楽しいMD5 Coll

KMCID:taisei

やること

- 一方向ハッシュ関数の前提知識
- MD5

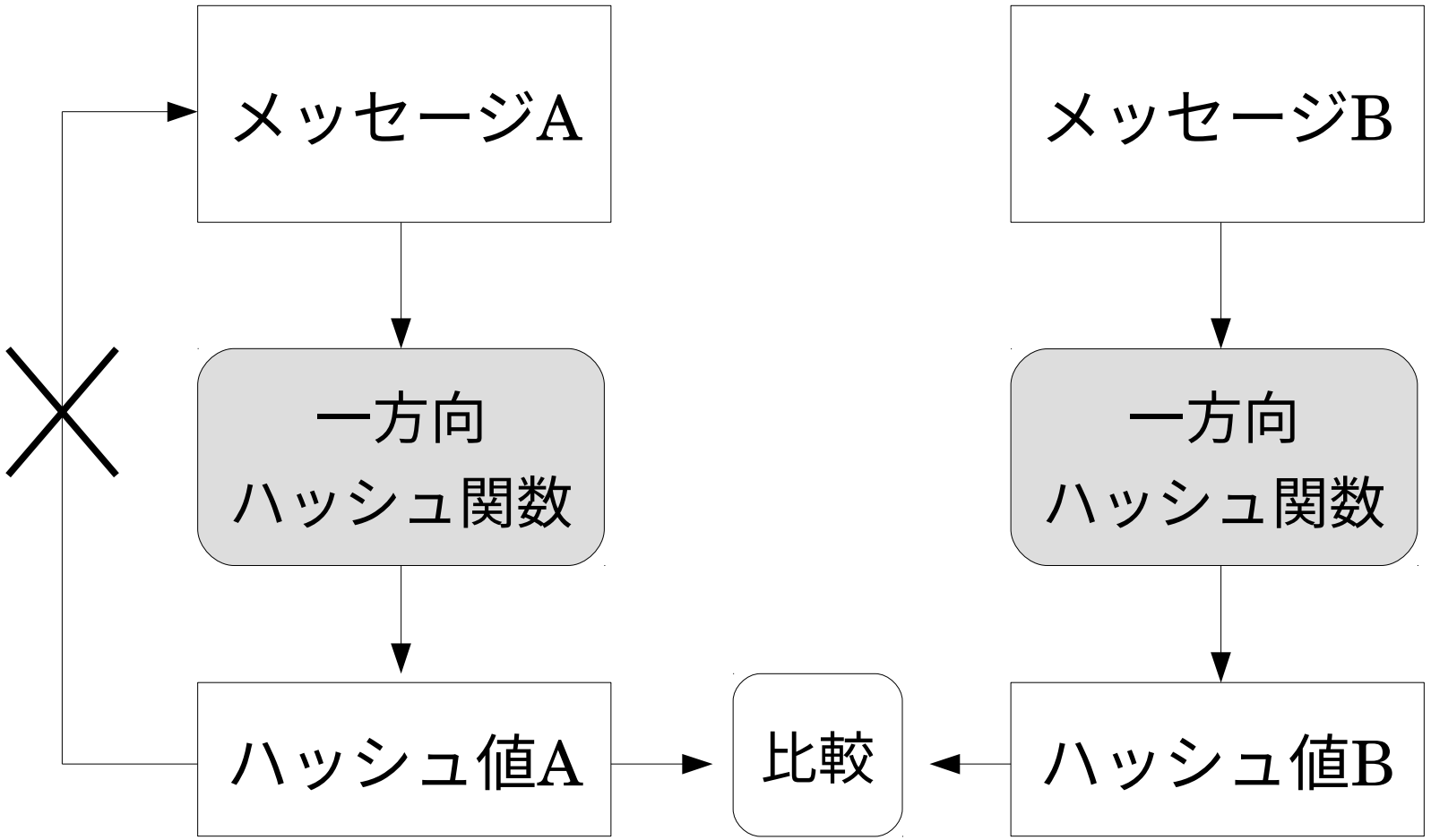
一方向ハッシュ関数

一方向ハッシュ関数

- 任意の入力を固定長のハッシュ値に変換する関数

- メッセージの正真性を検証するのに使われる
  - ハッシュ値が衝突するような異なるメッセージの組を見つけるのが非常に困難である必要がある
    (衝突が存在しないハッシュ関数は存在しない)

一方向ハッシュ関数



一方向ハッシュ関数

- 用途
  - ソフトウェアの改竄検出
  - パスワードを元にした暗号化(PBE)
  - デジタル署名
  - 擬似乱数生成器

一方向ハッシュ関数

- 弱衝突耐性
  - そのハッシュ値を持つ別のメッセージを見つけ出すのが困難
- 強衝突耐性
  - ハッシュ値が一致するような異なる2つのメッセージを見つけ出すのが困難

一方向ハッシュ関数

- 一方向ハッシュ関数は強衝突耐性をもつ必要 がある
  - 強衝突耐性がないと、ハッシュ値が一致するように メッセージを改竄できるかもしれない

一方向ハッシュ関数

- 原像攻撃
  - あるハッシュ値をもつメッセージを探索する
- 第二原像攻撃
  - あるメッセージと同じハッシュ値を持つ別のメッセージを探索する
- どちらも弱衝突耐性を破ろうとする攻撃

誕生日のパラドックス

- 誕生日が同一の2人が50%以上の確率で存在するには、何人集めればよいでしょうか?

誕生日のパラドックス

- 誕生日が同一の2人が50%以上の確率で存在するには、何人集めればよいでしょうか?

- 23人いればよい
- H個の値の集合からn個を無造作に選んだとき、同じ値が2度以上選ばれる確率が50%以上になるには、Hが大きいときおよそ$n \fallingdotseq \sqrt{H}$

一方向ハッシュ関数

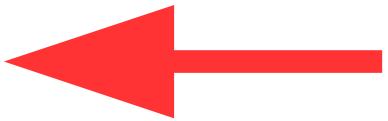
- 誕生日攻撃 (衝突攻撃)
  - 同じハッシュ値を持つ2つのメッセージを探索する
  - 強衝突耐性を破ろうとする攻撃
  - 誕生日のパラドックスに由来
  - 誕生日攻撃の試行回数は原像攻撃より
    ずっと少なくて済む

一方向ハッシュ関数

- MD4
- MD5
- SHA-1
- SHA-2
- SHA-3

一方向ハッシュ関数

- MD4
- MD5
- SHA-1
- SHA-2
- SHA-3



今回はこれで
遊びます

MD5

MD5 (Message Digest Algorithm 5)

- 任意のbit列に対し、128bit=16byteのbit列を出力する関数

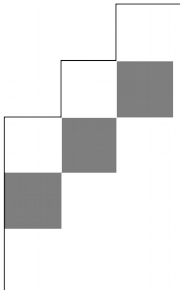
- 鍵空間(=鍵のとりうるパターン)は2^{128}通り
  - IPv6のアドレス空間と同じ
  - そこら辺の石ころに割り振っても尽きない程多い

- Linux系 … md5sum
- BSD系 … md5

MD5 (Message Digest Algorithm 5)

- 誕生日攻撃を50%以上の確率で成功させるには
  2^{64}通りの探索が必要
  - 普通のパソコンでは多く見積もって2^{30}通り/sしか探索できないので、このペースでも120年くらいかかる

- 数千万くらい投じられる資金力があれば何とかなるかもしれない



実演

- 次の実行ファイルをみてみましょう

実演

- 次の実行ファイルをみてみましょう
- md5の結果は同じ
  - 2つの実行ファイルは同じっぽい
- 実行結果は異なる
  - !???
- SHA-1の結果は異なる
  - !????????

実演

- 全く同じ実行ファイルで結果が異なるものを作ることは一応可能
  - ファイル名依存で挙動を変える等をする
- SHA-1は異なるので、2つの実行ファイルは絶対に異なる
- md5は同じなのは、ハッシュ値が衝突している
  - つまり、この2つのファイルはmd5への衝突攻撃が成功した例である

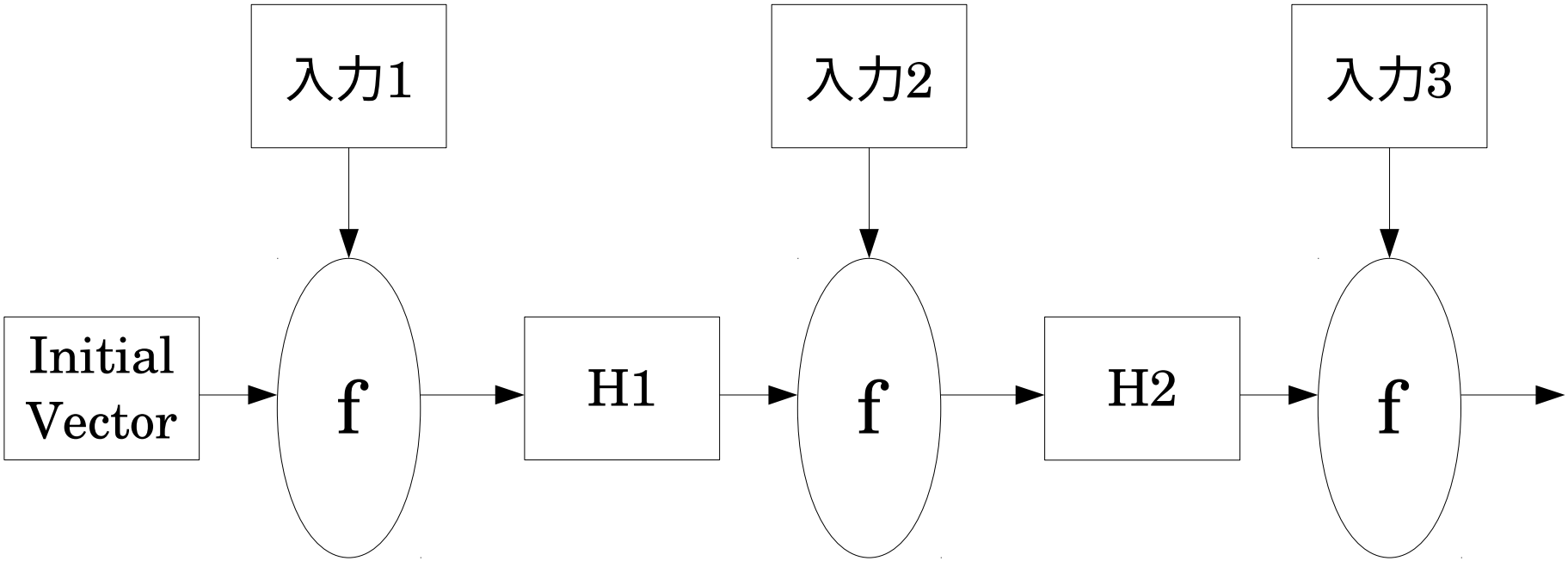
md5 collision Attack

- 2004年にWangという中国の無名女性研究者が最初に発表した

- 当初は2^{37}乗の探索だったが、その後高速化されて今では2^{29}乗で済む
  - 普通のパソコンでも1分かからない

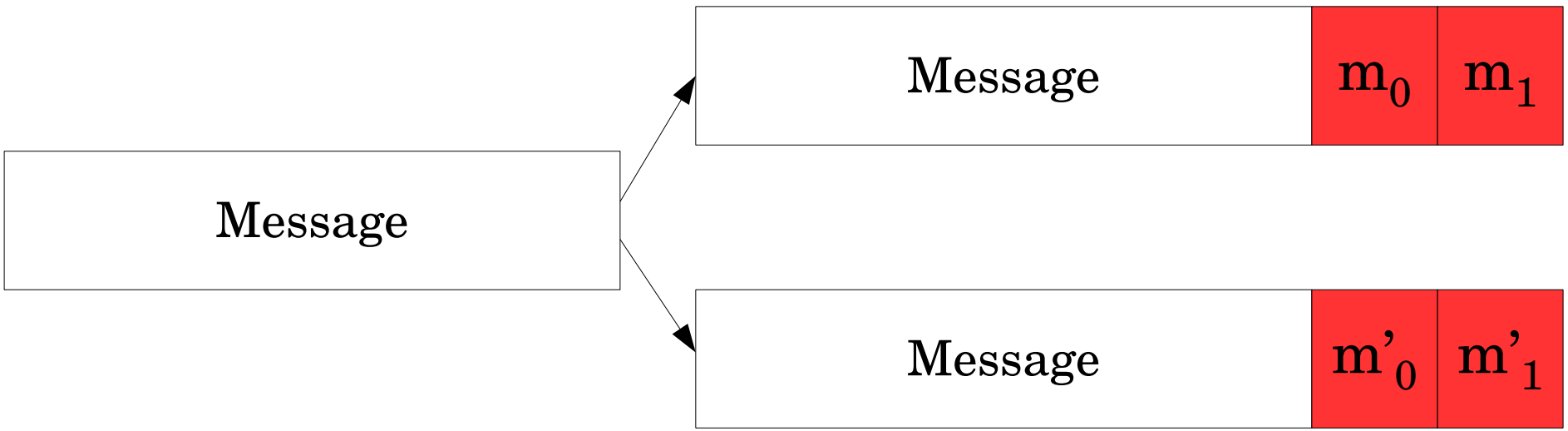
md5のしくみ

- 入力を512bit毎のブロックに分け、128bitの内部状態とともに関数fに適用することを繰り返す
- 内部状態はそのままハッシュ値として出力される
  - Merkle-Damgård constructionと呼ばれる



md5 collision Attack

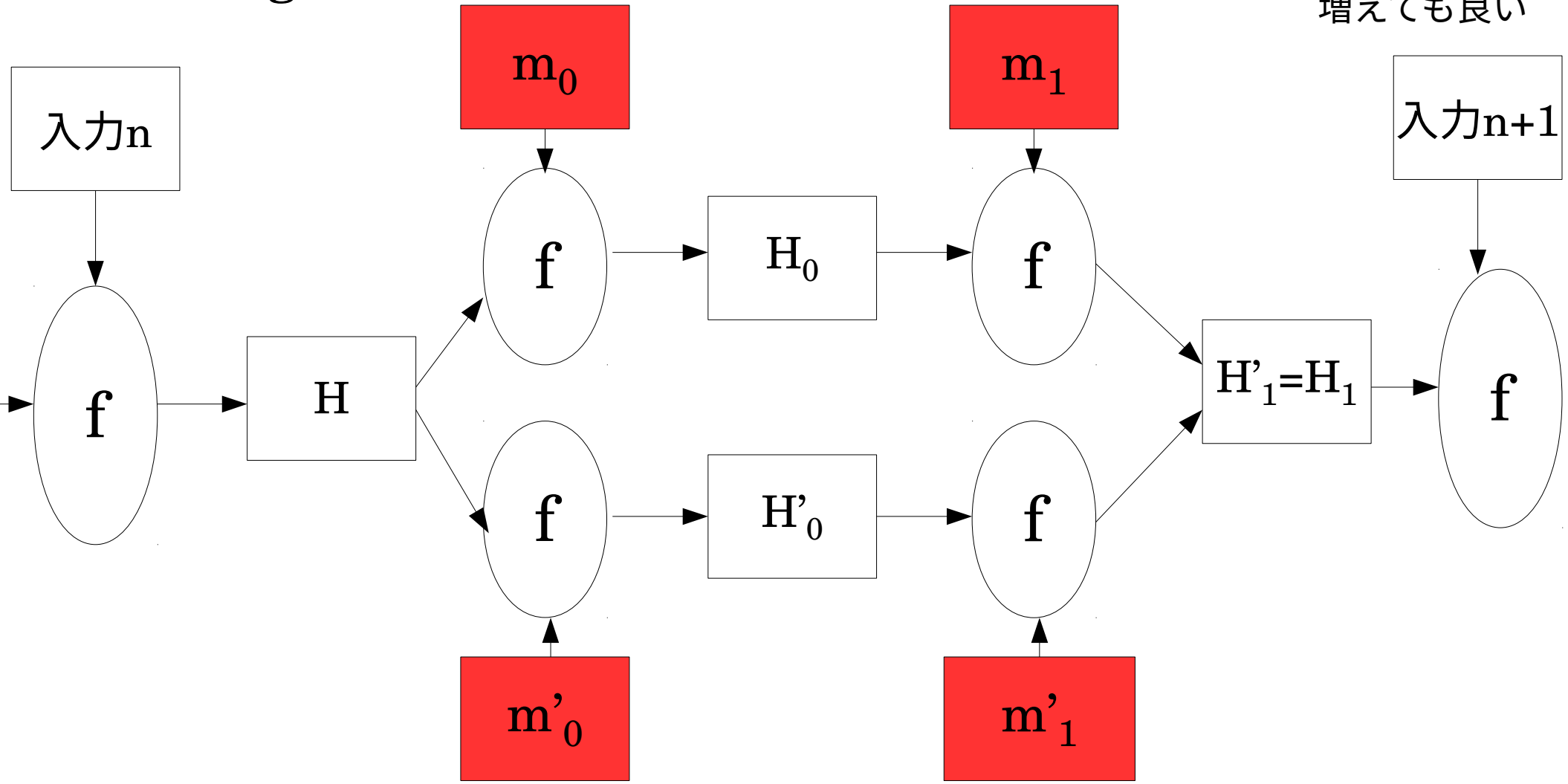
- Wangの攻撃では、あるメッセージに、異なる2種類の2ブロックを付け足し、2ブロックが付け足された2種類のメッセージを生成する



md5 collision attack

- Wangの衝突攻撃

その後に
メッセージが
増えても良い



md5 collision Attack

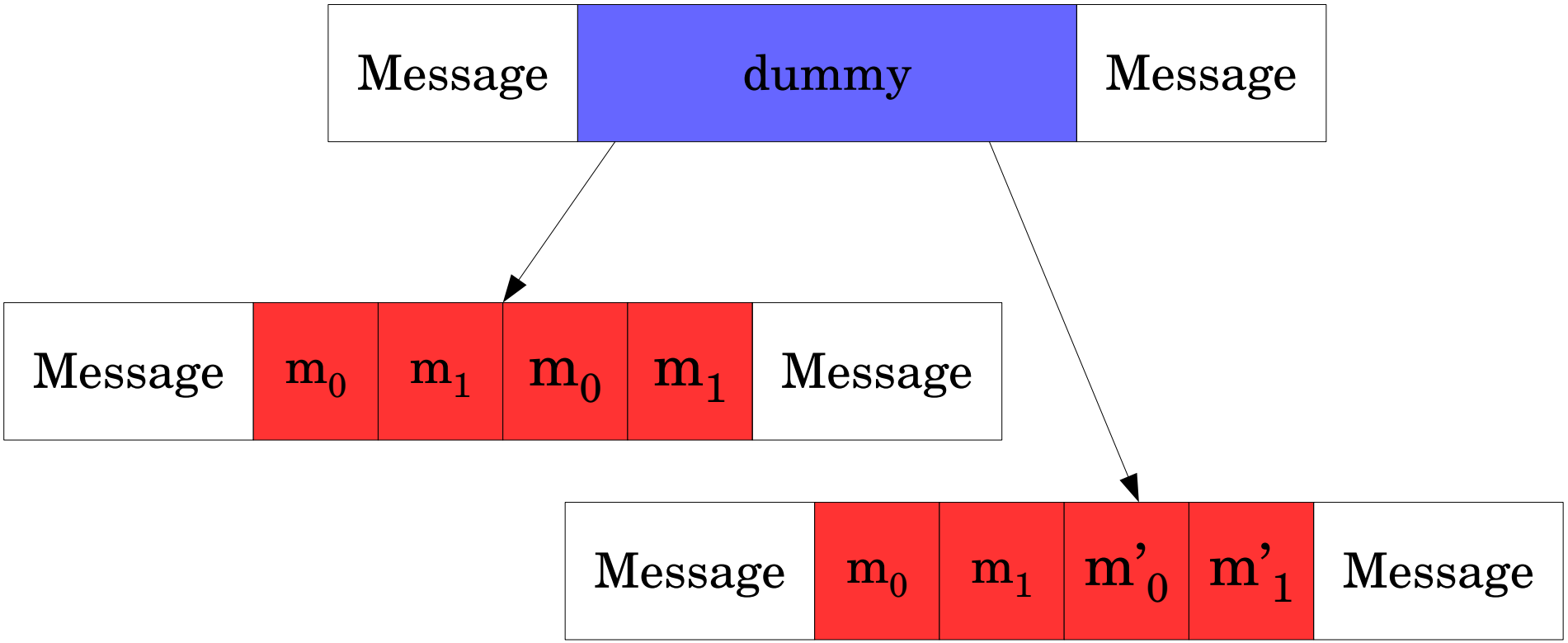
- つまるところ、差分攻撃
  - 入力を少し変化させて、結果の変化の仕方の偏りを解析
- ある差分を与えると、内部状態の差分がある値になりやすいという傾向を利用
- 1ブロックで一定の差分を作り、もう1ブロックで差分を打ち消す!!!

やってみよう?

- 今回使ったのはこのサイトのツール
  - www.mscs.dal.ca/~selinger/md5collision/
  - Download: evilize-0.2.tar.gz から
  - 高速化されていないので3時間くらいかかる
  - GPL

- プログラムにダミーの文字列がつまっていて、
  その文字列を比較して処理を分岐

やってみよう?


Message
dummy
Message
Message
m_0
m_1
m_0
m_1
Message
Message
m_0
m_1
m'_0
m'_1
Message

やってみよう?

- 高速なmd5衝突探索にはFastCollがある
  - https://marc-stevens.nl/research/
  - ページ一番下のSoftwareのfastcollから
  - 一分かからないことも

- ビルド手順
  - apt-get install libboost-all-dev lib32z1-dev libbz2-dev
  - g++ *.cpp *.hpp -lboost_program_options -lboost_filesystem -lboost_system

SHA-1、壊れる

- 2017年2月24日にGoogleがぶつけた
  - https://shattered.io
  - 2^{63}を探索したらしい
- 90日経ったらコードを公開してくれる(まだ)
- SHA-1は1ブロック64byteだが、衝突した2つのpdfの差分は大体128byteで、手法はMD5に同じ

参考文献

- MD5の仕様
  - ja.wikipedia.org/wiki/md5
  - www.ipa.go.jp/security/rfc/RFC1321JA.html

- MD5 Collision Attack
  - How to Break MD5 and Other Hash Functions. Xiaoyun Wang and Hongbo Yu.
    http://merlot.usc.edu/csac-f06/papers/Wang05a.pdf

  - On Collision For Md5. M.M.J Stevens.
    http://www.win.tue.nl/hashclash/On Collisions for MD5 – M.M.J. Stevens.pdf

参考文献

- MD5 Collision Attack
  - Katagaitai CTF勉強会 #5
    https://www.slideshare.net/trmr105/katagaitai-ctf-5-crypto

- SHA-1 Collision Attack
  - Google SHA-1のはなし
    https://www.slideshare.net/herumi/googlesha1

- 一方向ハッシュ関数
  - 結城 浩(2015)『暗号技術入門 第三版 秘密の国のアリス』
    SBクリエイティブ.