IC Activation and User Authentication for Security-Sensitive Systems

Jiawei Huang
Charles L. Brown Department of
Electrical and Computer Engineering
University of Virginia
Charlottesville, VA, USA
jh3wn@virginia.edu

John Lach
Charles L. Brown Department of
Electrical and Computer Engineering
University of Virginia
Charlottesville, VA, USA
jlach@virginia.edu

***Abstract*—A number of applications depend on the protection of security-sensitive hardware, preventing unauthorized users from gaining access to the functionality of the integrated circuits (ICs). Failure to protect such devices can have consequences ranging from the loss of financial revenue to the loss of human lives. The key to providing protection does not lie in the prevention of theft but in a secure IC activation and user authentication procedure so that an adversary has nothing to gain by acquiring the physical hardware.**

The proposed protection scheme is robust against various types of malicious attack, such as reverse engineering to extract the circuit layout, brute-forcing the access key, and FIBing. The scheme provides the capability both for one-time or every-powerup activation and for every-powerup user authentication. Given the resource constraints of many security-sensitive hardware systems (such as those deployed in remote locations or carried in combat arenas), this paper proposes and evaluates the cost of several techniques for achieving secure activation and authentication.

Keywords- IC activation, user authentication, security-sensitive hardware protection, reverse engineering, PUF, signature generation

I. INTRODUCTION

Security-sensitive systems, such as military equipment and weapons, classified scientific research equipment, and financial systems, require stringent protection measures. Although their design is never to be disclosed, it is still constantly subject to theft by adversaries and attackers. Even after the production of such hardware, these devices may be deployed out in the field, such as military operations and field experiments. Enemy scientists can study those deserted ICs and extract their design. Even worse, they can simply use the devices for their own purposes, including attacks against the rightful owners.

It is therefore desirable to have such ICs require activation upon use and have that activation be performed by a trusted party. In deployment scenarios with one-time activation (as opposed to every-powerup re-activation), an additional layer of security in the form of user authentication may be required. Both activation and authentication are usually achieved using passwords that are checked against values embedded inside the IC, and those passwords must be device specific to prevent an attacker from using one password to activate and/or authenticate himself on multiple copies of the same IC. However, unlike software, fabricated hardware cannot be easily modified, and given the extremely high cost of fabrication masks, it is impractical to fabricate-by-design each device so that it has a unique key. The activation and authentication techniques should be provably secure against common attacks (such as brute-forcing the password), and they should only introduce trivial design effort and cost to the original device.

This paper presents an approach to achieving this IC protection for security-sensitive systems. Under our scheme, before the activation is done (pre-activation stage), the IC is effectively locked (nonfunctional). The first-time user needs to contact the original designer or a trusted party for activation. The IC can be designed for one-time activation or so that it needs to be re-activated at every powerup. In the latter, the approach supports user authentication so that only the user who participated in the activation can access the IC without again contacting the activation controller.

The approach makes use of manufacturing variability (MV) as a method to generate a unique activation and authentication signature for each IC. MV is commonly seen as a major problem in the IC industry because it makes performance and power unpredictable and can negatively impact manufacturing yield. However, from the security perspective, MV allows unique signature generation to come for free without any design-time or manufacturing effort or cost. A unique signature can be extracted from delay and power characteristics [1, 2] for every single IC. In an attempt to further reduce the cost of the security features so that it is negligible compared to the main IC functional blocks; this paper suggests several implementation considerations.

The security scenarios addressed in this paper are actually special cases of the more general problem of IC intellectual property (IP) protection. Therefore, this approach can also be applied to protection of IP used in consumer electronics. Under this scheme, IP owners are guaranteed royalties for every working IC built using their IP, because any user has to contact the IP owner for device activation. It therefore also enables hardware metering, allowing a fabless IP designer to keep track of the actual number of ICs in use in the market. Thus, it is easier to detect if a manufacturer has produced an extra number

of ICs than the IP owner has requested. This is a common channel through which illegal copies of an IC enter the market, usually sold at a much lower price.

The rest of this paper is organized as follows: Section 2 presents related work performed by other researchers on IC protection, including a closely related paper on IP protection. This is followed by Section 3, which provides an overview of the activation and authentication approach, detailing its working mechanism. Section 4 discusses the implementation of the two key components in the mechanism – the cryptosystem and the signature extraction units, with a focus on area-efficiency. A security-based robustness assessment of the approach is detailed in Section 5, followed by a summary and future work discussion in Section 6.

II. RELATED WORK

As mentioned above, the security scenarios addressed in this paper are a special case of general IP protection. Roy et al. recently proposed EPIC (Ending Piracy in Integrated Circuits) to ensure that IP designers received royalties for every copy of the IC in use in the field [3]. The key idea of EPIC is to embed a combinational locking mechanism inside the IC circuitry so that all fabricated ICs are initially disabled. Direct interconnections on non-critical combinational paths are interrupted with logic gates so that only when the unlocking key is applied will the circuit behave correctly. A random IC key pair (RCK) is then generated on initial powerup and serves as the public and private key pair for the IC. Another public and private key pair (MK) is generated for the IP holder before fabrication, and that private key is known only to the IP holder. A secure communication channel can be setup between the two parties by exchanging each other's public key while keeping the private key to themselves. Through this channel, a common key (CK), which is generated before fabrication and is known only to the IP holder, can be transmitted from the IP holder to the IC to unlock the combinational lock.

This protection mechanism is robust against a number of typical attacks. Guessing an IP holder's private key is equivalent to breaking the RSA public-key crypto-system, which is known to be hard. Reverse engineering the IC will not help the adversaries because it is difficult to locate the combinational locking circuitry. Due to manufacturing variations, different ICs nearly always have different key pairs so that the correct input to unlock one IC will not work on any others. This makes eavesdropping on data communication useless. Focused Ion Beam (FIB) can be used to bypass the combinational lock and directly activate the IC, but layout techniques have been shown to be effective against FIBing.

Though the technique presented in this paper shares many similarities with EPIC (in fact, it was developed in parallel with the EPIC work to address similar problems), it is different in several major design choices and provides an example on how to implement the protection scheme in a resource-constrained environment. It also incorporates device activation and user authentication in a single protection scheme, a desirable feature for security-sensitive applications.

Simpson and Schaumont (S&S) presented a powerful technique for offline HW/SW mutual authentication for FPGA-based systems [4]. Despite sharing a number of similarities (a similar two-party activation scenario is a natural extension of S&S's three-party authentication scenario), the two protection schemes are actually fundamentally different in several key aspects. Similar to processor-based protection mechanisms running encrypted binaries, S&S protects what is configured onto the device (i.e. the FPGA bitstream, without which the system is useless), not the device itself. The encrypted bitstream (referred to as SW in [4]) is the message being transmitted from the IP provider to the system developer, and the key idea is that only the authorized FPGA device can decrypt the bitstream and configure the device correctly.

S&S has two key issues when transposed into the two-party activation scenario addressed in this paper. First, it is not clear from [4] if the system developer is able to easily generate its own challenge-response pairs (CRPs) for each device. If it is possible, the system developer could extract the device-specific key that was used to encrypt the bitstream and decrypt it externally with the defined symmetric algorithm. This attack could then be used to configure (i.e. activate) any copy of the device, something that poses a great risk in security-sensitive systems. Second, the activation control (a combination of S&S's trusted third party and IP provider in this two-party scenario) must send the system developer the challenge to be used to decrypt the bitstream. Therefore, an attacker overhearing the communication who is able to gain access to the physical device can then apply the challenge and activate the device himself. As discussed in the following sections, in addition to being used for fixed-logic ICs instead of FPGAs, our method avoids these issues in large part by protecting the device itself (not what is configured onto it), using asymmetric cryptographic techniques, and keeping a secret user key that is never transmitted.

III. DESIGN OVERVIEW

Figure 1 details the IC activation and user authentication schemes providing protection to security-sensitive systems in a variety of usage and attacker scenarios. Consider, for example, a scenario in which Bob is a friendly soldier given a security-sensitive system for use in a combat arena, and the potential consequences of enemy soldier Eve obtaining and using this system (or any copy of the system) are dire. Therefore, the technology should require (possibly remote) activation from Alice, the Activation Control at headquarters. In addition, once the technology is activated, it should only be usable by Bob so that Eve is not able to use an already-activated technology without user authentication.

Design-time: The densely shaded box in Figure 1 represents the boundaries of the physical IC, and the lightly shaded box embedded within the boundary represents the Protected Core that is not functional without the release from the Unlocking Mechanism. Therefore, the IC designer selects a number of non-critical combinational paths to be embedded with EPIC's combinational locking mechanism [3] so that only when the two inputs to the comparator block (=?) are equal will the Protected Core function correctly. In addition, a private key (CA), public key (DA) pair is generated. DA is hardwired on-chip, while CA is kept secret and only provided to Alice as the

Activation Control. Asymmetric cryptography is utilized because DA is susceptible to reverse engineering attacks [5].

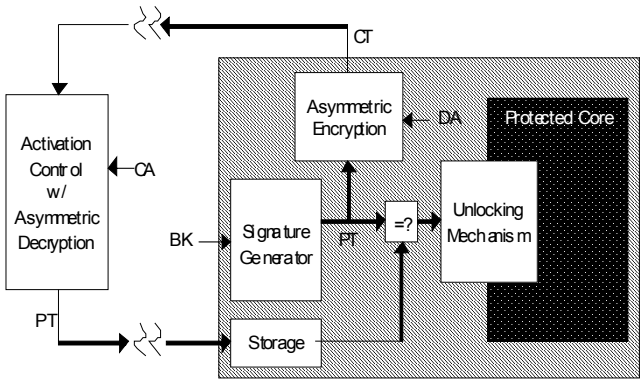

Figure 1. IC activation and user authentication

First-time activation: On initial startup (possibly in the field), Bob creates a secret key (BK) known to no one else, not even Alice. Taking BK as input, the Signature Generator creates an effectively random number (PT), based on BK and the non-functional characteristics of IC, such as combinational path delay. (As discussed in more detail in Section 4.2, even with the same BK input, the PT for each IC is unique based on manufacturing variations.) The PT is compared to the content of the Storage block, and the Comparator (=?) output is used to drive the embedded locking mechanism. Since the storage is by default empty (all zeros), the comparison will not produce the correct unlocking vector, as the content of the Storage is not equal to PT.

PT is not directly observable by Bob (circuit layout techniques have been shown to be effective preventing end-user FIB attacks that change interconnections on fabricated circuits to make PT observable), so Bob must send CT (PT encrypted with DA using an ECC encryption algorithm) to Alice, who is the only one who can determine PT from CT using CA. Once Alice has authenticated Bob as an approved user with existing techniques, she decrypts CT and sends PT back to Bob. Because PT is unique for each IC, Alice's response will only be useful to unlock the specific IC in question, and Eve gains no benefit from overhearing the exchange between Alice and Bob unless she is able to access BK, which is never transmitted.

User authentication: After receiving PT from Alice, Bob writes it to the Storage block. Now every time Bob wants to use the device, he simply enters the same BK that was used to generate PT. This establishes an agreement between PT from the Signature Generator output and the content within Storage, and the IC will resume normal operation. The Storage block can be volatile or non-volatile memory, depending on the application scenario. If it is volatile, Bob or another user can do another iteration of the activation process with Alice at every powerup.

However, many application scenarios make it impractical for users to contact Alice at every powerup, which gives rise to the need for user authentication of an already activated devise. Using volatile Storage (or no storage at all – the PT input pins can be routed directly to the Comparator), Bob and only Bob can apply PT to the memory and BK to the Signature Generator at powerup. Similarly, if PT is permanently stored in non-volatile Storage, Bob and only Bob can apply BK to the Signature Generator to unlock the circuit. As stated above, Eve gains no benefit from just overhearing PT, as she will not have the associated BK.

IV. RESOURCE-CONSTRAINED IMPLEMENTATION

Given that many application scenarios of interest have severe resource constrains, it is essential to consider the efficiency of the proposed security mechanisms.

A. Public-Key Cryptosystem

For most public-key crypto-systems, the decryption process is significantly more computationally intensive than encryption. In EPIC, two decryption procedures need to be integrated on-chip. Also on chip is a unit to generate a pair of public and private keys. Our design only requires an encryption module on-chip. The decryption is off-loaded to Alice, who presumably has more computational resources at headquarters. This approach allows for a more compact design on chip without sacrificing the level of security.

In addition, ECC crypto-system is preferable over RSA and Rabin because it offers a similar level of security but with much shorter keys (Table 1), even though ECC and RSA achieve approximately the same area efficiency [6][7][8][9]. However, PT and CT are also shorter as a result. This leads to a smaller design of the Signature Generator, because the random number (PT) to be generated is now shorter.

Table 1. Key size comparison of popular public-key cryptosystems (with approximately same key strength, equivalent to 112-bit symmetric-key algorithm)

Equivalent Key Size		
RSA	Rabin	ECC
2048	2048	224

We have implemented a 256-bit ECC logic unit over prime field. Synthesized onto a Xilinx Virtex II FPGA, it occupies 18,123 slices and has an estimated 349K equivalent gate count. The design runs at a maximum operating frequency of 34.4 MHz, and one ECC encryption takes 777K clock cycles, for a throughput of 44 encryptions per second. Note that performance is not a critical metric of interest with this security mechanism, as long as the encryption completes within an acceptable time span from a human perspective. It is therefore reasonable to sacrifice performance to achieve ultra-low area and power implementations. Other low area ECC implementations include Daneshbeh et al. [10], which details an ECC processor with an estimated area of 35K equivalent gates that performs 1300 256-bit scalar multiplications per second. Satoh et al. [11] designed a scalable ASIC ECC processor that can be used in both prime and binary fields. A more compact 256-bit version requires only 29.7K gates and performs 80 scalar multiplications per second. Satoh et al.'s work also has a good scalability – for a fixed-size multiplier, the circuit size grows linearly with field size, while performance decreases exponentially. The additional circuit size all comes from memory. Linear scalability of circuit size

with field size is desirable because the security of a key size becomes vulnerable over time and must eventually be enlarged.

B. *Unique Signature Generation*

The job of the Signature Generator is to produce a unique signature for every IC based on manufacturing variability. The possibility of two ICs having the same signature or one IC producing different signatures given the same input should be statistically negligible (uniqueness and stability). In addition, it should be difficult to obtain the input based on output (irreversibility).

Physical Unclonable Functions (PUFs) have been suggested to generate such signatures based on the physical characteristics of ICs rather than storing the signatures in addressable memory [12]. It is more difficult for an adversary to attack a PUF than a digital memory, because signatures generated by PUFs are volatile and the action of probing a PUF for values itself may destroy the device itself. However, the output of a PUF is strongly influenced by environmental factors and device aging, thus robustness becomes an issue. For example, the exact CRPs may not be replayed if the temperature is changed. Metastability of memory elements on initial powerup can also be used for unique signature generation. A fully integrated True Random Number Generator [13] hashed with user input, as shown in Figure 2, is a very simple example.

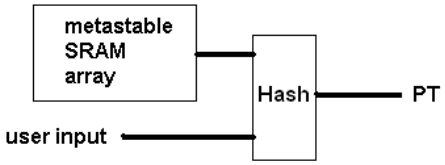

Figure 2. Signature generation using metastability of memory elements

The robustness issue with PUFs can also be resolved by writing its output to a one-time writable, non-volatile, non-addressable storage upon initial powerup, creating a random, automatically-generated (and therefore not susceptible to a supply chain attack) serial number for every device. Then we can hash that number with the user input (and possibly an internal counter or linear feedback shift register (LFSR) for additional dynamic behavior as discussed in Section 5) in a similar fashion as depicted in Figure 2, thus creating the capability for generating stable CRPs without risking exposure of the serial number. Extracting the serial number is possible through physical reverse engineering [5], but that destroys the device and does not help unlocking any other copy of the device.

V. ROBUSTNESS ASSESSMENT

Given the security-sensitive nature of the systems being protected by these activation and authentication techniques, it is important to assess the resilience against attacks.

Passive attacks rely purely on observation. The mathematical foundations of ECC and RSA guarantee that observing the public key will not reveal any information about its private counterpart. Brute-forcing the private key is unrealistic because of the intractable computational effort. Chosen ciphertext attack is useless because Alice will never respond to Eve's requests if she is not an authorized user. Unless decrypted with the private key (which is never transmitted), calculating PT from CT is equivalent to solving the elliptic curve discrete logarithm problem (ECDLP), which is believed to be hard.

Probing internal wires for the PT value is also difficult for several reasons. To expose wires embedded in deeper layers (done so intentionally during layout), an adversary has to remove all the above layers, thus destroying the IC and not providing any useful information for activating other copies of the IC.

Eavesdropping on the input of Bob and his communication with Alice is another potential threat. This threat can be easily avoided by making the challenge (BK) unique, so that a challenge is never asked a second time. This could be easily achieved by associating the challenge with an internal counter or LFSR, which increments on each activation attempt. A constraint of this technique is that even the authorized user now has to contact Alice upon every powerup.

Active attacks involve alteration of information being transmitted. Chosen ciphertext attack is possible by means of altering the communication between Alice and Bob. This issue can be resolved by encrypting the communication with a user specified key pair that is orthogonal to the proposed technique.

Physical attacks are the utmost threat to the integrity of our protection scheme. It is also the most expensive and technologically sophisticated. FIBing can alter the wiring of a manufactured IC so that the security features can be completely bypassed. However, as with probing attacks, the security interconnections can be buried under metal layers containing interconnections vital to the correct operation of the IC.

It is also possible to reverse engineer an IC by peeling off metal layers and extracting the circuit functionality [5]. Not only is this challenging given the proposed methods of integrating the security and functionality (as argued in [3]), but a successful reverse engineering would require the re-fabrication of new ICs, something that is extremely costly and time consuming.

VI. SUMMARY AND FUTURE WORK

In this paper, we presented an approach to security-sensitive hardware protection that combines device activation and user authentication while introducing only small area overhead. Future work includes a quantitative evaluation of the cost to upgrade to a larger key size to help make projections on the design cost scalability. The robustness of this design has also yet to be quantified, requiring extensive Red Team/Blue Team experiments.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grants CNS-0716443 and IIS-061204. The authors would like to thank Karsten Nohl from the University

of Virginia and the anonymous reviewers for their helpful suggestions.

REFERENCES

[1] B. Gassend, D. Clarke, M. van Dijk, S. Devadas, "Silicon Physical Random Functions," *ACM Conference on Computer and Communications Security*, pp. 148-160, 2002

[2] J. Li, J. Lach, "Negative-Skewed Shadow Registers for At-Speed Delay Variation Characterization," *IEEE International Conference on Computer Design*, pp. 354-359, 2007

[3] J. A. Roy, F. Koushanfar, I. L. Markov, "EPIC: Ending Piracy of Integrated Circuits," *Design Automation and Test in Europe*, pp. 1069-1074, 2008

[4] E. Simpson, P. Schaumont, "Offline HW/SW Authentication for Reconfigurable Platforms," *Workshop on Cryptographic Hardware and Embedded Systems*, pp. 311-323, 2006

[5] www.chipworks.com

[6] K. H. Leung, K. W. Ma, W. K. Wong, P. H. W. Leong, "FPGA Implementation of a Microcoded Elliptic Curve Cryptographic Processor," *IEEE Symposium on Field-Programmable Custom Computing Machines,* pp. 68-76, 2000

[7] T. Wolligner, J. Guajardo, C. Paar, "Cryptography on FPGAs: State of the Art Implementations and Attacks," *ACM Transactions in Embedded Computing Systems*, vol. 3, no. 3, pp. 534-574, 1999

[8] M. Ciet, M. Neve, E. Peeters, J. Quinsquater, "Parallel FPGA Implementation of RSA with Residue Number Systems," *IEEE Midwest Symposium on Circuits and Systems,* vol. 2, no. 2, pp. 806-810, 2003

[9] A. Mazzeo, L. Romano, G. P. Saggese, N. Mazzocca, "FPGA-based Implementation of a Serial RSA Processor," *Design, Automation and Test in Europe,* pp. 582-587, 2003

[10] A. K. Daneshbeh, M. A. Hasan, "Area Efficient High Speed Elliptic Curve Cryptoprocessor for Random Curves," *International Conference on Information Technology: Coding and Computing*, vol. 2, pp. 588-592, 2004

[11] A. Satoh, K. Takano, "A Scalable Dual-Field Elliptic Curve Cryptographic Processor," *IEEE Transactions on Computers*, vol. 52, no. 4, pp. 449-460, April 2003

[12] G. E. Suh, S. Devadas, "Physical Unclonable Functions for Device Authentication and Secret Key Generation," *Design Automation Conference,* pp. 9-14, 2007

[13] C. Tokunaga, D. Blaauw, T. Mudge, "True Random Number Generator with a Metastability-Based Quality Control," *IEEE Journal of Solid-State Circuits,* vol. 43, no. 1, pp. 78-85, Jan 2008