Recently I have been working a problem of with a difficult set of constraints, and I would welcome any input or insight into the problem.

I need to build a Merkle-tree on-chain, which later on I can prove membership of using a zkSNARK proof, this means I would be able to do a zero-knowledge proof of membership and other fun things.

The problem is: using SHA2-256 (or keccak256) is very expensive

to prove inside a zkSNARK circuit, especially so when you need to prove 29 of them for a merkle tree path, and even moreso when you're trying to make it run in reasonable time in WebAssembly (which is about 42x slower).

Anyway, this lead me to polynomial hash functions over $GF(p)$

, UOWHFs etc. which can be quite cheap to evaluate in EVM. e.g. the Hash127 function can compute 32 bits per loop iteration, requiring a MODEXP

and MULMOD

instruction each iteration (or just a MULMOD

, is using precomputation).

However, many of the algorithms I've found are either ciphers or MACs which require a key, and in some cases, with the key, a collision can be found in linear time. e.g. with output $f(?) = x$

I can find any number of inputs y

where $f(y) = x$

. This is fine if the input, say the leaf of the Merkle tree, is secured in another way, e.g. $f(H(w)) = x$

requires you to find a collision where the input to f

matches the output of another more secure function H

, but chaining f

together to authenticate a Merkle tree path is equivalent in security to a single f

until you get to the leaf and need to find a collision between both H

and f

. e.g. the collision resistance is finding a specific input to the function, not a specific output.

Here are some interesting papers which discuss the topic in more detail, but they make no claims about preimage resistance or their suitability as cryptographic hash functions:

- On an almost-universal hash function family with applications to authentication and secrecy codes
- Universal and Perfect Hashing
- Strongly universal string hashing is fast
- Software-Optimized Universal Hashing and Message Authentication
- Fast Universal Hashing with Small Keys and No Preprocessing: The PolyR Construction
- COMPOSITIONS OF LINEAR FUNCTIONS AND APPLICATIONS TO HASHING
- A Fast Single-Key Two-Level Universal Hash Function

This seems to be a widely studied topic, and finding a few candidate functions with strong security guarantees which can also be computed cheaply in EVM would significantly reduce the complexity of zkSNARK circuits - making them provable on mobile devices and the web.

Potential functions:

- VSH, an Efficient and Provable Collision Resistant Hash Function/ Security of VSH in the Real World
- MiMC: Efficient Encryption and Cryptographic Hashing with Minimal Multiplicative Complexity
- hash127 - a provably secure 127-bit secret-key authenticator of an arbitrarily long message

Any ideas?