

## 红包数据整理说明

---

本文件夹下面有两个csv文件（导成csv文件是为了方便于选课同学后面再reverse engineering的时候用一些编程语言进行导入），包含了两类不同的红包数据。

先说一下红包的元信息，一般的红包元信息包括：每次发放的总额和发放的人数。而由于上传上来的红包数据有各种总额与人数，于是我分成了两部分。

第一部分：“fixed-RMB1-10-p.csv”这个文件是搜集到的所有1元红包（各分给十个人）的记录，专门提取它们作为一个文件是因为他们的数量较大，比较适合作为regular的dataset。表格中的每一行对应的是每个1元红包的领取记录，前十个cell的每个cell是每个人领取的数额。最后一个cell是发放总额（原则上都为1元，除了有一次的红包领取总额为0.9元，比较weird）。

第二部分：“varying-RMB-varying-people.csv”这个文件记录的是那些（红包数额，人数）不为（1元，10人）的比较“不规则”的纪录。这部分记录“不规则”的原因在于：对于固定的某个（红包数额，人数），记录数较少。将这部分区分出来，意在这部分相对“不规则”的数据既可以作为辅助的样本集合，又或者可以做为测试集，测试逆向工程出来的效果。

以上。

Xi @ SIST, ShanghaiTech