# 基于统计学习的网络异常行为检测技术

**周 涛**
北京启明星辰信息安全技术有限公司 北京 100193

**摘要**
高级持续性威胁（APT）已经成为企业级安全用户的首要安全威胁。传统基于特征检测、边界防护的安全防范措施在应对APT攻击时存在不足。为此，介绍了网络异常行为检测方法的现状；分析了基于统计学习的检测方法的技术路线和体系架构，并以命令控制通道、获取行为等APT攻击中的典型环节为例，介绍了相关的参数提取和统计分析建模方法；总结了基于大数据的异常行为检测的特点，并指出了后续研究方向。

**关键词**
大数据；安全分析；异常行为检测；统计学习

doi: 10.11959/j.issn.2096-0271.2015039

# *Abnormal Network Behavior Detection Technology Based on Statistical Learning*

**Zhou Tao**
Beijing Venus Information Security Technology Incorporated Company, Beijing 100193, China

***Abstract***
In recent years, advanced persistent threat (APT) has become the chief threat to enterprise users. The traditional security protection methods, such as signature-based detection and perimeter protection, are insufficient in dealing with APT. Therefore, the status of network anomaly behavior detection method was described. The technology roadmap and system architecture of abnormal behavior detection based on statistical learning were introduced. The feature extract method and statistical modeling methods were proposed. The characteristic of abnormal behavior detection based on big data was concluded and the direction of future research was proposed.

***Key words***
big data, security analysis, abnormal behavior detection, statistical learning

## 1 引言

近年来，由高级持续性威胁（advanced persistent threat，APT）引发的信息安全事件层出不穷，使得信息安全成为大众关注的焦点。从已曝光的APT攻击案例可以看出，大量具备高度经济价值或特殊政治地位的机构成为APT攻击的目标，例如伊朗的核电站、互联网行业巨头谷歌公司、美国最大的武器制造商洛克希德马丁公司、信息安全行业的领跑者RSA和卡巴斯基、全美第二大零售商Target公司等。APT攻击已经超越传统的蠕虫、病毒、木马等恶意软件，成为企业级用户面临的首要安全威胁。

由于APT攻击具有攻击方法多样化、攻击技术复杂先进、攻击持续时间长等特点，传统基于特征匹配、边界防护的安全防范措施在应对时存在不足，新兴的针对APT攻击的检测都将重点放在了基于异常行为的检测上。从体现异常行为的主体来看，异常行为检测可分为终端异常行为检测和网络异常行为检测两大类，产业界和学术界在两个方面都取得了一定进展。

在信息安全产业界，FireEye研发了一种基于终端异常行为的APT检测系统，它将原本应用于安全厂商后台进行样本分析的沙箱技术，前置到用户环境的检测设备中[1]，开创了APT检测的新局面，但其检测过程建立在捕获恶意代码的基础上，存在被绕过的可能。RSA则提出了一种基于网络异常行为检测的原型Beehive[1]，通过对企业网络环境中各类日志的大范围收集和分析进行异常检测，但检测结果有效性的确认仍然需要大量的人工鉴定工作。当前对于APT攻击，产业界还尚未形成完整有效的解决方案。

在学术界，基于异常行为识别已知或未知的安全威胁一直都是研究的热点，相关研究成果包括：Kim等提出了一种基于Netflow数据的异常行为检测框架[2]，能够对蠕虫、DDoS攻击、网络扫描等行为进行检测，该技术侧重对大规模异常的检测，对于APT攻击这类注重自身隐蔽性、不引起网络流量显著异常的攻击行为检测能力有限。McCusker等建立了一套用于描述网络攻击的行为基元[3]，并基于Netflow数据设计了相应的检测原型，能够实现对APT攻击中信标等行为的检测，但在检测中采用了SVM算法进行分类模型训练，这在实际应用中往往会遇到训练样本难以采集的问题。Bhatt等提出了一种基于攻击链（kill chain）的APT攻击描述模型[4]，能够对攻击的过程和结果进行建模，并提出了基于Hadoop平台的系统框架，但对于如何进行攻击行为检测并未给出具体的算法描述。

本文从利用大数据检测APT攻击的应用背景出发，以网络异常行为检测为重点，提出了一种基于统计学习的异常检测技术。本文的主要贡献如下。

（1）提出了一种面向APT攻击的网络异常行为检测技术框架。本文对比了人工提取特征和深度学习两种方式的优缺点，根据现状确定了基于特征建模的技术路线，并建立了大数据异常检测平台框架。

（2）以APT攻击过程中的两个典型环节（命令与控制、信息获取）为例，详细介绍了特征提取方法，给出了完整的特征定义。

（3）结合真实网络环境数据，对算法的有效性进行了验证，结果表明：本文所述的检测算法，对APT攻击过程中不造成网络流量显著异常的攻击行为，仍然能够进行有效检测。

1 https://www2.fireeye.com/ppc-definitive-guide.html

## 2 基于统计学习的异常检测技术框架

### 2.1 技术路线选择

异常行为检测从本质上看是一个分类问题，即从行为数据中将正常行为和异常行为区分开。当前可供选择的技术路线包括：基于人工特征提取的传统统计学习方法以及不需要显式特征提取的深度学习方法。

基于人工特征提取的异常检测技术的技术路线是：分析人员首先以某种方式从原始数据中提取特征参数，然后基于特征进行建模和异常检测。人工特征提取的优点是：特征提取建立在安全分析人员的认知基础之上，对异常行为有较强的针对性；对训练样本数量的依赖度低，较少的样本训练即可得到相对准确的模型；模型的可解释度高，容易确定异常检测结果的有效性。但人工特征提取也存在着明显的缺点：对安全分析人员的依赖度高，特征的选取方法会对异常检测结果的有效性产生直接影响。当前已有的异常检测技术中，大都是基于人工特征提取的方法实现的。

深度学习技术是当前机器学习领域的研究热点，同传统的统计分析方法相比，深度学习提出了一种让计算机自动学习产生特征的方法，并将特征学习融入建立模型的过程中，从而减少了人为设计特征引发的不完备。虽然深度学习方法有诸多优点，但也有其局限性：需要有大量的训练样本进行训练，才能保证模型的准确度。当训练样本数量不足时，深度学习算法将不能够对数据的规律进行无偏估计，模型的识别效果可能还不如传统基于人工特征提取的统计分析方法。当前深度学习的成功应用大都集中在有大量训练样本的模式识别领域，如语音识别、图像识别、机器翻译等。

采取何种技术路线能够实现更有效的APT攻击检测，取决于应用的前提条件。当前与APT攻击相关的网络行为数据样本极为有限，从已曝光的APT攻击案例来看，截至2015年8月全球范围内相关的报告仅有200余篇[2]，而与攻击相关的网络行为数据更是无从获取。因此，基于当前有限的案例，很难通过深度学习的方法产生能有效识别APT攻击行为的分类器。从另一方面来看，现有的APT分析报告对攻击的过程和方法大都有详细的描述，这就有助于安全分析人员从中了解APT攻击各个阶段的特点，并有针对性地提取特征，以提高检测的准确度。

因此，通过对两种技术路线的比较，结合当前的实际情况，本文认为在现阶段通过网络异常行为检测技术进行APT攻击检测时，采用基于人工特征提取的方法，比深度学习的方法更符合当前企业级用户的应用场景。或许随着对APT攻击研究的进一步深入、攻击案例和攻击行为数据的持续积累，基于深度学习的训练方法能够更有效地识别攻击，但在当前阶段还是基于特征的建模方法更有效。基于大数据平台进行特征的提取和训练，能够使模型的准确度显著提升。

### 2.2 基于大数据的异常检测技术特点

异常检测并非是一项新技术，事实上在入侵检测概念产生初期，IDES、NIDES等原型系统都采用了异常检测技术[5]。但受限于当时的技术条件，异常检测的准确度较低，主要原因如下。

(1) 模型粒度问题

由于计算能力有限，在当时很难建立对异常行为较为敏感的细粒度模型，从而导致较高的漏报率。以异常流量检测为

2 https://github.com/kbandla/APTnotes

例，当时的建模对象往往基于安全域间的流量，这就使得个体间的攻击流量淹没在大量背景流量中，很难进行有效检测。

（2）特征数量问题

同样由于计算能力有限，在当时很难建立从不同维度描述网络行为的高维模型，从而导致较高的误报率。特征数量选取的限制，使得只能基于低维的特征判断网络行为的异常度，很难通过特征间的关联降低误报。

（3）模型训练问题

由于存储容量有限，在当时很难基于长期的数据对模型进行充分的训练，从而导致模型的准确度不足。模型的准确度与训练是否充分是有直接关系的，虽然安全分析人员的经验有助于提升特征选取的有效性，但仍然需要足够的样本对模型进行训练。

因此，虽然异常检测具有能够识别未知威胁的优势，但当时国内外商业化的入侵检测产品大都选择了基于攻击签名的误用检测技术。随着攻防博弈的发展，APT攻击成为首要的安全威胁，这就使得安全研究人员需要对技术路线的选择进行重新思考。

首先，误用检测在应对未知威胁检测方面的先天不足，使其无法成为APT攻击时代的支撑性技术；其次，原本导致异常检测技术准确度不足的各种障碍，都随着IT技术的发展、大数据时代的到来消失了。这就带来了检测技术的回归，异常检测重新成为了安全界研究的焦点。这种回归不是简单的反复，而是在一个更高层次上的螺旋式上升。

面向APT攻击的异常行为检测原理如图1所示，安全分析人员首先通过对APT攻击方法的分析和总结，提取出有针对性的特征；然后基于训练数据，采用有监督或无监督的方法，对模型进行训练；基于训练产生的模型，可利用测试数据或真实数据对模型的有效性进行验证；对于错误的检测结果，可通过调整特征参数的方式反馈到模型训练环节，直至产生满足准确度要求的模型。虽然该原理与传统异常检测技术并无本质区别，但基于大数据的异常检测技术具有如下特点。

（1）更细的模型粒度：与传统以安全域为建模对象不同，基于大数据的异常检测技术可以基于单个的主机，甚至是主机上的单个应用建立细粒度的模型，这就使得模型对异常的检测能力有足够的灵敏度。这样做的计算代价是很大的，以一个仅有数千台主机的小型企业为例，基于主机间的连接建模得到的模型数量将会是百万量级，这在过去是很难想象的，但目前大数据平台的性能足以支撑类似规模的计算。

（2）更高维的特征选取：基于大数据提取特征参数，可以从建模对象的时空维度、行为维度等方面抽取足够丰富的特征参数，使得对于任何可描述的攻击行为，总能体现在一组特征参数的异常上，真正做到让攻击者在大数据下无处可遁。

（3）更充分的模型训练：基于大数据平台的海量存储能力，可存储足够多的历史流量数据作为样本，对提取的参数和模型进行充分训练，使得模型对于异常行为具有足够精确的检测能力。

由此可见，基于大数据实现网络异常行为检测，克服了早期异常检测技术的不足，给检测技术带来了质的飞跃，但同时也

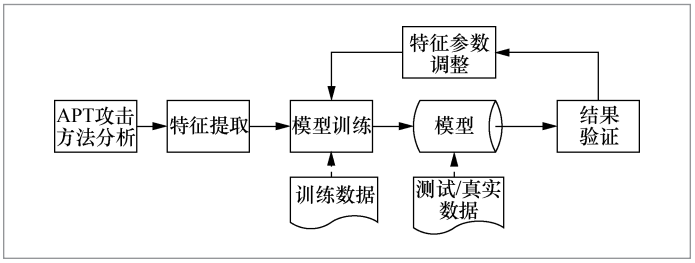



图1 面向APT攻击的异常行为检测原理

对存储和计算能力提出了新的挑战，需要有一个能提供有效支撑的技术平台。

### 2.3 大数据异常检测平台框架

本文提出的大数据异常检测平台可分为4层：数据采集层、存储管理层、入侵行为分析挖掘层和展示及配置管理层，其框架如图2所示。

一个完整的大数据异常检测平台，在数据源层面上要具备完整的数据采集能力，包括与网络行为相关的各类日志、网络流量以及情境数据和外部支持数据的采集。在存储层面上要能够支持异构数据存储，能够通过缓存应付突发的数据，具备弹性扩展能力。在分析层面上要能够支持灵活的特征提取、基于特征的统计分析和模型训练以及对检测结果的事后取证溯源和验证能力。在展示层面上要能够支持大数据平台集群配置管理和数据的交互式可视化分析。

## 3 网络异常行为检测算法

在对APT攻击进行描述时，攻击链是当前被广泛接受的模型[6]，它将APT攻击过程分为情报收集、恶意代码组装、投送、激活、安装植入、命令与控制、获取7个阶段。其中，除了恶意代码组装阶段是在攻击方实施，无法监控其行为之外，其他6个阶段均有可供检测的行为特征。本文以网络流量为检测对象，以命令与控制、获取两个典型阶段为例，详细介绍相关的网络异常行为建模算法。

### 3.1 命令与控制通道行为检测

命令与控制通道的作用是在攻击者和攻击目标之间建立网络连接，使得攻击目标处于攻击者控制下，能够接收攻击者的命令并返回执行结果。命令与控制通道的功能决定了其行为与正常的网络连接行为有一

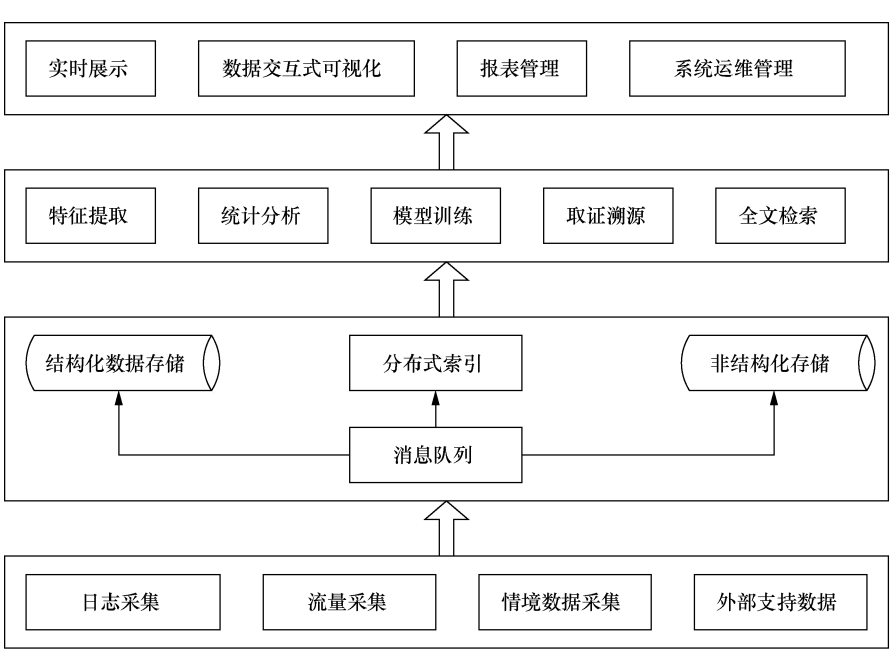



图 2 大数据异常检测平台框架

定差异，本文给出了相关特征的定义。

### 3.1.1 反向连接特征

**定义1** 活跃时间点。对于一条从源地址S到目的地址D的TCP流，按照从S到D以及从D到S方向，将数据传输的时间点划分为两个序列；对于每个序列$L$中的某个时间点$T_c$，如果与其前面的时间点$T_{c-1}$之间的间隔大于设定的阈值$\delta_a$，即满足：

$$T_c - T_{c-1} > \delta_a \quad (1)$$

则定义$T_c$为活跃时间点。否则，认为$T_c$为上次活跃数据传输的后续，而不是一个新的活跃时间点。

**定义2** 响应率。对于一条从源地址S到目的地址D的TCP流，计算从D到S的活跃时间点序列$L_{atv_D_S}$以及从S到D的活跃时间点序列$L_{atv_S_D}$；设定阈值$\delta_{resp}$；如果对于$L_{atv_D_S}$中的任意时间点$T_{D_S}$，在$L_{atv_S_D}$中存在一个对应的时间点$T_{S_D}$，使得满足：

$$T_{D\_S} < T_{S\_D}, \text{且} T_{S\_D} - T_{D\_S} < \delta_{resp} \quad (2)$$

则称$T_{D_S}$被响应，并称$T_{S_D}$为其响应点。响应率$R_{resp}$定义为$L_{atv_D_S}$中被响应的元素数量，与$L_{atv_D_S}$中总元素数量的比值。

**定义3** 激活率。参照定义2，对于一条从源地址S到目的地址D的TCP流，激活率$R_{actv}$定义为从S到D的活跃时间点序列$L_{atv_S_D}$中，能够作为响应点的元素数量与$L_{atv_S_D}$中总元素数量的比值。

**定义4** 反向连接特征。对于一条从源地址S到目的地址D的TCP流，如果满足：

$$R_{resp} > \delta_R, \text{且} R_{actv} > \delta_A \quad (3)$$

则称该TCP连接为反向连接，其中$\delta_R$和$\delta_A$为设定的阈值。反向连接是指一条TCP流建立连接的方向与后续活跃数据传输的方向相反的行为，本文通过计算响应率与激活率来检测反向连接。响应率高说明建立连接后，活跃的数据传输大都由连接的目标IP地址发起；激活率高说明源IP地址在建立IP地址后处于被动状态，只有等到接收到目标IP地址的数据后才进行数据传输。这种行为与正常的网络传输有明显的差异，是反弹端口型命令与控制通道的主要特征。

### 3.1.2 心跳特征

**定义5** 心跳行为。心跳行为是指一条从源地址S到目的地址D的反向连接TCP流中，存在的一系列从S发往D、大小相对固定、发送时间间隔平稳的数据传输行为。心跳行为通常用于APT攻击中被控端向控制端报告自身的存活状态，对该类行为可通过计算数据分组发送的平稳度特征来检测。

**定义6** 平稳度。对于一条从源地址S到目的地址D的TCP流，按照传输数据分组大小的不同，将从S到D的数据传输划分为不同的序列；计算每个序列中相邻两次数据传输的时间间隔$T_\Delta$，得到若干个时间间隔序列$L_{T_\Delta}$；计算每个时间间隔序列$L_{T_\Delta}$的均值$\mu$和标准差$\sigma$，并定义一个时间间隔序列平稳度为：

$$P = 1 - \frac{\sigma}{\mu} \quad (4)$$

取一条TCP流中所有时间间隔序列平稳度的最高值，作为整个TCP连接的平稳度。如果一个TCP连接的平稳度超过了设定的阈值$\delta_P$，则称该TCP连接存在心跳行为。

对于命令与控制通道，还存在着诸如可疑加密传输行为、上下行流量比异常、可疑恶意DNS域名解析等行为。对于可疑加密传输行为，可通过计算TCP流有效载荷部分的信息熵来实现检测；对于上下行流量比异常，可通过比较TCP流两个方向的有效载荷的大小来实现检测；对于可疑恶意DNS域名解析，可通过检测请求域名的文本特征和解析后的动态特征来检测。

## 3.2 获取行为检测

对于已经成功渗透到内网的APT攻击

行为，攻击者往往需要在内网进行横向转移，以获取目标数据，这就使得其网络访问行为与正常用户的网络访问行为有差异。本文以主机为对象，设计了一系列特征来检测主机的此类异常行为，并给出了如下的指标类型定义。

**定义7** 会话信息类指标。统计一台主机单位时间内不同协议类型的会话统计信息。如TCP连接次数、UDP连接次数、对应的流量、数据分组大小的均值和标准差等。

**定义8** 应用分布类指标。统计一台主机单位时间内不同应用类型的访问统计信息。如访问不同应用的流量分布、次数、目标地址位置、国别分布等。

**定义9** 指示位标识类指标。统计一台主机单位时间内收发的含特定协议标识位的数据分组数量及其比值。如含有TCP_SYN_send、TCP_SYN_ACK_receive、RST_send等标志位的会话数目；TCP_SYN_ACK_receive/TCP_SYN_send的比值；单位时间里的ICMP_T3、ICMP_Echo_Reply、ICMP_Echo_Request等报文数目。

**定义10** 地址分布指标。统计一台主机单位时间内访问的IP地址网段分布、内外网分布等参数。

通过以上指标的计算和提取，本文为主机的网络行为建立起了一个高维特征向量，并总结了各类攻击手法对相关特征的影响，使得对已知的各类获取行为，总能体现在一个或一组特征的异常上，从而实现了较为准确的APT攻击获取行为检测。

# 4 实验结果及分析

## 4.1 实验工作概述

为了验证本文算法的有效性，在北京启明星辰信息安全技术有限公司内网搭建了测试环境，通过真实数据进行检验。部署的网络流量捕获设备能够产生网络数据流的元数据和原始报文两类数据。通过该设备，获取了从2014年9月1日至2015年1月31日，与某个子网相关的约18.3亿条网络流量元数据以及与部分IP地址相关的总量约为2 TB的原始报文数据，并作为本文的实验数据。对异常行为的检测可分为两步。

- 特征提取：参照第3节的内容，分别为命令与控制通道行为和获取行为提取特征参数。
- 异常行为检测：基于特征参数的特点以及样本数量，可选取有监督或无监督的学习算法，进行特征值异常检测，具体算法限于篇幅不再详述。

通过模拟攻击的方法，验证了命令与控制通道检测算法的有效性；通过检测到的若干次真实的慢扫描行为，验证了获取行为检测算法的有效性。下面对检测过程进行详细描述。

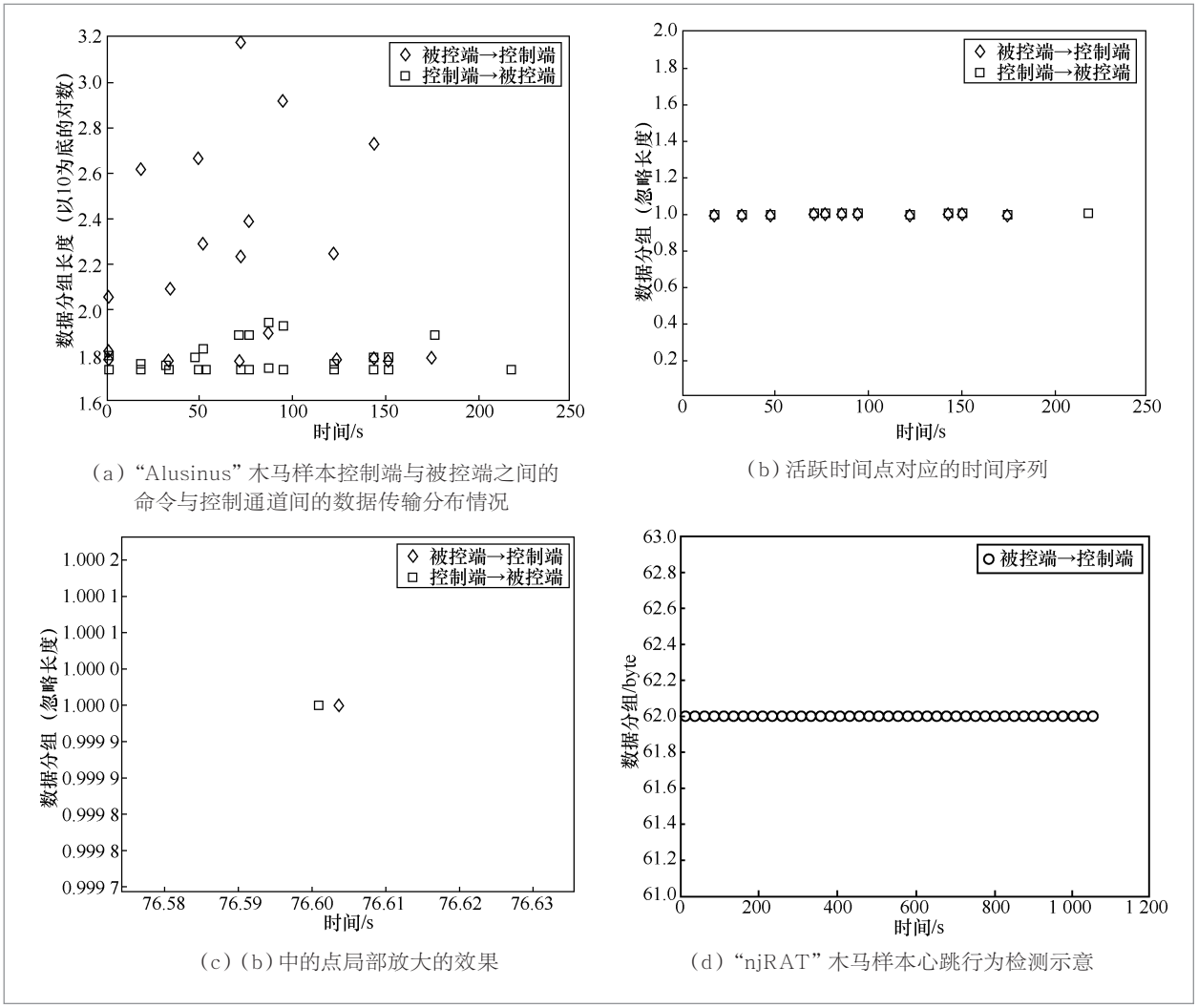
## 4.2 命令与控制通道检测实验结果

本文对命令与控制通道的检测采用了有监督的学习算法。由于样本数量有限，本文首先搭建测试环境，获取了6种典型的木马命令与控制通道的数据分组作为训练数据；然后通过密度估计的方法，利用训练数据分别计算正常网络连接和命令与控制通道的响应率、激活率、平稳度等特征的概率分布函数。

在测试中，本文利用“Alusinus”和“njRAT”木马样本，模拟了从某个受控主机到另一网段的控制端之间的命令与控制通道连接。通过提取的特征值，根据训练阶段产生的模型成功识别了这两次连接。两次连接对应的特征如图3所示。

图3（a）为“Alusinus”木马样本控制端与被控端之间的命令与控制通道间的数据传输分布情况。从图3（a）无法直接观测出是否存在反向连接行为、心跳等行为的特征，但可以看到从被控制端到控制端的数据分组要比反向的数据分组大1个数量级以上，存在明显的上下行流量不对称问题。按照定义1~定义4的计算方法，找出控制端和被控端之间两个方向的活跃时间点，并忽略数据分组大小的影响，可以得到活跃时间点对应的时间序列（如图3（b）所示）。可以看到，两个方向的活跃时间点总是成对出现（除了最后一个由控制端到被控端的活跃时间点之外，经验证，该次数据传输的命令为断开连接），响应率为91.7%，激活率为100%。图3（c）为将图3（b）中的点局部放大的效果，可以看到控制端发送命令在前，被控端响应命令在后。这些都是典型的命令与控制通道的特征。

图3（d）为“njRAT”木马样本心跳行为检测示意。本文根据定义5~定义6的计算方法，查找控制端与被控端之间平稳度最高的数据传输序列。从图3（d）中可以看到，由被控端发往控制端，大小固定为62 byte的数据分组，平均每19 s发送一次，平稳度高达99.8%，从而被本文所述的




（a）“Alusinus”木马样本控制端与被控端之间的命令与控制通道间的数据传输分布情况

（b）活跃时间点对应的时间序列

（c）（b）中的点局部放大的效果

（d）“njRAT”木马样本心跳行为检测示意

图 3 两次连接对应的特征

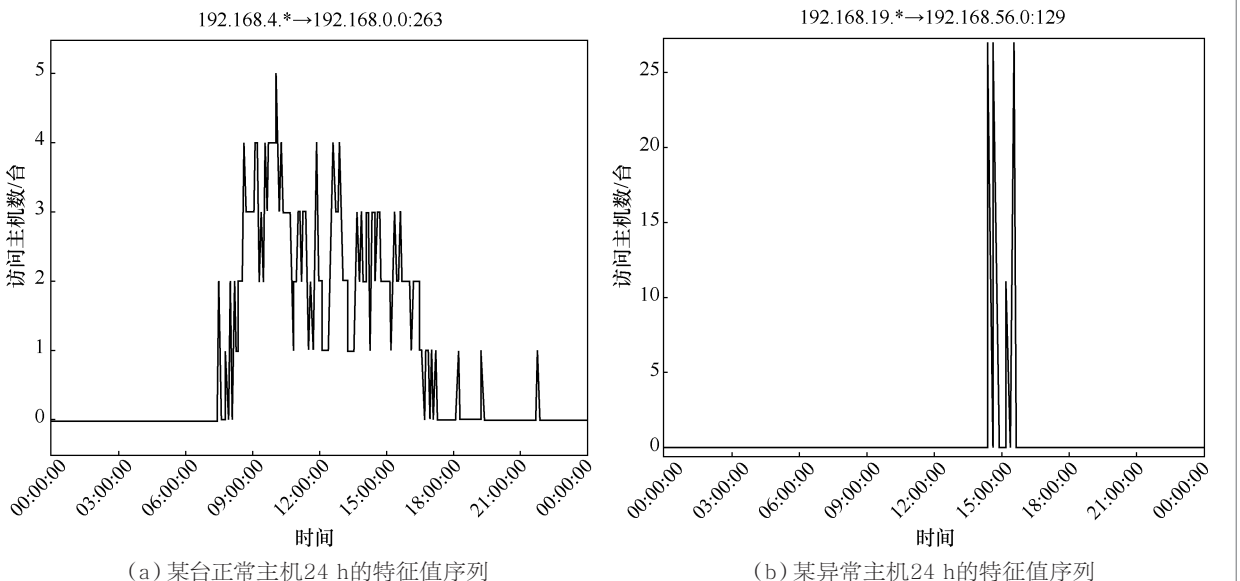
检测算法识别为异常行为。

### 4.3 获取行为检测结果

本文对获取行为的检测采用了无监督的学习算法。首先基于定义7~定义10的方法，计算出每个特征的时间序列；然后计算特征参数的密度函数，并进行异常点检测。异常点的定义准则为特征值明显偏离了均值的点。

本文以会话信息类指标中某具体指标，即主机—子网访问指标为例，说明算法的有效性。该指标定义为：在单位时间内，某台主机访问某个子网内不同独立主机的数量。以2014年9月10日为例，当天在测试环境中共产生7 240 086条流记录，从中可梳理出1 148个独立的IP和68个独立的子网以及11 165个存在着主机—子网访问的连接关系，从中只发现了一例异常行为，且该行为的有效性得到了验证。

图4（a）为某台正常主机24 h的特征值序列，从中可以看到192.168.4.0网段内的某台主机对192.168.0.0网段的访问情况。以5 min为时间窗，将每个时间窗内该主机访问该网段的IP地址数量相加，总共为263个。由于在不同时间窗内，一台主机站问过的IP地址集之间往往存在交集，这就导致按时间窗累加的IP地址数量，可能会超出该网段能够容纳的IP地址数量的上限。从图4（a）可以看到，该主机从上午约9点开始访问该网段，11点左右达到峰值，18点后迅速减少，这与用户的作息时间是高度相关的；每个时间窗内访问的不同主机数量为1~5个，相对平稳。对比图4（b），192.168.19.0网段的某台主机，15点起对192.168.56.0网段进行了低频扫描，每分钟尝试连接的主机数约为5个。通过该主机—子网访问特征，对此类行为能够进行准确识别。另外需要说明的是，P2P连接也会存在类似的特征，但P2P连接的地址分布指标与获取行为的连接有较大的差异。本文正是通过关联地址分布指标，在检测出异常时进一步判定了异常类型。




（a）某台正常主机24 h的特征值序列　　（b）某异常主机24 h的特征值序列

图4 不同情况下主机—子网访问指标对比

## 5 结束语

大数据安全分析是当前信息安全领域的研究热点，利用大数据进行异常行为检测，能够识别基于攻击签名无法检测的未知攻击行为。本文提出了一种基于统计学习的网络异常行为检测方法，其特点是提取有针对性的、细粒度的特征，并通过大数据进行模型训练和异常检测。这是一种把安全分析人员的经验，与大数据平台的存储和计算能力相结合的有效手段。实验表明，该方法对APT攻击过程中的命令与控制通道行为、获取行为有良好的检测效果。后续将对APT攻击过程其他阶段的特征进行总结和特征值提取，最终形成完整的、基于大数据安全分析的APT攻击检测方案。

## 参考文献

[1] Yen T F, Oprea A, Onarlioglu K, *et al*. Beehive: large-scale log analysis for detecting suspicious activity in enterprise networks. Proceedings of the 29th Annual Computer Security Applications Conference, New Orleans, Louisiana, USA, 2013: 199~208

[2] Kim A S, Kong H J, Hong S C, *et al*. A flow-based method for abnormal network traffic detection. Network Operations and Management Symposium, 2004(1): 599~612

[3] McCusker O, Brunza S, Dasgupta D. Deriving behavior primitives from aggregate network features using support vector machines. Proceedings of IEEE 5th International Conference on Cyber Conflict (CyCon), Tallinn,Estonia, 2013: 1~18

[4] Bhatt P, Toshiro Y E, Gustavsson P M. Towards a framework to detect multi-stage advanced persistent threats attacks. Proceedings of the 8th International Symposium on Service-Oriented System Engineering, Oxford, UK, 2014: 390~395

[5] Garcia-Teodoro P, Diaz-Verdejo J, Maciá-Fernández G, *et al*. Anomaly-based network intrusion detection: techniques, systems and challenges. Computers & Security, 2009, 28(1): 18~28

[6] Hutchins E M, Cloppert M J, Amin R M. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. Leading Issues in Information Warfare & Security Research, 2011(1): 80~106

作者简介

**周涛**，男，博士，教授级高工，就职于北京启明星辰信息安全技术有限公司，主要研究方向为大数据安全分析、事件关联分析、入侵检测等。

**收稿日期：** 2015-10-15

**论文引用格式：** 周涛. 基于统计学习的网络异常行为检测技术. 大数据, 2015039

Zhou T. Abnormal network behavior detection technology based on statistical learning. Big Data Research, 2015039