基于DFI和DPI技术的异常流量监控

绿盟科技　王卫东

摘要：本文从描述异常流量监控的应用场景入手，详细解释了基于DFI和DPI的两种异常流量检测手段，并进行了对比分析。然后提出了一种全新的异常流量控制方案的技术设想。

关键词：DFI； DPI；异常流量；流量监控；应用层访问控制策略

0 前言

随着企业各种应用系统的建设和完善，SNMP采集的信息无法描绘流量的细部特征，比如某种应用占到总带宽多少？某个IP地址的网络用量是多少？运维人员要想知道这类问题的答案，需要通过采集实际流量的方式进行统计。面对近年来网络攻击日趋频繁的趋势，只统计正常应用流量的比例也无法满足运维需求了。对异常流量进行分析监控成为流量分析系统主要功能。

1 基于DFI(Deep Flow Inspect)的异常流量检测

1.1 攻击与蠕虫传播的流量检测

基于DFI的检测技术一个主要优势就是可以高效准确的检测出网络攻击和蠕虫传播。在这里列出这个标题，是为了论述体系的完整性。因为DFI技术在这方面的应用涉及的问题比较多，需要单独成文论述，所以关于这个方面的检测原理详见本刊前一期中《电信IP网络异常流量及其检测》。

1.2 P2P流量检测

由于流数据(Netflow或sFlow)是经过汇聚的且通常都是抽样产生的，数据又仅包含IP层信息而没有应用层信息，因此很多人对基于DFI的P2P检测技术抱有一定成见，认为它不一定能很准确的检测到P2P流量。然而实际情况却是出乎这些人的意料，DFI技术不仅可以检测P2P流量，而且检测的准确度还相当高。这是因为P2P流量与其它网络应用有鲜明的区别，针对这些特征进行综合检测，便可以准确的检测出P2P流量。

1.2.1 P2P流量的统计特征

（1）流量大，符合“2/8原则”：P2P流量一般都会远远大于其它应用类型的流量，统计结果通常会出现20%的IP地址所相关的流量占到全部流量的80%，即符合所谓的“2/8原则”。

（2）并发端口数量：在一个终端上运行P2P应用程序之前，用netstat 命令检查网络状态，可以看到同时打开的端口一般在10～15个之间，如果启动P2P应用程序以后，再次检查网络状态，可以看到打开的端口数量一下激增到100多个。

（3）端口变化率：由于很多P2P应用程序为了逃避流量控制，会使用端口跳变技术，动态的变更通讯端口，因此造成端口变化率长时间保持很高的数值。

（4）拓扑特征值：由于P2P下载的端点都会用一些缺省的端口与其它端点通讯，通过分析流记录可以找到这些被高度疑似P2P端点间的拓扑关系，并使用一个人工定义的拓扑特征值来衡量这些拓扑关系。当特征值达到一定水平，即可确认该主机为P2P端点。

（5）封包字节数大：为了提高传输效率，P2P流量的封包字节数都会很大，除了基于P2P的IP语音包，一般P2P下载的数据包至少都在1200字节左右，这是与其它应用另一个明显的差异。

1.2.2 P2P流量的行为模式特征

（1）大量空闲连接：P2P端点通常都会有很多空闲连接，在流记录上就表现为很多流量非常少的记录。

（2）UDP/TCP 并存：有些特殊的应用，如DNS，NETBIOS，IRC，游戏和多媒体业务流量等，这些应用都有特定的端口,如135,137,139,445,53,3531等,可以通过端口匹配识别这些流量。

（3）同时充当客户端和服务器(角色分析）：通常服务器的通讯模式是接收资源请求信息，然后提供相应的数据资源。而数据资源的流量大小一般都远远大于请求信息的流量。因此，如果一个主机输送出的数据远远大于接受到的数据，我们就可以判断这个主机的角色是“服务器”。反之，则是“客户端”。P2P端点接收和发送的流量几乎大小相当，因此可以看作是同时充当“客户端”和“服务器”。

1.2.3 P2P流量检测效果释疑

基于DFI的P2P检测，不像基于DPI检测那样对P2P流量进行更细致的分类，甚至可以按照不同的P2P客户端软件进行分类。这看上去似乎是DFI技术的一个缺陷，而实际上并非如此。原因是有些P2P客户端软件，虽然名称不同，但使用的P2P协议是相同的，或者软件的核心代码是相同的。所以按照不同的软件对P2P协议进行分类没有太大意义。另外，检测P2P流量目的是控制这些流量，而对流量更详细的分类，无助于灵活准确的控制。

1.3 异常特征的自动提取

基于DFI的检测技术，还可以用于提取类型未知的异常流量的特征。在实际流量检测过程中，可能会遇到突发流量激增的情况，但是现有的检测算法又无法确认异常类型。这种情况下可以使用“异常特征提取技术”，将流量特征提取出

来。大致的步骤如下：

（1） 确定异常流量发生的位置(物理端口、IP地址、AS号)。

（2）聚合维度的选取：聚合之前，首先要确定聚合依据哪些字段，也就是聚合维度的选取。一般可供选取的维度包括：源端口、目的端口、协议、TOS、TCP-Flag。将每个聚合结果的总流量或总包数求和。

（3） 聚合结果的流量大小排序呈现(按包数或按字节数)。

（4）把聚合结果的特征导出。

2 基于DPI的异常流量检测

2.1 应用层攻击检测

由于应用层攻击具有代价小(带宽占用和攻击主机性能消耗小)、隐蔽性好、防御难度大三个特点，它已经演变为网络攻击的一个主要形式。我们这里讲的应用层攻击，是指完全模仿应用层访问行为的攻击。例如CC(Http Get Flooding)攻击，假人攻击、DNS Request Flooding等。应用层攻击很容易与两个概念混淆，一个是借助应用层手段发起的网络层攻击，例如DNS反射攻击。从被攻击者和防护方式的角度来看，后者仍然是流量型的网络攻击。另一个是网络入侵。入侵行为虽然看上去也是流量很小的破坏行为，但是入侵主要是利用系统的漏洞，以获取系统的控制权并窃取数据为目的，很少会造成服务中断和性能下降。因为那样势必会暴露入侵行为，而入侵者都希望入侵行为越隐蔽越好。

应用层攻击的概念清晰了，下面我们看看如何利用DPI技术检测应用层攻击。因为应用层攻击都是模仿正常的访问行为，审视单个的访问行为，往往无法判定是否为攻击。所以对于攻击的检测，需要对大量数据进行统计分析。DPI技术的优势是可以对应用层的信息进行分析，通常可供分析的内容有：

(1) 特征字段的统计分析：应用层协议消息（SIP、HTTP协议中的消息)、域名或URL。

（2）行为统计：登录数量增率、连接请求增率、连接请求的时间间隔。

2.2 非法业务的识别

非法业务是指影响运营商业务收入或降低网络收益的那些行为，通常包括：P2P流量、非法VoIP、宽带私接等。

（1）特征字段的统计分析：应用层协议消息、MAC地址、端口信息、协议类型、数据包校验和、操作系统和应用程序(浏览器、MSN等)版本及特征字符串信息、IP包的扩展属性。

（2）行为统计：端口增长率、流量增长率。

3 异常流量的控制手段

3.1 流量清洗

流量清洗是防护网络层攻击的最有效手段，其工作原理是将流向被攻击目标的全部流量牵引到智能清洗设备上，滤除攻击流量后，将净化后的流量回送到网络中，从而保证正常的访问。

3.2 行为验证

应用层攻击的仿真程度越来越高，因此很难检测和清洗。但是可以通过自动推送手工验证页面的方式，有效阻止这类攻击。比如某个Web服务器受到攻击的时候，先将访问请求牵引到防护设备，然后由防护设备强行推送一个认证页面，要求访问者按照验证码图片的内容输入认证码。攻击程序不会自动完成认证过程，攻击行为也就不会继续发挥作用。

3.3 访问控制策略

在路由器上，可以配置一些访问控制列表，作为临时应急的防护措施。在攻击防护设备上，还可以用更丰富的参数，如：源/目的IP地址、源/目的端口、协议类型、包大小、TCP-flag、TOS、应用层协议类型、应用层的特征字串等，组成应用层访问控制策略。这种访问控制策略使用了应用层信息作为参数，使得控制策略更加精细。当流量牵引到防护设备上时，访问控制策略可以发挥作用。

3.4 黑洞路由

黑洞路由的原理与流量清洗的原理类似，也是通过向路由器发送一条路由，只是这个路由没有下一跳地址，路由器将这些流量直接丢弃。这种方法虽然简单有效，迅速缓解，但也会阻止正常访问，用户体验很差。

3.5 干扰流量

干扰流量一般用于抑制非法业务，常用的手法包括：

（1）增加干扰包，降低通话质量。

（2）通过发送伪造的信令消息结束通话。

（3）通过发送伪装包拆断TCP会话，来抑制P2P流量。

4 技术实现及部署方案

4.1 部署图

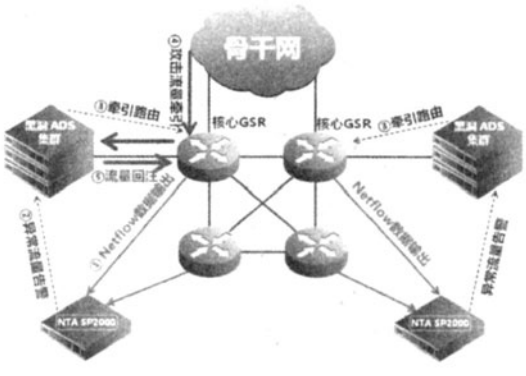

图1 部署图

4.2 工作流程

与攻击防护的工作流程一样，首先由NTA设备发现异常流量并产生告警，通知黑洞ADS集群。ADS集群向路由器发

[下转11页]

(4) TPM计算$s_v := r_v + c \cdot v$，$s_{f_0} := r_{f_0} + c \cdot f_0$，$s_{f_1} := r_{f_1} + c \cdot f_1$。主机计算$s_{e'} := r_{e'} + c \cdot (e - 2^{l-1})$，$s_{ee'} := r_{ee'} + c \cdot e'^2$，$s_w := r_w + c \cdot w$，$s_{ew} := r_{ew} + c \cdot w \cdot e$，$s_r := r_r + c \cdot r$，$s_{er} := r_{er} + c \cdot e \cdot r$主机输出签名$\sigma := (\zeta, (T_1, T_2), N_V, c, n_t, (s_v, s_{f_0}, s_{f_1}, s_{e'}, s_{ee'}, s_w, s_{ew}, s_r, s_{er}))$。

3.4 验证DAA签名流程

(1) 验证者根据签名σ计算$\hat{T}_1 := Z^{-c} T_1^{s_{e'} + c \cdot 2^{l-1}} R_0^{s_{f_0}} R_1^{s_{f_1}} S^{s_v} h^{-s_{ee'}} \bmod n$，$\hat{T}_2 := T_2^{-c} g^{s_{e'} + c \cdot 2^{l-1}} g^{s_{ee'}} \bmod n$，$\hat{T}_2' := T_2^{-(s_e - c \cdot 2^{l-1})} g^{s_{e'}} h^{s_{ee'}} g^{s_{er}} \bmod n$，$\hat{N}_v := N_v^{-c} \zeta^{s_{e'} + s_{f_0} 2^{l'}} \bmod \Gamma$。

(2) 验证者判断TPM根据签名信息m计算得到的HASH值c是否等于如下HASH值$H(H(H((n \| g \| g' \| h \| R_0 \| S \| Z \| \gamma \| \Gamma \| \rho) \| \zeta \| (T_1 \| T_2) \| N_v \| \hat{T}_1 \| \hat{T}_2 \| \hat{T}_3) \| \hat{N}_v) \| n_t)$，如相等则表明签名正确。

4 基于可信计算平台DAA协议的Web服务认证机制

基于可信计算平台的Web服务认证机制不仅认证服务提供者的用户身份证书还需要确认其软硬件配置环境的正确性。其认证流程如下：

(1) 首先服务请求者需要向服务提供者注册计算机正确启动后的完整性状态值(PCR寄存器值)。

(2) 服务请求者首先通过join协议向DAA证书发布者出示自己的假名及使用私有密钥加密的证据。如join过程成功则发布者向该服务请求者签发DAA证书DAACert。

(3) 服务请求者通过sign协议即根据、DAACert对该用户的AIK公钥信息进行签名，并使用该用户AIK私钥对计算机当前的PCR寄存器值进行签名。服务请求者向服务提供者出示假名，并将其用户AIK公钥信息，对AIK公钥信息的签名值、PCR寄存器值，对PCR寄存器的签名值发送至服务提供者。

(4) 服务提供者对服务请求者的验证过程分为以下四步：

第一步：验证假名的正确性。

第二步：然后验证DAA证书对AIK公钥信息签名的正确性。

第三步：验证AIK对验证PCR寄存器值的正确性。

第四步：验证服务请求者平台配置信息即PCR寄存器值。

5 结束语

可信计算平台的目标是要在现有安全措施，如数字证书、VPN、PKI、生物识别的基础上建立一个平台的可信根，它能惟一标识一个平台。可信平台能够将用户身份证明与平台信息紧密绑定，在确认用户身份的同时也确认了平台环境状态的完整性与正确性，为Web服务身份认证提供了可信的身份凭证。

基于可信计算平台的DAA认证机制的一个重要思想是不直接向验证者提供证书，而是向验证者提供拥有该证书后加了密的证据。DAA证书的只需签发一次，可由TPM制造商或平台买主签发，从而解决了传统Web服务身份认证过程中认证中心的瓶颈问题。

参考文献

[1]TCG Infrastructure Working Group. Use Cases Summary. Draft. Version 0.1[EB /OL].https://www.Trusted2 computing group.org.

[2]IBM Corporation Web Services Conceptual Architecture (WSCA1.0).May 2001.httpLs//www-900.ibm.com/developer WorksLcn/webservices/ws-wsca.

[3]Ernie Brickell, Jan Camenisch, Liqun Chen. Direct Anonymous Attestation[C] .ACM press.2004.

[4]羿苗 基于代理的Web服务认证机制研究[D] 华中科技大学 .2006.5.

[5]徐娜 基于可信赖计算平台的可信执行环境研究与实现[D] .中国科学院研究生院 .2006.

[上接 5页]

送牵引路由，路由器根据牵引路由把与告警相关的流量牵引到ADS集群上。ADS集群上的应用层访问控制策略会对异常流量和非法业务流量进行拦截，对正常合法的流量予以放行，送回网络中。

这个解决方案表面上看与以往的流量清洗的解决方案雷同，关键的区别在于系统的策略配置上，大量使用了应用层访问控制策略，可以有效的控制非法业务流量和应用层攻击。这种旁路的流量控制系统，较串联方式更加可靠，更加适合运营商的网络环境和运维要求。

4.3 防护设备的新应用

在前面的介绍中看出，可以把攻击防护设备变成一个流量控制设备。从而拓宽防护设备的应用场景。这种实现方式，有效地回避了用户对串联设备单点故障的担心，同时也增加了设备的复用性，提高了设备的性价比。如果把被攻击时候看作是“战时状态”，而无攻击的时候看成“和平时期”，那么这是一种相当完美的“平战结合”解决方案。

5 展望

随着互联网技术发展，网络应用呈现出丰富多彩的变化。层出不穷的新业务，以及花样翻新的攻击手法引发了对业务识别和异常检测的迫切需求。目前主要的流技术的是1996年的研发成果，流数据仅仅包含网络层和传输层信息，已经无法满足流量分析的需要。

按照IPFIX工作组的计划进度，预计在未来两年IPFIX将成为新的流量信息数据标准。新的流信息数据将提供包括应用层信息在内的更丰富的内容，依据这些数据得到分析结果将更加准确。同时也为流量控制提供了更加精准的控制策略参数。