2021 IEEE the 6th International Conference on Computer and Communication Systems

Password Cracking of PDF 2.0 documents on GPU

Fei Yu
Science and Technology on Communication Security Laboratory
Chengdu, China
feiyu80@gmail.com

Hao Yin
Science and Technology on Communication Security Laboratory
Chengdu, China
4181739@qq.com

2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS) | 978-1-6654-1256-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCS52626.2021.9449188

***Abstract*—PDF is one of the international documents standards and it is widely used in the world. Password cracking of the PDF documents has attracted the attention of forensic community. Some forensic tools can crack the password of PDF documents with GPU-acceleration only for PDF 1.x documents. This paper studies the password cracking of PDF 2.0 documents, which is the newest PDF standard, on GPU. We analyze the complexity of the password cracking algorithm, and implement it on GPU. We adopt many optimization techniques and our experiments shows that our GPU implementation has about 47 times speedup, compared with the CPU implementation.**

Keywords—PDF 2.0, password cracking, GPU

I. Introduction

PDF (Portable Document Format) is one of the most widely used file format. PDF was firstly proposed in 1993 by Adobe company. PDF has become a de facto global standard for more secure and dependable information exchange since then. In 2017, PDF 2.0 was approved as international standard ISO 32000-2[1].

A PDF document could be encrypted by two kinds of security handler: password-based encryption and public-key encryption. The former is more popular which uses password-based key derived schemes. PDF 2.0 introduced some new security features such as a new password-based key derived scheme which is much more complicated than before.

For the popularity of PDF, the password cracking of PDF documents has aroused people's interest, especially in the field of forensic. People have studied the use of GPU to accelerate the password cracking for PDF 1.x[2-4]. The famous forensic tools such as Elcomsoft Distributed Password Recovery, Passware Kit Forensic and John the Ripper can do so. However, there are no studies or tools for password cracking of PDF 2.0 documents with GPU-acceleration. And the cracking performance on CPU is very slow, which is about 1800 per second and is not practical.

Hence, in this paper we will study the password cracking of PDF 2.0 encryption documents on GPU, which will have a practical performance. First, we will give a summary of Password cracking of PDF 1.x. Then we will introduce the cracking algorithm of PDF 2.0 and give a complexity analysis. Finally, we will study the GPU implementation of the password cracking algorithm and give an optimization. Our experiments show that the password cracking of PDF 2.0 could be implemented on GPU and it is about 47 times faster than CPU.

This work was supported by the Science and Technology on Communication Security Laboratory Funding No.61421030305.

II. Password Cracking of PDF 1.X

Before the analysis of password cracking of PDF 2.0, we will give a summary of the password-based security handler and the password cracking of PDF 1.x.

An encrypted PDF document has two types of password: User's password and Owner's password. The User's password is also known as the open password, which is used to open an encrypted PDF document. The Owner's password is also known as the permissions password, which is used to change the permission settings such as printing, editing, and content copying permission of a PDF document.

For our purpose, we only need to crack the User's password. There are three password-based key derived schemes of PDF 1.x as shown in Table I. In the following we will give a brief description of them.

TABLE I. Cracking Algorithm of PDF 1.X

PDF Version	Acrobat Version	Password Cracking Algorithm
PDF 1.1 - 1.3	Acrobat 2 - 4	MD5:50-RC4-40
PDF 1.4 - 1.7 R4	Acrobat 5 - 8	MD5:50-RC4-128:20
PDF 1.7 R5	Acrobat 9	SHA256

In an encrypted PDF document, all the data should be encrypted except some data including the header which is always "%PDF-" and the encryption information named encryption dictionary. The encryption dictionary's entries indicate the encryption algorithm, key length and other information. The entries that are necessary in password cracking are listed in Table II.

TABLE II. Cracking Algorithm of PDF 1.X

Entry	Description
Filter	The name of the preferred security handler
V	A number specifying the encryption algorithm
Length	The length of the file encryption key (for V=2 or 3)
R	A number specifying the security handler's revision
O	A byte string (32bytes for R<5, 48 bytes for R=5 or 6)
U	A byte string (32bytes for R<5, 48 bytes for R=5 or 6)

978-0-7381-2604-3/21/$31.00 ©2021 IEEE