面向网络安全事件的入侵追踪与取证系统

郑飞飞，龚俭

（东南大学网络空间安全学院，南京，211189）

摘　要：为了提高安全事件应急响应的效率，设计并实现了一个入侵追踪与取证系统。该系统基于特定的安全事件信息，使用 OpenFlow 交换机实现报文的过滤和转发，利用 PF_RING ZC 零拷贝驱动工具采集报文，使用高效的报文分离算法离线分离报文，使用开源入侵检测软件 Suricata 完成对报文流量的入侵检测，使用开源系统 Bro 进行应用层协议分析，对警报和协议日志进行语义提取，同时结合追踪内容可视化，提升安全事件的应急响应效率。

关键词：安全事件；入侵追踪；语义提取；应急响应

Intrusion Tracking and Forensics System for Network Security Incidents

Zheng Feifei, Gong Jian

(School of Cyber Science and Engineering, Southeast University, Nanjing,211189)

Abstract: In order to improve the efficiency of emergency response to security incidents, an intrusion tracking and forensics system was designed and implemented. Based on the specific security incidents information, the system uses OpenFlow switches to filter and forward packets. The PF_RING ZC zero-copy driver is used to collect packets. The high-efficiency packet classification algorithm is used to separate packets offline. The open source intrusion detection software Suricata is used to detect packet traffic, using the open source system Bro for application layer protocol analysis, semantic extraction of alarms and protocol logs, and combined with tracking content visualization to improve the emergency response efficiency of security incidents.

Key words: Security incidents; Intrusion traceback; Data extraction; Semantic extraction; Emergency response

“十一五”期间 211 工程在 CERNET（China Education and Research Network）网络安全中心和 38 个核心节点上建设高性能网络管理与安全保障系统[1]，功能包括网络流量实时监控、网络安全异常检测和网络安全事件应急响应等。CHAIRS（Cooperative Hybrid Aided Incidence Response System）系统[2]是该项目中的一个大型分布式应急响应服务管理系统，主要是为 CERNET 网络中心以及各节点的安全管理人员提供应急响应辅助功能，帮助安全人员快速、便捷有效地处理事件，提高事件响应的效率。对于 CHAIRS 系统提供的安全警报，需要各节点或校园网的安全管理员进行应急响应，发现并解决网络中存在的安全问题或安全隐患[3]。

为了帮助安全分析员对 CHAIRS 系统提供的安全警报进行应急响应，我们设计并实现了 MONSTER（Monitor On Network Security and Tool for Emergency Response）系统，功能包括网络报文采集及过滤、入侵检测及协同响应等。当安全分析员对威胁源（经过威胁评估后的安全事件集合）进行调查分析时，CHAIRS 系统将产生待追踪的特定对象信息（以 IP 及相关特征来表示），并以对象信息自动生成追踪任务（包含待追踪的特定对象信息及相应控制字段等）发送给 MONSTER 系统进行追踪取证。MONSTER 系统需要对追踪任务中包含的特定对象信息进行通信活动的追踪与取证[4]，以此来获取足够的证据信息，达到给威胁源定性的目的。

随着 CERNET 主干网运行规律和安全保障内容的扩展，产生了新的应用要求和应急响应要求，使得现有的 MONSTER 系统不能满足这些新出现的需求。此外 MONSTER 系统在实际运行过程中也暴露出了一些缺陷，需要进一步改进完善。目前 MONSTER 系统主要有以下需求：第一，响应效率

作者简介：郑飞飞，（1992-），男，硕士研究生，E-mail: ffzheng@njnet.edu.cn；龚俭，（1957-），男，教授，博导，E-mail:jgong@njnet.edu.cn.

提升的需求。目前系统最多仅支持 32 个威胁源进行并行追踪取证，随着大量的安全事件被检出，这将严重影响安全保障系统的效率，因此需要提升威胁源的响应效率；第二，响应功能完善的需求。目前系统只将原始报文和报文检测产生的警报信息作为取证结果返回给 CHAIRS 系统，无法提供有价值的证据信息，因此需要完善系统的响应功能。

基于上述研究背景和系统现状，本文对现有的 MONSTER 系统进行了改进，使其满足安全保障系统的新需求。改进包括：基于特定的安全事件信息，使用 OpenFlow[5]交换机实现报文的过滤和转发，同时采用高效地报文捕获工具 PF_RING ZC[6]自动采集报文；将报文采集和报文分离松耦合，解决多追踪任务并发追踪取证；使用开源入侵检测软件 Suricata[7]完成对报文的入侵检测，使用开源系统 Bro[8]进行应用层协议分析，并对检测结果进行语义提取，形成证据力强的证据信息返回给 CHAIRS 系统进行应急协同响应。

1 入侵追踪与取证系统架构设计

根据安全保障系统的实际功能需求，入侵追踪与取证系统结构图如图 1 所示。其中，HYDRA（Hybrid Detection Response Agent）为基于 SDN 技术的入侵阻隔系统[9]，该系统经过 OpenFlow 交换机控制网络报文的转发，可以保证在网络正常运行的同时，实现对恶意流量的阻断、对攻击流量的样本采集。

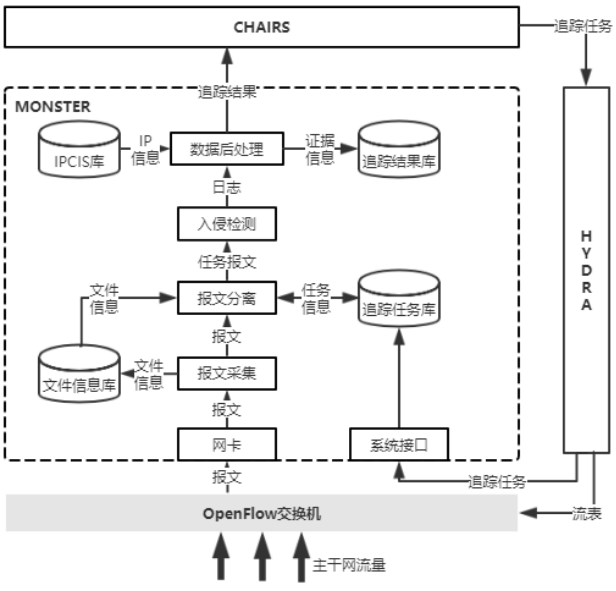

图 1 入侵追踪与取证系统结构图

HYDRA 系统接收到 CHAIRS 系统的追踪任务后，首先将追踪任务转发给 MONSTER 系统进行追踪取证，接着自动解析该追踪任务信息，生成相应的流表规则并将流表规则集合下发至 OpenFlow 交换机进行交换机流表项配置，交换机根据流表项规则过滤流量，并将匹配流表项的流量转发给 MONSTER 系统的网卡。

MONSTER 系统接口模块接收 HYDRA 系统发送过来的待追踪取证的追踪任务，对其进行生命周期管理，并将任务信息持久化在数据库中。

MONSTER 系统报文采集模块自动采集 OpenFlow 交换机转发给网卡的流量，并周期性地将报文输出并保存到磁盘 PCAP 格式文件中，同时将 PCAP 文件信息保存在文件信息库，该周期 PCAP 文件是当前采集周期内所有追踪任务的报文集合。

MONSTER 系统报文分离模块定时轮询文件信息库，查看是否有未被分离的周期 PCAP 文件，若有，则根据文件索引信息从磁盘读取文件，并根据各追踪任务的追踪需求，将周期 PCAP 文件中的所有报文依次分离给具体的追踪任务。报文分离结束后更新追踪任务追踪信息。

入侵检测模块使用开源检测软件 Suricata 检测任务报文，生成警报日志文件，同时使用开源流量分析器 Bro 分析任务报文，生成活动信息日志文件以及警报日志文件。

数据后处理模块对警报日志文件和活动信息日志文件进行语义提取，同时通过从 IPCIS（IP Comprehensive Information System，IP 综合信息系统）系统[10]富化 IP 相关信息，提取出有价值的证据信息，进而形成追踪结果回送给 CHAIRS 系统进行应急响应协同。

此外，威胁源的入侵追踪与取证是一个持续性的过程，当 MONSTER 系统提供的证据信息足以对威胁源定性时，则此威胁源的追踪取证工作完成，否则将一直持续被追踪与取证。

2 报文采集分离模块的设计与实现

当 MONSTER 系统接收到 OpenFlow 交换机转发来的流量时，需要考虑如下两个问题。一是要考虑如何高效地将网卡中的流量采集到本地磁盘中；二是要考虑追踪任务的并发问题，即如何满足多个追踪任务并发采集。根据威胁源追踪响应的实际功

能需求，本文分析并设计了一个基于生产者-消费者的报文采集分离模型，模型如下图 2 所示。

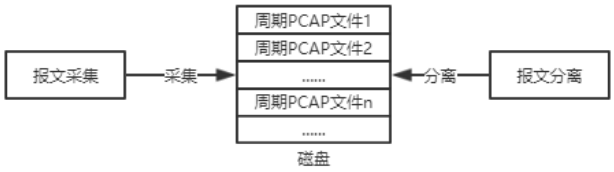

图 2 基于生产者-消费者的报文采集分离模型

该模型基于生产者-消费者模式，将报文采集模块和报文分离模块松耦合，使得追踪任务并发个数不受限制，同时方便系统进行后期维护。报文采集模块作为生产者持续采集 HYDRA 系统转发到网卡的流量，并周期性地将流量输出到磁盘，以 PCAP 文件格式保存，同时将该 PCAP 文件信息保存在数据库中。报文分离模块定时轮询文件信息数据库，如果数据库中有未被处理的周期 PCAP 文件索引信息，报文分离模块就会按照 FCFS（First Come First Served，先来先服务）的调度算法依次对磁盘中相应的周期 PCAP 文件进行离线分类，将每个周期 PCAP 文件中的报文依次分离到具体的追踪任务中。

2.1 报文采集

根据安全保障系统的采集需求，本文使用高效数据包捕获工具 PF_RING ZC 捕获网卡流量。PF_RING ZC 是实现了 PF_RING DNA（Direct NIC Access）技术，是一种映射网卡内存和寄存器到用户态的方法，因此除了由网卡的网络处理单元完成 DMA 传输外，没有任何额外的数据包复制，进一步节省了一次数据拷贝操作，报文捕获性能更好，因此报文采集模块使用 PF_RING ZC 网卡驱动工具，高效持续地采集 OpenFlow 交换机转发到网卡的流量。

为了让 PF_RING ZC 捕获的网卡流量尽快进行后续处理分析工作，本文以 5 分钟为时间窗口大小周期地将报文输出到磁盘并保存在 PCAP 格式文件中，PCAP 文件以当前周期时间命名。 同时，将该 PCAP 文件的文件索引信息保存在数据库中，以便进行后续的报文分离等流程。接着，报文采集模块继续进行周期性报文采集工作。

2.2 报文分离

由于报文采集模块只对网卡中的流量进行了周期采集并存储，所以周期 PCAP 文件中的报文是当前系统中所有正在追踪取证的追踪任务的报文集合。为了方便后续的分析处理工作，报文分离模块需要以追踪任务为单位，将周期 PCAP 文件中的报文依次离线分离给具体的追踪任务，得到各追踪任务的通信流量。报文分离模块定时轮询文件信息数据库，如果数据库中有未被处理的周期 PCAP 文件索引信息，报文分离模块就会按照 FCFS 的调度算法依次对磁盘中相应的周期 PCAP 文件进行离线分离。

报文分类的传统做法是对于每个数据包，依次遍历所有追踪任务，同时验证每个追踪任务的源 IP、宿 IP、源端口、宿端口等追踪条件是否匹配该数据包。这种分类方法简单并且易于实现，但是当追踪任务数目较大时这种线性分类方法速度较慢，导致磁盘中大量周期 PCAP 文件未能被及时处理而积压，严重降低了系统的可用性。

为了提高报文离线分离速度，本文先对所有的追踪任务进行数据预处理，将每个追踪任务中以点分十进制形式表达的 IP 地址都转换成十进制整数，并将该整数 IP 地址和追踪任务的映射关系维护在哈希表中。追踪任务数据预处理完成后，进入报文离线分类阶段。报文离线分离模块分为两层，如下图 3 所示。

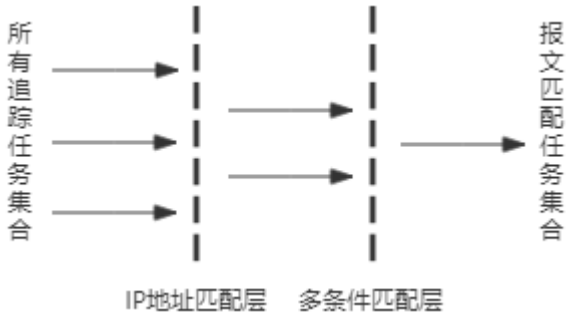

图 3 两层报文分离模型

第一层是 IP 地址匹配层，该层根据报文 IP 地址对所有追踪任务进行初步过滤。将报文中的点分十进制源、宿 IP 地址转换为十进制整数，根据源、宿整数地址分别从哈希表中取出追踪任务集合，两集合的并集即为 IP 地址匹配该报文的所有追踪任务。IP 地址匹配层通过初次过滤，已经将与报文 IP 地址相匹配的追踪任务从大量的追踪任务中过滤出来。第二层是多条件匹配层，该层根据追踪任务中

的源端口、宿端口、协议号等其他条件进行过滤，是一个更精确化的过滤层。经过两次过滤后的追踪任务集合即为与当前报文相匹配的追踪任务集合。

3 数据后处理模块的设计与实现

MONSTER 系统报文采集分离模块已经根据追踪任务的追踪条件，采集到追踪任务的网络流量并存储在磁盘中，但仅仅将采集到的任务报文返回给 CHAIRS 系统以供安全调查人员进行威胁源定性是不切合实际的。就如何从网络流量中提取出更具有语义的数据问题，本文分析并设计了一个追踪数据语义提取模型，模型如下图 4 所示。

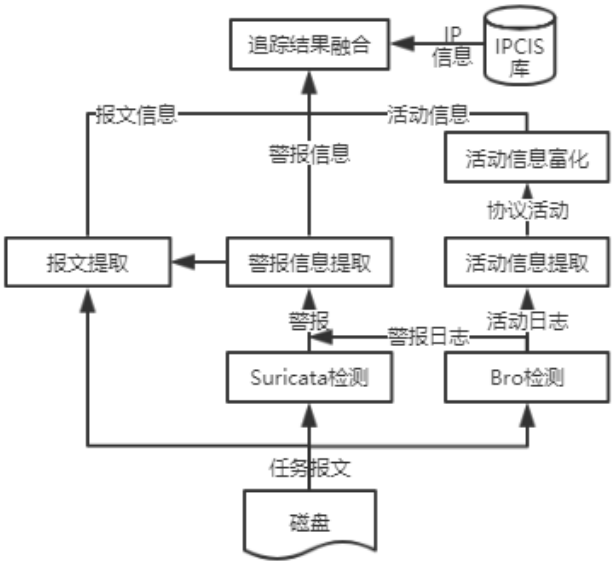

图 4 追踪数据语义提取模型

本文首先使用开源入侵检测软件 Suricata 检测任务报文产生警报日志文件，同时使用开源流量分析系统 Bro 分析任务报文产生警报日志文件和网络活动日志文件。接着，警报信息提取模块自动提取 Suricata 警报日志和 Bro 警报日志，并根据警报日志的 IP 地址及端口等信息从任务报文中提取相关报文形成警报信息；同时活动信息提取模块提取 Bro 活动日志形成活动信息，活动信息提取完成后进行数据富化；最后，通过融合警报信息、活动信息以及相关报文，形成追踪结果返回给 CHAIRS 系统。

3.1 报文检测

安全人员很难通过直接分析原始报文判断攻击意图，需要从原始报文中抽取有价值的语义信息，本文使用开源软件 Suricata 和开源流量分析系统 Bro 对报文进行检测。其中，Suricata 是一款高性能网络入侵检测、网络防御和网络安全监控引擎，根据配置的规则集，通过对网络运行状况进行监控，尽可能发现各种攻击企图、攻击行为或攻击结果；Bro 是一款被动的开源流量分析器，既可以用于对链路上的所有深层深层次的可疑行为进行安全监控，还可用于收集网络测量数据、进行网络取证调查等。当接收到报文采集分类模块的通知信息后，Suricata 自动对存储在磁盘中的任务报文进行基于规则的离线检测，将检测出的报警信息输出到警报日志文件 eve.json 并存储于系统磁盘中。同时，Bro 自动进行离线检测，将检测出的报警信息输出到警报日志文件 weird.log 中，将识别出的应用层协议活动信息输出到活动日志文件中，并存储于系统磁盘中。

3.2 警报信息语义提取

报文检测产生的警报日志文件异构并且难以分析，需要对警报信息进行语义提取。警报信息语义提取模块提取两部分内容，一是提取警报信息，该内容从 Suricata 检测产生的警报日志文件 eve.json 和 Bro 检测产生的警报日志文件 weird.log 中提取；二是提取警报关联的报文信息，该部分从任务报文中提取。警报信息提取模块自动调用脚本对 Suricata 警报日志文件 eve.json 和 Bro 警报日志文件 weird.log 进行处理。首先将两类异构警报的格式做统一化处理，统一化后的警报信息格式如表 1 所示，接着提取其中的时间戳、signature 和四元组相关信息，最后跟据警报的时间戳和四元组等相关信息，从任务报文中提取警报关联的报文信息。

表 1 警报信息格式

字段	字段描述
ts	时间戳
src_ip	源地址
src_port	源端口
dst_ip	目的地址
dst_port	目的端口
signature	警报特征值
sensor	检测出该警报的 IDS 名称

3.3 活动信息语义提取

开源流量分析系统 Bro 支持多种应用层协议活动的识别和解析，例如域名解析协议 DNS、超文本传输协议 HTTP 等。任务报文经过 Bro 检测后产生活动日志文件并存储于磁盘文件中。由于文件不方便后续的融合分析和数据可视化，且由于 Bro 系统自身的局限性和网络的复杂性导致部分通信活动未被识别，活动信息语义提取模块主要完成以下 3 个工作：（1）协议活动特征值提取。根据应用层协议活动的类型，在磁盘活动日志文件中提取不同的特征值，并将提取的内容持久化在数据库中。（2）活动信息富化。由于 Bro 系统自身的局限性，导致部分网络活动无法被识别，所以需要根据服务端口自动富化出未被识别的网络活动类型。（3）相关报文提取。对于 Bro 无法识别且无法自动富化的通信活动，需要从任务报文中提取出相关报文，以供安全分析员进行人工分析。

3.4 追踪结果融合

警报信息语义提取模块和活动信息语义提取模块已经从任务报文中提取出追踪任务在一段时间内的通信活动相关信息。但这些追踪数据互相独立且分布在系统中的不同位置，追踪结果融合模块需要融合这些追踪数据形成追踪结果，使其能够反映出追踪任务通信过程的全貌。同时，为了利于后续分析，需要从 IPCIS 库中富化出追踪任务中追踪 IP 的相关信息和对端 IP 的相关信息。融合后的追踪结果将被返回给 CHAIRS 系统进行威胁源定性分析。

4 实验与分析

4.1 实验过程

为了测试入侵追踪与取证的有效性，安全事件调查人员从 CHAIRS 系统中获取威胁源进行调查。CHAIRS 系统将根据威胁源的相关信息配置追踪任务，并将其自动发送给 HYDRA 系统进行流量的过滤转发。MONSTER 系统将在此流量信息的基础上自动进行报文采集、报文分离、报文检测、数据后处理等相关处理，以期获得追踪任务的警报信息和网络活动信息等证据信息。

如图 5 所示，MONSTER 系统对 HYDRA 系统转发过来的流量进行自动采集，并以 5 分钟为时间窗口大小周期性地将报文存储在磁盘中，文件存储格式为:年月日_时分秒。

```
-rw-r--r-- 1 root root  4867665 Jan 24 15:20 20190124_151500.pcap
-rw-r--r-- 1 root root  6932056 Jan 24 15:25 20190124_152000.pcap
-rw-r--r-- 1 root root  4276732 Jan 24 15:30 20190124_152500.pcap
-rw-r--r-- 1 root root  4595666 Jan 24 15:35 20190124_153000.pcap
-rw-r--r-- 1 root root 12431081 Jan 24 15:40 20190124_153500.pcap
```

图 5　报文自动采集结果

当磁盘中有未被处理的周期报文时，报文分离模块根据分离算法自动地将周期 PCAP 文件分离到具体的追踪任务中，每个追踪任务在不同周期内采集到的报文将存储在以时间周期命名的目录中，并以 cycle 统一命名。

报文分离结束后将自动进入报文检测模块，报文检测模块分别调用 Suricata 检测软件和 Bro 检测软件对任务报文进行离线检测。如图 6 所示，Suricata 检测软件生成警报日志文件 eve.json。Bro 检测软件生成警报日志文件 weird.log 及活动信息日志文件。

```
conn.log  eve.json  http.log  ssl.log  weird.log
```

图 6 报文检测结果

报文经过离线检测后，数据提取模块将自动提取警报语义信息和协议活动语义信息，并将提取的数据存储在数据库中。追踪数据信息如表 2 所示。

表 2 追踪数据信息

开始时间	持续时间	源 IP	源端口	宿 IP	宿端口	协议	活动
2019-01-24 15:21:29.841	75.515s	101. ***. ***. 13	37546	118. ***. ***. 84	8106	tcp	http
2019-01-24 15:21:42.206	31.919s	219. ***. ***. 254	57928	118. ***. ***. 84	443	tcp	ssl
2019-01-24 15:22:54.156	30.921s	219. ***. ***. 254	57943	118. ***. ***. 84	443	tcp	ssl
2019-01-24 15:23:32.111	30.931s	219. ***. ***. 254	57949	118. ***. ***. 84	443	tcp	ssl
2019-01-24 15:24:23.393	0.733s	61. ***. ***. 136	24135	118. ***. ***. 84	443	tcp	ssl
2019-01-24 15:24:38.266	1.366s	219. ***. ***. 254	57962	118. ***. ***. 84	443	tcp	ssl
2019-01-24 15:24:54.383	0.053s	61. ***. ***. 136	24135	118. ***. ***. 84	443	tcp	ssl

数据提取模块完成后，追踪结果融合模块将对数据提取出的追踪数据进行数据融合形成追踪结果，使其能够反映追踪任务的通信活动全貌。融合后的追踪结果如图 7 所示，在当前周期内，该追踪任务有追踪数据的追踪 IP 为 118.***.***.84，且该追踪 IP 共与 3 个不同的对端 IP 进行过通信，涉及的通信活动有 HTTP 和 SSL 两类。其中，追踪 IP 与对端 101.***.***.13 进行过一次 HTTP 通信；与对端 219.***.***.254 进行过 5 次 SSL 通信；与对端 61.***.***.136 进行过 1 次 SSL 通信。各通信活动的活动语义信息如图 7 所示。此外，任务报文通过开源软件 Suricata 和 Bro 检测没有生成警报信息。

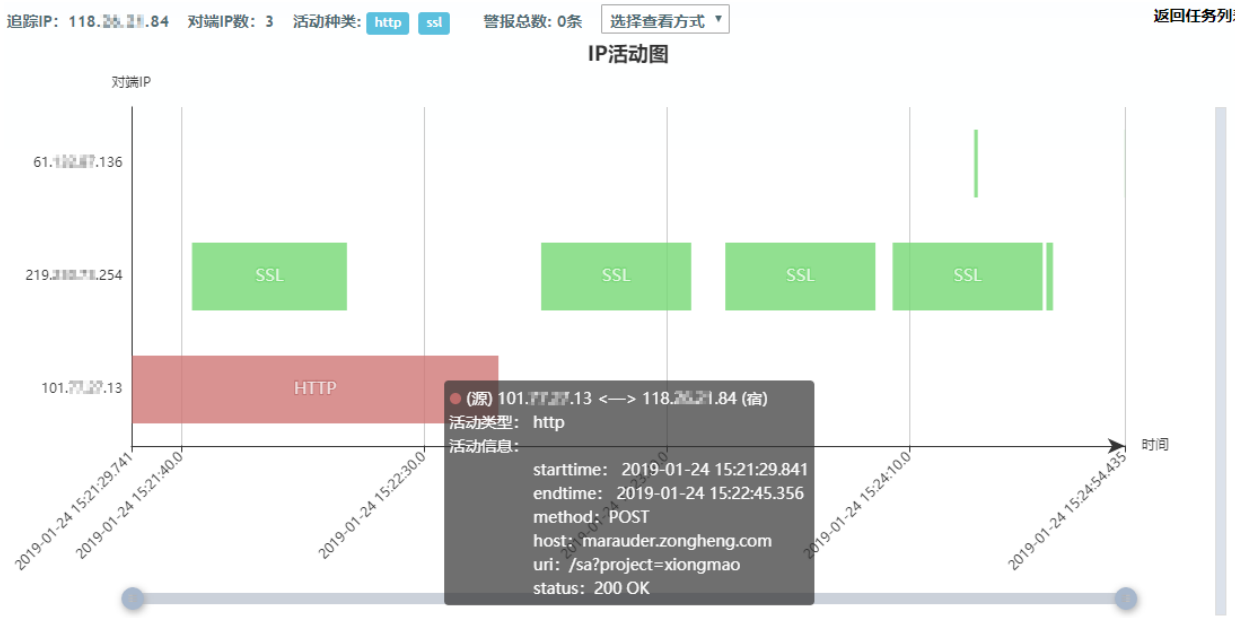

图 7 追踪结果

4.2 实验结果分析

为了提升安全事件的应急响应效率，本研究设计并实现了一个网络入侵追踪与取证系统。从实验结果来看，报文采集分离模块以 5 分钟为时间窗口大小周期性地将流量保存并存储在磁盘文件中，报文分离模块自动地对周期流量进行离线分离，从实验结果上看解决了多追踪任务并发追踪取证个数问题，提高了安全保障系统的应急响应效率。同时，入侵检测模块对追踪任务的原始报文进行入侵检测和协议分析，得到了追踪任务通信活动的警报信息和活动信息。数据后处理模块自动地将原始报文中的警报信息和活动信息提取出来，并通过数据融合形成追踪结果返回给 CHAIRS 系统以供安全调查人员调查分析，相比较原始报文信息，该追踪结果能使安全人员更加直观地了解网络入侵活动信息。整个追踪取证过程全部实现自动化，这在很大程度上提升了安全事件的应急响应效率。

参考文献

[1] 朱礼智, 龚俭. 分布式网络应急响应管理系统 CHAIRS 的设计与实现[D]. 南京: 东南大学计算机科学与工程学院, 2015.

[2] 吕少阳. CHAIRS 系统运行管理与离线检测的设计与实现[D]. 南京: 东南大学计算机科学与工程学院, 2013.

[3] 龚俭, 吴桦, 杨望. 计算机网络安全导论(第二版)[M]. 东南大学出版社, 2007.

[4] 杨泽明, 许榕生, 曹爱娟. 网络取证与分析系统的设计与实现[J],计算机工程, 2004, 30(13):72-74.

[5] ONF. OpenFlow Switch Specification[EB/OL]. [2016-06-10]. http://ONF.org.

[6] Ntop. PF_RING ZC(zero copy)[EB/OL]. [2016-06-10]. https://www.ntop.org/products/packet-capture/pf_ring/pf_ring-zc-zero-copy/.

[7] Suricata. Suricata open source IDS/IPS/NSM engine[EB/OL]. [2016-06-10]. https://suricata.ids.org/

[8] Nick Buraglio. Overview of the Bro intrusion detection syst

em[EB/OL]. [2016-06-10]. https://fasterdata.es.net/assets/20150522-Buraglio-Bro.pptx.

[9] 金磊, 龚俭. 基于 SDN 技术的网络入侵阻断系统 HYDRA 的设计与实现[D]. 南京: 东南大学计算机科学与工程学院, 2016.

[10] 李亚明. IPCIS 系统中的 IPv4 地址使用位置库的改进研究[D]. 东南大学, 2015.