# Authentication codes

Information security proposal for 2022 Wolfram summer school

Armando Benjamín Cruz Hinojosa
*Universidad Nacional Autónoma de México*
Mexico City, Mexico
aleph_g@ciencias.unam.mx

***Abstract*—Current authentication schemes are based on asymmetric cryptography protocols like RSA, with the arrival of quantum computing and fast factorization algorithms, security is endangered. A-schemes presents an alternative that remains secure regardless of computational power, but there are no tools to measure how secure they are. This proposal intends to create such tool.**

## I. Introduction

Authentication is a fundamental aspect of information security, it provides protection against impersonations and false messages. A strategy implemented in a communication system to achieve such protection is called an *authentication scheme*.

An authentication scheme is said to be secure if the if the chance that an impostor fools the system is so small that it doesn't represent a risk in the communication (comparable to channel noise).

A common scheme is to add a *digital signature* at the end of each message. This signature is generated with asymmetric cryptography: The message is compressed to a fixed size with a cryptographic hash function (like SHA-256), this hash is then encrypted with the transmitter's private key. In order to verify the message, the receiver compares the message's hash with the decrypted signature if they differ the message is forged.

Without the private key the best the enemy can do is try different keys until the message is accepted. This takes $2^{n-1}$ tries on average ($n$ is the size of the key). HTTPS specifies a maximum key size of 2048 bits and RSA encryption. A brute force attack would take $2^{2047} \approx 10^{616}$ tries, impossible task for current computers.

But if an efficient solution to the factorization into prime factors is used to break the key (Shor's algorithm [1]) or if the impostor's computational power drastically increases, the digital signature is useless for authentication protection.

With this in mind [2] proposes a different kind of authentication schemes, called *A-schemes*, that remains secure regardless of the impostor's computational power. Three parts interact in the communication system: $A$ an information source that wishes to communicate an element of the set of all possible source states $\mathcal{S}$, $B$ a receiver and $E$ the impostor.

$A$ will transform the source states into *ciphered messages* (sequences compatible with the communication channel) by means of an *encoding rule*. An encoding rule $r \in \mathcal{R}$ is a one-to-one mapping from $\mathcal{S}$ to $\mathcal{M}$. $B$ accepts a message $m$ as authentic if $m \in r[\mathcal{S}]$ and retrieves the original state with the pre-image $r^{-1}(m)$. This schema is represented in Figure 1.

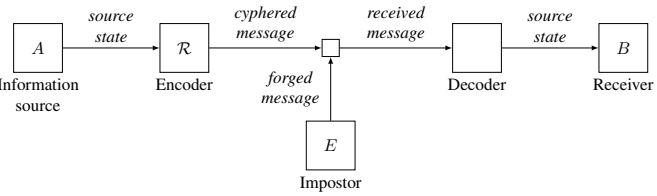


Fig. 1. A-scheme communication system.

An A-schema is a 5-tuple $(\mathcal{S}, \mathcal{M}, \mathcal{R}, S_{\mathbb{N}}, P_{\mathcal{R}})$, where $S_{\mathbb{N}}$ denotes a stochastic process over the set of all source states ie. the *information source*, and $P_{\mathcal{R}}$ is the probability distribution over the set of encoding rules. An *A-code* is the 3-tuple $(\mathcal{S}, \mathcal{M}, \mathcal{R})$ also referred as the A-scheme's combinatorial structure.

The impostor's best strategy is to listen to the first $t$ ciphered messages and with a cryptanalytical analysis, forge a message with the highest probability of success. The *unconditional probability of success of order* $t$ ($P_t$) is the expected percentage of successfully spoofed messages.

$P_t$ is the measurement of the security of an A-scheme, [3] demonstrates the theoretical lower bound for the unconditional probability of success, in terms of information entropy:

$$P_t \geq e^{H(R \mid M^{t+1}) - H(R \mid M^t)}$$

Finally [4] characterizes the A-schemes that achieve the theoretical lower bound. Such schemes satisfy two conditions.

- The information source $S_{\mathbb{N}}$ is message uniform.
- The underlying A-code structure is a strong partially balanced $t$-design (SPBD).

## II. Objective

The unconditional probability of success is the *expected* percentage of successfully spoofed messages. However very little is known about the probability distribution and variance of this random process.

On the other hand, the caracterization of perfect A-schemes imposes a restriction over the information source ($S_{\mathbb{N}}$ must be message uniform), that most information sources don't meet. Several unknown questions rise:

- How much $P_t$ differs from its theoretical lower bound, given an A-schemes with non message uniform source?
- Are SPBD the best A-code structure for such schemes?

Insight is needed to solve both problems. My proposal's objective is to implement in Wolfram language, a pseudo random simulation tool of the impostor's strategy, that will allow to generate data to statistically describe the spoof percentage, distribution and variance.

The simulations will be done over distinct information sources and distinct A-codes, focusing on the two questions related to non message uniform information sources.

## III. Implementation

Once the data points are generated, using Wolfram's *least square fitting* regression will describe the distribution and variance. Using Wolfram's descriptive statistics and plotting functionality, I hope to find a relation on the difference to the theoretical lower bound, in the non perfect A-schemes, this will lead to a mathematical hypothesis.

The problem lies in generating the data points ie. the simulation. This simulation must be as unbiased as possible, and optimal in order to generate a lot of data in short time. In order to achieve the simulation algorithm two things are needed.

### A. Information source simulation

Shannon defines an information source as a family of random variables dispensing sequences words in time: a discrete stochastic process over the set of all source states.

Last year, Mythreyi Namuduri's Wolfram summer school project [5], showed this statistical nature of language, by listing the most common words for less commonly spoken languages.

Wolfram's WordList function together with a Markov chain and a random number generator can be used to simulate an information source on a particular language.

### B. A-codes and SPBDs

A block design (Fig 2) is a specific type of hypergraph that can be defined in Wolfram language as a list of sets of vertices. On the other hand, an A-code is a list of two sets and a family of one-to-one mappings from the first set to the second one, something that can be defined in Wolfram's language as a relation.

[4] characterization of perfect schemes gives an algorithm to transform a given SPBD into an A-code. This will be implemented as a function that takes an SPBD block design, the set of source states and the set of encoding messages and returns an A-code.

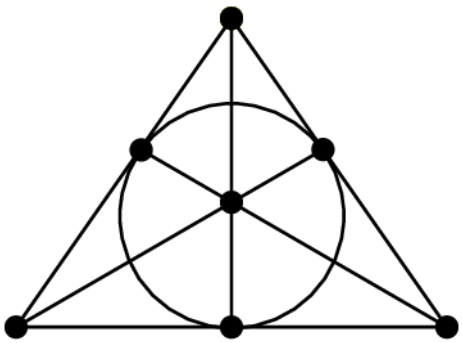

Fig. 2. Example of combinatorial block design.

## IV. Familiarity with the problem

I have been working with A-schsemes and A-codes for over a year for my bachelor degree's thesis (all UNAM's Math bachelors must present a thesis).

The content of the thesis is a proof (and explanation) of the theoretical lower bound described above, and a poof (and explanation) of the characterization of perfect A-schemes in terms of SPBDs block designs.

## References


[1] Shor, P.W. (1994). "Algorithms for quantum computation: discrete logarithms and factoring". Proceedings 35th Annual Symposium on Foundations of Computer Science. IEEE Comput. Soc. Press: 124–134. doi:10.1109/sfcs.1994.365700. ISBN 0818665807. S2CID 15291489.

[2] Gilbert, E.N., Macwilliams, F.J., and Sloane, N.J.A. Codes which detect deception, Bell System Technical Journal, 53, 405, 1974.

[3] Pei, D. Information-theoretic bounds for authentication codes and block designs. Journal of Cryptology, 8, 177, 1995.

[4] Pei, D. (2006). Authentication Codes and Combinatorial Designs. CRC Press.

[5] Namuduri, M. (2021, June 20). [WSS21] Curating common word lists for less commonly spoken languages - Online Technical Discussion Groups—Wolfram Community. Wolfram Community. https://community.wolfram.com/groups/-/m/t/2312867