Title:
A blockchain is fragile against the database approach

Yoshiyasu Takefuji, Ph.D., Professor, Keio University
contact:takefuji@sfc.keio.ac.jp

At least more than $1.3 billion has been globally invested to blockchain-based systems in 2018 (1). A blockchain is resistant to modification of the data where each block uses a cryptographic hash. SHA-256 is a hashing function used in the blockchain. SHA-256 cannot be reversed because it's a one-way function. US military states that SHA-256 is appropriate for protecting classified information (2). Researchers are building blockchain-based systems to encourage patients to securely share information (3). Many scientists are excited about blockchain for healthcare (4,5,6).

Although sha-256 is irreversible, it is possible to reverse. This sha-256 algorithm takes as input a 2^64 maximum length message, and outputs a 256 bits hash. Instead of cracking the algorithm, the relationship between a message and a hash can be stored in a database. The single keyword reversibility is demonstrated by the following three steps:

1. The following simple Python program generates a hash for a message composed of a single keyword "sciencemag". Note that you must install hashlib library in Python.

```
import hashlib
ho=hashlib.sha256(b'sciencemag')
print(ho.hexdigest())
```

generated hash:
3f0197a909ecaba9e0107455e69219e2ae7f98bdcc4fb950c0543a42aa238eba

2. Copy this hash string and paste it to the following site:
http://md5decrypt.net/en/Sha256/#answer
3. Click the Decrypt button.

A message "sciencemag" will be displayed within 0.022 second. If you have two keywords for demonstrating the reversibility, we must use the k x k database where k is the number of keywords. Even if k is million keywords with two keywords in a blockchain,

building a tera (10^12) database is still feasible for reversing the blockchain. Therefore, we should use at least three keywords or more for securing the blockchain-based system. However, the current blockchain system usually uses two keywords (username and password) which should be avoided. The database approach simply defeats its design goal of the blockchain without cracking the algorithm as long as the database can be created.

References:

1. https://techcrunch.com/2018/05/20/with-at-least-1-3-billion-invested-globally-in-2018-vc-funding-for-blockchain-blows-past-2017-totals/
2. https://www.acq.osd.mil/dsb/reports/2000s/ADA498577.pdf
3. https://www.nature.com/articles/d41586-018-03067-x
4. https://catalyst.nejm.org/decoding-blockchain-technology-health/
5. http://science.sciencemag.org/content/sci/361/6405/859.full.pdf
6. https://jamanetwork.com/journals/jama/article-abstract/2677995