

基于 GPU 的 Word 文档暴力破解

李 超

(南京邮电大学 计算机学院, 江苏 南京 230046)

摘 要: 本文首先分析了微软办公软件 word 文档安全加密机制, 对 RC4 算法及其在 word 文档加密中的应用进行研究, 提出了基于 GPU 的口令暴力破解算法. 该算法利用 GPU 的高速并行特性, 多线程并行执行 word 文档口令暴力破解任务, 实验结果表明现有 word 文档安全加密方法的脆弱性.

关键词: 暴力破解; 图形处理器; word 文档; RC4

中图分类号: TP309

文献标识码: A

文章编号: 1000-7180(2012)11-0130-04

Study of Brute Force for Word Document Based on GPU

LI Chao

(Computer College, Nanjing University of Posts and Telecommunications, Nanjing 210046, China)

Abstract: This paper firstly analyzes the Microsoft word document security encryption mechanism, the RC4 algorithm and the word document encryption application, and then a new brute force algorithm for password cracking based on GPU was proposed based on the GPU. The proposed algorithm utilized the high-speed parallel characteristics of GPU and executed the brute force task for word document with multi threads. The experimental results show that the existing security encryption method of word document is not safe.

Key words: brute force cracking; GPU; word document; RC4

1 引言

Microsoft word 是计算机用户常用的办公编辑软件, 具有强大的文字编辑功能, 是用户文档编辑和数据保存最通用的格式文件, 其安全性一直得到用户的高度重视. 通常情况下, word 软件可以为文档提供多种方式的密码保护, 但是无论哪种加密方式, 其核心算法均为 RC4 流加密算法, 使用的密钥可以为 40 位、64 位或者 128 位长度, 在加密时可以选择指定密钥的位数.

通常情况下破解 RC4 加密可以通过直接法^[1]和间接法^[2-4]两种. 直接法主要针对 RC4 算法的异或运算较为简单这个缺点展开攻击, 如果能够已知一对明密文对, 攻击者则可以轻松破解同一伪随机密钥加密过的等长密文. 间接法则采用暴力破解的方法去搜索其整个密钥空间. 间接法中使用较频繁

且有效的攻击方法为暴力破解攻击方法, 其实现平台有 CPU、FPGA、云平台、分布式平台等多种. 文献[2-6]采用 FPGA 实现 RC4 的暴力破解, 由于单引擎 RC4 暴力破解需要的 FPGA 资源仅需要 9000 个左右的 CLB slice 资源, 在较大型的 FPGA 设备中, 可以同时构建成百上千个破解引擎并行运行, 因此可以高速完成针对 RC4 算法的暴力破解. 但是文献[2-5]使用的 FPGA 逻辑均较大, 资源利用率不高, 文献[6]充分利用 FPGA 的特性, 提出的方案可以在 772 时钟周期内完成一次密钥匹配过程, 如果破解 40 比特的密钥, 采用该方案设计, 则可以在 28.5 小时内完成所有密钥的匹配搜索过程, 达到非常好的性能. 但是 FPGA 由于其功耗较高, 而且应用场合受限等缺点, 一般用于系统验证. 文献[7]利用分布式环境对 RC4 算法暴力破解, 结果表明如果要破解 40 比特密钥的 RC4 算法, 需要 100 台高性能计

收稿日期: 2012-05-20; 修回日期: 2012-06-15

基金项目: 江苏高校优势学科建设工程项目(yx002001); 嵌入式系统与开发-教育部英特尔精品课程(4700410G01)

算机 1 天的时间. 文献[8]利用云计算和预测判断算法对 WEP 进行攻击解密, 其性能相较于传统的方案, 可以提高百万倍, 能够在数分钟内完成密钥破解, 但是其仅应用于 WEP 加密协议, 不能推广到任何 RC4 算法使用场合.

本文利用 GPU 的高速并行性能, 根据 word 文档特点, 提出基于 GPU 的 word 文档暴力破解算法, 并在 NVIDIA Tesla™ C1060 GPU 平台上进行测试, 实验结果表明该方法破解率高, 破解时间短, 因此针对 word 文档的安全技术可靠性不高, 应该综合运用其它手段来提高 word 文档的安全.

2 相关工作

2.1 Word 文档加密机制及攻击

根据微软 word 文献规定, word 文档采用 RC4 算法进行加密, 密钥长度可以为 40bit、64bit 或者 128bit, 默认情况下为 40bit, 所以 word 文档的破解可以转化为对 RC4 算法的破解. 默认情况下, word 文档是按照 512 字节被划分成块, 所以应该产生 512 字节的密钥流来进行加密, 其加密流程如图 1 所示.

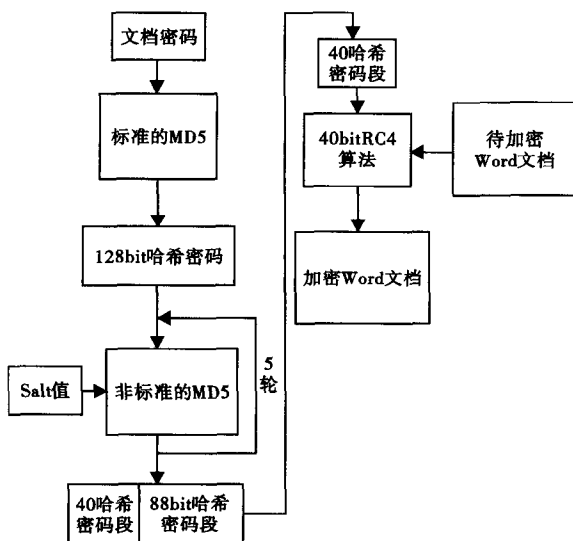


图1 word 文档加密机制

从图 1 可以看出, 对 word 文档有两种暴力攻击手段. 方案 1 就是先猜测一个用户口令字, 加上 salt 值计算出决定 RC4 算法的那个 40bit 数据, 然后验证是否正确, 如果正确则得到用户口令字, 停止攻击, 否则继续猜测下一个口令字; 方案 2 就是直接猜测决定 RC4 初始化向量的 40bit 数据值, 这样就绕过了前面多轮 MD5 运算, 仅相当于对 40 比特的 RC4 算法进行攻击, 从而大大加快破解速度.

2.2 RC4 算法

从图 1 可以看出 RC4 是 word 加密的关键算法. 下面就简单介绍 RC4 算法的原理及实现. RC4 算法包括密钥调度法 (Key Scheduling Algorithm, KSA) 和伪随机密钥流产生算法 (Pseudo-random Generation Algorithm, PRGA) 两个步骤, 其伪代码如下所示.

(1) 密钥调度法 (KSA). KSA 部分主要完成 8×8 的 S 盒初始化工作, 首先对 S 盒初始化, 然后利用初始密钥对 S 盒打散, 使其具有足够的随机性, 伪代码如下:

```

for i from 0 to 255 S[i]=i;
for i from 0 to 255
[j=(j+S[i]+key[i mod keylength]) mod 256
swap(S[i], S[j])]

```

(2) 伪随机密钥流产生算法 (PRGA)

利用 KSA 阶段所产生的 S 盒, 产生伪随机密钥流, 该伪随机密钥流与明文位异或得到密文, 伪代码如下:

```

i=j=0
while bitofPlaintext
[i=(i+1) mod 256; j=(j+S[i]) mod 256;
swap(S[i], S[j]); t=S[i]+s[j];
x=output S[t mod 256];
bitofCipherText=x⊕bitofPlaintext]

```

在密钥已知情况下, RC4 解密算法同加密算法非常类似, 用明文异或伪随机密钥流就可以得到原始明文.

2.3 基于 GPU 平台的暴力破解

暴力破解也称为穷举攻击, 其基本思想是穷举所有可能的密钥进行破解. 假设抽取穷举密钥空间的某个密钥 K_i , 对已知明文 P 进行加密得到密文 C' , 如果 C' 和 C 相同, 则密钥 K_i 即为所求密钥, 否则 K_i 不是所求密钥, 继续尝试下一个密钥 K_{i+1} . 暴力破解算法虽然简单, 但是却非常有效, 可以取得 100% 成功率, 其最大缺点就是耗时非常大, 尤其对于计算能力较弱的 CPU 平台, 很多时候仅有理论上的可行性, 没有实际应用价值. 但是由于硬件技术的高速发展, 越来越多的高速并行运算器件诞生, 其高速的并行计算能力使得暴力破解成为现实. RC4 暴力破解算法伪代码如下:

```

∀ Ki ∈ Ω
loop: C' = RC4Ki(P);
if (C' == C)
break;

```

```

else
goto loop

```

相对于其它具有并行运算能力的器件来说,图形处理器(GPU)内部具有成百上千个基本处理器,因此可以用于大规模并行处理任务。

3 基于 GPU 的 word 文档暴力破解算法原理与实现

3.1 方案设计

在上节论述了 word 文档暴力破解有两个方案,由于采用方案 1 还需设计实现 MD5 算法,故此处选择较节省资源和时间的方案 2. 图 2 显示了基于 GPU 的 word 文档暴力破解原理。

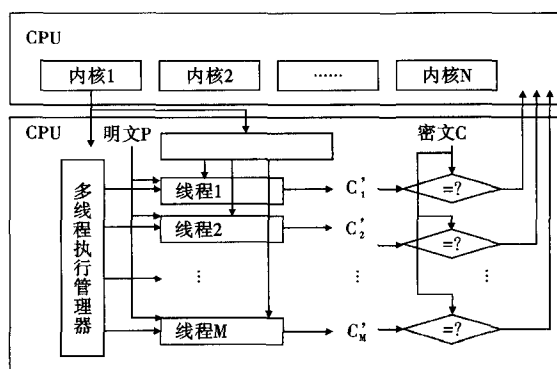


图 2 基于 GPU 的 word 文档密码并行攻击算法原理

从图 2 可以看出,可以在 GPU 平台上设计成千上万个独立破解线程,这些线程高速独立运行,整个平台可以达到很好的破解效果。

3.2 编程实现

CPU 平台是串行执行模式,而 GPU 平台是并行执行模式,因此两种平台下编程既有相同点也有不同点。由于现有的 GPU 处理器不能独立构成系统,需要 CPU 的支撑和管理,通常情况下 CPU 创建内核函数,管理 GPU 等任务,而 GPU 则完成并行任务,因此两者要相互配合,才能得到较好性能。基于 CPU+GPU 的异构计算编程通常有多种方案,本文选择基于 CUDA 架构来完成图 1 编程实现。具体分为如下几个步骤:

- (1) 设置 GPU 内核环境变量,初始化设备
- (2) 从主机平台搬运数据到 GPU 设备上
- (3) 创建所需存储空间并初始化
- (4) 创建内核函数并执行
- (5) 传递中间数据或结果到主机

下面详细描述上述过程。

(1) GPU 初始化和数据搬运

不同 GPU 平台其流处理器数量,内存大小等

内核属性都不同,因此在使用 GPU 平台之前,必须首先对 GPU 硬件资源进行初始化,可以调用 CUDA 的 API 函数来得到,其中 `cudaGetDeviceCount()` 和 `cudaGetDeviceProperties()` 函数用于枚举这些设备并检索其属性。得到设备属性后,可以调用 `cudaMalloc` 接口函数为 GPU 内核函数中的变量分配内存空间,最后调用 `cudaMemcpyToSymbol` 完成将数据从主机搬运到设备上去。伪代码如下:

```

__constant__ unsigned char g_CharSet[];
__constant__ unsigned long g_CharSetLength;
__constant__ unsigned long g_CipherKey;
unsigned int GpuNum;
cudaDeviceProp GpuDeviceProp;
cudaGetDeviceCount(&GpuNum);
cudaGetDeviceProperties(&GpuDeviceProp,0);
cudaMalloc((void *) &d_CompRC4Plaintext,4);
cudaMalloc((void *) &d_CompRC4Ciphertext,4);
cudaMalloc((void *) &d_Result,4);
cudaMemcpyToSymbol(g_CharSet, &CharSet, CharSetLength);
cudaMemcpyToSymbol(g_CharSetLen, &CharSetLength,4)

```

(2) 产生测试密钥

如图 2 所示,每个 GPU 线程均需要一个测试密钥,因此可以通过密钥产生器产生密钥并分配到不同的线程中。密钥产生器所产生的 RC4 密钥为 40bit,其基本要求就是能够均匀取自于密钥空间,每个密钥不重复,因此可以利用 GPU 线程号来进行分配。由密钥产生器来分配不同的密钥到 GPU 线程。

(3) 创建 GPU 内核函数

在 CPU 编程模型中,函数被调用时仅执行一次,但是 GPU 编程模型则不同,它需要编写 GPU 内核函数,并在 CPU 中调用该内核函数,该函数将会被 GPU 中 N 个不同的线程并行执行 N 次,且每个线程都会有唯一的线程 ID。GPU 内核函数调用方式为:

```

WordBFKerF<<blockNpG, threadNpB>>>
>(P1,C1);

```

其中 `WordBFKerF` 为 GPU 内核函数名,参数 `blockNpG` 表示内核的栅格数和大小,及线程块数量,参数 `threadNpB` 则表示每个线程块中的线程数量,所以每个 GPU 中内核线程的总数等于 `blockNpG * threadNpB`,参数 `P1` 和 `C1` 表示已知的明文对。

(4) word 文档密码暴力破解线程

按照图 2 所示编写 word 文档暴力破解线程,

其伪代码如下:

```
C=WordBruteForcesFun(P1, CipherKey);
```

其中 WordBruteForcesFun() 函数则为 RC4 算法加密函数,其具体实现见 2.2 节,C 为在密钥为 CipherKey 时经过 RC4 算法运算得到的明文 P1 所对应的密码。

4 测试结果及分析

为了测试 GPU 平台上暴力破解加密 word 文档的性能,搭建了 NVIDIA Tesla™ C1060 平台,该平台可将工作站转变成一个性能远超小型集群的高性能计算机,它的计算能力甚至超过了一个小的计算机集群。测试的密钥字符集为大小写字母和数字组合,随机产生多个加密 word 文档,加密算法选择 40bit 的 RC4 算法。在 C1060 平台上进行暴力破解实验,得到结果如图 3 所示。

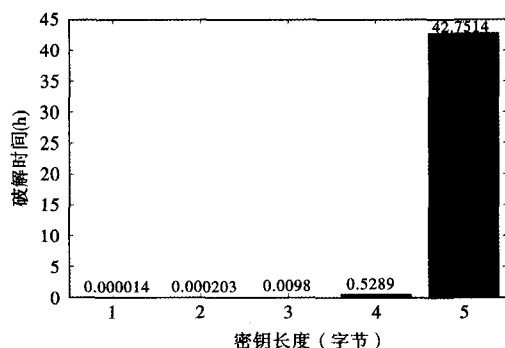


图3 基于 GPU 平台的 word 文档暴力破解

从图 3 可以看出,密钥空间的大小对实验结果影响很大,密钥长度为 3 字节时,密钥空间为 $61^3 = 226981$ 个密钥,在 C1060 平台上每秒大概能够测试 6458 个密钥,所以进行 35 秒左右就可以在 GPU 平台上测试完所有密钥。当密钥长度变为 5 字节时,密钥空间则为 $61^5 \approx 2^{30}$,仍然使用 C1060 平台来进行测试,则需要 42.7514 小时才能完成破解任务。因此加大密钥长度可以有效防范针对 word 文档的暴力破解,从而能够有效保护用户数据和应用的安全。

5 结束语

Word 文档被广泛应用在各种办公场合,其数据安全性非常重要。本文根据 word 文档的加密原理,结合 RC4 算法的实现特点,提出基于 GPU 平台的 word 文档暴力破解算法。该算法在 C1060 平台上进行了实验测试,测试结果表明 word 文档加密算法为 40bit 的 RC4 时,其文档安全性不高。为了提高文档安全性,应该使用 128bit 的 RC4 加密算法或者使用 winrar 软件进行二次加密等多种手段,当然

也可以扩大密钥的字符集,尤其在密钥中加入非大小写字母,非数字的 ASCII 码时,会使暴力破解难度成倍增大,有利于文档的安全性保护。

参考文献:

- [1] 孙宏,杨义先. 无线局域网协议 802. 11 安全性分析[J]. 电子学报, 2003 (7): 109821100.
- [2] Hamalainen P, Hannikainen M, Hamalainen T, et al. Hardware implementation of the improved WEP and RC4 encryption algorithms for wireless terminals[C]// proceedings of the Ecopcar Signal Processing Conference. 2000:2289-2292.
- [3] Kitsos P, Kostopoulos G, Sklavos N, et al. Hardware implementation of the RC4 stream cipher[C]// Proc IEEE Midwest Symp. Circuits and Syst[S. L.]; IEEE, 2003:1363-1366.
- [4] Couture N, Kent K. The effectiveness of brute force attacks on RC4[C]// Proc 2nd Annu Conf Commun Networks and Services Res (CNSR' 04) Couture; IEEE, 2004:333-336.
- [5] Tsoi K, Lee K, Leong P. A massively parallel RC4 key search engine[C]// Proc IEEE Symp Field-Programmable Custom Comput. Napa, Calif: IEEE, 2002: 13-21.
- [6] Kwok S. Effective Uses of FPGAs for Brute-force Attack on RC4 Ciphers[J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2008, 16 (8): 1096-1100.
- [7] 张丽丽. 基于分布式计算的 RC4 加密算法的暴力破解[J]. 计算机工程与科学, 2008, 30(7): 15-18.
- [8] 汪建. 云计算与无线局域网安全研究[J]. 重庆师范大学学报, 2010, 27(3): 64-68.
- [9] Anh-Duy Vu, Jea-Il Han, Hong-An Nguyen, et al. A homogeneous parallel brute force cracking algorithm on the GPU[C]// ICT Convergence (ICTC), 2011 International Conference on Digital Object, 2011: 561-564.
- [10] Martijn Sprengers. GPU-based Password Cracking [D]. Master's thesis, Radboud University Nijmegen, 2011.
- [11] Fabrizio Milo, Massimo Bernaschi, Mauro Bisson. A fast, GPU based, dictionary attack to OpenPGP secret keyrings[J]. The Journal of Systems and Software, 2011(84): 2088-2096.

作者简介:

李超 男, (1975—), 博士, 讲师。研究方向为嵌入式系统、信息安全。

