构建新型的 Web 应用交付基础架构

——WiseGrid 慧敏应用交付网关

背景概述

从全球范围来看，各种规模的企业都将其数据中心“web 化”,新型的 Web 应用程序框架开发方法日益盛行。现在，基本上所有类型的应用系统，如网站、电子商务、网银、企业 ERP、CRM 和 Web 2.0 等，均以 Web 应用模式向用户交付，我们早已不会对此感到惊讶，事实上，它已经成为标准。

基于‘HTTP/HTTPS’通信协议的 Web 应用交付模式具有天生的优势,企业都希望通过‘Web 化’来提高应用系统的灵活性和可接入性，同时节省成本并简化管理。然而，所有的 IT 机构都已发现，Web 应用交付的道路艰难崎岖，Web 应用的飞速发展并不能掩饰它存在的诸多不足，最主要的表现在于 Web 应用的性能问题、安全问题、可靠性问题和应用控制等方面的问题。

为了解决 Web 应用交付存在的问题，业界涌现出很多的技术和产品，有成功并走向成熟的，也有中途夭折的，但无论怎样，随着时间的推移，在我们的数据中心里面都留下了它们的印记。这时，企业的数据中心发现，在 Web 应用服务器前端，我们已经构建了一个庞大的优化和保护层。还记得这些熟悉的术语吗－“负载均衡交换机”、“SSL 加速网关”、“缓存服务器”、“广域网加速器”、“Web 入侵检测与防护”等等，这都是众多“点技术产品”所遗留下来的。事实上我们在 Web 应用服务器前端或多或少均部署了相关的产品，不仅让企业投入巨额资金，而且还带来集成、管理和维护工作上的烦恼。

在 Web 应用模式的高速发展的阶段，我们已经付出了巨大的代价，但是 Web 应用系统前端部署复杂的“设备层”对企业来说是无法承受的，这种趋势最终将出现逆转，因此构建一种新型的 Web 应用交付基础架构是必然的选择。新型的应用交付基础架构与传统解决方案具有明显的区别，新架构应该更加简洁、灵活和高效，确保应用系统安全、稳固、高效、可控的运行。

应用交付技术的发展历程

应用交付基础架构的核心就是应用交付控制器（Application Delivery Controller），也称之为应用交付网关。应用交付网关是一个达到网络设备标准的统一平台，这个平台将会集成智能流量管理、应用加速、应用安全防护和应用控制等诸多技术特点，并将最终替代 Web 应用前端庞大的“设备层”，从而帮助企业的数据中心构建全新的应用交付基础架构。应用交付网关承载应用系统的交互流量，它将能够深度感知应用内容，通过与应用系统紧密结合，智能地处理应用数据。因此，这个新平台可以提高应用性能、可靠性、可管理性和安全性，进而使企业数据中心能够快速而有效地实现应用交付基础架构。

然而，应用交付的概念并不是一蹴而就的，正是由于 Web 应用服务器的某些限制催生出许多技术解决方案，并通过时间的积累逐渐演变而来。事实上应用交付的概念来源于成熟的负载均衡技术，而负载均衡技术起源于利用多服务器承载更大的业务量，所以应用交付技术的发展也是经历了几代的更迭。

在互联网 Web 浏览时代的早期，利用集群软件或 DNS 轮询的办法协同服务器群为用户提供统一的应用访问，是应用交付的起源，它简单的解决了服务器扩展的燃眉之急，但是软件模式的性能瓶颈注定无法满足业务要求，所以第一代的应用交付专用设备－负载均衡器横空出世，负载均衡设备的出现是应用交付领域开始发展的一个重要标志。

第一代应用交付专用设备－负载均衡器的特点主要体现在三个方面：

1、丰富的负载均衡算法
2、强大的应用健康检查方法
3、会话保持技术

此时的负载均衡设备主要基于 TCP 层的信息对大数据提供有效的负载分担策略，它将多个服务器作为一个虚拟的大型服务器有效呈现给外界，外部用户通过发布的虚拟服务器的 IP 地址和应用端口实现应用访问。而众多客户请求则被负载均衡设备按照一定的策略转发到多台真实的应用服务器，同时提供对应用服务器的健康检查和对应用会话的保持功能，确保应用系统的可用性。在此阶段，负载均衡技术不断发展并逐渐成熟，负载均衡设备已经不仅仅满足服务器负载均衡领域，也扩展到防火墙负载均衡、链路负载均衡和全局负载均衡，另外为了满足了业务迅猛增长的需要，专用的 ASIC 芯片也被整合到负载均衡硬件平台之中。

如果说第一代负载均衡设备的侧重点还在于可用性保障和负载分发算法方面，那么第二代应用交付专用设备则是在负载均衡的基础上全面拓展，突出了应用安全防护、应用性能优化和对应用行为与内容的智能控制。主要特性如下，主要增加如下特性：

1、基于应用层智能的应用交换技术，通过深度识别应用层信息，实现基于内容信息的智能转发功能。

2、针对 Http 应用提供诸多性能优化功能，包括 TCP 连接复用、TCP 单边加速、Http 压缩、内容缓存、SSL 加速等。通过这些优化技术的引入，不仅提升系统的整体性能，还有效的卸载了真实服务器上的非内容处理任务，解决了服务器不断扩容的压力。

3、提升了 Web 应用安全的防护能力，包括网络层的 DDdos 防护、应用层的协议清洗、针对应用数据的 SSL 加密，甚至有些应用交付网关直接集成 Web 应用防火墙功能模块，在安全防护能力上第二代应用交付网关产品有了显著的提升。

4、基于应用层信息，依据策略实现对应用行为和内容进行干预和控制的能力。它通过定制化的策略脚本，来改变应用的请求或响应，而这些改变是在不需要更改程序代码的情况下实现的，强化了通过应用交付网关来达到控制应用的能力。

第二代应用交付网关的产品的特点是多功能紧密集成的统一平台，有效避免了多产品集

成解决方案的复杂度和相互制约的弊病，降低了用户的前期投入和后期管理维护的成本。应该说，我们目前正处于第二代应用交付网关的成熟期，在企业面对应用交付挑战的时候，应用交付技术快速发展，积极应对，为应用系统达到安全、稳固、高效、可控的目标奠定了坚实的基础。

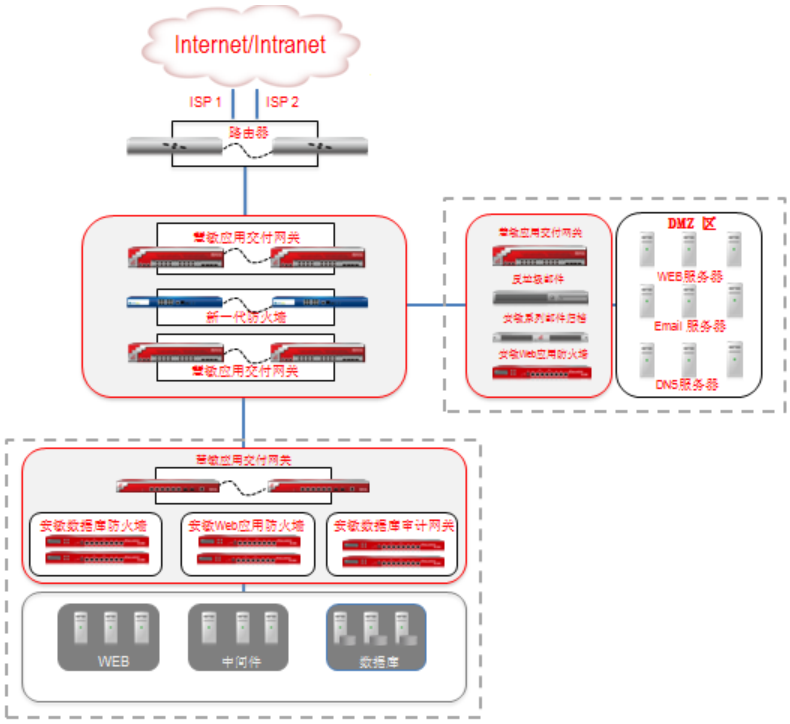
然而，应用交付技术的发展并没有放慢脚步，业界在 XML 交换、Web2.0 应用、基于网络的语义交换方面都在探索更加智能的应用交付技术。而且伴随着云计算模式的普及，应用交付领域又迈入更加广阔的空间，云计算环境下的应用交付、弹性计算、多租户服务等方面都出现了应用交付网关设备的身影，应用交付的未来空间广阔，令人憧憬。

WiseGrid 慧敏应用交付网关介绍

北京信诺瑞得软件系统有限公司的目标就是希望通过最新一代的应用交付技术平台为企业数据中心构建新型的应用交付基础架构。我公司最新推出的 WiseGrid 慧敏应用交付网关产品，在设计初始就基于实现全面的 Web 应用交付角度来考虑设计，WiseGrid 应用交付网关能够在深度理解用户、会话和业务内容上下文的情况下高效处理应用数据。

它缺省工作在反向代理的模式下（Reverse Proxy），实现在客户端与 Web 服务器之间的任务截断，同时在应用交付网关上实现基于应用智能的流量管理，应用性能提升，应用可靠性、安全性保障，应用行为的控制和云计算环境下计算资源的弹性调度等功能。虽然经历这些复杂的数据处理过程，但其性能及吞吐量依然达到网络设备线速高吞吐能力。

WiseGrid 应用交付网关是旨在适应新型的数据中心的产品，它在简化数据中心整体架构的同时提供了全面有效的应用交付功能，避免了多产品的复杂集成和较高投入。真正采用统一平台的方法，为客户提供完整并且性价比极佳的应用交付解决方案。



接下来，让我们详细了解一下 WiseGrid 慧敏应用交付网关的功能特点。

服务器负载均衡

负载算法

负载均衡算法就是负载调度的标准和决策依据，它可以使 WiseGrid 慧敏应用交付网关有效地管理应用流量。当 WiseGrid 部署在真实应用服务器群之前时，负载均衡算法指引客户端请求选择一台最合适的真实服务器来进行处理，决策方法将确保实现最佳的流量分配。

WiseGrid 提供的负载均衡算法可以根据 TCP 层信息（TCP 连接数）或 Http/Https 请求中的包头信息或正文信息（例如 URL、应用数据类型或 Cookie）将应用流量分类，结合服务器健康状况检查以及服务器处理性能权重确保将客户端请求送往适当的服务器，从而提高应用可用性。

WiseGrid 系统可以设置以下这些算法来实现服务器负载均衡：

- 轮询（Round Robin）

轮询算法按照请求的先后顺序将其依次循环地调度到不同的真实服务器，即每次调度执行 i = (i + 1) mod n，并选出第 i 台服务器。算法非常简洁，无需记录当前连接的状态，所以它是一种无状态调度。

- 加权轮询（Weighted Round Robin）

加权轮询算法是在轮询算法的基础上，根据真实服务器的不同处理能力，依照预设的权重按比例来调度访问请求。这样可以保证处理能力强的服务器能处理更多的请求任务。

- 最少连接（Least Connections）

最小连接算法是一种动态调度算法，它依据服务器当前活跃的 TCP 连接数信息来估计服务器的负载情况。应用交付网关会记录每个真实服务器活跃连接的数目，当一个请求来临时，会被调度到活跃的 TCP 连接数最少的那台服务器。

- 加权最少连接（Weighted Least Connections）

加权最小连接算法是在最小连接算法的基础上，通过为各个服务器设定相应的权值，在调度新连接时依据服务器已建立连接数和其权值的比例，然后把新的连接请求分配到当前比例最小的服务器上。

- 最小期望延迟(Shortest Expected Delay)

最小期望延迟算法是基于加权最小连接算法的基础上，针对连接数与权重比例相同的情况进行决策，例如：4 个真实服务器，当前活跃 TCP 连接数分别为 1、2、3、4，服务器权值也为 1:2:3:4，那么根据加权最小连接算法（WLC），他们的连接数与权值的比例都是 1，此时依据 WLC 算法，新请求可以随机分给任何服务器。

最小期望延迟算法希望在这种情况下，经请求调度给可能响应最快的服务器，因此它的连接数与权值比例计算公式调整为（当前活跃连接数+1）/权值，因此基于这个公式，最小期望延迟算法会将新请求发给权值为 4 的那台服务器。

- 源地址哈希（Source Hashing）

源地址哈希算法根据请求的源 IP 地址,进行哈希运算,得到此 IP 的哈希值(Hash Key)，从静态分配的哈希表找出对应的真实服务器,若该服务器是可用的,则将请求发送到该服务器。如果此时该服务器发生故障，源地址哈希算法将在现哈希值的基础上再次进行哈希运算，以获得新的服务器对应。

- URL 哈希（URL Hashing）

URL 哈希算法根据 HTTP 请求的 URL 信息,进行哈希运算，得到哈希值(Hash Key)，从静态分配的哈希表找出对应的真实服务器,若该服务器是可用的,则将请求发送到该服务器。如果此时该服务器发生故障，URL 哈希算法将在现哈希值的基础上再次进行哈希运算，以获得新的服务器对应。

- 不排队调度（Never Queue）

不排队调度算法相对比较简单，它基于最小延迟调度算法，但是如果有真实服务器没有活跃 TCP 连接，则请求不再进行最小延迟算法运算，直接将请求发送至此服务器。

服务器健康检查

应用交付网关是应用数据交互的载体，它需要解决系统可用性问题，而后台真实的应用服务器是真正处理应用业务请求并响应客户请求的宿主机,因此实时监控应用服务器的健康状态，避免将请求转发到不可用的服务器是保障整个系统可用性的关键。

WiseGrid 慧敏应用交付网关的高级健康检查功能，可以准确的做到应用层的健康检查。而服务器健康检查的结果是负载均衡决策的依据,健康状态为不可用的服务器将不再参与负载分发决策。WiseGrid 慧敏应用交付网关支持的服务器健康检测方法，包括：基础网络检查方法(Ping)、四层应用检查方法(TCP)、基于应用层内容的应用检查方法(HTTP/HTTPs/SMTP)以及利用自定义脚本的应用检查程序（检查范围可以从 2 层到 7 层）。

- 网络层健康检查方法是利用 Ping（ICMP）的手段来侦探目标 IP 是否，这种健康检测方法由于只是验证网络的可用性,因此常常用于网络设备或链路可用性的健康检查场景，并不适合做应用的健康检测。

- 四层（TCP 层）健康检查方法可以提供应用最基本的状态监视，以检验应用是可访问的。TCP 检查是通过向应用的 TCP 监听端口发送 TCP SYN 请求,以接收到 SYN ACK 回应消息做为健康的依据。

- 七层（应用层）健康检查方法，可以检测应用协议的工作状态。例如可以根据 HTTP 或 HTTPS 请求的响应代码（200 OK）判断 Web 服务的健康状态，通过查看发送 SMTP Hello 信息的响应获得 SMTP 服务的健康状态。如果发现应用层故障，用户即被透明地复位到可用的服务器上。

  另外对于 HTTP 或 HTTPS 应用，还可以自定义检测 Http 响应包的内容，通过检查返回内容中是否含有用户定义的字符串来判断应用的健康状态,这种检查方法的粒度更细，确保应用处于严格的健康状态之下。

- 自定义脚本的健康检查方法允许用户利用 Perl script，完全依据应用系统的特性定义发送和接收数据包的行为与内容。根据脚本程序执行的结果来确认应用的健康状态。

会话保持

WiseGrid 慧敏应用交付网关不仅可以为关键业务系统提供高可用性和智能负载均衡，与此同时，还可以满足用户固定访问特定服务器的要求，以支持用户会话持续建立到某台固定服务器上。会话保持意味着一旦一个真实服务器被选择处理某客户端请求，那么后续的从该用户发出的请求都被转发到同一服务器上。“会话保持”功能常用于需要检查会话状态的一致性的特定应用，例如：ERP 系统、电子商务系统等。

WiseGrid 主要支持的“会话保持”的方法包括：

- 基于源 IP 进行会话保持

  基于源 IP 会话保持，慧敏应用交付网关会对此客户端 IP 与虚拟服务器之间的连接创建一个计数器，只要持续性计数器尚未到时，它们之间新建的连接就会持续转发到同一台服务器。

- 基于 SSL Session ID 进行会话保持

  基于 SSL Session ID 的会话保持主要应用于 SSL 应用（例如 https），在这种模式下，判断持续性请求的依据是 SSL 的 Session ID，具有相同 Sesssion ID 的请求会在会话保持计时内，转发到同一台服务器。

- 基于 Cookie 进行会话保持

  基于 Cookies 的会话保持主要应用于 Http/Https 应用，在这种模式下，利用 Http 数据包存储的 coockie 信息进行持续性的判断，并把持续性请求转发到同一台服务器上。

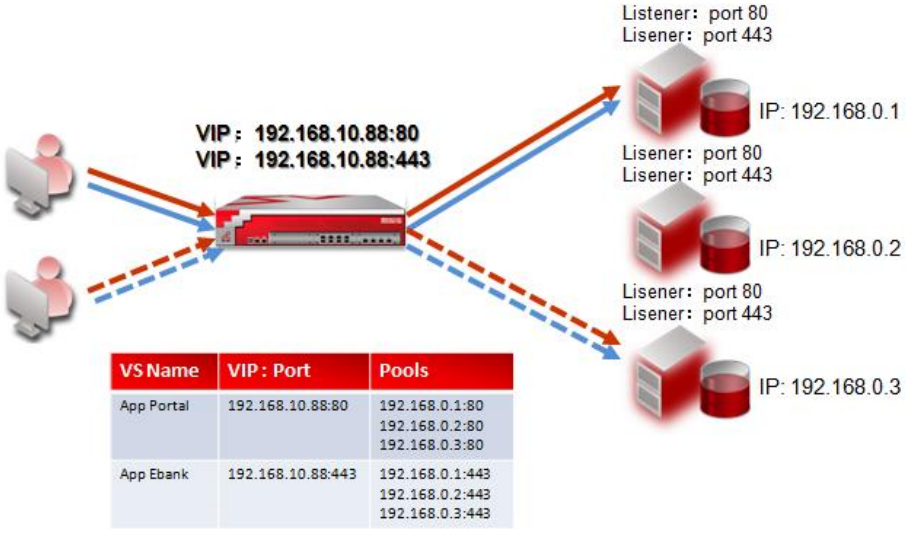
会话保持技术中，还有一个重要的控制参数：会话保持时间。这个时间就是相同用户前后两次持续性请求之间的间隔时间，如果相同用户前后两次持续性请求的时间间隔超过会话保持时间，会话保持机制将忽略其会话的关联性，这个会话请求会依据负载均衡算法进行决策调度。

会话保持组

会话保持技术的应用通常适用于单独的虚拟服务器，但是还有一种特殊的会话保持技术可以跨越多个虚拟服务器之间提供会话保持，这就是会话保持组的概念。WiseGrid 慧敏系列应用交付网关提供的会话保持组技术是会话保持机制的扩展，当一个复杂的应用系统场景，由多个虚拟服务器提供服务，例如一个 Http 页面嵌套了一个 Https 的界面，应用系统要求能够让某个客户端的所有特定请求（无论是 Http 还是 Https）都能保持在一个相同的真实服务器上处理，我们就需要做不同 Vserver 之间的会话保持。不过会话保持组技术实施的前提是不同虚拟服务器所绑定的真实服务器必须是相同的，只是真实服务器上均运行着不同的服务（例如 http 和 https）

以某网银应用系统举例来说，它的网站门户页面（http）中嵌有网银系统的登录接口（https），后台真实服务器上均启动 http 和 https 服务，应用业务要求相同用户的 Http 访问和 Https 访问必须转发到相同的真实服务器上。此时在 WiseGrid 应用交付网关上，我们就需要为 Http 和 Https 服务分别建立一个虚拟服务器，并且两个虚拟服务器之间做会话

保持组绑定。当某位用户的 Http 初始请求根据负载均衡算法被分发到服务器 A 上，此时如果我们启动了会话保持及会话保持组功能，那么该用户在设定的会话保持时间内所有 Http 请求仍然转发到该服务器上，当他通过门户登录网银系统时，它的 Https 服务请求也因为会话保持组的关系，而被转发至此台真实服务器。



VS Name	VIP : Port	Pools
App Portal	192.168.10.88:80	192.168.0.1:80 192.168.0.2:80 192.168.0.3:80
App Ebank	192.168.10.88:443	192.168.0.1:443 192.168.0.2:443 192.168.0.3:443

应用加速

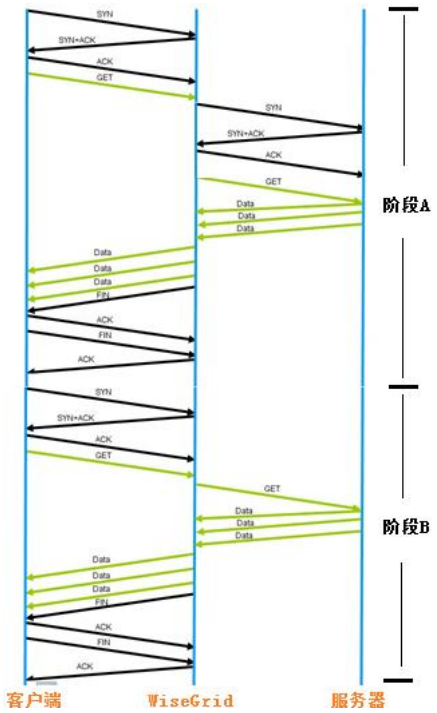
我们在使用负载均衡设备的过程中不难发现，90%的应用场景都在承载 Http 业务流量，这与数据中心的“web”化变迁密不可分。但是 Web 应用天生的性能缺陷使得仅仅利用传统负载均衡设备难以解决应用性能的问题，因此 WiseGrid 慧敏应用交付网关在改善 Web 应用系统性能方面提供诸多技术手段，能够帮助 Web 应用提升 5－10 倍的效率。

对于 HTTP/HTTPS 应用， WiseGrid 慧敏应用交付网关其实扮演着反向代理的角色，它终结来自客户端的连接，并替代客户端与真实服务器通过常连接进行应用层数据交互，正是这种机制让我们对 Web 应用性能进行优化提供可能。

TCP 连接复用

TCP 连接复用功能正是在反向代理模式下,WiseGrid 慧敏应用交付网关通过 Keep-alive 代理－保持服务器端的持久连接，可以为多个客户端连接的应用层请求所利用，高效率降低后端真实服务器上的 TCP 连接数量，同时避免 TCP 连接频繁拆建、慢启动等繁琐的过程。这样服务器可以将更多的资源用在对于 Http 请求的响应上，而不是去浪费在 TCP 连接的建立和拆除的管理上。慧敏应用交付网关提供这种连接管理技术令 Web 服务器真正实现 TCP 的有效卸载，从而提升服务器对应用层数据的处理性能。在实际的环境下，WiseGrid 慧敏应用交付网关可以实现从 50：1 甚至到 1000：1 的有效 TCP 复用率。

那么，TCP 连接复用是如何让服务器端 TCP 连接被多个客户端连接复用呢？让我们通过了解 WiseGrid 的 TCP 连接管理过程来理解 TCP 复用吧。首先 WiseGrid 会“中断”客户端 TCP 连接,将应用层的信息提取出来，并且异步选择一个与服务器端持久建立的空闲 TCP 连接提供服务。而当客户端接收到 RST 或 FIN 信号时客户端 TCP 连接会拆除，但服务器端连接仍然保持，并为另一个新建的客户端连接的应用请求提供服务，这样服务器端 TCP 连接被多个客户端连接串行复用。



如上图所示，时间阶段 A，假设某客户端 A 的 TCP 连接请求 SYN 发送至 WiseGrid，它们之间经过三次握手协商后，客户端开始发送应用层消息（Http Request），此时服务器端 TCP 连接开始启动（假设当时服务器端没有可以被复用的空闲连接存在），服务器端将通过 KeepAlive 机制建立持久连接，为客户端的应用层信息交互提供服务。随后 WiseGrid 将服务器响应的数据再传递给客户端 A，当 WiseGrid 通过响应内容长度确定响应数据传输完毕后，主动向客户端发送 FIN 信号，此时客户端 TCP 连接会关闭，而服务器端连接则继续保持。此时如果另外一个客户端 B 的 TCP 连接请求同时发送至 WiseGrid，客户端连接经过三次握手协商后建立起来，WiseGrid 则通过查找连接复用表，在 Reuse Pool 中发现刚刚为客户端 A 服务的 TCP 连接正处于可用的空闲状态，因此客户端 B 的应用层请求直接通过此连接转发至 Web 服务器，WiseGrid 与服务器之间无须再建立新的 TCP 连接。

与此同时，我们还可以看到，WiseGrid 对 HTTP 交易是深度感知并进行管理的，通过分析和了解 HTTP 请求和响应的头部信息，当客户端请求到达应用交付网关时，WiseGrid 会从 TCP 重用连接池中（Reuse pool）选择一个 TCP 连接，当这个 HTTP 交易结束后，WiseGrid 会将这个 TCP 连接放回 TCP 重用连接池（Reuse pool）。

TCP 单边加速

在典型的广域网环境中，一定存在丢包、延迟、抖动等限制，导致 TCP 连接的吞吐效率低下，数据传输耗时太长，应用响应缓慢甚至无法使用，而且常常因为延迟较大而无法有效地利用带宽，造成带宽的闲置和浪费。这完全是因为传统 TCP 协议导致的结果，因此针对传统 TCP 传输协议进行优化，提高应用数据在广域网上的传输效率，突破 TCP 的技术瓶颈，成为应用性能改善的关键任务之一。

TCP 加速技术中通常分为“双边加速”和“单边加速”，其显著区别在于：

- “TCP 双边加速”产品需要在 TCP 连接的两端部署硬件设备或者安装软件，TCP 加速的行为基本上需要双边设备协同完成，设备之间通常使用了与 TCP 不兼容的私有协议，破坏了网络透明性。而且，这种加速方法并不适合互联网应用。
- “TCP 单边加速”无需在客户端部署任何硬件设备或者安装软件，通过智能的流控

算法，保持兼容传统的 TCP 流控机制，但可以提升数据的网络传输效率，达到 TCP 加速的效果。

WiseGrid 慧敏应用交付网关作为服务器端部署的设备，更适于提供智能学习和自适应网络的 TCP 单边加速技术，帮助应用系统提升性能。该加速引擎采用智能流控算法，改善因网络丢包和超时所引起的 TCP 低效率，优化广域网络数据传输缓慢的问题，而且它与所有传统 TCP 协议栈兼容，不对 TCP/IP 头部字段进行任何修改，无需在客户端部署任何硬件设备或者安装软件。

WiseGrid 慧敏应用交付网关单边 TCP 加速引擎，改善的是传统的 TCP 拥塞控制机制，这个机制有四个阶段：慢启动(slow start)、拥塞避免(congestion avoidance)、快速重传(fast retransmit)和快速恢复（fast recovery)。在网络质量稍差的情况下，这种机制会产生令人难以忍受的性能问题。例如进入拥塞避免阶段，拥塞窗口增长速度过慢；超时后会重新开始慢启动阶段；丢包后的恢复过程较长等等。使用 TCP 单边加速技术之后，WiseGrid 能够提升对网络状况的识别和判断能力并自适应网络变化。

首先，它可以更加准确的判断拥塞的发生，并合理掌控流控机制，为数据传输提供更加合理的传输路径，避免因传统机制不合理而引起误判或判断滞后，而令网络的频繁拥堵。

其次，它可以利用全新的算法，敏捷的洞察网络状态的实时变化，提前预测出网络拥塞发生的可能，在网络上真正发生丢包行为之前进行干预，提前进行拥塞避免。虽然暂时降低数据传输速率，但能减少因拥塞而产生大量数据包的丢失后重传，数据传输效率更高。

另外，由于超时或丢包而导致拥塞发生之后，WiseGrd 在缓解拥塞的同时，尽可能收敛整个机制各个过程的时间，例如快速提高拥塞避免后窗口尺寸的恢复，避免过多的等待，提高带宽利用率。

最后，TCP 单边加速引擎在提升带宽利用率的同时，还兼顾链路延迟 RTT 的友好性，智能实现高速率（High Speed）和 RTT 的之间的公平性，从而真正令应用数据更为高效的通过网络传输环节。

WiseGrid 慧敏应用交付网关单边加速引擎共提供 7 种智能的拥塞管理算法，以适应复杂的广域网络环境，包括：BIC、Cubic、Vegas、Hybla、HTCP、Veno 和 Westwood 算法。

Http 数据压缩

WiseGrid 慧敏应用交付网关支持与客户端之间的 Http 数据响应以压缩的形式传输，从而减少数据传输时间，提升应用体验。WiseGrid 提供标准的压缩算法 Deflate 或 Gzip，通常情况下至少可以减少 30％－80％的数据流量。

WiseGrid 是按照 Http 对象类型定义压缩引擎的，缺省情况下针对 HTML、SHTML、DHTML、JHTML、PHTML、Javascript、J2EE、JSP、CSS 样式表单和 XML 等文本类对象实施压缩处理，包括 doc、xls、ppt 等微软 Office 文档，而忽略那些压缩率不高的对象，例如图片、Flash、视频等。

同时，WiseGrid 可以智能感知客户端浏览器是否支持压缩，如果浏览器发出不支持压缩的请求，那么响应则以明文的形式给出。

智能缓存

WiseGrid 慧敏系列应用交付网关采用基于 Http 对象的内容缓存功能，利用策略规则将

允许缓存的 Http 对象存放在系统内存空间或硬盘空间。当用户的 HTTP 访问请求发送到 WiseGrid 时，如果 Cache 中的内容能够匹配用户的访问请求则直接由 WiseGrid 来响应用户的访问，从而避免了对后台服务器的负载压力，在减小了后台服务器负载的同时，提高了对用户的响应速度和整体网站的处理能力。

WiseGrid 慧敏应用交付网关采取不同于其它厂商仅仅使用内存做缓存内容存储的模式，它增加了硬盘存储的能力，避免内寸空间的限制，有效扩展了缓存内容的数量和容量。同时又通过监视缓存内容的命中率变化情况，及时将命中率较高的对象迁移至内存中，保障绝大多数命中的内容对象始终存放在内存中，避免频繁的磁盘 I/O 读写操作，实现真正智能的内容缓存。WiseGrid 提供的智能缓存功能兼容 HTTP 1.0 和 HTTP 1.1 协议，缓存内容可以自动清除或刷新；

SSL 加速

随着诸如网购、网银、证券、网上办公、电子商务等类型的应用在人们的日常生活当中日益普及，这些应用的重要性也越来越高，尤其对这类应用的数据安全性要求格外严格，通常情况下，SSL技术已经广泛应用到这些互联网应用之中，通过SSL技术的安全保障机制，提升了应用的认证安全和数据安全。

但是SSL技术使数据安全性得到保障的同时，对应用服务器的性能也提出了更高的要求，由于SSL的加解密过程是CPU密集型任务，需要消耗的大量的服务器性能，从而影响了这类应用的性能体验。为了更好的解决SSL应用性能瓶颈问题，WiseGrid应用交付网关提供SSL加速能力，在提升用户访问体验的同时，还有效避免服务器硬件的投资。

SSL加速的核心解决思路，首先是将耗费服务器资源的SSL处理任务从后台服务器迁移到 WiseGrid 慧敏应用交付网关设备之上，其次要保障应用交付网关对SSL数据的处理性能。由于应用交付网关内置专业的SSL加速卡，其拥有超强的SSL协商和加解密能力，能够满足高并发访问应用的需求。通过借助WiseGrid在SSL业务处理上的优势，客户端与应用交付网关之间实现SSL化的应用数据交互，而后台服务器与WiseGrid之间采用明文处理，将原本在Web服务器上处理的SS任务全部卸载到应用交付网关上。

应用安全防护

Web 应用的安全性是目前业界的热点话题，WiseGrid 慧敏应用交付网关提供的安全防护机制能够帮助 Web 应用系统提升安全防护能力，它具有如下特点：

1、反向代理模式，屏蔽对服务器的直接攻击
3、提供全面的 DDos 防护能力
4、提供基于 Http 协议的协议清洗能力
5、定制安全过滤策略
6、利用 SSL 增强数据传输的安全性

隔离保护

由于 WiseGrid 慧敏应用交付网关缺省工作在反向代理模式，因此黑客无法直接与服务器联系，通常情况下，应用交付网关上的虚拟服务器只对外提供应用服务端口，从而有效地屏蔽了黑客攻击的第一步——端口扫描，增加了黑客攻击的难度。此时，服务器完全被应用交付网关与外界隔离，甚至对于提供服务的 Web 应用，WiseGrid 也会其尽量屏蔽响应信息中关于真实服务器的信息，如 Server 类型（Apache,IIS 等）信息，让黑客无法准确定位后台 Web 系统的构架和类型。

但是来自外界的攻击并不可能因此而消失，大量非法的攻击行为都落在应用交付网关的身上，因此大力提升应用交付网关系统的自身防护能力尤为重要。

DDos 防护

DDoS 攻击主要是黑客利用系统对外提供的正常服务，通过其控制的大量主机持续发送的数据请求，以达到阻止业务正常服务的目的。WiseGrid 慧敏应用交付网关为了抵御 DDoS 攻击，在内核中优化了 TCP 协议栈，利用独特的连接管理技术和其它限制手段精心打造一个全面的 DDos 防护体系。

正常的 TCP 连接遵循标准的 TCP 三次握手，然后发送数据，为了维持 TCP 连接的管理，系统需要为每一个 TCP 连接初始分配 512 字节的内存空间。典型的 DDos 攻击（例如 SYN_Flood），黑客就是利用 TCP 连接管理的机制，采用大量非法的半开连接来消耗系统资源，导致系统资源耗尽而服务崩溃。WiseGrid 对此进行了全新设计，它仅在可以确定客户端发起的是一个合法的应用层访问行为时，才为此 TCP 连接分配内存。WiseGrid 利用 syn-cookie 机制跟踪所有 TCP 连接信息，并为所有客户端 TCP 连接生成一个键值，并通过 SYN+ACK 返回给客户端，客户需要提交这个值在最终的 ACK 阶段去建立 TCP 连接，因此 WiseGrid 对于半开的 TCP 连接不消耗任何资源。但是 syn-cookie 机制可以使让应用交付网关在面对大量 syn 攻击流量时，仍然为合法流量和用户访问提供正常服务。

而对于 Web 应用，WiseGrid 即使与客户端完成 TCP 连接建立的三次握手过程，在收到 HTTP 有效请求（Get/Put/等）前，仍并不分配任何系统资源给这个 TCP 连接 — 这个特征能够有效阻止空闲连接攻击（idle client connection attack）。同时在服务器端，利用 TCP 连接复用技术，尽量减少服务器端的 TCP 连接数量，以便服务器把更多的资源放投入到 HTTP 交易处理上。

另外，应用交付网关还可以针对整个虚拟服务器或某个真实服务器提供带宽限制功能，通过限制流量吞吐，连接和请求速率来保护服务器，同时通过基于内容验证的健康检测机制，精确探测服务器的处理能力，从而在服务器处理能力饱和之前自动屏蔽新的连接请求，以防止整个系统过载。

Http 协议清洗

针对协议层攻击，WiseGrid 慧敏应用交付网关内置协议清洗功能引擎，该引擎实时检测 Http 请求内容是否符合 RFC 定义，防止 Http 协议范围外的行为，可以对 HTTP 协议整理、发送无缺陷的数据包、阻止非法请求等。对于黑客恶意制造的非法请求，直接进行有效阻断，避免无效的请求耗费服务器资源。

例如我们将正常的 Http 请求头中故意将 GET / HTTP/1.1 改写成 GET / HTTP/1.8，此类恶意的攻击请求基本可以骗过一般的防火墙产品，如果存在大量此类攻击，Web 服务器将会造成极大资源浪费来响应这些“bad request”，严重的甚至会造成内存溢出的异常现象。WiseGrid 通过协议清洗引擎，直接将这些协议异常行为及时阻断并丢弃。

基于策略的安全过滤规则

WiseGrid 慧敏应用交付网关提供基于策略脚本的应用行为和内容控制机制，其中基于请求的内容过滤功能，主要用于用户自定义非法请求特征，并予以告警和阻断。

这种能力来源于两个坚实的技术基础，首先，WiseGrid 具有详细的 HTTP 数据流检测能力，能够深度感知应用层信息。其次，WiseGrid 核心的策略控制引擎为外部提供一套完整的策略控制机制，它基于系统内部 API，允许用户通过简单易用的 SmartBiulder 图形工具

定制请求过滤策略规则（SmartRules）。

正是基于如上的技术基础，我们可以利用策略脚本自主应对对不断变化的安全威胁，迅速做出响应、采取措施，增强系统的安全防护能力。基于 WiseGrid 应用交付平台的 SmartRules，简单易用，结合对非法请求特征的分析和定位，可以轻松实现部分入侵检测与防护能力。

应用智能控制

在应用交付概念中，对应用行为和内容的智能控制是其核心理念的重要组成部分，因此应用智能控制也成为应用交付网关至关重要的功能。WiseGrid 慧敏应用交付网关的核心工作引擎－策略控制引擎允许通过SmartRules来实现Web应用层数据的深度检测和灵活控制。WiseGrid 应用交付网关可以识别 Http 应用层信息，包括：识别客户端类型及属性，了解请求使用的方法（Get 或 Post），了解响应的结果类型，甚至了解请求或响应的内容，基于对这些信息的细粒度识别，策略控制引擎可以制定规则，对应用流量的进行分类判断，并基于判断的结果进行策略操作。

例如某电商交易系统中，我们可以根据 Http 请求头中 cookie 的信息，识别商品交易的种类和金额，然后根据预先定义的策略，将购买贵重物品的高金额交易请求转发至服务质量保障较高的应用服务器上。

WiseGrid 的策略控制引擎就是根据预先定义的类似上面例子的规则脚本（SmartRules），实现对特定应用数据进行特殊操作。SmartRules 策略规则由两个重要部分组成的，即识别策略控制对象的“表达式”和针对此对象的“操作行为”。

SmartRules 中的“表达式”，实际上就是利用 Http 应用层的信息或信息之间的任意组合，准确定位具有期望特征的数据流。通过“表达式”的描述，我们可以识别哪些应用流量是希望被此规则进行管控的。例如我们可以根据 Http 请求头信息中 Accept-Language 属性就可以方便的区分那些浏览器是支持中文的，哪些是不支持中文的，而至于如何控制“表达式”判断出的结果，则由“操作行为”来决定。

宏观上，WiseGrid 规划了 4 类应用控制的“操作行为”，它们是：

1、基于 Http 请求的内容交换行为

2、基于 Http 请求的内容过滤行为

3、基于 Http 请求的内容改写行为

4、基于 Http 响应的内容改写行为

通过操作行为的定义，策略控制引擎就知道对相应的应用数据如何进行操作，以达到 SmartRules 定制的策略。

最终，策略规则是需要被虚拟服务器引用的，而且我们制定的策略规则可能是多个，那么每个策略都会定义优先级，策略执行的顺序就是根据规则的优先级来确定的。

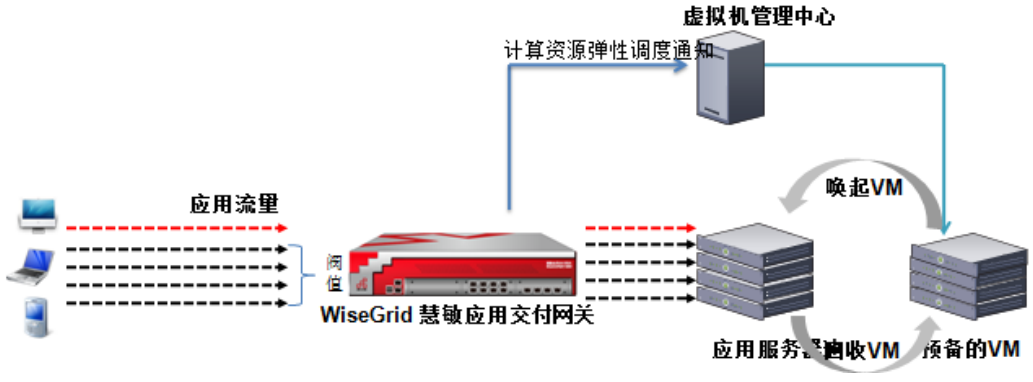
弹性计算

弹性计算是云计算环境中的一个重要特征，它可以根据工作负载的需求，动态、灵活调整服务器计算资源，弹性计算真正使云计算中心具备智能性，最终为用户提供按需所求的服务。WiseGrid 慧敏应用交付网关将流量管理、资源调度与虚拟化技术紧密结合，通过与云计算虚拟机管理平台的集成，为云计算环境提供一个完整的弹性计算解决方案，并成为解决方案中不可或缺的重要组成部分。

那么应用交付网关为什么会在弹性计算解决方案中如此重要呢？这是因为它具有得天独厚的天生优势，应用交付网关是应用交互数据的承载平台，对业务负载的变化非常敏感并最先感知。同时应用交付网关提供强大的服务器健康检测机制，可以实时掌握后台服务器的运行状态。最为关键的是，应用交付网关负责整体协调后台服务器计算资源的负载调度，那么对于业务的变化以及所需计算资源的多少，应用交付网关作为控制点是最为合适的。但问题的关键是，无论传统的负载均衡设备还是当前的应用交付网关，通常都是负责固定服务器群的负载调度，即使遇到业务的高峰或低谷，计算资源依然是固定的。除非人为的干预，通过调整服务器池中的参数，才能完成这种变化，可是这种管理方式是非常不灵活的，而且操作性较差。

不过我们也欣喜的发现，在云计算环境下，虚拟化技术的特点正好可以弥补上述的不足，虚拟机的灵活调度在云计算环境中根本不是什么问题，那么应用交付技术与虚拟化技术的完美结合，为弹性云计算提供坚实的技术基础。

WiseGrid 慧敏应用交付网关正是基于这种理念进行设计，其内置的弹性调度引擎是四大核心任务引擎之一。它首先利用 WiseGrid 对业务变化的实时跟踪数据，利用决策模型确定业务的高峰和低谷，并计算出满足对应业务情况下的合理计算资源配置，然后通过基于事件驱动的触发器（例如通过预置应用负载的阀值作为计算资源调配的触发器），向云计算资源管理中心发送指令，例如唤起或关闭虚拟机。通过云计算资源管理中心进行相应的计算资源调配，实现计算资源的弹性调度，保证用户以最佳的资源配置为应用提供服务，实现绿色 IT。

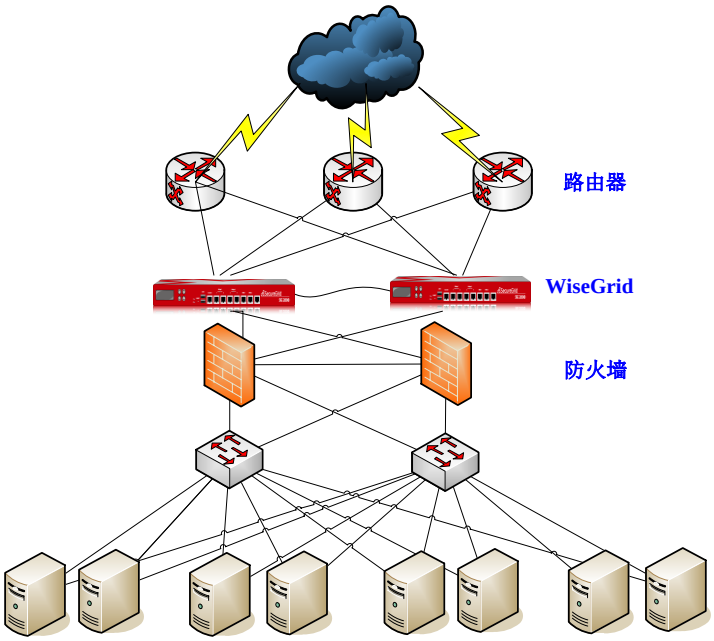


在实现计算资源弹性调度解决方案中，WiseGrid 慧敏应用交付网关将作为云计算资源管理中心的第三方应用者角色，通过资源管理中心提供的开放式 API，与资源管理中心进行集成，例如通过调用 VMware ESX server 或 Xen Server 提供的 API，远程执行唤起或关闭后台应用服务器策略，动态掌控应用服务器资源的规模。

部署方式

WiseGrid 慧敏应用网关部署非常简便，支持双臂/串接模式和单臂/旁路模式。

双臂/串接工作模式，应用交付网关串接部署在用户网络链路中，所有流量都通过 WiseGrid 处理，此时应用交付网关开启路由转发功能。系统将访问虚拟服务器的数据流量转发至任务引擎进行处理，与虚拟服务器无关的数据流量转发至路由模块进行处理。



单臂/旁路模式，应用交付网关以单臂模式部署到网络之中，此种模式不需要改变客户现有的网络结构，只要保证应用交付网关与客户端和服务器之间的网络连通性。同时设备上架也不会造成业务的中断，实现简洁、快速的部署，此种部署方式在实际应用中最为常用。