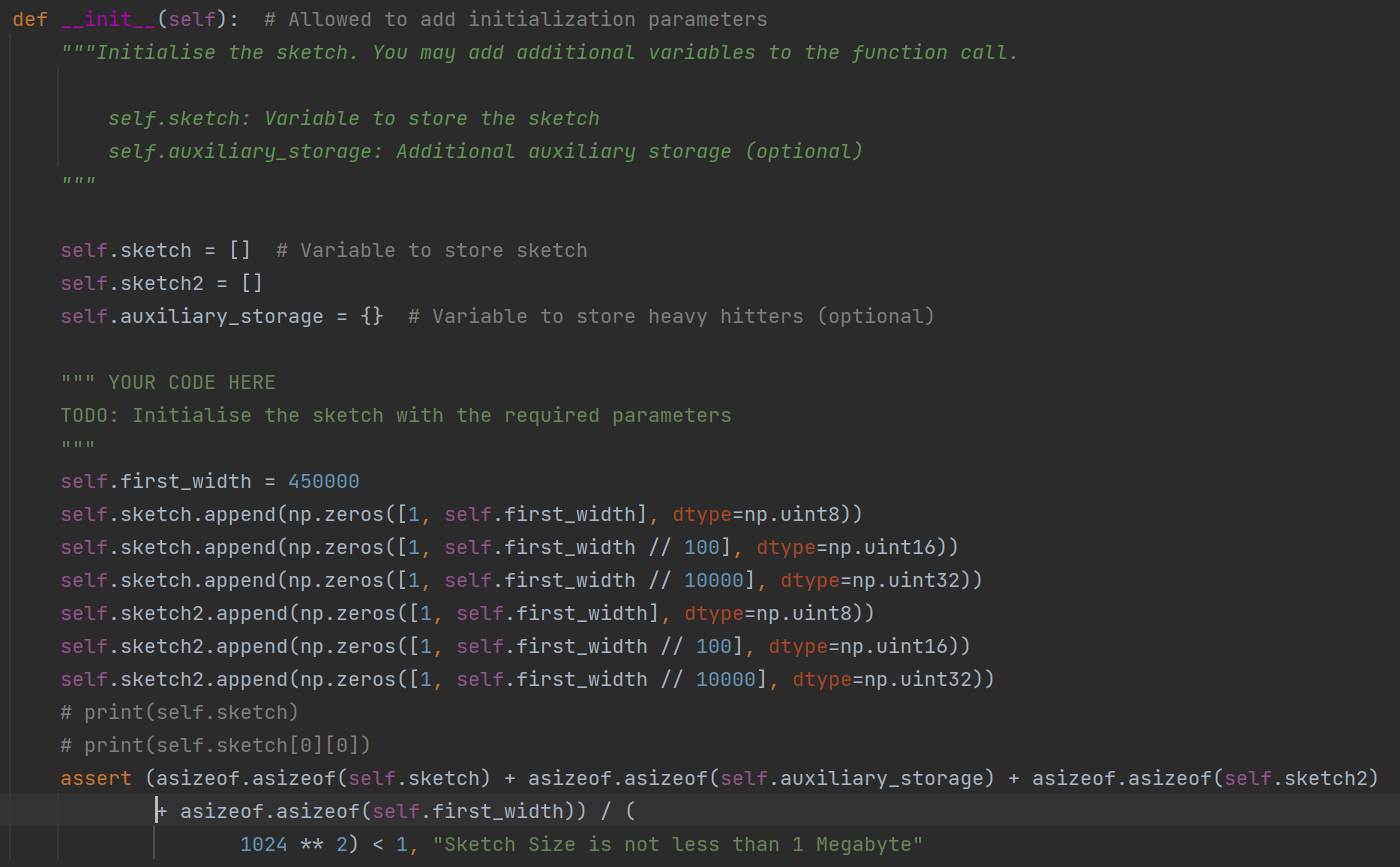
Name: WEI YILEI, Student ID: A0276571W

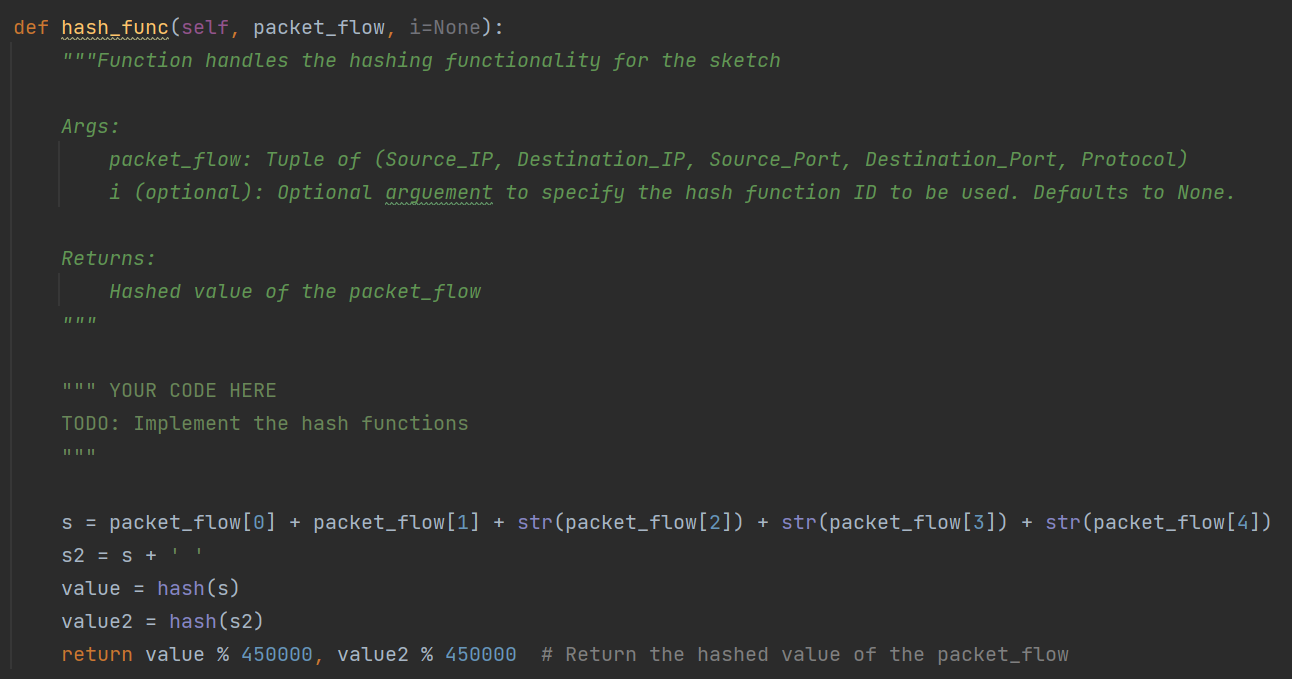
Part1: explanation and implementation of code

在尝试了basic Count Min Sketch、Elastic Sketch、Single-Tree FCM-Sketch后，mean error的结果始终不是很理想，最终选择尝试2-Tree FCM-Sketch。

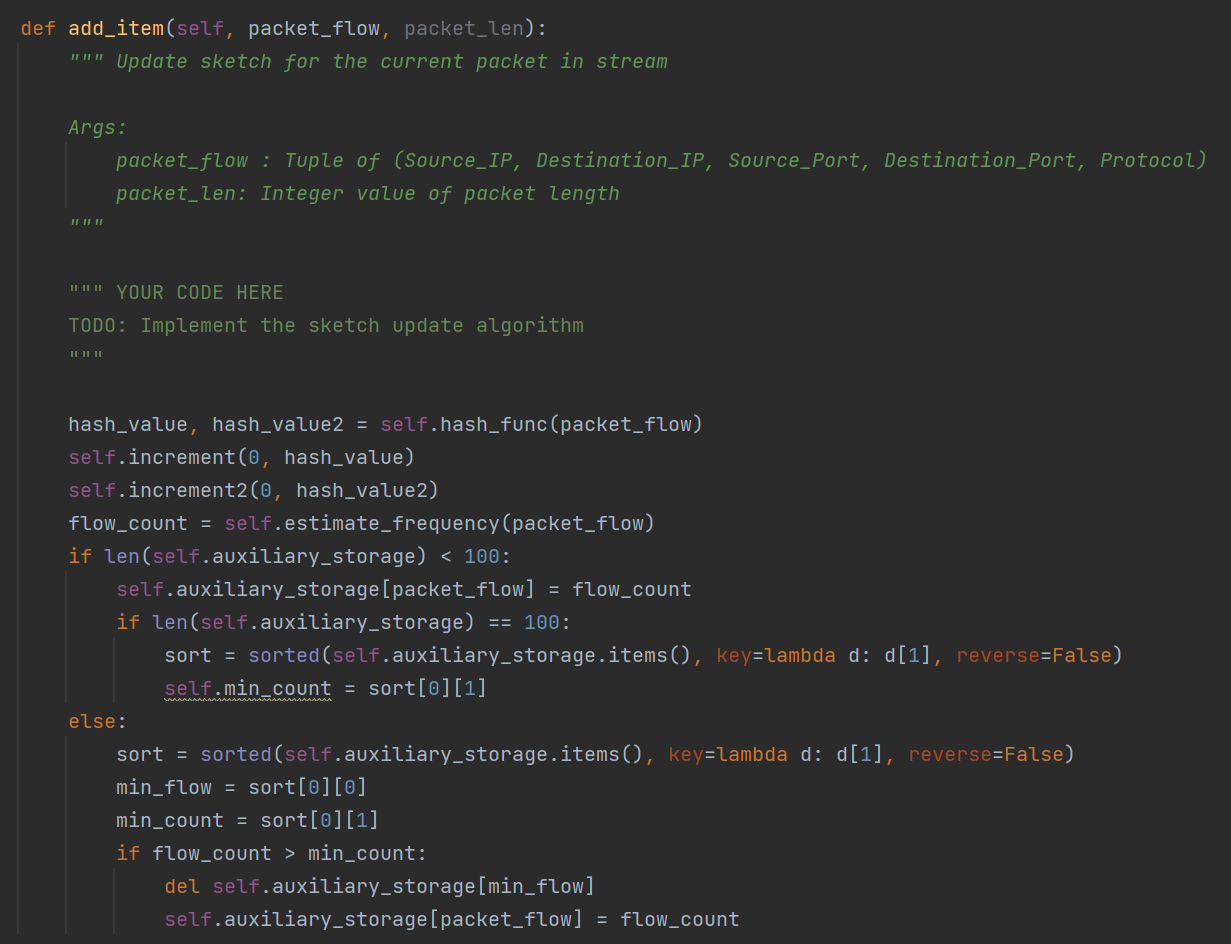
在初始化FCM-Sketch的时候，在sketch内部初始化两个结构相同的k-ary tree，因为之前尝试basic Count Min Sketch的时候设计的大小为（202000\*5）的uint8数组能够几乎用满1MB，且在网络中小包的数量远远多于大包，因此为了节省内存空间的同时又能尽可能地减少哈希碰撞的可能，经过相应计算我使用了第一层450000nodes， 第二层4500nodes， 第三层45nodes的100-ary tree. auxiliary\_storage用于在不断添加packet的过程中维护当前top-100大小的flow数据dict



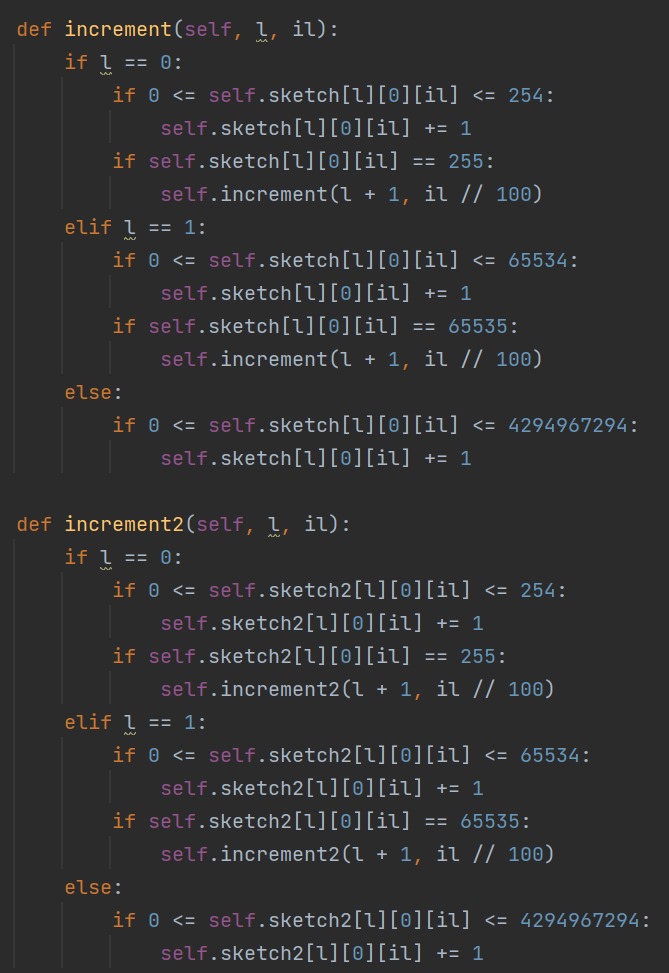
因为我有两个tree，所以相应地我需要两个哈希方程。在处理input tuple的过程中，我尝试过将所有的字符拼串起来，再用int（str， base=16）的方式转换成整型，但是最后的哈希结果并不理想，所以最后还是使用了字符串哈希，对所得的packet\_flow tuple，将其拼凑成一个字符串s，其中一个哈希计算hash（s）的结果value，另一个哈希计算hash（s + ‘ ’）的结果value2，最后输出（value mod 450000， value2 mod 450000）



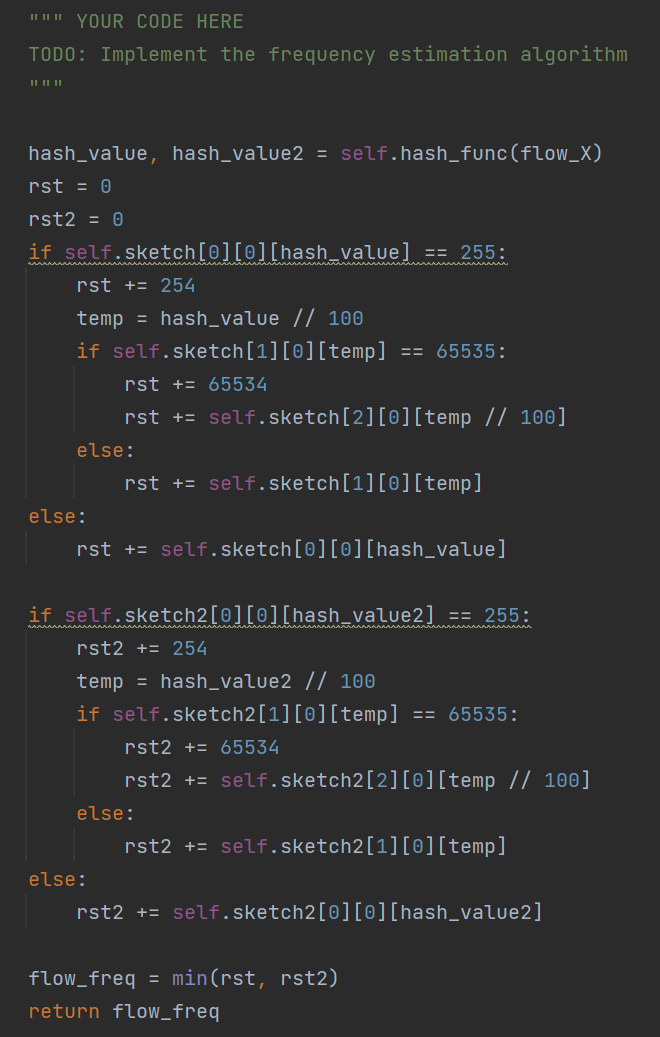
每添加一个packet，先通过hash\_func（）计算出两个哈希值hash\_value和hash\_value2，再分别通过increment(0, hash\_value)和increment2(0, hash\_value2)更新两个tree。在进程初期，sketch的top-100 dict还没有100个元素，此时遇到的packet\_flow可以全部添加进dict。当top-100 dict拥有100个元素后，我们就需要辨别每个packet被处理后，对应的新flow size是不是会成为新的top-100 dict中的一员，如果会的话，添加该flow并删除dict中最小的flow。



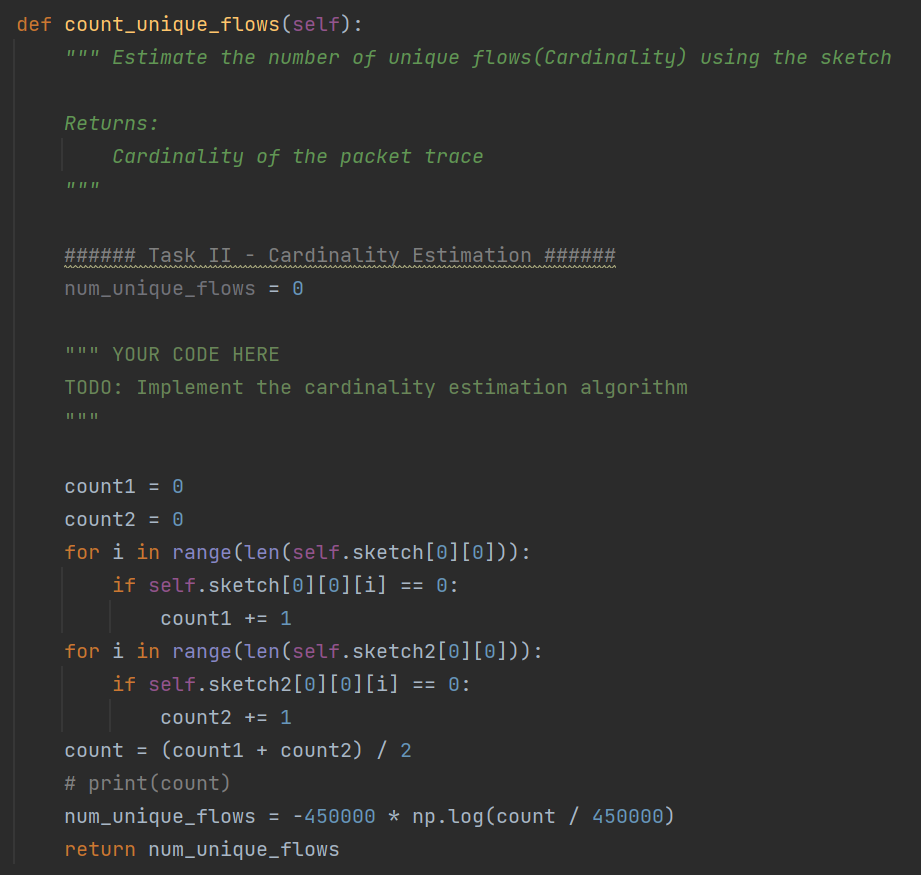
Increment（）和increment2完全按照paper上描述的方法实现，按照input的layer l和hash\_value il查询对应层的node的count情况，如果没有overflow，直接将node中的count++后退出，如果有overflow且当前layer并不是最终层，则调用increment（l + 1， il // 100）或者increment2（l + 1， il // 100），进入下一层操作，直至没有overflow发生。



estimate\_frequency(self, flow\_X)查询一个flow的size，首先调用hash\_func(flow\_X)获取两个哈希值hash\_value和hash\_value2,再分别去对应的tree里查询flow size，先查询layer 0 的node count，如果等于28-1（255），代表该node overflow，rst加上254后查询其父节点，如果layer1 的父节点count等于216-1（65535），代表该node overflow，rst加上65534后查询其父节点，以此类推。查询过程中如果遇到任何node没有overflow，立刻加上该node的count后退出查询。最后查询两个tree的结果分别为rst和rst2，取flow\_freq为min（rst，rst2）后输出



Count1为tree1 layer0的empty node counts，count2为tree2 layer0 的empty node counts， 取两者平均得到count，which is the average number of empty leaf nodes among those at stage 1。根据公式，求得num\_unique\_flows = -450000 \* np.log(count / 450000)



要计算top-100 heavy hitter，直接查询程序运行过程中维护的top-100 dict的keys和values即可，keys就是flow tuple，values就是flow size

