

Forschungspraxis, Bachelor's Thesis, Master's Thesis

[identification] Implementation of identification with universal hash functions

Identification is a communication scheme that allows rate doubly exponential in the blocklength, with the tradeoff that identities cannot be decoded (as messages do) but can only be verified.

- <https://ieeexplore.ieee.org/document/42172>
- <https://ieeexplore.ieee.org/document/42173>

The double exponential growth presents various challenges in the finite regime: there are heavy computational costs introduced at the encoder and decoder and heavy trade-offs between the error and the codes sizes.

The ultimate goal is to find a fast, reliable implementation while still achieving large code sizes.

Identification codes can be achieved by first removing the errors from the channel with regular transmission channel coding, and then sending a challenge through the corrected channel. For every identity i , the challenge is generated by picking a random input m and computing the corresponding output $T_i(m)$ using a function T_i that depends on the identity. The challenge is then the pair $m, T_i(m)$ and the receiver wanting to verify an identity j will verify whether $j=i$ by testing the challenge. This is done by recomputing the output with T_j and verifying whether $T_j(m) = T_i(m)$. The errors are reduced by ensuring that the various functions collide on a small fraction of the possible inputs.

It turns out that choosing good sets of functions $\{T_i\}$ is the same as choosing error-correction codes $\{c_i\}$ with large distance, where now each codeword c_i defines a function by mapping positions m (sometimes called code locators) to symbols c_{im} of the codeword.

We can thus construct identification codes by choosing error-correction codes where we are only interested in the performance of the error correction encoders (we are not interested in the error-correction decoder or error-correction codes).

Your task will be implementing the identification codes described in

- <https://ieeexplore.ieee.org/abstract/document/782144>

aiming at the fastest implementation, and testing their performance in comparison to other current implementations.

For reference, our previous work on identification based on Reed-Solomon and Reed-Muller code can be found at

- <https://arxiv.org/abs/2107.07649>
- <https://arxiv.org/abs/2107.06801>
- <https://arxiv.org/abs/2007.06372>

The coding will be in Python/Sagemath.
The working language will be in English.

Environment: we collaborate with LTI. At LNT and LTI there is currently a lot of funding for research in identification. Therefore you will find a large group of people that might be available for discussion and collaboration.

Advisors

Roberto Ferrara