Hashing

Topics covered today

- Homework and review
- What is hashing
- Class exercise

Homework review

Did anyone do any further digging?

Are there any questions over this?

What questions would you NOT want me to ask on a quiz?

What is Hashing?

You guys tell me:

Wikipedia

> *A cryptographic hash function is a one way mathematical function that maps data of an arbitrary size to a bit array of fixed size. The output bit array is commonly called a hash or digest.*

Key tenets of a hash algorithm:

1. ***Fast***, must not take a lot of resources or time to compute
2. ***Irreversible***, must not be able to retrieve the original data from the resulting hash (one way transformation)
3. ***Deterministic***, for the same input value, a hash function must always provide the same output
4. A small change to the input should result in large change to the hash value (***avalanche effect***)

What is Hashing used for:

- verify the integrity of data (block chain, downloads, git commits, secure software)
- digital signatures
- as a part of verifying a given users *Authenticity* (a part of *Authentication*)
- Proof-of-work (mining, defense against Denial of Service)

Class Exercise

Bogus quiz results are in!

But first, a pet peeve of mine. . .

Link to quiz result data `sha256sum` :

`7e320905f9480a0257abe1176da9b632322b9fc3f4135d6152e8c350d68`

Do you trust the above?

Follow along (in `bash`) for some data science fun:

What is the difference between the following:

- `echo "Hi!"`
- `printf "Hi!"'` <—— Hint: Use this one for the exercise

What do the following do?

- `|` The pipe character
- `awk -F ',' '{ print $1 }'` Hint: pipe the `cat` output of quiz-data.csv into this
- `grep <string>`

Now about the quiz data, can we reverse our hash?

What are some common (semi) unique identifiers for people?

- campus W-number
- campus UID
- Social Security Number
- First initial Last name combo ie. `MKijowski`

There is a quiz covering today's in class exercise and the homework!

Quiz is available till 11:59pm tonight! One chance only this time! Take the quiz! Note: your grade might be low on this one, I may have broken the auto-grader...

Day 2 Hashing

Don't believe everything you read on the internet.

Attacks on hashes

- Pass the hash: an authenticating system accepts hashes, and you have them
- Dictionary attack: you have a dictionary of likely input data used to compute hashes
- Rainbow table: you have a dictionary of pre-computed hashes and known input data
- Collision: you guess/compute ***different*** data that computes to the same hash value
- Brute force: hash everything

Nonce and salt

Both added to data prior to the hashing function to increase uniqueness. Used for different reasons though.

- Salts increase complexity and prevent several known attacks on hash values (dictionary and rainbow table attacks)
- Nonce's are unique (number used only once) and are used to prevent replay attacks (cannot use same nonce) and in proof of work
- These do not increase any guarantee of integrity!!

Proof of Work

Proof of an amount of work prior to participating.

Created to reduce email spam.

Example: include a nonce that hashes to X number of leading 0's (see Hashcash)

Lets talk lab 1

10 commits == 10 points style == 10 points

task 1 == 50 points (15 points for salted quiz data set) task 2 == 50 points (15 points for coins)

Git

Anatomy of a git commit

Linux password hashes are stored in `/etc/shadow`. Lets make a new user and see if we can verify their hash!