Problem 8: Collision

<u>Notations</u>:

1. A message M consisting of k blocks of n bits:
   $M_k = \{m_1, m_2, ... , m_k : m_i \in F_2^n\}$
2. Concatenate operation : a = “123” , b = “abc”
   a || b = “123” || “abc” = “123abc”
   => $M_k = M_{k-1} \| m_k$
3. $f(x)$ is a secret random function. $x, f(x) \in F_2^n$

<u>Scenario: With Constraints:</u>

Given message M_k and its hash H(M_k),
10 of n random different pairs ($x, f(x)$),
the structure of the hash function:

$$h_i = f(h_{i-1} \oplus m_i) \oplus m_i \quad with \quad h_0 = 0$$

and we are not allowed to calculate more hash.
(more hash calculations means more ($x, f(x)$) pairs because $f(x) = H(x) \oplus x$)

We have some interesting properties from the hash function:
With x is a n bits message:

$$H(x) = f(x) \oplus x \qquad with \quad x \in F_2^n$$

We can concatenate m to x:

$$\begin{aligned} H(x \| m) &= f(H(x) \oplus m) \oplus m \\ &= f(f(x) \oplus x \oplus m) \oplus m \end{aligned}$$

Replace m = $f(x)$, we get:

$$\begin{aligned} \Rightarrow \quad H(x \| f(x)) &= f(f(x) \oplus x \oplus f(x)) \oplus f(x) \\ &= f(x) \oplus f(x) \\ &= 0 \qquad (1) \end{aligned}$$

<u>From (1), we answered the Question1</u>. For any message $m_1 = x_1 \| f(x_1)$ and $m_2 = x_2 \| f(x_2)$, we have $H(m_1) = H(m_2) = 0$

Based on the 1st property (1), we can append any messages m behind $x \| f(x)$ to create a collision with H(m).

$$M'_3 = \{m'_1, m'_2, m'_3\} = x \| f(x) \| m$$

$$\begin{aligned} H(M'_3) &= H(x \| f(x) \| m) \\ &= f(H(x \| f(x)) \oplus m) \oplus m \\ &= f(0 \oplus m) \oplus m \\ &= H(m) \qquad (2) \end{aligned}$$

From (2), we proved that we can add as many “$x \| f(x)$” before any message and the hash will not change. We call this “Prefix collision”

Given $H(M_k \| x) = f(H(M_k) \oplus x) \oplus x$, it seems that we can concatenate $H(M_k)$ behind M_k in order to remove $H(M_k)$ from inside F function.

$$\begin{aligned} H(M_k \| H(M_k) &= f(H(M_k) \oplus H(M_k)) \oplus H(M_k) \\ &= f(0) \oplus H(M_k) \\ &= f(0) \oplus H(M_k) \qquad (3) \end{aligned}$$

From (3), we generalized it:

$$\begin{aligned} H(M_k \| (H(M_k) \oplus x)\ ) &= f(H(M_k) \oplus H(M_k) \oplus x) \oplus H(M_k) \oplus x \\ &= f(x) \oplus H(M_k) \oplus x \\ &= H(M_k) \oplus f(x) \oplus x \end{aligned}$$

The part $f(x) \oplus x$ resemble somewhat with property (1), doing similar to property (1), we concatenate "$H(M_k) \oplus f(x)$" into the message.

$$M'_{k+2} = M_k \| (H(M_k) \oplus x) \| (H(M_k) \oplus f(x))$$

$$\begin{aligned} H(M'_{k+2}) &= H(M_k \| (H(M_k) \oplus x) \| H(M_k) \oplus f(x)\ ) \\ &= f(\ H(M_k) \oplus f(x) \oplus x \quad \oplus \quad H(M_k) \oplus f(x)\ ) \oplus H(M_k) \oplus f(x) \\ &= f(\qquad x \qquad) \ \oplus H(M_k) \oplus f(x) \\ &= H(M_k) \qquad (4) \end{aligned}$$

From (4), we proved that we can add as many "$(H(M_k) \oplus x) \| (H(M_k) \oplus f(x))$" at the end of any message and the hash will not change. We call this "Postfix collision".

<u>Using "Prefix collision" and "Postfix collision", we answered Question2.</u>

From M_k, $H(M_k)$ and a number of pair of $(x, f(x))$, we can create preimage $M' \neq M_k$ but $H(M') \neq H(M_k)$ under the constraints of the problem.

<u>Question3:</u>

We converted string M = "A random matrix is likely decent" to hex.

Hex(M)=412072616e646f6d206d6174726978206973206c696b656c7920646563656e74

H(M)=D64ADBDF7458E158801EF33370D0A7C524065545447517D39CCBB0D8D9015BDC

For the sake of demonstration, we using pair $(13, f(13))$ for "prefix collision" and $(29, f(29))$ for "postfix collision", we created

M'= $13 \| f(13) \| M \| (H(M) \oplus 29) \| (H(M) \oplus f(29))$

Hex(13)=000000000000000000000000000000000000000000000000000000000000000d
Hex(f(13))=A219E692908D8F4F248488BD1643EC22A5DF31BAF43DBA9AC11920403E115C3B
Hex(29)=000000000000000000000000000000000000000000000000000000000000001d
Hex(f(29))=A09E43A8A80E174434AAA6D9AC3CF3BB586678BFC423D121150F7BC73BD223FB

Hex $(H(M) \oplus 29)$ =
d64adbdf7458e158801ef33370d0a7c524065545447517d39ccbb0d8d9015bc1

Hex $(H(M) \oplus f(29))$ =
76d49877dc56f61cb4b455eadcec547e7c602dfa8056c6f289c4cb1fe2d37827

M'=
000000000000000000000000000000000000000000000000000000000000000dA219E692908D8F4F
248488BD1643EC22A5DF31BAF43DBA9AC11920403E115C3B412072616e646f6d206d617472697
8206973206c696b656c7920646563656e74d64adbdf7458e158801ef33370d0a7c524065545447517d
39ccbb0d8d9015bc176d49877dc56f61cb4b455eadcec547e7c602dfa8056c6f289c4cb1fe2d37827

Testing H(M') =
96927213786383450510742783183865994351939043923041992908825186024229118761948
=D64ADBDF7458E158801EF33370D0A7C524065545447517D39CCBB0D8D9015BDC

Extended "Postfix collision" into "Middle Inject collision":

In the case of no constraint on the number of calculating the hash, given M_k and $H(M_k)$, we can use the hash function to break down $H(M_k)$. We can get all H_i for each M_i if we recalculate the hash from H_1, using M_1, till H_k:

$$H_i = f(H_{i-1} \oplus m_i) \oplus m_i$$

=> $$H(M_i) = f(H(M_{i-1}) \oplus m_i) \oplus m_i$$

We can put 2 blocks "$(H_i(M_i) \oplus x) \,||\, (H_i(M_i) \oplus f(x))$" between m_i and m_{i+1}

$$M'_{k+2} = \{m'_1, ..., m'_{k+2}\} = \{m_1, ..., m_i, (H_i(M_i) \oplus x),\ (H_i(M_i) \oplus f(x)), m_{i+1}, ..., m_k\}$$

$$H(M'_{i+2}) = H(M_i \,||\, (H_i(M_i) \oplus x) \,||\, (H_i(M_i) \oplus f(x))\,)$$

$$= H(M_i)$$

$$H(M'_{i+3}) = f(H(M'_{i+2}) \oplus m_{i+1}) \oplus m_{i+1}$$

$$= f(H(M_i) \oplus m_{i+1}) \oplus m_{i+1}$$

$$= H(M_{i+1})$$

=> $$H(M'_{k+2}) = H(M_k) \qquad (5)$$

From (5), we proved that we can put as many new blocks in the middle of the original message M_k to make the new message M'_k more indistinguishable from the original and the hash of M'_k is still equal to the hash of M_k.

However, it is required that we have no constraint on the number of times of calculating hash so we can breakdown $H(M_k)$ into smaller H_i and to calculate more pairs $(x, f(x))$