DataCon开放赛事平台

Coremail邮件安全竞赛

主办单位：清华大学 | Coremail论客

刘保君

清华大学

2020年10月26日

Coremail邮件安全竞赛背景

- 电子邮件是最常使用的攻击入口点：勒索、APT攻击

清华某教师收到的勒索邮件

我收到的APT攻击邮件

Coremail邮件安全竞赛背景

- FBI 统计：邮箱被盗案件已导致120亿美元的经济损失

Public Service Announcement

FEDERAL BUREAU OF INVESTIGATION

Jul 12, 2018

Alert Number
I-071218-PSA

Questions regarding this PSA should be directed to your local **FBI Field Office**.

Local Field Office Locations:
www.fbi.gov/contact-us/field-offices

Business E-mail Compromise The 12 Billion Dollar Scam

This Public Service Announcement (PSA) is an update and companion to Business E-mail Compromise (BEC) PSA 1-050417-PSA posted on www.ic3.gov. This PSA includes new Internet Crime Complaint Center (IC3) complaint information and updated statistical data for the time frame October 2013 to May 2018.

FBI 2018年评估所得数据

TECH

Homeland Security Chief Cites Phishing as Top Hacking Threat

BY JEFF JOHN ROBERTS
November 21, 2016 2:30 AM GMT+8

PHOTOGRAPH BY SAUL LOEB—AFP/GETTY IMAGES

Most Popular

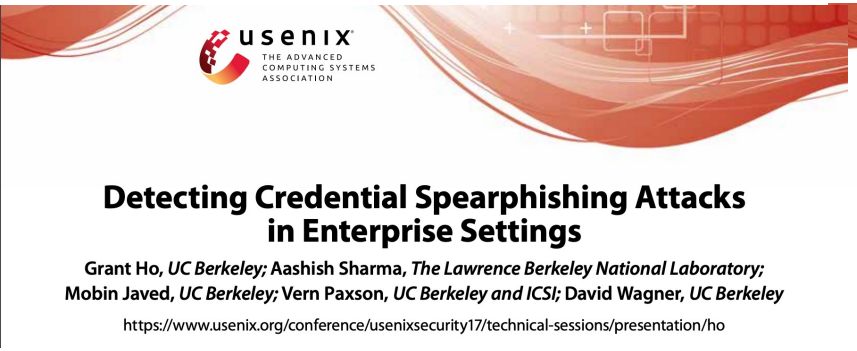
Trump's odds of winning rise or f based on one overriding force
BY SHAWN TULLY

When second stimulus checks co arrive if Congress reaches a deal
BY LANCE LAMBERT AND ANNE SRADERS

国土安全局：网络钓鱼仍是头号安全威胁

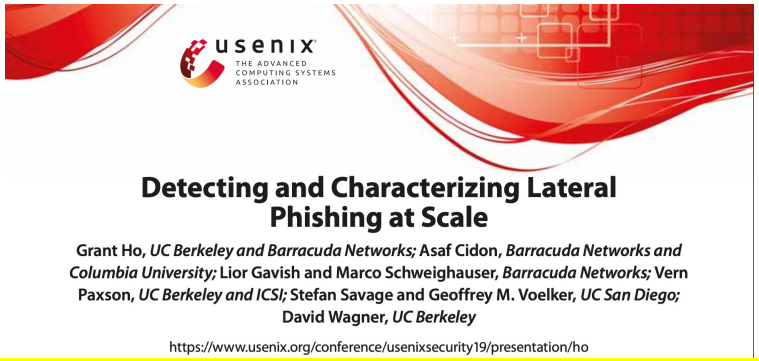
Coremail邮件安全竞赛背景

- 近年来，邮件安全已经成为学术研究的热点

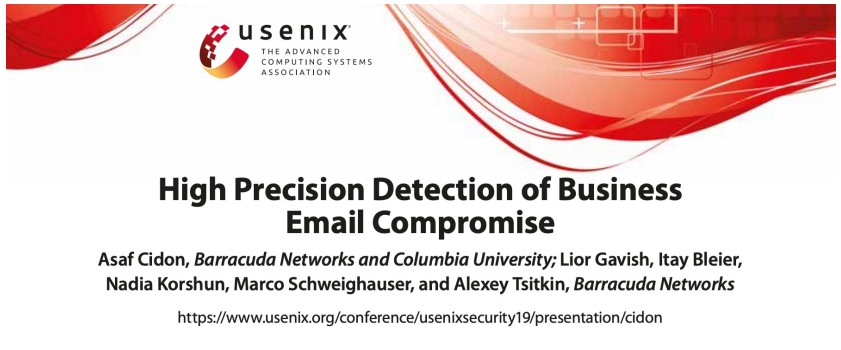


USENIX ’17 Distinguished Paper

USENIX Security ’18

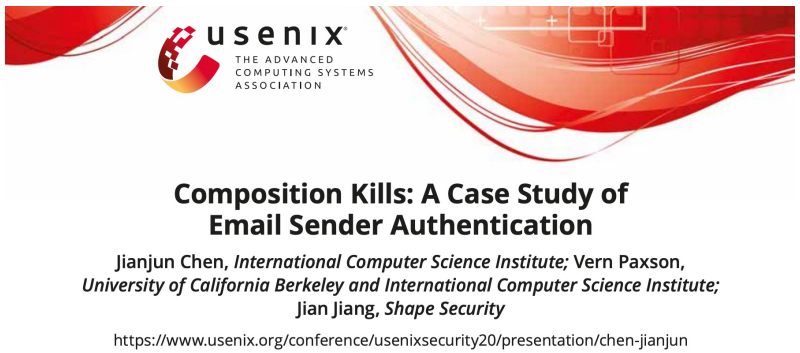


USENIX ’19 Distinguished Paper



USENIX Security ’19

USENIX Security ’19



USENIX ’20 Distinguished Paper

假冒发信人的影响的邮件服务商

QQ邮箱 mail.qq.com

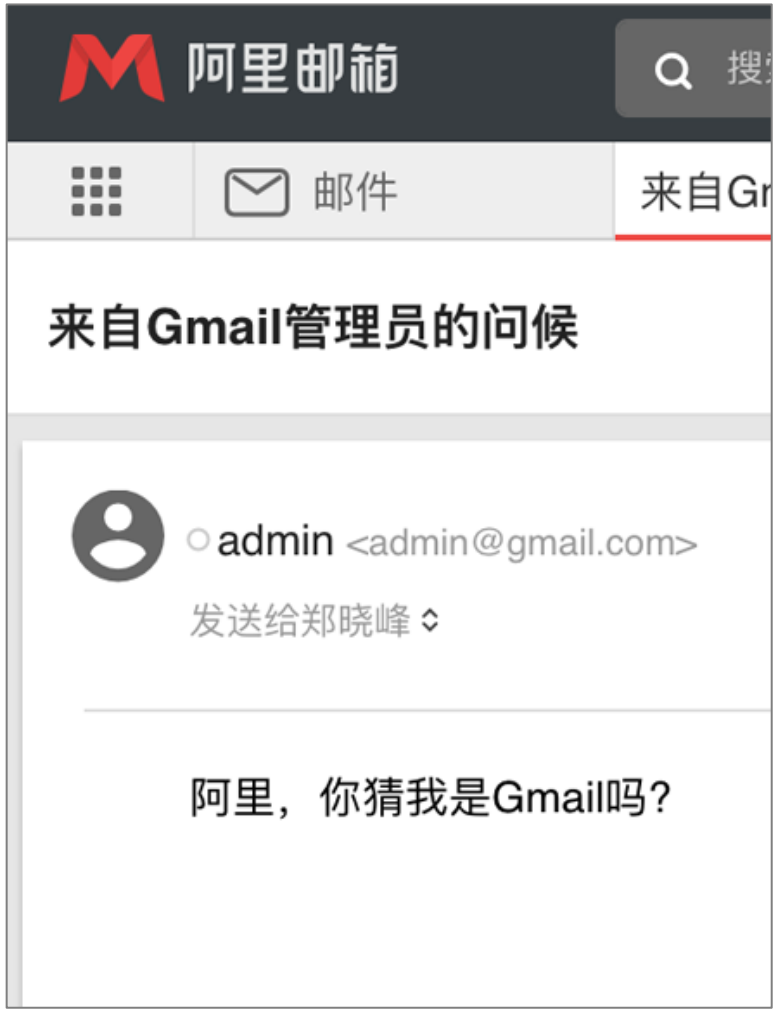
新浪邮箱

阿里邮箱

搜狐闪电邮箱 mail.sohu.com

中国电信 CHINA TELECOM | 189邮箱

中国移动通信 CHINA MOBILE | 139邮箱



清华-奇安信联合研究中心与Coremail合作发表的研究论文被USENIX Security '21录用

Weak Links in Authentication Chains:
A Large-scale Analysis of Email Sender Spoofing Attacks

Kaiwen Shen [1,*], Chuhan Wang [1,*], Minglei Guo [1], Xiaofeng Zheng [1,2,†], Chaoyi Lu [1], Baojun Liu [1,†], Yuxuan Zhao [4], Shuang Hao [3], Haixin Duan [1,2], Qingfeng Pan [5] and Min Yang [6]

[1]Tsinghua University [2]Qi An Xin Technology Research Institute [3]University of Texas at Dallas
[4]North China Institute of Computing Technology [5]Coremail Technology Co. Ltd [6]Fudan University

Abstract

As a fundamental communicative service, email is playing an important role in both individual and corporate communications, which also makes it one of the most frequently attack vectors. An email's authenticity is based on an authentication chain involving multiple protocols, roles and services, the inconsistency among which creates security threats. Thus, it depends on the weakest link of the chain, as any failed part can break the whole chain-based defense.

This paper systematically analyzes the transmission of an email and identifies a series of new attacks capable of bypassing SPF, DKIM, DMARC and user-interface protections. In particular, by conducting a "cocktail" joint attack, more realistic emails can be forged to penetrate the celebrated email services, such as Gmail and Outlook. We conduct a large-scale experiment on 30 popular email services and 23 email clients, and find that all of them are vulnerable to certain types of new attacks. We have duly reported the identified vulnerabilities to the related email service providers, and received positive responses from 11 of them, including Gmail, Yahoo, iCloud and Alibaba. Furthermore, we propose key mitigating measures to defend against the new attacks. Therefore, this work is of great value for identifying email spoofing attacks and improving the email ecosystem's overall security.

email depends on the weakest link in the authentication chain. Even a harmless issue may cause unprecedented damages when it is integrated into a more extensive system. Generally, the email authentication chain involves multiple protocols, roles and services, any failure among which can break the whole chain-based defense.

First, despite the existence of various security extension protocols (e.g., SPF [24], DKIM [2] and DMARC [31]) to identify spoofing emails, spoofing attacks might still succeed due to the inconsistency of entities protected by different protocols.

Second, authentication of an email involves four different roles: senders, receivers, forwarders and UI renderers. Each role should take different security responsibilities. If any role fails to provide a proper security defensive solution, an email's authenticity can not be guaranteed.

Finally, security mechanisms are implemented by different email services with inconsistent processing strategies. Besides, those security mechanisms are implemented by different developers, some of which deviate from RFC specifications while dealing with emails with ambiguous headers. Therefore, there are a number of inconsistencies among different services. Attackers can utilize these inconsistencies to bypass the security mechanisms and present deceptive results to the webmails and email clients.

2020/09/19，Coremail邮件安全竞赛启动

- 目标：在实战环境下培养和挖掘邮件安全领域的研究型人才

Coremail邮件安全竞赛 赛制安排

- 参赛人员：不做限制
  - 高校　　高等学校在校学生或教师
  - 科研院所　　科研人员及网络管理人员
  - 企业　　网络安全研究人员，邮件安全从业者

- 赛程设置：组队参加，为期一周
  - 9.19 - 10.8　　自由组队报名
  - 10.9 - 10.15　　线上积分赛
  - 10.26　　线下答辩 & 颁奖典礼 & 技术分享

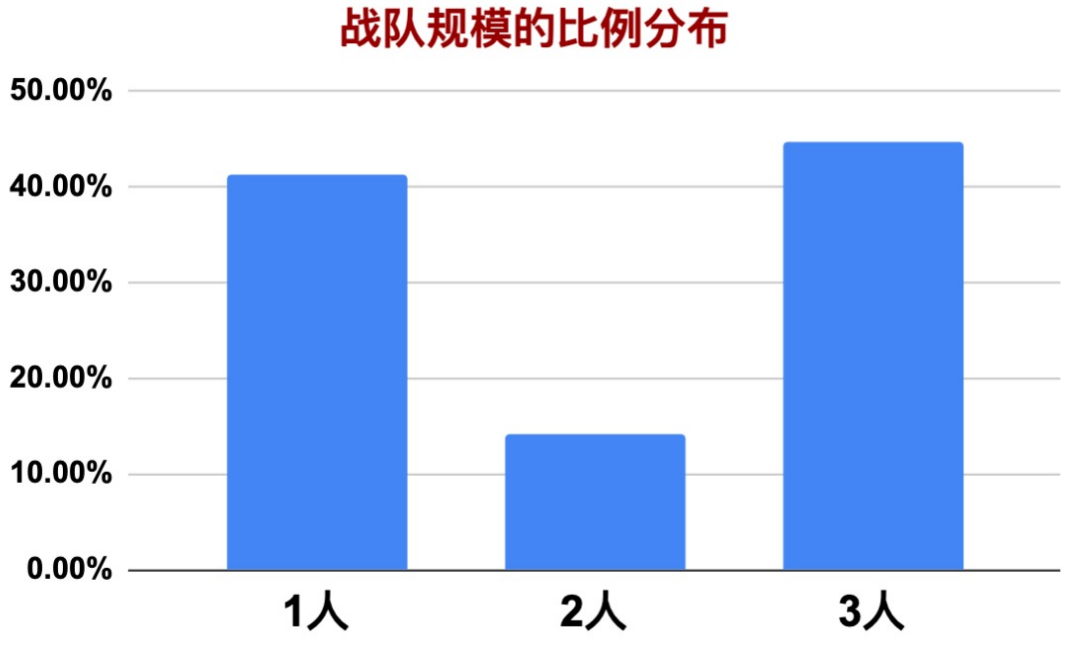
Coremail邮件安全竞赛 赛题设计

- 出发点：尽可能贴近实战，解决实际痛点，难度依次递增
  - 提供大量贴近于真实生产环境下的细粒度数据
  - 赛题均源自于企业邮件安全中的实际应用场景
  - 赛题难度梯度增加；统一采用Docker环境，保护数据

- 第一题：火眼金睛
  - 难度：较易　　识别“发件人伪造”攻击行为
- 第二题：明察秋毫
  - 难度：中等　　分析邮件正文的类型
- 第三题：孤胆猎手
  - 难度：较难　　判断邮件的安全威胁等级

线上积分赛 报名情况

- 累计 **114** 支队伍报名参赛
  - ❖ **高校** 清华 山大 武大 东南 …
  - ❖ **企业** 中国移动 启明星辰 …



战队规模

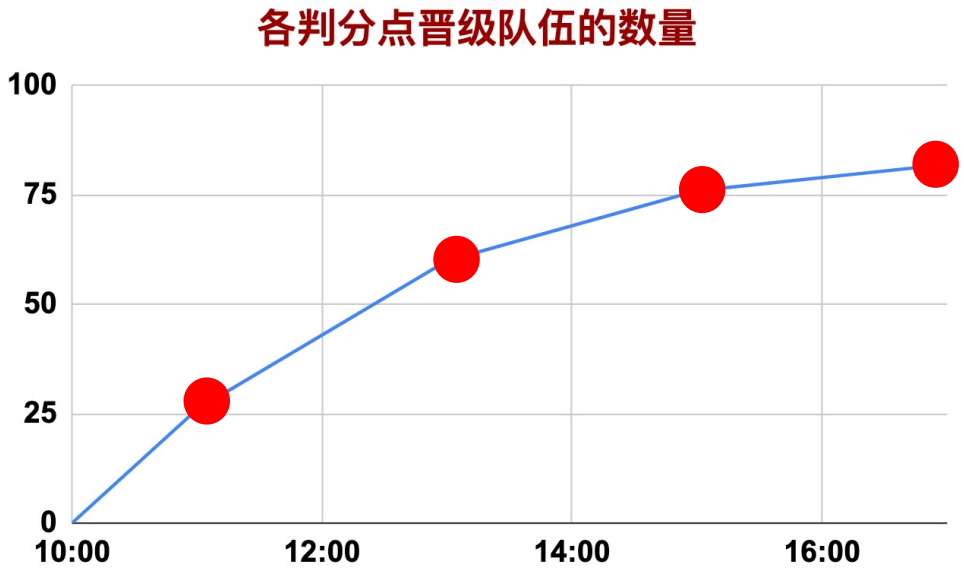
战队所属单位

抢位资格赛

- 资格赛：数据安全类竞赛的积极探索
  - ❖ 如何过滤僵尸队？ 如何提高竞赛的活跃度？
  - ❖ 限制晋级名额，引入抢位资格赛作为前置的热身环节

作为一个好男朋友，通过长期的学习，小明已经非常了解他的女朋友是否喜欢一种食物，小明这种神奇的能力属于：

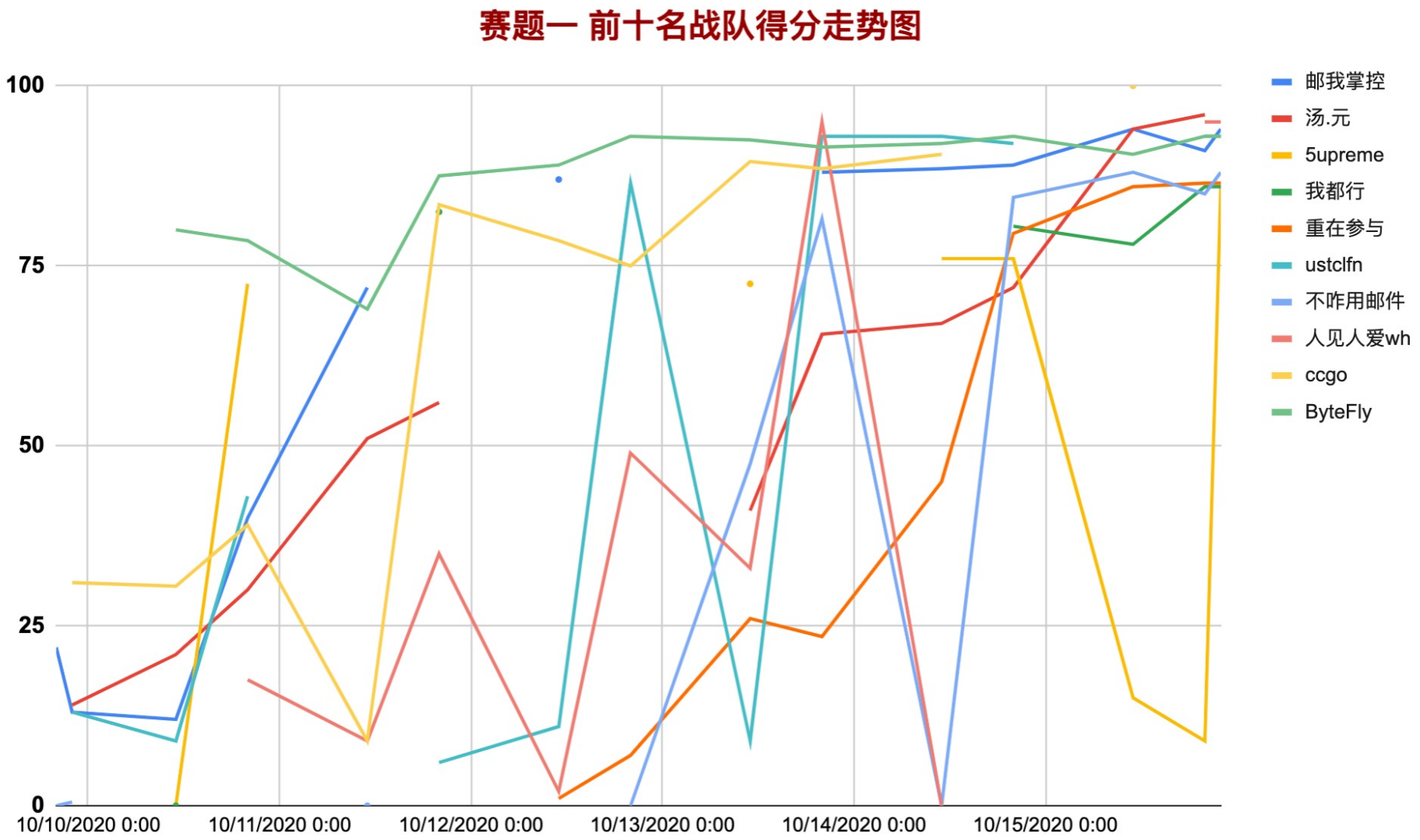
A：分类 （正确答案）
B：回归
C：聚类
D：随机猜测



抢位资格赛
在难度与趣味性之间取得平衡

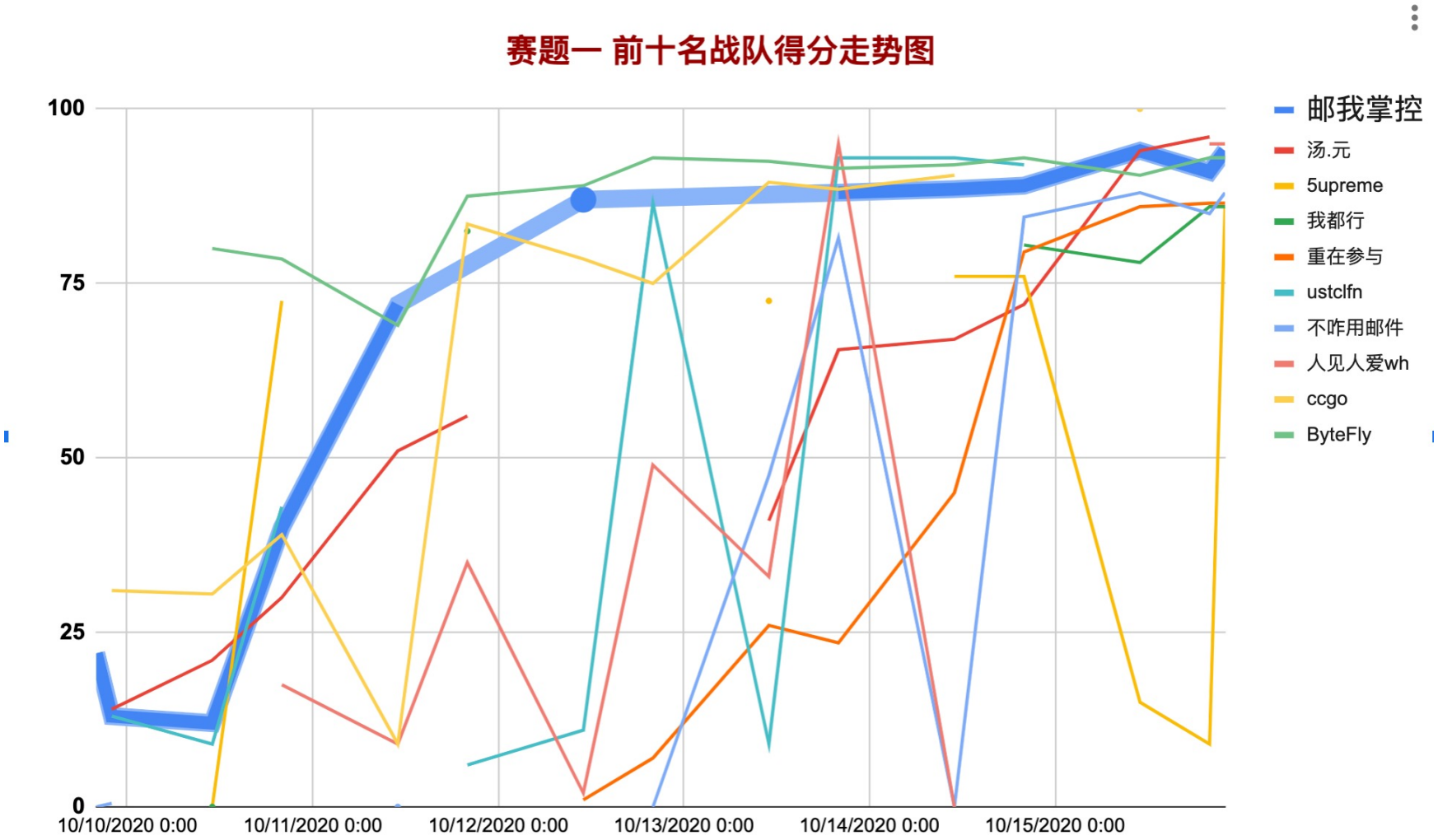
上午十点，竞赛开始，
下午五点，共有82支队伍通过资格赛！

赛题一 火眼金睛 得分走势图



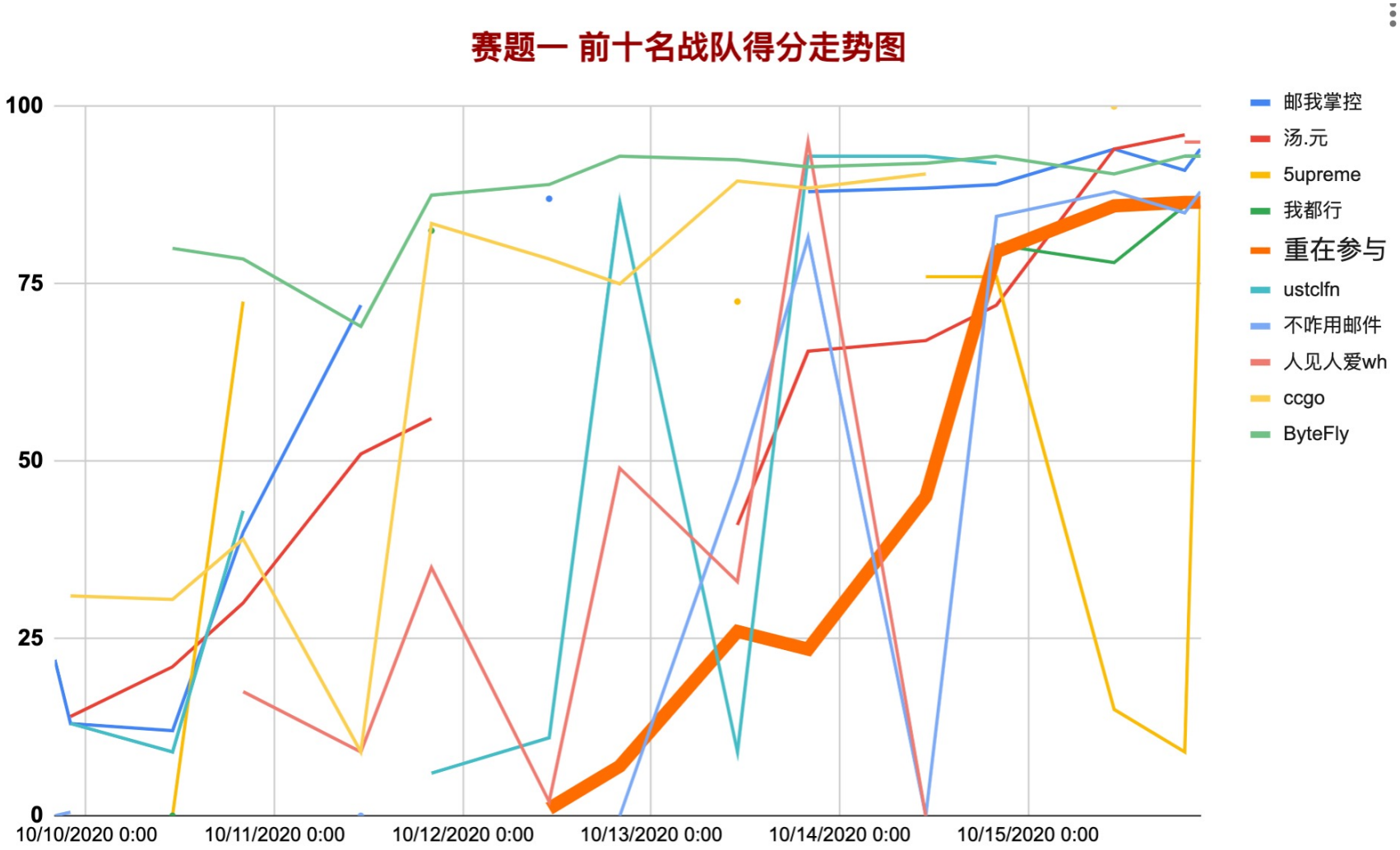
赛题一 火眼金睛 得分走势图

- 案例一： 邮我掌控战队，稳扎稳打



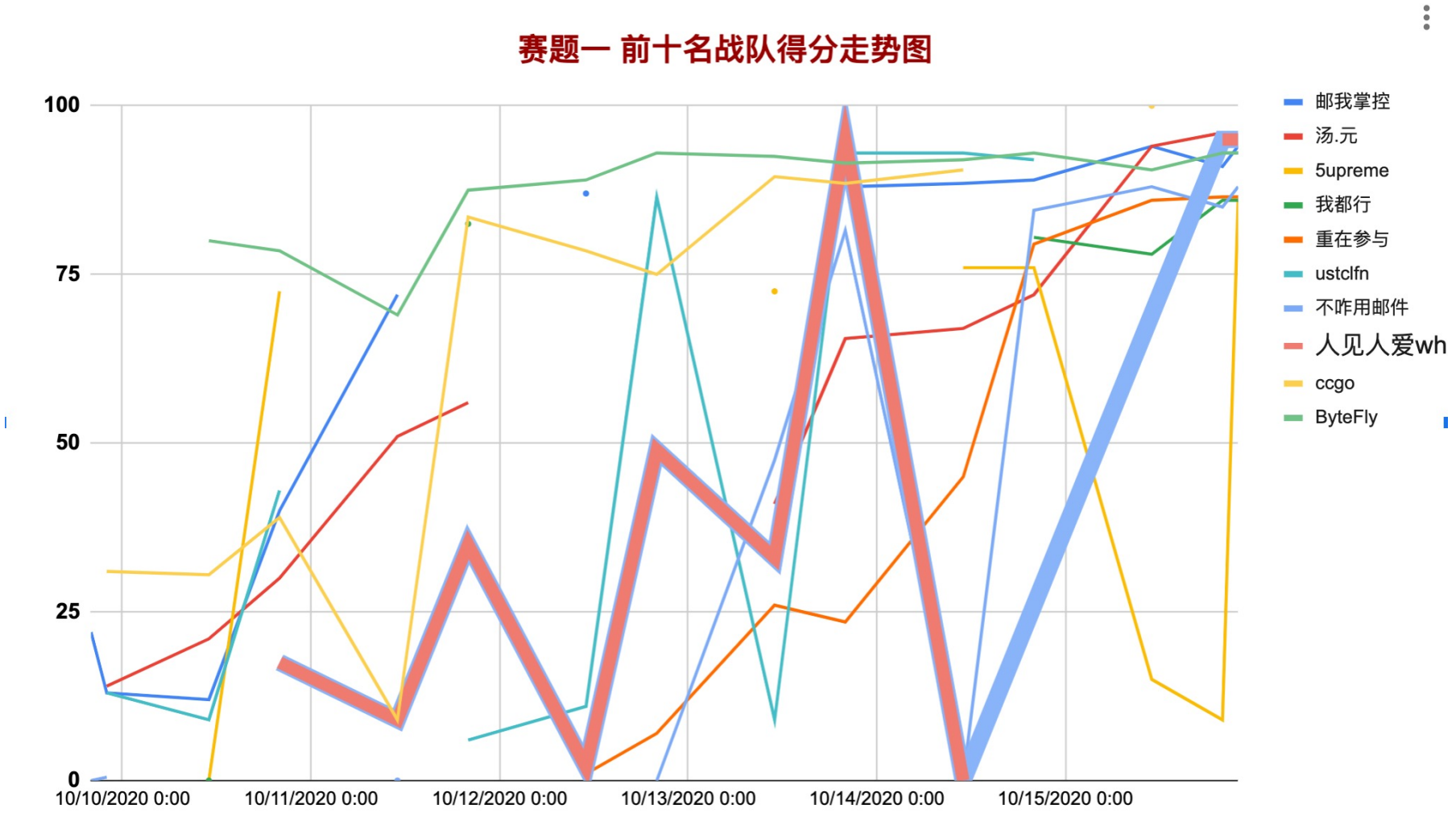
赛题一 火眼金睛 得分走势图

- 案例二： 重在参与，后半程开始发力

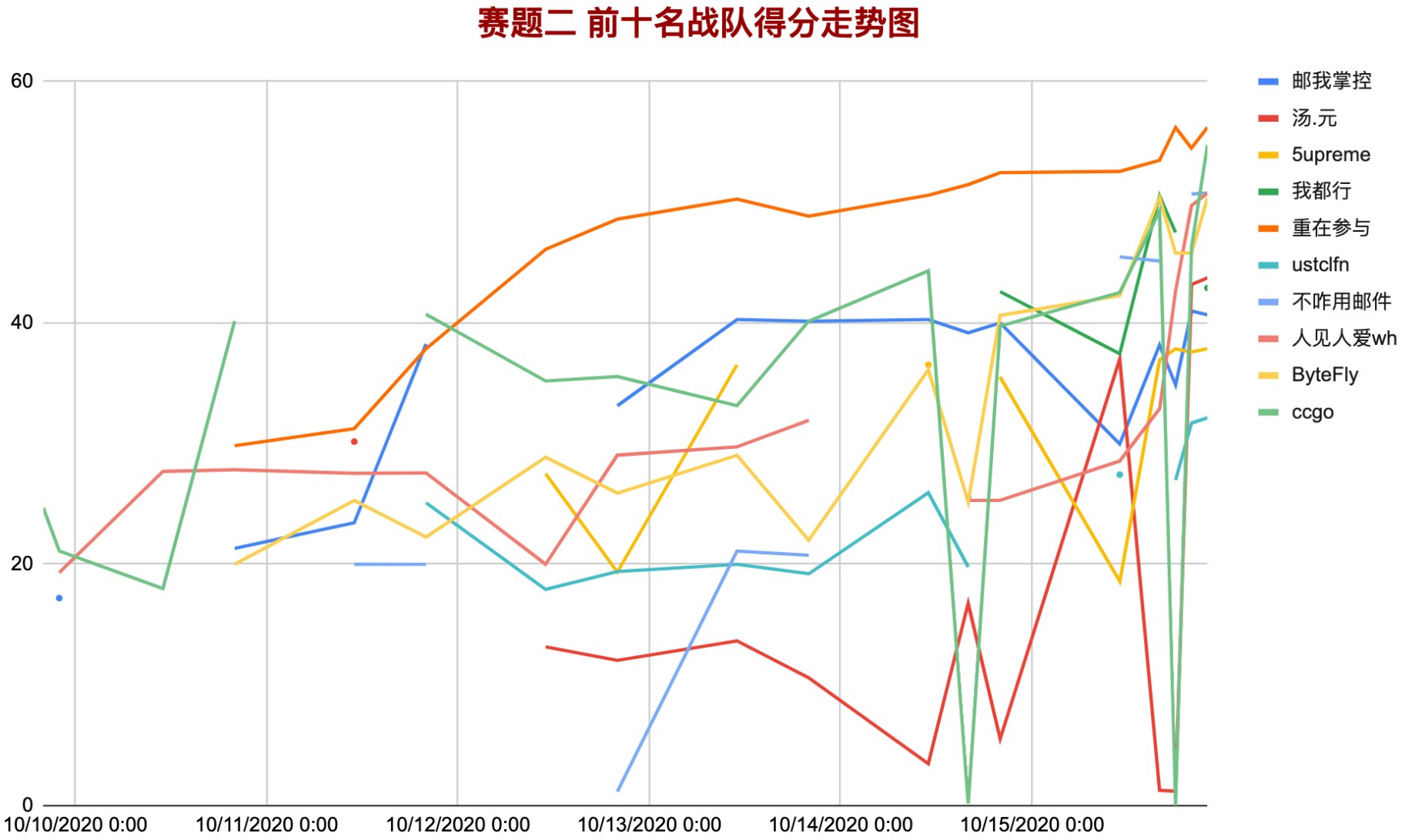


赛题一 火眼金睛 得分走势图

- 案例三： 人见人爱WH，利用判分反馈调整答案

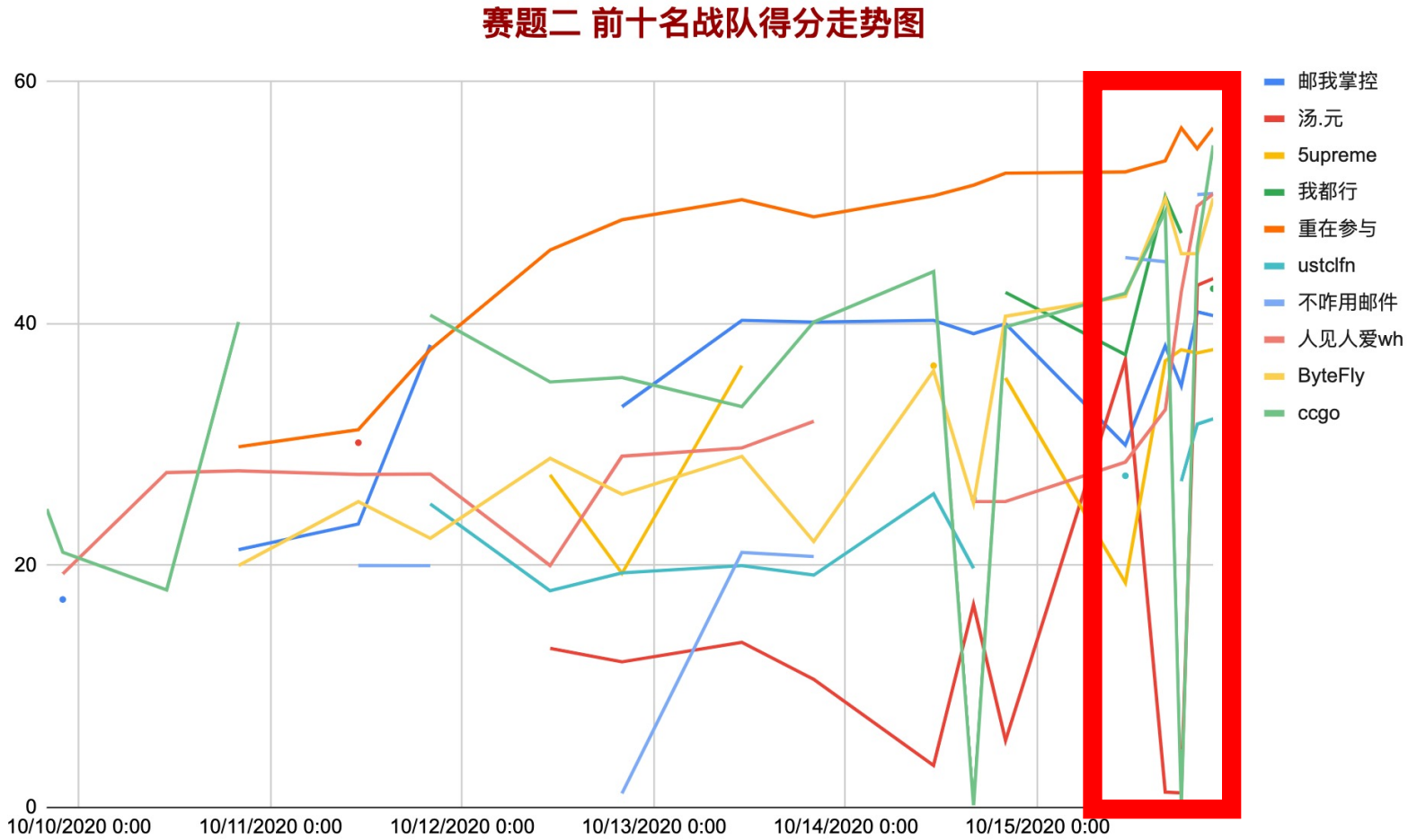


赛题二 明察秋毫 得分走势图



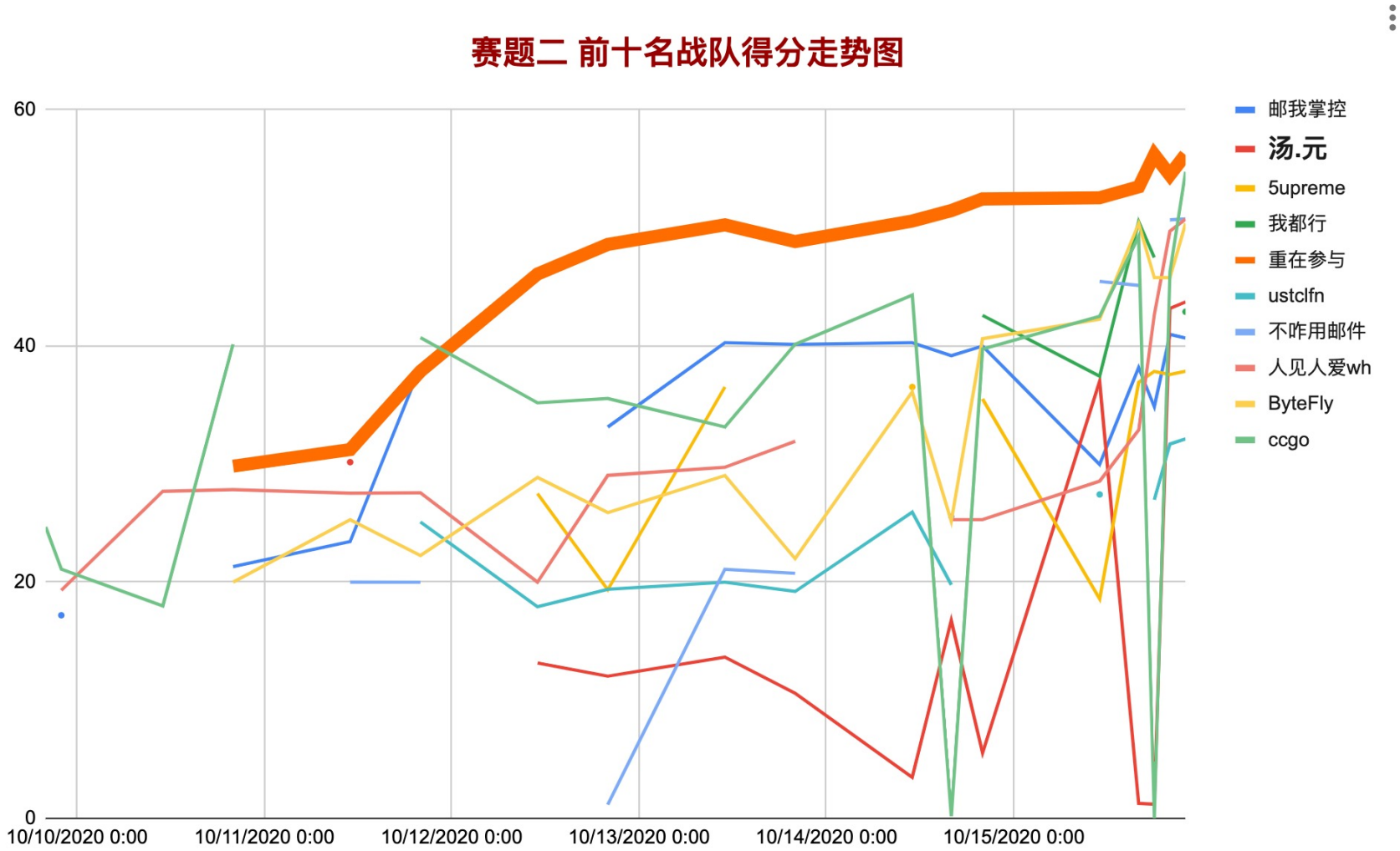
赛题二 明察秋毫 得分走势图

- 案例一：用放hint的形式，搅动一潭死水



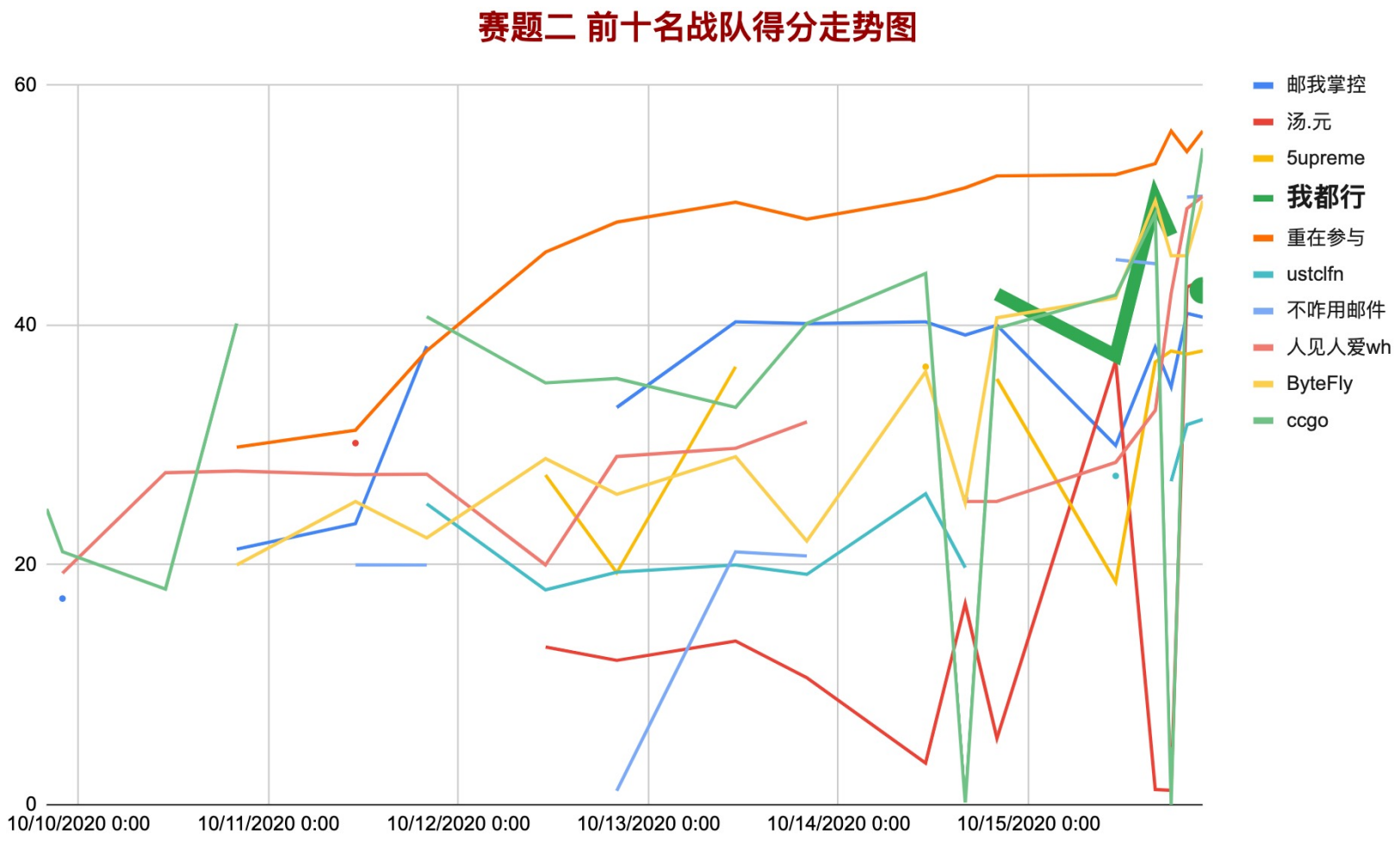
赛题二 明察秋毫 得分走势图

- 案例二： 汤. 元，赛题投入的时间较长，成绩最好

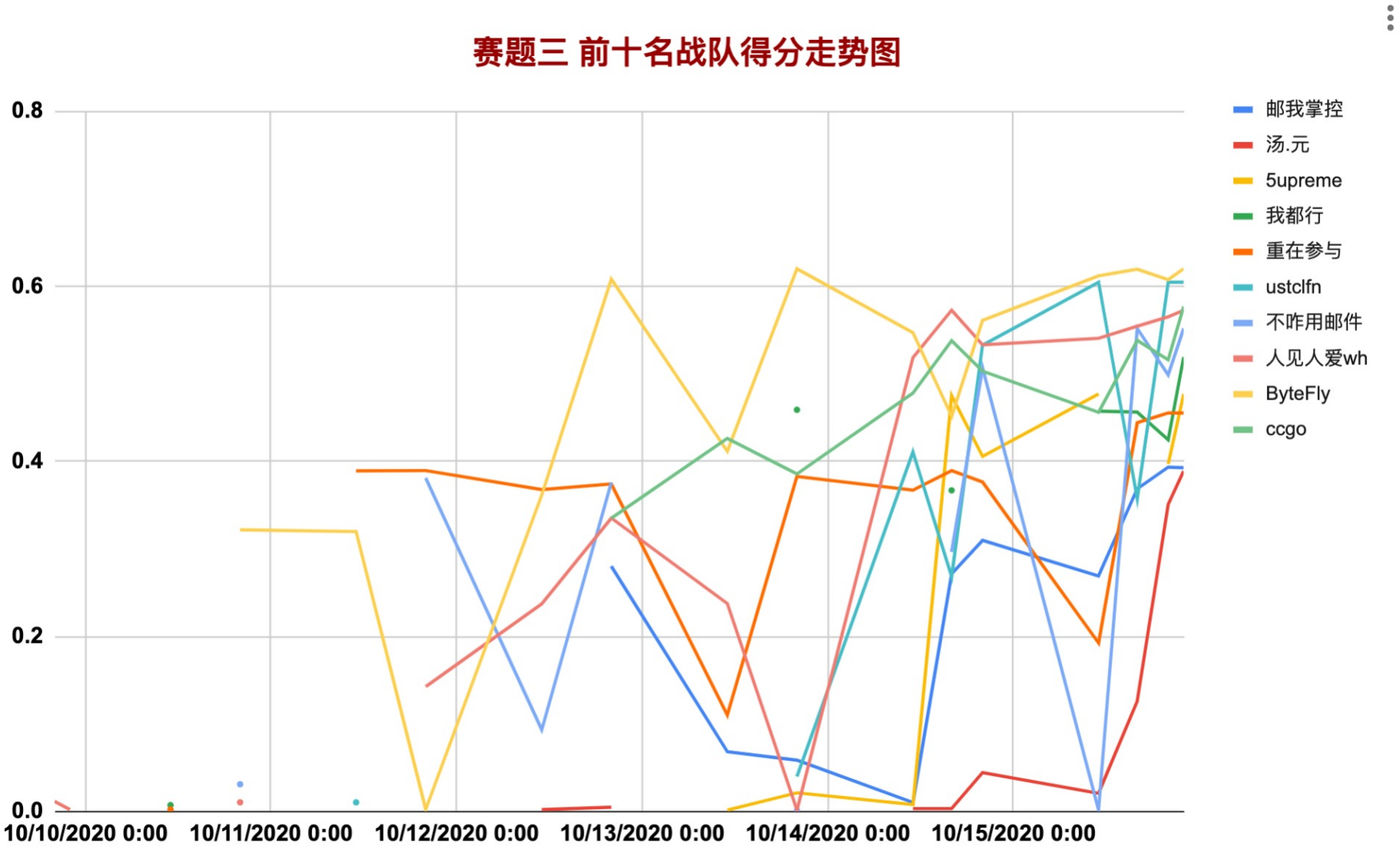


赛题二 明察秋毫 得分走势图

- 案例三： 我都行，百米冲刺型选手

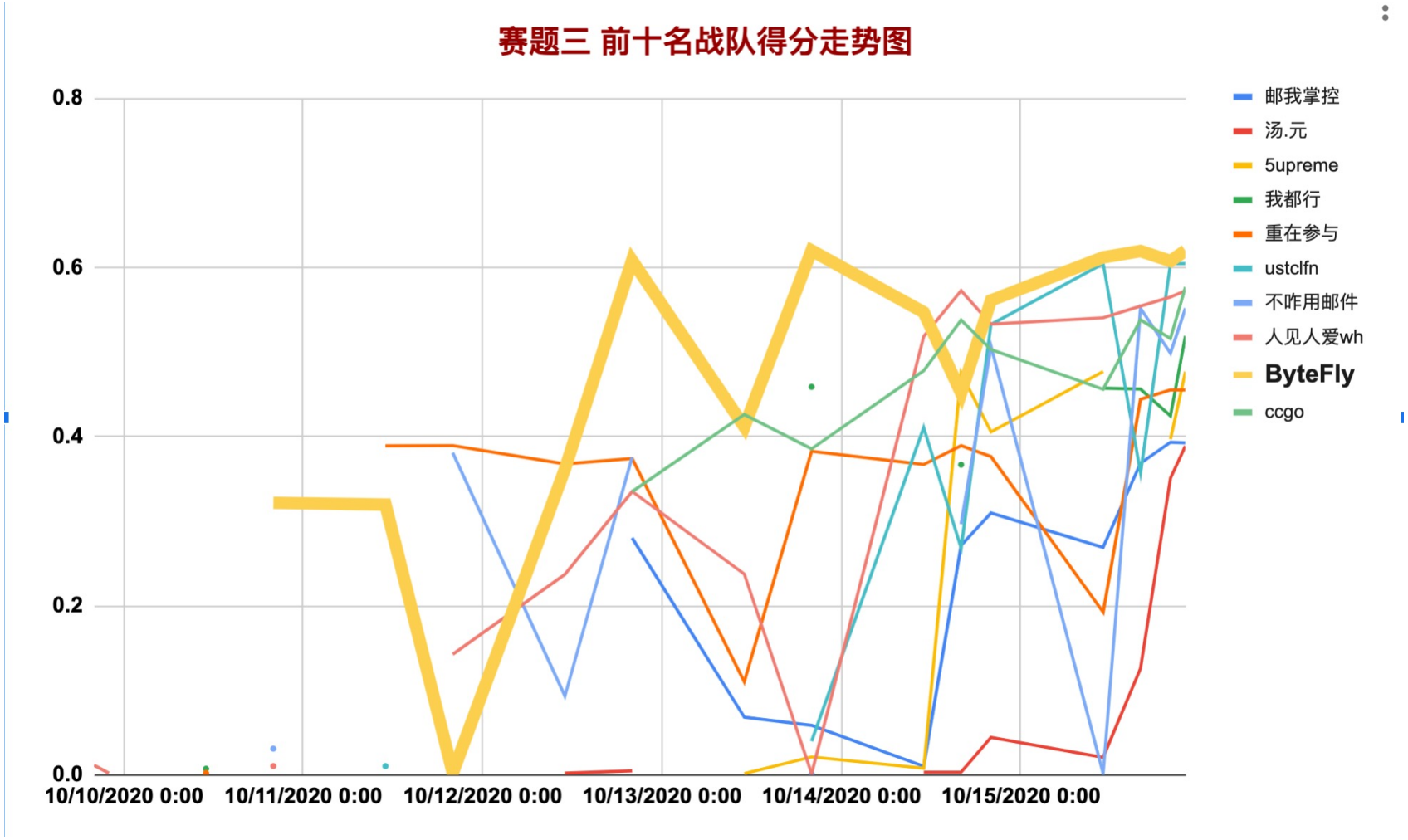


赛题三 孤胆猎手 得分走势图



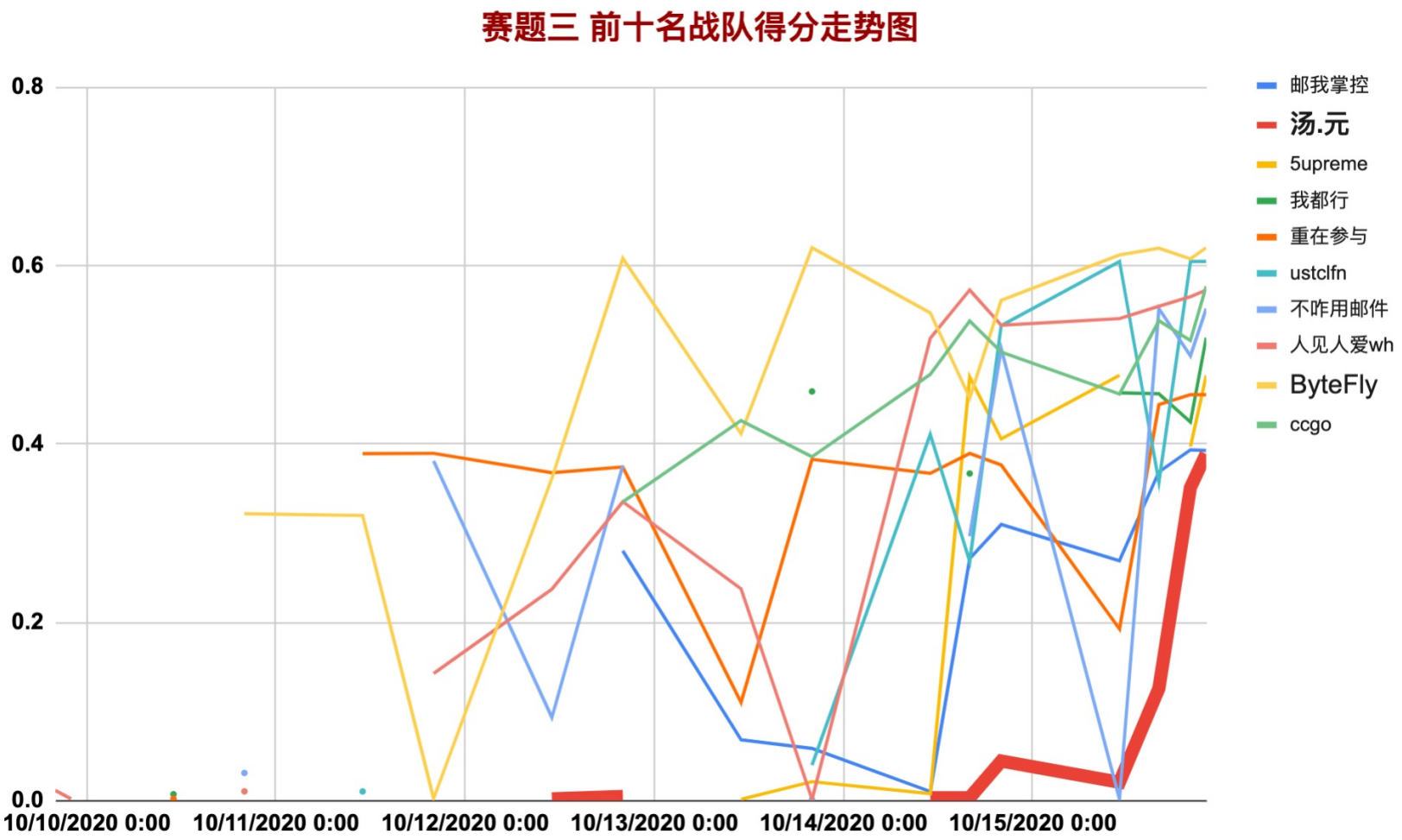
赛题三 孤胆猎手 得分走势图

- 案例一：ByteFly，最早开始提交，得分优于出题人预期



赛题三 孤胆猎手 得分走势图

- 案例二：汤. 元，最后时刻绝杀，逆袭竞赛排行榜



竞赛趣事分享

- 一场让主办方感到头皮发麻的竞赛
  - 尽管使用Docker环境保护数据，但“赛棍”总能设法绕过限制
  - 赛题一火眼金睛：最终被多支参赛队打穿，取得满分成绩
  - 赛题三孤胆猎手：出题人预期F1-Score最高约为0.4，实际超过0.6
- 乌龙：两支参赛队伍，在一个判分点得分完全相同
  - 审核代码后发现，两支队伍思路完全一致，非常简单，却也很巧妙

```
from_name =[]
susepect_mail_index=[]
with open('/home/datacon/coremail/challenge_3/player_data_for_team0_new.json','r') as fid:
    challenge3_data = fid.readlines()
    for mail_data_idx in range(len(challenge3_data)):
        if 'admin' in json.loads(challenge3_data[mail_data_idx][:-1])['fromname']:
            susepect_mail_index.append(mail_data_idx)
            from_name.append(json.loads(challenge3_data[mail_data_idx][:-1])['fromname'])
```

```
import json

result = []
date_list_json = []
count = 0
# Opening JSON file
with open('/home/datacon/coremail/challenge_3/player_data_for_team0_new.json') as f:
    for line in f:
        # returns JSON object as a dictionary
        data = json.loads(line)
        date_list_json.append(data)
        region = data['region']
        fromname = data['fromname']
        if 'admin' in fromname: #('越南' in region or '马来西亚' in region) and
            result.append(int(data['tag']))
```

参赛队伍的反馈与建议

如何看待2020Coremail邮件安全竞赛?

lvzzzz

段老师课程一个作业，然后就很神奇来参加了，并且在大佬的带领下进了前10，3道题目质量都很高，同实验室的小伙伴们前期出题验题也花了不少时间，感谢主办方提供这样的一次机会让我认识到了中文的博大精深，哈哈哈。唯一一点吐槽吧：判题能不能快点

发布于 10-17

如何看待2020Coremail邮件安全竞赛?

匿名用户

数据都是好数据，挺真实的，中文英文日文俄文西班牙语的垃圾邮件居然都有-0-

能有这类比赛已属不易，but，就不夸了吐槽一下吧。

既然是要吐槽，那么先说一下整体的吐槽的点吧。

时间实在是太紧张了，这又不是CTF，连着肝七天肝要坏了。然而比起参加过的数据方面的比赛 1 没遇见过数据没标的，不过这不是问题还可以搞搞聚类，但是也不知道聚成几类就瞎聚，或者干脆其实不用聚类，还是我菜 2 不能实时测评？不能实时测评？不能实时测评？ 3 不做A/B test？ 4 比赛情况居然根据答题情况变？最后几小时还给提示我也是妹想到

参赛队伍的反馈与建议

如何看待2020Coremail邮件安全竞赛?

123
学生

我觉得很有意思，第一次参加数据分析的题目，还挺好玩，就是比赛时间太长了！我这一周天天做这个题，都没时间写作业了。就是自己太菜了，但是比较有参与感，还有第一题怎么分都这么高?
菜哭

发布于 10-15

▲ 赞同 ▼ 添加评论 分享 收藏 喜欢 ···

lerry13579
计算机

一打开稍微大一点的文件就直接卡死，这个很难受。。。

发布于 10-15

如何看待2020Coremail邮件安全竞赛?

CheckSorry
程序员

2 人赞同了该回答

赛题一

我们是菜鸡也要试一下队，我们是第一次参加比赛，不是做安全出身，对邮件安全也不了解，过滤规则都是参加本次比赛现学现卖找出的。比赛一开始觉得Received字段是一个强特征，但是解析麻烦，后来也发现这个字段可以伪造，且过滤出结果太多，不太准确，同时前两天有些别的事情没有花太多时间阅读协议字段含义导致了没有成功得分。后来，第三天定下心来认真阅读协议头部字段含义后，逐渐找到了有效的过滤规则，分数逐渐提升，第四天晚上冲到了96分，同时那晚提交之后找出了最后一个规则，最后所有规则结果去重合出来是101，其他两题还没有太大进展，所以直接交了101个。最后发现是之前有一个规则代码写错导致结果出错，去除错误答案后，有一次服务器崩坏导致有一个错误答案残留在了之前的结果list里，当时不知道服务器崩坏导致之前的结果list变了，还以为就是101个，最后在和96分结果比对后找出了那个残留的错误答案，得到了100分。

通过做这道题还是学到了“发件人伪造”攻击原理和邮件协议头字段含义相关的邮件安全知识。菜鸡第一次参赛可以得到一题满分，很开心了！

总结

- 祝 2020 Coremail邮件安全竞赛圆满结束！
- 不过，这也仅仅只是一个开始
  - 学术研究：将竞赛中数据集，开放用于学术研究
  - 企业实习：Coremail论客 & 奇安信集团