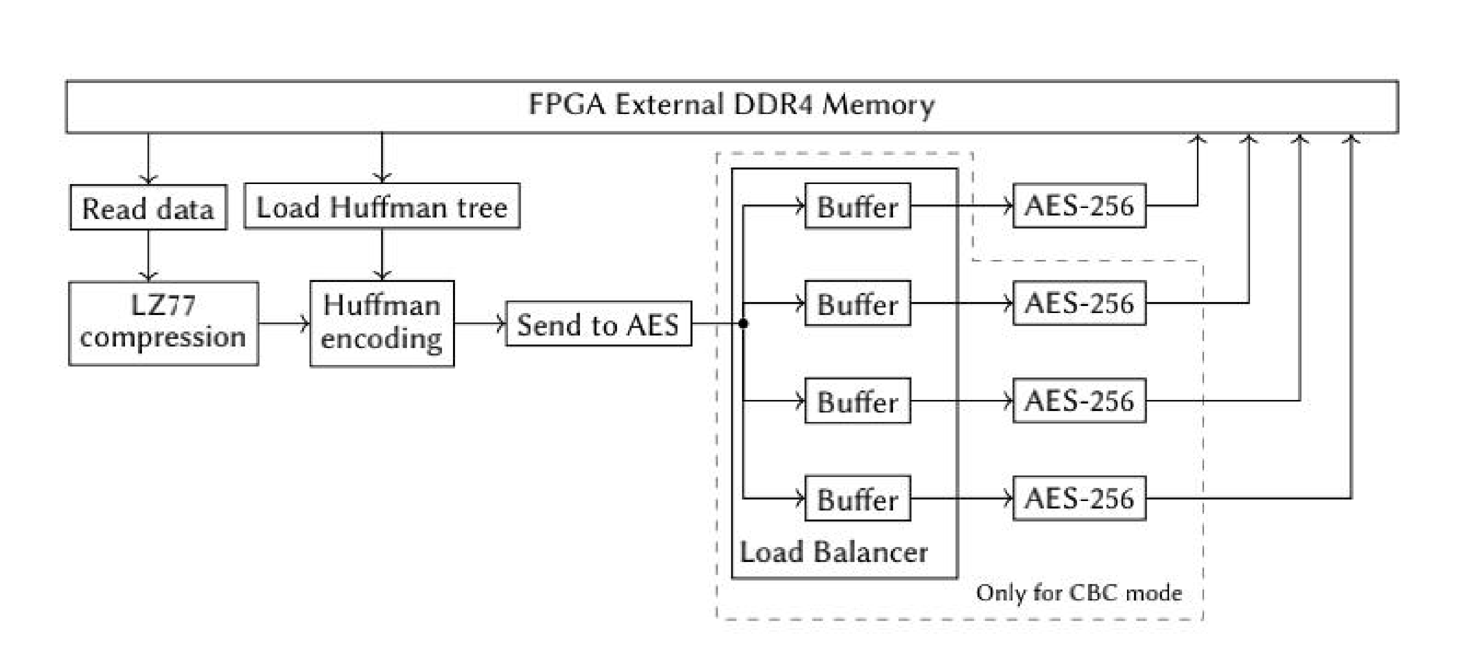
Universität Bamberg

Hardware Acceleration of Compression and Encryption in SAP HANA

June 28, 2023

DI XU

Compression and Encryption



Why compress and encrypt data?

- Compression reduces the amount of space and bandwidth needed to store and transmit data, which can improve efficiency, speed, and cost.
- Encryption transforms data into an unreadable form that can only be decrypted with a key, which can prevent unauthorized access, tampering, and leakage.

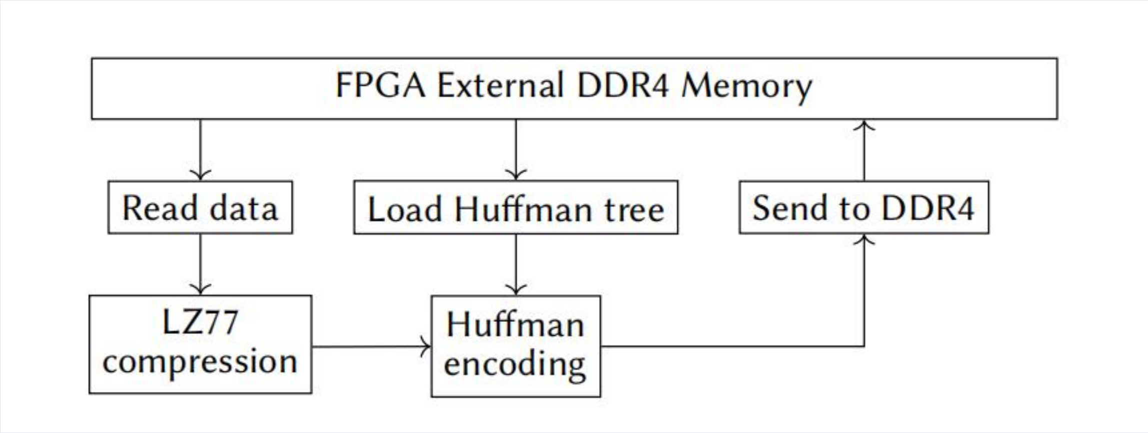
→ data security and performance

Compression

- Lossless Compression
  - Huffman encoding
  - LZ77 compression
  - DEFLATE method
- Lossy Compression
  - JPEG
  - MP3
  - MPEG

The DEFLATE Compression method

- Two parts:
  - LZ77 compression
  - Huffman encoding

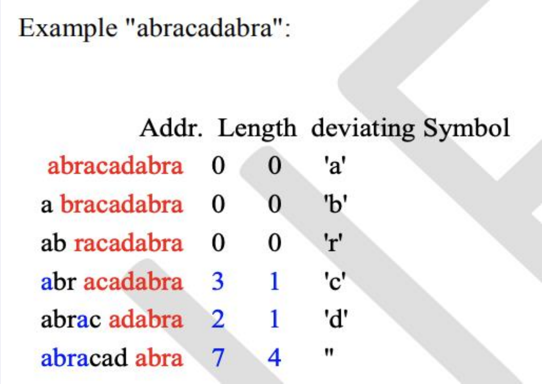


LZ77 compression

LZ77 is a dictionary based algorithm that addresses byte sequences from former contents instead of the original data.

In general only one coding scheme exists, all data will be coded in the same form:

- Address to already coded contents
- Sequence length
- First deviating symbol



Example "abracadabra":

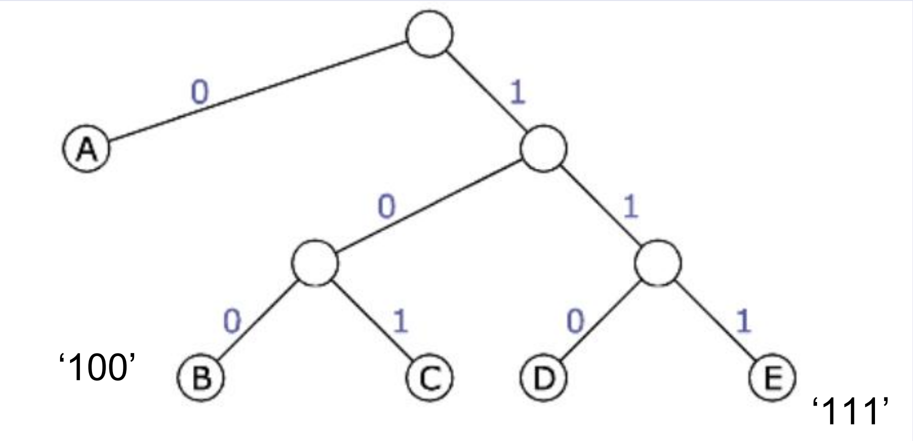
	Addr.	Length	deviating Symbol
abracadabra	0	0	'a'
a bracadabra	0	0	'b'
ab racadabra	0	0	'r'
abr acadabra	3	1	'c'
abrac adabra	2	1	'd'
abracad abra	7	4	''

The achievable compression rate is only depending on repeating sequences.

- ➢ If no identical byte sequence is available from former contents, the address 0, the sequence length 0 and the new symbol will be coded.
- ➢ Each byte sequence is extended by the first symbol deviating from the former contents,the set of already used symbols will continuously grow

Huffman encoding

The algorithm as described by David Huffman assigns every symbol to a leaf node of a binary code tree. These nodes are weighted by the number of occurences of the corresponding symbol called frequency or cost.



The character with the highest frequency is the closest to the root node of the tree

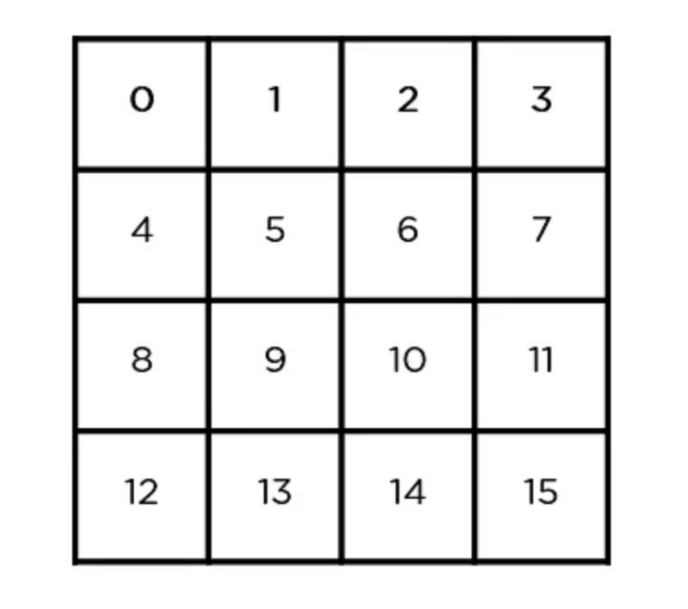
Encryption

- Symmetric Encryption：AES,DES
- Asymmetric Encryption: RSA

Symmetric Encryption	Asymmetric Encryption
Uses a single key to encrypt and decrypt data	Uses a public key to encrypt data and a private key to decrypt data
Faster encryption process	Slower encryption process
Example key sizes are 128 or 256-bit long	Example key sizes are 2048-bit or longer

The Advanced Encryption Standard(AES)

- AES is a block cipher.
- The key size can be 128/192/256 bits.
- Encrypts data in blocks of 128 bits each.



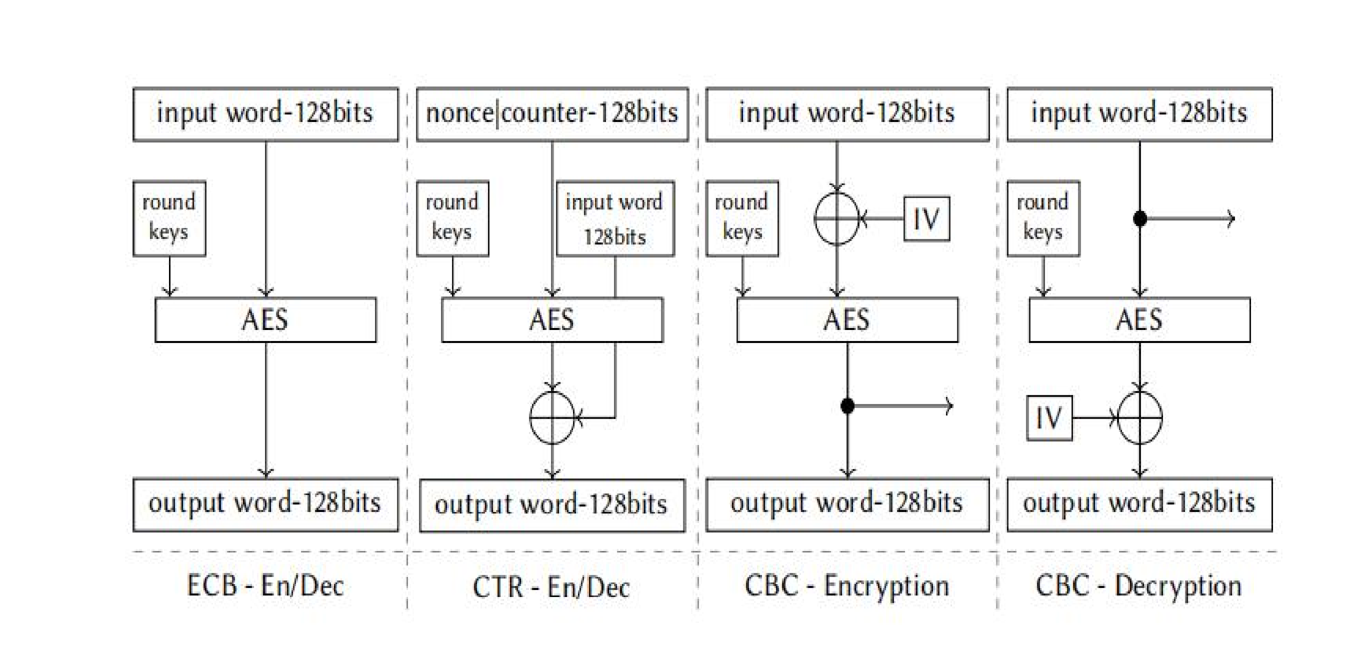
0	1	2	3
4	5	6	7
8	9	10	11
12	13	14	15

The Advanced Encryption Standard(AES)

- A round consists of several processing steps that include substitution, transposition and mixing of the input plaintext to transform it into the final output of ciphertext.
- The number of rounds depends on the key length as follows :
    - 128 bit key – 10 rounds
    - 192 bit key – 12 rounds
    - 256 bit key – 14 rounds

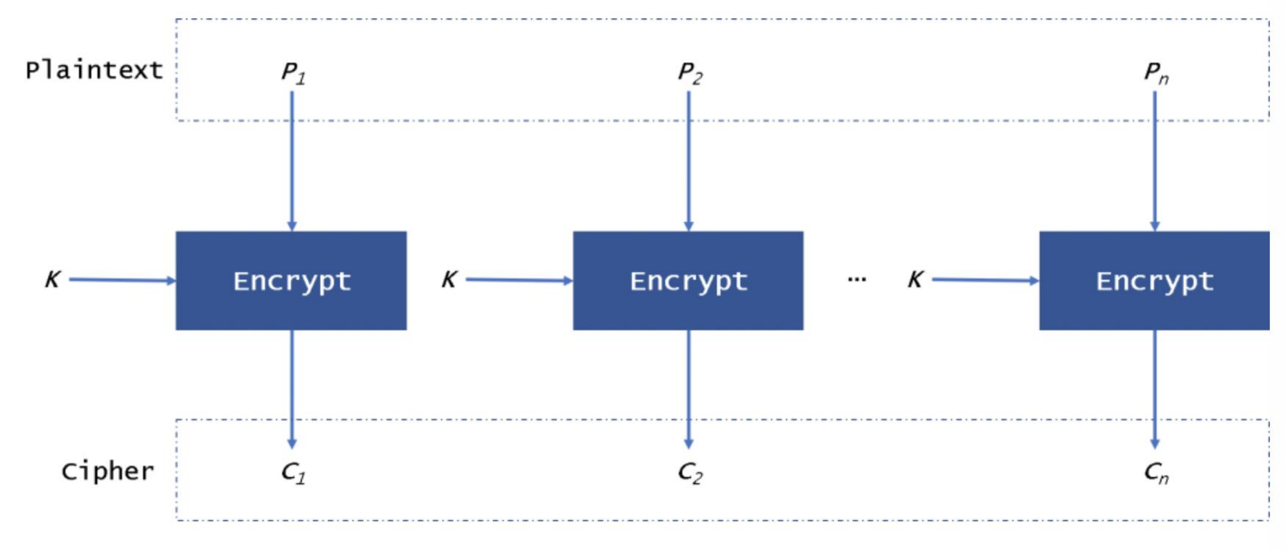
The Advanced Encryption Standard(AES)

- Three different AES block cipher modes

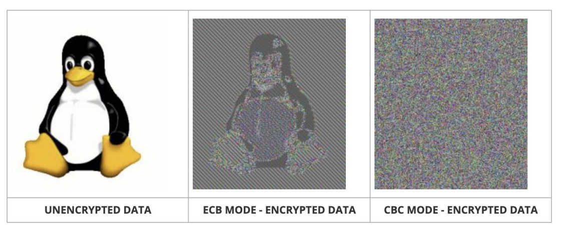


ECB(Electronic Code Book)

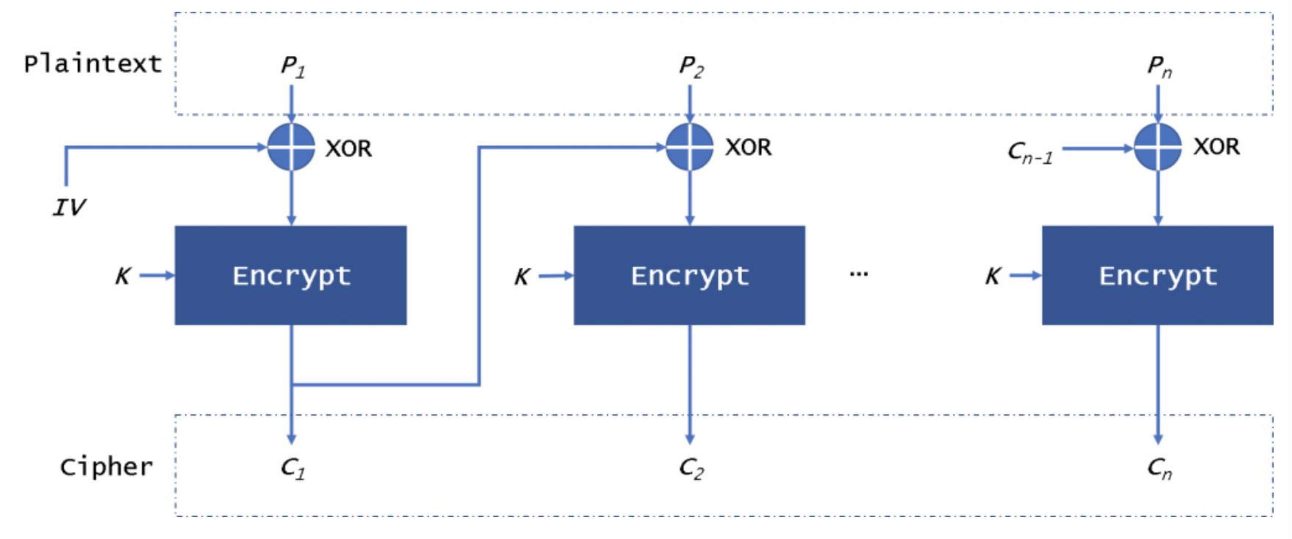
- The first generation of the AES. It is the most basic form of block cipher encryption.



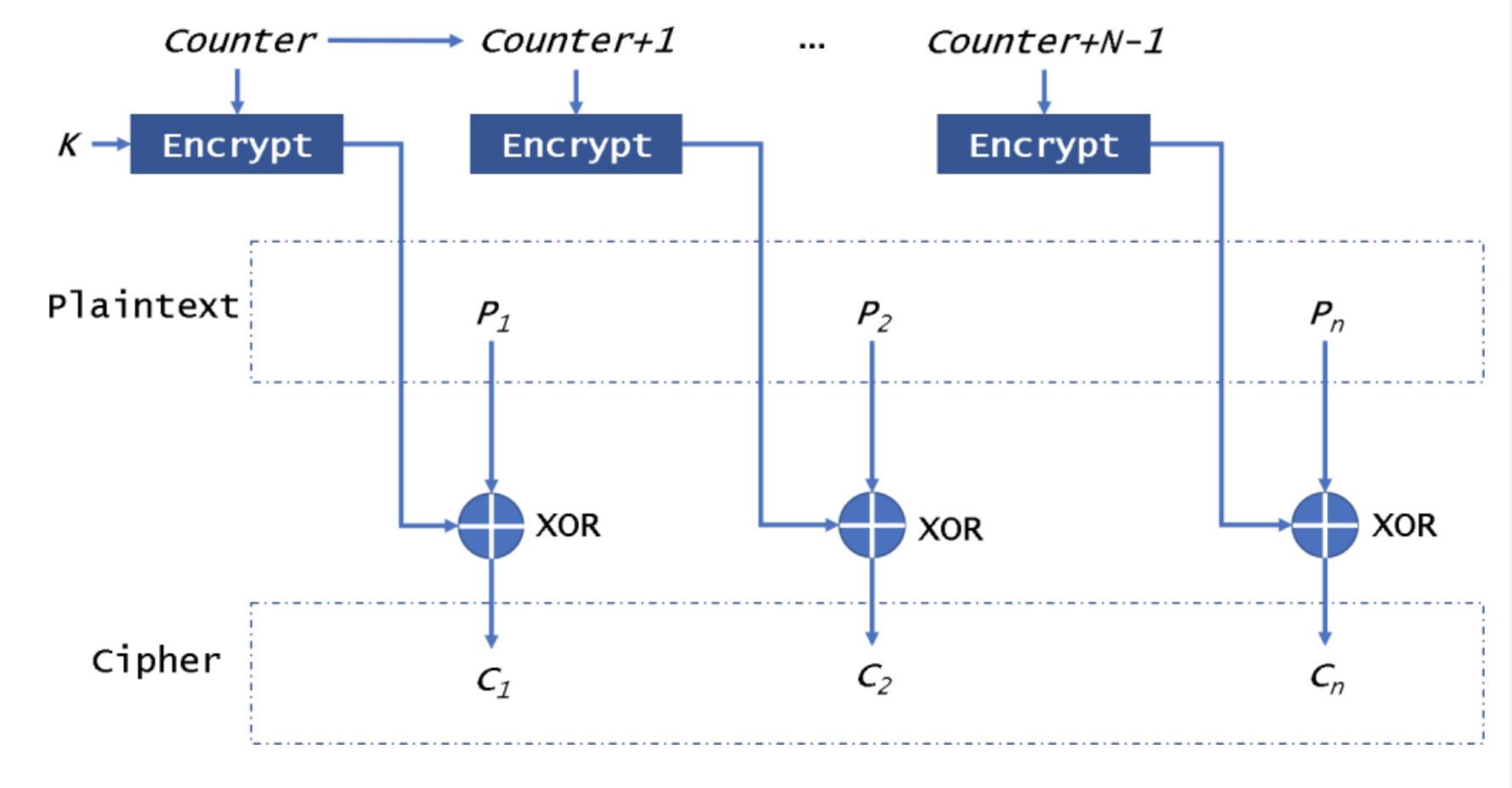
CBC(Cipher Block Chaining)



- An advanced form of block cipher encryption. With CBC mode encryption, each ciphertext block is dependent on all plaintext blocks processed up to that point. This adds an extra level of complexity to the encrypted data.



CTR(Counter)

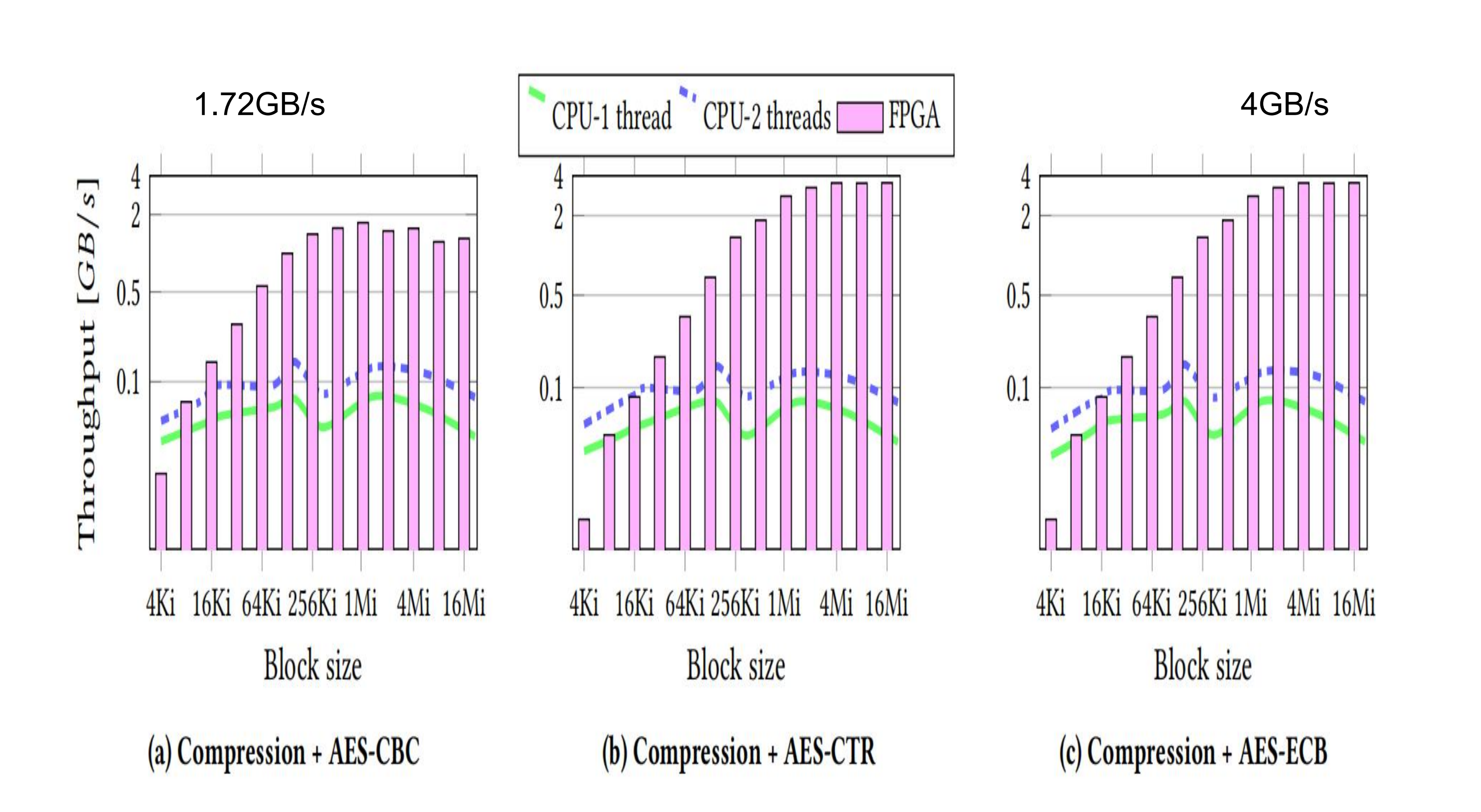

Counter
Counter+1
...
Counter+N-1
K
Encrypt
Encrypt
Encrypt
Plaintext
P_1
P_2
P_n
XOR
XOR
XOR
Cipher
C_1
C_2
C_n

Evaluation

SAP HANA
Block size: 4K-16M

1.72GB/s

4GB/s



(a) Compression + AES-CBC

(b) Compression + AES-CTR

(c) Compression + AES-ECB

Conclusion

- Data compression and encryption are two techniques that help improving the performance and security of database systems.

How to order compression and encryption ?

- Compressing before encrypting can offer better performance and security benefits.
- Compressing after encrypting can also have some advantages, such as avoiding compression attacks or preserving encryption metadata.

References

- The paper discussed:

Chiosa M, Maschi F, Müller I, et al. Hardware acceleration of compression and encryption in SAP HANA[C]//48th International Conference on Very Large Databases (VLDB 2022). 2022.

- Further references:

  - https://www.linkedin.com/advice/1/how-do-you-balance-data-security-performance
  - Oswal S, Singh A, Kumari K. Deflate compression algorithm[J]. International Journal of Engineering Research and General Science, 2016, 4(1): 430-436.
  - https://www.geeksforgeeks.org/advanced-encryption-standard-aes/

Thanks for your attention!

Questions?