多维度视角下APT挖掘实践

闫忠
深信服蓝军高级威胁研究组

2021 SDC分会场-公开课

APT 的定义

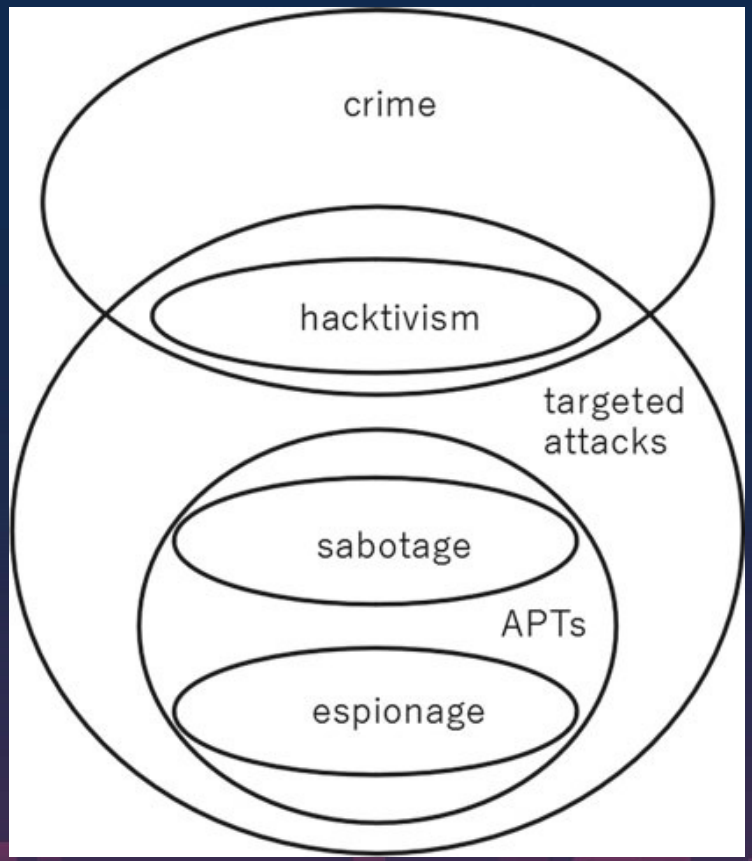
“高级持续性威胁（APT）”这个术语最初是由美国军方引入的，并很快在网络安全领域中被采用。

这个术语的日常使用导致了一定的歧义，随着时间的推移意思也在不断的变化。首先，2006年美国空军发明了该术语，主要的目的是在与外部人员进行交流时，避免透露关于攻击者来源的调查结果。同时还得向外部人员表达出这种攻击不是普通攻击，所以提出了高级持续性威胁的概念。

高级这个词表示与普通的互联网上不断进行扫描的攻击行为不同，意味着复杂；
持续性意味着攻击者会故意选择目标，并且长时间反复针对目标；
威胁指的是攻击者的属性与动机；

APT ，全称Advanced Persistent Threat，不是一种技术或某种类别的恶意软件，而是具有战略动机的参与者，往往有地区或政治背景，以情报搜集、破坏、或经济利益为目的，攻击环节可能使用各类社工、打点和内网渗透以及0day漏洞利用，作为一种非对称的攻击手段，往往能为攻击组织背后的政治或经济实体带来意想不到的地缘、情报、经济甚至军事利益或战术优势。

APT 与网络犯罪的交集



crime 犯罪

hacktivism 黑客主义

targeted attacks 针对性攻击

sabotage 破坏

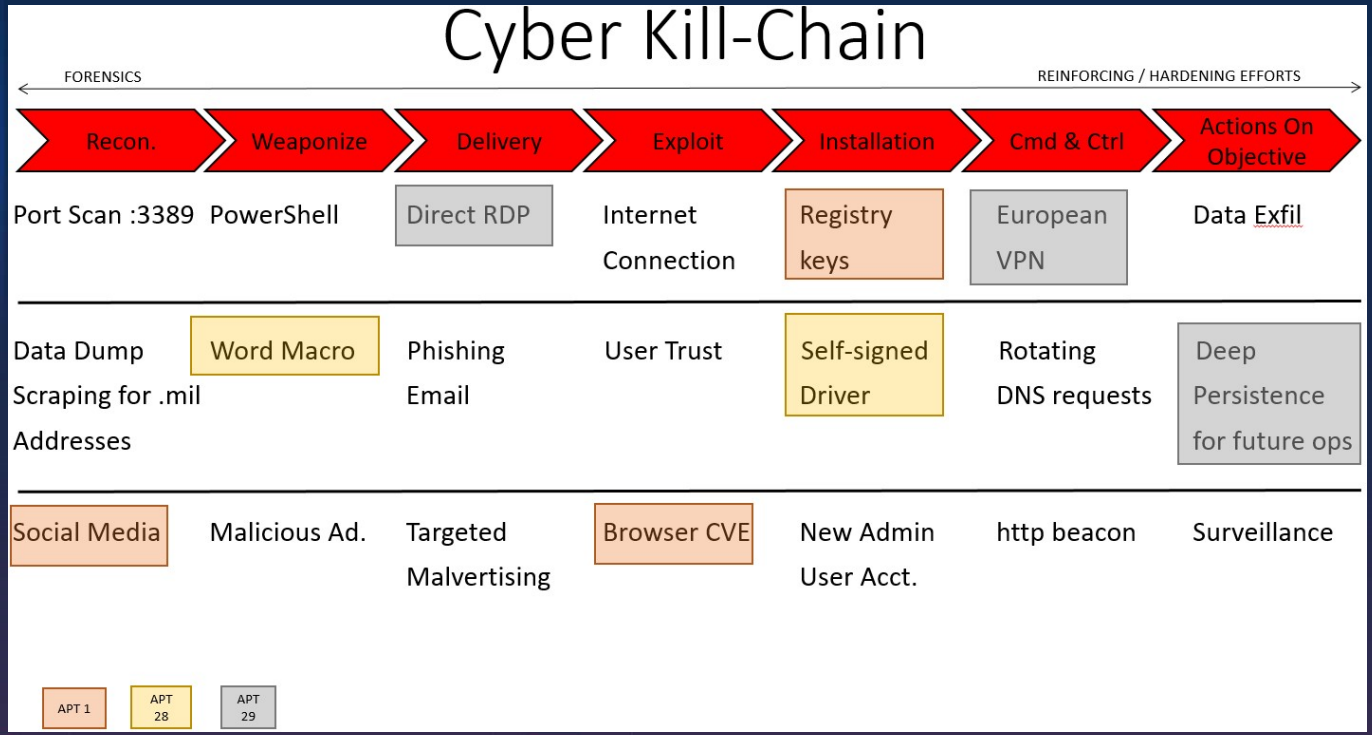
espionage 间谍

source:
Timo Steffens, Attribution of Advanced Persistance Threat

APT 的攻击阶段

杀伤链（Kill Chain）是 APT 攻击典型阶段的理想模型，杀伤链描述了攻击者通常经历的阶段，并不一定是顺序的，只是一个理想情况下的模型。

source：
MAJ Joe Marty,Cyber Threat Heat-Mapping



APT 的攻击目标

高级持续性威胁是一个不仅仅出于单纯利益动机，而是试图获取可以在战略或政治上使用的数据的团体。

APT 组织最有针对性的目标：

- 政府机构
- 能源公司
- 军事组织
- 媒体
- 国防承办商

主流思路

基于已有的 APT 组织（团伙）挖掘

海莲花（OceanLotus）、蔓灵花（BITTER）、Lazarus、肚脑虫（Donot）、响尾蛇（SideWinder） 等。

基于未知的 APT 组织（团伙）挖掘

由点到面，然而超高能力威胁组织由于公开的情报较少，挖掘难度大，已公开案例如震网（Stuxnet） 蠕虫、Duqu 病毒、火焰（Flame） 蠕虫等。

Malware Hunting

在之前列举的 APT 整个攻击链中，存留时间最长的数字类型证据是恶意文件，恶意文件在整个网空安全领域里存留时间最长，且因外因变化而改变的概率小，不易被篡改。同时，这也是我们与攻击者之间关联度最高的一种媒介，可以说恶意文件在 APT 挖掘过程中所占的比重最大。

我们如何挖掘到所属攻击者的恶意文件?

1、被动事件响应

2、主动事件挖掘

3、私有数据源挖掘

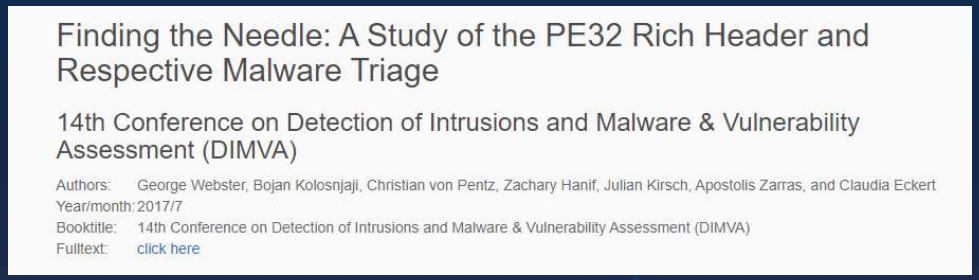
4、公共数据源挖掘

PE Metadata Hunting

充分利用 PE 文件元数据来追踪攻击者的更多样本，主要聚焦于以样本为核心进行研究扩展，目标是找到更多攻击者的恶意文件与 IOC 情报。

PE 文件元数据的多个维度：

- 文件名
- imphash 值
- Rich header 哈希值
- 数字证书
- 模糊哈希（SSDEEP、TLSH、VHASH）


Finding the Needle: A Study of the PE32 Rich Header and Respective Malware Triage

14th Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA)

Authors: George Webster, Bojan Kolosnjaji, Christian von Pentz, Zachary Hanif, Julian Kirsch, Apostolis Zarras, and Claudia Eckert
Year/month: 2017/7
Booktitle: 14th Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA)
Fulltext: click here


THREAT RESEARCH

Tracking Malware with Import Hashing

MANDIANT

JAN 24, 2014 | 10 MINS READ

背后的攻击者会犯错误，他们后续会改进。

必须与其他来源（基础设施、TTP、地缘政治）结合分析，以避免出现误报。

实践案例

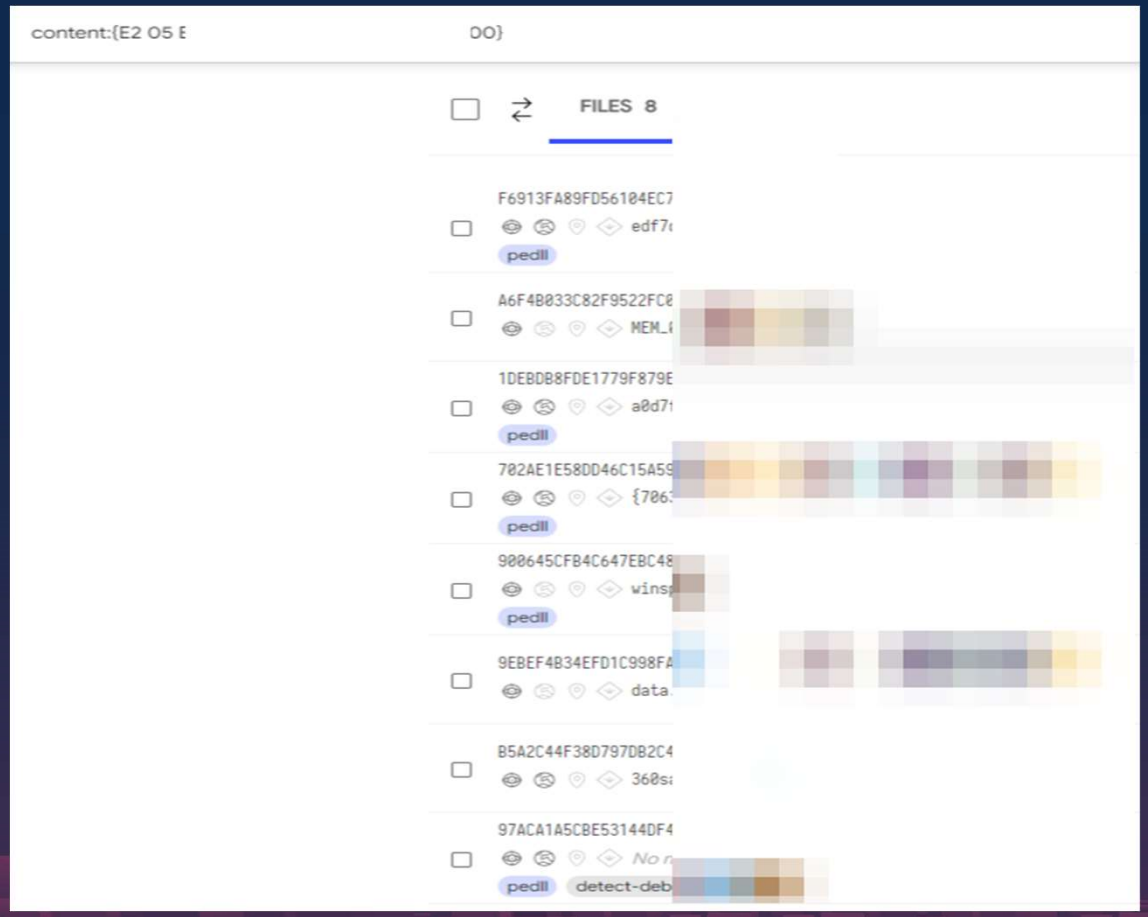
接下来，以东南亚地区活跃的 APT 组织海莲花为例，分享我们的部分技术分析思路。

OceanLotus（海莲花）又名 APT32、SeaLotus 等称号，是一个据称有东南亚背景的 APT 组织，OceanLotus 自首次披露以来持续活跃至今，其攻击所涉及的国家地区分布非常广泛，包括某东南亚国家周边国家（如柬埔寨、泰国、老挝等）和欧洲地区，该组织的攻击目标还包括其国内的异见人士、人权代表、媒体、环保组织等。

恶意代码同源性

挖掘特定攻击者的恶意文件并不是一件容易的事，我们以下实践的一个重要的前提：

代码相似性不会偶然发生，而是更有可能是由于同一位开发人员编写了代码，这个假设从逻辑推理上是合理的。



误导归因结果

代码相似性的归因判定难免出现误报

左侧归因的结果存在错误，非最早的代码来源出处，例如右侧在2010年的文章里出现的代码就已经存在很大的相似性，其中包括变量名与相关执行的命令。因此分析人员需要格外警惕，不能排除攻击者故意模仿其他攻击者引导溯源人员将攻击归因到错误的来源。

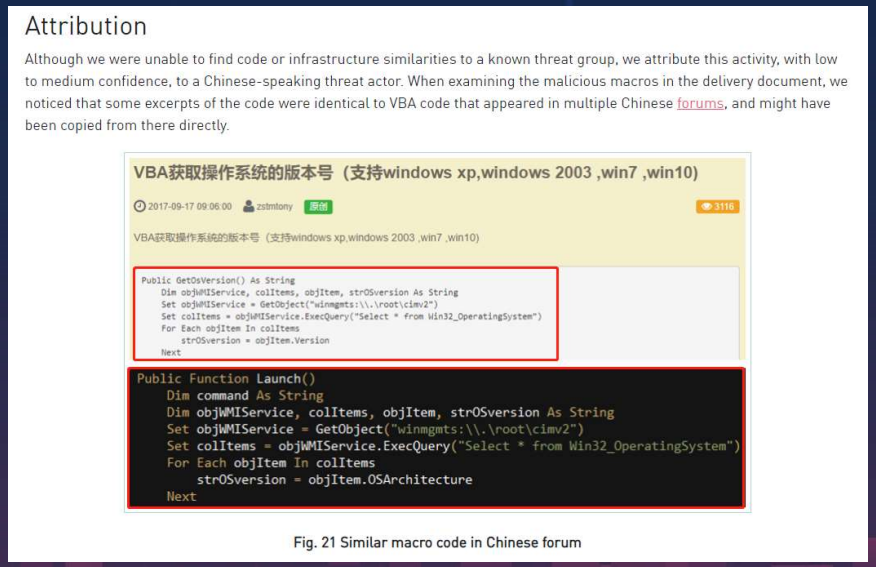
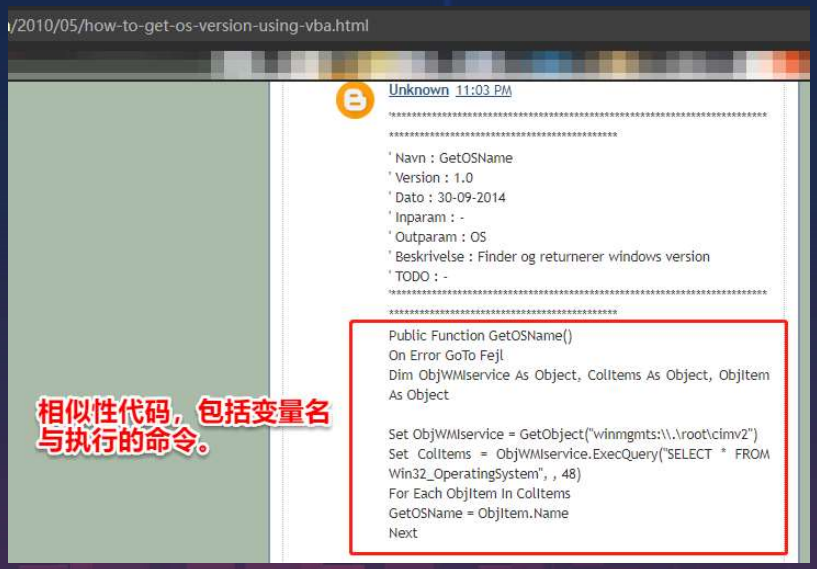


Fig. 21 Similar macro code in Chinese forum



字符串或特殊编码特征 hunting

除了 PE 文件的元数据外，同样可以基于样本的独特特征来进一步挖掘更多海莲花攻击者的恶意文件。

不寻常的字符串/常数

值得注意的加密/混淆算法

```
00001000D72A        0
00001000D80A        0   abcdefghijklmnopqrstuvwxyz
00001000D82A        0   ABCDEFGHIJKLMNOPQRSTUVWXYZ
00001000DDB0        0   @HAIBJCK
                        <?xml version="1.0" encoding="UTF
                        <assembly xmlns="urn:schemas-micr
                        sWND8kL[
```

```
                    LAB_1000edb0                    XREF[1]:    FUN_10002ab0:10002b04(*)
1000edb0 d9 e1          FABS
1000edb2 d9 74 24 f4    FNSTENV     [ESP + -0xc]
1000edb6 5f             POP         EDI
1000edb7 31 c9          XOR         ECX,ECX
1000edb9 66 b9 85 ee    MOV         CX,0xee85
1000edbd b8 57 43       MOV         EAX,0x7f954357
         95 7f          Shikata Ga Nai编码混淆器
1000edc2 31 47 1c       XOR         dword ptr [EDI + 0x1c],EAX
1000edc5 03 47 1c       ADD         EAX,dword ptr [EDI + 0x1c]
1000edc8 83 ef fc       SUB         EDI,-0x4
1000edcb e2 a2          LOOP        LAB_1000ed6f
1000edcd 98             CWDE
1000edce 41             INC         ECX
1000edcf c5             ??          C5h
1000edd0 a6             ??          A6h
1000edd1 68             ??          68h    h
1000edd2 f7             ??          F7h
1000edd3 98             ??          98h
1000edd4 ef             ??          EFh
1000edd5 e3             ??          E3h
1000edd6 23             ??          23h    #
1000edd7 29             ??          29h    )
1000edd8 51             ??          51h    Q
```

提取这一段汇编代码特征

shellcode解码指令

白+黑的样本案例

依据这些挖掘思路，我们利用内外部数据源挖掘到以下所属海联花组织的部分恶意攻击样本：

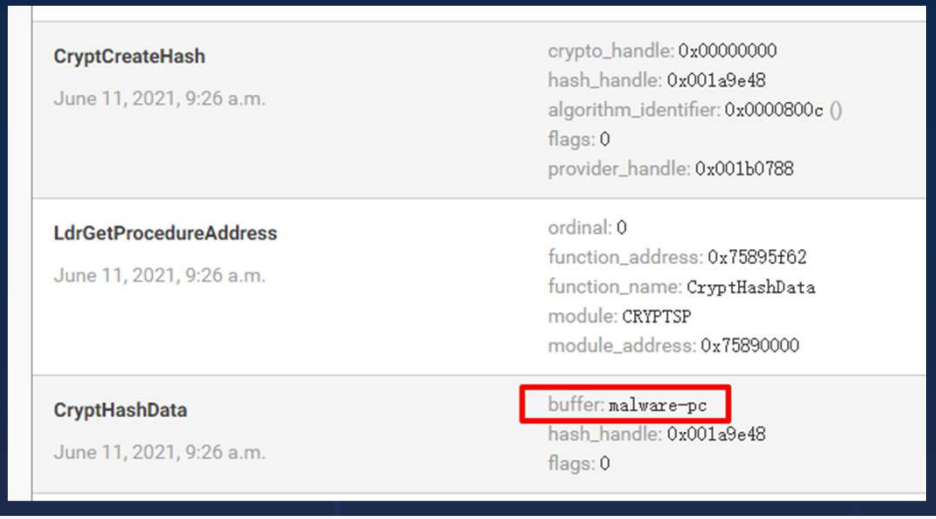
原始文件名	加载的母体文件	回连IP	来源地区
无	wmpnetwked.exe	无法确定	CN
123.txt	acrobatupdater.exe	无法确定	CN
oinfo11.ocx	无	185.225.19.22	CN
kdump.dll	无	无法确定	ID（印尼）
acSpecmain.sdb	无	185.225.19.22	CN
11coccocpdate.dll	无	无法确定	VN（越南）
OInfo11.OCX	oinfop11.exe	无法确定	CN
skhooks.dll	skdh8811.exe	无法确定	HK
gtn.dll	无	无法确定	VN
Ahnl2.dll	无	无法确定	CN
actxprxy.dll	无	无法确定	无法确定
winspc.dll	无	无法确定	VN
QMLogEx.dll	无	无法确定	CA（加拿大）
Ahnl2.dll	无	无法确定	CN
LBTServ.dll	无	无法确定	CN

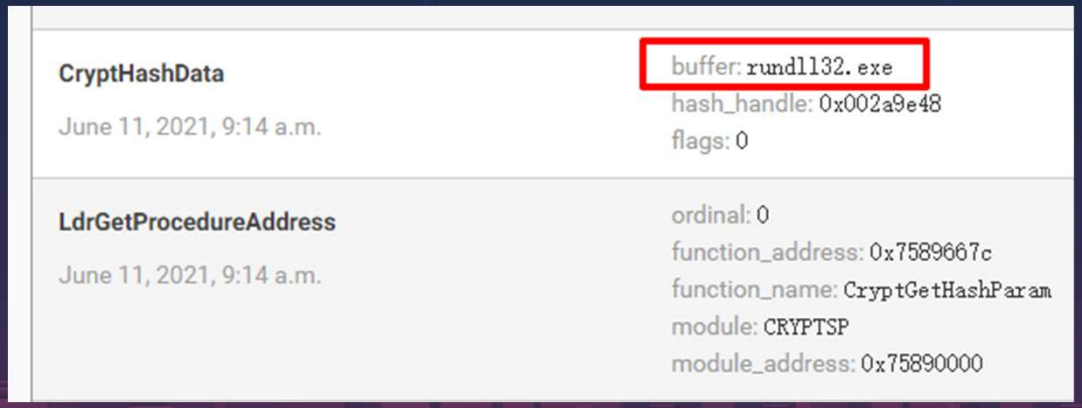
样本和攻击目标绑定

通过对获取的所有样本的分析结论表明，海莲花组织历史攻击样本执行流程里出现了四种密钥获取逻辑。

第一种：当前计算机名称作为输入密钥，用于解密后续的 shellcode 代码。

第二种：当前的系统 MAC 地址作为输入密钥，用于解密后续的 shellcode 代码。





样本和特定任务绑定

第三种：当前加载的母体文件名称作为输入密钥，用于解密后续的 shellcode 代码。

第四种：当前系统 IP 地址作为输入密钥，用于解密后续的 shellcode 代码。

基础设施 hunting

从定制化的攻击样本中无法挖掘到海莲花组织的更多情报，为了突破这个限制，只能寻求新的方向去挖掘，主要聚焦于以网络资产为核心进行研究挖掘扩展。

需要重点关注的点：

- C2（Domain、IP）
- Whois、DNS 解析记录
- 网络交互数据

最终的目标同样是找到更多 IOC 情报与攻击者恶意文件样本

挖掘前的准备

首先挖掘的前提是在于需要一个已经被分析人员多方论证判定属于攻击者的基础设施情报，之后依据此情报为起点进行后续挖掘。

网络侧挖掘主要挖什么?

攻击者的资产信息（历史运营，正在运营，准备用于攻击），基础设施（IP、域名、邮箱等）。

依据的普遍原理是什么?

网络空间仅仅只是一种利用媒介，其实归根结底还是人与人之间的对抗。

是人就总有弱点，攻击者会存在个人习惯，惯性思维。

从事 APT 活动的背后组织架构庞大，分工明确，无法面面俱到。

遭遇的困境

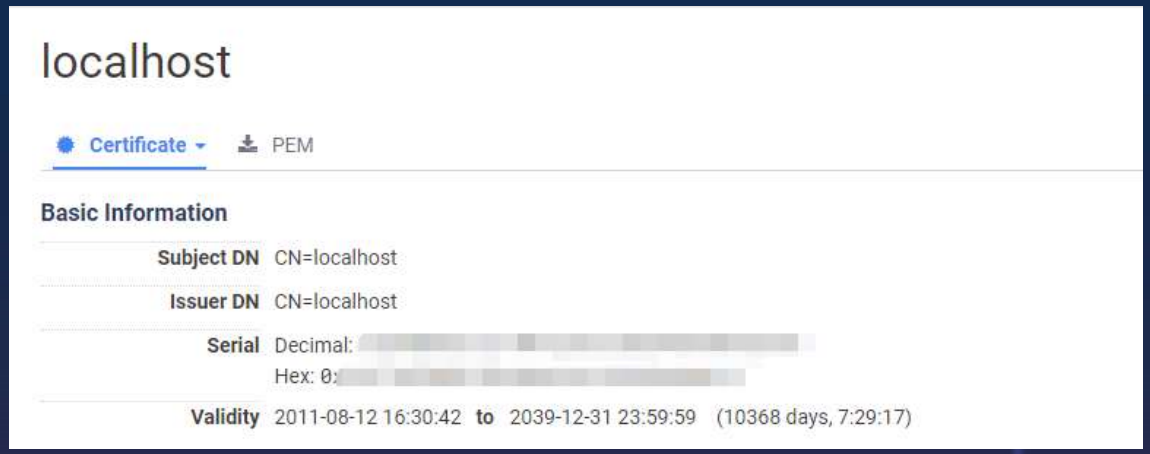
除了目前遭遇的 APT 组织的攻击者采用定制化攻击文件的策略使得分析人员情报挖掘推进举步维艰外，后续挖掘过程还遇到了以下问题：

- 我们内部是否有需要的资源？
- 如何判断 IP 地址在某个时间点是否为攻击者所拥有使用权？
- 如何判断域名当前是否还被攻击者所拥有运营权？
- 如何找到网络侧的攻击者的特征？
- 背后攻击者的特征有哪些？
- 哪些是我们重点关注的特征？

SSL证书的强关联特征

通过在文件侧实际挖掘到的攻击样本，我们拿到了确定为 APT 组织海莲花所属的网络资产信息，通过资产185.***.***.*** 获取到 2020 年 9 月份的 SSL 证书，之后通过 SSL 证书搜索，拿到绑定过相同证书的 20 个 IP 资产。

右侧截图中攻击者采用自签名的 SSL 证书，挖掘起始时间时间从 2018 年到如今，陆续找到总共 49 个独立 IP 资产。



异常特征关联

通过我们的持续观察，海莲花组织很喜欢使用 Red Hat 操作系统，网络交互数据 Server 字段值得注意“Apache/2.4.34 (Red Hat) OpenSSL/1.0.2k-fips”，但是数据里的标题为“Welcome to nginx!”，依据这个异常数据，最终我们挖掘到一批网络资产，经过分析人员的多方验证后确定了其中属于海莲花组织的网络资产。

```
1  HTTP/1.1 200 OK
2  Connection: close
3  Content-Length: 608
4  Accept-Ranges: bytes
5  Content-Type: text/html; charset=UTF-8
6  Date: Sat, 09 Oct 2021 13:06:26 GMT
7  Etag: "260-5a9c04d3da0a9"
8  Last-Modified: Mon, 06 Jul 2020 06:51:18 GMT
9  Server: Apache/2.4.34 (Red Hat) OpenSSL/1.0.2k-fips
```

跟踪中文标题特征

意外出现的中文标题内容“没有找到站点”引起了我们的警觉，结合之前的服务端组件特征“Apache/2.4.34 (Red Hat) OpenSSL/1.0.2k-fips”。

最后的挖掘结果收获很大，之后经过多方验证后确定了其中属于海莲花组织的网络资产。

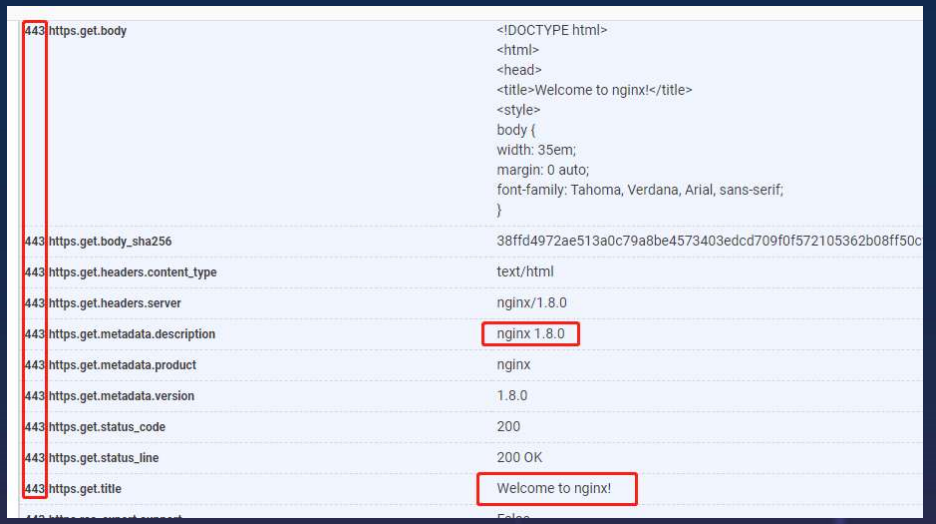
```
HTTP/1.1 200 OK
Connection: close
Content-Length: 1272
Accept-Ranges: bytes
Content-Type: text/html; charset=UTF-8
Date: Wed, 22 Sep 2021 10:02:20 GMT
Etag: "4f8-5bdb858b3db3a"
Last-Modified: Wed, 17 Mar 2021 09:45:43 GMT
Server: Apache/2.4.34 (Red Hat) OpenSSL/1.0.2k-fips

<!doctype html> <html>
<head> <meta charset="utf-8">
<title>没有找到站点</title>
<style> *{margin:0;padding:0;color:#444} body{font-size:14px;font-family:"宋体"} .main{wi
</head>
<body>
<div class="main"> <div class="title">没有找到站点</div> <div class="content">
<p class="t1">您的请求在Web服务器中没有找到对应的站点！ </p>
<p class="t2">可能原因： </p> <ol> <li>您没有将此域名或IP绑定到对应站点!</li> <li>配置文
<p class="t2">如何解决： </p> <ol> <li>检查是否已经绑定到对应站点，若确认已绑定，请尝试
</li> </ol> </div> </div> </body> </html>
```

从人性的角度关联

对于背后的 APT 组织的运营团队来说，他们很可能会依据人的习惯，选择特定区域部署相关资产。

经过我们的持续观察，网络资产出现在荷兰地区的概率较大，于是通过限定资产的区域，我们利用上述的标题“Welcome to nginx!”以及更改的服务端组件为“nginx 1.8.0”，通过限定端口为443，我们挖掘到一批数量上分析人员能分析处理的资产，经过分析人员的多方验证后确定了其中属于海莲花组织的网络资产。



443	https.get.body	<!DOCTYPE html>
		<html>
		<head>
		<title>Welcome to nginx!</title>
		<style>
		body {
		width: 35em;
		margin: 0 auto;
		font-family: Tahoma, Verdana, Arial, sans-serif;
		}

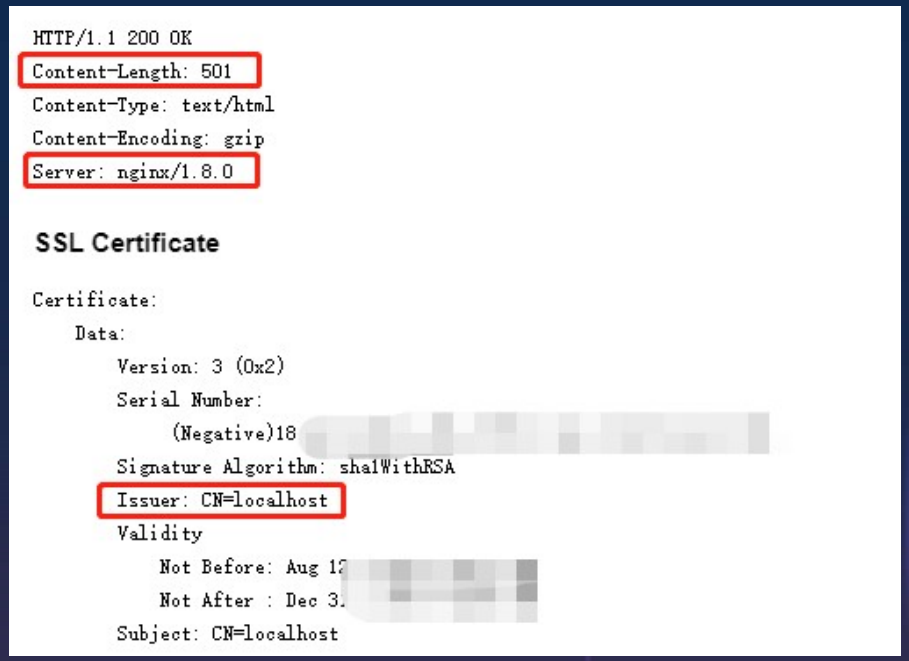
443	https.get.body_sha256	38ffd4972ae513a0c79a8be4573403edcd709f0f572105362b08ff50c
443	https.get.headers.content_type	text/html
443	https.get.headers.server	nginx/1.8.0
443	https.get.metadata.description	nginx 1.8.0
443	https.get.metadata.product	nginx
443	https.get.metadata.version	1.8.0
443	https.get.status_code	200
443	https.get.status_line	200 OK
443	https.get.title	Welcome to nginx!

多个特征结合关联

依据之前 SSL 证书的强关联特性，我们发现海莲花组织的 C&C 服务器往往采用自签名证书，依据这个特性我们之后结合之前多个特征进行限定挖掘。

我们关注自签名证书的网络资产，限定选择返回的数据长度 501 字节，且服务端组件为“nginx 1.8.0”。

最终分析并确定了其中属于海莲花组织的网络资产。



确定攻击者 IP 资产存活时间

如何确定攻击者 IP 资产的存活时间范围?

可利用历史数据挖掘并确定出攻击者真实历史 IP 资产

举例 Cobalt Strike 服务器特征识别（版本 3.X）

NanoHTTPD 404 Not Found 响应异常

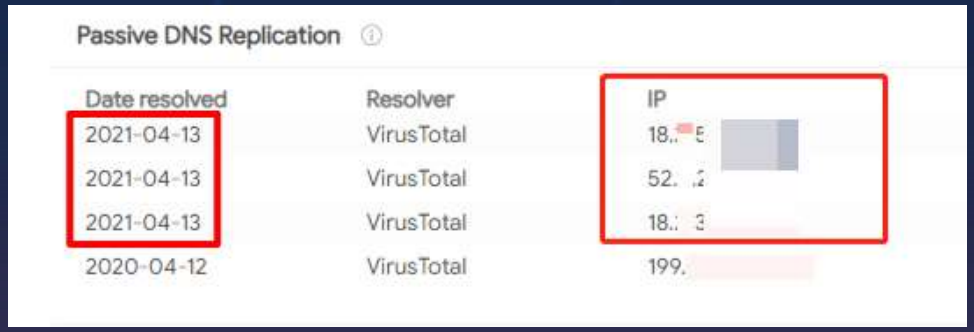
空字节异常

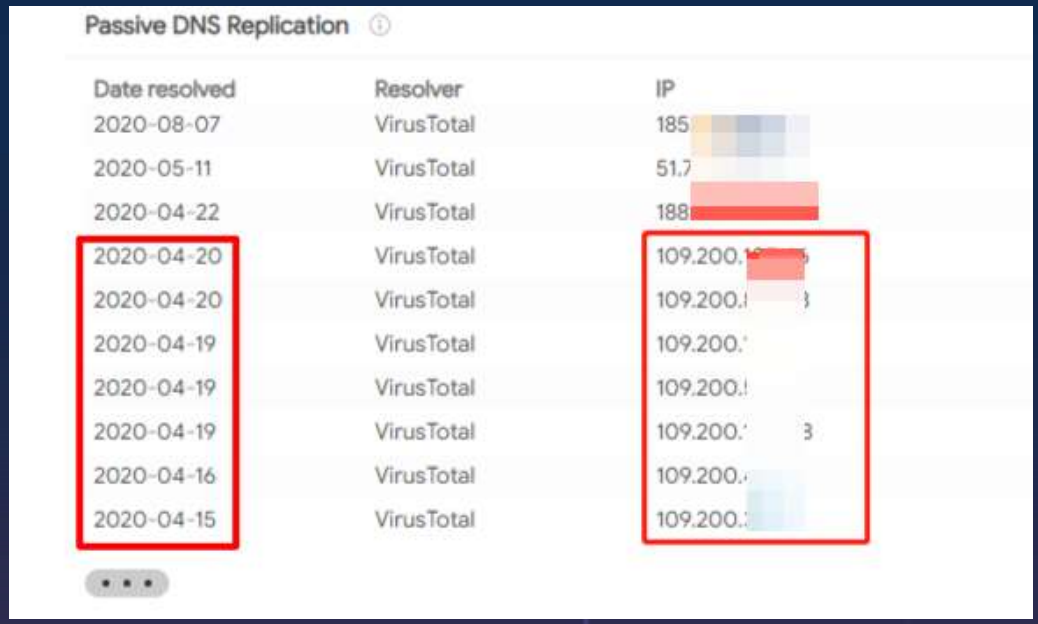
```
1  HTTP/1.1 404 Not Found
2  Connection: close
3  Content-Length: 548
4  Content-Type: text/html
5  Date: Fri, 24 Sep 2021 03:34:02 GMT
6  Server: nginx
7  Vary: Accept-Encoding
```

```
1  NanoHTTPD 404 Not Found响应异常
2  空字节异常
3
4  HTTP/1.1 404 Not Found
5  Date: Fri, 26 Jun 2020 23:02:09 GMT
6  Content-Type: text/plain
7  Content-Length: 0
```

循环 DNS（Round-robin DNS）技术

为了隐藏自身宝贵的 IP 资产，海莲花组织采用类似循环 DNS（Round-robin DNS）技术不断的切换 B 段 IP 替代域名的解析也是一个高维度的运营特征。





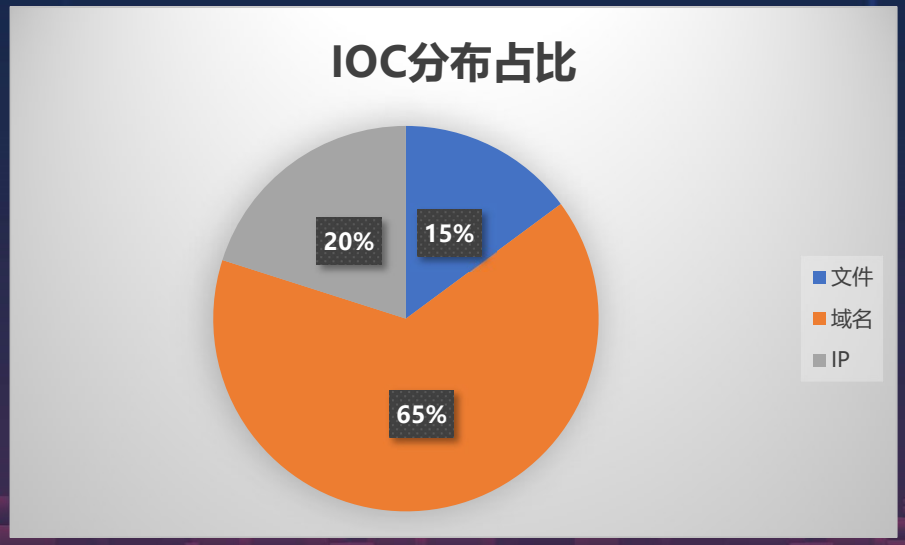
扩大战果

从文件侧挖掘到恶意攻击文件，提取到属于攻击者的网络资产，意外发现失陷主机，这些IP资产成为了攻击者的跳板，利用我们掌握的运营特征指纹，再次关联到其他失陷主机，从而扩大了战果。

区域	IP	类型
香港	223.***.***.***	企业专线
杭州	114.***.***.***	企业专线
蚌埠	60.***.***.***	企业专线

挖掘效果

依据这些特征，针对海莲花组织我们在几个月时间里挖掘到了大量文件侧与网络侧的 IOC 情报，其中后续经过文件侧挖掘的恶意文件里提取的信息也印证了从网络侧（基础设施）挖掘的情报准确度非常高，我们不仅拿到了大量攻击者正在运营还未失效的资产以及刚注册不久的域名（依据运营经验，后续大概率将用于攻击）、且未被公开的历史资产等情报。



总结-基于情报的挖掘思路

以上的挖掘思路，最终都会促使我们流向情报侧，而分析人员通常依赖三种情报源：

- 互联网上公开可用的开源情报（OSINT）；

通过已经掌握的网络侧特征，我们可以利用第三方情报源挖掘更多有价值的情报，例如：URL特征、交互的数据特征等。

- 依靠恶意文件分析的技术情报（TECHINT）；

通过已经掌握的文件侧特征，在各种数据源中利用这些特征（元数据、代码同源性、利用手法等）挖掘更多相关联的恶意文件，从而提高获取更多相关情报的概率。

- 仅涉事机构或者组织才有的专有数据；

在具体的真实安全事件中，如有条件可获取到第一手价值情报。


看雪·第五届
安全开发者峰会
Thank You
#include <stdio.h>
int main()
{
printf("Hello,World!);
return 0;
#include <stdio.h>
int main()
{
printf("Hello,World!);
return 0;