Secure management of the Intellectual Property in the Cloud

Octavian MACHIDON
Department of Electronics and Computers
"Transilvania" University, Brasov
ROMANIA
octavian.machidon@unitbv.ro

Abstract: The development of distributed computing systems – and especially cloud computing – has raised new vulnerability challenges regarding intellectual property (IP) generated by moving client's data to the cloud infrastructure. Sensitive data is thus located at a remote facility, where clients are not granted full control and administrative access, generating critical concerns about the security of their intellectual property. This paper proposes an approach based on securing the IP in the cloud using reconfigurable hardware architectures – FPGA (Field Programmable Gate Array)-based System-on-Chip (SoC): IP elements being integrated in the programmable logic as IP cores, benefiting from the security and management of on-chip integrated IP modules. Also, by transcending the cloud computing model from macro- to micro-structures, integrated IP cores gain a higher degree of adaptability and can be easily interconnected using Network-on-Chip architectures.

Key-Words: cloud computing, security, System-on-Chip, FPGA, partial reconfiguration, scalability, Network-on-Chip

1. Introduction

Cloud computing, one of the most promising and rapidly evolving technologies today, is a computing architecture that delivers various types of distributed resources – computing power, infrastructure and applications – to the customers, as services. Cloud services offer many advantages to customers, since they are based on a scalable delivery of hardware / software / netware resources over the Internet, thus adapting to the user’s demand and load. This leads to lower startup costs, higher availability and instant access to computing components or integrated platforms.

The rapid growth of cloud computing services leads to the shifting of more and more applications to the cloud, therefore requiring massive volumes of data to be transferred between servers and users, while guaranteeing the security and availability of the data. Being stored remotely from the customer’s location, data is exposed to security vulnerabilities. Consequently to this openness and liberalized access, data security is one of the main issues, regarding the cloud computing field that needs to be addressed today [1].

In cloud computing systems, the client is not granted exclusive administrative access to the system where the data is stored and manipulated – this could lead to a breach in security. Data protection is ensured by encryption in transit between the client and cloud servers, but while in storage and use at the remote location, it can be accessed by a malicious third party. These potential vulnerabilities need to be resolved by strong security guarantees provided to the clients as part of the Service Level Agreement.

As security attacks become more elaborated and specific, the need arises for “smarter” security models that go beyond the traditional software-based approach which offers limited protection to already-known threats. New hardware implemented security concepts are needed that can adapt quickly to changing and evolving security threats while meeting energy efficiency and reliability constraints.

The recent developments in computer architecture have enabled greater transistor densities and design complexity, leading to the rise of large multi-core processors as an alternative to overcoming the downsides of the single-core monolithic processors. Such state-of-
the art processors have multiple cores placed on a single die, integrating complexity and power management technologies [2]. Consequently the new generation of embedded systems (reconfigurable System-on-Chips) make the ideal candidate for hardware-based adaptive security modules that meet the growing security requirements in today’s systems [3]. The latest generation of FPGA-based SoCs can help improve security due to their isolated memory spaces, partial reconfigurability, bitstream protection and other built-in security mechanisms. Also, having a high degree of scalability, they may help overcoming the increased power consumption and complexity of the growing cloud infrastructure [4] by a better management of the Intellectual Property from the Cloud as integrated IP cores.

2. Intellectual Property in the Cloud Computing Environment

Intellectual property (IP), in both its legal dimension (materialized into patents) and technological form (modules/blocks as components of a larger system) represents a sensitive issue in the context of Cloud Computing due to the specific particularities of the two entities that interact when IP elements become integrated into cloud. In a cloud computing system, clients do not have administrative control over the remote system where their data (that may contain critical Intellectual Property elements) is being stored and/or processed. Their private data's location is unknown to the clients, and most of the time, it is a dynamic one, as is the implementation of the cloud services. This generates two key issues: transfer of control and transfer of responsibility from the client to the cloud provider. Thus, clients need solid and clear guarantees provided in the SLA (Service Level Agreement) regarding property rights over IP stored or created in the cloud, and regarding confidentiality and security of cloud services. The situation is worsened by the fact that, for the client, the cloud computing system he is using is perceived as a "black-box", having known inputs and outputs, but not allowing access to the internal components and processes. This makes any evaluations on how IP is protected and investigations in the case of suspicions of patent infringement very difficult to make. Situations may exist where the cloud provider would outsource some services to other companies in order to respond to special requests from some clients or to manage periods with very high loads. This scenario also generates concerns about the security and integrity of IP since third-party entities gain access to them, without the knowledge of the client, who is being "blinded" by the interface with the provider's services [5]. During the use of cloud computing services, by storing and processing the client's data, new IP elements can be created. In this case, a major controversy appears: which entity has property rights on the newly created IP - the client, or the provider - especially given the fact that the IP cannot be separated from the client's data, nor from the software and hardware infrastructure used to create it, which belongs to the provider. Another related issue is the fact that usually the physical hardware resources of a cloud provider are scattered on a wide geographical area, sometimes located in different countries, a situation that does not allow a unified legislative approach on IP elements in the cloud, since each country has its own laws regulating these issues. Also, due to the specific cloud computing model of providing services, information regarding client's data (locations where they are stored, what entities access them, how are they being used) are difficult to obtain. For example: a cloud storage service uses multiple physical resources that are situated at various locations, thus it is difficult to guarantee the complete deletion of private data (due to the built-in characteristics of the traditional storage devices, even if deleted, data can be recovered, so in order to guarantee a secure deletion, the
provider must overwrite that particular data from all its locations).

On a global level there is an increase of patent requests containing the term "cloud computing", which leads to an increase in research and debate regarding the secure management of IP elements in cloud computing systems [6].

Taking all these issues into consideration, this paper proposes that in a cloud computing environment, patents and other IP elements could materialize into integrated IP cores, thus transcending from macro- to integrated micro-structures, logic blocks or data from a reconfigurable circuit.

3. Using Network-on-Chip architectures for managing Intellectual Property as integrated IP cores

System-on-Chip architectures are very complex, integrating various components on a single die (processing cores, cache banks, memory controllers, etc...). With such architecture, the key for having high performance is the efficient communication between the on-die components of the chip, thus raising the issue of inter-connectivity.

The first multi-core systems, having only a small number of cores, solved this problem by using a bus for interconnecting all the components. However, in the existing or forecasted chips with tens or hundreds of cores (like [7]), using a bus to connect so many cores brings down scalability due to high electrical loading on the bus (thus reducing the speed) and high bandwidth demand (that cannot be met on a single shared bus with so many cores) [8].

Network-on-Chip represents a solution for efficient on-chip communication and integration of IP cores that grants scalability [9]. A NoC contains three fundamental components:

- Network adapters – implementing the interface for connecting to the network
- Routing nodes – routing the data according to the specific protocols
- Links – connecting the nodes

The network adapters are used by the IP cores to connect to the NoC, thus ensuring the decoupling of computation from communication. [10].

A topological illustration of a 3-by-3 grid structured NoC is presented in Figure 1, with the fundamental components being indicated, while Figure 2 shows a detailed view of the network adapter and its interfaces with the IP core and the network.

A key issue regarding the NoC from the cloud computing perspective is the re-configurability. This is a key issue for integrating the IP elements from the cloud into SoC architectures while granting the cloud scalability model, since the integrated IP cores must have the ability to adapt and re-configure their functionality during run-time.
The NoC, being a flexible communication structure, can be used to implement an application-specific re-configurable system by programming specific connection into the NoC. Modern SoC designs integrate tens of IP cores and processors, with high requirements of flexibility and the need of re-configuration during runtime. NoC represent a solution proposed for managing SoC reconfiguration complexity, offering scalability and adaptability in connecting the IP Cores. By favoring the design and implementation of reconfigurable circuits, NoC implementations are an important asset for having scalable and flexible hardware resources for cloud computing applications. Due to its internal architecture, the NoC makes integration of reconfigurable modules (RM) easier. The network adapter acts as the interface between the RM and the network, and all the reconfigurable IP cores sharing that specific region will present identical interfaces and placement pins [11].

Performance isolation between scale-out workloads has been credited as able to provide QoS and priority based services, critical for cloud computing providers [4]. Performance isolations requirements state that the resources (both computational – cores, memory, and communication – network) should not be blocked by some workloads, thus ensuring continuous and parallel work sessions. NoC implementations have the ability of playing a crucial role in ensuring a high degree of communication isolation, one the main components needed for performance isolation. Communication isolation implies that the NoC traffic belonging to different workloads should not block each other, in order to maintain the network latency in limits required for normal operation of all parallel applications. For achieving this isolation, an optimal synergy between the network topology and routing mechanism must be reached in the NoC implementation [4].

3. Enhancing IP Security using Reconfigurable SoCs

One of the latest trends in the industry of embedded systems is the development of chips that contain both reconfigurable components like programmable logic (PL) and programmable cores (embedded microprocessors). These new Systems-on-Chip that embed FPGAs and micro-processor cores offer major strongpoints when used in cryptography applications [12], but also due to their unique characteristics like isolated memory spaces, computing parallelism, dynamic reconfiguration and built-in security mechanisms (e.g. bitstream protection), they have a great potential of being considered as “trusted hardware” components. However, a thorough analysis of their strongpoints and weaknesses is needed in order to properly assess whether or not such SoCs can help enhance the security of cloud computing systems.

3.1 Advantages

One of the main built-in security features of a FPGA based SoC is the bitstream, which acts as a type of design clouding. The bitstream configuration file codes the original netlist according to a proprietary generation mechanism. The reverse operation – obtaining a netlist from a bitstream file is highly unlikely and requires huge computational resources – not even FPGA manufacturers are able to accomplish this task [13]. The FPGAs are by design secure platforms because design manufacture is separated from end-application design by an OEM (Original Equipment Manufacturer). This separation of confidential data between various parties contributes to the security level of such a platform. With regard to the bitstream encryption mechanism (provided by some manufacturers), the decryption key can be programmed into the device in a secure and distinct facility, other than belonging to the product manufacturer (e.g. a trusted third party – "key escrow"). The dynamic partial reconfiguration feature of these embedded systems also brings a series of benefits since it allows for a great degree of adaptability during runtime. From the security perspective, a system implementing sensitive
computations that include data protection by using cryptography is vulnerable to attacks that exploit eventual weaknesses in the running algorithms and protocols. Having a reconfigurable system means that in the case of such a security flaw, the system configuration can be remotely updated by uploading a configuration file that fixes the respective issue, during the runtime of the system, without physical access to it. Therefore, such devices become adaptive security modules that can update their functionality at any time (e.g. updating or replacing the cryptographic algorithm with a new and better one), thus achieving a higher level of security and being able to cope to wider range of challenges [12].

Another asset that a FPGA-based dynamic reconfigurable SoC brings on is the possibility to switch between different cryptographic algorithms during runtime by re-configuring only a part of the design with partial bitstreams (that can be stored locally in a bitstream repository accessed remotely).

3.2 Potential issues

Despite the promising features mentioned above, there are also many weaknesses regarding these systems that need to be identified and addressed.

For an embedded system, there are several types of security threats [13]:

- **Reverse Engineering**: an attacker can gain knowledge about the layout, devices used, interaction between the design components and logic running on the FPGA. Using this information, the malicious party can then attempt to reconstruct the design, with the goal of either producing his own similar products or obtaining more inside information and data.

- **Cloning**: This is another threat to the design security in which the attacker doesn’t need to understand the insides of the design, but instead builds copies of it for selling them and thus obtaining a higher profit since the product development and marketing stages have been skipped.

- **Tampering**: When a malicious party gains physical access to the system, he can try to extract or modify the operating data or firmware of the system with the criminal goal of compromising it or shutting it down. This is a special concern in the case of cloud computing systems, since the hardware platform that stores and manipulates the client’s data is located at a remote location where such a physical intrusion can easily happen unnoticed by the client.

- **Side-Channel attacks**: In such an attack, the malicious party attempts to insert faults and gain insight into the design by using its timing or power characteristics.

- **Fault insertion**: This is another type of attack that implies physical access to the design. The attacker varies the system’s operating parameters (voltage, temperature, clocks, and inputs) beyond their normal ranges attempting to make it malfunction and enter a debug or invalid mode or outputting specific data after introducing input glitches.

- **Readback**: By readback the bitstream data from a FPGA can be read, usually for debug purposes. The readback bitstream lacks configuration information but contains data about the internal memory components and registers states. From a security perspective it raises concerns because, if such an operation is performed by a malicious party, operational data can be recovered.

In order to develop a FPGA-based security module, solutions to the above problems need to be implemented. Many of the security features of a FPGA chip are built-in and developed by the manufacturer, others must be implemented during the design and programming of the FPGA [12].

Since every FPGA manufacturer offers different security features and functionality for their chips, a balanced approach is needed for choosing the best hardware solution for this task, taking
advantage of the FPGA’s dynamic reconfigurability and built-in security mechanisms in order to obtain the desired security level.

4. Xilinx Zynq-7000 All Programmable SoC-based solution

Apart from a high security level, cloud computing applications require features like: high-speed data processing, real-time responses to load and demand changes (high scalability and adaptability) and versatile functionality. In order for an embedded system to meet such high requirements it needs to offer high performance, low latency and low power consumption, together with a high level of scalability and security, thus merging both cost-effectiveness and high functionality in one solution.

In this paper we focus on providing such an embedded system solution for cloud computing applications based on the Xilinx Zynq™-7000 All Programmable SoC as a system feasible for deployment in cloud computing applications. This is an ideal platform for today's cloud computing applications, offering both software development and programmable hardware resources for real time data processing. This embedded platform offers the unique flexibility of changing or updating the design while in the field, due to the partial reconfigurability feature and high security level.

We take advantage of both built-in security and partial reconfiguration features in a synergy that offers an adaptive and scalable platform with improved security, thus matching the performance requirements needed for developing high-quality cloud computing applications while offering improved security solutions to the specific threats of this new computing technology.

4.1 Zynq built-in security

The anti-tamper (AT) features of the Zynq architecture offer designers a solid protection of the intellectual property (IP) and sensitive data existing in their system. The tamper-resistant protection applies on three levels: prevention, detection and response. These security mechanisms prevent or render useless cloning, reverse engineering and other security attacks, as a result protecting commercial designs and critical data [14]. The designer can choose which built-in AT features to include in his system, and whether or not to combine them with his own security measures implemented in the design.

The security measures implemented in the Zynq chip protect the design from the attacks identified above [14]:

- Bitstream encryption and authentication – with the decryption key being kept in the battery-backed RAM (BBRAM) for protection – thus securing the system against cloning or reverse engineering attempts

- Disabling of external readback circuitry after device configuration for protecting the user data

- Temperature and voltage monitoring for detecting physical tampering (as part of potential side channel or fault insertion attacks)

- Active response measures in case of a tamper event, including: clearing of the AES (Advanced Encryption Standard) key used for bitstream decryption, clearing the FPGA configuration memory and flip-flop contents, driving all FPGA outputs to high-Z state (to prevent further output data flow) and resetting the whole design.

Of particular interest to cloud computing applications are the implemented active counter-measures against physical tampering, as they help secure the SoC embedded platform located remotely against such attacks which represent an important vulnerability in cloud systems. Another important security feature unique to the Zynq architecture is the boot sequence: processor first, FPGA second, with support for user authentication – RSA (Rivest-Shamir-Adleman), encryption (AES-256) and data authentication – HMAC (Hash-based Message Authentication Code). Compared to other
solutions of its kind, Zynq is the only architecture that applies AES decryption to the processor boot code and boots the processor first [15]. This approach offers protection from Trojan attacks since by securely booting the PL such malicious code cannot be inserted into the device. Of particular interest to cloud computing applications are the implemented countermeasures against physical tampering – Zynq devices monitor their own environment (temperature level and power supply voltage) using on-chip thermal sensors and A/D converters for creating fail-safe mechanisms that respond to environmental challenges and attacks by “zeroizing” the device when detecting an intrusion (deleting sensitive data and cryptographic keys) [14].

4.2 Partial reconfigurable architecture

In order to enhance the security of cloud computing servers, a separation between sensitive (including Intellectual Property elements) and non-sensitive data should be accomplished. SoC platforms with improved security could consequently become trusted devices inside cloud computing servers dedicated to computations and other operations involving such sensitive data (Figure 3). From a design and resource effectiveness point of view this approach improves security without diminishing the overall performance and scalability since it allows the general-purpose servers and other cloud machines to handle the majority of data and related operations and only those needing added security and limited outside interaction – like information containing patents or other IP – would be transferred to such trusted embedded platforms [16].

With regard to our Zynq-based proposed implementation of a secure computing platform, a preventive approach should take advantage of the partial reconfiguration (PR) feature of the device. The partial reconfiguration feature comes as an enhancement for FPGAs, making them more adaptable and flexible in order to meet the growing performance, cost and power constraints for today’s systems. Partial reconfiguration allows specific regions of the FPGA to be re-programmed with new functionality during runtime, while the rest of the system continues to run uninterrupted. This offers a series of advantages:

- optimizing hardware resource usage: for a design that has components running sequentially, these can be dynamically time-multiplexed so that they are implemented one at a time - using logic cells only when they are needed;
- increasing the scalability;
- reducing system down-time, since the updates and new functions can be implemented dynamically during runtime.

One way to enhance the security level is by creating one or more partial reconfigurable regions in the design for loading the sensitive technology blocks – IP cores – only when needed and discarding them (by loading black box versions of the PR regions) after they are no longer needed or in case of a security attack to the system (Figure 4). The partial bitstreams involving sensitive IP modules can be encrypted using the built-in cryptography of the Xilinx technology but also by using a user-preferred encryption (and in this case the configuration files can be decrypted inside the FPGA logic). The cryptographic keys involved can be stored in the block-RAM or FPGA registers, and can be also erased in case of a tamper event. Also, in the case of an embedded design running sensitive code on the CPU, this can also be deleted in critical situations, leading to a complete zeroisation of the design (by deleting keys, sensitive data and code), only non-sensitive data remains in the external memory [14].

Such an integration of a FPGA-based SoCs in a Cloud computing system offers the system a secure reconfigurable hardware module for performing the sensitive client applications involving Intellectual Property elements that require an increased level of security.
Given its FPGA architecture, such a module can be used for dynamically switching between different encrypted bitstreams, thus accommodating a variety of designs and applications while solving problems like scaling and managing an infrastructure of several dedicated-hardware modules.

Also, as recent research on this subject [16] has identified, this type of integration may be considered as an emulation of homomorphic encryption, which is a technique under development that allows computations to be carried out on the encrypted data, with the decrypted result matching the same result of operations performed on the original data. In this case, the SoC is viewed as a secure and computationally closed environment that performs computations on encrypted I/O data, the decryption and re-encryption being done inside the secure device.

Such a perspective is viable only when the FPGA-based device is protected from security attacks, like the ones described above, with effective countermeasures. In this light, the implementation of a PR region inside the SoC programmable logic for deploying the sensitive and critical portions of the design adds an extra level of security and also allows for response actions in case of detecting a possible attack.

An important reason for choosing Zynq is the novel PCAP (Processor Configuration Access Port) that allows for both full and partial runtime reconfiguration of the PL, which is an important improvement over the existing ICAP (Internal Configuration Access Port – a component that involves a more complex reconfiguration procedure) [17].

In order to configure the PR region of the Zynq’s PL, the following steps are executed in an embedded C application running on the PS:

1. The Xilinx Device Configuration Interface (XDCfg) is initialized by enabling the PCAP interface and setting the control register for the corresponding PCAP mode (full PL configuration).
2. DMA (Direct Memory Access) and PCAP Done interrupts are cleared.
3. The bitstream is transferred from DDR memory to PL fabric.
4. PCAP and AXI (Advanced eXtensible Interface) Done interrupts are polled, so the function call returns only after both transfers are complete.

4.3 Military applications

Cloud computing has emerged as a field of great interest not only for the IT industry, but also for the military. The United States of America Department of Defense (DoD) is currently running a Cloud Computing implementation project with the stated goal of "implementing cloud computing as the means to deliver the most innovative, efficient, and secure information and IT services in support of the Department's mission, anywhere, anytime, on any authorized device."

The DoD Cloud Computing Strategy paper [18] credits cloud computing as a technology helping to achieve key benefits like increased mission effectiveness and operational efficiency. In the effort of implementing cloud computing services, the DoD underlined specific challenges in areas of cybersecurity, information assurance and resilience.

Consequently, our approach has potential applications in the military area of cloud computing also, especially since there is a defense-grade Zynq AP SoC (the 7000Q series). The Zynq-7000Q is the only defense-grade AP SoC in the industry, making it an ideal building block for secure military cloud services.

Besides the described features of the "commercial" version, the Zynq-7000Q offers key defense-grade features like: extended temperature range (-40 to +125°C), leaded (Pb) content, ruggedized packaging and anti-counterfeiting features.

Integrating the defense-grade Zynq-7000Q with its built-in security features and the partially reconfigurable architecture described above can help provide secure and high-quality cloud services for defense organizations.

5. Conclusion

Cloud computing security is an important subject under heavy debate and research today, since it raises specific issues related to the fact that clients do not have exclusive administrative or physical access to the machines storing and handling their data.

This paper has presented a novel approach that integrates the Intellectual Property elements as IP cores of reconfigurable SoC architectures. This brings on two important advantages: an optimized management of the IP complexity – by using Network-on-Chip solutions – and improved security that benefits from the built-in features and partial reconfiguration design flow of the SoC devices.

Several directions have been identified where NoC implementations have the potential of offering high scalability and flexibility to accommodate IP cores on a SoC device for cloud computing applications. The scalability potential of an integrated network on the chip meets the scalability requirements of the hardware resources needed in cloud computing systems. Thus, NoC provides efficient communication between on-die components of the chip (IP cores and other functional modules), lowering design complexity and improving power management.

The cloud computing perspective of NoC also underlines the huge potential of having reconfigurable hardware resources with NoC implementations. Being a flexible communication structure, such a network can easily scale and re-configure during runtime in order to accommodate different configurations of the same system. Such architectures implemented with reconfigurable network topology and embedded protocols, combine packet-switching with circuit-switching resources in order to offer a great degree of flexibility and energy efficiency.

All the strongpoints identified above make Network-on-Chip an excellent instrument in optimizing cloud computing hardware, by easing the increased stress on network components, and improving scalability and resource management, thus contributing to the integration of Intellectual Property blocks.

Besides the improved management of IP cores using NoC solutions, the paper has shown how security can be enhanced by integrating a SoC with dynamic partial reconfiguration feature in the cloud system. This device, given its quality
performance under high security conditions, may be used for running the computations on the sensitive/critical data, while the rest of the data, being non-critical, can be run on ordinary cloud machines. In order for the SoC to be regarded as a computationally closed and secure environment inside the cloud system, a synergy between the device’s built-in security and design concepts that enhance the overall protection against attacks is needed. We have proposed a design based on the Xilinx Zynq-7000 AP SoC, given its built-in passive and active security features and the potential of partial-reconfigurability, which in our perspective is a key asset that can add an extra level of security to the system. By combining the device’s security in a design with PR regions for critical modules – interconnected using an on-chip network, active countermeasures can be implemented in the case of an attack, including a total zeroisation of the device – thus deleting all critical information and cryptographic keys. Such architecture has great potential if integrated in cloud computing systems, improving the security of both commercial and defense-grade cloud services and IP cores, making them more flexible and scalable while granting high performance and quality-of-service.

Acknowledgement
This paper is supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), ID134378 financed from the European Social Fund and by the Romanian Government. Parts of this research have been published in the Proceedings of the 7th International Conference on Security for Information Technology and Communications, SECITC 2014

References

