《Python 程序设计高阶》
(2023-2024 学年第 1 学期)
作业 3

学号:2021329600006　姓名: 陈昊天　班级: 计算机科学与技术 21(4) 班
学号:2021329621213　姓名: 陈佳伟　班级: 计算机科学与技术 21(3) 班
学号:2021329621257　姓名: 冯佳钧　班级: 计算机科学与技术 21(4) 班

§1　题目 1

选择你认为最合适的包，读取 8.Files&Directories 目录下的 course.xls 这个 excel 文件，输出这个文件的每一行的内容。

§1.1　解答

```python
import pandas as pd

# 读取Excel文件
df = pd.read_excel('course.xls')

# 打印每一行的内容
for index, row in df.iterrows():
    print(row)
```

```
Unnamed: 10      推荐教材
Unnamed: 11     主要参考书
Name: 2, dtype: object
附件3                                                    NaN
Unnamed: 1                                             62008
Unnamed: 2                                            操作系统
Unnamed: 3                                    Operating System
Unnamed: 4                                                2.5
Unnamed: 5                                                 48
Unnamed: 6                                       信息科学与技术专业
Unnamed: 7                    数据结构、程序设计基础、计算机组成原理、汇编语言
Unnamed: 8                                       计算机科学与技术系
Unnamed: 9     操作系统是计算机科学与技术专业的专业必修课，是计算机类专业的主干课程。它研究在程序执行中，多...
Unnamed: 10    汤子瀛、哲凤屏、汤小丹编著：《计算机操作系统》（第二版），西安电子科技大学出版社，2001年出版。
Unnamed: 11    William Stallings著，魏迎梅、王涌等译：《操作系统—内核与设计原理》（第三...
Name: 3, dtype: object
附件3                                                    NaN
Unnamed: 1                                             62019
Unnamed: 2                                          计算机安全与保密
Unnamed: 3                                    Computer Security
Unnamed: 4                                                   2
Unnamed: 5                                                  32
Unnamed: 6                                       信息科学与技术专业
Unnamed: 7                                      高等数学、计算机基础
Unnamed: 8                                       计算机与科学技术系
Unnamed: 9     本课程较系统地介绍计算机信息（数据）保密与安全的基本理论和实用技术，既简明扼要地介绍国内外的...
Unnamed: 10    卢开澄编：《计算机密码学—计算机网络中的数据保密与安全》(第三版)，清华大学出版社，200...
Unnamed: 11    [1] 张焕国编：《计算机安全与保密技术》，机械工业出版社，1995年出版；[2] 飞天诚信...
```

§2 题目 2

构造两个不同的文件，但他们有相同的 md5 值，给出你的解决方案，如果你参考了相关的文献，请给出参考文献，如果你使用了其他的工具，请一并说明。

§2.1 解答

思路：使用 cr-marcstevens/hashclash 项目 [1] 构造两个 MD5 值相同的二进制文件，使用 Python 调用此项目并计算 MD5 值

项目环境需求：

g++ make autoconf automake libtool zlib1g-dev libbz2-dev

项目构建：

./build.sh

Python 调用

```python
import subprocess

bash_script = '../scripts/poc_no.sh'
arg1 = 'prefix.txt'

try:
    subprocess.run([bash_script, arg1], check=True)
except subprocess.CalledProcessError as e:
    print(f"Error occurred: {e}")
```

```
e     ************************************************
Found maxcond = 212
t=11: 0%   10   20   30   40   50   60   70   80   90   100%
      |----|----|----|----|----|----|----|----|----|----|
      ************************************************
Q-3:    |01100111 01000101 00100011 00000001|
Q-2:    |00010000 00110010 01010100 01110110|
Q-1:    |10011000 10111010 11011100 11111110|
Q0:     |11101111 11001101 10101011 10001001| ok p=1
Q1:     |11001110 11000010 10010001 00011110| ok p=1
Q2:     |.....11. ........ ........ ........| ok p=1
Q3:     |....V+-. ........ ........ ........| ok p=1
Q4:     |.....+-. ........ ........ .VV.....| ok p=1
Q5:     |....+--. ........ ........ ..0.....| ok p=0.993164
Q6:     |....0+-. ........ ........ .+-.....| ok p=1
Q7:     |....+--. ........ ........ .0+.....| ok p=1
Q8:     |....0+-. .....V.. ........ .1+.....| ok p=1
Q9:     |....1+-. V....1.. ....V... ..+.....| ok p=1
Q10:    |......+. .....+.. ........ ..+.....| ok p=0.979492
Q11:    |.....1+. +....+.. ....+... ..+.....| ok p=0.928711
Q12:    |...+..+. +....+.. +...+... ..+.....|
Saving 160000 paths...done.
Autobalance parameters: maxcond=211
Verified: 0 bad out of 208213
Runtime: 19.3957
Extend MD5 differential paths backward
Copyright (C) 2009 Marc Stevens
http://homepages.cwi.nl/~stevens/

delta_m[2] = [!8!]
Generated 1 new path.
t=32: 0%   10   20   30   40   50   60   70   80   90   100%
      |----|----|----|----|----|----|----|----|----|----|
      ************************************************
```

```
Block 1: ./data/coll1_3357515216
9b 87 04 41 cc a2 65 42 75 ed 3b 1b 89 b5 ca 73
21 70 a3 d4 79 44 e5 91 26 a0 bf da e0 91 c5 3a
87 da 10 61 79 fb 5f 94 ba a1 20 b4 d3 cb b5 a6
bc 71 0c 6e 4d b5 33 65 e1 4e ca 6d 57 24 ad 2c
Block 2: ./data/coll2_3357515216
9b 87 04 41 cc a2 65 42 75 ee 3b 1b 89 b5 ca 73
21 70 a3 d4 79 44 e5 91 26 a0 bf da e0 91 c5 3a
87 da 10 61 79 fb 5f 94 ba a1 20 b4 d3 cb b5 a6
bc 71 0c 6e 4d b5 33 65 e1 4e ca 6d 57 24 ad 2c
Found collision!
bef61d03adaab3bbd7360fb55336df38  collision1.bin
bef61d03adaab3bbd7360fb55336df38  collision2.bin
06a0d15ec8a4353ed02f9101f5eee4aa3c4558e0  collision1.bin
a4d3fdcb74fb6d53982a378aa90bb2b98e6c90e6  collision2.bin
4 -rw-r--r-- 1 root root 128 Nov 13 03:35 collision1.bin
4 -rw-r--r-- 1 root root 128 Nov 13 03:35 collision2.bin
root@STfEQzGo:~/hashclash/cpc_workdir# ls
cal.py  clash.py  collision1.bin  collision2.bin  data  logs  prefix.txt  upper_1_640000
```

Python 计算 MD5

```python
import hashlib

def calculate_file_md5(filename, block_size=4096):
    md5 = hashlib.md5()
    try:
        with open(filename, 'rb') as f:  # 打开文件以进行二进制读取
            for block in iter(lambda: f.read(block_size), b''):  # 分块读取数据
                md5.update(block)
    except IOError as e:
        print(f"无法打开或读取文件: {e}")
        return None
    return md5.hexdigest()

file1_path = 'collision1.bin'
file2_path = 'collision2.bin'
print(f"The MD5 of the file is {calculate_file_md5(file1_path)}")
print(f"The MD5 of the file is {calculate_file_md5(file2_path)}")
```

```
root@STfEQzGo:~/hashclash/cpc_workdir# python3 cal.py
The MD5 of the file is bef61d03adaab3bbd7360fb55336df38
The MD5 of the file is bef61d03adaab3bbd7360fb55336df38
root@STfEQzGo:~/hashclash/cpc_workdir#
```

§3 总结

§3.1 分工情况

以下说明小组的分工情况，每个人完成了什么功能，完成的情况如何

陈昊天

负责报告的整合撰写；

负责题目 2-cr-marcstevens/hashclash 项目环境准备和二进制构建；

陈佳伟

负责题目 1 的研究和解答；

冯佳钧

负责题目 2-Python 调用 Bash 脚本和 MD5 计算；

§3.2 解决方案评价

以下说明小组对解决方案的评价，是否完成了题目的要求，怎么证明完成了题目的要求，好的地方在哪里，哪里有不足。

3.2.1 题目 1

完成了题目要求，打印了 course.xls 文件每一行的内容

优点：完整显示了 excel 文件的内容；代码简单；

缺点：输出排版不美观；

3.2.2 题目 2

完成了题目要求，构造了两个具有相同 MD5 的不同文件；

优点：Marc Stevens 基于王小云的差分攻击方法，进一步改进了攻击技术，开发了 chosen-prefix collisions 方法，这允许攻击者为两个具有不同前缀的消息创建相同的 MD5 哈希值。

缺点：解决方案依赖于 Marc Stevens 的 HashClash 项目，该项目是用 C++ 开发的，Python 仅仅起到调用的作用；

参考文献

[1] Marc Stevens et al. Attacks on hash functions and applications. *Mathematical Institute, Faculty of Science, Leiden University*, 3, 2012.