Birthday Problem

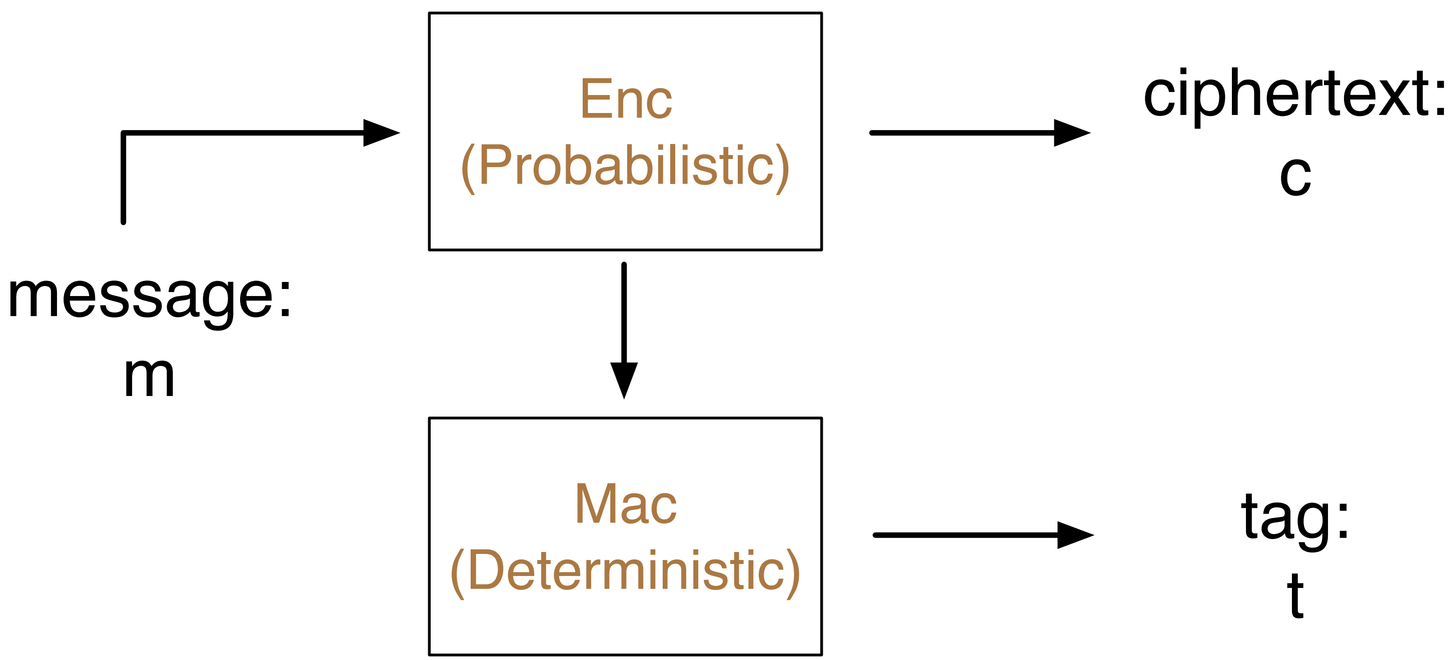
CS 5158/6058 Data Security and Privacy

Spring 2018

Instructor: Boyang Wang

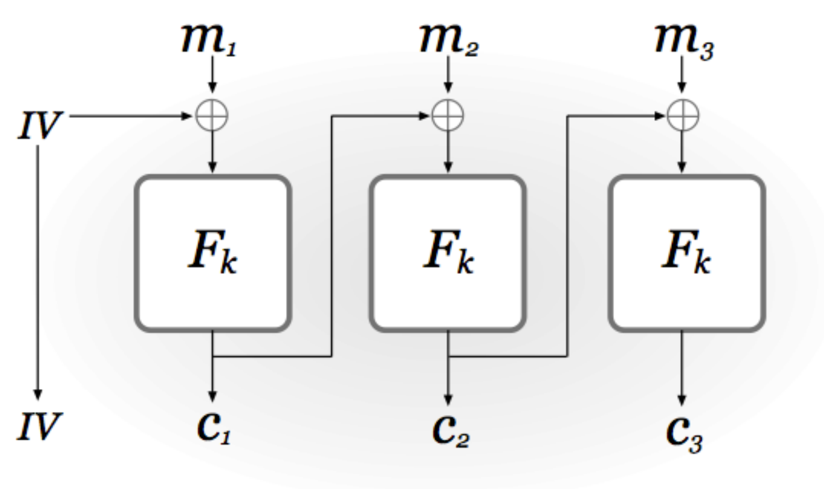
Encrypt-then-Authenticate

- Given Enc is probabilistic & Mac is deterministic
- Enc and Mac use two different keys
  - Output (c,t) is still probabilistic, still CPA-secure

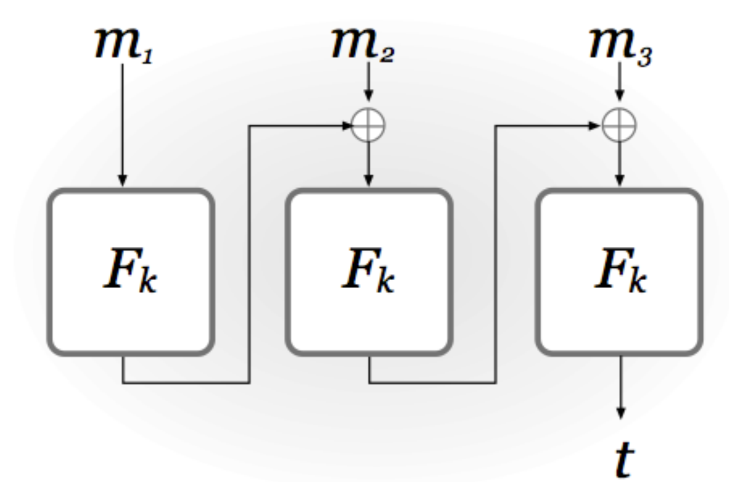


x	00	01	10	11
k=00,F(x)	11	00	01	10
k=01,F(x)	10	11	00	01
k=10,F(x)	01	10	11	00
k=11,F(x)	00	01	10	11

Example: given k_e=10, k_m=01
m=1011, and IV=01



Use Encrypt-then-authenticate
(CBC-Enc then CBC-MAC)



c=010000 and t=11
(or IV=01, c=0000, t=11)

Hash Function

- A mapping/function from an arbitrary long input to a fixed-length output (a digest or hash value)

- Hash Function: H(m) = h, deterministic
  - Collision: two inputs map to a same output
    - For x and y, x!=y, but H(x) == H(y)
    - Collision-resistance: collision must exist, but hard to find
- Crypto hash: MD5, SHA1, SHA2, etc.

Merkle-Damgard Transform

- MD transform: from fixed-length to arbitrary-length
  - If we have a hash function h: $\{0,1\}^{2n}$ —> $\{0,1\}^n$
  - then we can build a hash function H with arbitrary-length: $\{0,1\}^*$ —> $\{0,1\}^n$

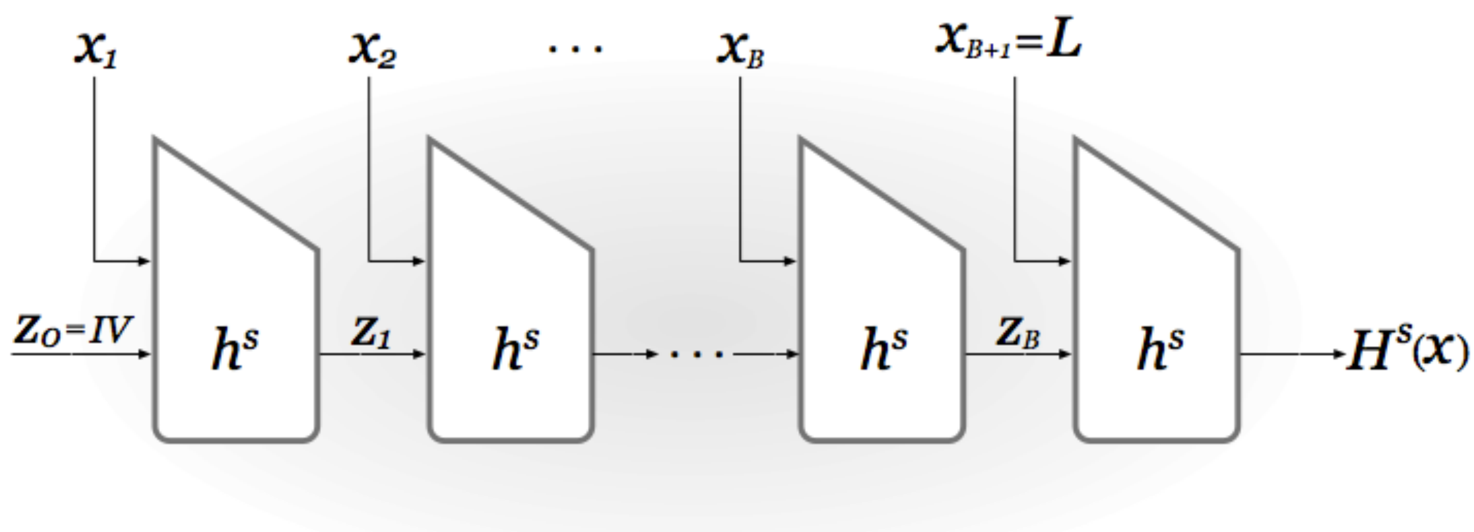


FIGURE 5.1: The Merkle–Damgård transform.

Merkle-Damgard Transform

- Arbitrary-length (KeyGen, Hash) is collision resistant if fixed-length (g, h) is collision resistant.

- Collision in Hash(m) could happen in two cases:
  - Messages are same, but lengths are different
    - m==m, L!=L', indicate h(z||L)==h(z||L')
  - Lengths are same, but messages are different
    - m!=m', L==L, indicate h(z||m)==h(z||m')
- h is a collision resistant, two cases both happen with negligible probability

Arbitrary-Length MAC

- Hash-and-MAC
  - Given hash function $H: \{0,1\}^* \longrightarrow \{0,1\}^n$
  - Given a fixed-length MAC for n-bit messages

  - Given a message m with arbitrary size
  - Compute $H(m) = h$, $Mac(h) = t$
  - t is the tag of message m
  - Security: finding a collision is hard, so generating a valid tag for a new message is hard.

Arbitrary-Length MAC

- Hash-and-MAC: H(m) = h, Mac(h) = t
  - Alice:
    - m="Cincinnati", h=H(m)=1011, t=Mac(h)=0101
    - Sends (m,t)
  - Attacker:
    - m'="Dayton", sends (m', t)
  - Bob:
    - m'="Dayton", h'=H(m')=0001, t'=Mac(h')=1110
    - t'!=t, then message is not correct/authenticate

Security of Hash Function

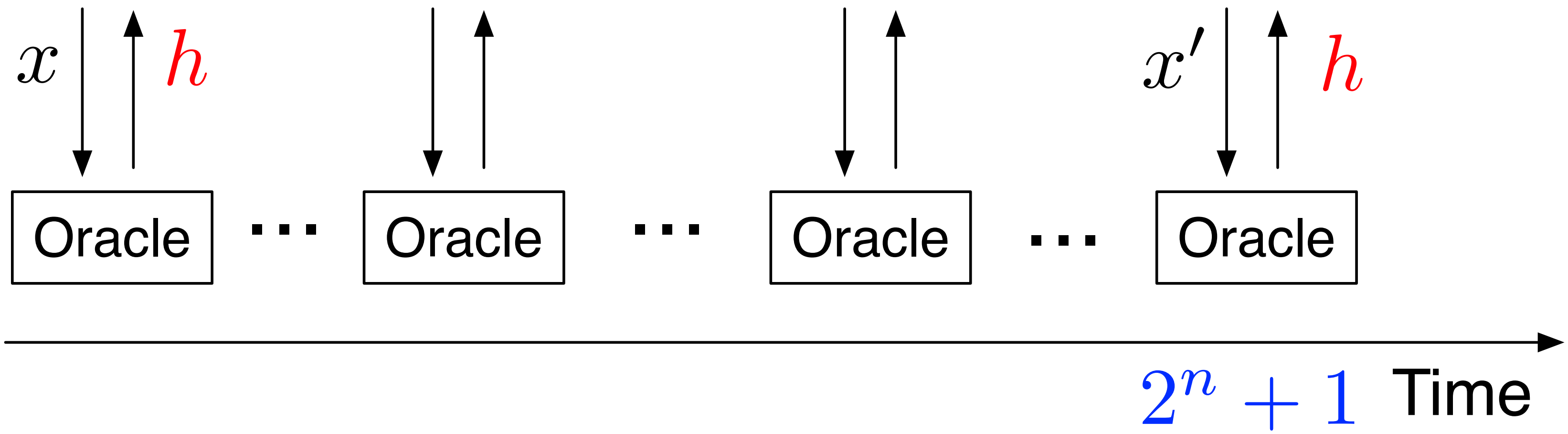
- The security of a n-bit hash function is (n/2)-bit
  - Find a collision after about $2^{n/2}$ tries with brute-force
    - Set S is empty
    - $x = 1$, $h_1 = H(x)$, add h_1 to S
    - $x = 2$, $h_2 = H(x)$, if h_2 not in S, add h_2 to S
    - ……
    - $x = m$, $h_m = H(x)$, if h_m not in S, add h_m to S
  - E.g., find a collision of SHA-1 (160-bit) with about 2^{80} tries

Security of Hash Function

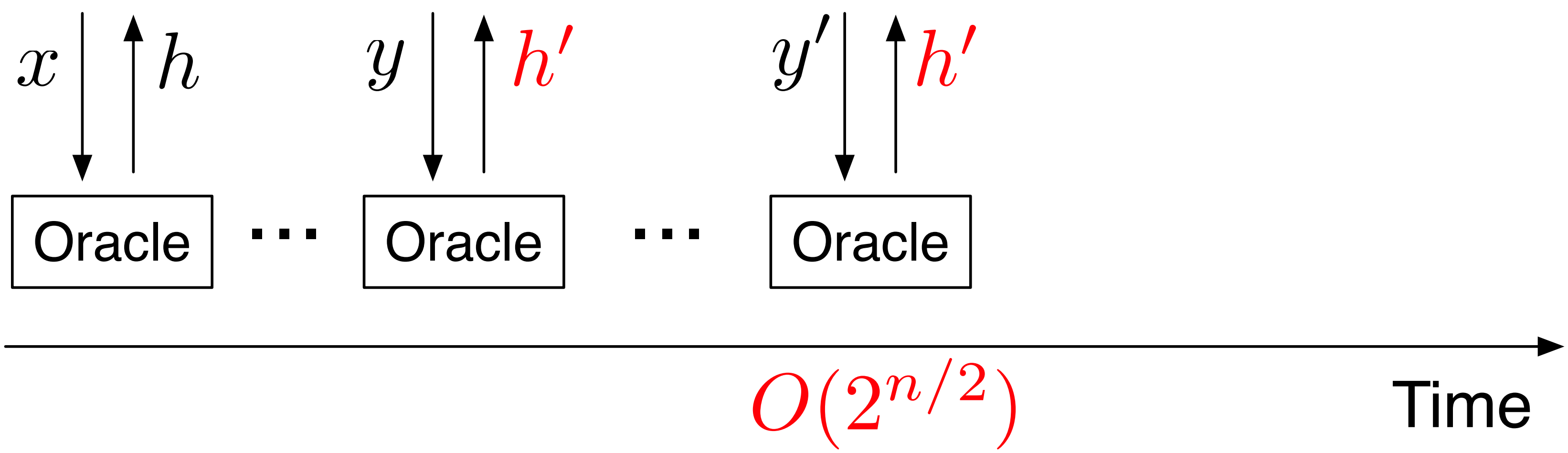
- The security of a n-bit hash function is (n/2)-bit
  - Find a collision after about $2^{n/2}$ tries with brute-force,
  - Wait…, it is not 2^n?

- Key: the two following problems are different
  - Problem 1: find a collision
  - Problem 2: given x, find a collision of x

Problem 2: Find a collision of x



Problem 1: There is a collision



Birthday Problem

- Assume 365 days per year, uniformly distribution
- There are n students in a room, p is the probability that there are two students have a same birthday
  - If $n=366$, then $p=100\%$
    - Each takes a different date, 366-th student will certainly have a same birthday with someone
  - What is n if $p=1/2$? Surprisingly, n is only 23.

Birthday Problem

- 1st student, probability of no same day is 1
- 2nd student, probability of no same day is 364/365
  - Choose any date, except 1st birthday
- 3rd student, probability of no same day is 363/365
  - Choose any date, except 1st and 2nd’s birthday
- 4th student, probability of no same day is 362/365
  - Choose any date, except 1st, 2nd & 3rd’ birthday

Birthday Problem

- 1st student, probability of no same day is 1
- 2nd student, probability of no same day is 364/365
- ……
- n-th, probability of no same day is (365 - (n-1))/365

- The probability of no same day for all n students is

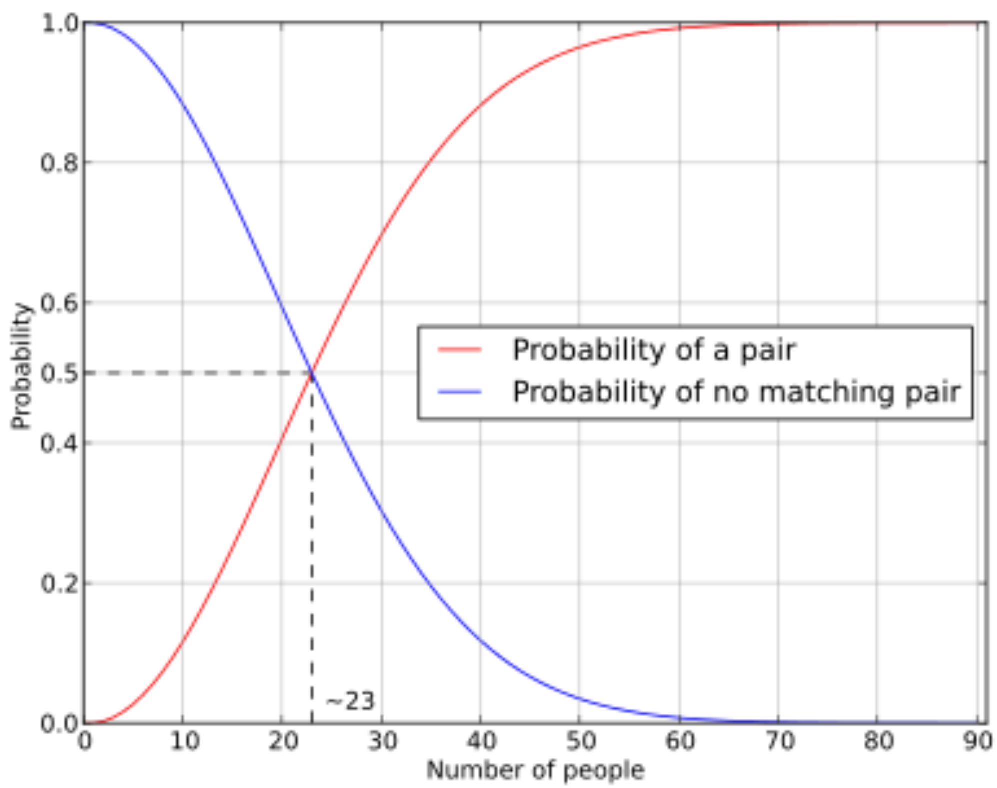
$$
\begin{aligned}
p' &= p'_1 \cdot p'_2 \cdot p'_3 \cdots p'_n \\
&= 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365 - (n-1)}{365} = \frac{365!}{365^n \cdot (365-n)!}
\end{aligned}
$$

Birthday Problem

- The probability that two having a same birthday is

$$p = 1 - p' = 1 - \frac{365!}{365^n \cdot (365-n)!}$$

- p: red curve
- p’: blue curve



Birthday Problem

- Taylor series $e^x \quad \approx \quad 1 + x \quad (\text{ if } x << 1)$
- Approximated probability of having a same birthday

$$p = 1 - p' \quad \approx \quad 1 - e^{\frac{-n^2}{2 \cdot 365}}$$
$$p(n, D) \quad \approx \quad 1 - e^{\frac{-n^2}{2D}}$$

- Example: what is the probability that two students in this class have a same birthday? (D=365, n=48, e=2.718)

$$
\begin{aligned}
p = 1 - p' &\approx 1 - e^{\frac{-n^2}{2\cdot 365}} \\
p(n, D) &\approx 1 - e^{\frac{-n^2}{2D}}
\end{aligned}
$$

- <u>Example</u>: what is the probability that two students in this class have a same birthday? (D=365, n=48, e=2.718)

  - n*n = 48*48 = 2304
  - 2*D = 2*365 = 730
  - -(2304/730) = -3.156
  - $e^{-3.156} = 0.043$
  - p = 1 - 0.043 = 0.957

$$
\begin{aligned}
p = 1 - p' &\approx 1 - e^{\frac{-n^2}{2\cdot 365}} \\
p(n, D) &\approx 1 - e^{\frac{-n^2}{2D}}
\end{aligned}
$$

- Practice: we have 20 students and 200 classes, each student (uniformly) chooses one class, what is the probability that two students choose a same class? (D=200, n=20, e=2.718)
    - n*n = 20*20 = 400
    - 2*D = 2*200 = 400
    - -(400/400) = -1
    - e^{-1} = 0.368
    - p = 1 - 0.368 = 0.632

$$p = 1 - p' \quad \approx \quad 1 - e^{\frac{-n^2}{2 \cdot 365}}$$

$$p(n, D) \quad \approx \quad 1 - e^{\frac{-n^2}{2D}}$$

- Brute-force Attack on Hash Function:
    - There is a collision s.t., x!=y, but H(x)==H(y)

- H() has $D=2^{160}$ outputs, if we try this H() n times, where each time we use a different input, p is the probability that there are two times having a same output ($D=2^{160}$, n, e=2.718, p)
    - If p=1/2, then n is about $O(2^{160/2})$
    - 160-bit hash provides 80-bit security

Problem 2

- There are n students, given Alice's birthday (e.g., Jan. 1st), the probability that there is another student having a same birthday as Alice

- 1st student is Alice
  - Alice's birthday is Jan. 1st

- 2nd student, probability of not the same is 364/365
  - Choose any date, except Jan. 1st

Problem 2

- 1st student is Alice, probability is 1
  - Alice’s birthday is Jan. 1st
- 2nd student, probability of not the same is 364/365
  - Choose any date, except Jan. 1st

- 3rd student, probability of not the same is 364/365
  - Choose any date, except Jan. 1st

- 4th student, probability of not the same is 364/365
  - Choose any date, except Jan. 1st

Problem 2

- 1st student is Alice, probability is 1
- 2nd student, probability of not the same is 364/365
- ……
- n-th student, probability of not the same is 364/365

- The probability that there is no student having a same birthday as Alice

$$
\begin{aligned}
p' &= p'_1 \cdot p'_2 \cdot p'_3 \cdots p'_n \\
&= 1 \cdot \frac{364}{365} \cdot \frac{364}{365} \cdots \frac{364}{365} = (\frac{364}{365})^{n-1}
\end{aligned}
$$

Problem 2

- The overall probability that there is another student having a same birthday as Alice

$$p = 1 - p' = 1 - (\frac{364}{365})^{n-1}$$

- Example: there are n=48 students, given Alice's birthday (e.g., Jan. 1st), what is the probability that there is another student having a same birthday as Alice? (D=365, n=48)
  - $(364/365)^{47} = 0.879$, $p = 1 - 0.879 = 0.121$

$$p = 1 - p' = 1 - (\frac{364}{365})^{n-1}$$

- Practice: there are n=20 students and D=200 classes, each student (uniformly) chooses one class. Given Alice chooses CS5158, what is the probability that there is another student chooses CS5158?

- $(199/200)^{19} = 0.909$, $p = 1 - 0.909 = 0.091$

- Problem 1: There are n students in a room, p is the probability that there are two students having a same birthday

- Problem 2: There are n students, given Alice's birthday (e.g., Jan. 1st), the probability that there is another student having a same birthday as Alice

	Problem 1	**Problem 2**
n=48, D=365	p = 0.957	p = 0.121

Public-Key Cryptography

CS 5158/6058 Data Security and Privacy

Spring 2018

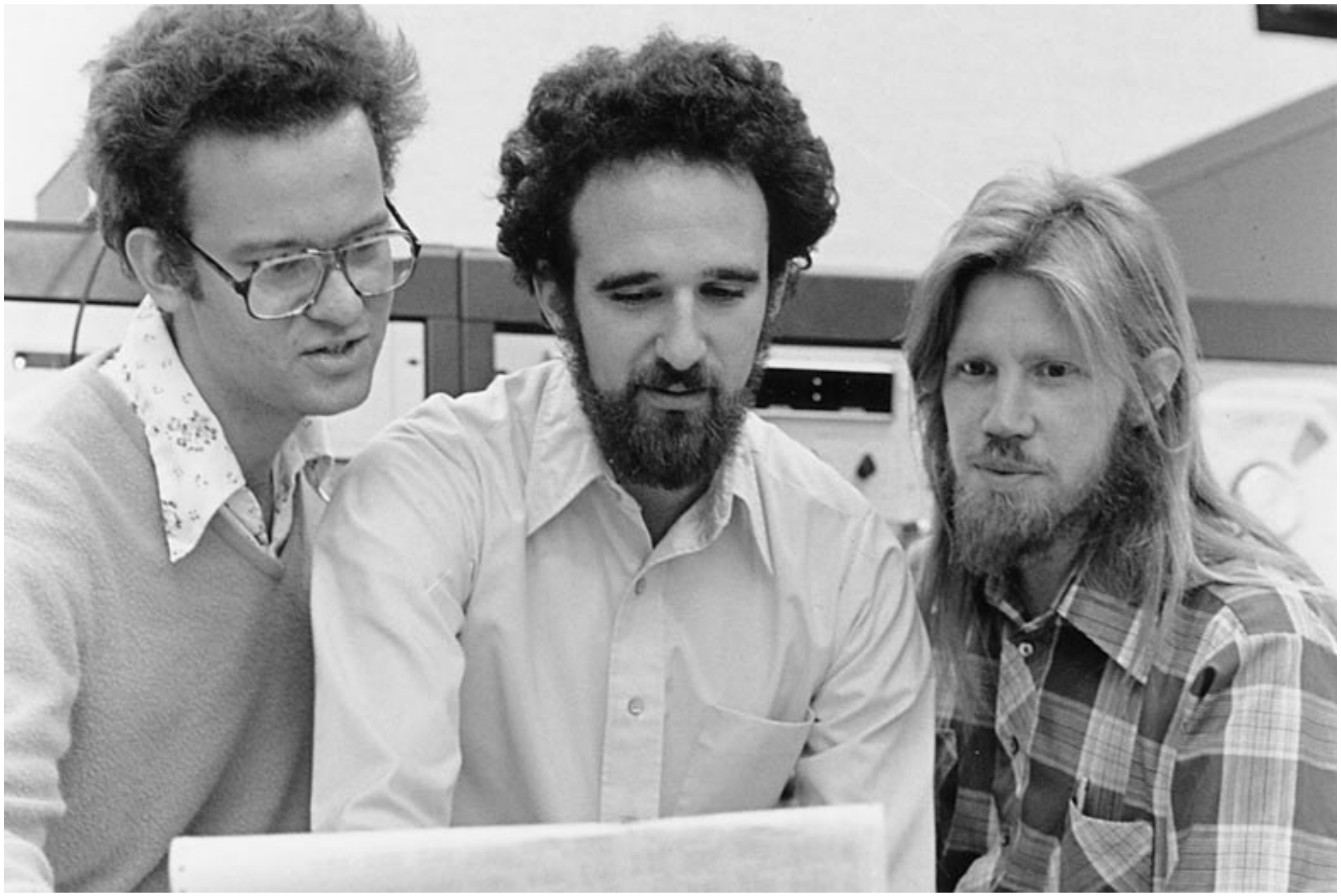
Instructor: Boyang Wang

Public-Key Revolution

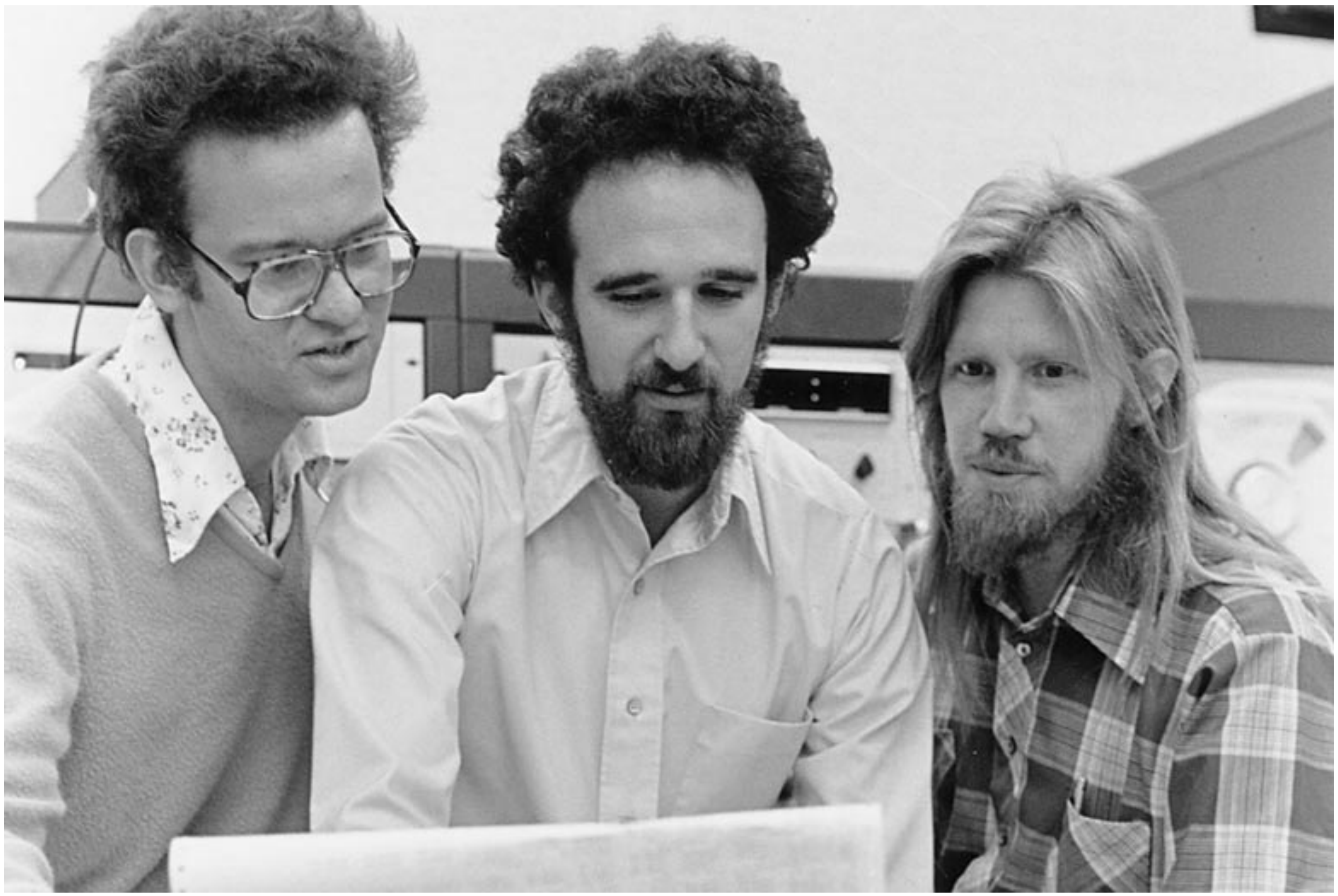
- Symmetric-Key Crypto:
  - Block Cipher and MAC
  - Need to share a private key in advance
  - Limit usage in practice: mainly military

- Public-Key Crypto (Diffie and Hellman,1976)
  - "*New Direction in Cryptography*"
  - Idea: No need to share a private key
  - Expend crypto usage to almost everywhere

- Whitfield Diffie (right) and Martin Hellman (middle), ACM Turning award in 2015



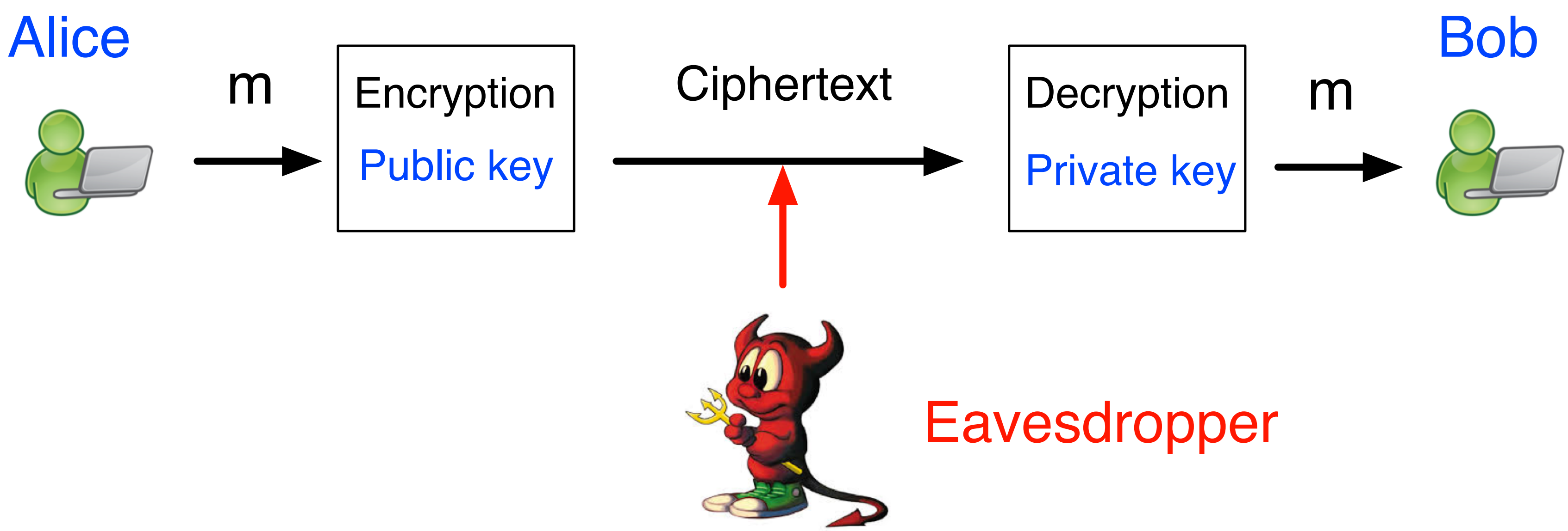
- Ralph Merkle (left)
  - Merkle-Damgard Transform, Merkle hash tree



Public-Key Revolution

- A key includes two parts,
  - <u>a public</u> (pk) & <u>a private</u> (sk)
  - Send public key on public channel for encryption
  - Use private key to decrypt
- Based on some hard problem (one-way function)
  - Easy to compute (with public key),
  - But hard to decrypt (without private key)
- Data Privacy: Public-Key Encryption
- Data Integrity: Signature

Public-Key Encryption



- Alice obtains Bob' public key from <u>public channels</u>
- Alice encrypts with public key
- Bob decrypts with private key

Number Theory

Mathematical Foundation of Public-Key Crypto

Divisor and Factor

- Z: the set of integers, {…, -2, -1, 0, 1, 2, 3, …}

- For a, b in Z, a divides b, written a | b
  - i.e., there is an integer c, s. t. ac = b
- E.g., a = 6, b = 12, a divides b
  - c=2, ac = 2*6 = 12 = b

- a | b and a is positive, a is divisor of b
  - In addition, if a is not 1 or b, a is factor of b

Prime and Composite

- $a \mid b$ and a is positive, a is divisor of b
  - In addition, if a is not 1 or b, a is factor
  - E.g., b = 12
  - divisors: 1, 2, 3, 4, 6, 12
  - factors: 2, 3, 4, 6

- A positive integer $p>1$ is a prime, if p has no factors, i.e., only two divisors, 1 and p; otherwise p is a composite

Prime and Composite

- A positive integer p>1 is a prime, if p has no factors, i.e., only two divisors, 1 and p; otherwise p is a composite
- Practice: Prime or Composite? 2, 3, 4, 5, 6,
  - 2: {1,2}, no factors, prime
  - 3: {1,3}, no factors, prime
  - 4: {1,2,4}, 1 factor, composite
  - 5: {1,5}, no factors, prime
  - 6: {1,2,3,6}, 2 factors, composite

Additional Reading

Chapter 5 & 8, *Introduction to Modern Cryptography, Drs. J. Katz and Y. Lindell, 2nd edition*