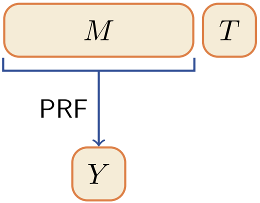
PMAC's Message Length Dependence

Atul Luykx, Bart Preneel, Alan Szepieniec,
Kan Yasuda

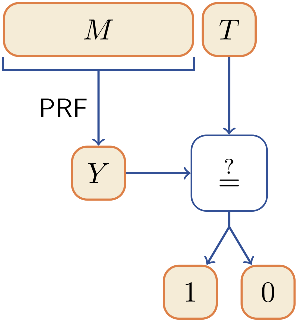
PRF-based MACs

1. Many MACs are PRF-based
2. Due to reduction from authenticity bound to PRF bound we can focus on PRF-bounds for MACs (Connection with verification queries)



PRF-based MACs

1. Many MACs are PRF-based
2. Due to reduction from authenticity bound to PRF bound we can focus on PRF-bounds for MACs (Connection with verification queries)



Upper and Lower Bounds

1. PRFs with state size n have a generic attack with success probability $q^2/2^n$, with q the number of queries made to the PRF
2. In contrast, the best-known MAC upper bounds are of the form $\ell q^2/2^n$: PMAC, EMAC, CBC-MAC, HMAC/NMAC, polynomial based MACs.

Generic Attacks and Optimal Bounds

1. Factor-ℓ gap: there is no known generic attack establishing a $\ell^\epsilon q^2/2^n$ lower bound for some $\epsilon > 0$
2. Two possibilities:
    1. There exists such a generic attack
    2. There exists a MAC of state size n with upper bound $q^2/2^n$

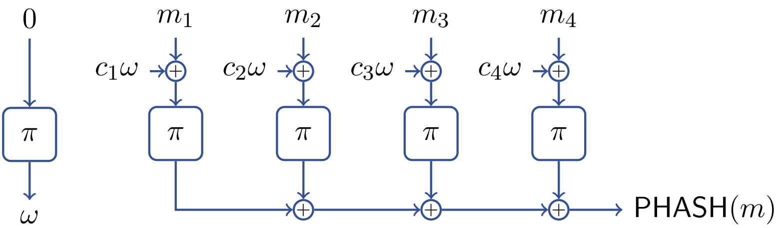
   We focus on the second possibility

PMAC's Role

1. None of the above mentioned MACs can be candidates: all have attacks establishing dependence on message length
2. Exception: PMAC. Has no attack establishing the bound
3. PMAC is interesting to analyze:
    1. It is simple
    2. It has a radically different structure from other MACs
    3. Many beyond the birthday bound variants, PMAC-with-Parity, PMACX

PHASH Definition

X	finite field
$\vec{c} \in \mathsf{X}^{\ell}$	Vector of masks
$\pi : \mathsf{X} \to \mathsf{X}$	URP



Standard argument to reduce to PRP

PMAC Definition

1. PMAC, add final block cipher call to PHASH, fix finite field, two types of masks:
   1. Gray codes
   2. Powering up

Connection Between PHASH and PMAC

1. A collision for PHASH implies a collision for PMAC $\rightarrow$ distinguishing attack
2. Open problem: to what extent does a distinguishing attack against PMAC imply a collision for PHASH?

Generic PMAC

Generic PMAC, with independent output transformation

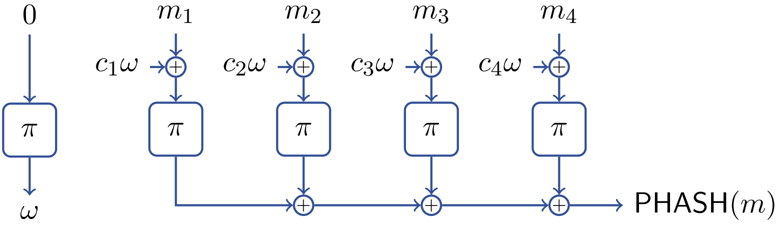
1. Tight connection between generic PMAC and PHASH
2. Allows us to focus on PHASH

Results

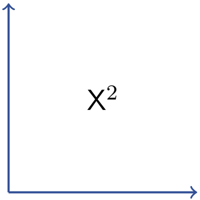
1. One of the following two statements is true:
   1. either there are infinitely many instances of PHASH for which it is impossible to find collisions with probability greater than $2q^2/2^n$,
   2. or finding a collision against PHASH with probability greater than $2q^2/2^n$ is computationally hard*

   *this statement relies on a conjecture
2. Collision for PHASH with Gray codes establishing roughly linear dependence on message length

Approach

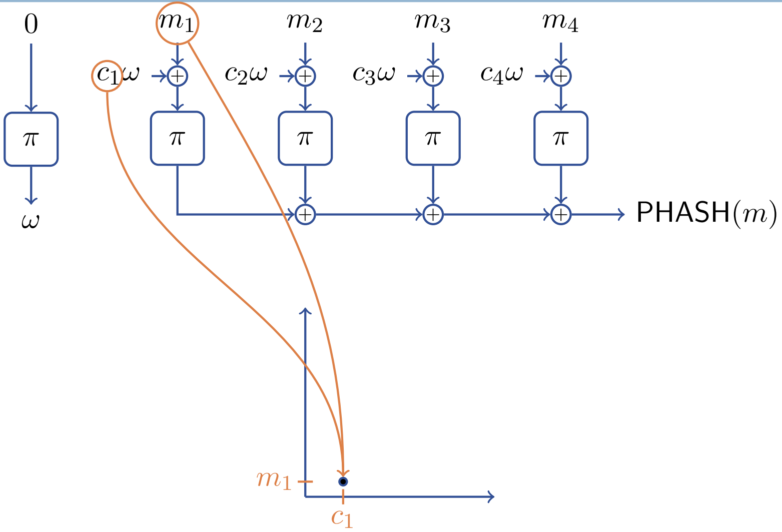


0
m_1
m_2
m_3
m_4
$c_1\omega$
$c_2\omega$
$c_3\omega$
$c_4\omega$
π
π
π
π
π
ω
PHASH(m)

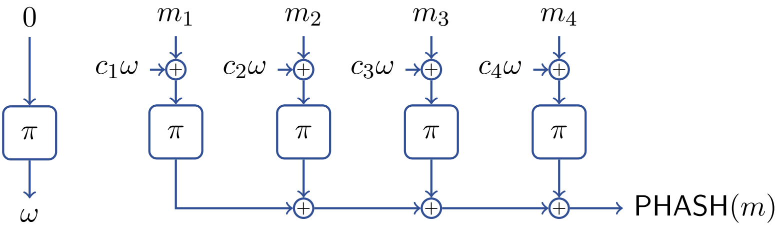


X^2

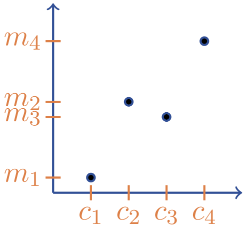
Approach


0
m_1
m_2
m_3
m_4
$c_1\omega$
$c_2\omega$
$c_3\omega$
$c_4\omega$
π
ω
PHASH(m)
m_1
c_1

Approach

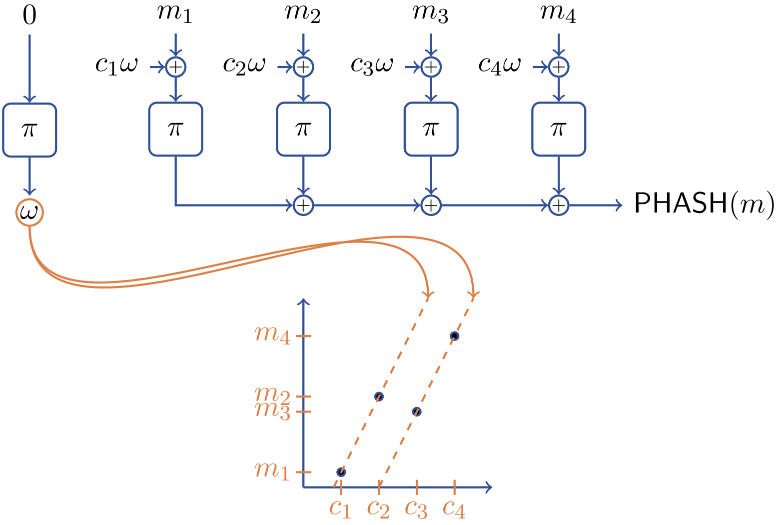


0
m_1
m_2
m_3
m_4
$c_1\omega$
$c_2\omega$
$c_3\omega$
$c_4\omega$
π
ω
PHASH(m)



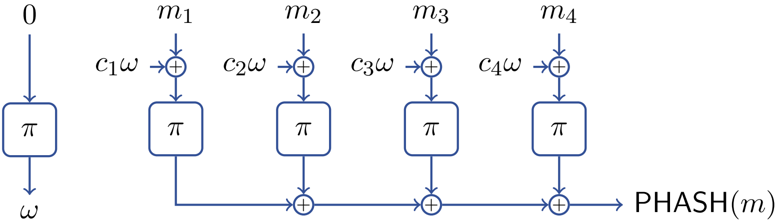
m_4
m_2
m_3
m_1
c_1 c_2 c_3 c_4

Approach

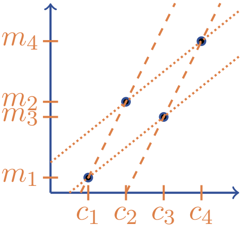


0
m_1
m_2
m_3
m_4
$c_1\omega$
$c_2\omega$
$c_3\omega$
$c_4\omega$
π
ω
PHASH(m)
m_4
m_2
m_3
m_1
c_1
c_2
c_3
c_4

Approach



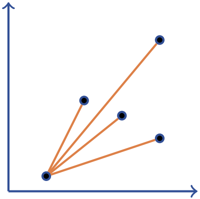
0
m_1
m_2
m_3
m_4
$c_1\omega$
$c_2\omega$
$c_3\omega$
$c_4\omega$
π
ω
PHASH(m)



m_4
m_2
m_3
m_1
c_1
c_2
c_3
c_4

Connection With PHASH Collision Probability

Two messages $\vec{m}_1$ and $\vec{m}_2$ collide with probability $k/2^n$ if the corresponding set in X^2 is evenly covered by k slopes.
Simple proof of ℓ-bound:



Set Evenly Covered by Two Slopes

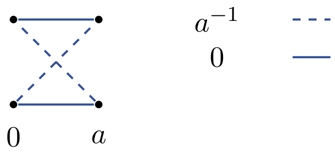


Figure: A set of four points evenly covered by the slopes 0 and a^{-1}. The x-coordinates of the points are 0 and a, and the y-coordinates are 0 and 1.

Guarantees a collision with probability $2/2^n$.

Set Evenly Covered by Three Slopes

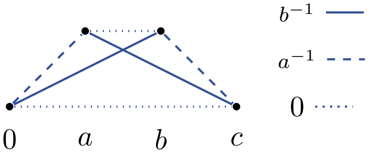


Figure: A set of four points evenly covered by the slopes 0, a^{-1}, and b^{-1}. The x-coordinates of the points are 0, a, b, and c, and the y-coordinates are 0 and 1.

Exists if and only if $a + b + c = 0$.

Another Set Evenly Covered by Three Slopes

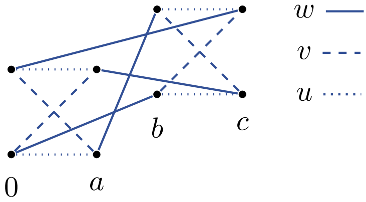


Figure: A set of points evenly covered by the slopes u, v, and w. Each point is accompanied by another point with the same x-coordinate. The x-coordinates of the pairs are indicated below the lower points.

Exists if and only if $a^2 + b^2 + c^2 + ab + ac = 0$.

Evenly Covered Sets in General

The x-coordinates of evenly covered sets satisfy one of the following:

1. They contain a subset summing to zero (NP-complete)
2. They are the solution to a non-trivial binary quadratic form (similar problem NP-complete)

Conjecture

Given $S \subset \mathsf{X}$, finding a subset of S satisfying either of the above requirements is computationally hard.

Searching for Evenly Covered Sets

Proposition

An evenly covered set with distinct x-coordinates forms a complete graph if and only if the x-coordinates are an additive subgroup of X.

1. For sufficiently long messages, the masks will always contain an additive subgroup
2. Finding additive subgroups in Gray codes is easy for every power of two.

Success probability of Gray code attack:

$$\frac{2^{k-1}-1}{2^n} \text{ for } \ell = 2^k$$