网康科技 NETENTSEC

应用层网络设备性能

白皮书

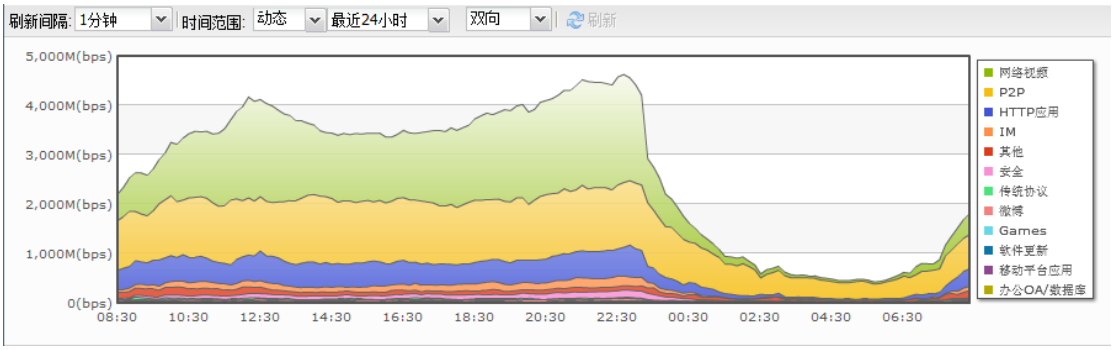
如何评估应用层网络设备性能

北京网康科技有限公司

目 录

1. 传统网络层性能指标不适合应用层

用户对网络业务的可视性、可管理性要求越来越高，要求能从业务视角理解网络流量，并针对应用制定管理策略。因此近十年来，各种创新的应用层网络产品雨后春笋般涌现出来，比如上网行为管理、智能流量管理、Web 应用防火墙（WAF）等，下图为某运营网络的流量走势图，通过以网络应用而非传统的 TCP/UDP 对流量分类，可以更直观地看到流量的构成与分布。另外，传统的网络层安全产品也在向应用层安全设备演进，如传统防火墙、IPS，由于缺乏应用识别与控制能力，正在被面向应用层的下一代防火墙（NGFW）、下一代 IPS 产品取代。



某运营网络流量走势

性能是网络设备选型必须考虑的因素，目前通用的性能指标和评估方法都是基于网络层的，不适合应用层的特点。由于缺乏相应标准，业内仍普遍以网络层性能参数来衡量应用层性能，这既不利于用户选型使用，也不利于产品规划发展。因此很有必要讨论适合应用层特点的网络设备性能指标与评估方法。

2. 网络层性能指标回顾

2.1 RFC 对网络设备性能的定义

传统网络设备性能标准主要来源于 3 个 RFC 文档：RFC1242，RFC2544 和 RFC3511。

其中，RFC1242 定义了网络性能基准测试及测试结果用到的基本术语，最重要的 4 个是：吞吐量，丢包率，延迟，背靠背。RFC2544 对上述性能评测参数的具体测试方法、结果提交形式作了较详细的规定。RFC3511 详细描述了防火墙设备的测试标准与方法，对 RFC2544 中的指标如何测试提供了更详细的指导。

RFC 文档定义的 4 个最重要的性能参数简述如下：

1. **吞吐量（Throughput）**

   被测设备在不丢包的情况下，所能转发的最大数据流量。通常使用每秒钟通过的最大的数据包数或者字节数来衡量。它反映了被测试设备所能够处理（不丢失数据包） 的最大的数据流量。

2. **丢包率（Loss Rate）**

   在一定的负载下，由于缺乏资源而未能被转发的包占应该转发的包数的百分比。它反映了被测设备承受特定负载的能力。

3. **延迟（Latency）**

   发送一定数量的数据包，记录中间数据包发出的时间 T1，以及经由测试设备转发后到达接收端口的时间 T2，然后取 T2 和 T1 差值为时延。它反映了被测设备处理数据包的速度。

4. **背靠背（Back-to-Back）**

   以所能够产生的最大的速率，发送一定长度的数据包，并不断改变一次发送的数据包数目，直到被测设备能够完全转发所有发送的数据包，这个包数就是此设备的背对背值。它反映了被测设备处理突发数据的能力（数据缓存能力）。

2.2 吞吐量是网络层性能的首要指标

RFC3511 针对防火墙设备（早期防火墙是典型网络层设备），进一步明确了吞吐量的测试方法（参见 http://tools.ietf.org/html/rfc3511 5.1.4.1 节），强调：

1. 在不发生丢包前提下，设备可承担的最大压力，其单位是比特每秒或包每秒；
2. 吞吐量只计算 IP 数据包的比特数（包含 IP 头和负载），不包括链路层的任何数据。

可见，吞吐量能够衡量网络设备承载流量压力（不允许丢包）的能力，同时，吞吐量只计算 IP 数据包的比特数，说明吞吐量被明确定位为网络层指标。吞吐量概念清晰，易于测量，因此广泛用于描述各种网络设备的关键性能指标，成为网络设备最重要的性能标准。

3. 应用层性能指标

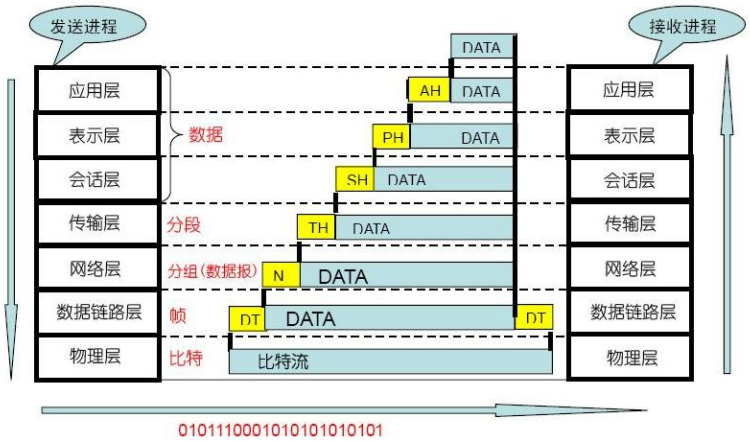
随着应用需求和网络技术的发展，用户不再满足于基本的数据通信能力，而是更多关注业务本身的运营效率与安全，希望看到网络流量中用户、应用、内容等可理解的信息，并针对性完善管理策略。以网络安全为例，根据权威咨询机构 Gartner 的统计，近年来 75%的网络攻击都发生在应用层，甚至在网络层流量“制胜”的 DDoS 攻击，近年来也有向应用层上移的趋势。截止 2014 年，30%以上的 DDoS 攻击都是基于应用程序的，这个比例还在逐年提高。

正是由于用户需求和网络技术越来越向应用层发展，应用层网络设备近年来不断涌现。然而，有别于业内广泛认可使用“网络层吞吐量”衡量网络设备性能，目前并无统一标准来评估应用层设备的性能。

3.1 应用层网络设备特点

为了澄清应用层设备性能指标，首先需要了解它与传统网络层设备的区别。OSI 模型是

国际标准化组织（ISO）和国际电报电话咨询委员会（CCITT）联合制定的开放系统互连参考模型，为开放式互连信息系统提供了一种功能结构的框架。它从低到高分别是：物理层、数据链路层、网络层、传输层、会话层、表示层和应用层。数据包在每一层进行层层封装，来实现互联网上的正常转发。



应用层网络设备，和传统网络设备相比，几乎所有的特性都是聚焦于应用层协议。其不再简单的关注数据报文的转发，而是关注和应用协议相关的内容，如协议识别、应用特征识别、应用威胁识别等，并基于此开发功能特性。针对这些特性，通常应用层网络设备都有各种各样的识别引擎，无论何种算法，最基本的一个要求就是：必须把数据包解封到第七层，即应用层。而网络层设备以数据包转发为主要目的，只关心数据包转发能力，只需要解封到第三层，找到对应的 IP 地址和端口信息即可。

数据包封装和解封，层次越多，CPU 的计算负载就越高，这会直接体现在性能上的衰减。同样处理能力的 CPU，处理网络层数据转发和应用层识别，所呈现的性能参数是完全不同的，不能放在一起横向比较，不能使用传统网络层性能指标来评价应用层设备的性能。

3.2 NSS Labs 对应用层性能的定义

国际领先网络设备厂商，如 Cisco，Palo Alto Networks 都或多或少提出了针对自有

产品的性能指标，但更多属于一家之言，局限性较大，相比之下，NSS Labs 的测试标准具有更严谨的定义而被广泛认可。NSS Labs 是全球最知名的独立安全研究和评测机构，总部设在美国，其认证的权威性来源于它的定位 -- 模拟用户真实应用场景，严格的产品评测过程，以及公正无私的数据分析。

针对应用层设备，NSS 提出了相应的评估指标与测试方法。

1. **裸包处理能力（RAW PACKET PROCESSING PERFORMANCE）**

   测试方法：通过发送包长在 128 字节到 1518 字节之间的 UDP 报文来计算最大吞吐量，此项指标对应 RFC 定义的网络层吞吐量。

   测试目的：对于应用层网络设备，测试网络层吞吐量的目的是衡量设备对数据报文的基本转发能力。虽然应用层设备主要关注应用数据，但如果某些具有攻击特征的数据包严重影响了设备的处理能力，则应用层的性能也将受到显著影响，应用层处理引擎能力再强也无法发挥作用。因此，尽管这种裸包数据对于应用层引擎来说像是“无用包”，但对于“无用包”的高效处理能力是保证应用层引擎正常工作的基础。

2. **HTTP 性能（HTTP CAPACITY WITH NO TRANSACTION DELAYS）**

   测试方法：通过发送平均大小在 1.7K 至 44K 之间 HTTP 页面来计算最大 HTTP 吞吐，且必须是成功获得 HTTP 响应的连接。业界将此项指标理解为设备的应用层吞吐量。

   测试目的：通过给应用引擎施加最大的压力，来获得设备应用引擎的最大工作能力。该测试提供了实验室中尽量接近“真实世界“的流量模型，以保证测试准确性和可重复性。

3. **最大 TCP 新建连接速率（MAXIMUM TCP CONNECTIONS PER SECOND）**

   测试方法：通过正常建立和销毁 1 字节负荷的 TCP 连接，来计算最大 TCP 新建连接数。业界将此项指标理解为设备的网络层新建。

   测试目的：1 字节负荷的 TCP 连接在真实流量中几乎没有，但可以通过这个经过简化和

抽象的方法来衡量最大 TCP 连接建立速率。

4. **最大 HTTP 新建连接速率（MAXIMUM HTTP CONNECTIONS PER SECOND）**

测试方法：发送一个 1 字节大小的 HTTP 页面，且必须获得正常的 HTTP 响应，计算每秒可以建立的最大 HTTP 连接数。业界将此项指标理解为设备的应用层新建。

测试目的：由于应用层设备需要维护应用协议的各种状态，因此需要通过给应用引擎施加应用计算压力，来衡量应用引擎的能力。

3.3 小结

NSS 建议采用 4 个指标评估应用层设备性能参数：网络层吞吐量、网络层新建速率、应用层吞吐量、应用层新建速率。

对于应用层设备，引入网络层吞吐和网络层新建指标主要是衡量基础的数据转发能力，以确保工作引擎在攻击流量下仍然有足够的应用层处理能力；应用层吞吐和应用层新建指标，是为了衡量应用引擎能力的高低，代表应用层处理技术的有效性和先进性。高应用层性能可以保障单位计算资源处理更多的应用层数据包，更好地满足应用识别与控制需求。

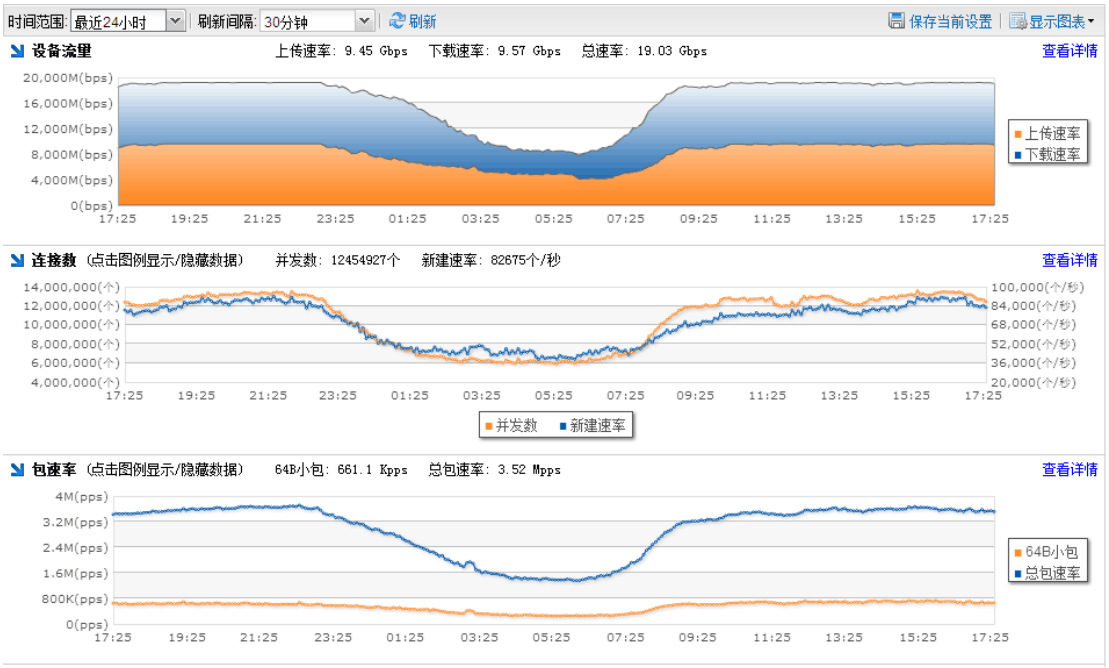
4. 应用层性能测试方法

4.1 逼近“真实世界”的流量模型

NSS Labs 给出了应用层性能指标与测试指导，由于其测试产品品类多、覆盖广，为了具有理论和方法的普适性，未充分考虑产品实测的流量模型，且国内的热点应用、网络条件、使用方式有着不同的特点，也要求测试模型更符合本地化要求。因此，如果能采用贴近中国市场“真实世界”的流量模型，将能更准确地评估性能指标，也会提高测试过程的效率。

网康科技致力于网络应用层技术的研究，选择了近 30 家高校客户与运营商客户的网络进行长期跟踪研究，发现网络高峰期每 Gbps 流量的平均包速率在 170-220Kpps 之间，平均包长在 580-750 字节之间，平均新建速率在 4200-5500 个/秒之间，这为构建流量模型提供了真实的数据。经过进一步抽象与标准化，可以认为：对于每 Gbps 的流量，包速率采用 185Kpps，平均包长 670 字节，新建连接速率 5000 个/秒，能够比较准确地描述实际流量，可以作为实际性能测试的流量模型。

下图是一个实际的例子，显示了某运营商核心节点的流量、连接数、包速率的关系。



某运营网络流量、连接数、包速率关系

4.2 应用层性能指标与测试方法

网康科技参考国际标准，并结合自己对应用层性能测试的理解和实践，给出应用层性能指标和测试方法的建议。

1. 应用层吞吐量

测试方法：在开启应用识别引擎的前提下，通过发送平均 21K 大小 HTTP 页面来测试

设备应用层最大吞吐量，且只能计算成功获得 HTTP 响应的连接。

说明：选取大小为 21K 的页面，是因为按照上一节讨论的最接近实际流量模型，将流量换算为 HTTP 页面，恰好为 21K。

2. 开启 IPS 的应用层吞吐量

测试方法：在开启应用识别引擎、IPS 引擎的前提下，通过发送平均 21K 大小 HTTP 页面来测试设备应用层最大吞吐量，且只能计算成功获得 HTTP 响应的连接。本项测试主要针对下一代防火墙（NGFW）设备。

说明：下一代防火墙（NGFW）已开始全面替代传统防火墙和 UTM，IPS 是标配功能模块，大多数场景下需要开启。由于 IPS 开启会较大影响整体性能，因此有必要考察开启 IPS 功能后设备的性能表现。

3. 应用层新建速率

测试方法：发送 64 字节大小的 HTTP 页面，且必须获得正常的 HTTP 响应，计算每秒可以建立的最大 HTTP 连接数。

4. 网络层吞吐量

测试方法：发送 1518 字节 UDP 数据包来测试设备网络层最大吞吐量，与传统方法相同。

5. 网络层新建速率

测试方法：正常建立和销毁 1 字节负荷的 TCP 连接，来计算最大 TCP 新建连接数。

5. 结论

本文从需求、使用、技术等多个角度探讨了网络层设备和应用层设备的特点与差异，指出传统的网络层性能标准无法有效衡量应用层网络设备的能力。

通过参考 RFC 标准、业界最权威测试机构的测试方法，并结合网康长期对应用层技术的研究，网康建议使用 4 项通用指标来评估应用层网络设备性能：应用层吞吐量，应用层新建速率，网络层吞吐量，网络层新建速率。对于下一代防火墙（NGFW）设备，还需要考察开启 IPS 的应用层吞吐能力。此外，应用层指标应作为评估应用层设备性能的主要指标，网络层性能参数可作为参考指标。

6. 关于网康科技

北京网康科技有限公司（以下简称网康科技）成立于 2004 年，总部位于北京，是中国上网行为管理理念的缔造者和下一代网络安全管理的引领者。网康科技拥有员工近千人，建立了遍及全国的销售与服务网络。产品线涵盖网络安全、应用安全、网络管理及优化，形成了完整的网络安全解决方案，并在业内率先发布云管端下一代网络安全架构与方案，实现云管端智能协同、主动防御，有效应对未知威胁及 APT 攻击。网康的产品和方案广泛应用于政府、金融、能源、教育、通信、企业等众多行业，客户数量超过 20000 家。

网康科技是业界最具技术创新和产品研发实力的企业之一，拥有一流的互联网内容研究实验室和网络安全攻防团队，并缔造了“全球最大的中文网页分类数据库”和“中国最大的网络应用协议库”。

网康科技承担了国家工业和信息化部技术改造项目、国家发展和改革委员会信息安全专项、下一代互联网专项、北京市发改委工程实验室技术创新能力建设项目、北京市科学技术委员会高新技术成果转化等项目，并被评为“国家级高新技术企业”，“北京市知识产权局专利试点单位”。网康科技还入选德勤“亚太区高技术、高成长企业 500 强”，并荣获“福布斯 2014 非上市潜力企业 100 强”，“中关村 2014 高成长企业 TOP100”等荣誉。