can produce $H_{\text{Adv}}$, such that $\text{Verify}(H_{\text{Adv}}, \text{Chal}) = 1$, without causing $P_{rv}$ to reset.

3.2 POR Security Definition: A POR scheme is considered secure if for all PPT adversaries $\text{Adv}$, there exists a negligible function negl such that:

\[
P_r[\text{Adv, POR-game}] \leq \text{negl}(1)
\]

5.1 Security Proof for Construction 1

\textbf{Theorem 1.} Construction 1 is a secure POR implementation according to Definition 3 provided that HMAC is a secure MAC.

\textbf{Construction 1 (PURE POR).} Suppose PoR.C is the program memory region inside the TCB storing PoR software binary PoR.C $\in$ TCB. Construction of PURE PoR is defined as follows:

- \text{Request}^{\text{rf} \rightarrow \text{P}_r}() : \text{Vrf generates a random challenge Chal} \leftarrow \{0, 1\}^l \text{ and sends it to } P_{rv}.

- \text{Reset}^{\text{P}_r \rightarrow \text{Vrf}}(\text{Chal}) : \text{(1) Use VRASED’s HMAC to compute HMAC($K$, Chal), and write the result at } RSP_{\text{RESULT_ADDR}}, \text{ where } RSP_{\text{RESULT_ADDR}} \text{ is in a persistent storage (flash) memory region.}

(2) Enforce the following LTL invariant:

\[
G \cdot (PC \rightarrow (PC \in \text{TCP} \land \neg \text{DMA}_{\text{en}}) \land \text{reset})
\]

- \text{Verify}^{\text{Vrf}}(H, \text{Chal}) : \text{Vrf returns } 1 \text{ if and only if } H = \text{HMAC($K$, Chal).}
which breaks the existential unforgeability of HMAC [33].

5.2 PoR Verified Implementation

In Section 5.1 we show that Construction 1 is a secure PoR scheme as long as we can formally verify that the underlying implementation adheres to LTL 1 defined in Construction 1. To guarantee LTL 1 with minimal modification to VRASED, we rely on Lemma 1.

**Lemma 1.**

\[
V R A S E D \land \forall G : ([PC = f s t(PoR.C)] \land (PC = I st(TCB)) \rightarrow (PC = I st(TCB)) \lor reset) \quad (2)
\]

where \( V R A S E D \) denotes the set of LTL properties already guaranteed by VRASED.

**Proof.** The intuition behind Lemma 1 is that \( V R A S E D \) already guarantees that TCB code can not be interrupted before reaching its last instruction. This property is referred to as "atomicity" and specified in LTL as:

\[
G : (\neg reset \land (PC \in TCB) \land (X(PC) \in TCB) \rightarrow PC = I st(TCB) \lor X(reset)) \quad (3)
\]

\[
G : (\neg reset \land (PC \in TCB) \land (X(PC) \in TCB) \rightarrow X(PC) = f s t(TCB) \lor X(reset)) \quad (4)
\]

LTLs 3 and 4 express that the only valid entry point for VRASED software is the first instruction in the TCB, the only legal exit is via the last instruction in the TCB. Since \( f s t(PoR.C) \in TCB \) and LTL 4 is guaranteed, it suffices to enforce that:

\[
G : ([PC = f s t(PoR.C)] \land (PC = I st(TCB))) \rightarrow (PC = I st(TCB)) \lor reset) \quad (5)
\]

**Remark:** The above discussion provides some intuition for the proof of Lemma 1. The actual proof was automated and machine-checked using Spot [34], an LTL theorem prover. The formal computer proof is available at [11].

Figure 3 shows a formally verified FSM corresponding to the Hardware module enforcing LTL 5. This FSM works by monitoring the PC value. Normally \( PC = I st(TCB) \) should not trigger a reset. In the PoR case, however, once \( PC = f s t(PoR.C) \), this state machine transitions into the PoR state. Once there, a future state with \( PC = I st(TCB) \) always triggers a reset. Formal verification of such hardware module adopts the same methodology as used in VRASED. We design the FSM in Verilog HDL and automatically translate into SMV using Verilog2SMV [29]. Finally, we use the NuSMV model checker [28] to automatically prove that the state machine complies with LTL 5.

5.3 PoR Correctness

We design the PoR scheme to co-exist with VRASED remote attestation functionality. In Figure 3, a reset is only triggered at \( PC = I st(TCB) \) after the Reset functionality is invoked. This way, the original VRASED remote attestation is unaffected by the added rules.

PoR security implies that it is impossible for untrusted code to produce H without resetting, and any attempt to do so triggers a reset. We now consider PoR correctness. Figure 4 shows the C implementation of the software TCB, including both VRASED RA
Figure 4: VRASED software TCB extended to incorporate PoR functionality

and PoR. Line 7 corresponds to Ist(PoR). If nothing goes wrong, HMAC executes and terminates at Ist(PoR). In PoR’s FSM of Figure 3, PC = Ist(PoR) always triggers a reset, which is the expected behavior after completion of HMAC. After a reset, H is accessible at RST_RESULT_ADDR persistent memory address (Line 6 in Figure 4), and Prv can forward it to Vrf.

5.4 Practical Considerations

After a reset, the (unprivileged) MCU application always begins execution anew. It must start by checking whether something (H) is stored at RST_RESULT_ADDR. If so, the first action is to send H to Vrf and zero-out RST_RESULT_ADDR, thus completing the PoR protocol. We note that this code is malleable and is part of the TCB. In particular, it can always decide not to reply to Vrf. However, lack of response would cause Vrf to realize that something is wrong with Prv.

6 PROOFS OF (REMOTE SOFTWARE) UPDATE (POU)

We consider a Proof of (Remote Software) Update (POU) scheme involving two parties: Prv and Vrf. The goal is to allow Vrf to request a remote software update for Prv, and to obtain an unforgeable proof upon a successful update. The generic description of POU scheme is presented in Definition 4.

Definition 4 (syntax). A POU scheme consists of a tuple (Request, Install, Verify, Request) satisfying the following:

- Request\textsuperscript{Vrf} → Prv (S): algorithm initiated by Vrf to request software S to be installed on Prv at a memory range UR, where |UR| = |S|. Prv also receives a challenge Chal as a part of the request.
- Install\textsuperscript{Prv} → Vrf (Chal, S): algorithm executed by Prv to update software such that UR = S. Upon successful update, it outputs a proof H that the update has completed.
- Verify\textsuperscript{Vrf} (H, Chal, S): algorithm executed by Vrf upon receiving H. It outputs 1 if H is a valid proof in response to Request. Otherwise, it outputs 0.

Based on this syntax, we define a security game in Definition 5 that captures our Adv model (discussed in Section 4). Intuitively, security of POU means that Adv cannot compute a valid proof of successful update for software that is not installed on Prv. In other words, Adv wins the game if it can generate a valid proof of update without actually updating UR memory on Prv with S.

5.1 POU Security Game (POU-game):

Notation:
- UR is the memory region on Prv that should be updated, and UR(t) denotes the content of UR at time t.
- Challenge plays the following game with Adv:
  1. Adv is given full control over Prv software state and oracle access to Install.
  2. Adv is presented with Chal and S, and continues to have oracle access to Install.
  3. Eventually at time t, Adv queries Install and outputs H.
  4. Adv wins if and only if Verify(H, Chal, S) = 1 and UR(t) ≠ S.

5.2 POU Security Definition:

A POU scheme is secure if, for all Probabilistic Polynomial-Time (PPT) adversaries Adv, there exists a negligible function negl such that:

\[ Pr[Adv, POU-game] ≤ negl(1) \]

6.1 Security Proof for Construction 2

Theorem 2. Construction 2 is secure according to Definition 5.2 as long as VRASED is secure according to Definition 1.

Proof. The proof is by reduction of VRASED security in Definition 1 to Construction 2’s security. Suppose there exists adversary Adv such that \( Pr[Adv, POU-game] > negl(l) \). We then show that such Adv can be used to construct \( \text{Adv}' \) that breaks VRASED security in Definition 1. \( \text{Adv}' \) acts by selecting the same S and output σ used by Adv to win POU-game. Recall from Definition 5 and Construction 2 that, since S and σ allow Adv to win POU-game, it must hold that:

\[ σ = \text{HMAC}(\text{KDF(K, Chal)}, S) \text{ and } \text{UR}(t) ≠ S \]

According to Definition 1 and the fact that \( UR = AR \), (S, σ) is a valid response in RA-game. Therefore, \( \text{Adv}' \) is able to win RA-game.
with the same non-negligible probability that \( \mathcal{A}_{\text{adv}} \) has of winning PoU-game.

\[ \square \]

Note that Construction 2 uses VRASED as-is and involves no additional trusted software or hardware. Therefore, there is no additional effort for verifying its implementation. Also, Construction 2 requires unprivileged helper code to receive \( \text{Chal} \) and \( S \), write them to memory, and invoke VRASED remote attestation functionality. This helper code is outside the TCB and has no impact on PoU security. If \( \mathcal{P}_{rv} \) is infected and malware refuses to perform the update, such anomalous behavior is always detected by \( \mathcal{V}_{\text{rf}} \) due to either lack of any response, or an invalid response \((UR \neq S)\).

### 6.2 Practical Considerations

PoU security relies on \( \mathcal{A}_{\text{adv}} \)'s inability to produce a valid \( H \) without installing correct new software on \( \mathcal{P}_{rv} \). However, it does not prohibit a network \( \mathcal{A}_{\text{adv}} \) from spoofing PoU's request. Such an attack could lead \( \mathcal{P}_{rv} \) to inadvertently install malicious software chosen by \( \mathcal{A}_{\text{adv}} \), while believing that the request came from \( \mathcal{V}_{\text{rf}} \). This can be easily prevented by authenticating \( \mathcal{V}_{\text{rf}} \)'s request before performing PoU. Fortunately, VRASED already supports \( \mathcal{V}_{\text{rf}} \) authentication functionality as an optional part of its design; see Appendix B in [9]. This functionality can be seamlessly incorporated into PURE PoU construction to prevent installation of unauthorized software.

Another practical consideration is the need to protect the helper code. Recall that our construction requires it to receive the challenge, install the update and send \( H \) back to \( \mathcal{V}_{\text{rf}} \). As we argue in Section 6.1, this helper code does not have to be verified or trusted, since it does not impact PURE PoU security. Nonetheless, if helper code is not write-protected, malware can infect \( \mathcal{P}_{rv} \), modify or erase it, and simply leave \( \mathcal{P}_{rv} \). This simple attack would prevent \( \mathcal{P}_{rv} \) from receiving and performing any subsequent update, and consequently put \( \mathcal{P}_{rv} \) into a non-updatable state, which would require physical access to remedy. This might be problematic in settings where physical access to \( \mathcal{P}_{rv} \) is hard. This attack can be easily circumvented by making helper code immutable, by either placing it in ROM, or adding MCU access control rules to enforce its immutability.

### 7 PROOFS OF MEMORY ERASURE (PoE)

Several scenarios require a proof of memory erasure (PoE). For example, it can be used to ensure deletion of sensitive information when \( \mathcal{P}_{rv} \) is no longer of use, or whenever there is a change of its ownership. Alternatively, PoE can be used in conjunction with a PoU to ensure that \( \mathcal{P}_{rv} \) is in a pristine (malware-free) state before installing a software update.

PoE is also a two-party protocol between \( \mathcal{V}_{\text{rf}} \) and \( \mathcal{P}_{rv} \), where the former requests erasure of a specific memory region on \( \mathcal{P}_{rv} \) to be erased, i.e., zeroed-out. Based on this informal definition, PoE can be viewed as a special case of PoU, where newly installed software \( S \) is a priori known default value, e.g., a string of zeros \( S = \{0\}^{UR} \). \( S \) can be compressed and represented using a single bit, i.e., \( 0 \), during transmission in Request.

Since PoE is a subset of PoU, the PoE security definition is derived from Definition 5 by substituting arbitrary \( S \) with \( \{0\}^{UR} \). We present PURE PoE scheme in Construction 3. Because its security proof follows the same arguments as that of Theorem 2, we omit it.

#### 7.1 Practical Considerations

A typical use case of PoE is secure erasure of \( \mathcal{P}_{rv} \) data memory, ensuring that no sensitive data remains. Nonetheless, it can also be used to provide a proof of absence of malware on \( \mathcal{P}_{rv} \) by setting \( UR \) to \( \mathcal{P}_{rv} \)'s entire writable memory, both program and data segments. Since, akin to PoU, PoE needs helper code to receive the challenge and return the proof to \( \mathcal{V}_{\text{rf}} \), wiping out all of \( \mathcal{P}_{rv} \)'s writable memory would imply erasing this helper code as well, which is obviously problematic. Naïvely, \( \mathcal{P}_{rv} \) can simply avoid erasing helper code. However, this does not guarantee malware-free state, since malware might manage to hide itself in the helper code region. The most intuitive remedy is to make helper code immutable, as with PoU, e.g., by housing it in ROM. Alternatively, we can leave helper code be, and attest its presence (along with zeros everywhere else). However, this alternative approach would require incorporating this helper code into PoE's definitions as well as modifying PURE PoE construction and its proof accordingly.

### 8 EVALUATION

We implemented PURE by extending VRASED with the ability to perform PoU, PoE, and PoR schemes. Our implementation is synthesized and executed using the commodity FPGA prototyping board Basys3. We report PURE additional overhead compared to an unmodified VRASED architecture.

**Hardware.** Table 2 shows PURE hardware overhead. Compared to VRASED, the number of Look-Up Tables (LUT) grows by 4 and the number of registers – by 3. With respect to unmodified OpenMSP430 architecture, this represents an overhead of 0.2% for LUT and 0.4% for registers. This is in addition to VRASED overhead of \( \approx 6.6\% \) and \( \approx 5.8\% \) for LUT and registers, respectively.

**Memory.** VRASED requires \( \approx 4.5\text{KB} \) of ROM, most of which is for storing HACL* HMAC-SHA256 code. Since PURE re-uses the same HMAC implementation, additional memory is minimal: 50 bytes of ROM. Unprivileged helper code, for installing update and erasing memory requires additional 26 bytes of program memory. Execution of PURE services requires only slightly more RAM: 4
UC Irvine authors’ work was supported in worthy applications, protocols, and systems that use these services (2) implementing and formally verifying these services for medium- and higher-end embedded systems; and (3) introducing new trust-worthy: (1) extending proposed services to swarms/groups of devices; \( P \) and \( O \) runtimes are similar, linear in terms of updated/erased memory size. Due to the execution of their helper codes, their runtimes are slightly higher (at most 1.6% higher) than the attestation runtime.

### 9 CONCLUSIONS

This paper described PURE: a set of three provably secure services targeting low-end embedded devices: system-wide reset, erasure, and software (more generally, memory) update. We argue that serial composition of these services allows 'Vrf to ensure that a remote \( P \) has been re-configured to a functional and malware-free state, and properly re-initialized. PURE was implemented on a popular commercially available low-end embedded system, MSP430, and synthesized using a commodity Basys3 FPGA. Its additional hardware footprint overhead is \(\approx 0.4\%\) and it takes on the order of 100-s milliseconds to generate proofs of update, erasure, and reset, altogether.

This paper motivates several directions for future work, including: (1) extending proposed services to swarms/groups of devices; (2) implementing and formally verifying these services for medium- and higher-end embedded systems; and (3) introducing new trustworthy applications, protocols, and systems that use these services as building blocks.

**Acknowledgments:** UC Irvine authors’ work was supported in part by ARO under contract: W911NF-16-1-0536, and NSF WiFUS Program Award #: 1702911. Authors thank ICCAD anonymous reviewers for their valuable comments.

### REFERENCES