

协助设计高级持续性未歇检测系统的分析框架

非官方中文译本 · 安天实验室 译注

文档信息			
原文名称	An Analysis Framework to Aid in Designing Advanced Persistent Threat Detection Systems		
原文作者	J.A.de Vries, J. van den Berg, M.E. Warnier, H.Hoogstraaten	原文发布日期	2012 年 7 月 5 日
作者简介	J.A.de Vries, J. van den Berg, M.E. Warnier 是荷兰代尔夫特理工大学科技、政策与管理学院的教授。H.Hoogstraaten 是荷兰 Fox-IT 公司的员工。 --参见脚注 a 和 b。		
原文发布单位	荷兰代尔夫特理工大学 Fox-IT 公司		
原文出处	http://repository.tudelft.nl/assets/uuid:090446d3-7562-41ce-94d3-ab7153cd05d5/Scientific_paper_J.A._de_Vries.pdf		
译者	安天技术公益翻译组	校对者	安天技术公益翻译组
免责声明	<p>本译文译者为安天实验室工程师，本文系出自个人兴趣在业余时间所译，本文原文来自互联网的公共方式，译者力图忠于所获得之电子版本进行翻译，但受翻译水平和技术水平所限，不能完全保证译文完全与原文含义一致，同时对所获得原文是否存在臆造、或者是否与其原始版本一致未进行可靠性验证和评价。</p> <p>本译文对应原文所有观点亦不受本译文中任何打字、排版、印刷或翻译错误的影响。译者与安天实验室不对译文及原文中包含或引用的信息的真实性、准确性、可靠性、或完整性提供任何明示或暗示的保证。译者与安天实验室亦对原文和译文的任何内容不承担任何责任。翻译本文的行为不代表译者和安天实验室对原文立场持有任何立场和态度。</p> <p>译者与安天实验室均与原作者与原始发布者没有联系，亦未获得相关的版权授权，鉴于译者及安天实验室出于学习参考之目的翻译本文，而无出版、发售译文等任何商业利益意图，因此亦不对任何可能因此导致的版权问题承担责任。</p> <p>本文为安天内部参考文献，主要用于安天实验室内部进行外语和技术学习使用，亦向中国大陆境内的网络安全领域的研究人士进行有限分享。望尊重译者的劳动和意愿，不得以任何方式修改本译文。译者和安天实验室并未授权任何人士和第三方二次分享本译文，因此第三方对本译文的全部或者部分所做的分享、传播、报道、张贴行为，及所带来的后果与译者和安天实验室</p>		

	无关。本译文亦不得用于任何商业目的，基于上述问题产生的法律责任，译者与安天实验室一律不予承担。
--	---

协助设计高级持续性未歇检测系统的分析框架

J.A. de Vries, J. van den Berg, M.E. Warnier^a, H. Hoogstraaten^b

2012 年 7 月 5 日

a 荷兰代尔夫特理工大学，科技、政策与管理学院。

b 荷兰 Fox-IT 公司

摘要

针对企业和政府的网络攻击的复杂性、持续性和数量都在不断增加。与先前已知的多步骤攻击相比，攻击者付出了更多的时间和精力来规避检测。常见的入侵检测方法无法检测到这种复杂的攻击。因此，需要一种新的基于攻击特征的检测方法，以便逐步分析新型威胁。我们提出了一个分析框架：将攻击（例如攻击步骤和特征）与检测和业务联系起来。该框架可以作为检测系统设计的路线图。将该框架作为系统设计的路线图能够分析网络中多个位置的网络流量和客户端数据。这些分析采用签名和异常检测方法。

1.引言

计算机已经成为我们日常生活的一部分，互联网将全球范围内的用户和企业联系起来。对网络基础设施的恶意攻击可以追溯到 80 年代，这些攻击导致了防御病毒和禁止未授权访问。目前，全球范围内网络基础设施攻击带来的损失为每年 1000 亿至 10000 亿美元[1]。网络犯罪活动很受罪犯的青睐，因为他们被抓和判刑的风险很低。其结果是，网络犯罪已经发展为一个完整的产业。另一方面，各国政府也发现网络空间可以用于监视其他国家，因此也是一个战争舞台[1]。

互联网中的病毒、蠕虫和其他恶意活动导致了防御系统的建立。病毒扫描、防火墙和入侵检测系统建立的目的是减少网络犯罪带来的经济损失。反过来，网络罪犯和间谍创造了更先进的手段来规避安全措施。今天，这项你死我活的竞争仍在继续。攻击者的目标明确，试图在不被察觉的前提下窃取专利信息。这些攻击通常被称为高级持续性威胁（APT）。APT 是一个新型的多步骤攻击，其运行更隐蔽，并具有特定目标（最常见的是间谍活动[2]）。正如普通的多步骤攻击

一样，APT 也采取多个步骤来实现目标。但是 APT 也有不同之处，因为攻击者更多地利用零日漏洞，零日漏洞是指软件中未知的安全漏洞以及其他先进的手段（如社会工程学[2]）。APT 是目前最严重的企业和政府威胁，因为当前的防御方法往往无法检测到 APT [3]。

本文提出了一种新的分析多步骤攻击（如 APT）的方法，旨在将攻击特征与检测方法联系起来，这些检测方法包括网络入侵检测系统（NIDS）或主机入侵检测系统（HIDS）。这些方法中的智能数据分析算法是检测网络攻击活动的关键。我们提出的框架考虑到攻击方法、检测方法和对业务的影响。从框架中吸取的教训已应用于设计检测 APT 的系统。

本文的结构如下。第 2 章介绍了多步骤攻击（特别是 APT）的背景，以及智能数据分析在入侵检测中的应用。第 3 章介绍了本文提出的分析 APT 的框架。第 4 章提出了采用智能数据分析方法检测 APT。第 5 章反思了设计方法，第 6 章总结了全文。

2.攻击与数据分析方法

2.1 网络攻击

计算机网络包含了大量的信息。大部分信息受到保护，以确保机密性、完整性和可用性。故意破坏安全的行为被称为网络攻击。网络攻击有多种不同的形式，范围从简单的拒绝服务攻击到复杂的网络间谍攻击。网络攻击的分析是一个持续的过程，以便了解攻击者的方法。攻击分类法可以帮助我们基于攻击特征将攻击进行分类。攻击和恶意代码家族的特征可以用于设计检测特征[4]。分类法也可以用作入侵检测系统的检查表，确保特征涵盖所有的已知攻击。攻击分类法并不一定针对单一的低级攻击方法，如病毒或漏洞利

用。分类法也可以包含高级攻击，高级攻击是低级攻击方法的组合序列[4]。防御系统，如杀毒软件和防火墙，用来对付低级攻击。自本世纪初，我们就开始研究多步骤的更复杂的网络攻击。这些多步骤攻击采用不同的攻击方法，以实现具体的目标。例如，侦察步骤可以将端口扫描用作攻击方法。在科学文献中，这些攻击通常被称为多步骤攻击或攻击场景。例如，Ning 等人试图将低级攻击联系起来，以减少入侵检测系统发出警报的数量[5]。Chuang 等人进一步扩展了这一理念，利用针对低级攻击的不同系统的信息来检测多步骤攻击场景[6]。Yang 等人提出了另一种检测多步骤攻击的方法。他们融合多个入侵检测系统的警报来识别多步骤攻击。他们还提出了一个包括 7 个攻击阶段的指导模板。第一阶段是来自外部网络的侦察攻击，最后阶段是在内部网络中实现攻击目标[7]。

2.2 高级持续性威胁

上述多步骤攻击检测方法假设大部分（甚至全部）攻击步骤都能被检测到。多步骤攻击的一个新变种，通常被称为高级持续性威胁，可以认为是多步骤攻击的新形式[2]。这些攻击与上文所述的攻击不同，这些攻击更隐蔽，攻击者的技术更高超且执着于实现目标。大量零日漏洞的使用使得检测更加困难。利用社会工程学和电子邮件将用户引导至恶意网站并安装恶意软件也是 APT 的共同特征。APT 通常被认为有侦察阶段、获得网络据点的阶段、寻找资源并最终获取专利数据的阶段 [2] [3] [8]。一个众所周知的 APT 案例是“极光行动”。这种攻击的目标是多个高价值公司并利用多个零日漏洞。社会工程学和加密混淆技术的使用使得我们难以检测和防御此类攻击[2] [3]。

防御 APT 首先需要保持软件和防御措施的

随时更新。但是，鉴于未知漏洞的使用，这还不够。我们需要一种改进的方法来检测 APT。

2.3 入侵检测中的智能数据分析

一般来讲，有 3 种不同的方法来检测入侵[9]。第一种方法是特征检测。特征检测系统将数据样本与系统中的特征相比较，如果有特征匹配，则发出警报。这样的系统是可靠的，误报率低（误报是一种分类错误，当未发生攻击时发出警报；漏报则正好相反，发生攻击时没有发出警报）。问题是，这样的系统并不能够检测未知攻击[9]。

第二种方法是异常检测。异常检测方法了解网络或计算机系统中的正常行为，并上报异常行为。有两个不同的方法用来了解正常行为。第一个被称为监控学习方法。这些方法使用标记的数据库来了解什么是正常行为、什么有可能是攻击行为。这些方法被认为是比较成功的，没有太多误报。第二个是无监控学习算法。这些方法使用无标记数据来发现异常。这些方法会产生大量的误报[9]。

第三种方法结合了特征和异常检测：特征检测用于确保已知攻击的检测；异常检测则用来检测未知攻击[9]。

2.3.1 文献中的异常检测方法

Tavallaee 等人的异常检测方法的研究表明，分类方法是最常用的异常检测方法。最常用的分类算法是神经网络、隐马尔可夫模型、支持向量机和贝叶斯网络。其他方法包括基于统计的方法、聚类方法等[10]。在研究中，异常检测方法的成功率往往在 95% 以上 [11] [12]。这一成功主要是因为分类方法（以及监控学习）用于众所周知的 1999 年创建的 DARPA 99 数据库。这个数据库提供了标记的攻击数据库和非攻击数据库。其结果是：可以采用更准确的监控学习算法进行异

常检测，而无需创建昂贵的标记数据库。通过 DARPA 数据库测试的结果是有争议的，因为 DARPA 数据库被认为是过时的，将其作为基准并不可靠[13]。其他基于统计的方法（如频率时间序列数据或聚类方法）不太受欢迎，但它们可以用于无监控学习。文献中流行的聚类方法是共享近邻、K-均值和自组织映射。

将机器学习算法应用于异常检测的最大挑战是数据和数据特征的选择。数据类型的选择（例如选择 IP 数据包而非流数据）决定着是否可以检测到攻击。数据特征的选择更加重要，例如地址、协议、持续时间等。大量数据特征会减慢分析，而过少的数据特征则无法检测攻击[14]。

3.攻击分析框架

第 2 章介绍了分类法，该方法能够创建用于入侵检测系统开发的检查表。分类法能够涵盖 APT 中所使用的零日漏洞等先进手段，但这通常不能向基于特征的检测系统（分类法支持该系统）提供足够的信息。因此，我们需要一个新的分析框架，以更好地洞察 APT 的结构和检测。APT 的结构和攻击方法可用来确定该框架的结构。APT 的步骤数量和攻击方法提供了可检测的特征和特征的可能检测位置。

框架中包含可能的检测和分析方法，这些与攻击方面和可能的检测位置有关。最后，攻击的业务方面和检测方法被添加到框架中，以确定业务方面对 APT 检测设计的影响。

3.1 分析框架

本文提出的新框架包含 7 列（图 1）。该框架试图给出与检测相关的检测方法。前 3 列包含攻击有关的方面。它们提供了详细的攻击描述。描述提供了检测特征（作为检测相关列的输入）。

第 1 列包含了不同的攻击步骤。此列中的步骤数量决定了框架中行的数量。第 2 列包含每个攻击步骤中使用的低级攻击方法。第 3 列包含第 2 列中攻击方法的特征。这些特征可以用于检测。对于未知方法，例如零日漏洞利用，这些特征可能无法确切地得知。在这种情况下，攻击步骤的目标和攻击方法可以用来识别信标或行为变化。该框架中列的内容应该遵循一定的顺序，以便不同列的信息可以相互关联：攻击特征到攻击方法，攻击特征的位置等。绘制行之间的树状结构可使列之间的关系更为明显。

第 4 列包含特征的可能检测位置。这些位置可能是 DMZ、服务器日志或工作站。位置决定了检测方法和分析方法用于攻击检测的可能性。有些攻击可能有多个检测特征。设计检测系统时，这是很有用的。

第 5 和第 6 列涉及检测方面。第 5 列包含检测方法，包括网络入侵检测、主机入侵检测或日志分析。第 6 列则列出了在检测方法中使用的分析方法。这是智能数据分析算法被放置在框架中的位置。前 5 列的内容决定了分析算法的输入数据。第 6 列列出了可用于检测攻击特征（位于前一系列提出的位置）的方法。

框架中的最后一列包含与攻击和检测方法有关的业务方面。右侧的影响刻度显示可能的影响随着攻击的进展而增加。应尽早地检测到攻击，在攻击后续阶段检测到攻击会减少防御者的可用时间并提高攻击者提取信息的机会。影响可以视为检测的激励手段。业务方面也对检测系统的设计提出了限制。例如，有些检测方法可能涉及到隐私问题；或者系统的成本可能过高。

3.2 应用于 APT

将该框架应用于 APT 攻击首先需要选择攻

击步骤数量。在本文中，我们选择了 8 个步骤。这 8 个步骤描述了不同的活动。该些步骤类似于 Yang 等人提出的 7 个步骤[7] ,以及 Tankard [2] 和 GOVCERT 描述的步骤[8]。第 1 步是外部侦察。第 2 步是获得网络访问权限。第 3 步是内部侦察。第 4 步扩大权限，例如获取管理员权限。这一步可以与第 3 步同时进行。下一步是收集网络中的某个位置的信息，并准备提取信息。第 6 步将收集到的信息发送到网络以外的位置，这是一个单独的步骤，因为它具有明显不同的目标和

更大的影响。第 7 和第 8 步涉及控制和执行攻击，以防止检测。最后两个步骤是活跃在整个攻击过程中。

其他列的内容基于前 3 列的攻击分析。例如：包含链接的电子邮件，链接指向包含恶意软件的网站，这样的邮件可以用来获得对网络的访问。可以主动扫描电子邮件，看看他们是否包含链接。可以在网络中的不同位置对电子邮件进行扫描：工作站、邮件代理服务器或网络流量。不同位置可以采取不同的检测方法。

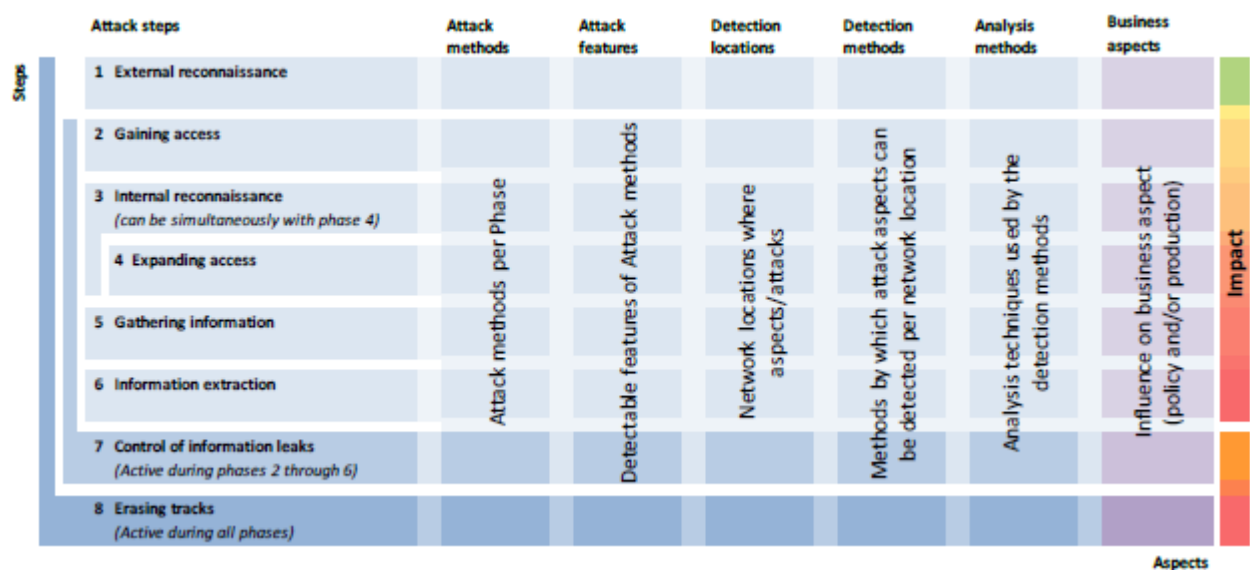


图 1：框架的概览

反过来，这些方法又可以使用不同的分析方法。其结果是，该框架为 APT 检测系统的设计提供了选择。

4.APT 检测设计

如前文所述，多步骤攻击（如 APT）的每一步可以单独分析，每一步可以采用不同的攻击方法。APT 与多步骤攻击的不同之处在于：APT 经常利用未知漏洞，而且目标明确。APT 攻击者努力做到不被察觉。其结果是，普通的检测方法可能无法检测到 APT。本章提出了采用智能数据分析的入侵检测系统来检测 APT。

4.1 框架作为设计的路线图

第 3 章提出的框架有助于我们了解需要检测什么、在哪里检测、如何检测、为什么需要检测。这 4 个问题影响着 APT 检测系统的设计。该框架的攻击相关列回答了需要检测什么：APT 攻击的步骤，可使用的方法和可检测的攻击特征。检测位置列回答了在哪里检测攻击相关的特征。攻击特征和检测位置的组合限制了检测方法和分析方法的选择。如何检测受到前两个问题的影响。检测和分析方法列包含如何检测的可能答案。业务方面回答了为什么需要检测。业务方面提出的检测理由也限制着分析和检测方法的选择。

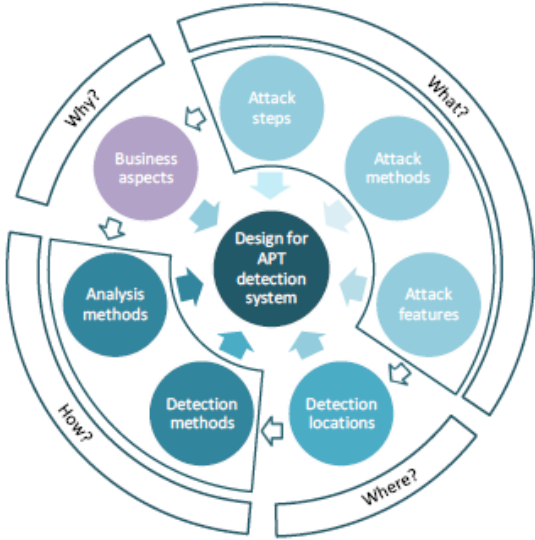


图 2：框架列和设计输入

使用该框架来分析攻击和检测可能性为 APT 检测系统的设计提供了路线图。需要检测什么、在哪里检测、如何检测、为什么需要检测，这四个问题的答案为系统设计提供了输入。这种系统设计方法将攻击分析作为设计选择的驱动，并支持设计选择，同时确保满足业务需求。

4.2 使用路线图进行系统设计

下文将该框架作为路线图来设计 APT 检测系统。图 2 中的 4 个问题被用作设计步骤。

4.2.1 需要检测什么？

第 2.2 节介绍了 APT，这是一种新的高级威胁，现有的防御方法不足以抵御这种威胁。本节中的设计应当能够检测 APT。APT 是多步骤攻击，每一步都有不同的目的并使用不同的攻击方法。区分这些步骤有助于了解攻击的进展。结合事件也有助于识别 APT。APT 可以使用已知的攻击方法，但它们也经常利用未知的零日漏洞来获取访问权限。因此常见的防御措施无法检测到 APT，但可以检测到被成功攻击的客户端和服务器的行为的改变。行为变化指数据库访问频率或互联

网连接的变化。行为变化可以出现在网络中，也可以出现于工作站和服务器的软件中。并非 APT 的所有攻击方法都产生网络流量。例如，工作站中的提权攻击并不一定产生网络流量。

为了识别 APT，我们不应该只关注已知的攻击方法，还要关注行为变化；不仅要监控网络流量，还要监控网络客户端。

4.2.2 在哪里检测？

对成本效益的网络入侵检测系统的研究显示，网络上的多个传感器能够很好地检测变化[15]。对检测复杂攻击（如 APT）的分布式系统的研究显示，多种分析方法和相关性是最成功的检测方法[16]。APT 攻击一步一步深入网络。在不同位置收集数据增加了检测 APT 不同步骤的可能性。检测位置可能是网络上某个位置、工作站或服务器。为了获取网络数据，我们应该在不同的物理网段部署多个探针，以监控流量。探针可以是捕获网络数据包的物理设备，也可以是监控计算机程序行为的软件。

局部行为分析和特征检测能够创建分布式系统，而且无单点故障。它也消除了对高性能系统的（能够处理所有探针收集的所有数据）的需求。局部分析能够对检查到的特征和行为变化发出警报。这些警报需要提交给安全分析人员。但是还需要对这些警报进行分析，以确定 APT 攻击的当前序列。不同分析元件的警报和数据需要进行组合和分析，以检测 APT 的攻击序列。收集数据能够减少网络流量，但也会引发单点故障。中央分析元件的冗余可以减少故障风险，但是增加了系统成本。使用局部分析元件来检测 APT 要求所有局部元件之间的警报共享。这急剧增加了网络流量，由于性能问题可能无法应用于工作站。

另一种方法是让局部分析元件检测攻击序列的一部分（可见部分）。其结果是，序列分析也应用于整个网络中，减少了中央分析元件故障的影响。

4.2.3 为什么要检测？

网络攻击导致的经济损失可能是非常巨大的。攻击的预期经济损失是安全措施投资的主要原因[17]。入侵检测的投资回报取决于系统减少攻击影响的能力。根据 Iheagwara 等人，这种能力取决于系统设计和分析方法的选择[15]。他们的研究表明，部署了涵盖所有物理网段的多个传感器的系统能够提供最好的检测。Zhou 等人的研究表明，使用多个数据分析算法进一步提高了分布式系统中的检测率[16]。系统的有效性（即检测攻击的能力）需要尽可能地高。最好应具有高精度，这意味着系统误报率低。

具有多个算法的分布式设计符合减少攻击影响的要求。另一方面，这种系统的成本可能太高。系统的最大可接受成本可以通过成本/收益算法进行计算[15]。从理论上讲，预期的经济损失是投资的最大金额（如果攻击可以被系统阻止）。其结果是，像 APT 这样导致重大损失的攻击也会导致更多的投资。但这也受到系统准确性和攻击自身影响的限制。

当显示用户个人信息或行为时，捕获和创建行为模式被认为能够侵犯用户隐私。如果这样的详细程度对于检测 APT 是必要的，则系统设计需要处理这些隐私问题。

4.2.4 如何检测？

前面的步骤表明，能够检测 APT 的系统需要检测已知和未知的攻击方法。低级的攻击方法可

以在网络中的不同位置执行。捕获的流量可以用于识别很多攻击方法,但是有的 APT 攻击步骤并不一定产生网络流量。系统的设计也应该注意工作站和服务器中的 APT 的痕迹。分布式系统能够给出最有效、最准确的结果。这意味着,需要用不同的方法分析不同的数据类型。分析方法的效率和准确度必须足够高,以保证对该系统的投资是值得的。

第 2.3 节表明,异常检测仍然遭受着大量的误报,特别是当使用无监控学习算法时。监控学习算法的异常检测更有有效,但是它们需要攻击免费或标记数据库进行训练,才可以检测异常。为每次安装和每个局部元件创建这样的数据库是很难完成的任务。基于通用特征的特征检测已被证明是可靠的,能够检测攻击[9]。利用人工确定的特征作为基准能够确保系统更可靠,而不需要太高的安装成本。大多数特征可以在不同的设备和多个系统中重复使用。

检测未知的攻击方法(流行于 APT 攻击)确实需要异常检测。无监控学习方法不需要创建训练数据库,并且可以增加检测特征。无监控学习法的一个优点是它们确定什么是正常行为,并识别网络行为的变化。这也带来了风险:攻击者可以误导算法,即慢慢地开始攻击让算法习惯于流量的变化[9]。

4.2.4.1 异常检测

已知的攻击可以通过特征检测被识别。也可以通过在特征中描述正常行为来检测行为的改变,但是这需要许多具体特征,使得该方法不具吸引力。异常检测方法可以使用为无监控学习方法描述行为的数据。例如,可以通过聚类算法比较网络客户端的行为。如果探针的输入数据包含

客户端的不同的正常行为,则这种方法会产生误报。例如:行为不同的客户端可能属于不同的部门。了解网络并认真选择探针安装的位置能够防止这类问题。

有效的聚类算法是 K-均值聚类和自组织映射。为了防止误报,可以采用半监控方法。半监控方法使用有限数量的标记事件,而非完全标记的训练数据库。标记事件应确定不同聚类并创建聚类算法的开始[9]。

中央元件中的异常检测更为困难。需要结合通过匹配特征和行为变化而产生的警报,以识别可能的 APT。多步骤攻击(如 APT)中低级攻击方法的大量可能序列使得我们难以确定该攻击的事件序列[16]。大量可能序列的结果是难以定义正常行为。利用聚类算法的无监控学习仍然可以用于识别异常行为序列,但它们会产生大量误报。通过结合不同聚类算法(共享近邻和 k-均值)的结果,可以降低误报率。相比于一种算法认为事件序列异常,两种算法都认为事件序列异常的准确率更高。这种方法被称为 Boosting [9]。

更复杂的方法如 Yang 等人提出的方法[7],利用低级攻击的信息关联事件并创建攻击场景。Yang 等人尝试将警报序列与已知的攻击序列相匹配,并将结果与信息曝光序列相匹配。信息曝光序列是 7 个阶段,从外部侦察到实现内部网络目标。这些阶段与第 2 章中介绍的框架 8 步骤非常相似。Yang 等人表示警报关联法仍处于起步状态,还需要大量的研究。

根据所属步骤标记事件并使用 APT 结构信息,这一方法有助于创建用于异常检测的更好的事件序列。

4.2.4.2 智能数据分析的其他应用

智能数据分析也可用于改善特征检测以及自动创建特征。例如，如果系统中有很多规则，可以为规则应用创建决策树以减少分析时间[18]。另一种选择是规则学习方法。一个例子是模糊的基于规则的异常检测[9]。这种方法使用标记的数据库来创建规则，而规则定义了正常和异常行为的聚类。标记的数据库可以来自于异常检测模块收集的数据。根据警报手动标记数据可以提高该数据库的准确度。这种方法可以提高系统局部分析元件的精度。

5. 框架作为路线图的反思

将该框架作为系统设计的路线图必须能够获得检测 APT 的系统。本章反思了系统设计的能力。

创建分布式系统使我们能够分析来自不同数据源的数据。分析元件中同时使用签名和异常检测提高了检测攻击特征的可能性。需要异常检测来检测未知的攻击方法。然而，零日漏洞攻击的一般特征是可取的，因为这些规则更清楚地表明了为什么要上报异常。这也使系统更加可靠，因为所选择的无监控学习算法仍具有相对较高的误报率。

序列分析对中央分析元件检测 APT 至关重要。另一方面，研究表明，当前对多步骤攻击的序列分析远不够精确[5] [7]。该系统设计能够识别 APT 攻击的各个步骤。但该框架并没有说明如何将低级攻击与高级攻击联系起来。如果在一小段时间内在同一个位置检测到不同的步骤，则将攻击步骤联系起来是比较容易的。

该系统可在局部分析层面做到这一点。但是

在中央分析层面更难以做到这一点。低级攻击的智能过滤可以减少数据库，有助于提高检测结果。

所有警报，从局部分析元件到中央分析元件，都上报给专家进行分析。这些专家可以采取适当的措施。这些专家做出的决策可以作为该系统的输入，以提高系统的精度。这种方法可以用来实现半监控学习。

对专家的依赖也要求高效的界面设计。更智能的提交警报和数据的方法能够提高系统的效率。

6. 结论

APT 是一种新的、持续性和针对性攻击，时多步骤攻击的一种。APT 攻击为当前的防御方法带来了挑战，因为当前的防御措施主要依赖于已知特征，而 APT 利用大量的未知安全漏洞。本文提出的框架用来分析攻击，将低级攻击方法与检测方法和智能数据分析方法结合起来。

该框架被用作设计检测 APT 系统的路线图。这一系统根据对 APT 的分析选择适当的检测方法。其结果是，业务方面以及攻击相关方面指向分布式系统的设计和多种分析算法的使用。特征检测用来更精确地检测已知攻击。异常检测对于检测未知攻击是必要，因为特征检测无法检测到未知攻击。异常检测的问题是具有较高的误报率。异常检测对高级攻击的预计检测错误率甚至更高。

即使具有较高的误报率，异常检测仍然是必要的。提高精度的方法（如 boosting）可以减少误报的数量。但对警报进行人工分析仍有必要。

本文提出的框架有助于分析攻击，以确定检

测需要什么攻击方法。

6.1 研究建议

用于分析的特征决定着异常检测算法能否检测到攻击。因此，数据的预处理也是检测中最重要的一步。因此，选取好的特征有助于改善异常检测。

本文的设计方法仍然需要专家对警报进行分析。创造更好的用户环境需要专家进行更多的研究。需要什么样的信息以及什么时候需要这些信息？这样的问题应得以解决，以便创建自适应的用户界面。

最后，在入侵检测中我们需要新的数据库，以便获得更多的相关信息，并提高算法的成功率。攻击是不断变化的，特别是 APT，这使得我们难以创建代表性的数据库。另一方面，DARPA 数据库已经创建了 10 多年，不能代表现今的攻击。

参考书目

- [1] N. Kshetri, The global cybercrime industry: economic, institutional and strategic perspectives, Springer, 2010.
- [2] C. Tankard, "Persistent threats and how to monitor and deter them," Network security, vol. 2011, no. 8, pp. 16-19, 2011.
- [3] Symantec, "Symantec Internet Security Threat Report," Symantec, 2011.
- [4] V. Igure and R. Williams, "Taxonomies of Attacks and Vulnerabilities in Computer Systems," IEEE Communications Surveys & Tutorials, vol. 10, no. 1, pp. 6-19, 2008.
- [5] P. Ning, Y. Cui and D. Reeves, "Constructing attack scenarios through correlation of intrusion alerts," in Proceedings of the 9th ACM conference on Computer and communications security, New York, 2002.
- [6] S. Cheung, U. Lindqvist and M. Fong, "Modeling Multistep Cyber Attacks for Scenario Recognition," in Proceedings of the DARPA Information Survivability Conference and Exposition, Washington, 2003.
- [7] S. Yang, A. Stotz, J. Holsopple, M. Sudit and M. Kuhl, "High level information fusion for tracking and projection of multistage cyber attacks," Information Fusion, vol. 10, pp. 107-121, 2009.
- [8] GOVCERT.NL, "Nationaal Trendrapport Cybercrime en Digitale Veiligheid 2010," GOVCERT.NL, The Hague, 2011.
- [9] S. Dua and X. Du, Data mining and machine learning in cybersecurity, Taylor & Francis Group, 2011.
- [10] M. Tavallaee, N. Stakhanova and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, vol. 40, pp. 516-524, 2010.
- [11] S. Mukkamala and A. Sung., "A Comparative Study of Techniques for Intrusion Detection," in 15th IEEE International Conference on Tools with Artificial Intelligence, Sacramento, 2003.
- [12] S. Wu and W. Banzhaf, "The use of computational intelligence in intrusion detection systems: A review," Applied Soft Computing, vol. 10, pp. 1-35, 2010.
- [13] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez and E. Vazquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," Computers & Security, vol. 28, pp. 18-28, 2009.
- [14] J. Davis and A. Clarck, "Data preprocessing for anomaly based network intrusion detection: A review," Computers & Security, vol. 30, pp. 353- 375, 2011.
- [15] C. Iheagwara, A. Blyth, T. Kevin and D. Kinn, "Cost effective management frameworks: the impact of IDS deployment technique on threat mitigation," Information and Software Technology, vol. 46, pp. 651-664, 2004.
- [16] C. Zhou, C. Leckie and S. Karunasekera, "A survey of coordinated attacks an collaborative intrusion detection," Computers & Security, vol. 29, pp. 124- 140, 2010.
- [17] T. Rakes, J. Deane and L. Rees, "IT security planning under uncertainty for high-impact events," Omega, vol. 40, pp. 79-88, 2012.
- [18] C. Kruegel and T. Toth, "Using Decision Trees to Improve Signature-Based Intrusion Detection," in RAID 2003, Pittsburgh, 2003.