2019 年 CCF 大数据与计算智能大赛

企业网络资产及安全事件分析与可视化赛题

复赛答卷

参赛队名称： 飞向月球

团队成员： 董博，西安交通大学，dong.bo@mail.xjtu.edu.cn，指导老师

张发，西安交通大学，fazhang@stu.xjtu.edu.cn，队长

查志超，西安交通大学，810240960@qq.com，队员

吴映潮，西安交通大学， xjtuwuyc@163.com，队员

杨育婷，西安电子科技大学，944866457@qq.com，队员

是否学生队（是或否）： 是

使用的分析工具或开发工具（如果使用了自己研发的软件或工具请具体说明）：Echarts，Jupyter Notebook，MySQL，Pycharm，flask

共计耗费时间（人天）： 10 人天

（灰色字为参赛信息填写模板，请参赛者在提交时参照模板填写）

1.1：通过对两周的网络安全日志数据的协同可视分析，找出这两周公司内部网络中可能存在的异常通信模式（异常事件），异常标准不限，如访问或被访问量突变、周期行为模式、连接模式变化、活跃时段变化、行为不符合主机角色等等。（请将回答限制在 1200 个字和 10 张图片，请总结不超过 10 个异常通信模式）

1）周期行为异常模式

我们对 db 表格进行统计，发现所给的数据都是呈现成对出现，并具有一定的周期行规律，大多数是 SET 与 SELECT 操作与 AUTOCOMMIT 相结合；且相邻时间段趋势相似，考虑到都是通过 mysql 协议进行控制，因此我们忽略该因素的影响。因此我们结合操作时间和星期对 db 的操作信息进行可视化分析。如图：1-1.

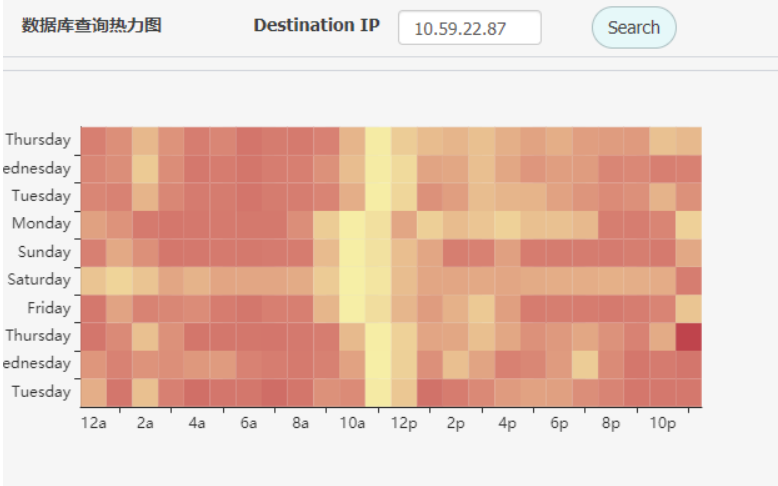

图 1-1

该热力图横轴表示一天 24 小时，纵轴表示天，图中块表示访问频次。鼠标悬浮可以看到当天的该时间段内 mysql 的访问量。我们可以清楚的观察到，红色块在晚上 23 点开始有大量的数据库访问。我们点击红色块，可以对具体信息进行观察，如图 1-2.

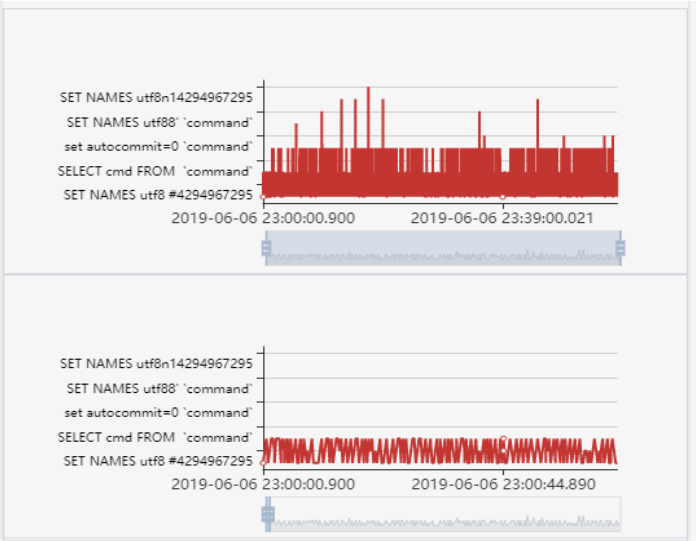

图 1-2

横轴表示时间，纵轴表示 MySQL 操作命令，正常的操作命令周期性变化如图 1-2 下图所示，而纵观当前时间后一段时间，如图 1-2 上图所示，出现了访问周期异常行为，我们认为这样的突变访问存在一定的嫌疑。

2）http 访问注入异常

我们对 flow 表进行分析，判断是否有注入访问异常。我们可以根据访问连接中的 uri 进行检测，我们发现存在大量异常不明确字符，且 POST 传递参数混杂，因此，我们对 method，uri，host，useragent 进行特征提取。首先使用停用词进行分词处理，其次对所得分词进行 tfidf，然后，对 method 进行 one_hot 编码，最终得到特征向量表，然后我们利用异常检测常用的三个算法对提却到的特征进行预测，预测的结果如图所示 1-3 所示。

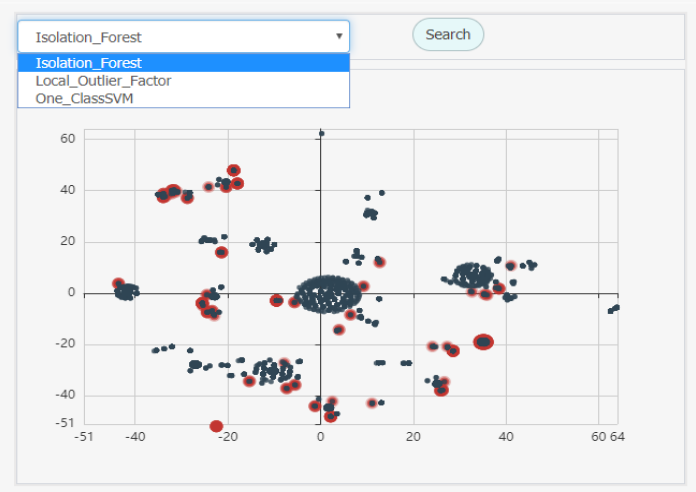

图 1-3

我们选择 Isolation_Forest 为例，主要是利用集成学习的思想结合决策树进行异常点检测。显示结果如图所示，红色节点表示算法认为是异常的节点，鼠标悬浮，我们可以得到具体的异常信息，如图 1-4 所示。

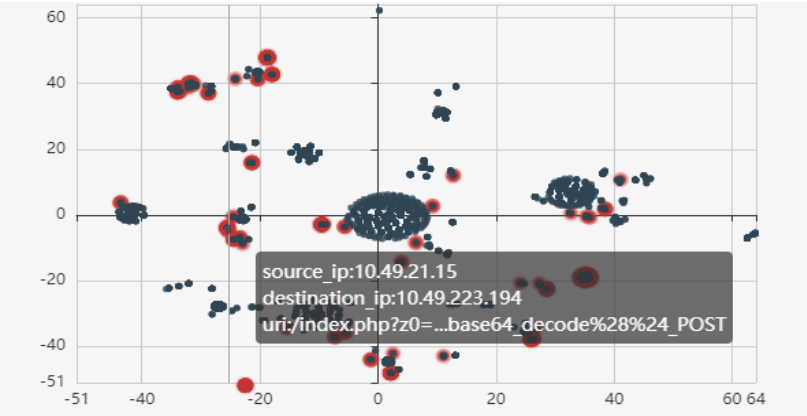

图 1-4

悬浮的信息中包括 base64，JavaScript 等信息，查阅相关的资料可知该编码是可执行网页脚本的注入方式，因此该条访问属于 http 访问注入异常。

3）访问量突变

我们对 tcpflow 表进行分析，发现包含多种不同的协议进行通信，我们选取其中的 TOP5 进行分析，对不同的通信协议进行下行流量的分析，对一天不同通信协议的下行总流量进行课时化分析。如图 1-5 所示.

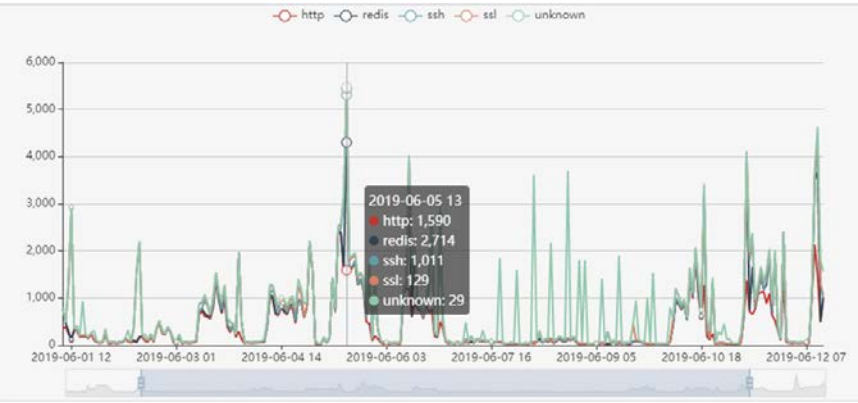

图 1-5

我们可以看到在 13 时 http 和 redis 访问属于大量异常访问时间，这些时间段访问量的突变我们认为它是嫌疑异常通信。

4）登陆异常

我们对 login 表进行分析，根据统计分析可知，一共有 17 种登陆日志记录。我们去除成功登陆的登陆日志信息，对失败或者异常授权等登陆信息进行展示，如图 1-6.

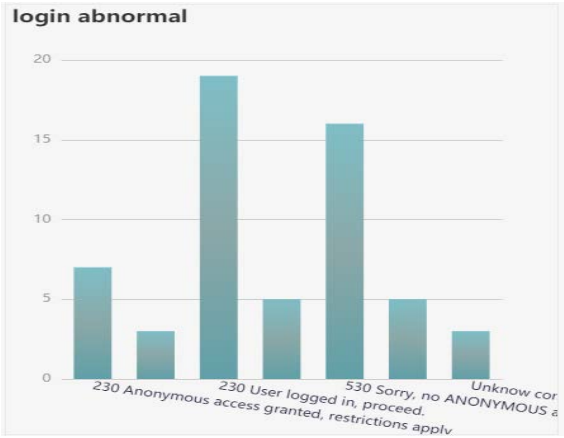

图 1-6

我们对登陆用户异常信息进行展示，如图 1-7 所示。

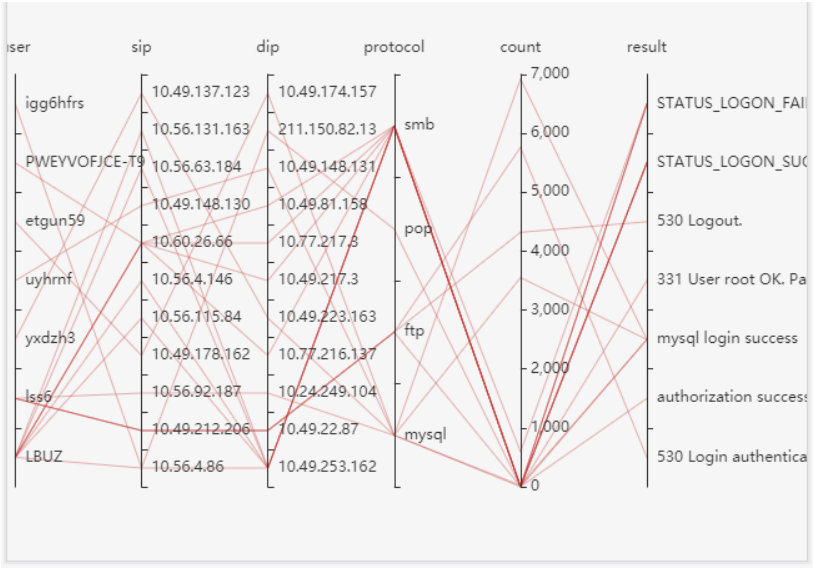

图 1-7

该图中线条颜色越深，表明该访问过程次数较多。可以看到登陆失败操作机器的 ip 及其对应的协议。

5）流量访问异常

由于该公司发生了数据泄露，因此会出现流量访问的异常，我们分析 tcpflow 协议，并对其进行可视化展示，如图 1-8.

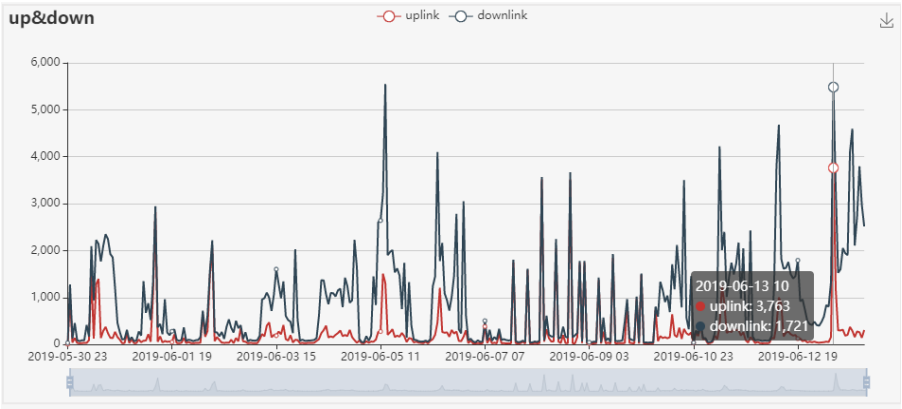

图 1-8

我们可以看到 2019 年 6 月 13 日晚上 10 点，上行流量和下行流量都出现了较大的变化，不同于正常访问的流量，因此，我们推测，该日晚上可能发生了数据泄露。我们认为这是属于流量异常。

6）通信协议异常

我们对指定端口和协议进行筛选搜索，如图 1-9，左图的桑基图是指定通信协议下前 top100 结果图，宽度表示通信的流量大小，我们提供交互功能，点击可以在右侧显示源和目标 IP 具体的访问信息，鼠标悬浮可以看到上行流量和下行流量及控制协议.

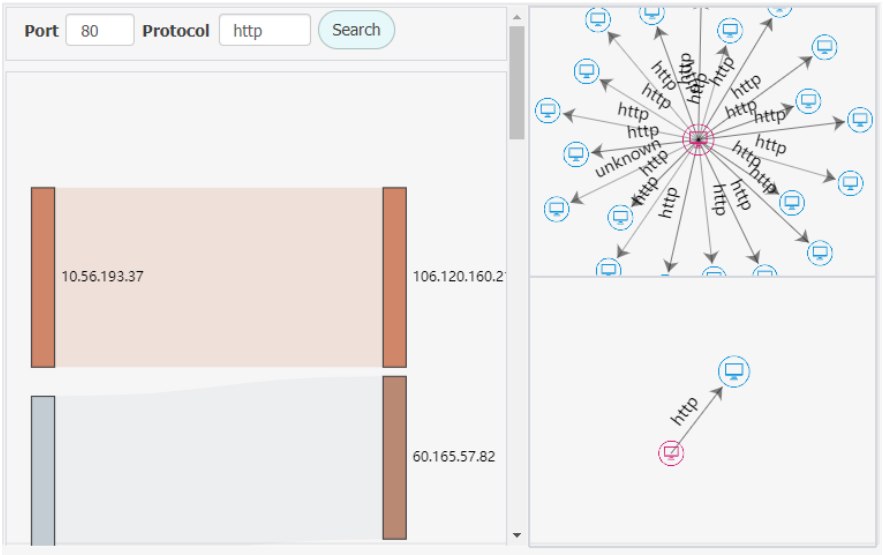

图 1-9

1.2：找出这些异常通信模式中最可能与本次“安全问题”相关的模式，并综合分析本次“安全事件”发生的时间、相关网络资源、事件的演化过程、原因和结果。（请将回答限制在 1200 个字和 10 张图片）

根据上述分析，我们可以推断，该公司最有可能的安全问题模式是属于“http 注入异常”及“流量访问异常”的异常通信模式。

1）安全事件的发生变化：

我们根据算法所展示的异常结果信息，图 1-4 的提示信息对 uri 进行分析推测。我们对 uir 种包含 base64 及其 script 注入的结果进行筛选，得到的结果如图 2-1：

record_time	source_ip	destination_ip	method	uri
2019-06-11 21:39:03.672	10.49.21.15	10.49.223.194	POST	/index.php?z0=QGluaV9zZXQoImRpc3BsYXlfZXJyb3JzIi
2019-06-11 21:39:03.672	10.49.21.15	10.49.223.194	POST	/index.php?z0=QGluaV9zZXQoImRpc3BsYXlfZXJyb3JzIi
2019-06-11 21:39:44.267	10.49.21.15	10.49.223.194	POST	/index.php?z0=QGluaV9zZXQoImRpc3BsYXlfZXJyb3JzIi
2019-06-11 21:39:44.267	10.49.21.15	10.49.223.194	POST	/index.php?z0=QGluaV9zZXQoImRpc3BsYXlfZXJyb3JzIi
2019-06-11 21:39:13.921	10.49.21.15	10.49.223.194	POST	/index.php?z0=QGluaV9zZXQoImRpc3BsYXlfZXJyb3JzIi
2019-06-11 21:39:13.921	10.49.21.15	10.49.223.194	POST	/index.php?z0=QGluaV9zZXQoImRpc3BsYXlfZXJyb3JzIi
2019-06-11 21:39:48.269	10.49.21.15	10.49.223.194	POST	/index.php?x=%40eval%01%28base64_decode%28%24
2019-06-11 21:39:48.269	10.49.21.15	10.49.223.194	POST	/index.php?x=%40eval%01%28base64_decode%28%24
2019-06-11 21:39:43.274	10.49.21.15	10.49.223.194	POST	/index.php?z0=QGluaV9zZXQoImRpc3BsYXlfZXJyb3JzIi
2019-06-11 21:39:43.274	10.49.21.15	10.49.223.194	POST	/index.php?z0=QGluaV9zZXQoImRpc3BsYXlfZXJyb3JzIi
2019-06-11 21:39:47.370	10.49.21.15	10.49.223.194	POST	/index.php?x=%40eval%01%28base64_decode%28%24
2019-06-11 21:39:47.370	10.49.21.15	10.49.223.194	POST	/index.php?x=%40eval%01%28base64_decode%28%24
2019-06-11 21:39:07.917	10.49.21.15	10.49.223.194	POST	/index.php?x=%40eval%01%28base64_decode%28%24
2019-06-11 21:39:07.917	10.49.21.15	10.49.223.194	POST	/index.php?x=%40eval%01%28base64_decode%28%24
2019-06-11 21:39:17.907	10.49.21.15	10.49.223.194	POST	/index.php?x=%40eval%01%28base64_decode%28%24
2019-06-11 21:39:17.907	10.49.21.15	10.49.223.194	POST	/index.php?x=%40eval%01%28base64_decode%28%24
2019-06-13 10:10:37.705	10.49.212.162	10.49.223.194	POST	/index.php?action=submitlogin&returnto=%E9%A6%9E
2019-06-13 10:10:37.705	10.49.212.162	10.49.223.194	POST	/index.php?action=submitlogin&returnto=%E9%A6%9E
2019-06-13 10:16:06.701	10.49.212.162	10.49.223.194	POST	/index.php?action=submitlogin<iframe%20src%3d'data
2019-06-13 10:16:06.701	10.49.212.162	10.49.223.194	POST	/index.php?action=submitlogin<iframe%20src%3d'data
2019-06-13 14:40:14.548	10.56.144.126	10.49.253.233	POST	/index.php?action=submit&title=1<iframe%20src%3d'c
2019-06-13 10:16:01.721	10.49.212.162	10.49.223.194	POST	/index.php?action=submitlogin<iframe%20src%3d'data
2019-06-13 10:16:01.721	10.49.212.162	10.49.223.194	POST	/index.php?action=submitlogin<iframe%20src%3d'data
2019-06-13 14:47:13.278	10.56.144.126	10.49.253.233	POST	/index.php?action=submitlogin&returnto=%E9%A6%9E
2019-06-13 14:47:13.278	10.56.144.126	10.49.253.233	POST	/index.php?action=submitlogin&returnto=%E9%A6%9E
2019-06-13 10:01:12.107	10.49.212.162	10.49.223.194	POST	/index.php?action=submit&title=1<iframe%20src%3d'c
2019-06-13 10:16:07.853	10.49.212.162	10.49.223.194	POST	/index.php?action=submitlogin<iframe%20src%3d'data

图 2-1

我们可以发现，主要是在 2019-06-11 21：39 和 2019-6-13 14：47 和 10：10 分左右，从 10.49.21.15 ---> 10.49.223.194 和 10.49.212.162 --->10.49.223.194 和 10.56.144.126 ---> 10.49.253.233 为嫌疑 ip，也就是说 10.49.223.194 和 10.49.253.233 可能被控制。

因此我们对上述 ip 进行分析，在 tcpflow 表中查询嫌疑 ip，如图 2-2，2-3，2-4 我们可以发现，在 10.49.233.194，10.49.212.162IP 被劫持后，流量突变，上行流量和下行流量变得异常的大且对应 http 协议。

record_time	source_ip	destination_ip	protocol	destination_port	uplink_length	downlink_length
2019-06-11 21:38:05.303	10.49.21.15	10.49.223.194	http	80	899	757
2019-06-11 21:38:05.312	10.49.21.15	10.49.223.194	http	80	831	757
2019-06-11 21:38:05.312	10.49.21.15	10.49.223.194	http	80	870	757
2019-06-11 21:38:05.321	10.49.21.15	10.49.223.194	http	80	831	757
2019-06-11 21:38:06.195	10.49.21.15	10.49.223.194	http	80	1389	18802
2019-06-11 21:38:06.717	10.49.21.15	10.49.223.194	http	80	1389	18794
2019-06-11 21:38:09.813	10.49.21.15	10.49.223.194	http	80	2117	18827
2019-06-11 21:38:10.596	10.49.21.15	10.49.223.194	http	80	1535	18794
2019-06-11 21:38:12.864	10.49.21.15	10.49.223.194	http	80	782	18985
2019-06-11 21:38:13.331	10.49.21.15	10.49.223.194	http	80	1391	18794
2019-06-11 21:38:13.818	10.49.21.15	10.49.223.194	http	80	1395	18802
2019-06-11 21:38:14.808	10.49.21.15	10.49.223.194	http	80	1390	18802
2019-06-11 21:38:14.943	10.49.21.15	10.49.223.194	http	80	2130	18819
2019-06-11 21:38:17.706	10.49.21.15	10.49.223.194	http	80	1377	18794
2019-06-11 21:38:17.977	10.49.21.15	10.49.223.194	http	80	719	18946
2019-06-11 21:38:21.387	10.49.21.15	10.49.223.194	http	80	1395	18802
2019-06-11 21:38:22.340	10.49.21.15	10.49.223.194	http	80	721	18952
2019-06-11 21:38:23.312	10.49.21.15	10.49.223.194	http	80	1401	18802
2019-06-11 21:38:24.249	10.49.21.15	10.49.223.194	http	80	709	18934
2019-06-11 21:38:24.438	10.49.21.15	10.49.223.194	http	80	1379	18794
2019-06-11 21:38:25.031	10.49.21.15	10.49.223.194	http	80	1410	18794
2019-06-11 21:38:26.148	10.49.21.15	10.49.223.194	http	80	712	18927
2019-06-11 21:38:28.515	10.49.21.15	10.49.223.194	http	80	1471	18802
2019-06-11 21:38:33.384	10.49.21.15	10.49.223.194	http	80	779	19022
2019-06-11 21:38:35.553	10.49.21.15	10.49.223.194	http	80	111972	149359
2019-06-11 21:38:37.245	10.49.21.15	10.49.223.194	http	80	1402	18802
2019-06-11 21:38:38.343	10.49.21.15	10.49.223.194	http	80	711	18928
2019-06-11 21:38:38.613	10.49.21.15	10.49.223.194	http	80	1388	353
2019-06-11 21:38:40.728	10.49.21.15	10.49.223.194	http	80	1392	18802

图 2-2

record_time	source_ip	destination_ip	protocol	destination_port	uplink_length	downlink_length
2019-06-11 21:38:40.728	10.49.21.15	10.49.223.194	http	80	1392	18802
2019-06-11 21:38:42.113	10.49.21.15	10.49.223.194	http	80	728	18961
2019-06-11 21:39:44.259	10.49.21.15	10.49.223.194	http	80	44716	30343
2019-06-11 21:40:30.050	10.49.21.15	10.49.223.194	http	80	2147	259
2019-06-11 21:40:30.384	10.49.21.15	10.49.223.194	http	80	2134	259
2019-06-11 21:40:31.031	10.49.21.15	10.49.223.194	http	80	2142	259
2019-06-11 21:40:31.158	10.49.21.15	10.49.223.194	http	80	1981	259
2019-06-11 21:40:31.220	10.49.21.15	10.49.223.194	http	80	1983	259
2019-06-11 21:40:36.027	10.49.21.15	10.49.223.194	http	80	1916	259
2019-06-11 21:40:37.765	10.49.21.15	10.49.223.194	http	80	1996	259
2019-06-11 21:40:41.896	10.49.21.15	10.49.223.194	http	80	1884	259
2019-06-11 21:40:42.831	10.49.21.15	10.49.223.194	http	80	1937	259
2019-06-11 21:40:44.442	10.49.21.15	10.49.223.194	http	80	1966	259
2019-06-11 21:40:46.602	10.49.21.15	10.49.223.194	http	80	1898	259
2019-06-11 21:40:47.278	10.49.21.15	10.49.223.194	http	80	1884	259
2019-06-11 21:40:48.033	10.49.21.15	10.49.223.194	http	80	1916	259
2019-06-11 21:40:48.240	10.49.21.15	10.49.223.194	http	80	1889	259
2019-06-11 21:40:50.012	10.49.21.15	10.49.223.194	http	80	1905	259
2019-06-11 21:40:50.283	10.49.21.15	10.49.223.194	http	80	1866	259
2019-06-12 10:26:04.513	10.49.21.15	10.49.223.194	http	80	222	758

图 2-3

2019-06-11 21:12:32.303	10.49.212.162	10.49.223.194	http	80	456	151
2019-06-11 21:12:32.475	10.49.212.162	10.49.223.194	http	80	457	151
2019-06-11 21:12:32.691	10.49.212.162	10.49.223.194	http	80	457	151
2019-06-11 21:13:06.468	10.49.212.162	10.49.223.194	http	80	357	18742
2019-06-11 21:13:30.696	10.49.212.162	10.49.223.194	http	80	377	18758
2019-06-11 21:13:31.299	10.49.212.162	10.49.223.194	http	80	415	18792
2019-06-11 21:13:32.225	10.49.212.162	10.49.223.194	http	80	409	18786
2019-06-11 21:13:32.540	10.49.212.162	10.49.223.194	http	80	407	18784
2019-06-11 21:13:33.170	10.49.212.162	10.49.223.194	http	80	663	19008
2019-06-11 21:13:38.238	10.49.212.162	10.49.223.194	http	80	750	19105
2019-06-11 21:13:41.199	10.49.212.162	10.49.223.194	http	80	392	18771
2019-06-11 21:13:41.801	10.49.212.162	10.49.223.194	http	80	395	18774
2019-06-11 21:13:42.378	10.49.212.162	10.49.223.194	http	80	395	18774
2019-06-11 21:13:44.402	10.49.212.162	10.49.223.194	http	80	398	18777
2019-06-11 21:13:51.747	10.49.212.162	10.49.223.194	http	80	380	18759
2019-06-11 21:13:58.209	10.49.212.162	10.49.223.194	http	80	453	18830
2019-06-11 21:13:58.498	10.49.212.162	10.49.223.194	http	80	383	18762

图 2-4

我们对上述找到的 ip 进行筛选，如图 2-5，确定了可以连接的 mysql 服务器的机器 ip

record_time	source_ip	destination_ip	protocol	destination_port	sql_info
2019-06-11 21:20:39.418	10.49.223.194	10.49.253.35	mysql	3306	SET NAMES gbkr #4294967295
2019-06-11 21:20:39.418	10.49.223.194	10.49.253.35	mysql	3306	SHOW DATABASES #4294967295
2019-06-11 21:21:06.554	10.49.223.194	10.49.253.35	mysql	3306	SET NAMES gbkr #4294967295
2019-06-11 21:21:06.555	10.49.223.194	10.49.253.35	mysql	3306	SHOW TABLE STATUS294967295
2019-06-11 21:21:06.555	10.49.223.194	10.49.253.35	mysql	3306	SHOW DATABASES #4294967295
2019-06-11 21:21:10.944	10.49.223.194	10.49.253.35	mysql	3306	SET NAMES gbkr #4294967295
2019-06-11 21:21:10.944	10.49.223.194	10.49.253.35	mysql	3306	SHOW DATABASES #4294967295
2019-06-11 21:21:18.186	10.49.223.194	10.49.253.35	mysql	3306	SET NAMES gbkr #4294967295
2019-06-11 21:21:18.187	10.49.223.194	10.49.253.35	mysql	3306	SHOW DATABASES #4294967295
2019-06-11 21:21:18.188	10.49.223.194	10.49.253.35	mysql	3306	SHOW TABLE STATUS294967295
2019-06-11 21:21:26.404	10.49.223.194	10.49.253.35	mysql	3306	SET NAMES gbkr #4294967295
2019-06-11 21:21:26.405	10.49.223.194	10.49.253.35	mysql	3306	SHOW TABLE STATUS294967295
2019-06-11 21:21:26.405	10.49.223.194	10.49.253.35	mysql	3306	SHOW DATABASES #4294967295
2019-06-11 21:21:35.070	10.49.223.194	10.49.253.35	mysql	3306	SET NAMES gbkr #4294967295

图 2-5

同时，我们使用可以端口对 tcpflow 表进行查询，并结合图 2-6

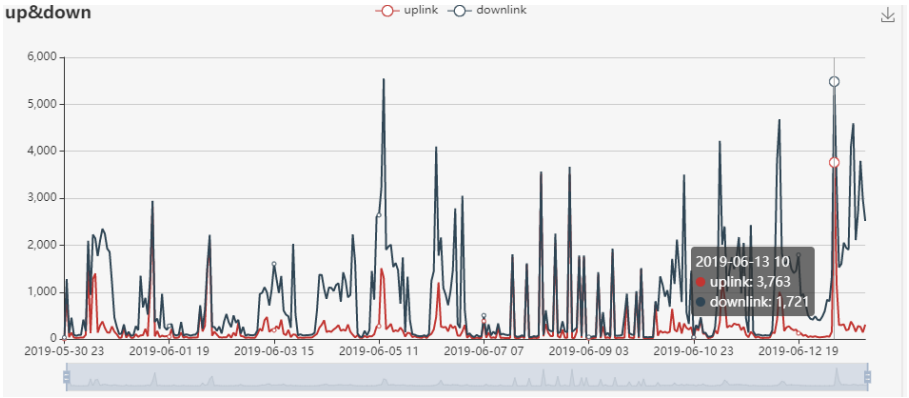

图 2-6

找出了大流量的可疑服务器，他们是 10.30.150.54，10.49.223.194，10.49.137.87，10.49.223.163.

总结上述可以 ip 机器及其访问量，我们做如图 2-7 结论：

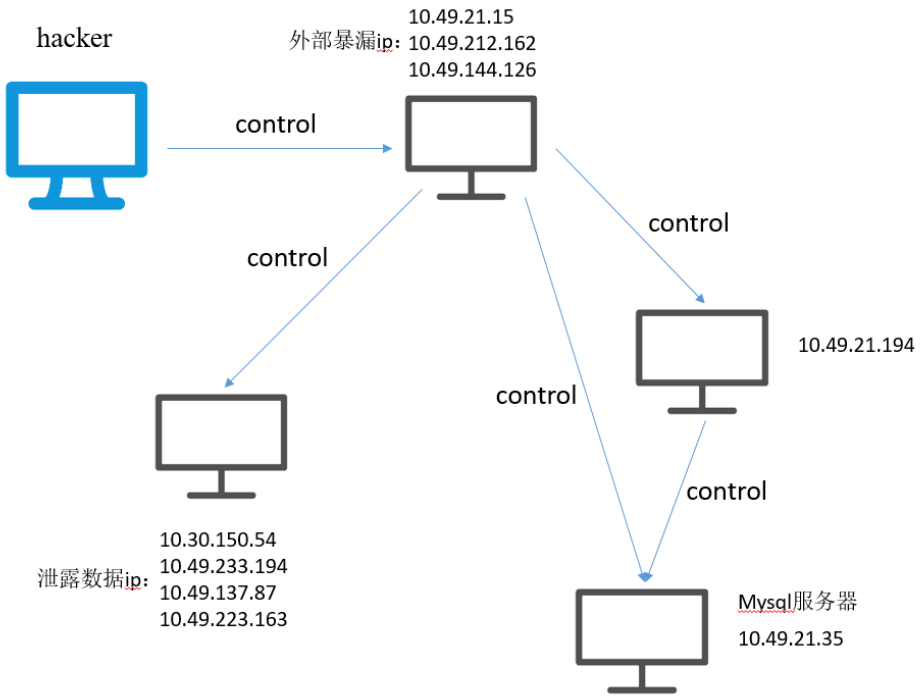

图 2-7