Overview of research results on hardware-accelerated cryptography and security

Alessandro Cilardo
Department of Computer and System Engineering, University of Naples Federico II
via Claudio 21, 80125 Napoli, Italy, Email: acilardo@unina.it

***Abstract*—This paper provides an overview of the research findings related to cryptographic hardware, acceleration of cryptanalytical algorithms, FPGA design automation and testing, as well as security service provisioning achieved by the author until the time of writing. The paper also refers to a few results developed in the framework of funded research projects which involved the author as a team member. The text briefly describes the implications of the main research results, indicating the corresponding publication and the essential insights behind each work.**

I. Introduction

This document is meant to report on the main research activities by the author until year 2011, and it also covers results developed in the context of a couple of funded research projects indicated at the end of the paper. The contents of the report can be grouped in three main areas:

- Design of advanced hardware blocks for cryptographic and cryptanalytical applications, in most cases targeting Field Programmable Gate Arrays (FPGAs), although a few results relied on Application Specific Integrated Circuits (ASICs).
- Investigation of new techniques for FPGA design automation and testing, including methodologies for early estimation of hardware complexity in high-level synthesis toolflows as well as application-dependent testing of hardware-reconfigurable devices.
- Implementation, deployment, and evaluation of security services, particularly public-key certification and digital time stamping, provisioned to heterogeneous mobile devices in distributed environments.

This paper is organized as follows. Section II deals with cryptographic hardware and acceleration of cryptanalytical algorithms. Section III addresses FPGA design automation and testing. Section IV presents the main results on security service provisioning. Section V gives a few conclusive comments.

II. Acceleration of security-related operations

One important branch of the activity addressed the high-performance and secure implementation of cryptographic routines. In [1] the author presented a hardware implementation of the Rivest-Shamir-Adleman (RSA) public-key cryptography algorithm, relying on a novel solution applying Montgomery's multiplication to carry-save represented numbers. The work presented an FPGA-based implementation using a modular bit-sliced hardware scheme for the execution of modular exponentiation on large integers, the key operation in RSA, achieving considerable performance improvements compared to other solutions in the literature.

In a subsequent work [2], the author presented an architecture and an FPGA-based implementation exploiting a *digit-serial* approach, enabling shorter critical paths and lower area requirements, although operations are serialized and hence take more clock cycles. The work explored several area cost vs. time performance trade-offs with various key lengths and serialization factors, influencing the design latency and size. The best designs compared well against a few alternative works in the literature.

The exploration of the design space for the FPGA implementation of RSA was also addressed by the journal paper [3], which presented two solutions at the extremes of the design space (word serial vs. fully parallel approach). We systematically analyzed the results for different serialization factors, evaluating and comparing the execution time and the FPGA resource usage, offering a deep understanding of the trade-offs behind the two design alternatives.

The research also covered Elliptic Curve Cryptography (ECC) as the other prominent public-key cryptosystem. The review presented in [4] touched the essential aspects behind ECC implementation, showing that implementation efficiency and security requirements affect not only mere low-level technological aspects but also higher level choices, ranging from finite field arithmetic up to curve mathematics and protocols.

Both RSA and ECC use modular multiplication, performed in one of two different algebraic structures, $GF(N)$ and $GF(2^m)$, which normally require distinct hardware solutions for speeding up performance. The work in [5] pursued the idea of developing a single circuit supporting the two algebraic structures, achieving a *unified* solution for acceleration of public-key cryptography. The proposed system outperformed similar unified solutions available at the time in terms of clock count and critical path.

A similar idea was presented in [6], addressing again Montgomery multiplication for the long operands used in cryptographic applications. Unlike the previous proposal, however, this work relied on a modified Booth recoding scheme for the multiplicand along with a radix-4 scheme for the modulus, enabling reduced time delays, low clock counts, and hence high throughput levels. The experimental evaluation was based on $0.18\ \mu$m CMOS technology and outperformed similar

Manuscript revised December 31, 2011.

previous works.

Operations on long operands, particularly integer addition, were also tackled by novel approaches, including speculative techniques. These enable shorter combinatorial delays and faster circuits, which however might occasionally generate wrong results needing a correction step. The work in [7] introduced an original speculative scheme suitable for two's complement addition of signed integers. Signed addition can be tricky to implement speculatively because, due to sign extension, there might be long carry chains propagating through the operands, unlike the assumption of uniformly distributed bits that most previous schemes made. The work in [7] introduced an architecture based on an innovative technique for speculative global carry evaluation, addressing the case of long carry chains due to sign extension. The solution was evaluated on real-world benchmarks and achieved significant improvements compared to both standard adders and alternative architectures for speculative addition.

The activity on building blocks for cryptographic computing also led to theoretical results. The letter presented in [8] formalized a technique, effectively a change of representation in $GF(2^m)$, transforming a standard $GF(2^m)$ multiplication into a Montgomery multiplication at nearly no implementation costs. The idea can facilitate the implementation of unified $GF(2^m)$/$GF(N)$ Montgomery multiplication used for elliptic curve cryptography.

As a further theoretical result, the author of this report explored an unconventional, non-Boolean computational model for the implementation of Montgomery multiplication used in cryptography [9]. In particular, the work addressed the so-called *threshold logic*, the model natively supported by emerging electronic technologies like Resonant Tunneling Diodes, Single-Electron Tunneling, Quantum Cellular Automata, and Tunneling Phase Logic. The work in [9] presented a study of the computational power of linear threshold functions for modular reduction and multiplication, establishing an optimal bound to the delay of a threshold logic circuit implementing fixed-modulus Montgomery modular reduction and multiplication. The work also proposed a circuit scheme ensuring feasible $O(n^2)$ area requirements with constant latency and limited architectural critical path, independent of the input size n.

Last, in [10] the author presented an efficient bit-parallel $GF(2^m)$ multiplier. The work assumed that $GF(2^m)$ elements are represented using a Shifted Polynomial Basis (SPB), and specifically targeted the case where an irreducible pentanomial is used to build the representation. The theoretical part of the work derived a closed expression of the reduced SPB product, then applying the formulation to the case of pentanomials. The resulting multiplier was at least as efficient as previous proposals in the literature, but it was suitable for a larger class of pentanomials, allowing the choice of pentanomials optimizing further field operations (for example, inversion).

A. Acceleration of hash collision search

Besides hardware circuits for the fast execution of underlying mathematical operations in cryptographic algorithms, the work also explored the use of advanced solutions, based either on software-programmable processors or reconfigurable hardware solutions, for the acceleration of cryptanalysis, aimed at demonstrating possible weaknesses in cryptographic algorithms.

The conference paper [11] describes the results of an activity partially carried out at the Barcelona Supercomputing Center (BSC), which made available the prototypical MariCel HPC cluster based on the IBM CellBE processor. The activity was aimed at developing an advanced HPC application for fast collision search against the SHA-1 cryptographic hash function, which was a hot topic in cryptanalysis. While developing the application, a number of original techniques for speeding up the search process were introduced. They ultimately resulted in an actual collision for a 71-round version of SHA-1, the first ever at the time. This experience was then exploited to study the benefits of FPGA platforms for the cryptanalysis of SHA-1. In [12] we identified a number of techniques inherently based on hardware reconfigurability, ranging from algorithm/architecture exploration to circuit specialization and static/dynamic reconfiguration, which can effectively improve the SHA-1 collision search process. The ultimate benefits were expressed in terms of speed/cost ratio and, based on a small-scale prototypical FPGA cluster, demonstrated by a collision for a 72-round SHA-1.

III. Field Programmable Gate Array design and test methodologies

As implied by the above summary, most of the activities involved the use of hardware-reconfigurable devices. In fact, a part of the research work was also aimed at studying new methodologies for FPGA testing as well as new techniques for FPGA design automation.

A first work along this line was developed in the context of the PRIN2005 project *COMMUTA: Hardware/software mutant components for distributed, dynamically reconfigurable systems* and presented in [13], where the mobile agents software concept was explored as an enabling technique for self-healing hardware reconfigurable systems deployed in heterogeneous networked infrastructures. Subsequent activities specifically addressed testing techniques for FPGA devices and applications, particularly focusing on online testing, which is relevant in many applications including avionics, space, railway transportation, etc. The work in [14] dealt with embedded devices including a hardware reconfigurable fabric along with a software-programmable system, used to run a reconfigurable hardware abstraction layer exposing to the software a high-level view of the physical resources. This allowed FPGA testing to be restricted to those hardware resources that are actually covered by the user design and the test configurations to be handled in a compact way, saving much memory – a limited resource in typical embedded systems– compared to conventional approaches storing static sets of FPGA testing configurations. The work in [15] addressed application-dependent testing (ADT) techniques, an approach inherently enabled by hardware reconfigurability, and specifically focused on a few aspects that were not properly handled by existing proposals, e.g. bridging fault coverage and test configuration generation.

In addition to testing methodologies, the activity also explored techniques supporting FPGA design automation. In particular, the work in [16] introduced an approach and software tool to support design flows based on *high-level synthesis*, which allows programs written in a high-level language (HLL) to be translated to hardware description language (HDL) modules. The work focused on the problem of anticipating the hardware cost of the translated modules, in order to drive hardware/software partitioning choices at an early stage of the design flow, avoiding time-consuming try-and-error loops. Notice that the work did not propose a specific high-level synthesis approach. Rather, it was meant to be used with generic HLL-to-HDL toolchains as a framework for characterizing third-party tools and identifying the most suitable early prediction metrics. The framework, relying on the LLVM compiler infrastructure along with the R statistical package, was evaluated with a large base of benchmarks, confirming the accuracy of the hardware cost prediction.

IV. Security service provisioning in distributed environments

Because of the central focus on security applications, the activity also covered the implementation, deployment, and evaluation of security services in distributed environments, particularly addressing time stamping and public-key certification services provisioned to lightweight/mobile platforms, possibly augmented with dedicated hardware acceleration. Partly based on a collaboration with industry, the work in [17] dealt with the implementation of time stamping services in inter-enterprise settings. In fact, the paper specifically addressed integration issues posed by a potentially large community of enterprises accessing remote third-party time stamping services. The work proposed a practical architecture, oriented to Web Services technologies, offering relative and absolute temporal authentication, the former based on linear linking schemes, the latter relying on publishing mechanisms and a trusted time source. Targeted at a different audience of potential users, the work in [18] presented the design and implementation of a digital time stamping service provisioned to mobile devices with limited resources. The work also included the presentation of a case-study system and an experimental evaluation.

As mentioned above, the activity also considered hardware acceleration and its interplay with the design of security services. The work in [19] presented a multi-tier architecture combining a hardware-accelerated back-end and a Web Services based web tier for public-key certification and digital time stamping services delivered to mobile devices. Similarly, [20] presented a solution combining hardware acceleration with a Web Services tier, also including a case study and a thorough experimental campaign.

As a different development involving hardware support, [21] focused on a specific component, i.e. an FPGA-based key-store allowing secure storage and retrieval of RSA keys, meant to improve the dependability of a wide class of security services. The implementation relied on a programmable hardware board, a Celoxica RC1000 product equipped with a Xilinx Virtex-E 2000 FPGA device. In addition to describing the key-store architecture and implementation, the paper also evaluated the security and performance gain achieved by a real-world application using the proposed solution.

To gain a deeper understanding of architectural features and their impact on the efficient deploying of security-enabled applications on mobile devices, the work in [22] examined the characteristics of mobile device platforms and their implications on performance of security. In line with the previous works, the case study was a Web Services solution providing public-key certification services to mobile devices. Experiments were conducted on three different mobile terminals, spanning a large range of characteristics representative of resource-constrained devices. The analysis highlighted that performance figures can vary significantly even with similar hardware platforms, while security and performance are often conflicting requirements.

A different development, based again on a collaboration with an ICT company and presented in [23], focused on heterogeneous data stream parsing in distributed enterprise environments, e.g. for handling stock-market buy/sell orders, purchase records, web server log entries, etc. Such streams are usually processed locally in ad-hoc, heterogeneous formats, although they often share a common semantics and need a combined analysis to infer global events. The work introduced a solution enabling ad-hoc parsers to be generated automatically based on an appropriate grammar written in a common format. The approach was also demonstrated by a working implementation successfully integrated into a telecommunication billing system inherently requiring real-time processing of information flows.

V. Conclusions and Acknowledgments

This document summarized the main findings stemmed from the author's research activity, including a few funded research projects, until the time of writing. The document described the scientific works derived from the above activities and briefly commented on the essential insights brought by each of them.

The described activities have been partially funded by PRIN2005 project *COMMUTA: Hardware/software mutant components for distributed, dynamically reconfigurable systems*, and POR Campania FESR 2007-2013 project *Progetto Metadistretto del Settore ICT - Sistema di comunicazione per l'integrazione delle informazioni nella distribuzione commerciale e nei punti vendita*.

Following is a complete list of publications derived from the research activities described in this report.

References

[1] A. Cilardo, A. Mazzeo, L. Romano, and G. Saggese, "Carry-save Montgomery modular exponentiation on reconfigurable hardware," in *Proceedings - Design, Automation and Test in Europe Conference and Exhibition*, 2004, pp. 206–211.

[2] A. Cilardo, A. Mazzeo, L. Romano, and G. Saggese, "Architecture and FPGA implementation of a digit-serial RSA processor," in *New Algorithms, Architectures and Applications for Reconfigurable Computing*. Springer US, 2005, pp. 209–218.

[3] A. Cilardo, A. Mazzeo, L. Romano, and G. Saggese, "Exploring the design-space for FPGA-based implementation of RSA," *Microprocessors and Microsystems*, vol. 28, no. 4, pp. 183–191, 2004.

[4] A. Cilardo, L. Coppolino, N. Mazzocca, and L. Romano, "Elliptic curve cryptography engineering," *Proceedings of the IEEE*, vol. 94, no. 2, pp. 395–405, 2006.

[5] A. Cilardo, A. Mazzeo, N. Mazzocca, and L. Romano, "A novel unified architecture for public-key cryptography," in *Proceedings -Design, Automation and Test in Europe, DATE '05*, vol. 2005, 2005, pp. 52–57.

[6] A. Cilardo and N. Mazzocca, "Time efficient dual-field unit for cryptography-related processing," *IFIP Advances in Information and Communication Technology*, vol. 313, pp. 191–210, 2010.

[7] A. Cilardo, "A new speculative addition architecture suitable for two's complement operations," in *Proceedings -Design, Automation and Test in Europe, DATE*, 2009, pp. 664–669.

[8] A. Cilardo, A. Mazzeo, and N. Mazzocca, "Representation of elements in F_{2^m} enabling unified field arithmetic for elliptic curve cryptography," *Electronics Letters*, vol. 41, no. 14, pp. 798–800, 2005.

[9] A. Cilardo, "Exploring the potential of threshold logic for cryptography-related operations," *IEEE Transactions on Computers*, vol. 60, no. 4, pp. 452–462, 2011.

[10] A. Cilardo, "Efficient bit-parallel $GF(2^m)$ multiplier for a large class of irreducible pentanomials," *IEEE Transactions on Computers*, vol. 58, no. 7, pp. 1001–1008, 2009.

[11] A. Cilardo, L. Esposito, A. Veniero, A. Mazzeo, V. Beltran, and E. Ayguad, "A cellBE-based HPC application for the analysis of vulnerabilities in cryptographic hash functions," in *Proceedings - 2010 12th IEEE International Conference on High Performance Computing and Communications, HPCC 2010*, 2010, pp. 450–457.

[12] A. Cilardo, "The potential of reconfigurable hardware for HPC cryptanalysis of SHA-1," in *Proceedings -Design, Automation and Test in Europe, DATE*, 2011, pp. 998–1003.

[13] A. Benso, A. Cilardo, N. Mazzocca, L. Miclea, P. Prinetto, and E. Szilrd, "Reconfigurable systems self-healing using mobile hardware agents," in *Proceedings - International Test Conference*, vol. 2005, 2005, pp. 468–476.

[14] A. Cilardo, N. Mazzocca, and L. Coppolino, "Virtual scan chains far online testing of FPGA-based embedded systems," in *Proceedings - 11th EUROMICRO Conference on Digital System Design Architectures, Methods and Tools, DSD 2008*, vol. 2008-January, 2008, pp. 360–366.

[15] A. Cilardo, C. Lofiego, A. Mazzeo, and N. Mazzocca, "Revisiting application-dependent test for FPGA devices," in *Proceedings - 16th IEEE European Test Symposium, ETS 2011*, 2011, p. 213.

[16] A. Cilardo, P. Durante, C. Lofiego, and A. Mazzeo, "Early prediction of hardware complexity in HLL-to-HDL translation," in *Proceedings - 2010 International Conference on Field Programmable Logic and Applications, FPL 2010*, 2010, pp. 483–488.

[17] A. Cilardo, A. Mazzeo, L. Romano, G. Saggese, and G. Cattaneo, "Using Web Services technology for inter-enterprise integration of digital time stamping," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 2889, pp. 960–974, 2003.

[18] A. Cilardo, D. Cotroneo, C. Di Flora, A. Mazzeo, L. Romano, and S. Russo, "Design and implementation of a high performance architecture for providing digital time stamping services to mobile devices," *Computer Systems Science and Engineering*, vol. 22, no. 3, pp. 103–112, 2007.

[19] A. Cilardo, L. Coppolino, A. Mazzeo, and L. Romano, "High-performance and interoperable security services for mobile environments," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3726 LNCS, pp. 1064–1069, 2005.

[20] A. Cilardo, L. Coppolino, A. Mazzeo, and L. Romano, "Combining programmable hardware and Web Services technologies for delivering high-performance and interoperable security," in *Proceedings - 15th EUROMICRO International Conference on Parallel, Distributed and Network-Based Processing, PDP 2007*, 2007, pp. 381–386.

[21] A. Cilardo, A. Mazzeo, L. Romano, and G. Saggese, "An FPGA-based key-store for improving the dependability of security services," in *Proceedings - International Workshop on Object-Oriented Real-Time Dependable Systems, WORDS*, 2005, pp. 389–396.

[22] A. Cilardo, L. Coppolino, A. Mazzeo, and L. Romano, "Performance evaluation of security services: An experimental approach," in *Proceedings - 15th EUROMICRO International Conference on Parallel, Distributed and Network-Based Processing, PDP 2007*, 2007, pp. 387–394.

[23] F. Campanile, A. Cilardo, L. Coppolino, and L. Romano, "Adaptable parsing of real-time data streams," in *Proceedings - 15th EUROMICRO International Conference on Parallel, Distributed and Network-Based Processing, PDP 2007*, 2007, pp. 412–418.