CSE 484: Computer Security and Privacy

Cryptography
[MACs and Hash Functions]

Winter 2021

David Kohlbrenner

dkohlbre@cs.washington.edu

Thanks to Franzi Roesner, Dan Boneh, Dieter Gollmann, Dan Halperin, Yoshi Kohno, Ada Lerner, John Manferdelli, John Mitchell, Vitaly Shmatikov, Bennet Yee, and many others for sample slides and materials ...

Admin

- Homework 2
  - Out soon™

When is an Encryption Scheme "Secure"?

- Hard to recover the key?
  - What if attacker can learn plaintext without learning the key?
- Hard to recover plaintext from ciphertext?
  - What if attacker learns some bits or some function of bits?

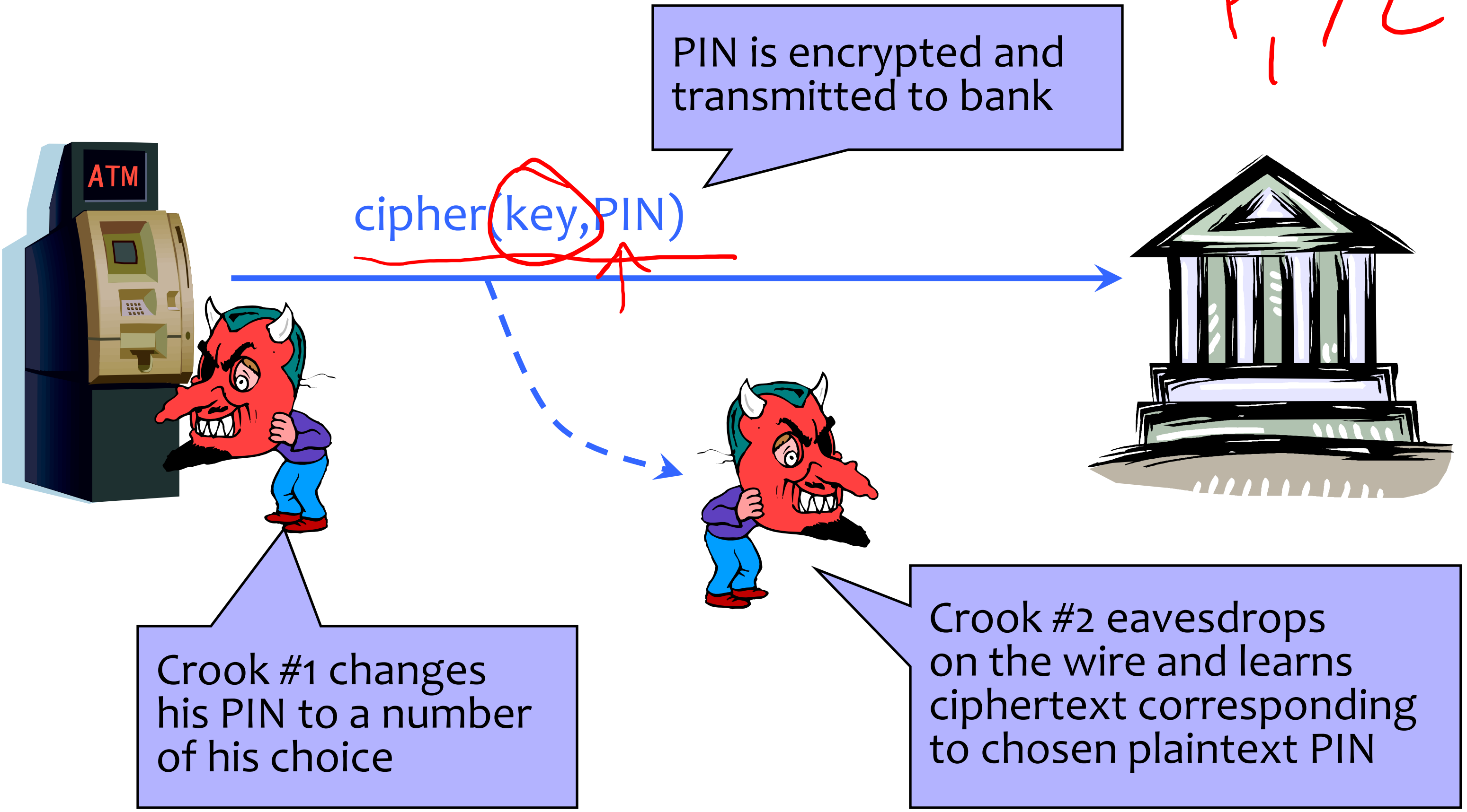
How Can a Cipher Be Attacked?

10x

- Attackers knows ciphertext and encryption algthm
  - What else does the attacker know? Depends on the application in which the cipher is used!

don't know keys

Chosen Plaintext Attack



… repeat for any PIN value

How Can a Cipher Be Attacked?

- Attackers knows ciphertext and encryption algthm
  - What else does the attacker know? Depends on the application in which the cipher is used!
- Ciphertext-only attack
- KPA: Known-plaintext attack (stronger)
  - Knows some plaintext-ciphertext pairs
- CPA: Chosen-plaintext attack (even stronger)
  - Can obtain ciphertext for any plaintext of his choice
- CCA: Chosen-ciphertext attack (very strong)
  - Can decrypt any ciphertext except the target

CCA2

Very Informal Intuition

Minimum security requirement for a modern encryption scheme

- Security against chosen-plaintext attack (CPA)
  - Ciphertext leaks no information about the plaintext
  - Even if the attacker correctly guesses the plaintext, he cannot verify his guess
  - Every ciphertext is unique, encrypting same message twice produces completely different ciphertexts
    - Implication: encryption must be randomized or stateful
- Security against chosen-ciphertext attack (CCA)
  - Integrity protection – it is not possible to change the plaintext by modifying the ciphertext

P_{10} / C_{10}

P_{10}/C_{12}

C_{10}

The shape of the formal approach

IND-CPA
IND-CCA
CCA2

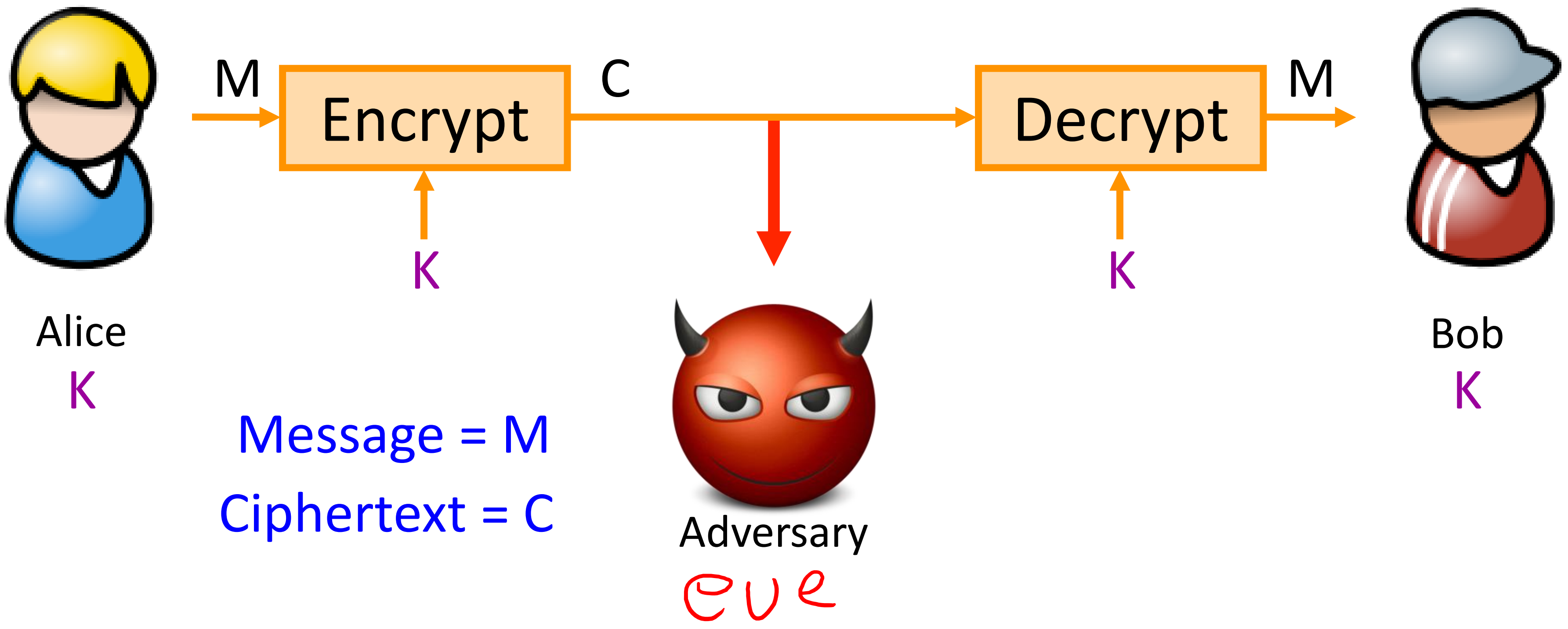
- INDistinguishability under Chosen Plaintext Attack
  - IND-CPA
- Formalized *cryptographic game*
- Adversary submits pairs of *plaintexts* (M_a, M_b)
  - Gets back ONE of the *ciphertexts* (C_x)
- Adversary must guess which ciphertext this is (C_a or C_b)
  - If they can do better than 50/50, they win

M_a M_b,
M_a M_c
M_a M_d

CIA

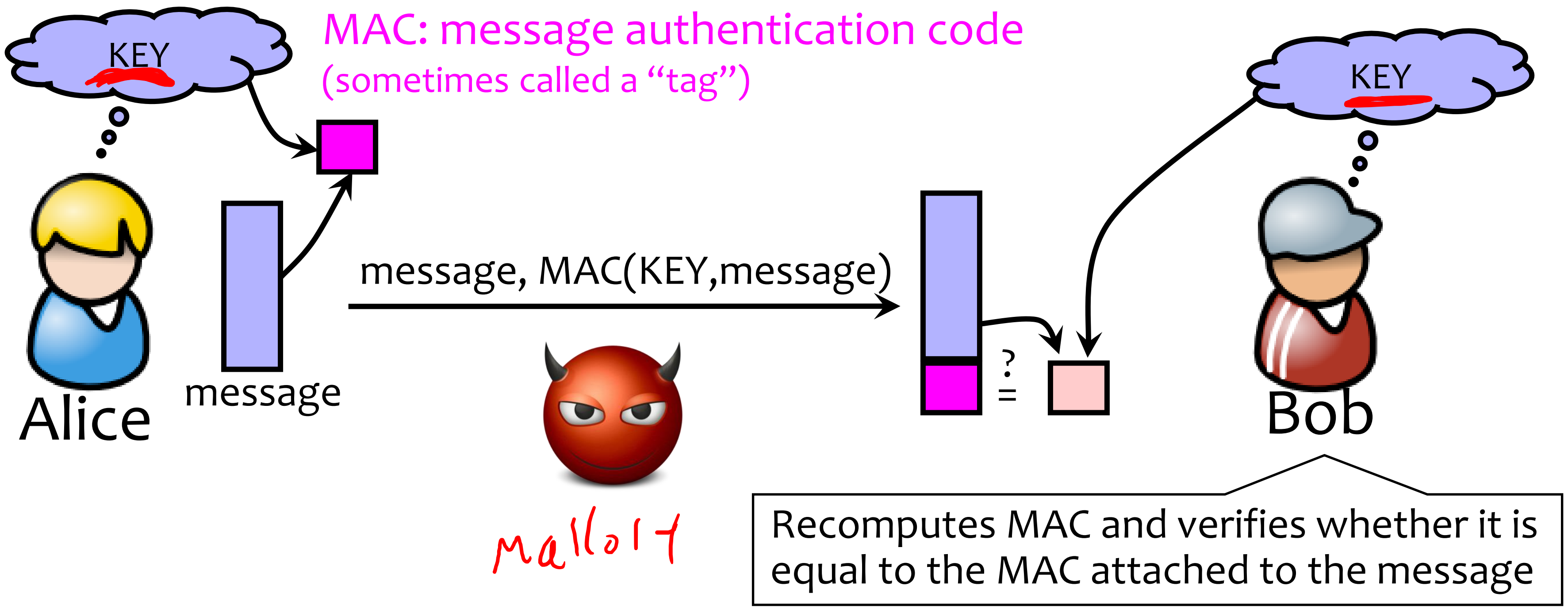
So Far: Achieving Privacy

Encryption schemes: A tool for protecting privacy.



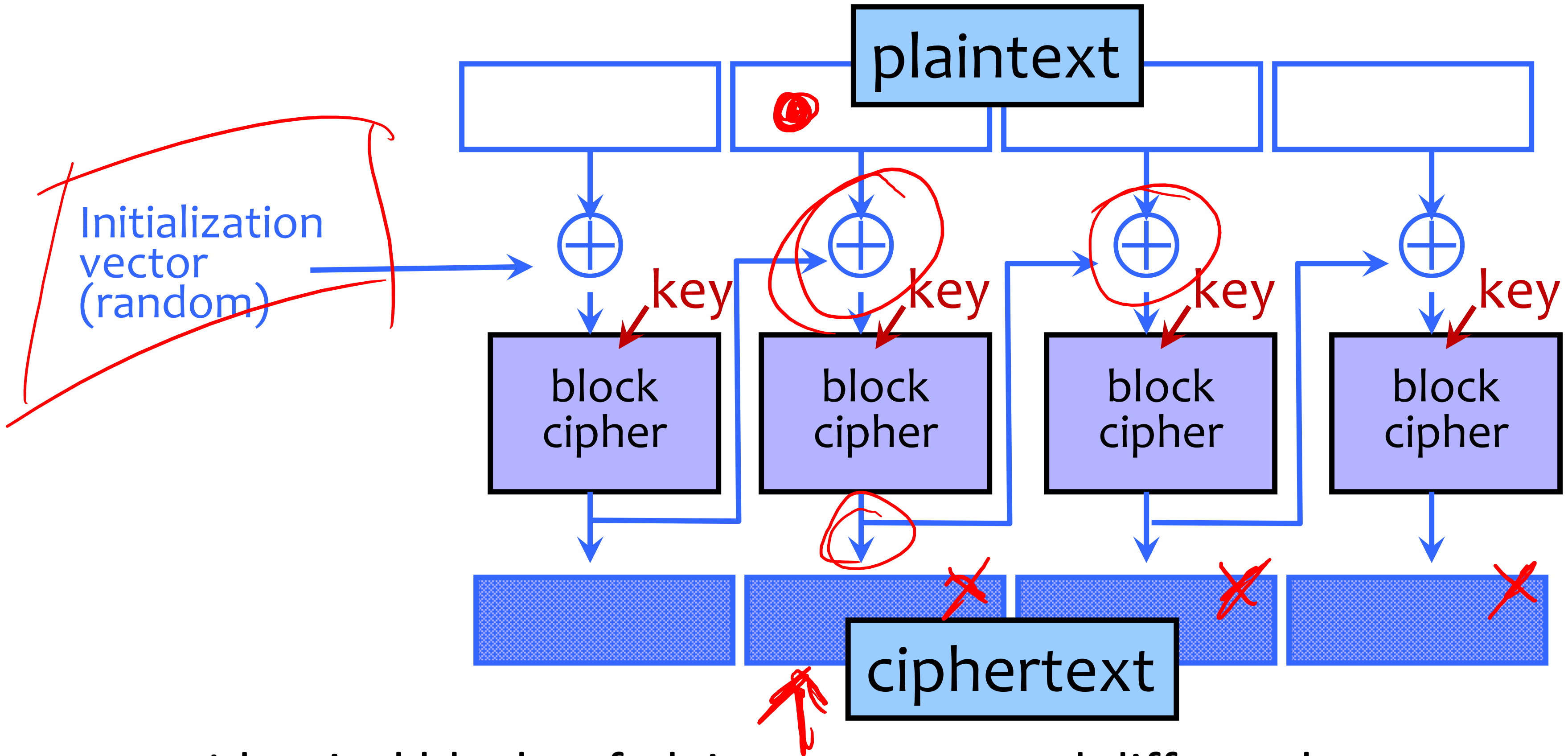
Now: Achieving Integrity

Message authentication schemes: A tool for protecting integrity.



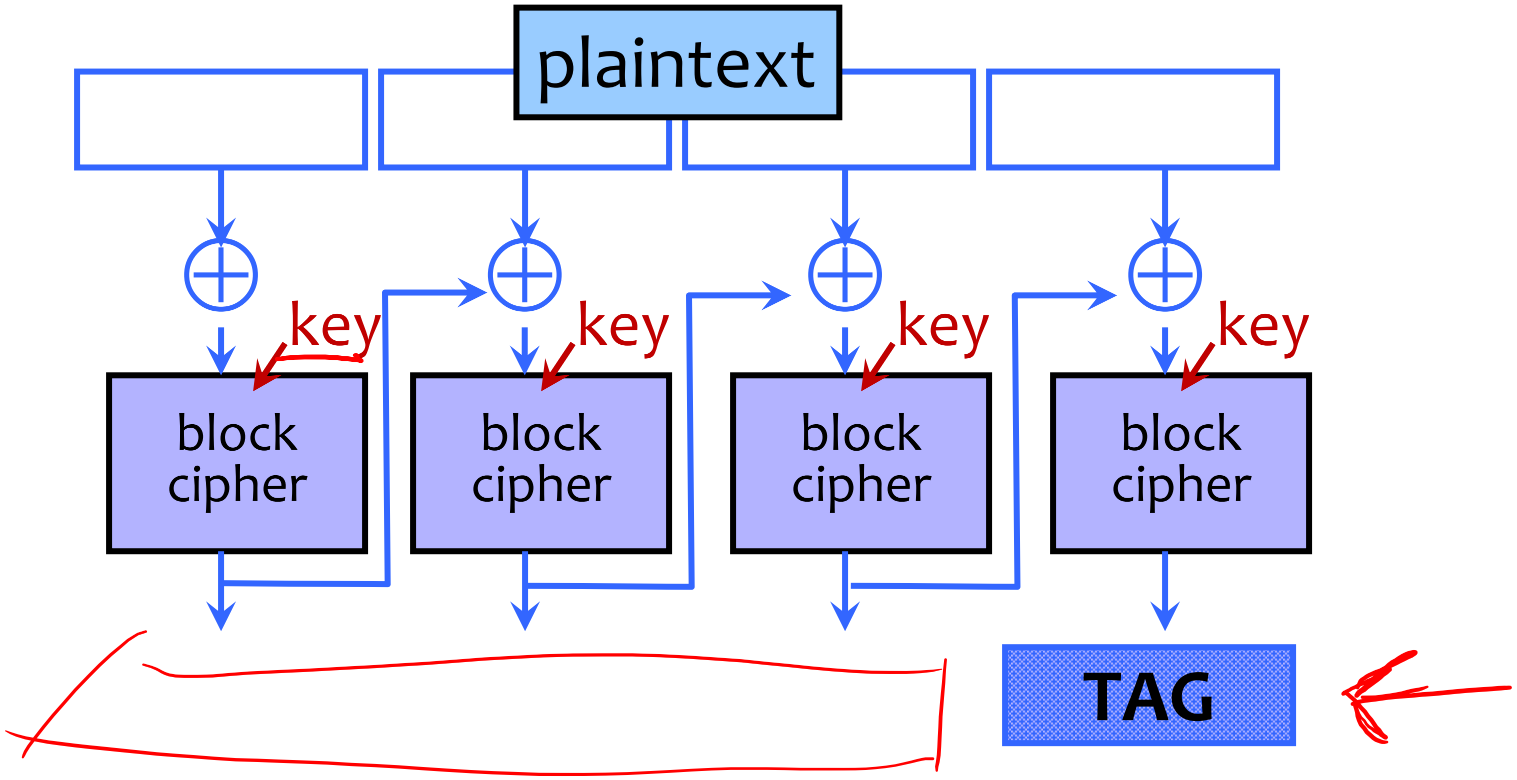
Integrity and authentication: only someone who knows KEY can compute correct MAC for a given message.

Reminder: CBC Mode Encryption



- Identical blocks of plaintext encrypted differently
- Last cipherblock depends on entire plaintext
  - Still does not guarantee integrity

CBC-MAC



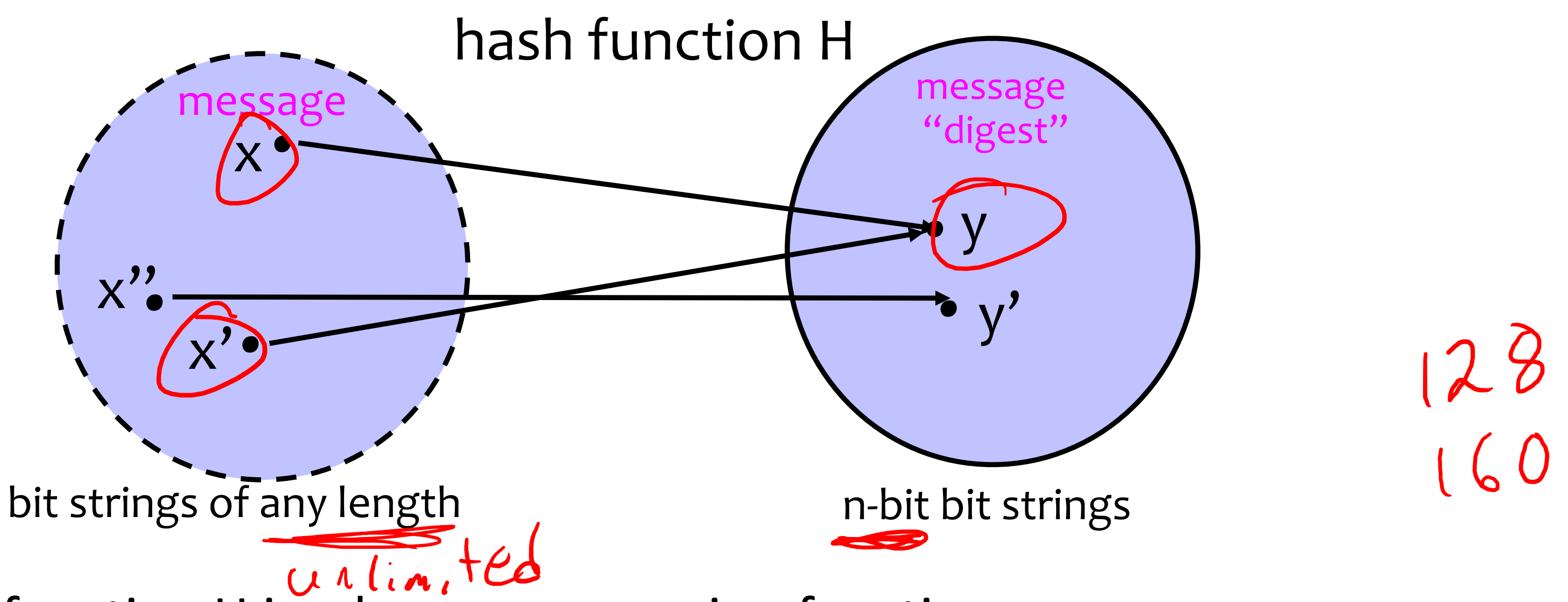
- Not secure when system may MAC messages of different lengths *(more in section!).*
- NIST recommends a derivative called CMAC [FYI only]

Another Tool: Hash Functions

You Just Did This

```
franzi@codered:~/sploits$ md5sum sploit0.c
3a2e6ce795bce4d06df1ff6835d25cea  sploit0.c
franzi@codered:~/sploits$ 
```

Hash Functions: Main Idea



- Hash function H is a lossy compression function
  - Collision: h(x)=h(x') for distinct inputs x, x'
- H(x) should look "random"
  - Every bit (almost) equally likely to be 0 or 1
- Cryptographic hash function needs a few properties…

Property 1: One-Way

H(x) no key

- Intuition: hash should be hard to invert
  - "Preimage resistance"
  - Let $h(x') = y \quad \{0,1\}^n$ for a random x'
  - Given y, it should be hard to find any x such that h(x)=y
- How hard?
  - Brute-force: try every possible x, see if h(x)=y
  - SHA-1 (common hash function) has 160-bit output
    - Expect to try 2^{159} inputs before finding one that hashes to y.

Property 2: Collision Resistance

- Should be hard to find x≠x' such that h(x)=h(x')

2^{154}

Birthday Paradox

- Are there two people in the first 1/8 of this class that have the same birthday?
  - 365 days in a year (366 some years)
    - Pick one person. To find another person with same birthday would take on the order of 365/2 = 182.5 people
    - **Expect birthday "collision" with a room of only 23 people.**
    - For simplicity, approximate when we expect a collision as **sqrt(365)**.
- Why is this important for cryptography?
  - 2^{128} different 128-bit values
    - Pick one value at random. To exhaustively search for this value requires trying on average 2^{127} values.
    - **Expect "collision" after selecting approximately 2^{64} random values.**
    - **64 bits** of security against collision attacks, not 128 bits.

x' $h(x') = h(x)$

Property 2: Collision Resistance

- Should be hard to find x≠x' such that h(x)=h(x')
- Birthday paradox means that brute-force collision search is only $O(2^{n/2})$, *not* $O(2^n)$
  - For SHA-1, this means $O(2^{80})$ vs. $O(2^{160})$

One-Way vs. Collision Resistance

One-wayness does **not** imply collision resistance.

Collision resistance does **not** imply one-wayness.

You can prove this by constructing a function that has one property but not the other.

Property 3: Weak Collision Resistance

- Given randomly chosen x, hard to find x' such that h(x)=h(x')
  - Attacker must find collision for a specific x. By contrast, to break collision resistance it is enough to find any collision.
  - Brute-force attack requires $O(2^n)$ time
- Weak collision resistance does not imply collision resistance.

Hashing vs. Encryption

- Hashing is one-way. There is no "un-hashing"
  - A ciphertext can be decrypted with a decryption key… hashes have no equivalent of "decryption"
- Hash(x) looks "random" but can be compared for equality with Hash(x')
  - Hash the same input twice → same hash value
  - Encrypt the same input twice → different ciphertexts
- Crytographic hashes are also known as "cryptographic checksums" or "message digests"

md

Application: Password Hashing

CIA

- Instead of user password, store hash(password)
- When user enters a password, compute its hash and compare with the entry in the password file

- Why is hashing better than encryption here?

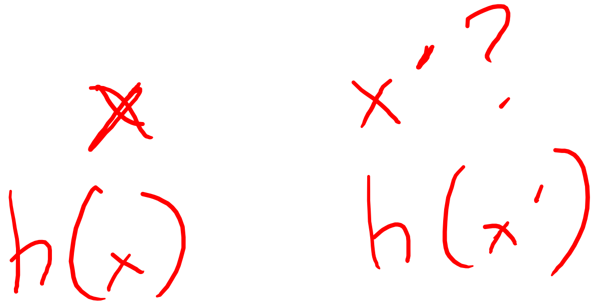
Application: Password Hashing

- Instead of user password, store hash(password)
- When user enters a password, compute its hash and compare with the entry in the password file
- Why is hashing better than encryption here?

- System does not store actual passwords!
- Don’t need to worry about where to store the key!
- Cannot go from hash to password!

Application: Password Hashing

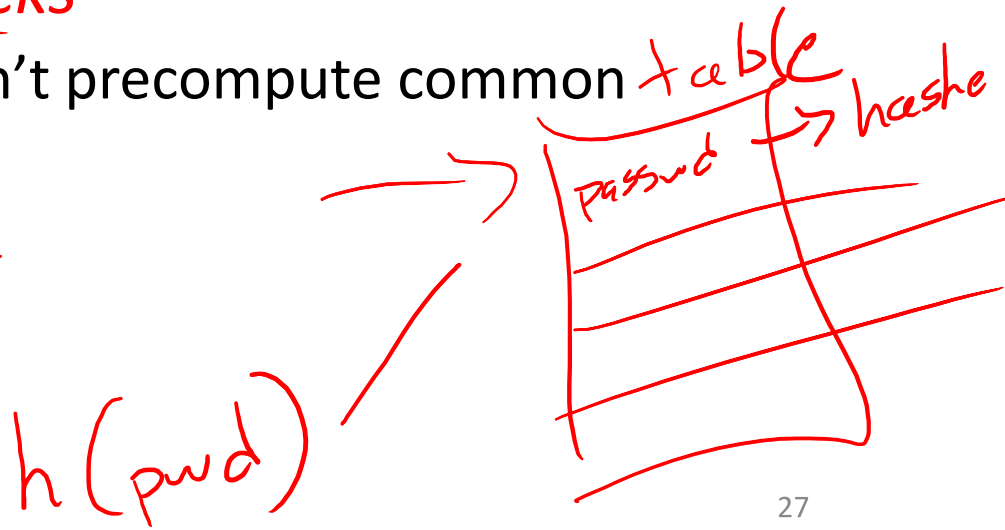
- Which property do we need?
  - One-wayness?
  - (At least weak) Collision resistance?
  - Both?



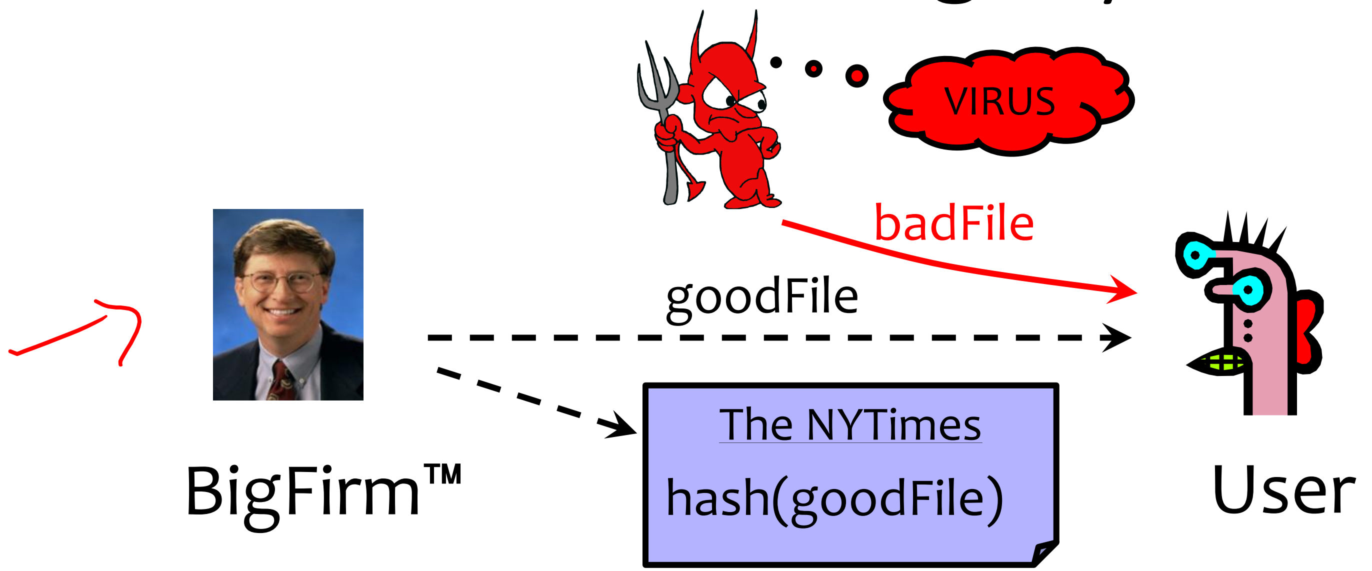
Application: Password Hashing + Salting

- Salting
  - We ‘salt’ hashes for password by adding a randomized suffix to the password
    - E.g. Hash(“coolpassword”+”35B67C2A”)
  - We then store the salt with the hashed password!

- The goal is to prevent *precomputation attacks*
  - If the adversary doesn’t know the salt, they can’t precompute common passwords



Application: Software Integrity



Goal: Software manufacturer wants to ensure file is received by users without modification.

Idea: given goodFile and hash(goodFile), very hard to find badFile such that hash(goodFile)=hash(badFile)

Application: Software Integrity

- Which property do we need?
  - One-wayness? ?
  - (At least weak) Collision resistance?
  - Both?

Which Property Do We Need?

One-wayness, Collision Resistance, Weak CR?

- UNIX passwords stored as hash(password)
  - **One-wayness:** hard to recover the/a valid password
- Integrity of software distribution
  - **Weak collision resistance**
  - But software images are not really random… may need **full collision resistance** if considering malicious developers

Which Property Do We Need?

- UNIX passwords stored as hash(password)
  - **One-wayness:** hard to recover the/a valid password
- Integrity of software distribution
  - **Weak collision resistance**
  - But software images are not really random… may need **full collision resistance** if considering malicious developers
- Commitments (e.g. auctions)
  - Alice wants to bid B, sends H(B), later reveals B
  - **One-wayness:** rival bidders should not recover B (this may mean that she needs to hash some randomness with B too)
  - **Collision resistance:** Alice should not be able to change her mind to bid B' such that H(B)=H(B')

Common Hash Functions

- MD5 – Don't Use!
  - 128-bit output
  - Designed by Ron Rivest, used very widely
  - Collision-resistance broken (summer of 2004)
- RIPEMD
  - 160-bit version is OK
  - 128-bit version is *not* good
- SHA-1 (Secure Hash Algorithm) – Don't Use!
  - 160-bit output
  - US government (NIST) standard as of 1993-95
  - Theoretically broken 2005; practical attack 2017!
- **SHA-2: SHA-256, SHA-512, SHA-224, SHA-384**
- SHA-3: standard released by NIST in August 2015

SHA-1 Broken in Practice (2017)

100,000x

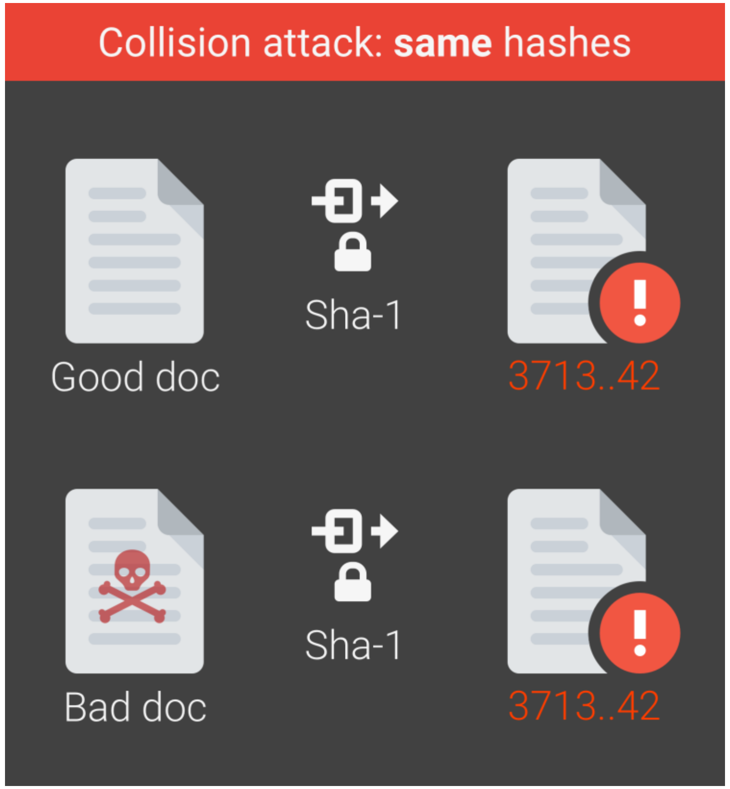
Google just cracked one of the building blocks of web encryption (but don't worry)

It's all over for SHA-1

by Russell Brandom | @russellbrandom | Feb 23, 2017, 11:49am EST

https://shattered.io

110 years GPU



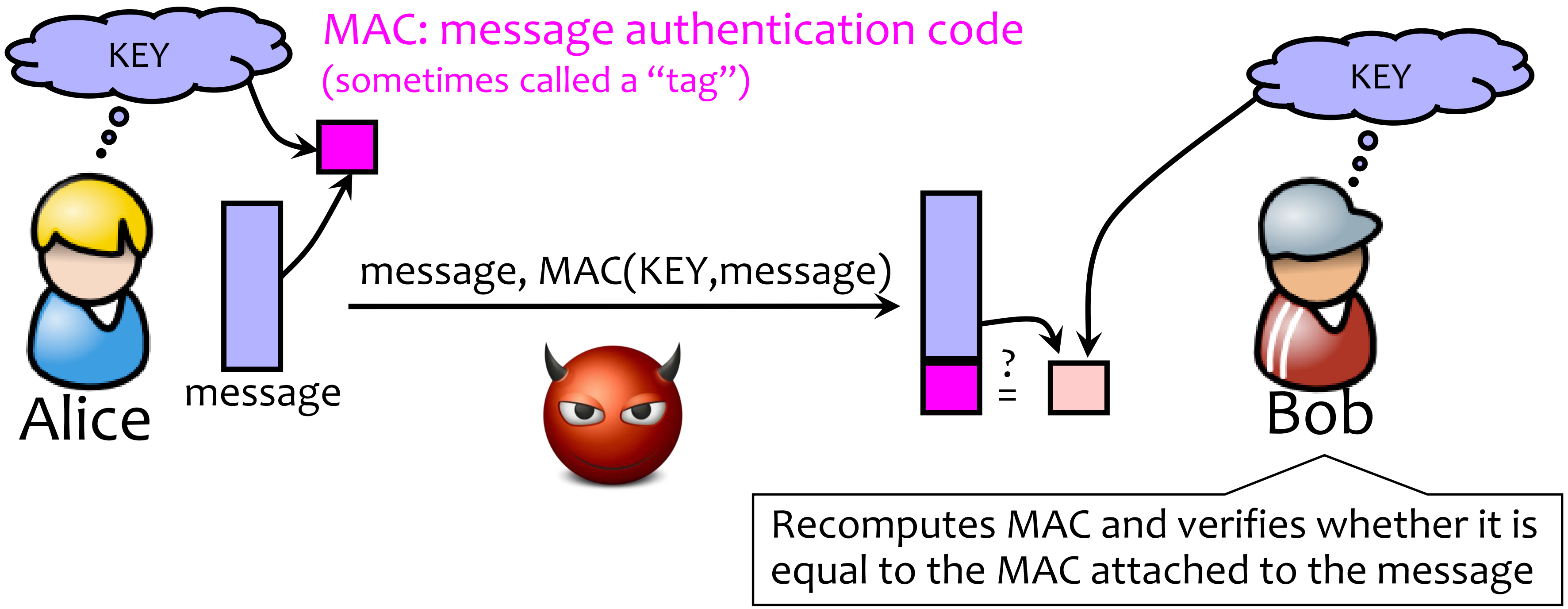
Aside: How we evaluate hash functions

sha-2
sha-3

- Speed
  - Is it amenable to hardware implementations?
- Diffusion
  - Does changing 1 bit in the input affect all output bits?
- Resistance to attack approaches
  - Collisions?
  - Length extensions?
  - etc

Recall: Achieving Integrity

Message authentication schemes: A tool for protecting integrity.



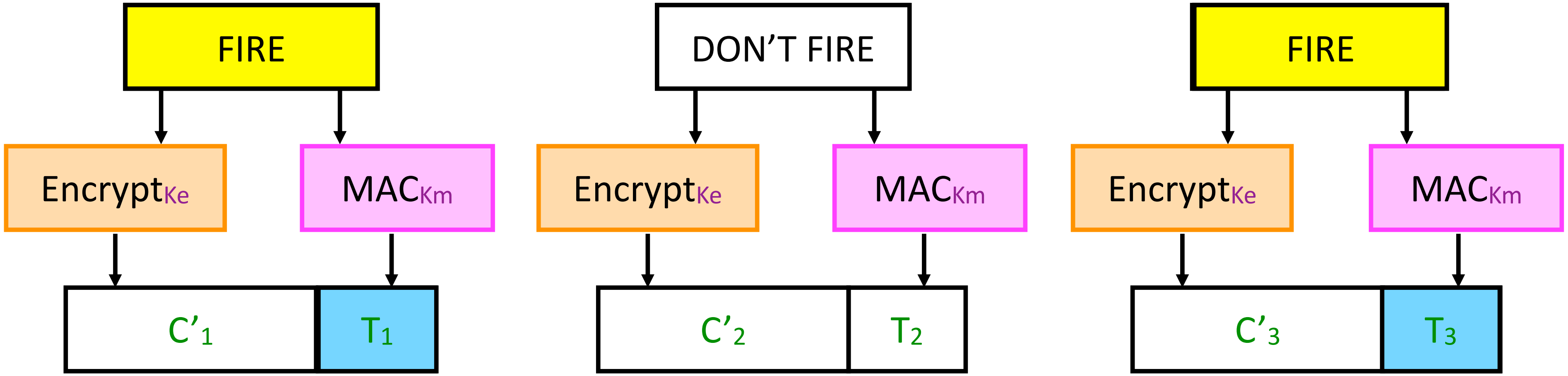
Integrity and authentication: only someone who knows KEY can compute correct MAC for a given message.

HMAC

- Construct MAC from a cryptographic hash function
  - Invented by Bellare, Canetti, and Krawczyk (1996)
  - Used in SSL/TLS, mandatory for IPsec
- Why not encryption?
  - Hashing is faster than block ciphers in software
  - Can easily replace one hash function with another
  - There used to be US export restrictions on encryption

Authenticated Encryption

- What if we want both privacy and integrity?
- Natural approach: combine encryption scheme and a MAC.
- But be careful!
  - Obvious approach: Encrypt-and-MAC
  - Problem: MAC is deterministic! same plaintext → same MAC

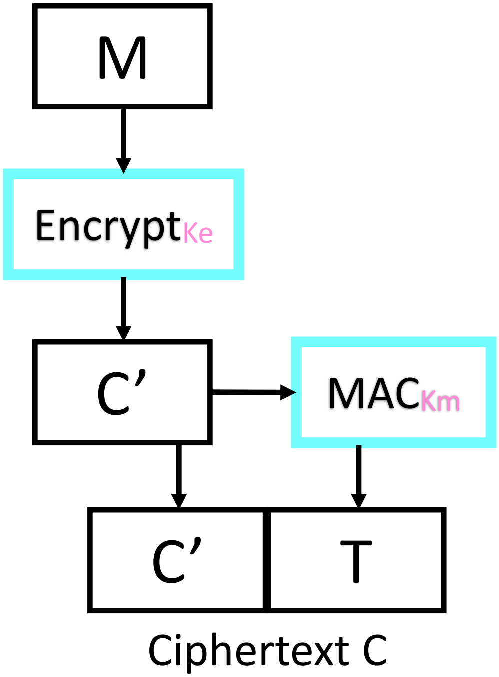


Authenticated Encryption

- Instead:

  Encrypt *then* MAC.

- (Not as good:
  MAC-then-Encrypt)



Encrypt-then-MAC