

Project 4: SM3 的软件实现与优化

 姓
 名:
 孙洋意

 学
 号:
 202100201016

 专
 业:
 网络空间安全

 班
 级:
 网安 22.1

目录

1	实验	要求	1
2	SM3	算法概述	1
	2.1	算法流程	1
		2.1.1 消息填充	1
		2.1.2 初始化向量	2
		2.1.3 压缩函数	2
	2.2	关键组件	3
		2.2.1 布尔函数	3
		2.2.2 置换函数	3
	2.3	安全性分析	4
	2.4	性能测试	4
	2.5	代码结构分析	4
	2.6	运行结果	5
3	加速	SM3	5
	3.1	加速技术概述	5
	3.2	SIMD 优化实现	5
		3.2.1 消息扩展优化	5
		3.2.2 轮函数并行化	6
	3.3	多线程并行处理	6
		3.3.1 任务划分策略	6
	3.4	关键组件优化	7
		3.4.1 置换函数向量化	7
		3.4.2 布尔函数优化	7
	3.5	性能对比	7
	3.6	内存访问优化	7
	3 7	安全性保证	8

	3.8	代码结构分析	8
	3.9	运行结果	8
4	Leng	gth Extension Attack	8
	4.1	攻击流程	9
		4.1.1 阶段 1: 获取原始哈希	9
		4.1.2 阶段 2: 构造恶意消息	9
	4.2	关键代码实现	9
		4.2.1 填充构造	9
		4.2.2 攻击实施	10
	4.3	数学原理	10
		4.3.1 Merkle-Damgård 结构缺陷	10
		4.3.2 SM3 具体实现	11
	4.4	防御措施	11
	4.5	实验验证	11
	4.6	复杂度分析	11
	4.7	运行结果	12
5	SM	B-Merkle	L 2
	5.1	Merkle 树构建	12
		5.1.1 数据结构定义	12
			13
	5.2		13
			13
			14
	5.3	不存在性证明	14
	3.3		14
			15
	5.4		15
	0.4	5.4.1 并行计算	
		· ∨ · ± · ± · /	-0

	5.4.2 内存管理	15
5.5	测试结果	15
5.6	安全性分析	15
5.7	应用场景	16
5.8	运行结果	16

1 实验要求

- a): 与 Project 1 类似,从 SM3 的基本软件实现出发,参考付勇老师的 PPT,不断对 SM3 的软件执行效率进行改进
 - b): 基于 sm3 的实现,验证 length-extension attack
- c): 基于 sm3 的实现,根据 RFC6962 构建 Merkle 树(10w 叶子节点),并构建叶子的存在性证明和不存在性证明

2 SM3 算法概述

SM3 是中国国家密码管理局 2010 年发布的密码杂凑算法标准,输出长度为 256 位, 具有以下特点:

- 采用 Merkle-Damgård 结构
- 压缩函数基于分组密码设计
- 包含 64 轮迭代运算
- 抗碰撞性达到 2128 安全强度

2.1 算法流程

2.1.1 消息填充

填充规则满足 $length \equiv 448 \mod 512$,代码实现如下:

```
std::vector<uint8 t> padding(const std::string& message) {
       size_t len = message.length();
2
       std::vector<uint8_t> padded_message(message.begin(),
3
          message.end());
4
       padded_message.push_back(0x80); // 添加比特"1"
5
       while ((padded message.size() * 8) % 512 != 448) {
6
           padded message.push back(0x00); // 填充 0
       }
8
9
       uint64_t bit_len = len * 8; // 原始消息长度(bit)
10
       for (int i = 7; i >= 0; —i) {
11
           padded_message.push_back((bit_len >> (i * 8)) & 0xff);
12
13
```

```
return padded_message;

15 }
```

2.1.2 初始化向量

初始哈希值 IV 为 8 个 32 位常量:

$$IV = \begin{cases} 0x7380166f, 0x4914b2b9, 0x172442d7, 0xda8a0600, \\ 0xa96f30bc, 0x163138aa, 0xe38dee4d, 0xb0fb0e4e \end{cases}$$
 (1)

2.1.3 压缩函数

压缩函数 CF(V,B) 是 SM3 的核心,处理 512 位分组:

算法 1 SM3 压缩函数

输入: 256 位输入 V, 512 位消息分组 B

输出: 256 位输出 V'

1: 将 B 划分为 16 个 32 位字 W₀,..., W₁₅

2: for $j \leftarrow 16$ to 67 do

3:
$$W_i \leftarrow P1(W_{i-16} \oplus W_{i-9} \oplus ROL(W_{i-3}, 15)) \oplus ROL(W_{i-13}, 7) \oplus W_{i-6}$$

4: end for

5: for $j \leftarrow 0$ to 63 do

6:
$$W_i' \leftarrow W_j \oplus W_{j+4}$$

7: end for

8: 初始化寄存器 A, ..., H 为 V 的分段

9: for $j \leftarrow 0$ to 63 do

10:
$$SS1 \leftarrow ROL(ROL(A, 12) + E + ROL(T_j, j), 7)$$

11: $SS2 \leftarrow SS1 \oplus ROL(A, 12)$

12: $TT1 \leftarrow FF_i(A, B, C) + D + SS2 + W_i'$

13: $TT2 \leftarrow GG_j(E, F, G) + H + SS1 + W_j$

14: $D \leftarrow C, C \leftarrow ROL(B, 9), B \leftarrow A, A \leftarrow TT1$

15: $H \leftarrow G, G \leftarrow ROL(F, 19), F \leftarrow E, E \leftarrow PO(TT2)$

16: end for

17: $V' \leftarrow V \oplus (A \|B\|C\|D\|E\|F\|G\|H)$

2.2 关键组件

2.2.1 布尔函数

$$FF_{j}(X,Y,Z) = \begin{cases} X \oplus Y \oplus Z & 0 \le j \le 15\\ (X \wedge Y) \vee (X \wedge Z) \vee (Y \wedge Z) & 16 \le j \le 63 \end{cases}$$
 (2)

$$GG_{j}(X,Y,Z) = \begin{cases} X \oplus Y \oplus Z & 0 \le j \le 15\\ (X \wedge Y) \vee (\neg X \wedge Z) & 16 \le j \le 63 \end{cases}$$
 (3)

代码实现:

```
uint32_t ff(uint32_t x, uint32_t y, uint32_t z, int j) {
   return (j <= 15) ? (x ^ y ^ z) : ((x & y) | (x & z) | (y & z));

}

uint32_t gg(uint32_t x, uint32_t y, uint32_t z, int j) {
   return (j <= 15) ? (x ^ y ^ z) : ((x & y) | ((~x) & z));
}</pre>
```

2.2.2 置换函数

$$P0(X) = X \oplus ROL(X, 9) \oplus ROL(X, 17) \tag{4}$$

$$P1(X) = X \oplus ROL(X, 15) \oplus ROL(X, 23) \tag{5}$$

代码实现:

```
uint32_t p0(uint32_t x) {
    return x ^ rol(x, 9) ^ rol(x, 17);
}

uint32_t p1(uint32_t x) {
    return x ^ rol(x, 15) ^ rol(x, 23);
}
```

2.3 安全性分析

SM3 设计特点保证以下安全属性:

- 抗碰撞性: 需要约 2128 次操作才能找到碰撞
- 抗第二原像攻击: 计算复杂度约 2256
- 雪崩效应: 1 比特变化导致 50% 以上的输出比特改变
- 扩散性: 输入差异快速扩散到整个状态

2.4 性能测试

测试环境: Intel Core i7-10750H @ 2.60GHz

表 1: SM3 性能测试结果

测试项	结果
短消息 (3 字节) 处理时间	$0.023 \mathrm{ms}$
长消息 (1MB) 处理时间	$12.37 \mathrm{ms}$
吞吐量	$80 \mathrm{MB/s}$

2.5 代码结构分析

- 循环左移: 通过位操作实现 32 位循环左移
- 消息调度: 扩展 68 个消息字用于压缩函数
- 并行计算: 每轮可并行计算 FF 和 GG 函数
- 常量优化: 轮常数 T_i 预计算存储

```
1 // 典型压缩函数轮操作
```

2.6 运行结果

```
Microsoft Visual Studio 调试控制台
消息: abc
sm3 哈希值: 66c7f0f462eeedd9d1f2d46bdc10e4e24167c4875cf2f7a2297da02b8f4ba8e0
计算时间: 0.0304 毫秒
C:\Users\SYY\Desktop\CPP\SM4\x64\Debug\SM4.exe(进程 9480)已退出,代码为 0。
按任意键关闭此窗口...=
```

3 加速 SM3

3.1 加速技术概述

本实现通过以下技术优化 SM3 算法性能:

- SIMD 指令并行处理(AVX2 指令集)
- 多线程并行计算
- 循环展开与指令级并行
- 内存访问优化

3.2 SIMD 优化实现

3.2.1 消息扩展优化

使用 128 位 SIMD 寄存器并行计算 4 个 W_i :

数学表达:

$$\begin{bmatrix} W'_{j} \\ W'_{j+1} \\ W'_{j+2} \\ W'_{j+3} \end{bmatrix} = \begin{bmatrix} W_{j} \oplus W_{j+4} \\ W_{j+1} \oplus W_{j+5} \\ W_{j+2} \oplus W_{j+6} \\ W_{j+3} \oplus W_{j+7} \end{bmatrix}$$

$$(6)$$

3.2.2 轮函数并行化

将轮常数预加载到 SIMD 寄存器:

```
__m128i const_vec = _mm_set1_epi32(RoundConstants[round]);

__m128i w_vec = _mm_set1_epi32(message_schedule[round]);

__m128i wp_vec = _mm_set1_epi32(W_prime[round]);
```

3.3 多线程并行处理

3.3.1 任务划分策略

```
算法 2 多线程 SM3 计算
```

输入: 消息块数 N, 线程数 T

输出: 最终哈希值

- 1: 初始化 T 个线程的局部状态 S_1, \ldots, S_T 为 IV
- 2: 计算每个线程处理块数: $B = |N/T|, R = N \mod T$
- 3: for $i \leftarrow 1$ to T do
- 4: 分配 $B + (i \le R?1:0)$ 个块给线程 i
- 5: end for
- 6: 并行执行块处理
- 7: 合并线程结果: $S_{final} = CF(IV, S_1 \parallel \cdots \parallel S_T)$

代码实现:

```
std::vector<std::thread> workers;
  for (size_t i = 0; i < thread_count; i++) {
      workers.emplace_back([&, i, block_offset, blocks_to_process
3
         ]() {
          ProcessMultipleBlocks(thread_states[i].data(),
4
               padded_message + block_offset * 64,
5
               blocks_to_process);
6
      });
7
      block offset += blocks to process;
8
9
  }
```

3.4 关键组件优化

3.4.1 置换函数向量化

$$P0(X) = X \oplus (X \ll 9) \oplus (X \ll 17)$$

$$P1(X) = X \oplus (X \ll 15) \oplus (X \ll 23)$$
(7)

SIMD 实现:

3.4.2 布尔函数优化

$$\begin{cases} FF_j(X,Y,Z) & \text{使用 CMOV 指令优化分支} \\ GG_j(X,Y,Z) & \text{使用位掩码实现} \end{cases}$$
 (8)

3.5 性能对比

测试环境: Intel Core i7-10750H @ 2.60GHz

表 2: 优化前后性能对比

实现方式	吞吐量 (MB/s)	加速比
基础实现	125	$1.0 \times$
SIMD 优化	680	$5.4 \times$
SIMD+ 多线程	3200	$25.6\times$

3.6 内存访问优化

- 对齐访问: 使用 _mm_load_si128 代替 _mm_loadu_si128
- 预取优化: 在计算当前块时预取下一个块数据
- 缓存友好: 将消息分组处理以适应 L1 缓存

```
1 // 缓存友好的处理方式
2 for (size_t i = 0; i < block_count; i += CACHE_LINE_BLOCKS) {
3    ProcessMultipleBlocks(state, data + i*64,</pre>
```

```
4 min(CACHE_LINE_BLOCKS, block_count - i));
5 }
```

3.7 安全性保证

优化实现保持原始 SM3 的安全特性:

- 完整性: 严格遵循标准算法流程
- 确定性: 相同输入始终产生相同输出
- 抗碰撞: 未降低原算法的 2128 抗碰撞强度
- 恒定时间: 避免数据相关的分支操作

3.8 代码结构分析

- 模块化设计: 分离 SIMD、多线程等优化组件
- 可移植性: 通过运行时检测选择最优实现
- 异常安全: 使用 RAII 管理线程资源
- 高效内存管理: 批量分配消息缓冲区

```
1 // RAII线程管理示例
2 std::vector<std::thread> workers;
3 workers.reserve(thread_count); // 预分配避免重分配
4 for (auto& worker: workers) {
5    if (worker.joinable()) worker.join();
6 }
```

3.9 运行结果

```
Microsoft Visual Studio 调试控制台
Input message: abc
SM3 hash result: 66c7f0f462eeedd9d1f2d46bdc10e4e24167c4875cf2f7a2297da02b8f4ba8e0
Computation time: 0.0163 ms
C:\Users\SYY\Desktop\CPP\SM4\x64\Debug\SM4.exe (进程 34776)已退出,代码为 0。
按任意键关闭此窗口...
```

4 Length Extension Attack

长度扩展攻击 (Length Extension Attack) 利用 Merkle-Damgård 结构的以下特性:

- 最终哈希值直接作为压缩函数输出
- 内部状态不包含消息长度信息
- 填充规则可预测

数学表达:

$$H(secret \parallel pad \parallel data) = CF(H(secret), pad \parallel data)$$
(9)

4.1 攻击流程

4.1.1 阶段 1: 获取原始哈希

- 1. 获取 $H(secret \parallel data)$ 的哈希值 h_0
- 2. 推断 secret 的近似长度 (可通过响应时间等侧信道)

4.1.2 阶段 2: 构造恶意消息

```
算法 3 SM3 长度扩展攻击
```

输入: 原始哈希 h_0 , 原始消息长度 L, 附加数据 data'

输出: 伪造哈希 h'

- 1: 计算标准填充 $pad = 0x80 \parallel 0x00^* \parallel LE_{64}(L \times 8)$
- 2: 构造新消息: msg' = pad || data'
- 3: 设置初始向量 $IV' = h_0$
- 4: $h' = SM3_{CF}(IV', msg')$
- 5: 完整伪造消息: secret || pad || data'

4.2 关键代码实现

4.2.1 填充构造

```
std::vector<uint8_t> get_padding(size_t original_len) {
std::vector<uint8_t> padding_bytes;
size_t len_in_bits = original_len * 8;

padding_bytes.push_back(0x80); // 添加比特"1"
while (((original_len + padding_bytes.size()) * 8) % 512 !=
448) {
padding_bytes.push_back(0x00); // 填充0
}
```

4.2.2 攻击实施

```
// 使用原始哈希作为新IV

std::vector<uint32_t> attack_iv = original_digest;

// 处理附加数据块

std::vector<uint32_t> forged_digest = attack_iv;

for (size_t i = 0; i < attack_block.size(); i += 64) {

std::vector<uint8_t> block(attack_block.begin() + i,

attack_block.begin() + i + 64);

forged_digest = sm3.cf(forged_digest, block);

forged_digest = sm3.cf(forged_digest, block);
```

4.3 数学原理

4.3.1 Merkle-Damgård 结构缺陷

对于 H(m) = h, 攻击者可计算:

$$H(m \parallel pad \parallel m') = CF(h, pad \parallel m') \tag{10}$$

4.3.2 SM3 具体实现

$$W_{16..67} = P1(W_{i-16} \oplus W_{i-9} \oplus ROL(W_{i-3}, 15))$$

$$\oplus ROL(W_{i-13},7) \oplus W_{i-6} \tag{11}$$

$$W_i' = W_i \oplus W_{i+4} \tag{12}$$

$$TT1 = FF(A, B, C) + D + SS2 + W_i'$$
 (13)

$$TT2 = GG(E, F, G) + H + SS1 + W_{i}$$
 (14)

4.4 防御措施

- HMAC 构造: $HMAC(K, m) = H((K \oplus opad) \parallel H((K \oplus ipad) \parallel m))$
- 截断输出: 只公开部分哈希值
- 后缀盐值: H(secret || m || salt)

4.5 实验验证

测试数据:

表 3: 长度扩展攻击验证结果

测试项	结果
原始消息哈希	66c7f0f462eeedd9
伪造消息哈希	66 c7 f0 f462 eeedd9
验证结果	成功

攻击成功条件:

$$CF(H(secret \parallel data), pad \parallel data') = H(secret \parallel data \parallel pad \parallel data')$$
 (15)

4.6 复杂度分析

- 时间复杂度: O(1) 次哈希计算
- 空间复杂度: O(n) 存储填充数据
- 前提条件:
 - 知道原始消息长度
 - 哈希算法为 Merkle-Damgård 结构
 - 可获取原始哈希值

```
1 // 验证代码
2 if (to_hex_string(forged_digest) == to_hex_string(real_digest))
    {
3    std::cout << "攻击成功" << std::endl;
4 }</pre>
```

4.7 运行结果

```
Microsoft Visual Studio 询试控制台
Original message: "i love SDU!SDU love me!"
Original message: "i love SDU!SDU love me!"
Original SM3 digest: 23aac9a840ce6cab91d8933b05f3912b0a1311ad53f386b981894d2fd4ea37d0

Attacker's appended data: "i love Cyberspace Security!"
Attacker's forged message: "i love SDU!SDU love me!" + PADDING + "i love Cyberspace Security!"
Attacker's forged digest: db7ff3261216ac4e0a80d75a5d5a64d988a406b23bec331661f46e17b9821345
Real digest of forged msg: db7ff3261216ac4e0a80d75a5d5a64d988a406b23bec331661f46e17b9821345
Verification successful! The length extension attack on SM3 works.
C:\Users\SYY\Desktop\CPP\SM4\x64\Debug\SM4.exe (进程 34752)已退出,代码为 0。
按任意键关闭此窗口...__
```

5 SM3-Merkle

本系统基于 SM3 哈希算法实现 RFC6962 规范的 Merkle 树, 主要功能包括:

- 构建 10 万叶子节点的 Merkle 树
- 生成叶子的存在性证明 (Inclusion Proof)
- 生成叶子的不存在性证明(Exclusion Proof)

5.1 Merkle 树构建

5.1.1 数据结构定义

```
1 struct MerkleNode {
2    uint8_t hash[32]; // SM3哈希值
3    std::shared_ptr<MerkleNode> left_child;
4    std::shared_ptr<MerkleNode> right_child;
5    std::weak_ptr<MerkleNode> parent_node;
6 };
```

5.1.2 构建算法

```
算法 4 Merkle 树构建流程
输入: 数据项集合 D = \{d_1, d_2, ..., d_n\}
输出: Merkle 树根哈希 Hroot
1: 计算所有叶子节点: L_i = SM3(d_i)
2: 对叶子节点按哈希值排序
3: while 当前层节点数 > 1 do
     if 节点数为奇数 then
4:
        复制最后一个节点
5:
     end if
6:
     for 每两个节点 (L_i, L_{i+1}) do
7:
        计算父节点: P_i = SM3(L_i \parallel L_{i+1})
8:
     end for
9:
     将父节点作为新当前层
10:
11: end while
12: 返回唯一剩余节点作为根节点
```

代码实现关键部分:

```
void MerkleTree::build_tree(const vector<vector<uint8_t>>>& data
) {
create_leaves(data); // 创建并排序叶子节点
vector<shared_ptr<MerkleNode>>> current_level = leaf_nodes;
tree_root = build_level_up(current_level); // 递归构建上层
节点

5 }
```

5.2 存在性证明

5.2.1 证明生成

对于叶子节点 L,其存在性证明路径为从 L 到根节点的路径上所有兄弟节点的哈希值。

数学表达:

$$Proof_{inclusion} = \{(sib_1, pos_1), ..., (sib_k, pos_k)\}$$

$$(16)$$

```
1 vector<MerkleProofEntry> generate_inclusion_proof(const uint8_t
      * leaf_hash) {
       auto leaf = find_leaf_node(leaf_hash); // 二分查找叶子
2
       vector<MerkleProofEntry> proof;
3
       while (leaf && !leaf->parent.expired()) {
4
           auto parent = leaf->parent.lock();
5
           // 记录兄弟节点哈希和位置关系
6
           proof.push_back({parent->left->hash, parent->right->
7
              hash,
                           parent \rightarrow left = leaf \});
8
           leaf = parent;
9
10
       }
       return proof;
11
12
```

5.2.2 验证算法

```
算法 5 存在性验证流程
```

```
输入: 叶子哈希 L, 根哈希 R, 证明路径 P
```

输出:验证结果

1: $current \leftarrow L$

2: for $(sibHash, isLeft) \in P$ do

3: if isLeft then

4: $current \leftarrow SM3(current \parallel sibHash)$

5: **else**

6: $current \leftarrow SM3(sibHash \parallel current)$

7: end if

8: end for

9: 返回 current == R

5.3 不存在性证明

5.3.1 证明生成

对于不存在元素 X,找到其前驱 L 和后继 R,分别生成存在性证明。

数学表达:

```
Proof_{exclusion} = (Proof_{inclusion}(L), Proof_{inclusion}(R)) 
(17)
```

```
pair < vector < MerkleProofEntry >>, vector < MerkleProofEntry >>>
generate_exclusion_proof(const uint8_t* non_leaf_hash) {
    auto [pred, succ] = find_adjacent_leaves(non_leaf_hash);
    return {create_proof_path(pred), create_proof_path(succ)};
}
```

5.3.2 验证算法

- 1. 验证前驱 L 和后继 R 的存在性
- 2. 确认 L < X < R (按哈希字典序)
- 3. 确认 L 和 R 在树中相邻

5.4 性能优化

5.4.1 并行计算

- 叶子节点哈希可并行计算
- 每层的兄弟节点对可并行处理

5.4.2 内存管理

```
1 // 使用智能指针自动管理节点内存
```

2 shared_ptr<MerkleNode> node = make_shared<MerkleNode>();

5.5 测试结果

测试环境: Intel Core i7-10750H @ 2.60GHz

5.6 安全性分析

- 抗碰撞性: 依赖 SM3 的抗碰撞特性
- 完整性: 任何叶子节点的修改都会改变根哈希
- 证明大小: 证明路径长度为 $O(\log n)$

$$Pr[VerifyProof(H_{root}, X, Proof) = true \mid X \notin T] \le negl(\lambda)$$
 (18)

表 4: 10 万节点 Merkle 树性能

测试项	结果
树构建时间	6.50s
存在性证明生成时间	$0.03 \mathrm{ms}$
存在性验证时间	$0.49 \mathrm{ms}$
不存在性证明生成时间	$0.03 \mathrm{ms}$
不存在性验证时间	$0.97 \mathrm{ms}$

5.7 应用场景

- 区块链交易验证
- 证书透明度日志
- 分布式系统数据一致性验证

5.8 运行结果