암호화된 파일의 비밀번호 복구 연구 동향

한성대학교 융합보안학과 (석사과정) 윤세영
유투브 주소: https://youtu.be/_g31cM_Qn4w

목차

1. 서론

1. 서론

- 복잡하게 설정된 비밀번호는 복구 과정에서 상당한 시간이 소요되며, 경우에 따라 비밀번호를 복구하는 것 자체가 불가능해 정보에 접근하지 못할 수 있다.
- 본 논문은 현재에도 사용되고 있는 주요 비밀번호 복구 도구들과 함께 높은 사용률을 보이는 압축 파일, PDF 및 Excel 문서를 대상으로 비밀번호를 복구하는 기존의 연구에 대해서 살펴본다.

2. 관련 연구

2.1 비밀번호 복구 기술

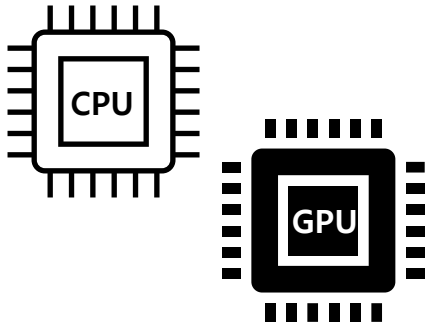
2.2 압축 파일

2.3 비밀번호 복구 관련 도구

2.1 비밀번호 복구 기술

2.1 비밀번호 복구 기술

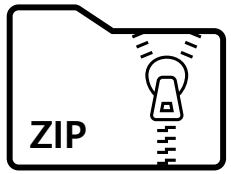
- 무차별 대입 공격(Brute-force attack) : 사용할 수 있는 모든 문자와 숫자를 조합하여 대입하는 방식이므로 많은 시간과 자원을 필요로 한다.
- 사전 공격(Dictionary attack) : 비밀번호로 자주 쓰이는 단어를 사전 파일로 만들어 두는 방식이다.
- 레인보우 테이블 공격(Rainbow Table attack) : 특정 암호 알고리즘으로 미리 해시된 해시값을 저장해 두는 방식이다.
- 따라서 길이가 길고 복잡한 비밀번호도 빠르게 복구하기 위해 CPU 대신 GPU만 사용하거나, CPU와 GPU를 함께 사용하여 연산 과정의 처리 속도를 높이는 연구가 진행되고 있다.



2.2 압축 파일

2.2 압축 파일

- 압축 파일은 압축 알고리즘을 사용하여 아카이브 파일의 크기를 줄인 것이다.
- 데이터 압축 시 무손실 압축 (Lossless compression)을 수행한다.
- RAR, ZIP, 7z 등의 파일 포맷을 가지고 있다.



- 주어진 비밀번호로부터 해시 함수 PBKDF2 (Password-Based Key Derivation Function 2)를 사용하여 AES 키를 생성
- PBKDF2(pw, salt, dkLen), HMAC-SHA1 알고리즘 1000번 수행
- 생성된 AES 키를 사용하여 암호화된 ZIP 파일을 복호화
- 압축 파일에 저장된 MAC 값과 비교되어 비밀번호가 올바른지 여부를 결정하는 확인 값을 생성

2.3 비밀번호 복구 관련 도구

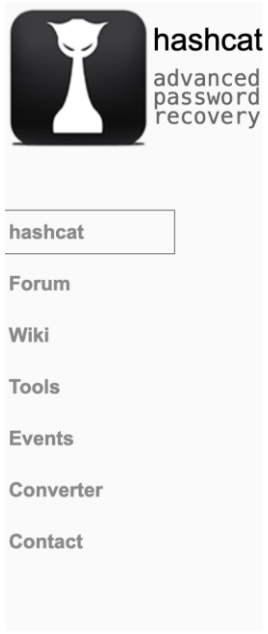
2.3.1 hashcat

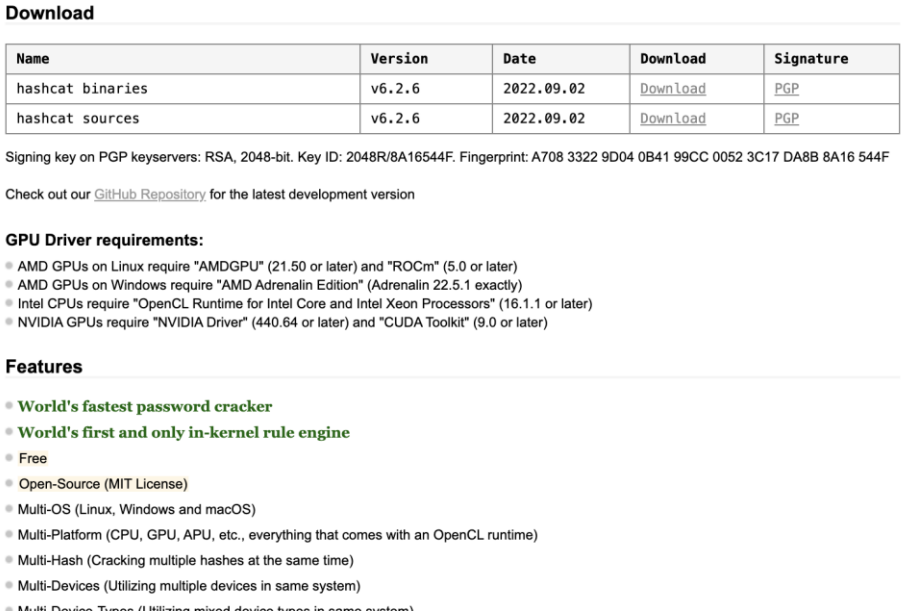
2.3.2 John The Ripper

https://hashcat.net/hashcat/

2.3.1 비밀번호 복구 관련 도구 - hashcat

- MIT 라이센스에 따라 오픈소스로 공개되어 있는 비밀번호 복구 도구이다.
- MD5, SHA512 등 350개 이상의 다양한 해시 알고리즘을 이용하여 암호화된 파일 및 시스템의 비밀번호를 복구하는 데 사용된다.
- 공식 웹에서 다운로드할 수 있는 최신 버전은 v6.2.6이며, 2022년 9월에 업데이트되었다.





Download

Name	Version	Date	Download	Signature
hashcat binaries	v6.2.6	2022.09.02	Download	PGP
hashcat sources	v6.2.6	2022.09.02	Download	PGP

Signing key on PGP keyservers: RSA, 2048-bit. Key ID: 2048R/8A16544F. Fingerprint: A708 3322 9D04 0B41 99CC 0052 3C17 DA8B 8A16 544F

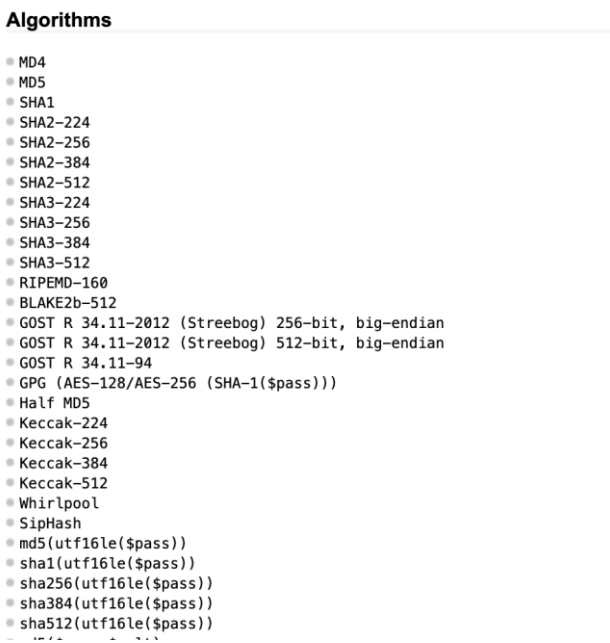
Check out our GitHub Repository for the latest development version

GPU Driver requirements:

- AMD GPUs on Linux require "AMDGPU" (21.50 or later) and "ROCm" (5.0 or later)
- AMD GPUs on Windows require "AMD Adrenalin Edition" (Adrenalin 22.5.1 exactly)
- Intel CPUs require "OpenCL Runtime for Intel Core and Intel Xeon Processors" (16.1.1 or later)
- NVIDIA GPUs require "NVIDIA Driver" (440.64 or later) and "CUDA Toolkit" (9.0 or later)

Features

- **World's fastest password cracker**
- **World's first and only in-kernel rule engine**
- Free
- Open-Source (MIT License)
- Multi-OS (Linux, Windows and macOS)
- Multi-Platform (CPU, GPU, APU, etc., everything that comes with an OpenCL runtime)
- Multi-Hash (Cracking multiple hashes at the same time)
- Multi-Devices (Utilizing multiple devices in same system)



Algorithms

- MD4
- MD5
- SHA1
- SHA2-224
- SHA2-256
- SHA2-384
- SHA2-512
- SHA3-224
- SHA3-256
- SHA3-384
- SHA3-512
- RIPEMD-160
- BLAKE2b-512
- GOST R 34.11-2012 (Streebog) 256-bit, big-endian
- GOST R 34.11-2012 (Streebog) 512-bit, big-endian
- GOST R 34.11-94
- GPG (AES-128/AES-256 (SHA-1($pass)))
- Half MD5
- Keccak-224
- Keccak-256
- Keccak-384
- Keccak-512
- Whirlpool
- SipHash
- md5(utf16le($pass))
- sha1(utf16le($pass))
- sha256(utf16le($pass))
- sha384(utf16le($pass))
- sha512(utf16le($pass))

https://www.openwall.com/john/

2.3.2 비밀번호 복구 관련 도구 – John The Ripper

- Unix 버전(Linux, AIX, QNX, Solaris, BSD)과 MacOS, Windows의 운영 체제에서 사용할 수 있는 오픈 소스 비밀번호 복구 도구이다.
- 기존의 John the Ripper는 CPU 플랫폼에서만 이용할 수 있었지만, 2023년 3월 NVIDIA GPU 드라이버와 함께 Amazon Linux 2에서도 사용할 수 있도록 업데이트되었다.
- 현재 John the Ripper에서는 RAR, ZIP, 7-Zip에 대한 포맷을 지원하고 있으며, AES로 암호화된 WinZip은 JtR 1.7.8-jumbo-2 이상에서 지원하고 있다. 문서 파일로는 PDF, docx, iWork 포맷을 사용할 수 있다.

3. 암호화된 파일에 대한 비밀번호 복구 연구 동향

3.1 WinRAR3(RAR)

3.2 PDF(version 1.4-1.6)

3.3 Microsoft Excel(version 2003)

3.1 WinRAR3(RAR)

3.1 WinRAR3(RAR)

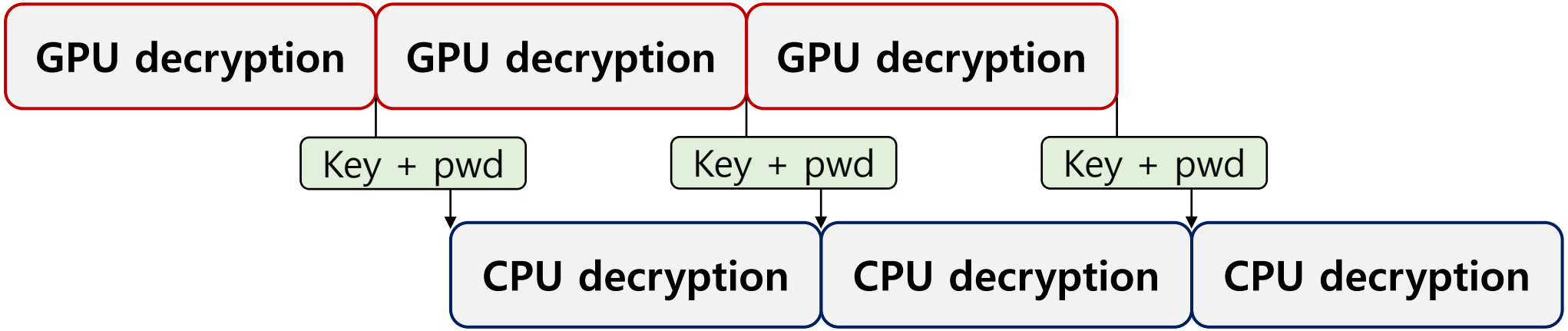
Qingbing Ji, Hao Yin, "Speedup and Password Recovery for Encrypted WinRAR3 without Encrypting Filename on GPUs.", Journal of Physics: Conference Series, Vol.1673, 012047, 2020.

- 해당 연구는 WinRAR 버전 3에서 파일 이름이 암호화되지 않은, 기본 암호화 모드를 대상으로 연구를 진행했다.
- WinRAR3의 암호화 방법으로는 AES 암호화 알고리즘 및 SHA-1 해시 함수가 사용되었다.

CPU	Xeon(R)E5-2620
GPU	NVIDIA 1080Ti
CUDA	10.2
operating system	Linux CentOS7

<표 1> 실험 환경

3.1 WinRAR3(RAR)



<그림 1> GPU와 CPU를 이용한 파이프라인 모드

Size of compressed file	Speed before optimization	Speed after optimization
1K	10981/s	24423/s
10M	9738/s	22423/s
100M	6235/s	16423/s

<표 2> 최적화 전 후 속도 비교

3.2 PDF(version 1.4-1.6)

3.2 PDF(version 1.4-1.6)

Hyun Jun Kim, Si Woo Eum, Hwa Jeong Seo, "PDF Version 1.4-1.6 Password Cracking in CUDA GPU Environment.", KIPS Trans. Comp. and Comm. Sys, Vol.12, No.2, pp.69-76, 2023.

- 해당 연구는 PDF 문서 1.4-1.6 버전의 암호 해독 알고리즘을 CUDA GPU 상에서 최적화 구현하였다.
- 해당 버전의 PDF 암호화 알고리즘에서 반복적으로 사용되는 MD5와 RC4의 최적화를 중심으로 연구를 진행했다.
- MD5 알고리즘에서는 변경되지 않는 일부 메시지 워드의 덧셈 연산을 제거했으며, RC4는 32비트 워드를 통합하여 8비트 워드로의 변환 과정을 없애, 덧셈 및 XOR 연산 횟수가 줄어듦으로써 알고리즘의 연산 속도가 높아졌다.

Reference	Environment	Speed
hashcat 6.2.5	RTX 3060	25,693 kp/s
H Kim	RTX 3060	31,460 kp/s
hashcat 6.2.5	RTX 3090	57,601 kp/s
H Kim	RTX 3090	66,351 kp/s

<표 3> 초당 계산 횟수 비교

3.3 Microsoft Excel(version 2003)

3.3 Microsoft Excel(version 2003)

Zhang, Lijun, Cheng Tan, and Fei Yu, "Fast Decryption of Excel Document Encrypted by RC4 Algorithm.", 2020 IEEE 20th International Conference on Communication Technology (ICCT), IEEE, pp.1572-1576, 2020.

- Microsoft의 Excel 문서에 대해 암호화 중간 키(The intermediate key)를 복구하여 암호화된 문서 자체를 복호화 하는 방법을 제시했다.
- 본 연구는 암호화된 파일의 정보에 접근하기 위해 비밀번호를 크래킹(Cracking) 하는 것이 아니라, 레인보우 테이블 공격을 이용하여 암호화 중간 키를 복구하고, 해당 키로 문서를 복호화하여 정보에 접근하는 방식으로 진행되었다.

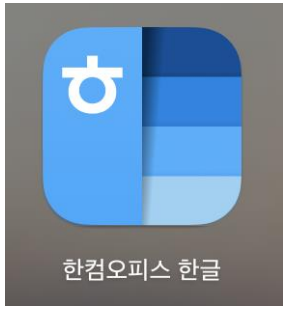
CPU	Intel core i5-8265U @1.6GHz 1.80GHz
RAM	8GB
Operating system	Windows 10, 64 bit.

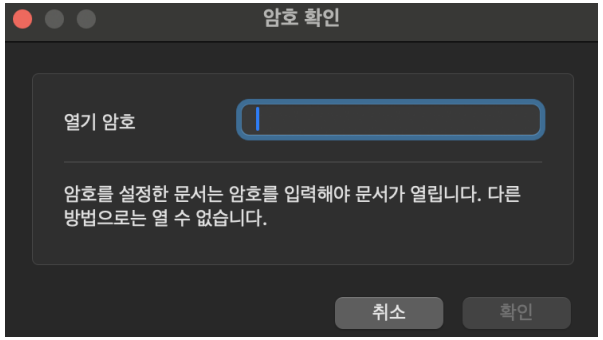
<표 4> 실험 환경

4. 결론

4. 결론

- 암호화된 파일에 접근하기 위해 기존 암호화 알고리즘을 분석하여 최적화 혹은 파일 자체를 복호화 하는 방식을 사용하거나, CPU나 GPU의 아키텍처를 활용하여 비밀번호 복구 성능을 개선하는 연구들이 수행되었다.
- 현대 컴퓨팅 환경에서 무차별 대입 공격을 위한 단순 영문자 및 숫자 조합의 수가 11자리를 넘어가면, 5천억 개 이상의 경우의 수가 필요하다. 여기에 특수문자까지 섞인다면 다항 시간 내에 비밀번호를 찾아낼 수 없을 것이다.
- 이러한 한계점을 극복하기 위해서는 향상된 성능의 프로세서를 사용하거나 파일 암호화에 사용되는 알고리즘에 대한 고속 구현 등의 추후 연구가 더 필요할 것으로 생각된다.
- 한글 파일은 국내의 기업 및 개인 사용자가 많은 만큼 이에 대한 비밀번호 복구 연구 또한 추가적으로 이루어져야 할 것으로 보인다.





감사합니다