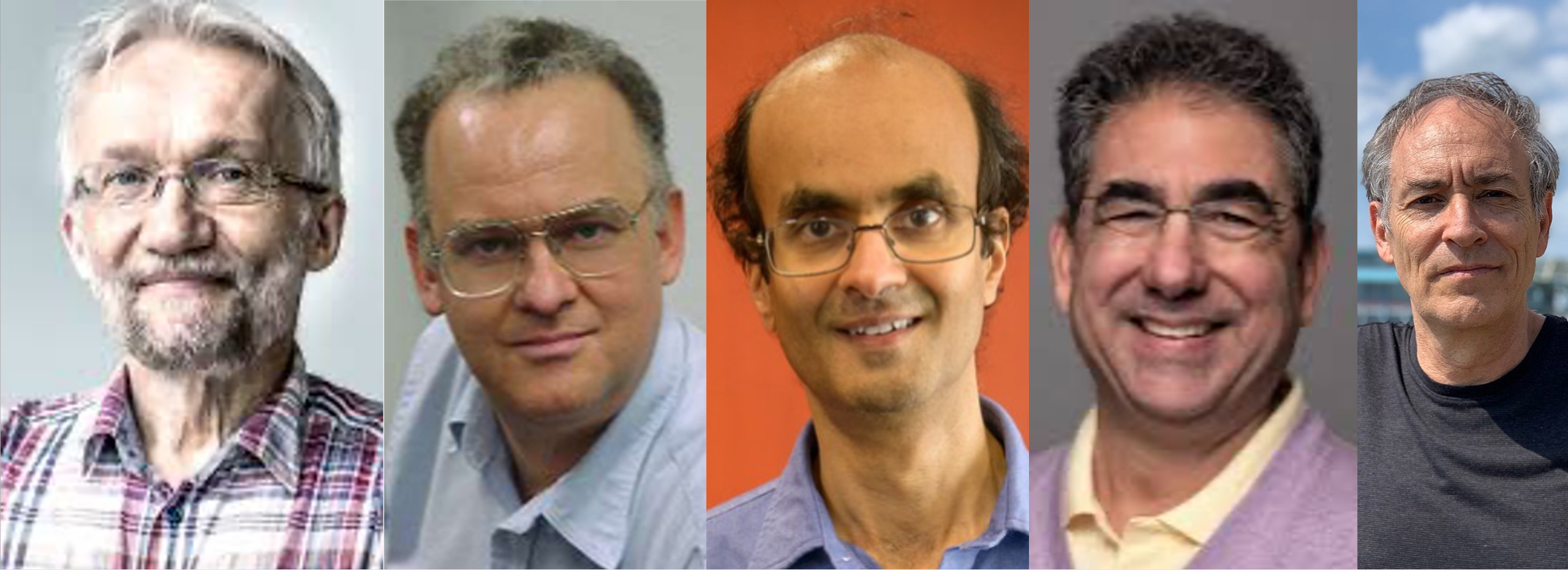
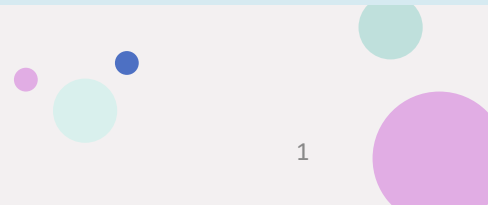




Lecture 12

Hash Functions

CS409/6011: Introduction to Cryptography
Sruthi Sekar



Today

RECALL

Cryptographic Hash Function

$H: \mathcal{M} \rightarrow \mathcal{T}$ where $|\mathcal{M}| \gg |\mathcal{T}|$

Collision for H A pair of messages $m_0, m_1 \in \mathcal{M}$ such that $H(m_0) = H(m_1)$ and $m_0 \neq m_1$

Collision Resistance

H is collision-resistant if, for all (explicit) efficient algorithms $\mathcal{A}$
 $\text{Adv}_{\text{CR}}[\mathcal{A}, H] = \Pr[\mathcal{A} \text{ outputs collision for } H] \leq \text{negligible}$

Example SHA-256 takes arbitrary long inputs and outputs 256 bits



$|\mathcal{M}| \gg |\mathcal{T}| \Rightarrow$ there will be a lot of collisions

Example of a Compression Function

Davies-Meyer Construction

F be a block cipher with n -bit keys and n -bit input length



$h: \{0,1\}^{2n} \rightarrow \{0,1\}^n$

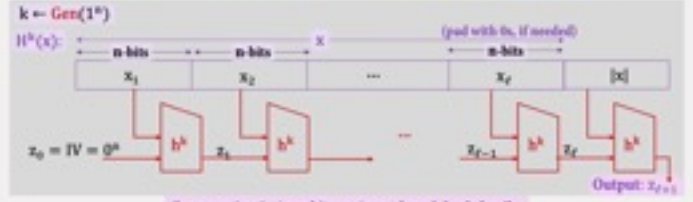
- There is no known proof that Davies-Meyer is collision resistant, just assuming an SPRP.
- To prove its security, need to assume a strong assumption about F ("ideal cipher").
- Hard to instantiate F (any block cipher cannot be plugged in naively!) (E.g. MD5, SHA-1, SHA-2 all use Davies Meyer + Merkle Damgård Transform)

Merkle Damgård Transform

(Gen, h) : a compression function with input length $2n$ and output length n

$k \leftarrow \text{Gen}(1^n)$

$H^k(x)$:



$z_0 = \text{IV} = 0^n$

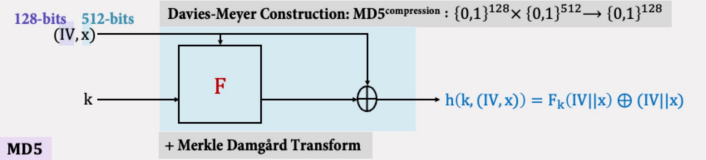
Output: z_{L+1}

Construction 1: An arbitrary input-length hash family

Theorem 1 [Katz and Lindell (Third edition), Theorem 6.4]
If (Gen, h) is collision resistant, then so is (Gen, H) .

MD5: Construction and Attack

Davies-Meyer Construction: MD5^{compression} : $\{0,1\}^{128} \times \{0,1\}^{512} \rightarrow \{0,1\}^{128}$



MD5 + Merkle Damgård Transform

[Wang et al., 2004]'s attack:

- For any fixed IV, they find x and y such that $\text{MD5}^{\text{compression}}(\text{IV}, x) = \text{MD5}^{\text{compression}}(\text{IV}, y)$ in 2^{39} steps. Hence, choosing this IV, $\text{MD5}(x) = \text{MD5}(y)$. But attacker here does not have control over x, y .
- How to get collisions on inputs of my choice? (this is what I showed demo for in last class with two postscript files of my choice that had the same MD5 hash!)
- Key Idea to do this:** For the fixed IV that gives $\text{MD5}(x) = \text{MD5}(y)$ for same length x and y , it is also true that $\text{MD5}(x||z) = \text{MD5}(y||z)$ for any message z !

RECALL

Cryptographic Hash Function

$$H: \mathcal{M} \rightarrow \mathcal{T} \text{ where } |\mathcal{M}| \gg |\mathcal{T}|$$

Collision for H

A pair of messages $m_0, m_1 \in \mathcal{M}$ such that $H(m_0) = H(m_1)$ and $m_0 \neq m_1$

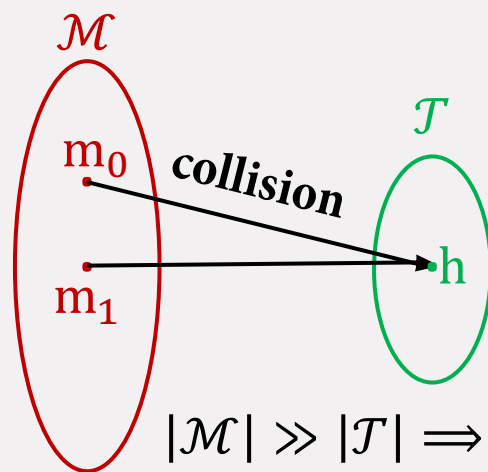


Collision Resistance

H is collision-resistant if, for all (explicit) efficient algorithms $\mathcal{A}$

$$\text{Adv}_{\text{CR}}[\mathcal{A}, H] = \Pr[\mathcal{A} \text{ outputs collision for } H] \leq \text{negligible}$$

Example: **SHA-256** takes arbitrary long inputs and outputs 256 bits



$|\mathcal{M}| \gg |\mathcal{T}| \Rightarrow$ there will be lot of collisions

RECALL

Hash Functions: Theory vs Practice

Theoretical Constructions

$$\{H_k\}_{k \in \{0,1\}^n}$$

It is typically a keyed family of functions.
A Gen algorithm picks one k

There is a security parameter n

Security proved under well-studied assumptions

If you have a family of functions, you do not have the space to hardcode collisions for all functions!

Practical Constructions

E.g.: SHA3, SHA256, SHA1, MD5, MD4

Single function and not a family of functions
("unkeyed")

No security parameter knob!

Use until some finds a collision!

Not secure, theoretically! Hardcode the collision into algorithm A, it runs in constant time!

Secure: no
collision found

Broken:
collision found

RECALL

Collision Resistant Hash Functions

$\Pi = (\text{Gen}, H)$: $\text{Gen}(1^n) \rightarrow k$; H^k takes input $x \in \{0,1\}^*$ and outputs $H^k(x) \in \{0,1\}^{\ell(n)}$

If H^k only defined for $\{0,1\}^{\ell'(n)}$ for $\ell'(n) > \ell(n)$ called a **compression function**.



Challenger

$k \leftarrow \text{Gen}(1^n)$

Can define collision resistance similarly for a single hash function as well.

k



Collin

wins

x, x'

Collin outputs a collision

$\text{HashColl}_{\text{Collin}, \Pi}(n)$:

Outputs 1, if and only if $H^k(x) = H^k(x')$ and $x \neq x'$

Definition 1

A hash function $\Pi = (\text{Gen}, H)$ is **collision resistant** if, for all **PPT** adversary **Collin**, there exists a negligible function **negl** such that, $\forall n$

$$\Pr[\text{HashColl}_{\text{Collin}, \Pi}(n) = 1] \leq \text{negl}(n)$$

Collision Resistant Hash Functions

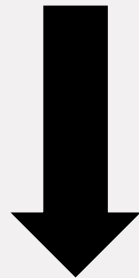
$\Pi = (\text{Gen}, H): \text{Gen}(1^n) \rightarrow k; H^k$ takes input $x \in \{0,1\}^*$ and outputs $H^k(x) \in \{0,1\}^{\ell(n)}$

If H^k only defined for $\{0,1\}^{\ell'(n)}$ for $\ell'(n) > \ell(n)$ called a **compression function**.

How do we build such CRHFs?

Step 1: Build a compression function (will give an example in the end)

(Gen, h) : a compression function with input length $2n$ and output length n



Merkle-Damgård Transform

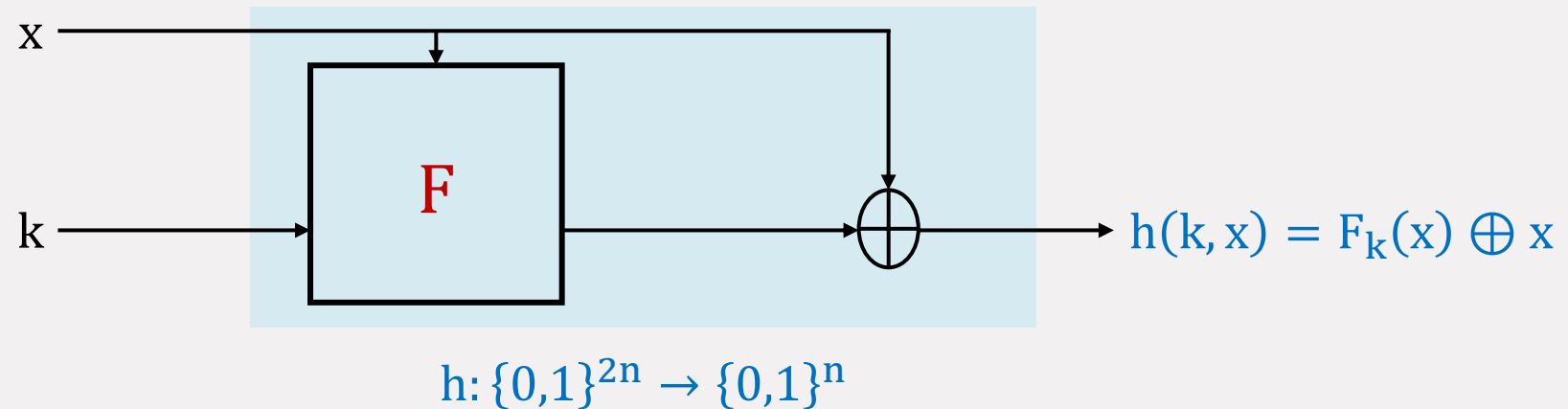
Step 2: Build an arbitrary input-length hash family using this.

(Gen, H) : a hash family with arbitrary input length and output length n

Example of a Compression Function

Davies-Meyer Construction

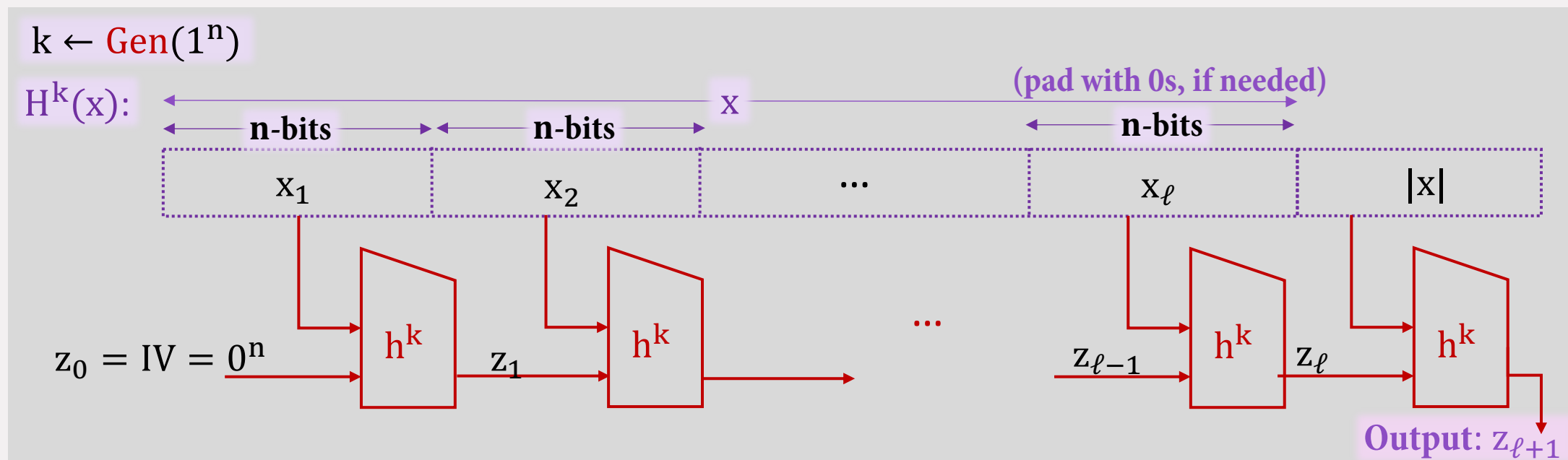
F be a block cipher with **n-bit keys** and **n-bit input length**



- There is no known proof that Davies-Meyer is collision resistant, just assuming an SPRP.
- To prove its security, need to assume a strong assumption about F (“*ideal cipher*”).
- Hard to instantiate F (any block cipher cannot be plugged in naively!)
(E.g MD5, SHA-1, SHA-2 all use Davies Meyer + Merkle Damgård Transform)

Merkle Damgård Transform

(Gen, h) : a compression function with input length $2n$ and output length n



Construction 1: An arbitrary input-length hash family

Theorem 1



[Katz and Lindell (Third edition), Theorem 6.4]

If (Gen, h) is collision resistant, then so is (Gen, H) .

Merkle Damgård Transform : Proof Idea

Theorem 1

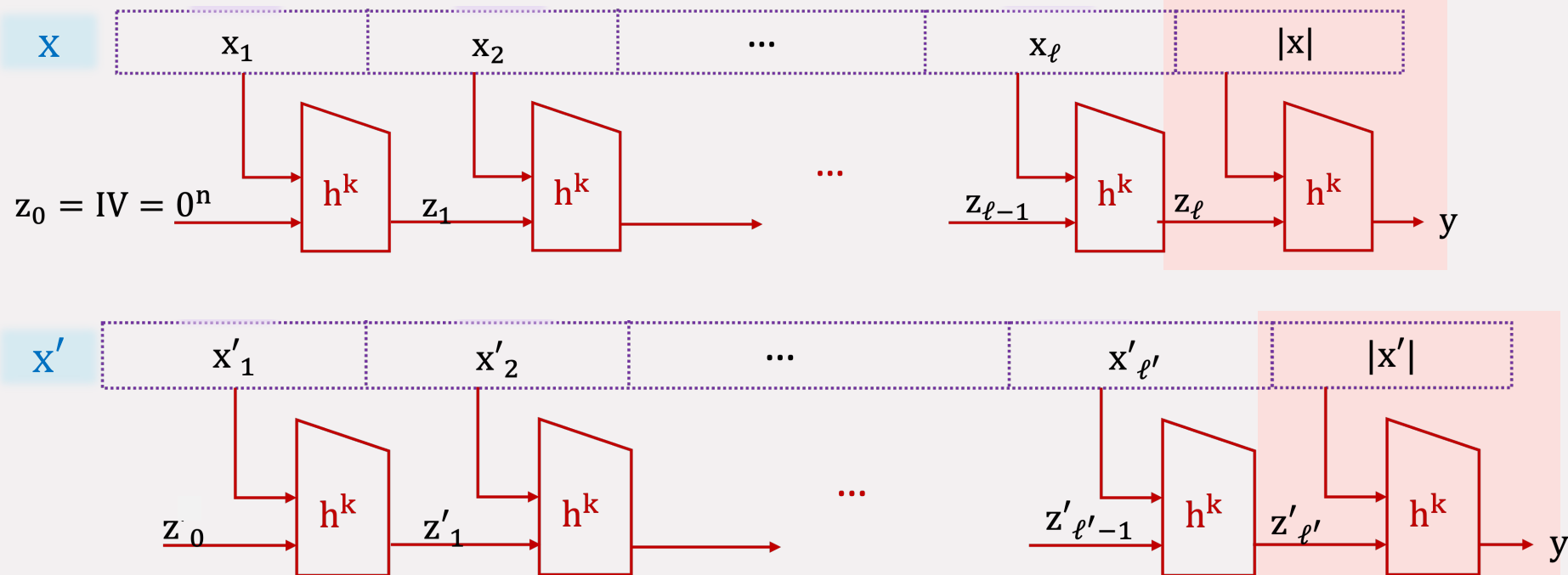


[Katz and Lindell (Third edition), Theorem 6.4]

If (Gen, h) is collision resistant, then so is (Gen, H) .

Proof Sketch: Assume to contrary that $\exists$ a collision x, x' for H^k , we show a collision for h^k

Case 1: x, x' are of different lengths ($|x| \neq |x'|$) $h^k(z_\ell || |x|) = h^k(z'_{\ell'} || |x'|) = y$
the collision for h^k



Merkle Damgård Transform : Proof Idea

Theorem 1



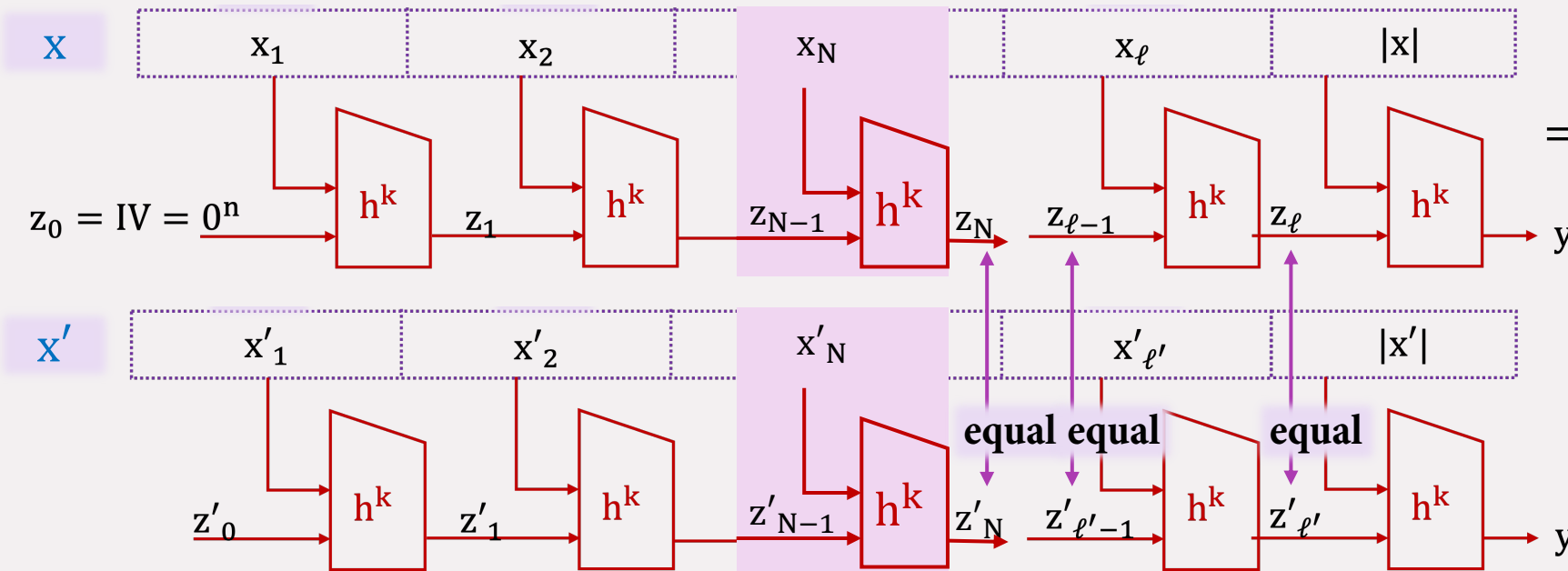
[Katz and Lindell (Third edition), Theorem 6.4]

If (Gen, h) is collision resistant, then so is (Gen, H) .

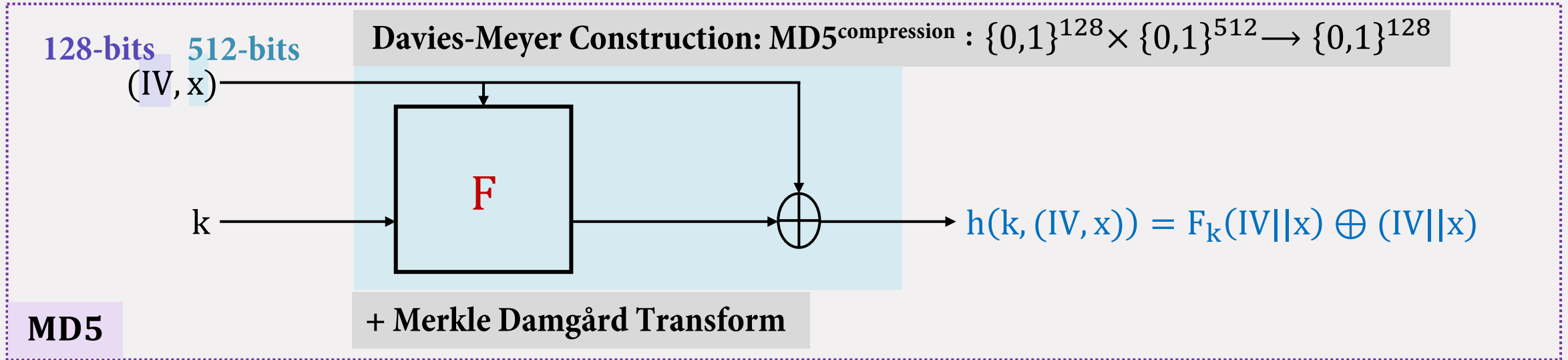
Proof Sketch: Assume to contrary that $\exists$ a collision x, x' for H^k , we show a collision for h^k

Case 1: x, x' are of different lengths ($|x| \neq |x'|$) $h^k(z_\ell || |x|) = h^k(z'_{\ell'} || |x'|) = y$

Case 2: x, x' are of same length ($|x| = |x'|$) $i = N$ be the largest index for which $z_{i-1} || x_i \neq z'_{i-1} || x'_i$
 (Since $x \neq x'$ and $|x| = |x'|$ such an N exists)
 $\Rightarrow z_N || x_{N+1} = z'_N || x'_{N+1}$



MD5: Construction and Attack



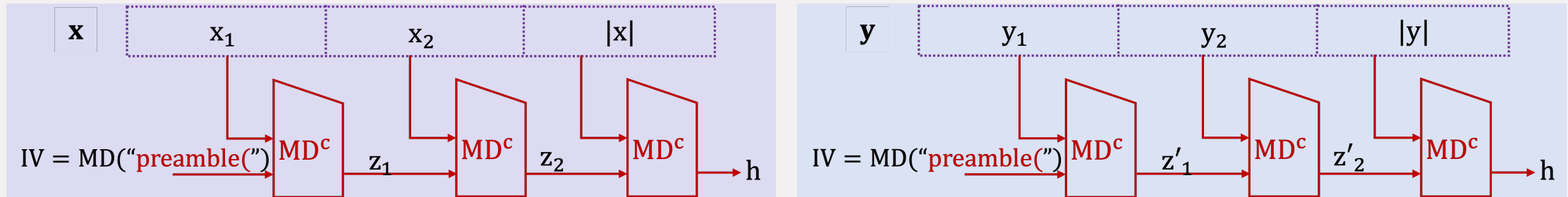
[Wang et al., 2004]’s attack:

- For any fixed IV, they find x and y such that $\text{MD5}^{\text{compression}}(IV, x) = \text{MD5}^{\text{compression}}(IV, y)$ in 2^{39} steps. Hence, choosing this IV, $\text{MD5}(x) = \text{MD5}(y)$. But attacker here does not have control over x, y .
- How to get collisions on inputs of my choice? (this is what I showed demo for in last class with two postscript files of my choice that had the same MD5 hash!)
- **Key Idea to do this:** For the fixed IV that gives $\text{MD5}(x) = \text{MD5}(y)$ for same length x and y , it is also true that **$\text{MD5}(x||z) = \text{MD5}(y||z)$ for any message z !**

MD5: Construction and Attack

Poisoned Message Attack [Daum, Luck, 2005] for MD5-Hash-and-MAC

- Start with $IV = MD5(\text{"preamble"})$, for a postscript preamble "preamble".
- Use [Wang et al., 2004]'s attack for this IV to find x, y such that $MD5(x) = MD5(y) = h$



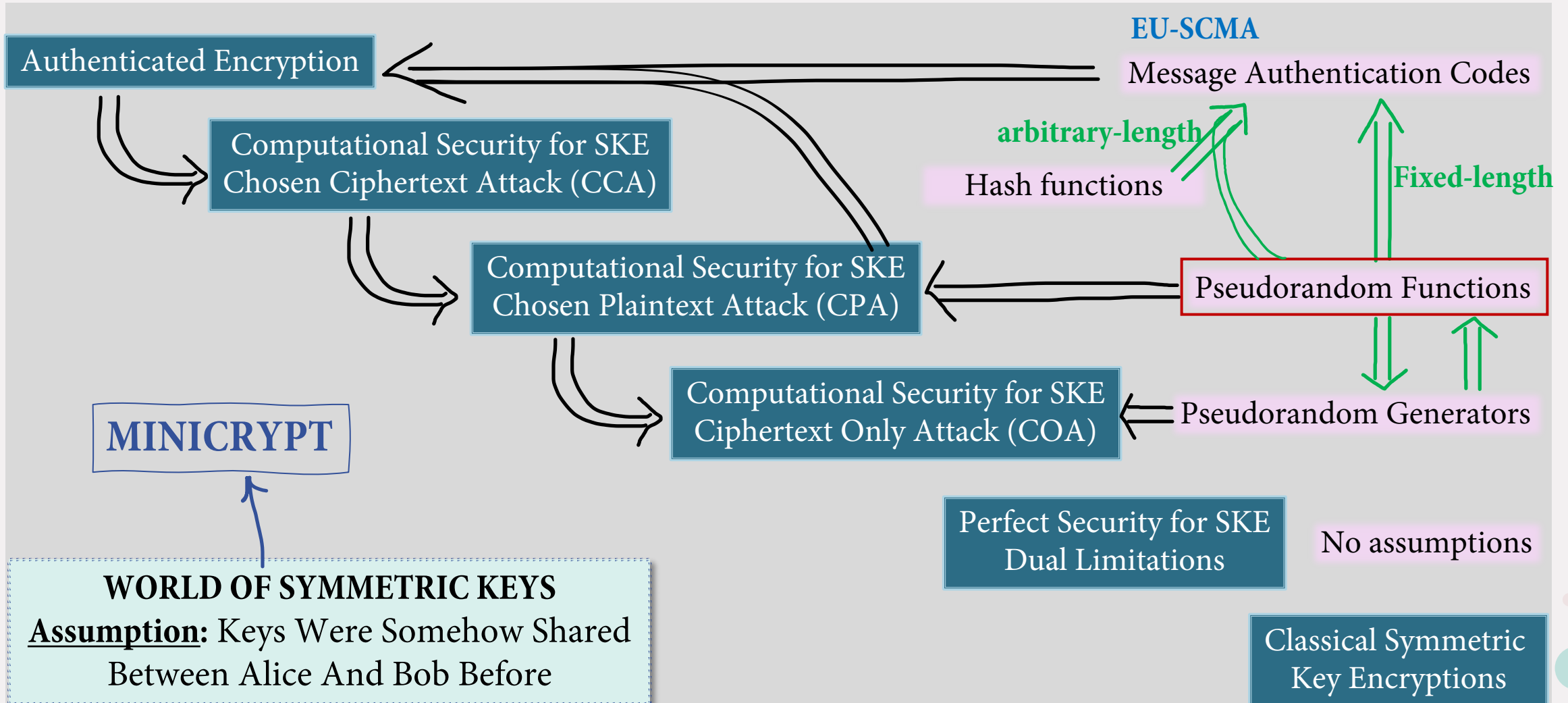
- Select your choice of postscript documents $Doc_1.ps$ and $Doc_2.ps$ (executed by postscript commands T_1 and T_2 respectively).
- Construct the messages:

$$Doc'_1 = \text{preamble}(x)(x)\text{eq}T_1T_2$$

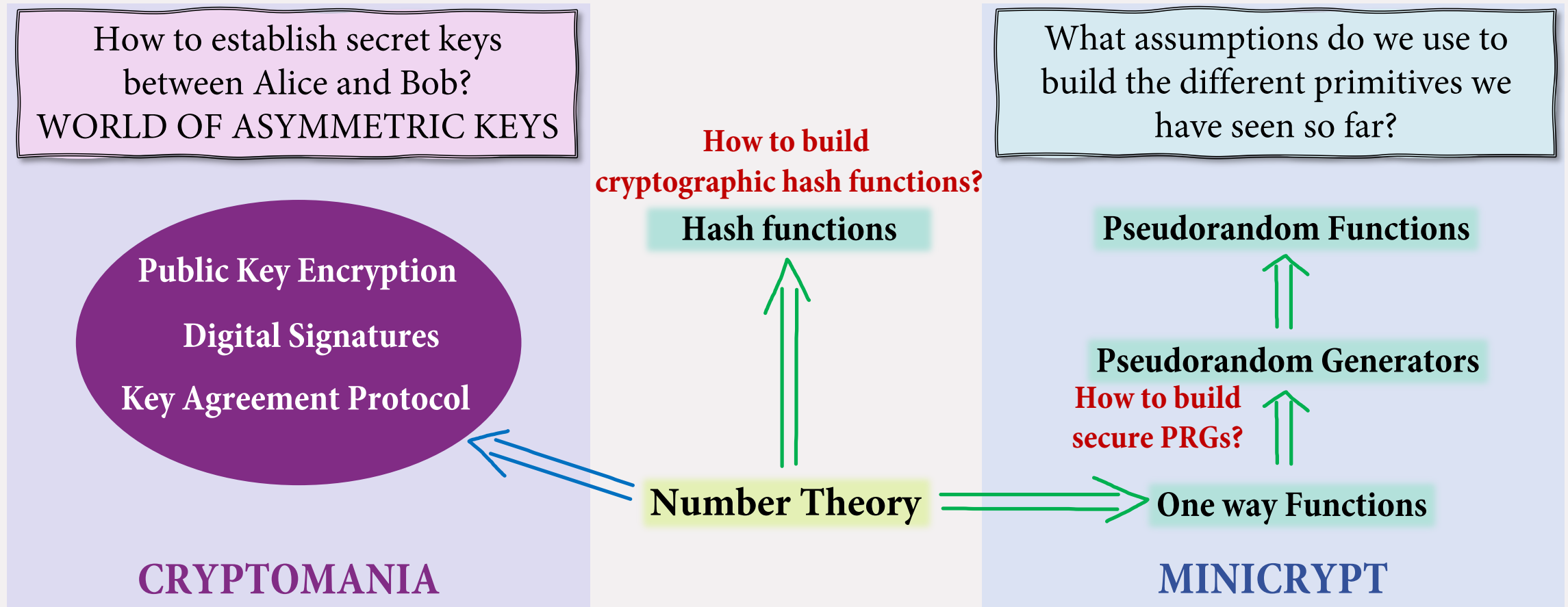
$$Doc'_2 = \text{preamble}(y)(x)\text{eq}T_1T_2$$

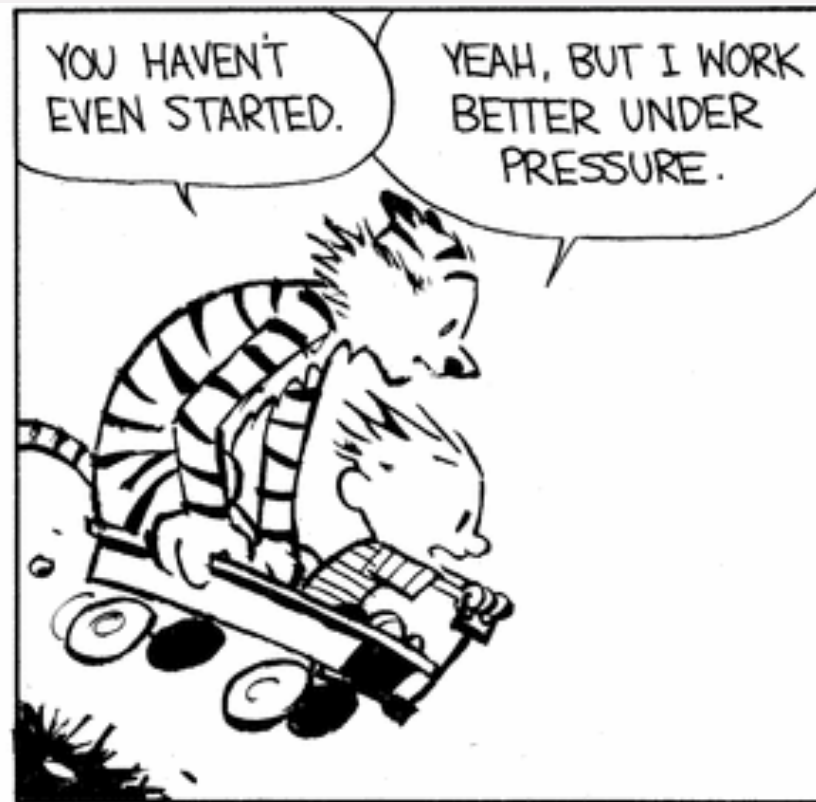
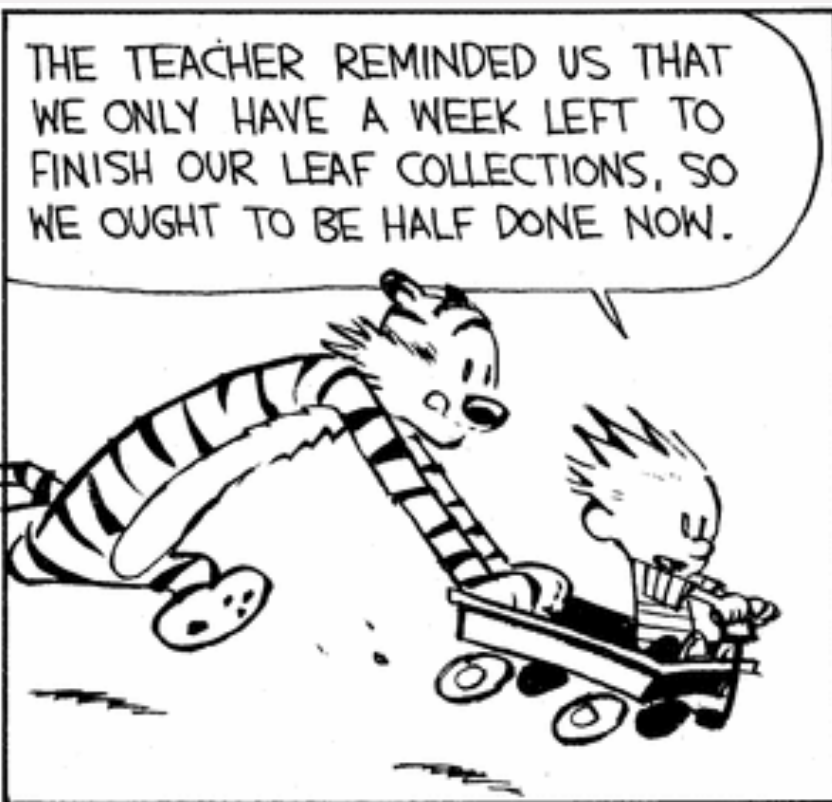
- Then $MD5(Doc'_1) = MD5(Doc'_2)$ (the key idea!).
- Postscript executes $(S_1)(S_2)\text{eq}T_1T_2$ as: if $S_1 = S_2$ execute T_1 ; else execute T_2 .

Journey so far



Journey Ahead: After MidSem





THANK YOU