

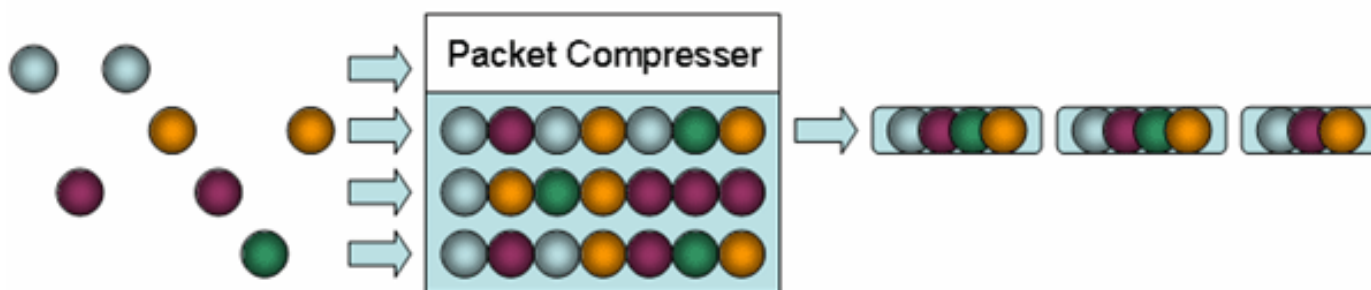
概述

几乎所有广域网优化技术都充分利用高级压缩例程来提高应用性能。此外，这些技术以更高水平存储并利用以前传输的网络数据，以实现高压缩比率。然而，在执行进一步检查时，如何获得这些优势以及随之产生的限制，将会出现很大差异。

解决方案

数据包与会话之比较

迄今为止大多数网络压缩系统都是基于数据包。基于数据包的压缩系统缓冲数据包都通过解压器引导至远程网络。此后，用户可一次压缩一个数据包，或一次压缩多个数据包，然后再发送至在其中反向进行该流程的解压器。基于数据包的压缩功能已应用多年，并且路由器、VPN 客户端、Juniper Networks 的 WX 和 WXC 设备都具备这种功能。



基于数据包压缩应用的主要问题是压缩时它将多种数据类型混合在一起。所有压缩例程在处理同类数据时将获得更大的压缩比。在处理异质数据时（例如，多种协议的大量数据包），压缩比率会大大降低。

基于数据包的压缩系统会存在其它问题。压缩数据包时，这些系统必须在网络中编写小数据包，并进行其它工作以集合并封装多个数据包。仅有其中一项操作不可能达到最佳效果。在网络中编写小数据包会增加 TCP/IP 标头的开销。另外，集合并封装数据包会为该数据流增加封装标头。

与以前的压缩解决方案不同，F5 WANJet 在对话层中运行。这会支持 WANJet 在处理所有应用类型时能够在完全同类的数

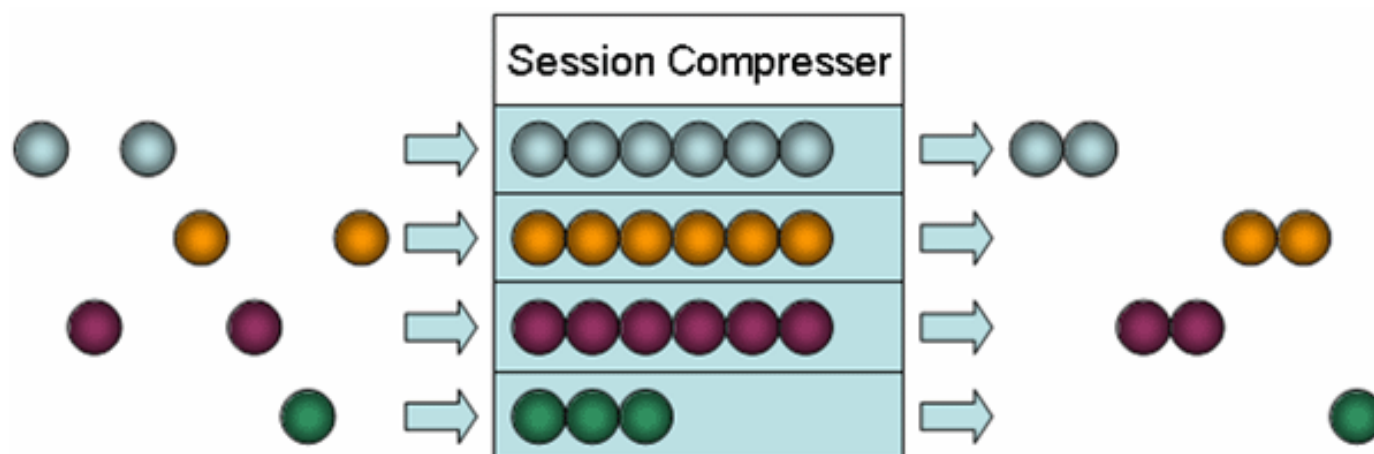


图 2: 基于会话的压缩

此外，在会话层中运行时，数据包界限与重新分组的问题也得以解决。这能够使 WANJet 在数据流中轻松找到匹配的数据，在第三层中这些数据可能是许多分开的字节，但在第五层中可能就是邻近的字节。由于取消了封装阶段，因此在对话层中执行压缩时系统的吞吐率会增加。

字典大小

所有压缩例程共同存在的局限性是存储空间有限。许多例程，例如 **gzip**，只能存储 **64 Kb** 的数据。其它技术，例如基于磁盘的压缩系统，可以存储 **1 TB** 的数据。为了理解字典大小的作用，需要对高速缓存管理内容有一个基本的了解。

请求 **web** 站点类似，并非所有网络中传输的字节会在同一个频率下重复。有时系统会通过高频率传输一些字节，因为这些字节是常用文件或通用网络协议中的一部分。其它字节只会出现一次并且不会重复出现。压缩和堆积定律 (**Zipf's Law and Heaps' Law**) 中描述了频繁重复字节序列和非频繁重复字节序列之间的关系。

所有基于当前字典的压缩系统会通过存储频繁访问的数据并删除非频繁访问的数据以进行不均等的分配。通过这种优化方式，存储少于 **10%** 的所有字节方式会使命中率超过 **50%**。这种字节方式的不均等分布效果充分证明了公共压缩程序的效率。**Gzip** 仅存储 **64kb** 的历史记录，但平均能够压缩近 **64%** 的内容。**Bzip2** 能够存储 **100kb** 至 **900kb** 的历史记录，平均压缩了 **66%** 的内容。尽管数据存储空间不足，但 **Gzip** 和 **Bzip2** 仍能出色运行的原因在于频繁出现的字节序列能够表示网络中的大多数字节。

块与字节之比较

基于块的系统（例如 **Juniper Networks WXC** 和 **Riverbed** 技术的 **Steelhead** 设备）可存储以前在广域网中传输数据流部分。再次遇到这些块时，其参考数据会传送到远程设备中，该远程设备继而会重组原始数据。

基于块的系统主要缺点是反复出现的数据和块的长度永远不会完全相同。因此，匹配仅是部分匹配，还会留下一些重复数据不被压缩。下图详细描述了使用 **256** 字节块大小压缩 **512** 字节数据时的情况。

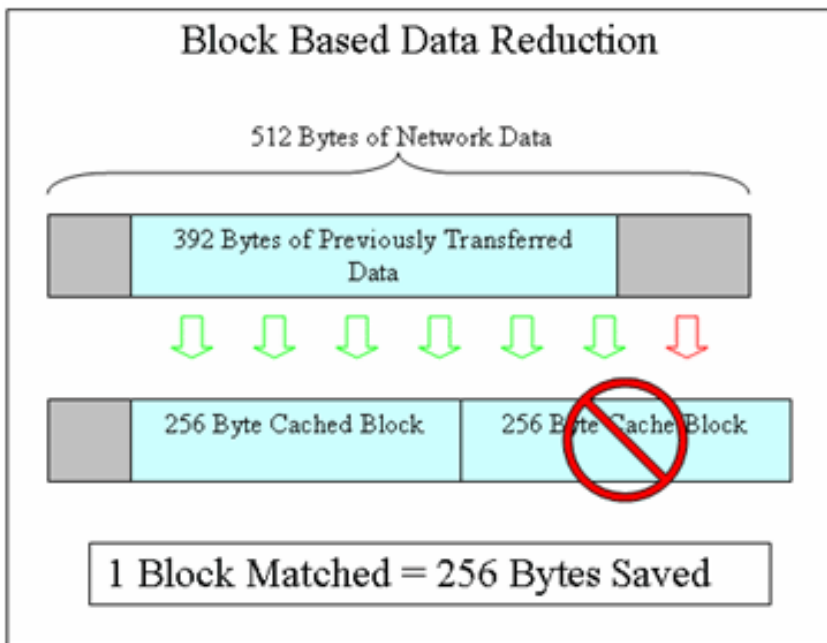


图 3：基于块的压缩示例

与 Riverbed 和 Juniper 方法类似（通过以前传输的数据来减少网络利用率），F5 WANJet 采用 TDR（透明数据压缩）技术为以前传输的字节建立字典。但与 WXC 和 Steelhead 设备不同，WANJet 匹配并发送带有字节级粒度 (byte level granularity) 的参考数据。下图说明了 WANJet 如何处理相同 (512) 字节的数据。

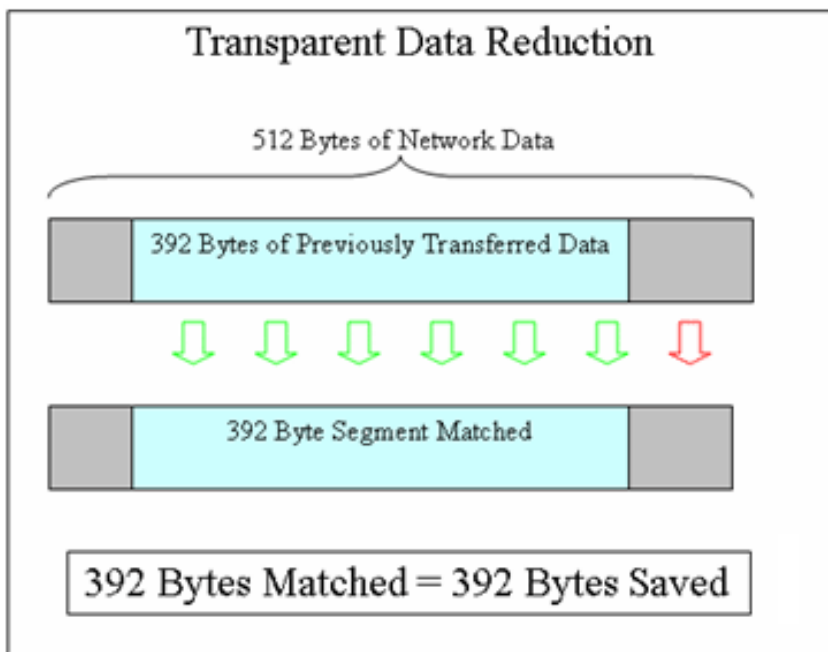


图 4：WANJet 压缩示例

与基于块的系统不同，所有重复模式都匹配并且都是通过 WANJet 压缩。在上述实例中，WANJet 能够匹配并减少所有重复出现的 392 字节数据。与基于块的系统相比，粒度级别无论对于文档还是对于应用层协议标头，均能提高其压缩级别。

是吞吐量造成的问题吗？

实现高压缩比率对提高带宽受限网络的应用性能非常重要，系统的吞吐率也发挥着举足轻重的作用。通过技术预期压缩率、设备峰值压缩吞率和网络带宽可以评估指定压缩技术的性能增益。如果压缩比率过低，那么该网络就会处于饱和状态并且性能增益会降至最低。同样，如何压缩速度过慢，那么压缩器本身就会称为瓶颈。

在下图中，链接速度低于 10 mbps 时，压缩比率较高 (75%) 的压缩器性能领先于速度较快的例程。这是因为速度高于 5 mbps 并低于 20 mbps 时，低速压缩例程无法处理用以继续充分利用网络的足够数据。

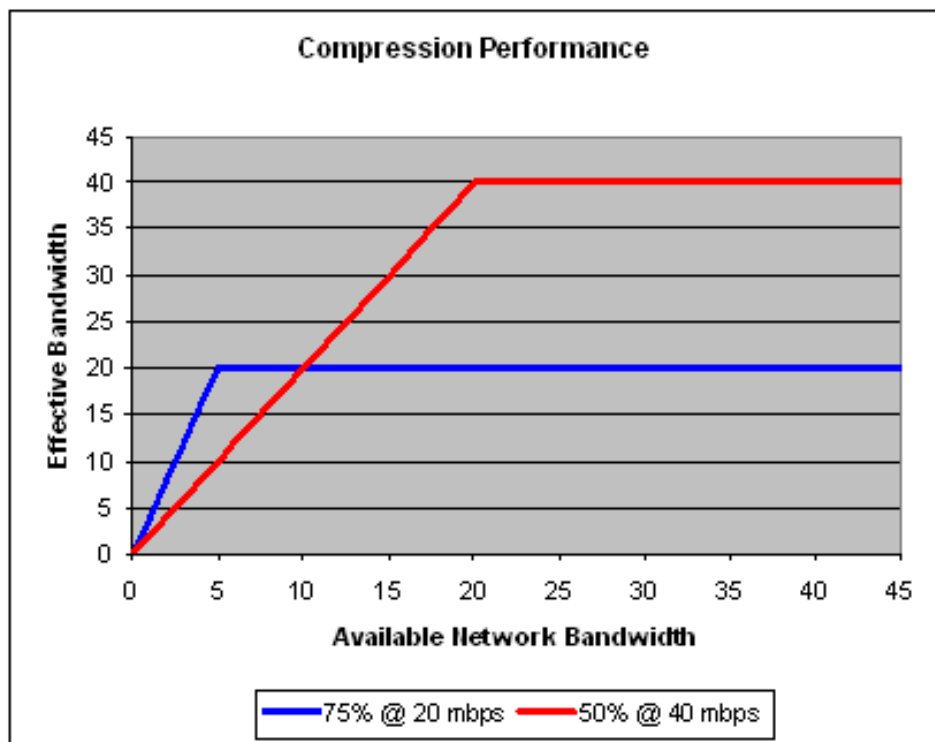


图 5: 压缩性能图

已优化在 F5 WANJet 中实施的 TDR (透明数据压缩) 以维持高吞吐率。Juniper WXC 500 速度高达 20 mbps 并且 Riverbed Steelhead 5010 速度高达 45 mbps 时，WANJet 可以通过单一设备维持高达 400 mbps 的速度。这种性能级别能够使 WANJet 为 T3 网络提供 8 倍的性能——而 WXC 和 Steelhead 设备则无法实现这一性能。

总结

通过压缩获得应用性能增益不仅要求适当的压缩算法，而且还要求设计适用于性能的系统架构。这种压缩系统必须正好与重复出现的模式匹配以实现高压缩比率。必须既可管理存储数据又可管理输入应用流量以实现最佳的效率。最后是必须快速完成压缩以降低延迟并继续填充网络。必须从整体对 F5 Networks WANJet 和 TDR 进行设计以满足这些需求。