If passwords have high entropy, then hard to recover pwd from H(pwd) [by one-wayness of H]
↳ But not true in practice...

Users often choose weak passwords (e.g., 123456, password, 123456789, ...)
↳ With a dictionary of 360 million entries, can cover about 25% of user passwords
(3% choose 123456)
(10% choose among top 25 common passwords)
} Based on password hashes that have been leaked from compromised databases

Simple hashing vulnerable to "offline dictionary attack":
adversary computes table (pwd, H(pwd)) for common passwords — completely offline
given H(pwd), can now invert with a single lookup if pwd is contained in the database
for LinkedIn breach in 2012, attacker stole password file with ~6 million passwords
(all passwords hashed using single iteration of unsalted SHA-1) → 90% of passwords recovered in ~6 days!

Problem: One-time precomputation (computing the lookup table) can be reused to compromise many passwords
Overall cost of attack: $O(m+n)$ where m is the dictionary size and n is the number of passwords to attack

Defense #1: Salt passwords before hashing: namely when storing password pwd, sample salt $\xleftarrow{R} \{0,1\}^n$ and store (salt, H(salt ‖ pwd)) on the server
↑ typically, $n \geq 64$

Note: Salt is a public value (needed for verification)

Offline dictionary attack no longer effective since every salt value induces different set of hash values
Overall cost of dictionary attack: $O(mn)$ — need to re-hash dictionary for every salt

Defense #2: Use a slow hash function [SHA-1 is very fast — enables fast brute-force search]
- PBKDF2 (password-based key-derivation function): iterate a cryptographic hash function many times:
(or bcrypt)
PBKDF2(pwd, salt): H(H(··· H(salt ‖ pwd) ···))
can use 100,000 or 1,000,000 iterations of SHA-256
honest user only needs to evaluate hash function once per authentication; adversary evaluates many times

Drawback: custom hardware can evaluate SHA-256 very fast
- scrypt (more recent: Argon2i): slow hash function that needs lots of memory (space) to evaluate
↳ custom hardware do not provide substantial savings (limiting factor is space, not compute)

Can also use a keyed hash function (e.g., HMAC with key stored in HSM)
↳ ensures adversary who does not know key cannot brute force at all!

Best practice: Always salt passwords
Always use a slow hash function (e.g., PBKDF2, scrypt) or keyed hash function or both!

```
$cur  = 'password'
$cur  = md5($cur)
$salt = randbytes(20)
$cur  = hmac_sha1($cur, $salt)
$cur  = remote_hmac_sha256($cur, $secret)
$cur  = scrypt($cur, $salt)
$cur  = hmac_sha256($cur, $salt)
```

md5: raw MD5 hash — not secure!
hmac_sha1: salted, keyed hash function
remote_hmac_sha256: (key on remote service)
scrypt: slow hash function

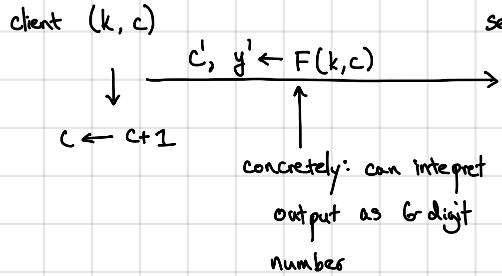
Facebook password onion (circa 2014)
↓
layers gradually added over time to achieve better security
(and probably to avoid password rehashing)

Password-based protocol <u>not</u> secure against eavesdropping adversary
(adversary sees vk and transcript of multiple interactions between honest prover + honest verifier)

One-time passwords (OTP) (SecurID tokens, Google authenticator, Duo)

<u>Construction 1</u>: Consider setting where verification key vk is <u>secret</u> (e.g., server has a secret)

- Client and server have a shared PRF key k and a counter (initialized to 0):



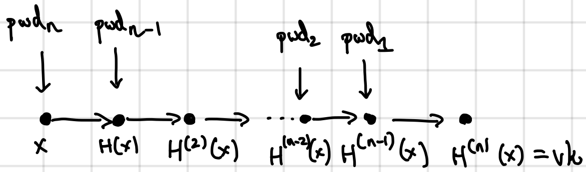
check that $y' = F(k, c')$ and $c' > c$ (no replaying)
if successful, update $c \leftarrow c'$
} car key authentication

- <u>RSA SecurID</u>: stateful token (counter incremented by pressing button on token)
  - ↳ State is cumbersome — need to maintain consistency between client/server
- <u>Google Authenticator</u>: time-based OTP: counter replaced by current time window (e.g., 30-second window)

If PRF is secure ⟹ above protocol secure against eavesdroppers (but requires <u>server</u> secrets)
↳ can be problematic: RSA breached in 2011 and SecurID tokens compromised and used to compromise defense contractor Lockheed Martin

<u>Construction 2</u>: No server-side secrets (S/Key)

- Relies on a hash function (should be one-way) ← "under composition"
- Secret key is random input x and counter n;
  Verification key is $H^{(n)}(x) = \underbrace{H(H(\cdots H(x)\cdots))}_{n \text{ evaluations of } H}$



to verify y: check $H(y) \stackrel{?}{=} vk$
if successful, update $vk \leftarrow y$
} attacker has to invert H in order to authenticate

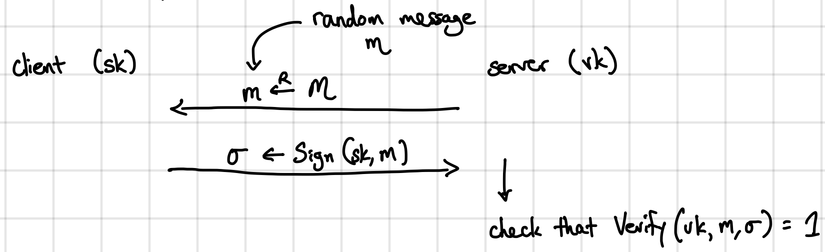
- Verification key can be public (credential is preimage of vk)
  - ↳ Can support <u>bounded</u> number of authentications (at most n) — need to update key after n logins
  - ↳ Output needs to be large (~80 bits or 128 bits) since password is the <u>input</u>/<u>output</u> to the hash function
- Naively, client has to evaluate H many times per authentication (~O(n) times)
  - ↳ Can reduce to O(log n) hash evaluations in an amortized sense by storing O(log n) entries along the hash chain

Thus far, only considered <u>passive</u> adversaries, but in reality, adversaries can be <u>malicious</u>

- Adversary can impersonate server (e.g., phishing) and then try to authenticate as client (but cannot interact with client during auth.) ← no man-in-the-middle protection
- All protocols thus far are vulnerable [all consist of client sending token that server checks, which can be extracted by active adversary]
- For active security, we use <u>challenge-response</u>

Signature-based challenge-response

- Server stores a verification key vk for digital signature scheme
- Client holds signing key sk



Server asks client to sign a random message
↳ Client's signature indicates proof of possession of sk associated with vk
↳ Active adversary that interacts with the client <u>before</u> interacting with the prover cannot forge signatures

Provides active security but signatures are long (~ 384 bits)

Signature-based challenge response: client "demonstrates knowledge" of signing key
↳ we will generalize this to "proving" <u>arbitrary</u> statements