基于蜜罐的 APT 攻击防御方法

邱焕然 唐宾徽

四川大学锦城学院 计算机与软件学院 四川 成都 611731

【摘 要】：APT 攻击是近年来出现的一种危害性极高的网络攻击形式。传统的被动防御技术早已不能用于抵御 APT 攻击。针对现有的问题，提出了一种基于蜜罐和改进的 K-Means 算法的 APT 攻击防御方法。首先介绍了相关的技术，然后，针对传统的 k-means 算法的不足进行了改进，分析了改进的思路；最后，给出了利用蜜罐和改进的 K-Means 算法的防御方法。

【关键词】：蜜罐；APT 攻击；K-Means

引言

自从谷歌极光事件曝光后，APT 攻击便开始进入公众视野中。近年来，APT 攻击次数显著上升，APT 攻击的手法也在不断变化，对各国家各行业的威胁也在不断增加。

APT 攻击的危害性如此之高，如何有效的防御 APT 攻击是急需解决的问题。传统的防御技术，例如防火墙，是通过包过滤的方式进行访问控制，而入侵检测系统是基于自身的规则进行检测，而且多数依赖于专家系统，存在误报率高等问题。两种方法都不能有效的进行 APT 防御。本文给出了一种基于蜜罐和改进的 K-Means 算法的 APT 攻击防御方法,分析攻击者的攻击流程，以此提高防火墙和入侵检测系统的防御能力。

1 相关技术研究

1.1 APT 攻击

1.1.1 APT 简介

APT 是指将 0day 漏洞、渔叉攻击、网络钓鱼、社会工程学等多种高级攻击手法相结合，进行的危害性高、持续时间长，隐蔽好的网络攻击。其目标多是军工，政府，能源等重要行业，目的在于盗取目标系统的机密资料，破坏目标系统的正常活动。

1.1.2 攻击步骤

APT 攻击的手法多样，总体来说，APT 攻击可以分为情报收集、载荷投递、C&C 通信、横向移动、回传数据五个步骤，如图一所示[1]。

图一

1、情报收集

APT 攻击者在发动攻击之前会收集目标系统的信息，并且在入侵过程中持续收集更多的敏感信息，动态调整入侵方案，以便于最终的顺利入侵。这一阶段常使用的技术有：开源情报，主机扫描、端口扫描、网络嗅探、社会工程学、网络钓鱼等技术。

2、载荷投递

攻击者会利用第一阶段收集到的信息，编写特定的恶意程序，并且制作完美的诱饵。利用水坑攻击、鱼叉攻击、网络劫持、物理设备等方法，将恶意程序传送到目标系统。

3、C&C 通信

攻击者进入内网后，会在目标主机上安装恶意程序，恶意程序会与 C&C 服务器通信，为了防止不被安防系统检测到，攻击者会将与 C&C 服务器通信过程封装在 DNS、SSL、SSH 等隧道技术中，以此来规避检测。然后，攻击者会使用 C&C 服务器向木马发出各种命令，继续收集内网的敏感信息，为下一步的横向移动做好准备。

4、横向移动

攻击者会以被控制的设备为跳板，进而对内网中的其他设备发起攻击,同时结合 SQL 注入和缓冲区溢出等提权手段，不断提高攻击者在内网中的权限，最终获得对核心主机的控制，拿下整个内网。

5、回传数据

攻击者在收集到敏感数据后，首先会将其临时存储在内网的合法服务器上，然后对数据进行压缩、加密、使用加密通道或模拟正常的通信服务过程向 C&C 服务器回传数据，同时会清除自己在内网中留有的痕迹，并长期潜伏内网中。

1.2 蜜罐

1.2.1、蜜罐简介

蜜罐是一种主动防御技术，可用于安全的不同方面，如预防、检测和信息收集，常用来网络取证和入侵检测，蜜罐的价值在于被检测、攻击或被破坏[2]。

1.2.2 蜜罐的分类

1.低交互蜜罐

低交互蜜罐和攻击者的交互很有限，没有部署真实的操作系统。这就使得蜜罐被攻破的可能性非常低，捕获到的数据也非常有限，很难用来对攻击者进行分析。

2.中交互式蜜罐

中交互蜜罐的结构比低交互蜜罐复杂一些。但是，中交互蜜罐不提供操作系统，会模拟出各项服务。攻击者发现安全漏洞的机会增加了，但系统仍然不太可能受到威胁。攻击者可以与它进行足够的交互。因此，中交互蜜罐可以分析更复杂的攻击[2]。

3.高交互蜜罐

高交互蜜罐提供了一个真实的操作系统来与攻击者交互，所有的服务都是真实的。因此，使用这种蜜罐能收集到大量信息，可以记录和分析所有操作。由于攻击者有更多的资源可以支配，因此应该不断监控高度交互的蜜罐，确保它不会成为攻击者访问内部网的绝佳跳板[2]。

4、蜜网

蜜网是蜜罐的衍生形式。蜜网项目的创始人 L.Spitzner 认为蜜网本质上是收集攻击者的信息。它通过使用真实的系统和应用程序创建一个多漏洞的系统网络来做到这一点。它的目的是允许攻击者侵入蜜网的内部系统，并在他们不知情的情况下捕获和控制他们的一举一动[3]。

5、蜜场

蜜场是一种分布式的蜜网，通过在不同的子网中安装重定向器，将攻击流量转向一个集中的蜜罐进行监控。这种形式能降低安全风险，同时维护也较为容易。

2 K-Means 算法介绍

2.1 传统 K-Means 算法思想

对于一个数据集,随机选取 K 个点,作为聚类中心,计算样本中每个点与聚类中心的距离,将每个点分配到与其距离最近的簇中,对新的簇重新计算聚类中心,直到聚类中心不变化或者达到最大迭代次数。

2.2 传统 K-Means 算法优点与缺点

2.2.1 优点

K-Means 算法可以根据少量的已知的聚类样本,确定样本的分类,算法实现较为简单。

2.2.2 缺点

聚类中心的数量 K 值是事先给定的，事先是不知道分为几个类是合适的。而且，算法需要不断计算聚类中心,在大数据的情况下,特别是网络中的巨大数据流量，算法的时间开销会很大，需要的运行时间会特别多。

2.3 K-Means 算法改进

2.3.1 针对运行时间的改进思路

由于网络中的数据流量是极大，特别是当蜜罐受到 DDOS 攻击的时候，其流量可以达到数十 G，甚至几百 G，蜜罐捕获到的流量数据将会非常庞大。因此，对于经过数据处理之后得到的流量数据集，首先选取一部分数据集，利用传统的 K-Means 算法进行聚类，然后再选取剩下的数据集的一部分，将其与到前一部分的聚类结果合并，再次计算聚类中心，然后不断迭代以上步骤，直到聚类中心不变或者达到迭代次数。算法执行结束[4]。

2.3.2 针对 k 值选取的改进思路

由于 k 值对聚类结果影响很大，而且蜜罐中的数据流量非常之多。首先暂时依靠专家系统手动指定聚类中心，然后计算样本中各维度的最大值和最小值，找出离样本点最近的聚类中心，并将样本点将聚类中心移动，在移动聚类中心的时候，采用随机梯度下降算法，避免陷入局部最优，最后不断迭代计算新的聚类中心，直到聚类中心不变或者达到最大迭代次数[4]。

2.3.3 运行步骤

结合两个问题的改进思路，改进后的总体算法运行步骤如下：

1、对于处理好的数据集，首先选取部分数据集，对于这一部分数据集,暂时依靠专家系统人为给定 K 个聚类中心。

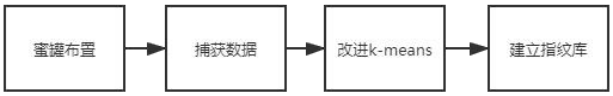
2、在选出的部分数据集中，运用改进后的 K-Means 算法，计算出聚类中心。

3、选取剩余的数据集的一部分，并将其和第一步得到的结果合并，运用改进的 K-means 算法，重新计算聚类中心。

4、再次迭代 2，3 步，直到聚类中心不变或者达到最大

迭代次数[4]。

3 APT 攻击的防御方法设计



图二

3.1 蜜罐布置

首先防御方要根据自身的网络架构和自身需求部署选择的蜜罐形式,并且要保证能给攻击者提供足够多的交互,并且要在蜜罐中放置虚假的数据等待攻击者盗取，防止攻击者识破蜜罐。并且最重要的是，要在蜜罐的出口上要时刻密切监控流量,以确保蜜罐不会成为攻击者进入内网的跳板。

3.2 捕获数据

从攻击者的开始扫描蜜罐,到攻击完成的每一步都应该捕获完整数据。而且，蜜罐捕获到的数据不应该存储在蜜罐中，应采用其他方法存储,一旦攻击者识破蜜罐，捕获到的数据将会被攻击者销毁。

3.3 改进 K-Means

在捕获到数据之后,首先对数据进行处理。然后运用改进的 K-Means 算法进行聚类,找出每个攻击特征的聚类中心。

3.4 建立指纹库

在找到一个攻击类型的聚类中心后,防御方应该根据自身的情况,改善防火墙和入侵检测系统的规则,分析攻击者的攻击流程，加强相应基础设施的安全防御，提升整体的防御能力。

4 预想结果

防御方根据自身的需求和实际网络状况，选取合适的蜜罐形式，能捕获到攻击者的足够的数据。运用给出的方法得到每种攻击特征的聚类中心。

相比传统的 K-Means 算法，本文提出的 K-Means 算法，首先会在针对庞大的数据集的处理上，运行时间会减小，能提高算法的运行效率。采用随机梯度下降算法，能较好的在不断迭代的过程中，确定最佳的 K 值，对聚类结果会有较好的帮助。最后，运用聚类得到的结果，分析攻击流程，改进防火墙和入侵检测系统的检测规则，针对性加强防御能力。

结语

APT 攻击是随着网络的不断发展而出现的一种新型的高危害性攻击形式,传统的网络防御形式早已不能用来抵抗 APT 攻击。蜜罐作为一种诱导攻击的主动防御形式,对于 APT 攻击具备独特的优势。可以用来捕获攻击者的信息,对攻击者能很有效地进行研究,有针对性地加强安防建设。APT 攻击仍然在不断发展,攻击方法也在不断演变,如何对 APT 进行实时监控,更为主动的去防御 APT 攻击,仍需要大量研究。

参考文献：

[1] 贺诗洁,黄文培.APT 攻击详解与检测技术[J].计算机应用,2018,38(S2):170-173+182.

[2] Mokube I, Adams M. Honeypots: concepts, approaches, and challenges[C]//Proceedings of the 45th annual southeast regional conference. 2007: 321-326.

[3] L. Spitzner, "The Honeynet Project: trapping the hackers," in IEEE Security & Privacy, vol. 1, no. 2, pp. 15-23, March-April 2003, doi:10.1109/MSECP.2003.1193207.

[4] 张晓峰. 基于聚类和支持向量机的入侵检测方法研究[D].兰州理工大学,2018.