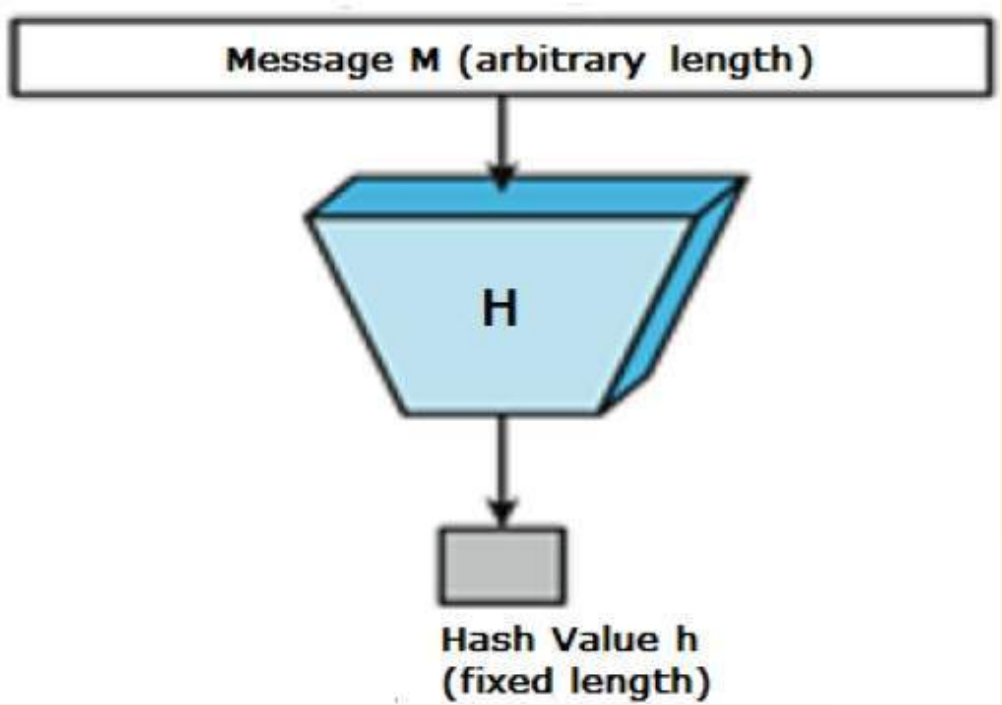
Hashing

- CEG 6430/4430 Cyber Network Security
- Junjie Zhang
- junjie.zhang@wright.edu
- Wright State University

Hashing

- Hashing is a function that takes as input a message of arbitrary length and generates as output a message of fixed length.



Hashing

- A hash function needs to satisfy three properties simultaneously
  - Deterministic: the same input -> the same output
  - One-Way: it is computationally infeasible to find the input given an output.
  - Collision Resistant: it is computationally infeasible to find two inputs that lead to the same an output.

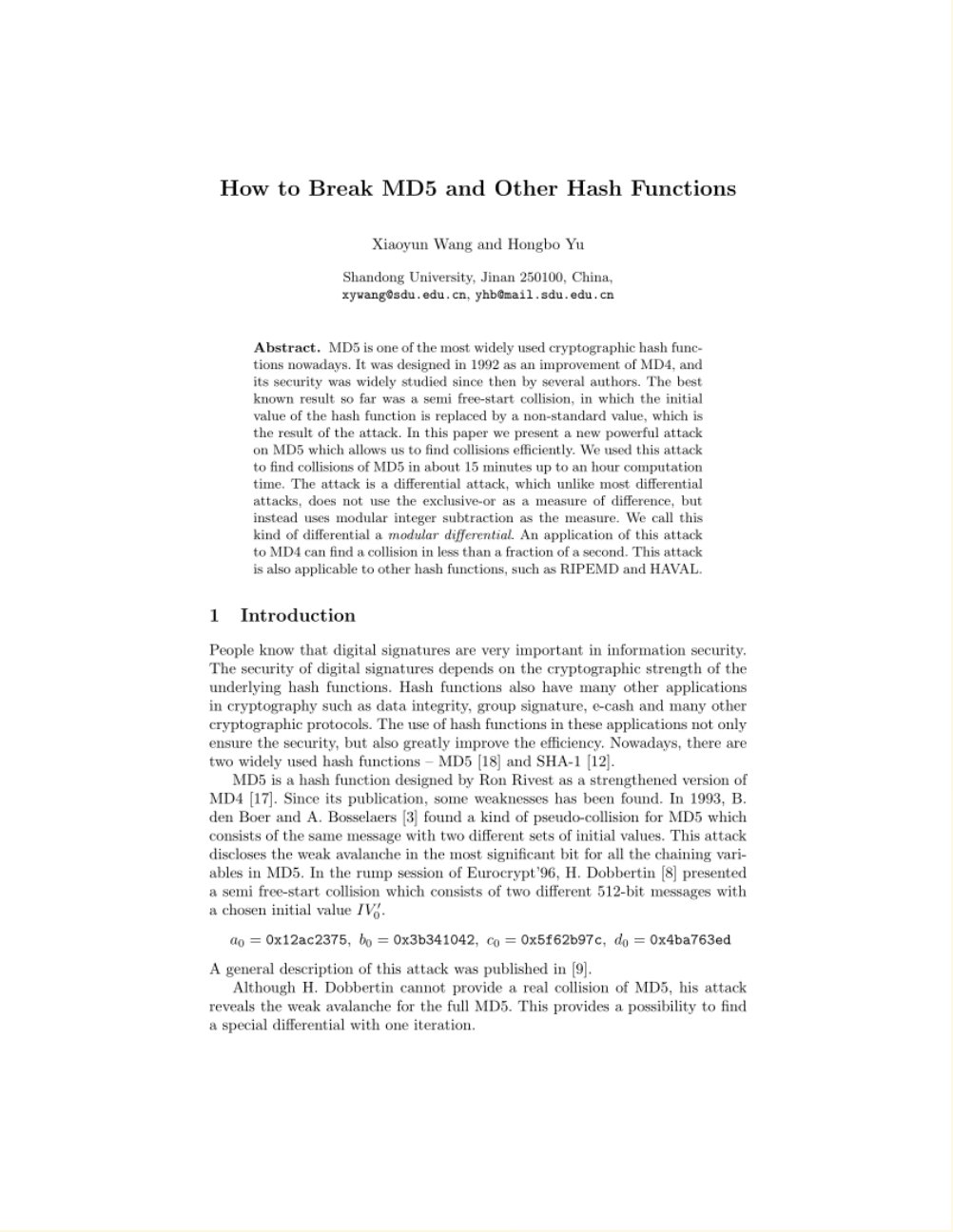
Quizz

Is $f(x)$ a hash function?

$$f(x) = x \ mod \ 16$$

Algorithms

- MD5 128 bits
- SHA/SHA1 160 bits
- SHA2 family:
  - SHA-224 224 bits
  - SHA-256 256 bits
  - SHA-384 384 bits
  - SHA-512 512 bits



How to Break MD5 and Other Hash Functions

Xiaoyun Wang and Hongbo Yu

Shandong University, Jinan 250100, China,
xywang@sdu.edu.cn, yhb@mail.sdu.edu.cn

Abstract. MD5 is one of the most widely used cryptographic hash functions nowadays. It was designed in 1992 as an improvement of MD4, and its security was widely studied since then by several authors. The best known result so far was a semi free-start collision, in which the initial value of the hash function is replaced by a non-standard value, which is the result of the attack. In this paper we present a new powerful attack on MD5 which allows us to find collisions efficiently. We used this attack to find collisions of MD5 in about 15 minutes up to an hour computation time. The attack is a differential attack, which unlike most differential attacks, does not use the exclusive-or as a measure of difference, but instead uses modular integer subtraction as the measure. We call this kind of differential a *modular differential*. An application of this attack to MD4 can find a collision in less than a fraction of a second. This attack is also applicable to other hash functions, such as RIPEMD and HAVAL.

1 Introduction

People know that digital signatures are very important in information security. The security of digital signatures depends on the cryptographic strength of the underlying hash functions. Hash functions also have many other applications in cryptography such as data integrity, group signature, e-cash and many other cryptographic protocols. The use of hash functions in these applications not only ensure the security, but also greatly improve the efficiency. Nowadays, there are two widely used hash functions – MD5 [18] and SHA-1 [12].

MD5 is a hash function designed by Ron Rivest as a strengthened version of MD4 [17]. Since its publication, some weaknesses has been found. In 1993, B. den Boer and A. Bosselaers [3] found a kind of pseudo-collision for MD5 which consists of the same message with two different sets of initial values. This attack discloses the weak avalanche in the most significant bit for all the chaining variables in MD5. In the rump session of Eurocrypt'96, H. Dobbertin [8] presented a semi free-start collision which consists of two different 512-bit messages with a chosen initial value IV_0'.

$a_0 = \texttt{0x12ac2375},\ b_0 = \texttt{0x3b341042},\ c_0 = \texttt{0x5f62b97c},\ d_0 = \texttt{0x4ba763ed}$

A general description of this attack was published in [9].

Although H. Dobbertin cannot provide a real collision of MD5, his attack reveals the weak avalanche for the full MD5. This provides a possibility to find a special differential with one iteration.

MD5

```
[jzhang@DESKTOP-DSVPHPI system32]$echo "I am here"| md5sum
2551b4c1e373e5fa322e84dd4c74de05  -
[jzhang@DESKTOP-DSVPHPI system32]$echo "I am there"| md5sum
d069c6fc7fe8802b32fd964812c01320  -
```

Hashing and Integrity

Hashing itself cannot protect integrity. For example, it can be used as the checksum of a message transmited over the Internet.

[data|hash(data)]

If the attacker can alter the data, attacker can also recalculate the hash value using modified data.