

Fișa suspiciunii de plagiat / Sheet of plagiarism's suspicion	Indexat la: 32/06
--	------------------------------

Opera suspicionată (OS)	Opera autentică (OA)
Suspicious work	Authentic work

OS	Mang I., "Considerations on hardware implementations of encryption algorithms", <i>Visnik Sumskovo derjavnovo universitety, Seria Technichni nauki</i> 12(71), p.41-50, 2004 Disponibil la: http://essuir.sumdu.edu.ua/handle/123456789/10536 .
OA	Elbirt, A.J., Yip,W., Chetwynd, B., Paar, C., „An FPGA Implementation and Performance Evaluation of the AES Block Cipher Candidate Algoritm Finalists”, The Third Advance Encryption Standard (AES3) Candidate Conference, New York, April 13-14, 2000. Disponibil la: http://www.ei.rub.de/media/crypto/veroeffentlichungen/2011/01/21/aelbirtetales3.pdf

Incidența minimă a suspiciunii / Minimum incidence of suspicion	
p.41:6 – p.41:17	p.1:24 – p.2:7
p.41:31 – p.42:36	p.3:5 – p3:28
p.43:Table 1	p.5:Table 1
Fișa întocmită pentru includerea suspiciunii în Indexul Operelor Plagiate în România de la www.plagiate.ro	

An FPGA Implementation Performance Evaluation of the AES Block Cipher Candidate Algorithms*

AJ Elbirt¹, Waiyi¹, B Chetwyn², C Paar¹
Electrical and Computer Engineering Department
Worcester Polytechnic Institute
100 Institute Road, Worcester, MA 01609, USA

¹ Email: aelbirt, waiyi, christf@ece.wpi.edu

² Email: sunge@alum.wpi.edu

The Third Advanced Encryption Standard (AES3) Candidate Conference,
April 13-14, 2000, New York, USA.

Abstract

The technical analysis used in determining which of the Advanced Encryption Standard candidates will be selected as the Advanced Encryption Algorithm includes efficiency testing of both hardware and software implementations of candidate algorithms. Reconfigurable devices such as Field Programmable Gate Arrays (FPGAs) are highly attractive candidates for hardware implementations of encryption algorithms as they provide cryptographic algorithm agility, physical security, and potentially much higher performance than software solutions. This contribution investigates the significance of FPGA implementations for the Advanced Encryption Standard candidate algorithm finalists. Multiple architectural implementations of each algorithm are explored for each algorithm. A strong focus is placed on high throughput implementations, which are required to support security for current and future high bandwidth applications. The implementations of each algorithm will be compared in an effort to determine the most suitable candidate for hardware implementation within commercially available FPGAs.

Keywords: cryptographic, algorithm-agility, FPGA, block cipher, VHDL

1 Introduction

The National Institute of Standards and Technology (NIST) has initiated a research level Federal Information Processing Standard (FIPS) for the Advanced Encryption Standard (AES), specifying an Advanced Encryption Algorithm to replace the Data Encryption Standard (DES) which expires in 1988 [1]. NIST has solicited candidate algorithms for inclusion in AES, resulting in fifteen official candidate algorithms of which five have been selected as finalists. Unlike DES, which was designed specifically for hardware implementations, one of the design criteria for AES candidate algorithms is that they can be efficiently implemented in both hardware and software. Thus, NIST has announced that both hardware and software performance measurements will be included in their efficiency testing. However, virtually all performance comparisons have been restricted to software implementations in various latfirms [2].

*This research was supported in part through NSF CAREER award #CCR-9733246.

The advantages of a software implementation include ease of use, ease of upgrade, portability, and flexibility. However, a software implementation offers only limited physical security, especially with respect to key storage [3] [4]. Conversely, cryptographic algorithms (and their associated keys) that are implemented in hardware are, by nature, more physically secure as they cannot easily be reprogrammed by an outside attacker [4]. The weaknesses of traditional (ASIC) hardware implementations are the lack of flexibility with respect to algorithm and parameter switch. A promising alternative for implementation of lock cipher are reconfigurable hardware devices such as Field Programmable Gate Arrays (FPGAs). FPGAs are hardware devices whose function is not fixed and which can be reprogrammed in-system. The potential advantages of encryption algorithms implemented in FPGAs include:

Algorithm Agility This term refers to the switching of cryptographic algorithms during operation. The majority of modern security protocols, such as SSL, IPsec, all use a multi-encryption algorithm. The encryption algorithm is negotiated in a pre-session phase; e.g., IPsec allows negotiating DES, 3DES, Blowfish, CAST, IDEA, RC4 and RC6 as algorithms, and future extensions are possible. Whereas algorithm agility is costly with traditional hardware, FPGAs can be reprogrammed on-the-fly.

Algorithm Update It is conceivable that field devices are upgraded with a new encryption algorithm which did not exist (or was not standardized!) at design time. In particular, it is very attractive for numerous security products to be upgraded for use of AES once the selection process is over. Assuming there is some kind of (temporary) connection to a network such as the Internet, FPGA-equipped encryption devices can update the new configuration on-line.

Algorithm Modification There are applications which require modification of a standardized algorithm, e.g., by using proprietary - versus - standard implementations. Such modifications are easily made with reconfigurable hardware. Similarly, a standardized algorithm can be swapped with a proprietary one. Also, modifications can be easily changed.

Architecture Efficiency In certain cases, a hardware architecture can be much more efficient if it is designed for a specific set of parameters; e.g., constant multiplication (of integers) in Galois fields is far more efficient than general multiplication. With FPGAs it is possible to design and optimize an architecture for a specific parameter set.

Throughput Although typically slower than an ASIC implementation, FPGA implementations have the potential of running substantially faster than software implementations.

Cost Efficiency The time and costs for developing an FPGA implementation for a given algorithm are much lower than for an ASIC implementation. (However, for high-volume applications, ASIC solutions usually become the more cost-efficient choice.)

Note that algorithm agility remains an open research issue in regards to physical security, and the cost associated with current high-end FPGA devices. However, we believe that cost is not a long-term limiting factor, as will be discussed in section 3.3. For these reasons, this paper describes a thorough comparison of the AES finalist algorithms RC6, Rijndael, Serpent, and Twofish with respect to implementation in state-of-the-art FPGAs. One aspect that seems to be especially relevant is the investigation of achievable encryption rates for FPGA-based implementations. We demonstrate that FPGA solutions encrypt at rates in the Giga-bit range for all four algorithms investigated, which is at least one order of magnitude faster than most real-time software implementations [1].

What follows is an investigation of the AES finalists to determine the nature of their underlying components. The characterization of the algorithms' components will lead to a discussion of the hardware architectures best suited for implementation of the AES finalists. A performance metric to measure the hardware cost for the throughput achieved by each algorithm's implementations will be evaluated and an

target FPGA will be chosen as the implementation that are optimized for high-throughput operation within the commercially available device. Finally, multiple architecture designs of the algorithms within the target FPGA will be discussed and the overall performance of the implementations will be evaluated versus typical software implementations.

2 Previous Work

As the target custom hardware is software implementations, little work exists in the area of clocking implementations within existing FPGAs. DE, the most common clocking implementation targeted to FPGAs, has been shown to operate at speeds of up to 400 M it/s [6]. We believe that this performance can be greatly enhanced using today's technology. These speeds are significantly faster than the best software implementations of DE [7] [8] [9], which typically have throughput of less than 100 M it/s, although a 137 M it/s implementation has been reported as well [7]. This performance differential is an expected result of DE having been designed in the 1970s with hardware implementations in mind.

Other clocking schemes have been implemented in FPGAs with varying degrees of success. A typical example is the IDEA clocking scheme which has been implemented at speeds ranging from 2.8 M it/s [10] to 28 M it/s [11]. Note that while the 28 M it/s throughput was achieved in a fully pipeline architecture, the implementation requires four Xilinx XC4000 FPGAs.

Some FPGA implementations of the AE canisters have been shown to be far slower than their software counterparts. Hardware throughput of about 12 M it/s [12] [13] have been achieved for CA T-2.6. However, software implementations have resulted in throughput of 37.8 M it/s for CA T-2.6 on a 200 MHz PentiumPro PC [14], a factor of three faster than FPGA implementations. When scaled to a more current 600 MHz PentiumPro PC, it is expected that the same software implementation would outperform FPGA implementations by an even larger factor. While an FPGA implementation of RC6 achieves data rates of 37.8 M it/s [13], our findings indicate that considerably higher data rates are achievable.

When examining the AE finalists, it is important to note that they do not necessarily exhibit similar behavior to DE when comparing hardware and software implementations. One reason for this is that the AE finalists have been designed with efficient software implementations in mind. Additionally, software implementations may be executed on processors operating at frequencies as high as 800 MHz while typical implementations of target FPGAs reach a maximum clock frequency of 100 MHz.

3 Methodology

3.1 Design Methodology

There are two basic hardware design methodologies currently available: language based (high level) design and schematic based (low level) design. Language based design relies upon synthesis tools to implement the desired hardware. While synthesis tools continue to improve, they rarely achieve the most optimized implementation in terms of both area and speed when compared to a schematic implementation. As a result, synthesizer designs tend to be (slightly) larger and slower than their schematic based counterparts. Additionally, implementation results can greatly vary depending on the synthesis tool as well as the design being synthesized, leading to potentially increased variances in the synthesizer results when comparing synthesis tool outputs. This situation is not entirely different from a software implementation of an algorithm in a high-level language such as C, which is also dependent on coding style and compiler quality. As shown in [14], schematic based design methodologies are not longer feasible for supporting the increase in architectural complexity envisioned by modern FPGAs. As a result, a language based design methodology was chosen as the implementation for the AE finalists with VHDL being the specific language chosen.

3.2 Implementations — General Considerations

Each AE finalist was implemented in VHDL using a top-down design and test methodology. The same hardware interface was used for each of the implementations. In an effort to achieve the maximum efficiency possible, note that key scheduling and encryption were not implemented for each of the AE finalists. Because FPGAs may be reconfigured in-system, the FPGA may be reconfigured for key scheduling and then later reconfigured for either encryption or decryption. This is a major advantage of FPGAs implementations over classical ASIC implementations. Run keys for encryption are loaded from the external key bus and are stored in internal registers and all keys must be loaded before encryption may begin. Key loading is possible until encryption is complete. Each implementation was simulated for functional correctness using the test vectors provided in the AE submission package [1] [16] [17] [18]. After verifying the functionality of the implementations, the VHDL code was synthesized, placed and routed, and re-simulated with annotated timing using the same test vectors, verifying that the implementations were successful.

3.3 Selecting the Target FPGA

When examining the AE finalists for hardware implementation within an FPGA, a number of key aspects emerge. First, it is obvious that the implementation will require a large amount of I/O instantaneously support the 128-bit data stream at high speeds where bus multiplexing is not an option. It is desirable to emulate the 128-bit input and output data streams to all with a fully pipeline architecture. Since the run keys cannot change during the encryption process, they may be loaded via a separate key input bus prior to the start of encryption. Additionally, to implement a fully pipeline architecture requires 128-bit wide pipeline stages, resulting in the need for a register-rich architecture to achieve a fast, synchronous implementation. Moreover, it is desirable to have as many register bits as possible for each of the FPGA's configurable units to all with a regular layout of design elements as well as to minimize the routing requirements between configurable units. Finally, it is critical that fast carry-chaining be provided between the FPGA's configurable units to maximize the performance of AE finalists that utilize arithmetic operations [13] [12].

In addition to architectural requirements, scalability and cost must be considered. We believe that the chosen FPGA should be the best choice available, capable of providing the largest amount of hardware resources as well as being highly flexible as to yield optimal performance. Unfortunately, the cost associated with current high-end FPGAs is relatively high (several hundred US dollars per device). However, it is important to note that the FPGA market has historically evolved at an extremely rapid pace, with larger and faster devices being released into the market at a constant rate. This evolution has resulted in FPGA cost-curves that decrease sharply over relatively short periods of time. Hence, selecting a high-end device provides the closest model for the typical FPGA that will be available over the expected life span of AE.

Based on the aforementioned considerations, the Xilinx Virtex XCV1000BG 60-4 FPGA was chosen as the target device. The XCV1000 has 128K bits of embedded RAM divided among thirty-two RAM blocks that are separate from the main body of the FPGA. The 60-pin ball grid array package provides 12 usable I/O pins. The XCV1000 is comprised of a 64×6 array of 1 k-bit 2×2 slice architecture. Each slice (CLB), each of which acts as a 4-bit element comprised of two 2-bit slices for a total of 12288 CLB slices [1]. This type of configuration results in a highly flexible architecture that will accommodate the run functions' use of wide combinational functions. Note that the XCV1000 also bears a strong resemblance to the current FPGA and that devices from other vendors are not fundamentally different. It is thus hoped that our results carry over, within limits, to other devices.

3.4 Design Tools

FPGA Express by Synopsys, Inc. and Synplify by Synlicity, Inc. were used to synthesize the VHDL implementations of the AE finalists. As this study places a strong focus on high throughput implementations,

the synthesis tools were set to minimize resources. As will be discussed in section 6, the resultant implementations exhibit the expected throughput with the associated cost being an increase in the area required in the FPGA for each of the implementations. Similarly, if the synthesis tools were set to minimize area, the resultant implementations would exhibit reduced area requirements at the cost of decreased throughput.

XACTstep 2.1 by Xilinx, Inc. was used to place and route the synthesized implementations. For the sub-cell architectures, a 40 MHz timing constraint was used in the synthesis and place-and-route processes as it resulted in significantly higher system clock frequencies. However, the 40 MHz timing constraint was found to have little effect on the other architecture types, resulting in nearly identical system clock frequencies to those achieved without the timing constraint.

Finally, the wave by Viewlogic Systems, Inc. and Active-HDLTM by ALDEC, Inc. were used to perform behavioral and timing simulations for the implementations of the AE finalists. The simulations verified the functionality and the ability to operate at the designated clock frequencies for the implementations.

4 Architecture types in the AES Finalists

Before attempting to implement the AE finalists in hardware, it is important to understand the nature of each algorithm as well as the hardware architectures most suited for their implementation. What follows is an investigation into the key components of the AE finalists. Based on this background, a discussion is presented on the hardware architectures most suited for implementation of the AE finalists.

4.1 Characteristics of the AES Finalist Algorithms

Algorithm	XOR	Mod 2^{32} Add	Mod 2^{32} Subtract	Fixed Shift	Variable Rotate	Mod 2^{32} Multiply	GF(2 ⁸) Multiply	LUT
MAR	•	•	•	•	•	•		•
RC6	•	•		•	•	•		
Rijndael	•			•			•	•
Serpent	•			•				•
Twofish	•	•		•			•	•

Table 1: AE finalists characteristics [20]

Modern FPGAs have a structure composed of a two-dimensional array of configurable function units interconnected via horizontal and vertical routing channels. Configurable function units are typically composed of look-up tables and flip-flops. Look-up tables may be configured as either combinational logic or memory elements. Additionally, many modern FPGAs provide variable-size RAM blocks that may be used as either memory elements or look-up tables [21].

In terms of complexity, the characteristics detailed in Table 1 that require the most hardware resources as well as computation time are the modular 2^{32} multiplication and the variable rotation operations [20]. Implementing wide multipliers in hardware is an inherently difficult task that requires significant hardware resources. Additionally, algorithms that employ large variable rotations require a massive amount of multi-lexing hardware if carefully designed (see section 4.1 for further discussion). Besides being implemented in either combinational logic or memory elements—each having advantages of these techniques are discussed in section 4.2. Fast rotations such as bit-wise XOR, modular 2^{32} addition and subtraction, and fixed value shifting are constructed from simple hardware elements. Additionally, the Galois field multiplications required in Rijndael and Twofish can also be implemented very efficiently in hardware as they are multiplications by a constant. Galois field constant multiplications require far less resources than general multiplications [22].

Based on our evaluation of the AE finalists, the MAR algorithm appeared to be the most resource intensive based on its use of large S-boxes, an modular 2^{32} multiplication. As a result, it was conjectured

that the MAR algorithm would exhibit lesser performance when compared to the other AE finalists. Due to this evaluation and a lack of development resources, the MAR algorithm was mitte from this study.

4.2 Hardware Architectures

The AE finalists are all comprised of a basic looping structure (some form of either Feistel or substitution-permutation network) where data is iteratively processed through a round function. Based on this looping structure, the following architectures were investigated as a way to optimize implementations:

- Iterative Looping
- Loop Unrolling
- Partial Pipelining
- Partial Pipelining with Full-Pipelining

Iterative looping over a cipher's round structure is an effective method for minimizing the hardware required when implementing an iterative architecture. When only one round is implemented, an n -round cipher must iterate n times to perform an encryption. This approach has a low register-to-register delay but requires a large number of clock cycles to perform an encryption. This approach also minimizes in general the hardware required for round function implementation but can be costly with respect to the hardware required for round key and n -bit multi-lexing. Iterative looping is a subset of loop unrolling in that only one round is unrolled whereas a loop unrolling architecture allows for the unrolling of multiple rounds, up to the total number of rounds required by the cipher. As opposed to an iterative looping architecture, a loop unrolling architecture where all n rounds are unrolled and implemented as a single combinational logic block maximizes the hardware required for round function implementation while the hardware required for round key and n -bit multi-lexing is completely eliminated. However, while this approach minimizes the number of clock cycles required to perform an encryption, it maximizes the worst case register-to-register delay for the system, resulting in an extremely slow system clock.

A partially pipelined architecture offers the advantage of high throughput rates by increasing the number of blocks of data that are being simultaneously processed. This is achieved by replicating the round function hardware and registering the intermediate data between rounds. Moreover, in the case of a full-length pipeline (a specific form of a partial pipeline), the system will output a 128-bit block of ciphertext at each clock cycle once the latency of the pipeline has been met. However, an architecture of this form requires significantly more hardware resources as compared to a loop unrolling architecture. In a partially pipelined architecture, each round is implemented as the pipeline's atomic unit and are separated by the registers that form the actual pipeline. However, many of the AE finalists cannot be implemented using a full-length pipeline due to the large size of their associated round function and n -bit multi-lexing, both of which must be replicated n times for an n -round cipher. As such, these algorithms must be implemented as partial pipelines. Additionally, a pipeline architecture can be fully exploited only in cases of operations which do not require feedback of the encryption data, such as Electronic Codebook or Counter Mode [3, Section 4.1]. When operating in feedback modes such as Ciphertext Feedback Mode, the ciphertext of one block must be available before the next block can be encrypted. As a result, multiple blocks of plaintext cannot be encrypted in a pipeline fashion when operating in feedback modes. For the remainder of our discussion, feedback will be required as FB and non-feedback will be referred to as NFB.

Sub-pipelining a (partially) pipelined architecture is a advantage when the round function of the pipeline architecture is complex, resulting in a large delay between pipeline stages. By adding sub-pipeline stages, the atomic function of each pipeline stage is subdivided into smaller functional blocks. This results in a decrease in the pipeline's delay between stages. However, each subdivision of the atomic function increases the number of clock cycles required to perform an encryption by a factor equal to the number of

subdivisions. At the same time, the number of clocks for data that may be generated within NFB mode is increased by a factor equal to the number of subdivisions. Therefore, for this technique to be effective, the worst case delay between stages will decrease by a factor of m where m is the number of subdivisions. However, if the atomic function of the partially pipeline architecture has a small stage delay, subdividing the stage will achieve no significant decrease in the worst case stage delay. In this case, subdividing will result in no significant increase in the system's clock frequency but will increase the logic resources and clock cycles required to perform an encryption, resulting in reduced throughput.

Many FPGAs provide embedded RAM which may be used to replace the runkey and $-B \times$ multilexing hardware. By storing the keys within the RAM blocks, the appropriate key may be addressed using the current runkey. However, due to the limited number of RAM blocks, as well as their restricted width, this method is not feasible for architectures with many pipeline stages or unrlels. These architectures require more RAM blocks than are typically available. Additionally, the switching time for the RAM is more than a factor of three longer than that of a standard CLB slice element, resulting in the RAM element having a lesser speed-up effect on the overall implementation. Therefore, the use of embedded RAM is not consistent for this study to maintain consistency between architectural implementations.

5 Architectural Implementation Analysis

For each of the AE finalists, the four architecture descriptions in section 4.2 were implemented in VHDL using a top-down design and methodology. The same hardware interface was used for each of the implementations. Runkeys are stored in internal registers and all keys must be loaded before encryption may begin. Key loading is isolated until encryption is complete. These implementations yield a great deal of knowledge in regard to the FPGA suitability of each AE finalist. What follows is a discussion of the knowledge gained regarding each algorithm when implemented using the four architecture types.

5.1 Architectural Implementation Analysis — RC6

When implementing the RC6 algorithm, it was first determined that the RC6 modular 2^{32} multiplication was the dominant element of the run function in terms of required logic resources. Each RC6 run requires two cycles of the modular 2^{32} multiplier. However, it was found that the RC6 run function does not require a general modular 2^{32} multiplier. The RC6 multipliers implement the function $A(2A + 1)$ which may be implemented as $2A^2 + A$. Therefore, the multiplication operation was replaced with an array squarer with summation products, requiring fewer hardware resources and resulting in a faster implementation. The remaining components of the RC6 run function — fixed and variable shifting, bit-wise XOR, and modular 2^{32} addition — were found to be simple in structure, resulting in these elements of the run function requiring few hardware resources. While variable shifting operations have the potential to require considerable hardware resources, the bit-wise variable shifting required by the RC6 run function requires few hardware resources. Instead of implementing a 32-bit multilexer for each of the thirty-two rotations of the shifts (controlled by the five shifting bits), a five-level multilexing approach was used. The variable rotation is broken into five stages, each of which is controlled by one of the five shifting bits. For each rotation of a given stage, a 2-bit multilexer controlled by the stage's shifting bit is used. This implementation requires a total of 160 2-bit multipliers as opposed to the thirty-two 32-bit multipliers required for a one-stage implementation. However, using 2-bit multipliers for the five-stage barrel-shifter results in an overall implementation that is smaller and faster when compared to the one-stage barrel-shifter implementation as described in [18, section 3.4]. Finally, it was found that the synthesis tool is able to minimize the overall size of a RC6 run sufficiently to allow for a fully unrlelled fully pipeline implementation of the entire twenty runs of the algorithm within the target FPGA.

As discussed in section 4.2, implementing a single run of the RC6 algorithm requires the greatest area-time utilization. Further unrlelling requires only minor throughput increases as the decrease in

the number of cycles per encryption block was offset by the rapidly decreasing system clock frequency. 2-stage partial i elining was found to yield the highest throughput when operating in FB mode, outperforming the single run iterative learning implementation by achieving a significantly higher system clock frequency.

When operating in NFB mode, a partially i eline architecture with two additional su - i eline stages was found to offer the advantage of extremely high throughput rates once the latency of the i eline was met, with the 10-stage partial i eline implementation laying the best throughput results. Based on the delay analysis of the partial i eline implementations, it was determined that nearly two thirds of the run function's associated delay was attributed to the 2^{32} multiplier. Therefore, two additional i eline su - stages were implemented as to subdivide the multiplier into smaller blocks, resulting in a total of three i eline stages per run function. As a result, an increase by a factor of more than 2 was seen in the system's clock frequency, resulting in a similar increase in throughput when operating in NFB mode. Further su - i elining was not implemented as this would require subdividing the adders used to sum the partial results (a non-trivial task) to balance the delay between su - i eline stages.

5.2 Architectural Implementation Analysis — Rijndael

When implementing the Rijndael algorithm, it was first determined that the Rijndael -B boxes were the dominant element of the run function in terms of required logic resources. Each Rijndael run requires sixteen copies of the -B boxes, each of which is an 8-bit to 8-bit lookup table, requiring significant hardware resources. However, the remaining components of the Rijndael run function — byte swapping, constant Galois field multiplication, and key addition — were found to be simple in structure, resulting in these elements of the run function requiring few hardware resources. Additionally, it was found that the synthesis tools could not minimize the overall size of a Rijndael run sufficiently to allow for a fully unraveled fully i eline implementation of the entire ten runs of the algorithm within the target FPGA.

Surprisingly, a near run partially i eline implementation with one su - i eline stage revealed the most area-efficient solution. As compared to a one-stage implementation with no su - i elining, the addition of a su - i eline stage afforded the synthesis tool greater flexibility in its optimizations, resulting in a more area-efficient implementation. While 2-stage unravelling was found to yield the highest throughput when operating in FB mode, the measured throughput was within 10% of the single stage implementation. Due to the structural nature of the lace-and-run algorithms, one can expect a variance in performance based on differences in the starting point of the process. When performing this process multiple times, known as multi-pass lace-and-run, it is likely that the single run implementation would achieve a throughput similar to that of the 2-stage unraveled implementation.

When operating in NFB mode, partial i elining was found to offer the advantage of extremely high throughput rates once the i eline latency was met, with the 2-stage partial i eline implementation laying the best throughput results. While Rijndael cannot be implemented using a fully i eline architecture due to the large size of the run function, significant throughput increases were seen as compared to the unraveled architecture.

Subdividing of the partially i eline architectures was implemented by inserting a i eline su - stage within the Rijndael run function. Based on the delay analysis of the partial i eline implementations, it was determined that nearly half of the run function's associated delay was attributed to the -B box substitutions. Therefore, the additional i eline su - stage was implemented as to separate the -B boxes from the rest of the run function. As a result, an increase by a factor of nearly 2 was seen in the system's clock frequency, resulting in a similar increase in throughput when operating in NFB mode. Further su - i elining was not implemented as this would require subdividing the -B boxes (a non-trivial task) to balance the delay between su - i eline stages.

5.3 Architectural Implementation Analysis — Serpent

When implementing the Serpent algorithm, it was first determined that since the Serpent -Boxes are relatively small (4-bit to 4-bit), it is possible to implement them using combinational logic as sequential memory elements. Additionally, the -Boxes map extremely well to the Xilinx CLB slice, which is comprised of 4-bit look-up tables, allowing the -Boxes to be implemented in a total of two CLB slices, yielding a compact implementation which minimizes routing between CLB slices. Finally, the components of the Serpent round function — key masking, -Box substitution, an linear transformation — were found to be similar in structure, resulting in the round function requiring few hardware resources.

Implementing a single round of the Serpent algorithm requires the greatest area-optimized solution. However, a significant performance improvement was achieved by performing 8-round unrolling, removing the need for -Box multi-lexing hardware as needed by each serial -Box grouping is now included within one of the eight rounds. This amount of unrolling achieves a significant performance increase with little increase in hardware resources due to the compact nature of the Serpent round function. As expected, unrolling thirty-two rounds of the Serpent algorithm results in a lesser performance when compared to the eight-round implementation. Implementing the thirty-two rounds of the algorithm in a combinational logic severely hampers the overall clock frequency of the system, verifying the performance increase caused by the removal of the multi-lexing hardware required to switch between keys.

When operating in NFB mode, a full-length pipeline architecture was found to offer the advantage of extremely high throughput rates once the latency of the pipeline was met, but performing smaller partial pipeline implementations. In the fully pipeline architecture, all of the elements of a given round function are implemented as combinational logic. Other AE finalists cannot be implemented using a fully pipeline architecture due to the larger round functions. However, due to the small size of the Serpent -Boxes (4-bit look-up tables), the cost of -Box relocation is minimal in terms of the required hardware.

Finally, substituting of the partially pipeline architectures was determined to yield a throughput increase. Because the round function components are all similar in structure, there is little performance to be gained by substituting them with registers in an attempt to reduce the delay between stages. As a result, the increase in the system's clock frequency would not outweigh the increase in the number of clock cycles required to perform an encryption, resulting in a performance degradation.

5.4 Architectural Implementation Analysis — Twofish

When implementing the Twofish algorithm, it was first determined that the synthesis tools were unable to minimize the Twofish -Boxes to the extent of the AE finalist algorithms due to the -Boxes being key-dependent. Therefore, the overall size of a Twofish round was too large to allow for a fully unrolled fully pipeline implementation of the algorithm within the target FPGA. Moreover, the key-dependent -Boxes were found to require nearly half of the delay associated with the Twofish round function.

As expected, implementing a single round of the Twofish algorithm requires the greatest area-optimized solution in terms of total CLB slices required for the implementation. Additionally, unrolling requires a minor throughput increase as the decrease in the number of cycles per encryption clock was offset by the relatively decreasing system clock frequency. However, single stage partial pipelining with one substituting stage was found to yield the best throughput when operating in feedback mode. With a small increase in the required hardware resources, the substituting architecture was able to reach a significantly faster system clock frequency as compared to the fully unrolled and partially pipeline implementations.

When operating in NFB mode, a partially pipeline architecture was found to offer the advantage of extremely high throughput rates once the latency of the pipeline was met, with the 8-stage partial pipeline implementation laying the best throughput results. While Twofish cannot be implemented using a fully pipeline architecture due to the large size of the round function, significant throughput increases were seen as compared to the fully unrolled architecture.

Finally, su - i elining f the artially i eline architectures was im lemented y inserting a i eline su -stage within the Tw fish r un functi n. Base n the elay analysis f the artial i eline im lementati ns, it was etermine that nearly half f the r un functi n's ass ciate elay was attri ute t the -B x su stituti ns. Theref re, the a iti nal i eline su -stage was im lemented s as t se arate the -B xes fr m the rest f the r un functi n. As a result, an increase y a fact r f nearly 2 was seen in the system's cl ck frequency, resulting in a similar increase in thr ough ut when erating in NFB m e. Further su - i elining was n t im lemented as this w ul require su - ivi ing the -B xes (a n n-trivial task) t alance the elay etween su - i eline stages.

6 Perf rm nce Ev lu ti n

Tables 2 an 3 etail the thr ough ut measurements f r the im lementati ns f the three architecture ty es f r each f the AE finalists f r th NFB an FB m e. The architecture ty es — l un r lling (LU), full r artial i elining (PP), an artial i elining with su - i elining (P) — are liste al ng with the num er f stages an (if necessary) su - i eline stages in the ass ciate im lementati n; e.g., LU-4 im lies a l un r lling architecture with f ur r un s, while P-2-1 im lies a artially i eline architecture with tw stages an ne su - i eline stage er i eline stage. As a result, the P-2-1 architecture im lements tw r un s f the given ci her with a t tal f tw stages er r un . Thr ough ut is calculate as:

$$Throughput := (128 \text{ Bits} * \text{Cl ck Frequency}) / (\text{Cycles Per Encry te Bl ck})$$

Note that the im lementati n f a ne stage artial i eline architecture, an iterative l lling architecture, an a ne r un l un r lle architecture are all equivalent an are theref re n t liste se arately. Also note that the c m ute thr ough ut f r im lementati ns that em l y any f rm f har ware i elining (as discusse in ecti n 4) are made assuming that the i eline latency has een met.

The num er f CLBs require as well as the maximum erating frequency f r each im lementati n was taine fr m the Xilinx re rt files. Note that the Xilinx t ls assume the a s lute wrst ssi le erating c n iti ns — highest ssi le erating tem erature, l west ssi le su ly v ltage, an wrst-case fa ricati n t lerance f r the s ee gra e f the FPGA [23]. As a result, it is c mm n f r actual im lementati ns t achieve slightly etter erf rmance results than th se s ecifie in the Xilinx re rt files.

While this stu y f cuses n high thr ough ut im lementati ns, the har ware res urces require t achieve this thr ough ut is als a critical arameter. No esta lishe metric exists t measure the har ware res urce c sts ass ciate with the measure thr ough ut f an FPGA im lementati n. Tw area measurements f FPGA utilizati n are reaily a arent — l gic gates an CLB slices. It is im rtant t n te that the l gic gate c unt es n t yiel a true measure f actual FPGA utilizati n. Har ware res urces within CLB slices may n t e fully utilize y the lace-an -r ute s ftware s as t relieve r uting c ngesti n. This results in an increase in the num er f CLB slices with ut a c rres n ing increase in l gic gates. T achieve a m re accurate measure f chi utilizati n, CLB slice c unt was ch sen as the m st relia le area measurement. Theref re, t measure the har ware res urce c sts ass ciate with an im lementati n's resultant thr ough ut, the Thr ough ut Per lice (TP) metric is use . We efine TP as:

$$TPS := (\text{Encry ti n Rate}) / (\# \text{ CLB lices Use })$$

Theref re, the timal im lementati n will is lay the highest thr ough ut an have the largest TP . Note that the TP metric ehaves inversely t the classical time-area (TA) r uct.

Alg.	Arch.	Throughput (Gbit/s)	Slices	TPS
RC6	P-10-2	2.40	108 6	220881
Rijndael	P-1	1.4	10 2	1762 7
Serpent	PP-32	4.86	004	3 778
Twofish	P-8-1	1.5	34	16 63

Table 4: AE finalist performance evaluation — throughput — network fee — network efficiency — throughput — network efficiency

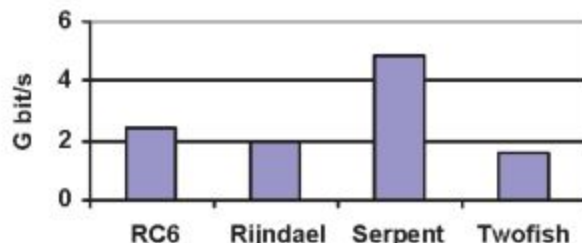


Figure 1: Best throughput — network fee — network efficiency

Alg.	Arch.	Throughput (Mbit/s)	Slices	TPS
RC6	PP-2	126.5	318	3 662
Rijndael	LU-2	300.1	302	660
Serpent	LU-8	444.2	7 64	771
Twofish	P-1-1	11 .6	30 3	3 16

Table 5: AE finalist performance evaluation — throughput — network fee — network efficiency — throughput — network efficiency

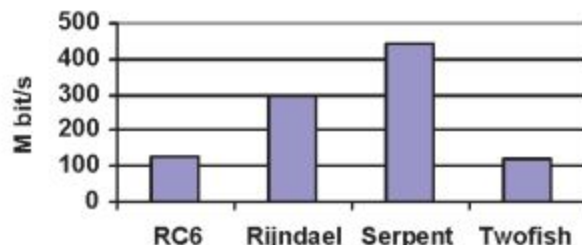


Figure 2: Best throughput — network fee — network efficiency

Tables 4 and 5 detail the optimal implementations of the AE finalists in the FB and NFB modes. Additionally, TPS is also shown for each of the implementations. It is critical to note that for the purposes of this study, the optimal implementation for an AE finalist is defined to yield the highest throughput. As previously discussed, the synthesis tools were not optimized for speed to guarantee that the high throughput would be achieved for each implementation. However, should an optimal implementation be found based on either TPS or area, the implementation results shown in Tables 2 and 3 (and, as a result, the ones shown in Tables 4 and 5 as well) are no longer representative of the best possible implementations for the architectures studied. To achieve a true representation that defines optimality based on either TPS or area, synthesis must be performed with the tools optimized for area. While an area-efficiency analysis of the AE finalists warrants investigation, it is beyond the scope of this study.

Based on the data shown in Tables 4 and 5, the Serpent algorithm clearly outperforms the other AE finalists in the metrics of throughput. As compared to its nearest competitor, Serpent exhibits a throughput increase of a factor 2.2 in NFB mode and a factor 1.5 in FB mode. Interestingly, RC6, Rijndael, and Twofish

all exhibit similar performance results in NFB mode. However, Rijndael exhibits significantly improved performance in FB mode as compared to RC6 and Twofish, although it is still 10% slower than Serpent.

One of the main findings of our investigation, namely that Serpent appears to be especially well suited for an FPGA implementation from a performance perspective, seems especially interesting considering that Serpent is clearly not the fastest algorithm with respect to most software comparisons [1]. Another major result of our study is that all four algorithms consistently and easily achieve Giga-bit encryption rates with standard commercially available FPGAs. The algorithms are at least one order of magnitude faster than the best reported software realizations. These speed-ups are essentially achieved by parallelization (i.e., eliminating and/or unrolling) of the loop structure and by wide operand processing (e.g., processing of 128 bits in one clock cycle), both of which are not feasible in current processors. We would like to stress that the pipeline architectures cannot use to their maximum advantage fast operations which require feedback (CFB, FB, etc.) However, we believe that for many applications which require high encryption rates, non-feedback modes (i.e., modes such as interleaved CFB [3, section 12]) will be the most effective. Note that the Counter Mode grew out of the need for high speed encryption of ATM networks which require parallelization of the encryption algorithm.

7 Conclusions

The importance of the Advanced Encryption Standard and the significance of high throughput implementations of the AE finalists has been examined. A design methodology was established which in turn led to the architectural requirements for a target FPGA. The characteristics of the AE finalists were identified and multiple architecture options were discussed. The implementation of each architecture option for each of the AE finalists was analyzed to determine their suitability for hardware implementation. Based on the implementation results, the best speed-optimized implementations were identified for each AE finalist in both non-feedback and feedback modes. Unfortunately, it was determined that the Serpent algorithm yields the best performance in both modes, where best performance was defined strictly as the highest throughput. The Serpent algorithm outperforms its nearest competitor by a factor of 2.2 in non-feedback mode and by a factor of 1.1 in feedback mode.

8 Acknowledgement

We would like to thank Pawel Chodowiec and Kris Gaj from George Mason University for their helpful discussions and the VHDL code modules that were provided to assist in the implementation of some of the AE finalists. We would also like to thank Alan Martell from the University of Pittsburgh for his useful VHDL code module that was used in implementation of the AE finalists.

References

- [1] D. Tinsley, *Cryptography, Theory and Practice*. Boca Raton, FL: CRC Press, 1996.
- [2] National Institute of Standards and Technology (NIST), *Second Advanced Encryption Standard (AES) NIST FIPS*, (Rome, Italy), March 1999.
- [3] B. Schneier, *Applied Cryptography*. John Wiley & Sons Inc., 2nd ed., 1996.
- [4] R. Du, "Hardware Cryptolite: A Fast VPN," *EETimes*, pp. 7-64, April 1999.
- [5] B. Golan, "Implementation Experience with AE Candidate Algorithms," in *Proceedings: Second AES Candidate NIST FIPS (AES 2)*, (Rome, Italy), March 1999.