Message authentication codes and cryptographic hashing

- Message authentication codes
- CBC-MAC
- Cryptographic hashing
- Birthday paradox
- Merkle-Damgard construction
- HMAC

Class of A+IP mechanisms with

$x = (u, t)$ $\quad t = T(k, u)$

that offer computational security

it must be hard to find t given u, but without knowing k

- even if F has observed previous $x_1 \dots x_i$ signed with the same K (known message)
- or even if F can choose $u_1 \dots u_i \neq u$, have them signed with the same k and observe $x_1 \dots x_i$ (chosen message)

Same requirements as a decryption function

Can we use some $D(\cdot,\cdot)$ as the $T(\cdot,\cdot)$?

Differences

in D $\quad H(u) \leq H(x)$, typically $=$

in T $\quad H(t)$ depends on the security level

$H(t) \geq \log_{1/2} \varepsilon$

Possible solutions

Use a block cipher $D(\cdot,\cdot)$
with output entropy $H(t) = \log_2 |\mathcal{T}|$
original message u may be longer/shorter

(1) → split u into blocks of same length or pad

$u_1 \ldots u_n$ sign $t_1 \ldots t_n$ separately

$t_i = T(k, u_i)$

NOT GOOD: attacker may rearrange or drop blocks

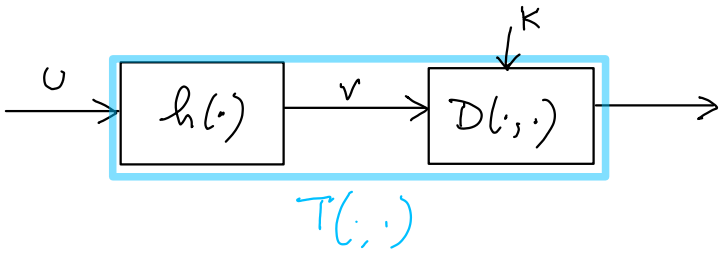
→ use CBC mode (see next page)

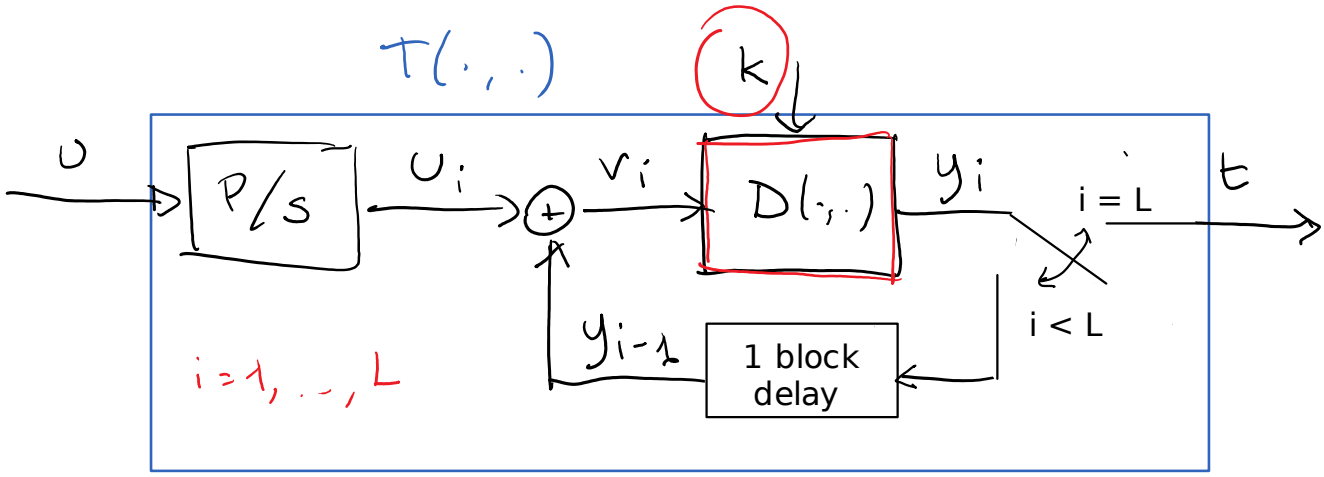
(2) compress the message with a hash function $h(\cdot)$, $h : \mathcal{M} \to \mathcal{T}$

$v = h(u)$, $t = D(k, v)$

$T_k = D_k \circ h$, $D_k : \mathcal{T} \to \mathcal{T}$

hash and sign mechanism





Choose $\mathcal{T}$ tag space according to security

" $D: K \times \mathcal{T} \to \mathcal{T}$

$\mathcal{T}$ must be a group $(\mathcal{T}, +)$

$\mathcal{M} = \mathcal{T}^L$

Commonly used with AES block cipher as D

Security

If $D(\cdot,\cdot)$ is secure (computationally indistinguishable from a random permutation):

- If $\mathcal{M}$ is fixed length, CBC-MAC is secure
  (even modifying a single U_i would change all y_j, $j \geq i$ and $t = y_L$)
- If $\mathcal{M}$ is variable length, possible Known message attack:
  F observes two messages $x = (U, t)$ and $x' = (U', t')$ and constructs U''
  as $U''_i = U_i$ $i = 1, .., L$, $U''_{L+1} = U'_1 - t$, $U''_{L+i} = U'_i$, $i = 2, .., L'$
  then $y''_i = y_i$ $i = 1, .., L$, $y''_{L+1} = y'_1$, $y''_{L+i} = y'_i$, $t'' = t'$
  attack can be prevented by prepending message length

A hashing function is $h : X \to Y$

with $X = A_x^*$ arbitrary length

$Y = A_y^\ell$ fixed length

with the properties

P1) h is easy to compute $y = h(x)$

P2) given p_x, y should be uniform in Y

A cryptographic hash function has the additional properties

P3) one-way (preimage resistance)

given $y_0 \in Y$ it is hard to find $x_0 \in h^{-1}(y_0)$

i.e. $x_0 : h(x_0) = y_0$

P4) weak collision (or 2nd preimage) resistance

given $x_0 \in X$ it is hard to find $x_1 \in X$

$x_1 \neq x_0$ such that $h(x_1) = h(x_0)$

P5) (strong) collision resistance

it is hard to find any pair $(x_1, x_2) \in X^2$

$x_1 \neq x_2$ such that $h(x_1) = h(x_2)$

(listed in increasing order of strength)

Ex.: P4 $\Rightarrow$ P3

can be shown by contradiction ($\not P3 \Rightarrow \not P4$)

assum P3 does not hold, ie. there is a way to invert h start for x_0, from P1 it is easy to compute $y_0 = h(x_0)$ then from y_0 find x_1

P5 $\Rightarrow$ P4

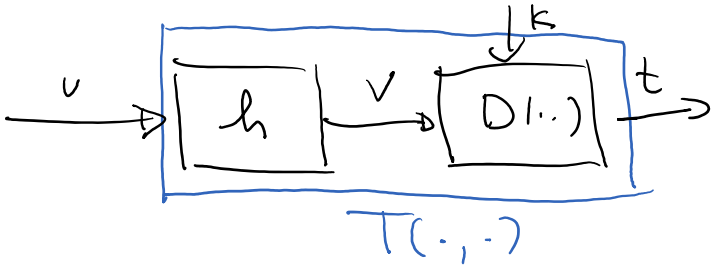
Similarly

Although h is a <u>deterministic function</u>

an ideal model for h is a <u>random function</u>

i.e. $\{h(x)\}$ are iid random variables uniform in Y

Why are P3, P4, P5 needed?



if P4 does not hold modification attack

F intercepts $x = (u, t)$, from u compute $v = h(u)$

find $u' \in h^{-1}(v)$ replace u with u' and have

$$t = T(k, v) = D(k, h(v)) = D(k, v)$$

$$= D(k, h(v')) = T(k, v')$$

send $x' = (v', t)$ will be verified ad accepted

if P3 does not hold same modification attack

if P5 does not hold attacker finds v_1 and v_2 such that $h(v_1) = h(v_2) = v$, lures A into signing $v_1 \Rightarrow x_1 = (v_1, t)$ then reuse t to authenticate $v_2 \Rightarrow x_2 = (v_2, t)$ [chosen message attack]

It is easier to find a <u>collision pair</u> rather than a <u>colliding element</u> to a given one

Consider a random function $h : X \to Y$ and L distinct inputs $x_1 \dots x_L$, plus a special input x_0

probability to find a colliding element to x_0

$$P\left[\bigcup_{i=1}^{L} \{h(x_i) = h(x_0)\}\right]$$

let $y_i = h(x_i)$ iid rv uniform in Y

Let $N = |Y|$

$$P\left[\bigcup_{i=1}^{L} \{y_i = y_0\}\right] = 1 - P\left[\bigcap_{i=1}^{L} \{y_i \neq y_0\}\right]$$

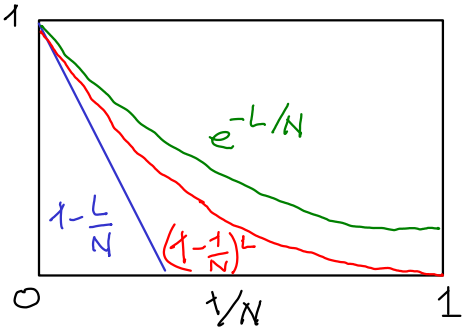
$$= 1 - \sum_{a \in Y} P\left[\bigcap_{i=1}^{L} \{y_i \neq a\} \mid y_0 = a\right] P_{y_0}(a)$$

$$= 1 - \frac{1}{N} \sum_{a} \prod_{i=1}^{L} P[y_i \neq a]$$

$$= 1 - \frac{1}{N} \sum_{a} \prod_{i=1}^{L} \frac{N-1}{N}$$

$$= 1 - \prod_{i=1}^{L} \left(1 - \frac{1}{N}\right)$$

$$= 1 - \left(1 - \frac{1}{N}\right)^{L}$$



$$1-e^{-L/N} \leq 1-\left(1-\frac{1}{N}\right)^L \leq \frac{L}{N}$$

$\backsimeq$ (approx. if L << N) $\backsimeq$

probability to find a collision pair

$$P\left[\bigcup_{i=1}^{L}\bigcup_{j=i+1}^{L}\{y_i = y_j\}\right]$$

$$= 1 - P\left[\bigcap_{i=1}^{L}\bigcap_{j=i+1}^{L}\{y_i \neq y_j\}\right]$$

$$= 1 - P\left[y_2 \neq y_1, y_3 \neq y_1, y_3 \neq y_2, \ldots, y_L \neq y_1, \ldots, y_L \neq y_{L-1}\right]$$

$$= 1 - \sum_{a_1}\sum_{a_2 \neq a_1}\sum_{a_3 \neq a_2, a_1}\cdots\sum_{a_L \neq a_1 \ldots a_{L-1}} P\left[y_1 = a_1, y_2 = a_2, \ldots, y_L = a_L\right]$$

$$= 1 - \sum_{a_1} P[y_1 = a_1] \sum_{a_2 \neq a_1} P[y_2 = a_2] \sum_{a_3 \neq a_2, a_1} P[y_3 = a_3] \cdots \sum_{a_L \neq a_1 \ldots a_{L-1}} P[y_L = a_L]$$

$$= 1 - \left(\sum_{a_1} P_{y_1}(a_1)\right)\left(\sum_{a_2 \neq a_1} P_{y_2}(a_2)\right)\cdots\left(\sum_{a_L \neq a_1 \ldots a_{L-1}} P_{y_L}(a_L)\right)$$

Recall that $P_{y_i}(a) = \frac{1}{N}$ $\forall i, \forall a \in Y$

The first sum has N terms, the second sum has N-1 terms, and so on

$$= 1 - \prod_{i=0}^{L-1}\left(1 - \frac{i}{N}\right)$$

observe that $\sum_{i=0}^{L-1}\frac{i}{N} = \frac{L(L-1)}{2N} = \frac{L^2-L}{2N}$

then $1 - e^{-\frac{L^2-L}{2N}} \leq 1 - \prod_{i=0}^{L-1}\left(1-\frac{i}{N}\right) \leq \frac{L^2-L}{2N}$ $\backsimeq$ if $L^2 \ll N$

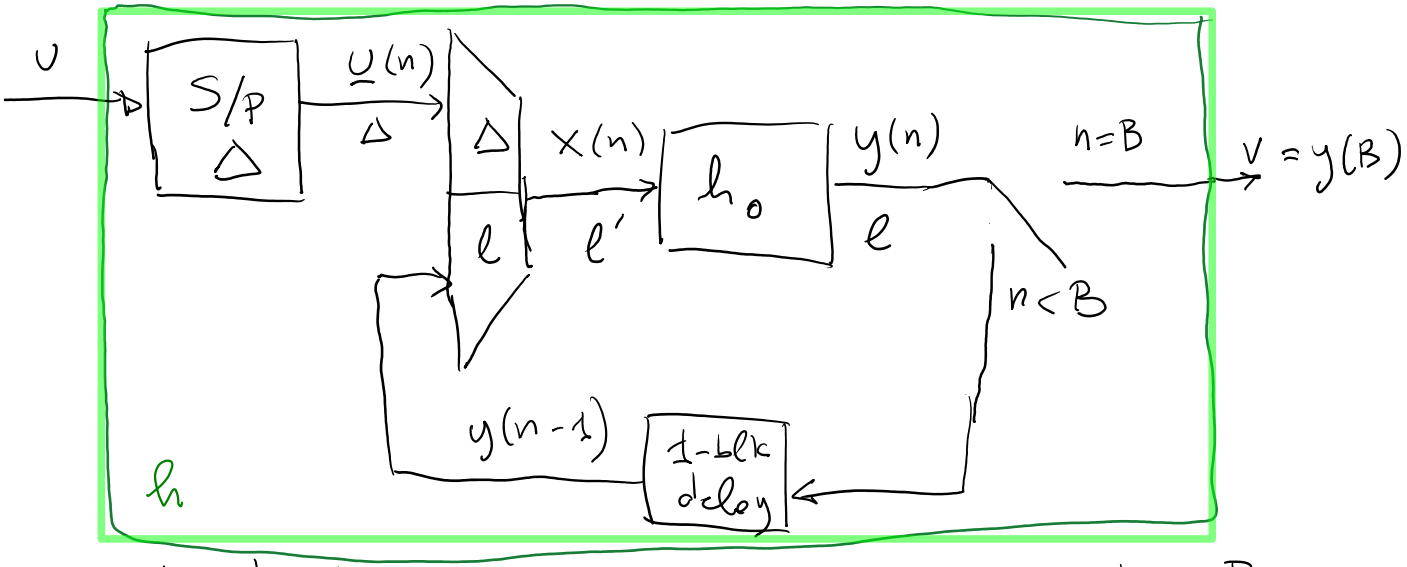
Problem: given a good hash function $h_0 : A^{\ell'} \to A^{\ell}$, $\ell' > \ell$
build some $h : A^* \to A^{\ell}$ that is just as good

Solution: [Merkle-Damgård]

let $\Delta = \ell' - \ell$, $v \in \mathcal{M} = A^*$, $L = \text{length}(v)$, $b_L = \log_{|A|} L$

$B = \lceil (L + b_L)/\Delta \rceil$ will be # of blocks

for $n = 1$ to B



v split into blocks of Δ symbols, $v(n)$, $n = 1, \ldots, B$

last b_L symbols of $v(B)$ are the representation of L
with $b_L \leq \Delta$ (all in the last block)

$$L \leq |A|^{\Delta}$$

$h(v) = V = y(B)$

$y(0) = \beta_0$ initialization vector

$$y(n) = h_0(x(n))$$
$$x(n) = [v(n), y(n-1)]$$

Theorem: if h_0 is a cryptographic hash function (satisfies P1 - P5) in the asymptotic computational sense, then so is h (wrt security parameter ℓ)

Proof: We only prove for P1 and P3

P1: h_0 easy to compute : $\exists$ algorithm A_0 such that $\forall x \in A^{\ell}$, $A_0[x] \to h_0(x)$ and $T_{A_0} < p(\ell)$ for poly p

- replace h_0 with A_0 in definition of h call A the resulting algorithm

$$T_A = B(T_{A_0} + \delta) \quad \text{if } B < q(\ell)$$
$$T_A < p(\ell)\, q(\ell) \quad \text{polynomial}$$

P3: h_0 is preimage resist. $\Rightarrow$ h is preimage resist.

proof by contradiction :

- suppose h is not, show h_0 would also not be

suppose $\exists A : \forall v \; A[v] \to v_0 \in h^{-1}(v)$

$$T_A < p(\ell)$$

define an attack A_0 as follows

start from $V \rightarrow$ apply $A[V] \rightarrow U_0$

from $U_0 \rightarrow$ apply constriction $(B-1)$ times

obtain $y(B-1)$ and hence $x(B)$

since $h_0(x(B)) = y(B) = V$, $x(B) \in h_0^{-1}(V)$

and thus we have a preimage of V for h_0

As regards the computational complexity, let T_0 be the complexity for computing one round

$$T_{A_0} = T_A + (B-1)T_0 \text{ is also poly in } \ell$$

(T_A: polynomial; $(B-1)T_0$: polynomial)

Examples $A = \{0,1\}$

	MD-5	SHA-1	SHA-224	-256	-374	-512
ℓ	128	160	224	256	374	512
D	512	512	—	512	1024	1024
b_L	64	64	—	64	128	128
B_{max}	80	80	64	64	80	80

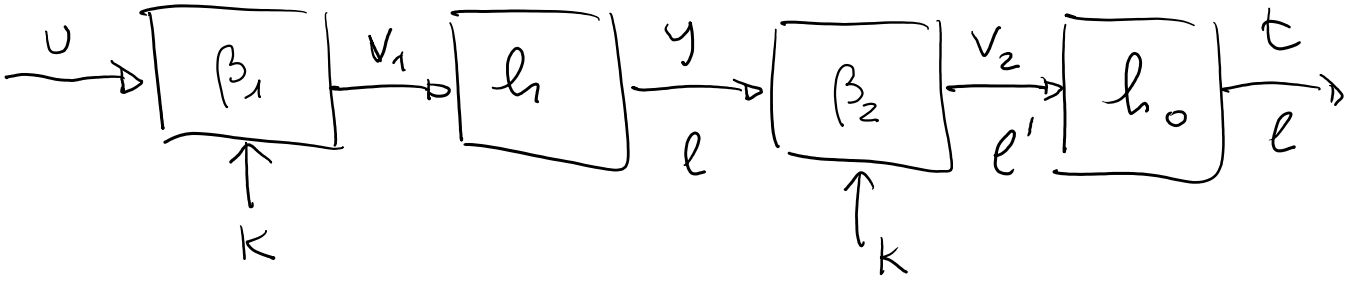
Message Auth through cryptographic hashing

Given $h_0 : A^{\ell'} \to A^{\ell}$

the M-D $h : A^* \to A^{\ell}$

Let $(A, +)$ be a group

Tag computation $T(k, u)$



$K \curvearrowleft \mathcal{M}(K)$ $\quad K = A^{\Delta}$ $\quad \Delta = \ell' - \ell$

$\beta_1, \beta_2 \in A^{\Delta}$ publicly known

$v_1 = [K + \beta_1, u] \in A^{L+\Delta}$

$y = h(v_1)$

$v_2 = [K + \beta_2, y]$

$t = h(v_2)$

Strong even with non collision resistant h, h_0

$\Rightarrow$ if the attacker finds a collision v_1, v_1'
it is unlikely that they both start with $k + \beta_1$