流量包特征提取软件功能需求

核心功能：

1、从pcap文件中过滤某种长度范围的报文，不超过M个，将传输层以上的业务数据提取N个字节，计算MD5值，并记录该报文在pcap包中的位置，报文长度等基本信息。

2、将MD5值记录在每个流量包对应的MD5文件中

3、交叉比对MD5文件中的MD5值，寻找每个文件中与其它文件不同的MD5值（不用在本文件中比对）——特征值，X个，记录下帧长，位置，提取的数据内容。

4、如果有某个文件中没有特征值，软件报错。

5、如果可以，当软件报错后，自动调整报文长度范围（一般是累增），直到所有的文件中都有一个特征值，记录下长度范围。

辅助功能：

1、能连续自动读取目录下的pcap文件

2、自动生成存放MD5值文件的目录（以长度范围+N命名），存放MD5文件，与流量包文件同名

3、自动生成存放特征值文件的目录（以长度范围+N+“特征值”命名）。

概要方案：

1、采用python代码编写

2、采用bat文件向python代码供给和调整运行参数

3、所有记录的数据均采用HEX字符流形式。程序内部采用什么格式根据程序自身逻辑需要而定

4、入选的包目前限定为TCP，UDP，过滤掉DNS。

可参考的代码包括

1、能分析pcap包的python代码和相关说明文件

2、MD5值计算的代码，可能是C的，可能需要转成python

3、目录和文件处理的python代码。

4、MD5值比对方法，采用穷举法。每一轮找出1个，然后开始下一轮找另外一个。