Breaking the ICE -
Multicollisions in Iterated Concatenated and Expanded (ICE) Hash Functions

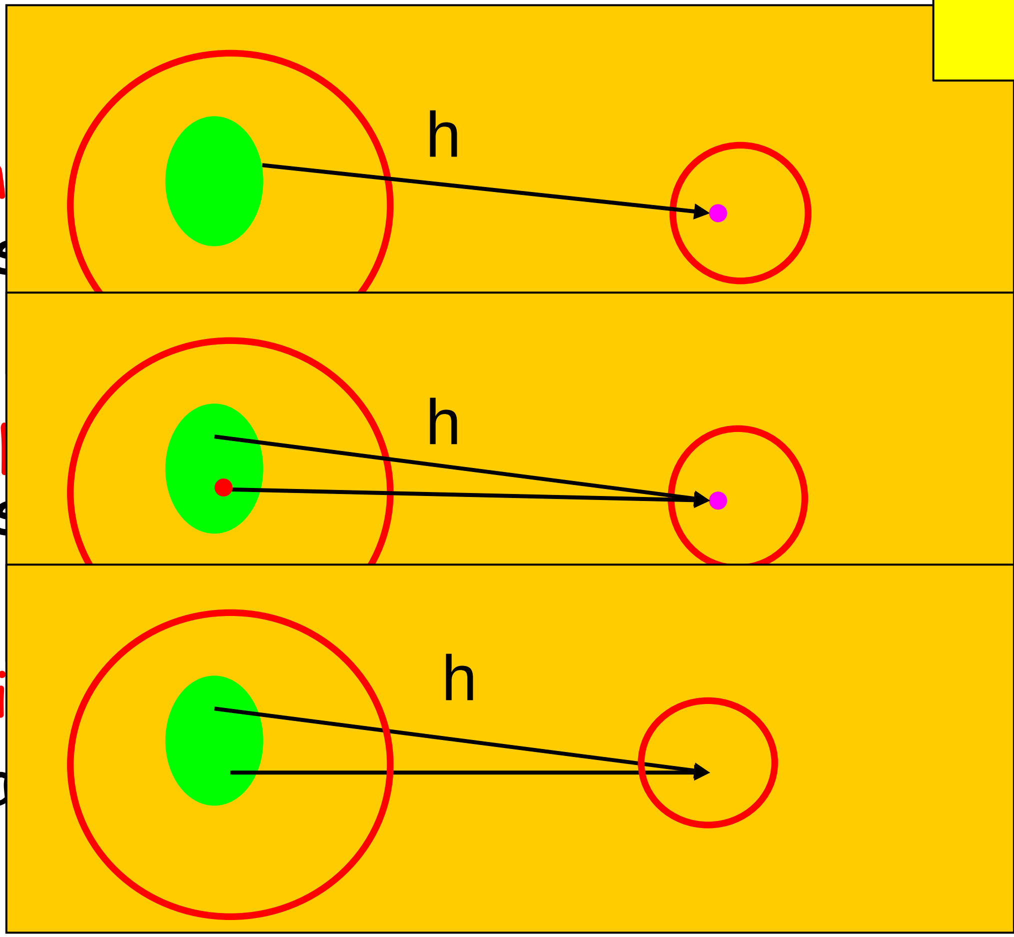
Adi Shamir

Joint work with
Ya'akov Hoch

IPA - 5/10/06

Classical Properties of hash functions

n – the output size of h



- Preim ... rationally infeas ...

  $O(2^n)$

- 2-nd ... omputationally infeas ... h(x)

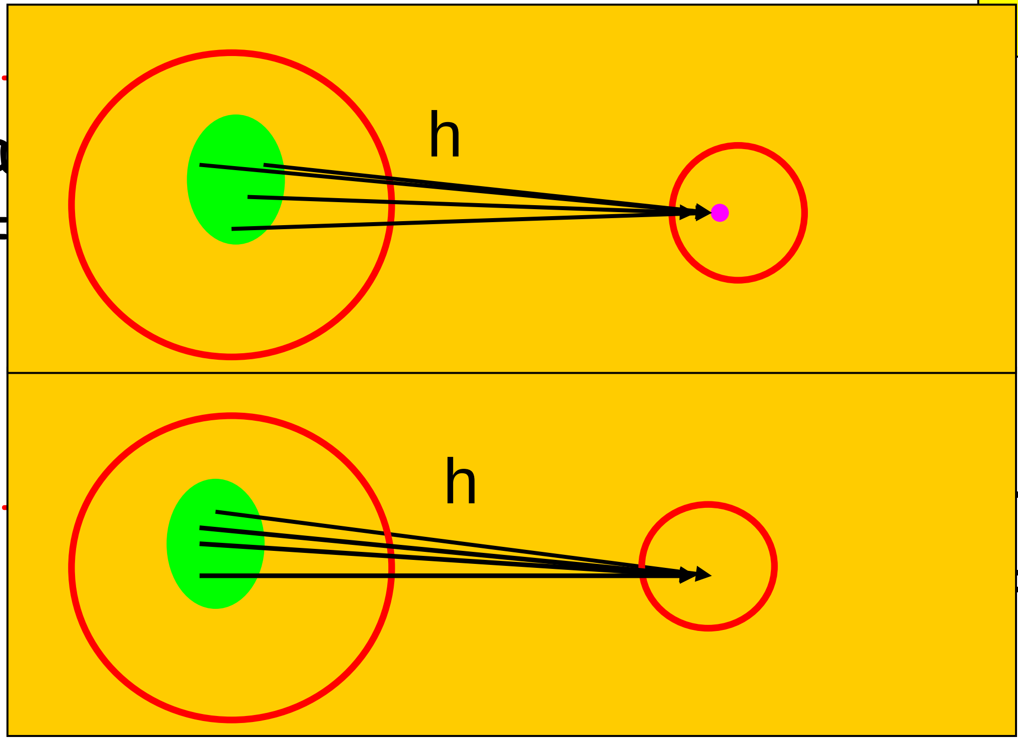
  $O(2^n)$

- collisi ... infeasible to find c ...)=h(x)

  $O(2^{n/2})$

More properties…

- K(multi)-
  computa x_i s.t.
  $h(x_1)$=…=



n – the output size of h

$O(k2^n)$

- K(multi)-   tionally
  infeasib $=h(x_k)$

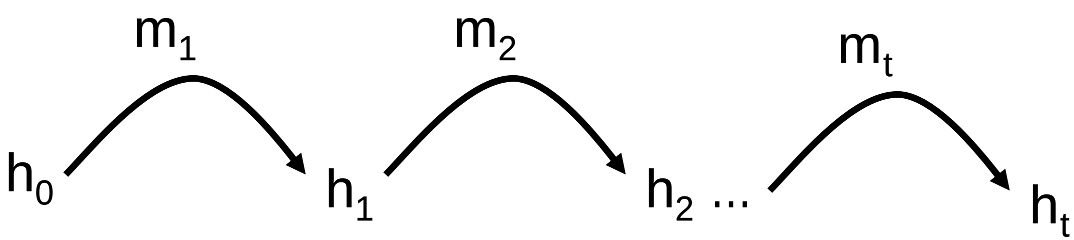
$O(2^{n(k-1)/k})$

Iterated Hash Functions

- A standard way to construct hash functions is as follows:
- Start from an initial hash value h_0
- Calculate $h_i=f(h_{i-1},m_i)$

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

- Output the last hash value h_t



Concatenated Hash Functions

- Concatenate the outputs of a number of independent hash functions
- H(M)=F(M)||G(M)
- Want to enlarge the output size – to protect against birthday attacks
- [illegible] the construction against discovery of [illegible] attack in one of the hash functions
- Secure against collisions if F and G are random oracles

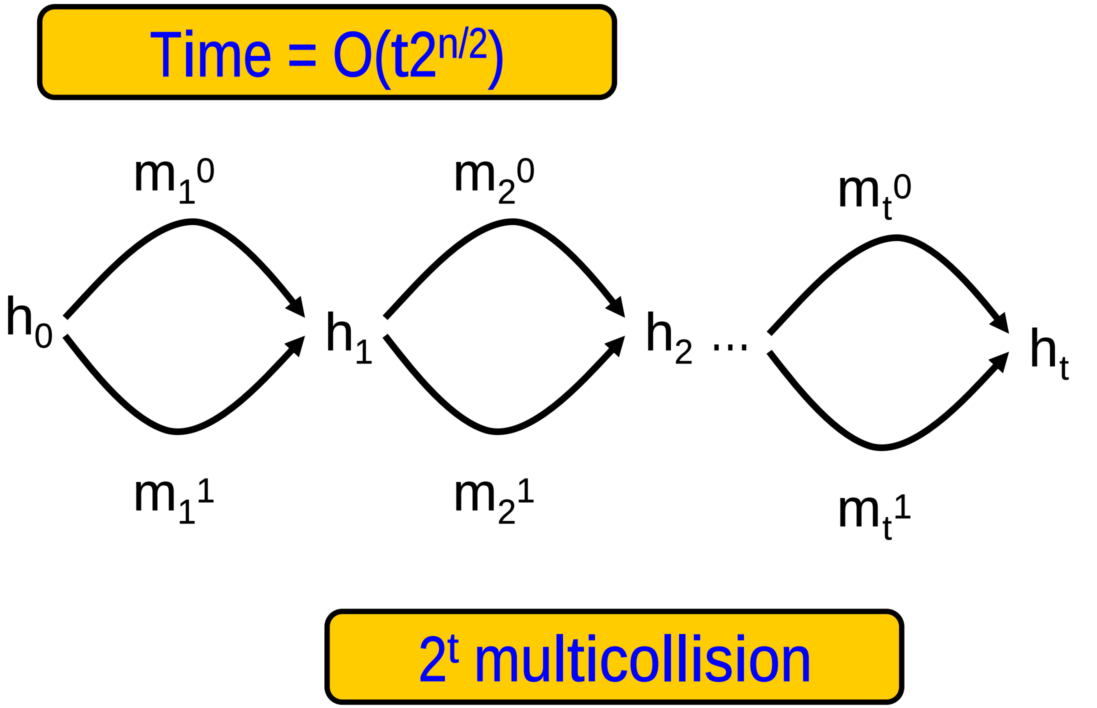
$F,G:\{0,1\}^* \rightarrow \{0,1\}^n$

$H:\{0,1\}^* \rightarrow \{0,1)^{2n}$

$O(2^n)$

Joux Multicollisions in Iterated Hash Functions

- Use iterated structure to create large multicollisions



Attacking a concatenated construction

- Form a $2^{n/2}$ multicollision in the first hash function
- We expect to find a collision in the second function among the $2^{n/2}$ colliding messages
- The attack can be generalized to attack
  - multiple concatenations
  - produce multi-preimages (in time 2^n)

M_i	$F(M_i)$	$G(M_i)$
M_1	X	Y_1
M_2	X	Y_2
...	...	...

$H(M)=F(M)||G(M)$

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

Possible Countermeasures

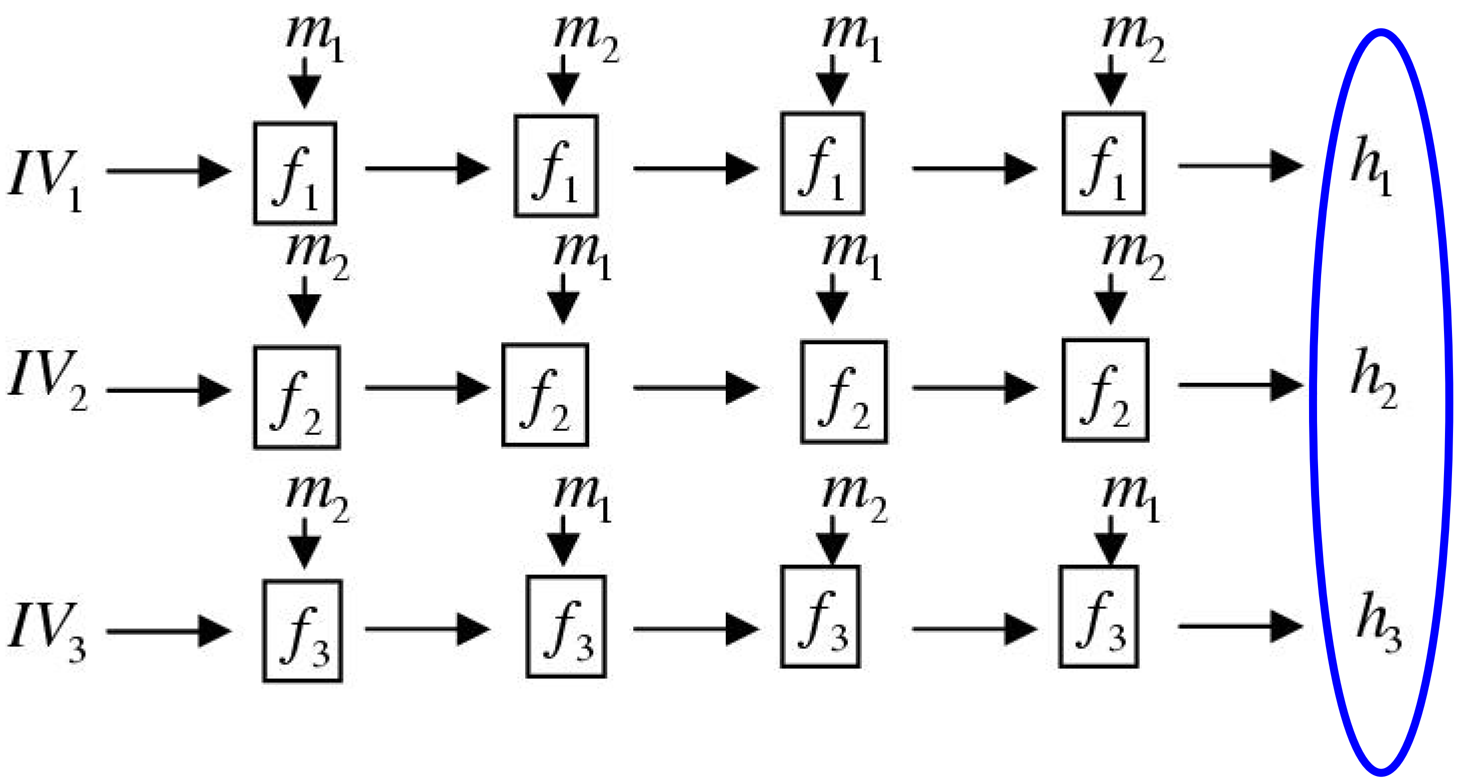
- Larger internal state - Lucks' proposition of a double width pipe
- Expansion - Using message blocks more than once

$$M=m_1m_2\ldots m_t \longrightarrow M=m_1m_2m_1m_5m_1\ldots m_tm_2m_5m_{t-1}\ldots$$

Problem Statement

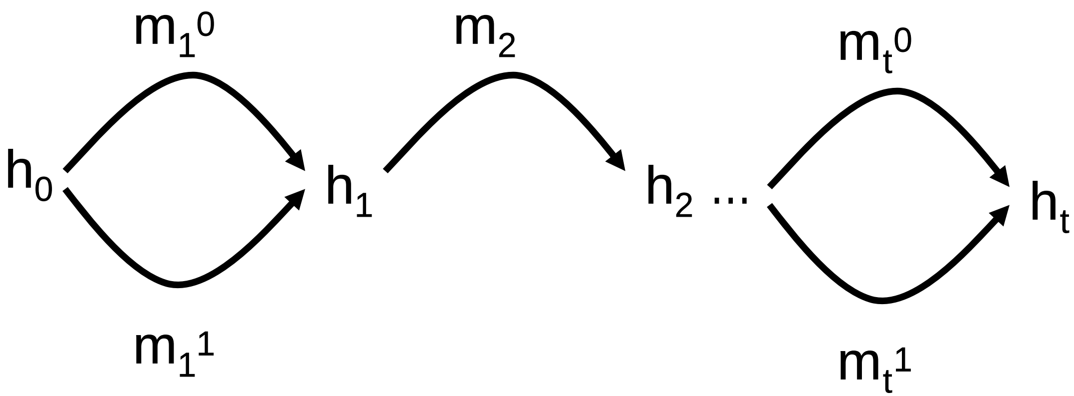
- Given a hash function H – find a 2^k multicollision in H
- Iterated and Concatenated – solved by Joux
- Iterated, Concatenated and Expanded – a special case solved by Nandi & Stinson
- Iterated, Concatenated and Expanded (by any constant factor)–solved in this presentation

Example of an ICE Hash function



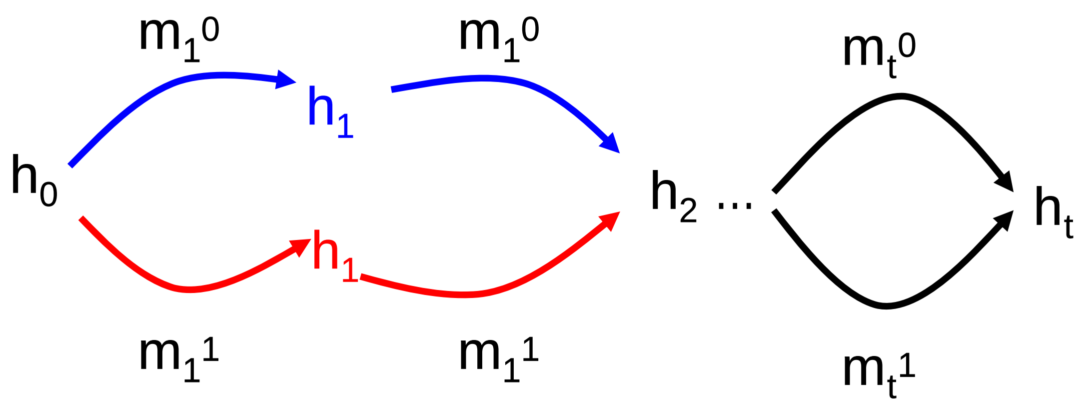
Some warm up examples

- Can have a fixed value for some message blocks



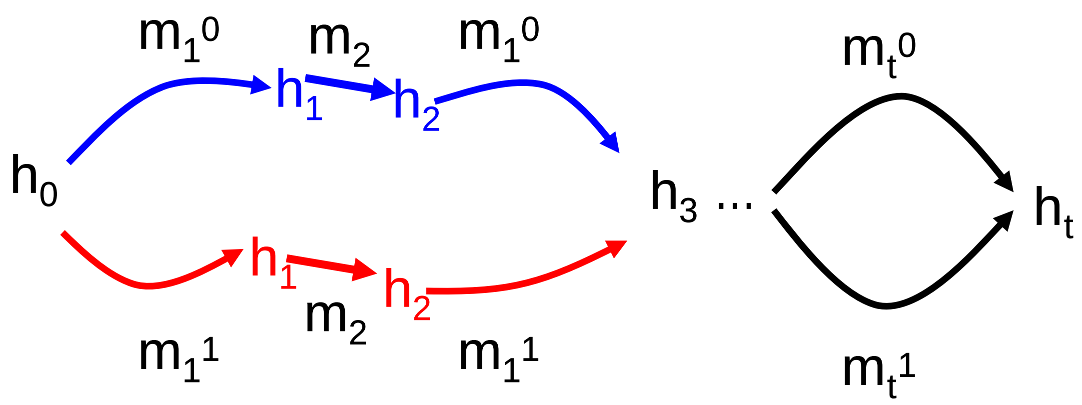
Some warm up examples

- Can have consecutive stretches of the same message block



Some warm up examples

- Can have consecutive stretches of the same message block

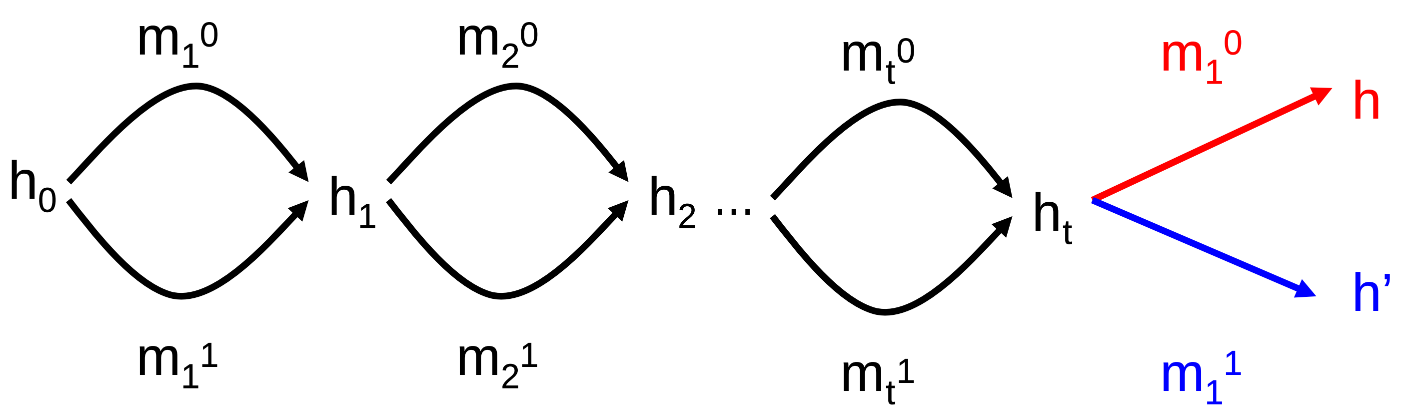


Some warm up examples

- Message expansion takes a message M and outputs M||M
- Find a 2^k multicollision in the iterated hash function based on the expanded message

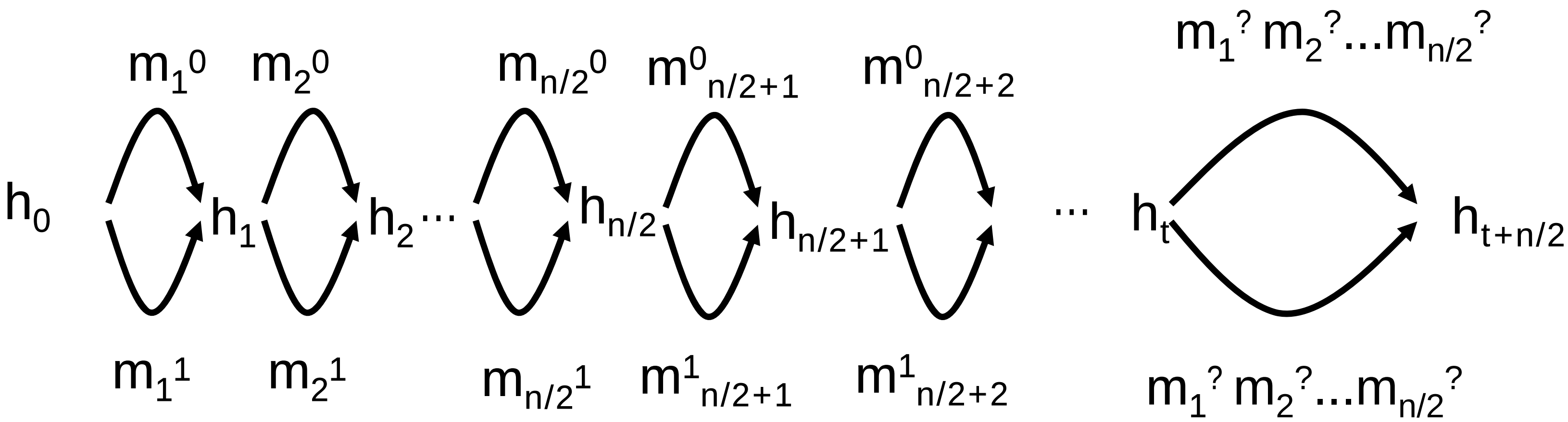
Example I

$$H(M)=F(M||M)=F(m_1m_2m_3...m_tm_1m_2...m_t)$$



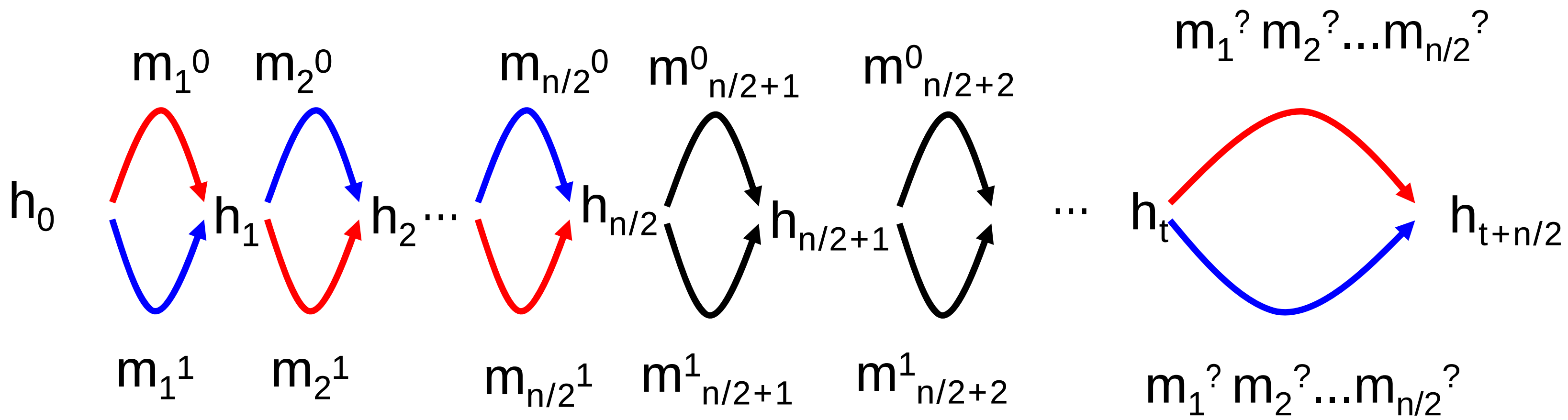
Example I

$$H(M)=F(M||M)=F(m_1m_2m_3\ldots m_tm_1m_2\ldots m_t)$$



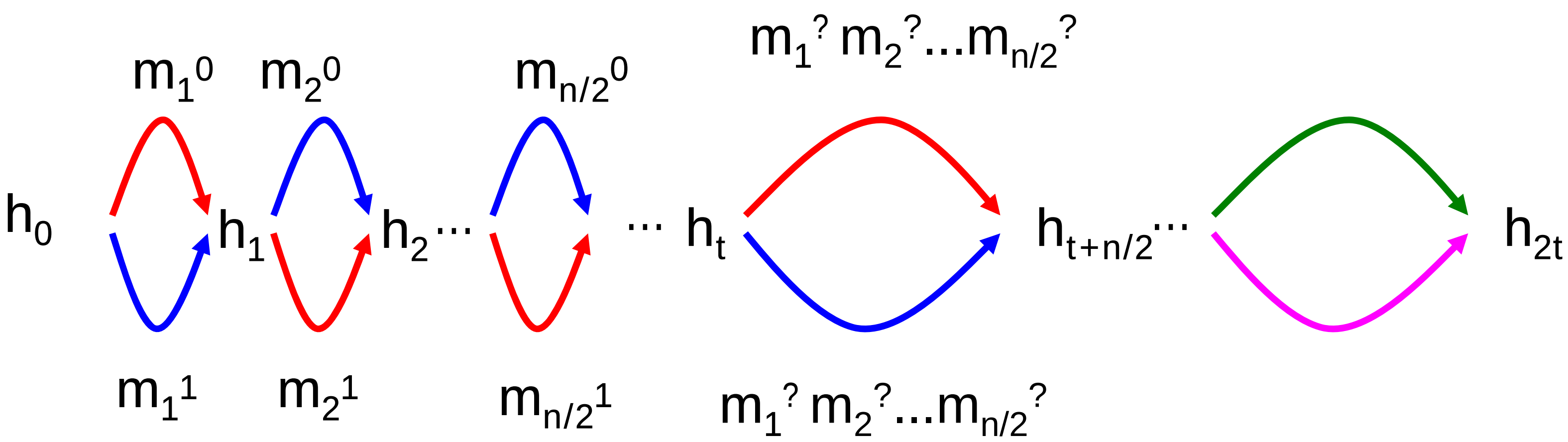
Example I

$$H(M)=F(M||M)=F(m_1m_2m_3...m_tm_1m_2...m_t)$$



Example I

$$H(M)=F(M||M)=F(m_1m_2m_3...m_tm_1m_2...m_t)$$

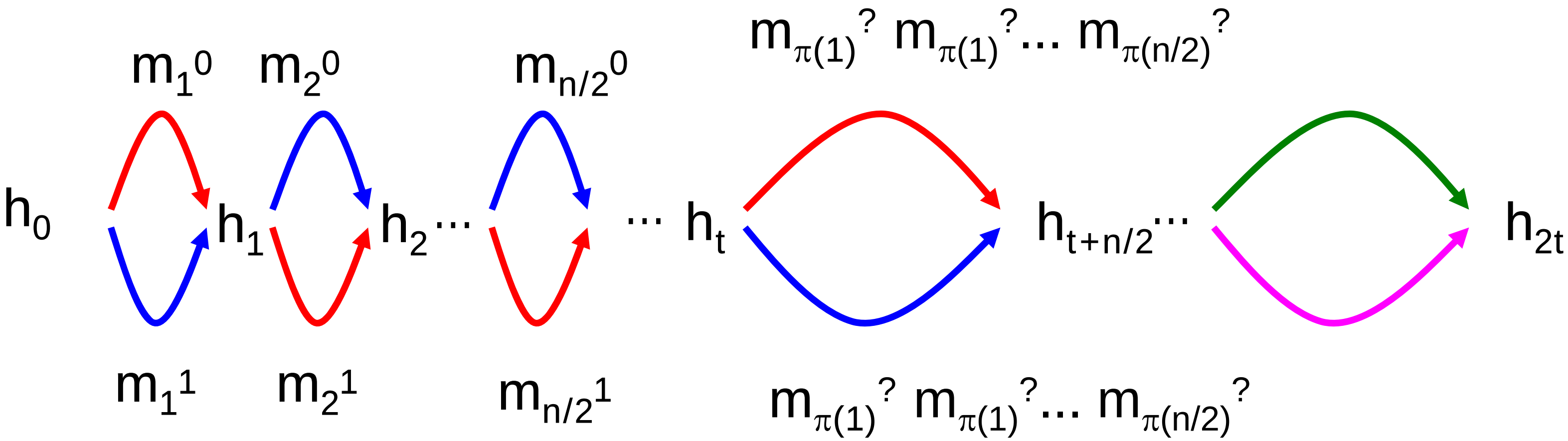


Works for any f [...] etitions

$2^{2t/n}$ multicollision

Example II - 2 successive permutations

- Message expansion adds a permutation of the original message blocks
- $E(M) = m_1 m_2 \ldots m_t m_{\pi(1)} m_{\pi(2)} \ldots m_{\pi(t)}$
- Use the same procedure as before



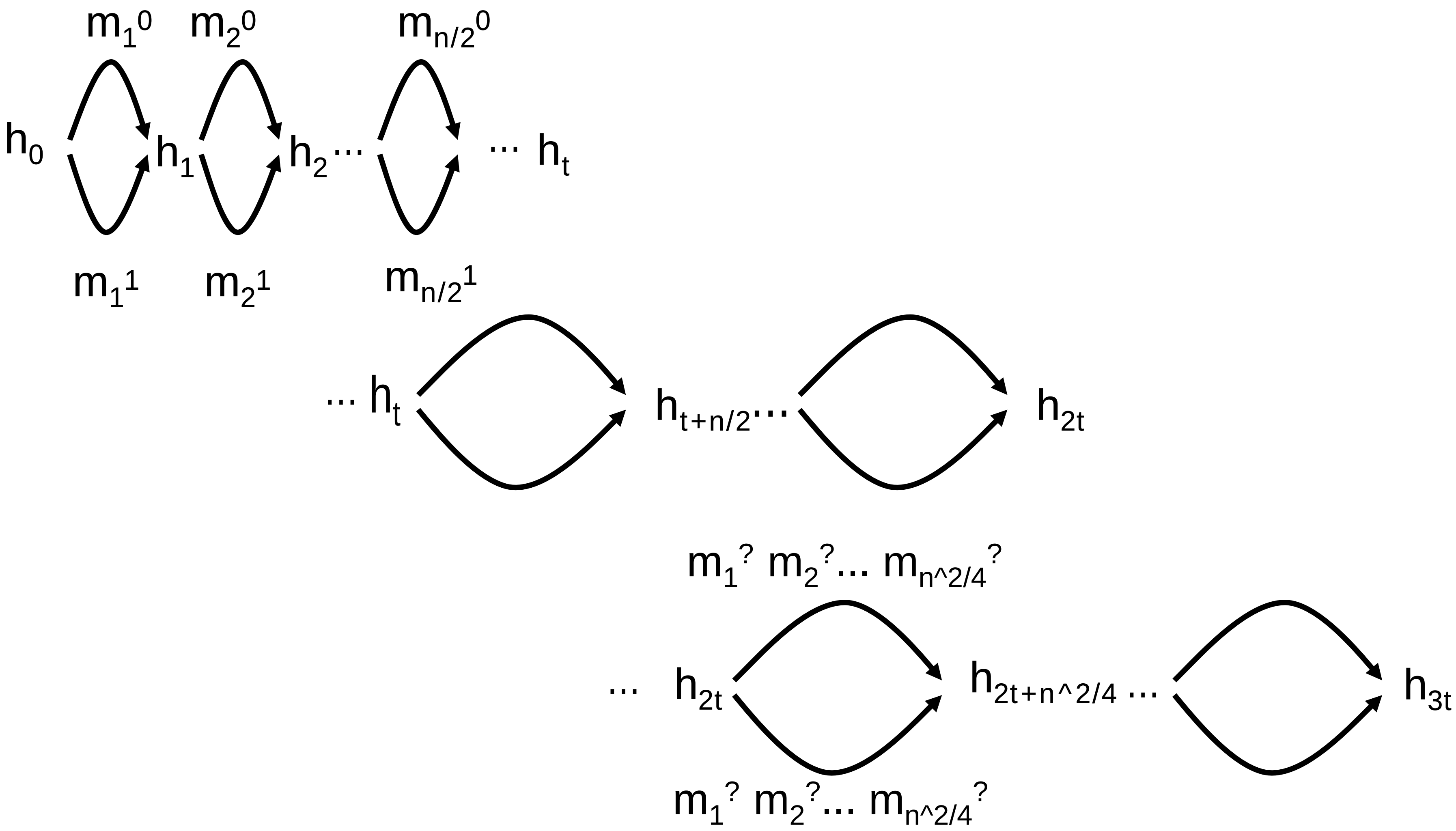
Previous results (Nandi & Stinson)

- If the message expansion contains each message block at most twice, can find a 2^k multicollision in time $2^{n/2}C(n,k)$ where $C(n,k)$ is polynomial in n, k

Our results

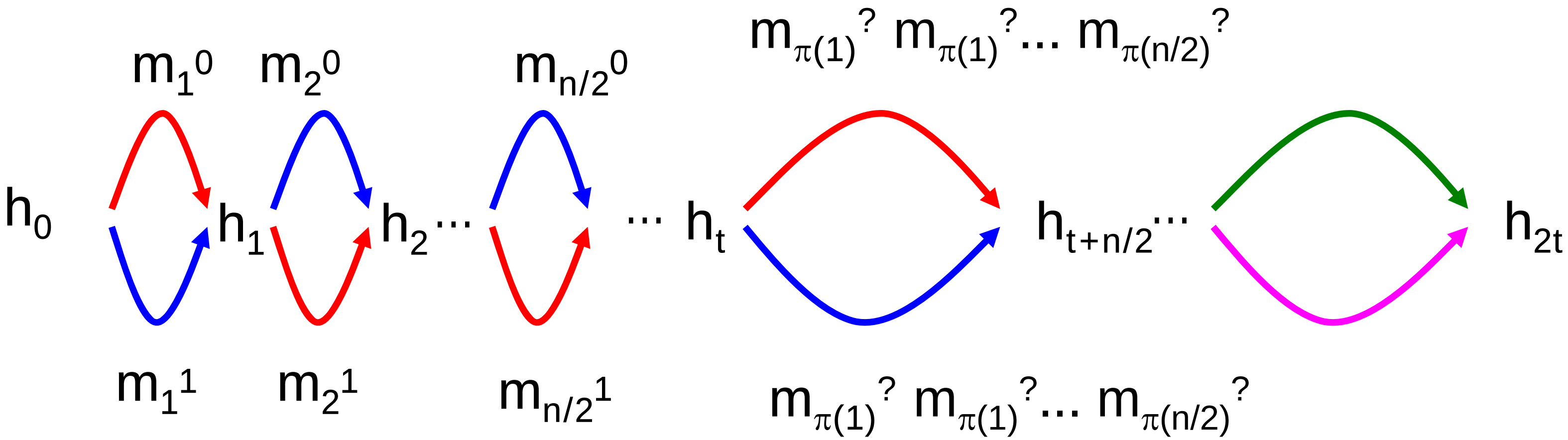
- If the message expansion expands by a constant factor e (by duplicating message blocks) can find a 2^k multicollision in time time $2^{n/2}C(n,k,e)$ where $C(n,k,e)$ is polynomial in n, k (but exponential in e)

Example III - 3 successive copies



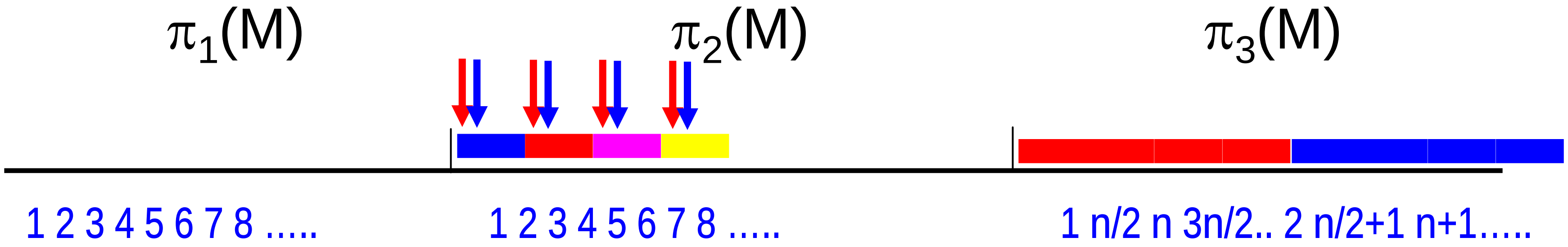
Example IV - 3 successive permutations

- $E(M) = \pi_1(M)\pi_2(M)\pi_3(M)$



Example IV - 3 successive permutations

- $E(M) = \pi_1(M)\pi_2(M)\pi_3(M)$

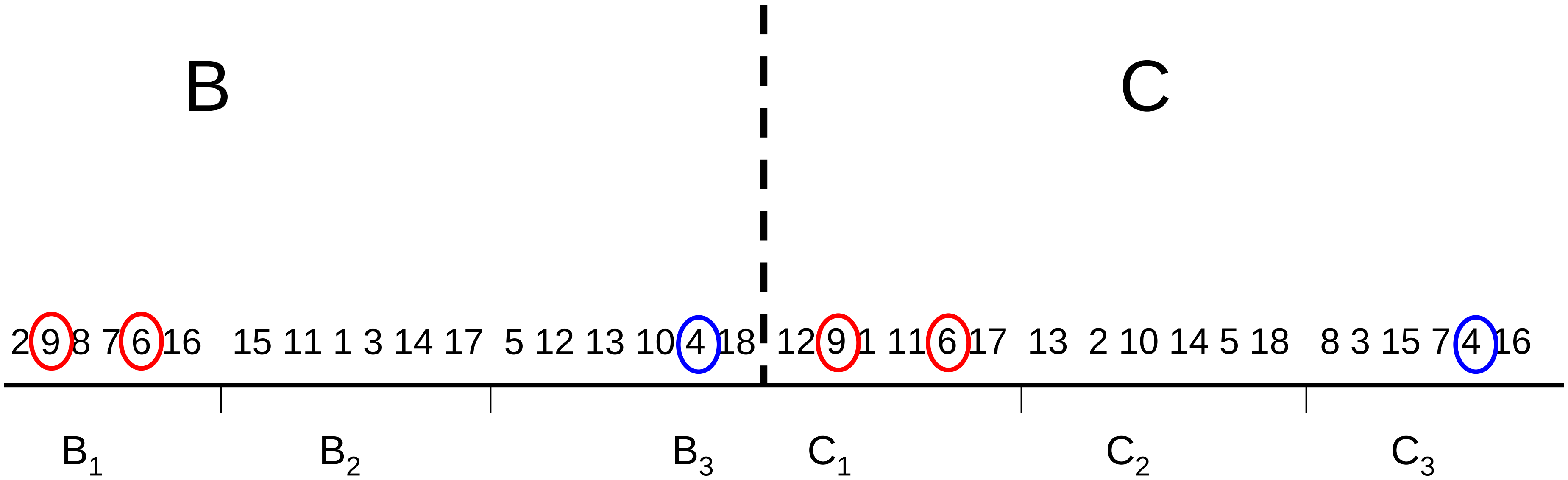


Proof of the 3-permutations case: Getting started

- Lemma 1:

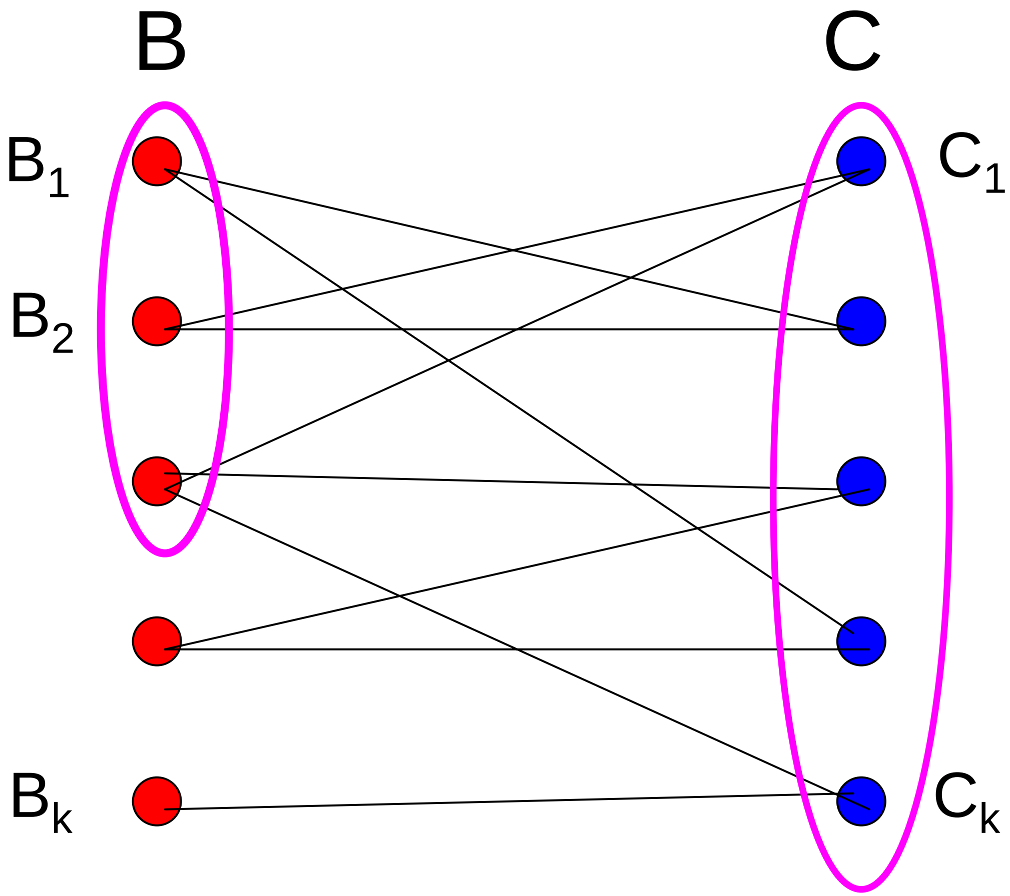
Let B and C be two permuted sequences of [L]. Divide B into k consecutive groups $B_1,\ldots,B_k$ and C into $C_1,\ldots,C_k$ of size n/k.
Then for x>0 and L¿ k^3x there exists a perfect matching of B_i's and C_j's such that $|B_i \cap C_j|$ ¿ x

Lemma 1

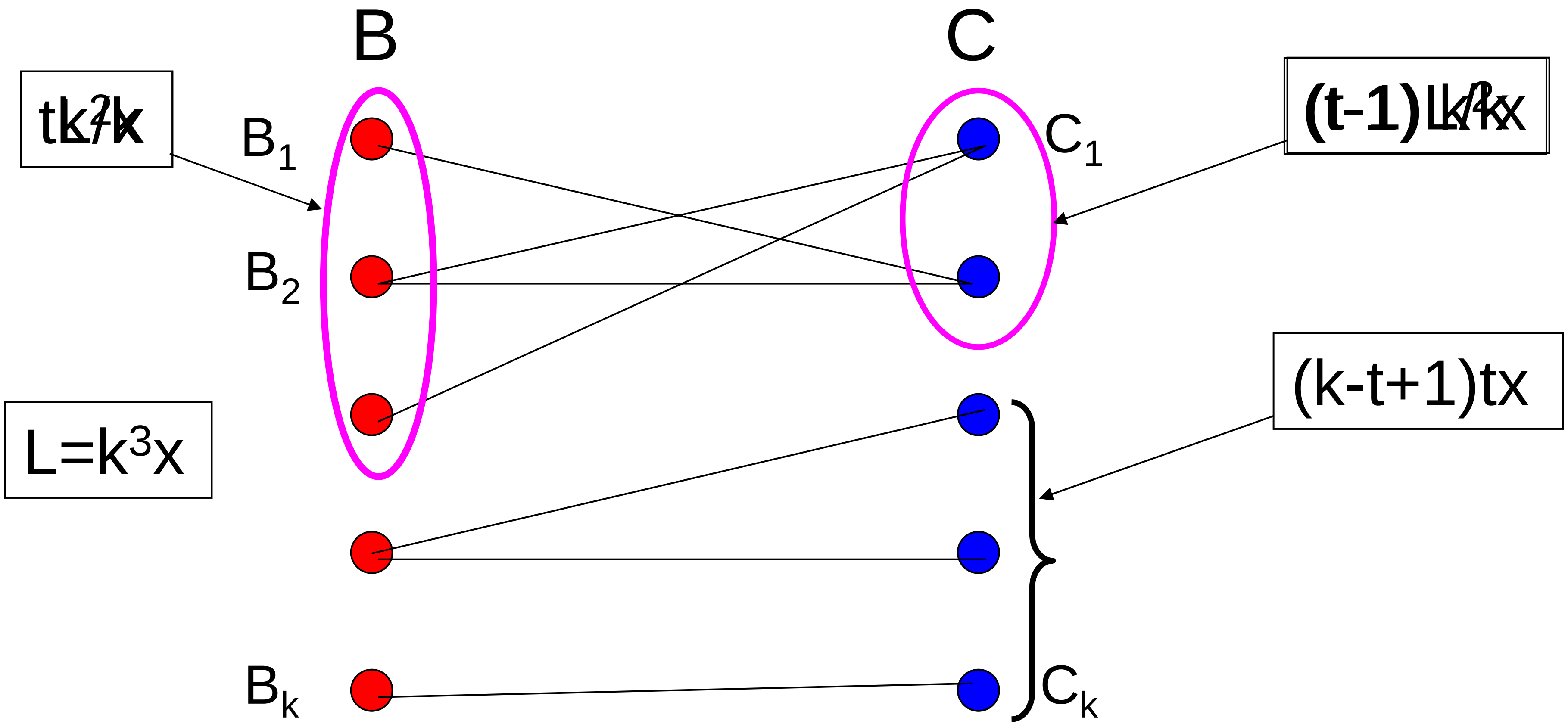


Given large sets - we expect the intersection between them to be large

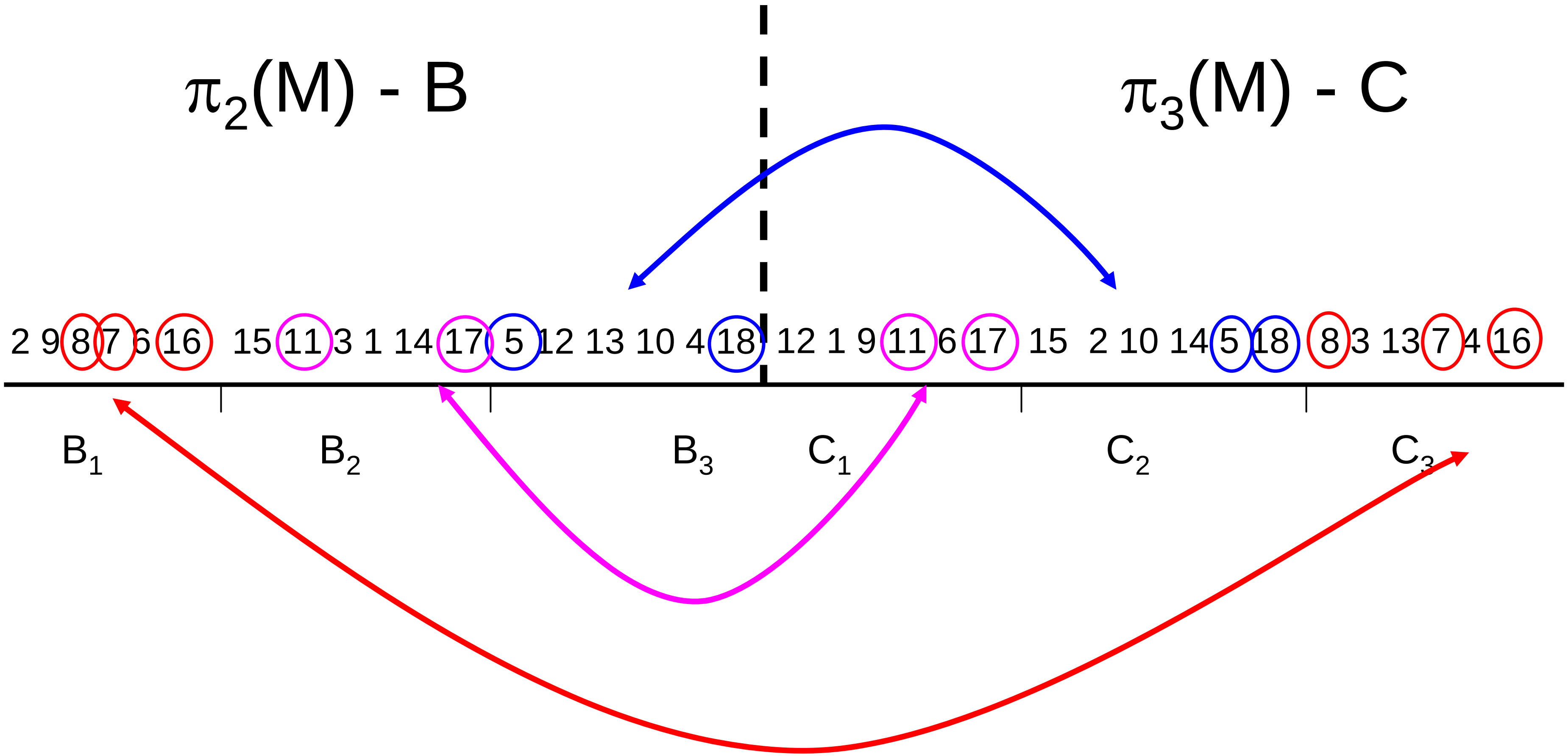
Lemma 1



Lemma 1



Lemma 1

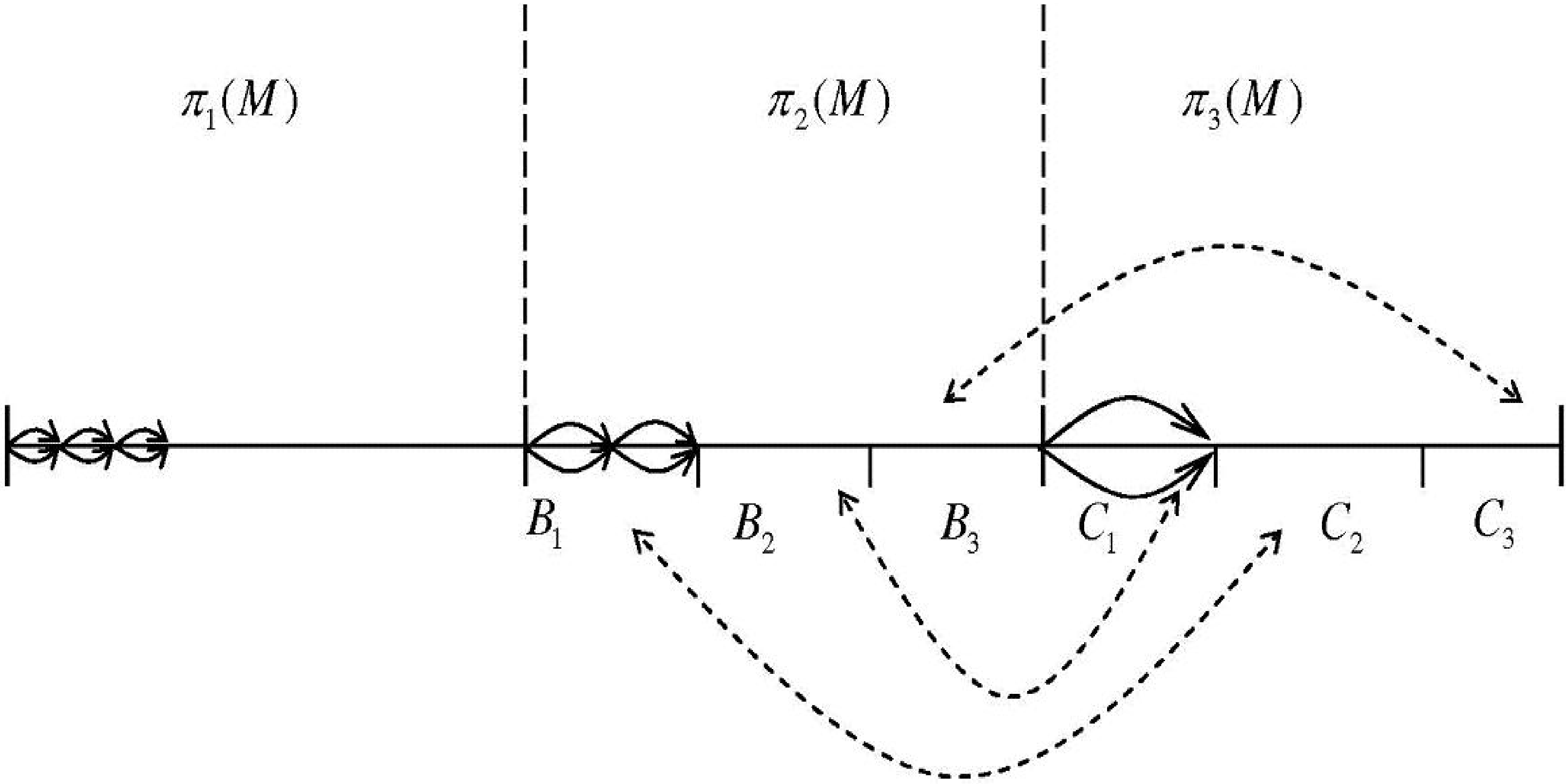


π2(M) - B
π3(M) - C
2 9 8 7 6 16 15 11 3 1 14 17 5 12 13 10 4 18
12 1 9 11 6 17 15 2 10 14 5 18 8 3 13 7 4 16
B1
B2
B3
C1
C2
C3

3 consecutive permutations

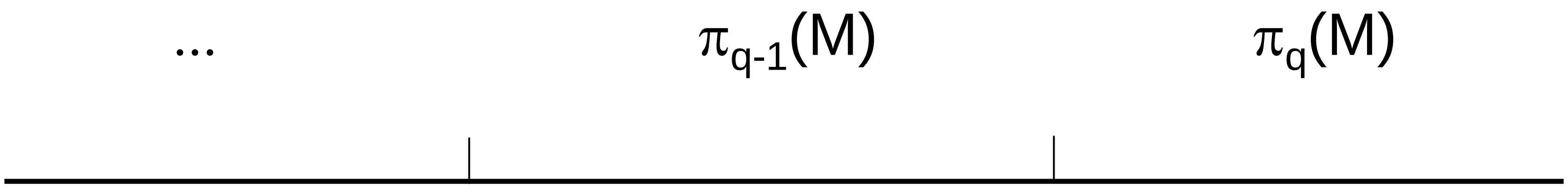
- Find a matching for $x=n^2/4$ in the last two permutations
- Set all non active message blocks to 0
- Build the multi-collision in 3 stages using larger blocks in each stage
- Requires a message of length $O(k^3n^2)$

3 successive permutations



Many successive permutations

- $E(M) = \pi_1(M)\pi_2(M)\ldots\pi_q(M)$



q consecutive permutations

- Find a matching for $x=O(n^{3(q-3)+2})$ in the last two permutations
- Set all non active message blocks to 0
- Find a matching for $x=O(n^{3(q-6)+2})$ in the two second to last permutations
- …
- Build the multi-collision in q stages using larger blocks in each stage
- Requires a message of length $O(k^3 n^{3(q-3)+2})$

Reduction from the general case

- So far proved for any constant number of permutations
- Reduction from general case to succesive permutations:
  - Choose a set of active message indices such that the resulting sequence is in successive permutations form

Case of expansion factor 2

- At least half the indices appear at most twice
- Given a sequence in which each index appears at most twice either
  - There exists a subset of variables which 'appears' once
  - There exists a subset of variables which are in successive permutation form

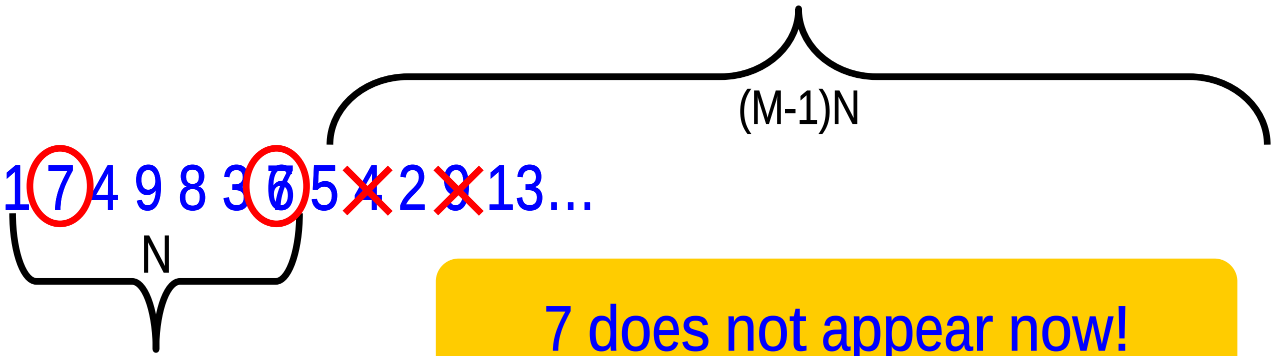
Case of expansion factor 2

- Lemma: for any 2-sequence over 1..l where l=MN either
  - There exists a subset of M variables which 'appears' once
  - There exists a subset of N variables which are in successive permutation form

Case of expansion factor 2

Case 2 : N elements appear in concatenated permutation form

- Proof: by induction on l=MN



7 does not appear now!

If each element appears at most once we are done!!

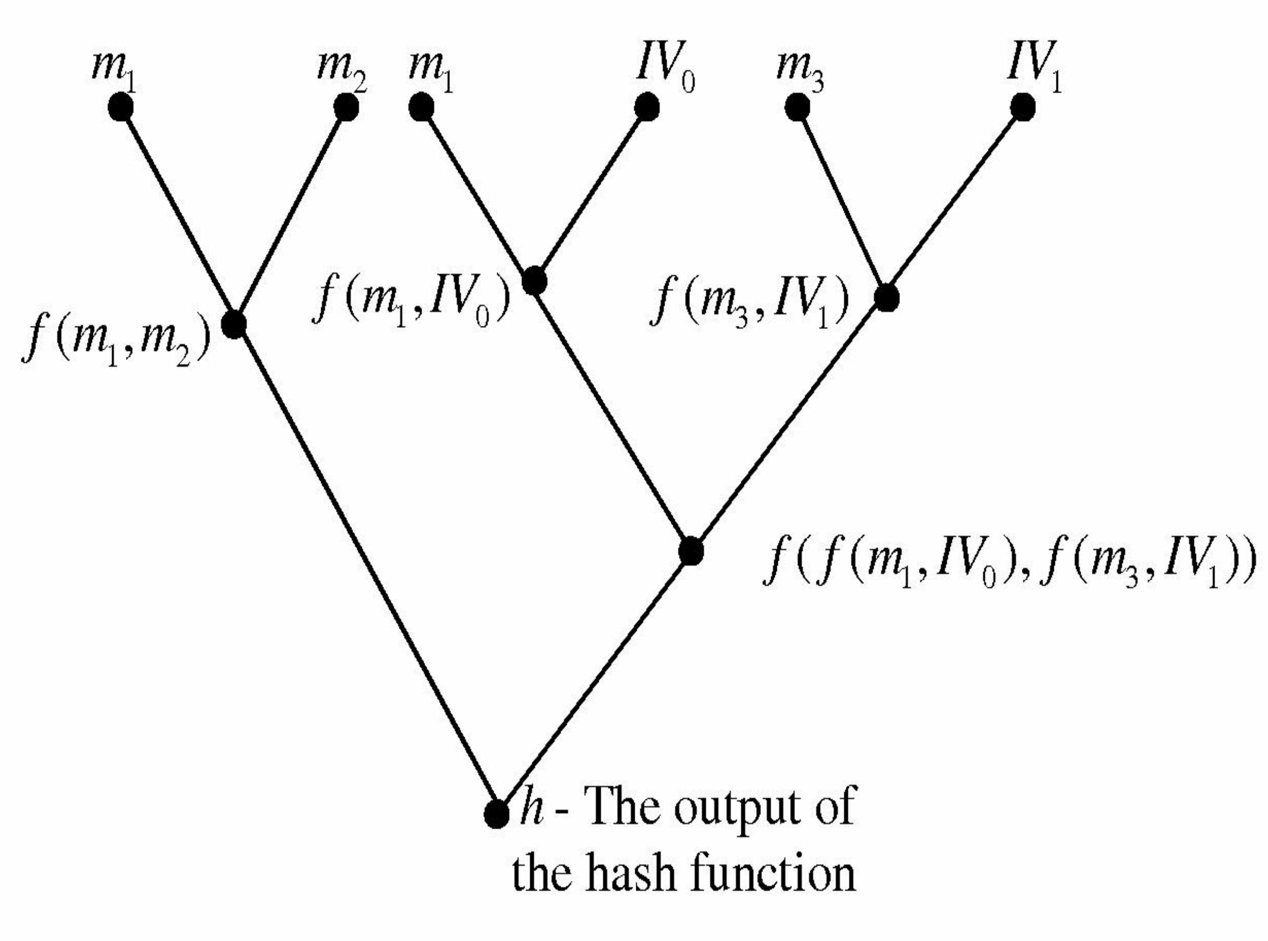
General Case

- At least half the indices appear at most twice the expansion rate e
- Given a sequence in which each index appears at most 2e either
  - There exists a subset of variables which 'appears' once
  - There exists a subset of variables which are in successive permutation form
- We already solved the successive permutation case

General Case

- If the message expansion expands by a constant factor e (by duplicating message blocks) can find a 2^k multicollision in time $2^{n/2}C(n,k,e)$ where $C(n,k,e)$ is polynomial in n, k but exponential in e)

Example of an Tree Based Hash function



Further research

- Other message expansion procedures
  - Linear combinations
  - LFSRs
  - …
- Keyed hash functions
- Tree based hash functions
- Other uses of multicollisions