哈希函数的设计原则

Ivan Bjerre Damgård,

翻译：李晓峰 (cy_lxf@163.com)[‡]

V1.0

2025 年 5 月 9 日

摘要

我们证明：若存在从 m 比特到 t 比特（m>t）的计算上抗碰撞函数 f，则存在计算上抗碰撞的函数 h，可将任意多项式长度的消息映射到 t-比特字符串。

设消息长度为 n。函数 h 的构造方式有两种：

单处理器线性时间：计算时间复杂度为 O(n)。

多处理器对数时间：使用 O(n) 个处理器时，时间复杂度为 O(log(n))（每次 f 的调用计为一步）。此外，对于任意常数 k 和较大的 n，使用 k 个处理器可将第一种构造的速度提升 k 倍。

本文不仅提出了哈希函数的通用设计原则，还统一了此前提出的几种看似无关的哈希构造方法，并建议对其他方案进行修改以简化安全性证明。

我们给出了三个具体构造示例，分别基于：模平方运算、Wolfram 的伪随机比特生成器 [Wo]、背包问题

I 引言与相关工作

抗碰撞哈希函数 h 的定义是：将任意长度消息映射到固定长度的字符串，且找到满足 h(x)=h(y) 的 $x \neq y$ 是困难问题。本文关注公开可计算的哈希函数（即无需密钥控制的函数）。

此类函数在数据完整性保护和数字签名中的应用已被广泛认知（参见 [Da]、[De]、[DP]）。已有多种哈希构造方案被提出，例如基于 DES[Wi][DP]、

[*]译者目前为北京联合大学智慧城市学院信息安全老师。

[†]译文来自于经典文献翻译项目 https://gitee.com/uisu/InfSecClaT, 欢迎大家加入经典翻译项目，为更多的人能够获取这些经典文献所传递信息做一点贡献。

RSA[DP][Gir] 等。其中，[Da] 首次提出可证明抗碰撞的构造（假设底层原子操作的安全性，如模平方的单向性）。但此类方案效率较低（计算时间与用 RSA 处理整个消息相当），因此如何构造既高效又可证明抗碰撞的哈希函数仍是一个开放问题。

现有构造的安全性证明困难源于消息长度增加导致复杂度上升。本文提出一种新构造方法，通过缩短消息长度 1 比特的抗碰撞能力，即可实现任意长度消息的哈希。该原则为设计新哈希函数和改进现有方案提供了指导。

本文构造与 Merkle 的“元方法”[Me] 类似，但通过额外设计使得形式化证明无需额外假设。此外，与 Naor-Yung[NaYu] 的构造相比，本文方案更高效，因其无需为每个消息块选择新的独立哈希实例。

近期，Impagliazzo-Naor[ImNa] 证明了基于背包问题的哈希构造（与本文第 4.3 节类似）在 [NaYu] 定义下的安全性。

II 预备知识

Definition 1 (固定长度抗碰撞哈希函数族) 设 $\mathcal{F} = \{F_m\}$ 为函数族,$t(m) < m$。每个 $f \in F_m$ 满足：

1. 实例生成：存在多项式时间算法生成随机实例 f
2. 高效计算：对任意输入 x，$f(x)$ 可在多项式时间内计算
3. 抗碰撞性：任何多项式时间算法找到碰撞的概率可忽略

Lemma 1 若 $\mathcal{F}$ 是抗碰撞函数族，则对随机选择的 x，任何算法反转 $f(x)$ 的成功概率不超过 $1/2 + 1/poly(m)$。若 f 的像分布均匀或 $m{-}t = O(m)$，则反转概率可忽略。

Definition 2 (抗碰撞哈希函数族) **抗碰撞哈希函数族** $\mathcal{H}$ 是一个由有限集合构成的无限族 $\{H_m\}_{m=1}^{\infty}$，以及一个多项式有界函数 $t : \mathbf{N} \to \mathbf{N}$。

- H_m 中的成员是一个函数 $h : \{0,1\}^* \to \{0,1\}^{t(m)}$，称为 $\mathcal{H}$ 的**大小为 m 的实例**。

$\mathcal{H}$ 需满足以下条件：

1. **实例生成**：给定 m，存在概率多项式时间（以 m 为参数）算法 Θ，输入 m 后随机选择 $\mathcal{H}$ 中大小为 m 的实例。

2. **高效计算**：对于任意实例 $h \in H_m$ 和输入 $x \in \{0,1\}^*$，$h(x)$ 可高效计算，即在 m 和 $|x|$ 的多项式时间内完成。

3. **抗碰撞性**：对于按条件 *1* 随机选择的实例 $h \in \mathcal{H}$，难以找到 $x, y \in \{0,1\}^*$ 使得 $h(x) = h(y)$ 且 $x \neq y$。

   形式化表述：对任意概率多项式时间算法 Δ 和多项式 P，定义满足以下条件的实例集合：

$$\left\{ h \in H_m \;\middle|\; \Pr_{\Delta}\left[{}^{x \neq y}_{h(x)=h(y)}\right] \geq \frac{1}{P(m)} \right\}$$

   设 $\epsilon(m)$ 为算法 Θ 选择此类实例的概率，则：

$$\forall \text{多项式} P,\ \epsilon(m) = o\left(\frac{1}{P(m)}\right)$$

注：定义 *2.2* 与定义 *2.1* 的核心区别在于，前者**不对输入长度施加限制**（仅隐含要求输入长度不超过多项式规模）。

III 基础构造

Theorem 1 (任意长度抗碰撞哈希构造) 设 $\mathcal{F}$ 为将 m 比特映射到 $t(m)$ 比特的**固定长度抗碰撞哈希函数族**，则存在一个**抗碰撞哈希函数族** $\mathcal{H}$，可将任意长度的字符串映射到 $t(m)$ 比特。

进一步地，对 $\mathcal{H}$ 中大小为 m 的实例 h 和长度为 n 的输入，其计算复杂度满足：

$$\text{计算步数} \leq \frac{n}{m - t(m) + 1} + 1 \quad \text{（单处理器，每次调用 } \mathcal{F} \text{ 计为 } 1 \text{ 步）}$$

证明概要：

1. 消息分块为 $m - t - 1$ 比特，填充至 512 比特倍数
2. 迭代处理：通过链式调用 f 更新状态 h_i
3. 抗碰撞归约：若存在碰撞 $h(x) = h(x')$，则必存在 f 的碰撞

Theorem 2 (并行抗碰撞哈希构造) 设 $\mathcal{F}$ 为将 m 比特映射到 $t(m)$ 比特的**抗碰撞函数族**，则存在一个**抗碰撞哈希函数族** $\mathcal{H}$，可将任意长度字符串映射到 $t(m)$ 比特，并满足以下性质：

对 $\mathcal{H}$ 中大小为 m 的实例 h 和输入长度 n，其计算复杂度为：

$$\text{步数} = O\left(\log_2\left(\frac{n}{t}\right) \cdot \frac{t}{m-t}\right)$$

所需处理器数为：

$$\text{处理器数} = \frac{n}{2t}$$

其中，每次调用 $\mathcal{F}$ 的函数计算计为 *1* 步。

参考文献

1. R. Merkle, “Secure communication over an insecure channel,” submitted to Communications of the ACM.

2. D. Kahn, The Codebreakers, The Story of Secret Writing. New York: Macmillan, 1967.

3. C. E. Shannon, “Communication theory of secrecy systems,” Bell Syst. Tech. J., vol. 28, pp. 656-715, Oct. 1949.

4. M. E. Hellman, “An extension of the Shannon theory approach to cryptography,” submitted to IEEE Trans. Inform. Theory, Sept. 1975.

5. W. Diffie and M. E. Hellman, “Multiuser cryptographic techniques,” presented at National Computer Conference, New York, June 7-10, 1976.

6. D. Knuth, The Art of Computer Programming, Vol. 2, Semi- Numerical Algorithms. Reading, MA.: Addison-Wesley, 1969.

7. –, The Art of Computer Programming, Vol. 3, Sorting and Searching. Reading, MA.: Addison-Wesley, 1973.

8. S. Pohlig and M. E. Hellman, “An improved algorithm for computing algorithms in GF(p) and its cryptographic significance,” submitted to IEEE Trans. Inform. Theorv.

9. M. V. Wilkes, Time-Sharing Computer Systems. New York: Elsevier, 1972.

l0. A. Evans, Jr., W. Kantrowitz, and E. Weiss, “A user authentication system not requiring secrecy in the computer,” Communications of the ACM, vol. 17, pp. 437-442, Aug. 1974.

11.G. B. Purdy, “A high security log-in procedure,” Communications of the ACM, vol. 17, pp. 442-445, Aug. 1974.

12.W. Diffie and M. E. Hellman, "Cryptanalysis of the NBS data encryption standard" submitted to Computer, May 1976.

13.A. V. Aho, J. E. Hopcroft, and J. D. Ullman, The Design and Analysis of Computer Algorithms. Reading, MA.: Addison- Wesley, 1974.

14.R. M, Karp, "Reducibility among combinatorial problems,"in Complexity of Computer Computations. R. E. Miller and J. W. Thatcher, Eds. New York: Plenum, 1972, pp. 855104.