性能管理:最佳实践白皮书

目录

[简介](#)
[背景信息](#)
[关键成功因素](#)
[绩效管理指标](#)
[性能管理流程](#)
[制定网络管理运营概念](#)
[衡量绩效](#)
[执行主动故障分析](#)
[绩效管理指标](#)
[记录网络管理业务目标](#)
[记录服务级别协议](#)
[创建基线的变量列表](#)
[查看基线和趋势分析](#)
[记录假设分析方法](#)
[记录用于提高网络性能的方法](#)
[摘要](#)
[相关信息](#)

简介

性能管理涉及优化网络服务响应时间以及管理单个和整体网络服务的一致性和质量。最重要的服务是需要衡量用户/应用响应时间。对于大多数用户而言，响应时间是关键性能成功因素。此变量影响用户和应用管理员对网络成功的看法。

背景信息

容量规划是确定对未来网络资源需求的过程，目的是防止对业务关键型应用的性能或可用性产生影响。在容量规划方面，网络基线（CPU、内存、缓冲区、输入/输出八位字节等）可能会影响响应时间。因此，请记住，性能问题通常与容量有关。在网络中，这通常是必须在队列中等待才能通过网络传输的带宽和数据。在语音应用中，这种等待时间几乎肯定会影响用户，因为延迟和抖动等因素会影响语音呼叫的质量。

另一个使性能管理复杂化的主要问题是，尽管高网络可用性对大型企业和服务提供商网络来说都是任务关键型的，但通常趋势是在冒成本长期增加（通常为不可预测）的风险下寻求短期经济收益。在每个预算周期中，网络管理员和项目实施人员都在努力寻求性能与快速实施之间的平衡。此外，网络管理员还面临着诸多挑战，包括为了满足狭小的市场窗口而快速开发产品、复杂的技术、业务整合、竞争性的市场、计划外的停机时间、缺乏专业知识以及工具通常不足等问题。

鉴于这些挑战，性能如何适合网络管理框架？理想的网络管理系统的主要功能是优化网络的运行能

力。一旦您接受这一点作为网络管理的最终目标，网络管理的重点就在于保持网络以最佳性能运行。

理想的网络管理系统包括以下基本操作：

- 通知操作员即将出现性能下降。
- 当出现性能下降或故障时，提供简单的备用路由和解决方法。
- 提供用于查明性能下降或故障原因的工具。
- 作为网络恢复能力和生存能力的主站。
- 实时传达性能。

基于此定义，性能管理对于网络管理至关重要。这些性能管理问题至关重要：

- 用户性能
- 应用性能
- 容量规划
- 主动故障管理

需要注意的是，对于语音和视频等较新的应用程序，性能是成功的关键变量，如果您无法实现一致的性能，服务会被视为价值低而失败。在其他情况下，用户会因为应用程序间歇性超时而遭受性能可变的困扰，从而导致工作效率和用户满意度下降。

本文档详细介绍最关键的绩效管理问题，包括关键成功因素、关键绩效指标和绩效管理的高级流程图。它还讨论了可用性、响应时间、准确性、利用率和容量规划等概念，并简要讨论了主动故障分析在性能管理和理想网络管理系统中的作用。

关键成功因素

关键的成功因素确定了实施最佳实践的要求。要作为关键的成功因素，流程或过程必须提高可用性，否则缺少过程必须降低可用性。此外，关键的成功因素应可衡量，以便组织能够确定成功的程度。

注：有关详细信息，请参阅绩效管理指标。

以下是绩效管理的关键成功因素：

- 收集网络和应用程序数据的基线。
- 对您的网络和应用执行假设分析。
- 对容量问题执行异常报告。
- 确定所有建议或潜在网络管理服务的网络管理开销。

- 分析容量信息。
- 定期审核网络和应用程序的容量信息，以及基线和异常。
- 设置升级或调整程序，以便从被动和长期角度处理容量问题。

绩效管理指标

绩效指标为组织提供了衡量关键成功因素的机制。绩效计划的绩效指标包括：

- 记录网络管理业务目标。这可以是网络管理的正式操作概念，也可以是所需功能和目标的非正式说明。
- 制定详细且可衡量的服务级别目标。
- 使用图表或图表提供服务级别协议文档，其中显示了如何在一段时间内满足这些协议的成败。
- 收集基线的变量的列表，例如轮询间隔、产生的网络管理开销、可能的触发阈值、变量是否用作陷阱的触发条件，以及针对每个变量使用的趋势分析。
- 定期举行会议，审查对基线和趋势的分析。
- 记录假设分析方法。这应包括建模和验证（如果适用）。
- 超过阈值时，制作有关用于增加网络资源的方法的文档。需要记录的项目是添加额外的WAN带宽和成本表所需的时间线。

性能管理流程

这些步骤为性能管理提供高级流程：

1. 制定网络管理运营概念
    a. 定义所需功能：服务、可扩展性和可用性目标
    b. 定义可用性和网络管理目标
    c. 定义性能SLA和指标
    d. 定义SLA
2. 衡量绩效
    a. 收集网络基线数据
    b. 评估可用性
    c. 测量响应时间
    d. 测量精度

e. 测量利用率

f. 容量规划

3. 执行主动故障分析

a. 使用阈值进行主动故障管理

b. 网络管理实施

c. 网络运营指标

制定网络管理运营概念

在定义网络的详细性能和容量变量之前，必须了解组织内网络管理的整体运营概念。在定义此总体概念时，它提供了业务基础，您可以基于此基础对网络中所需的功能建立精确的定义。如果您未能制定网络管理的操作概念，可能会导致缺少目标或目标，而目标又会因客户需求而不断变化。

您通常会将网络管理操作的概念作为网络管理程序系统定义阶段的第一步来生成。其目的是从操作角度描述所需的整体系统特性。本文档用于协调网络运营、工程、设计、其他业务部门和最终用户的总体业务（非定量）目标。本文的重点是形成网络管理和运营的远程运营规划活动。此外，它还为随后所有定义文件的编制提供指导，如服务级别协议。显然，这组初始定义不能过于狭隘地关注特定网络问题的管理，而是关注那些强调对整个组织的重要性，以及与必须管理的成本相关的项目。一些目标是：

- 确定高效使用网络基础设施所需的那些特征。
- 确定网络支持的服务/应用。
- 启动端到端服务管理。
- 启动基于性能的指标以改进整体服务。
- 收集和分发性能管理信息。
- 支持根据用户的反馈对网络进行战略评估。

换句话说，网络管理运营概念应侧重于整体组织目标以及实现这些目标的理念。主要组成部分包括任务的更高层次定义、任务目标、系统目标、组织参与和总体业务理念。

作为网络管理员，您能够统一用户通常不一致的性能预期。例如，如果网络的主要要求是将大型文件从一个位置传输到另一个位置，则您需要将重点放在高吞吐量上，而不是交互式用户的响应时间上。请注意不要限制您的绩效视图，除非您考虑各种各样的问题。例如，当您测试网络时，查看所使用的负载级别。负载通常基于非常小的数据包和非常大数据包的吞吐量。这些性能测试中的任何一个都可能产生非常积极的结果，但是根据您的网络流量负载，这些测试可能不会呈现真实的性能情况。研究在尽可能多的工作负载条件下网络性能并记录性能。

此外，虽然许多网络管理组织都采用有效的警报技术来通知技术人员设备发生故障，但定义并实施端到端应用程序性能评估流程则更为困难。因此，尽管网络运行中心(NOC)可以快速响应路由器或交换机故障，但可能破坏网络性能并影响用户感知的网络条件可能很容易被忽略，直到感知变为负

值。尽管非常困难，但此第二个过程可以为业务组织和网络管理提供巨大的好处。

最后，请确保不要对网络性能产生不切实际的期望。当您误解网络协议或应用的详细信息时，通常会产生不切实际的期望。通常情况下，性能不佳不是网络的故障，而是应用设计不佳的结果。记录和衡量应用程序性能的唯一方法是在安装应用程序之前建立网络性能基线。

定义所需功能：服务、可扩展性和可用性目标

性能管理、持续容量规划和网络设计的第一步是定义所需的功能和/或服务。此步骤要求您了解应用、基本流量、用户和站点计数以及所需的网络服务。此信息的第一个用途是确定应用程序对组织目标的重要性。您还可以应用此信息创建一个知识库，用于逻辑设计，以了解带宽、接口、连接、配置和物理设备要求。此初始步骤使网络架构师能够创建网络模型。

制定解决方案可扩展性目标，以帮助网络工程师设计满足未来增长需求的网络，并确保提议的设计不会因网络的增长或扩展而受到资源限制。资源限制可能包括：

- 整体流量
- 音量
- 路由数
- 虚电路数量
- 邻居计数
- 广播域
- 设备吞吐量
- 媒体容量

网络规划人员应确定所需的设计寿命、设计寿命期间所需的预期扩展或站点、新用户数量以及预期流量或变化。此计划有助于确保推荐的解决方案在设计的预计使用寿命内满足增长要求。

如果不调查解决方案的可扩展性，您可能不得不实施重大的被动设计变更。此设计更改可能包括额外的层次结构、介质升级或硬件升级。在那些依靠相当精确的预算周期来购买主要硬件的组织中，这些变更会成为整体成功的主要制约因素。在可用性方面，网络可能会遇到意外的资源限制，从而导致不可用和被动措施周期出现。

互操作性和互操作性测试对新解决方案部署的成功至关重要。互操作性可能指不同的硬件供应商，或者网络实施期间或之后必须集成在一起的不同拓扑或解决方案。互操作性问题可能包括通过协议栈向上传递硬件信令来导致路由或传输问题。互操作性问题可能发生在网络解决方案迁移之前、期间或之后。互操作性规划应包括不同设备之间的连接以及迁移过程中可能出现的拓扑问题。

解决方案比较是指您将不同的潜在设计与其他解决方案要求实践进行比较。此实践有助于确保解决方案最适合特定的环境，并且个人偏见不会推动设计过程。比较可以包括不同的因素，例如成本、恢复能力、可用性、风险、互操作性、可管理性、可扩展性和性能。一旦实施设计，所有这些都会对整体网络可用性产生重大影响。您还可以比较介质、层次结构、冗余、路由协议以及类似功能。创建一个图表，其中包含X轴上的因素和Y轴上的潜在解决方案，以总结解决方案比较。在实验室环

境中对解决方案进行详细比较，还有助于客观地调查与不同比较因素相关的新解决方案和功能。

作为网络管理操作概念的一部分，以所有用户都能够理解的方式定义网络目标和支持的服务至关重要。业务概念发展之后开展的活动在很大程度上受到该文件的质量的影响。

以下是标准性能目标：

- 响应时间
- 使用
- 吞吐量
- 容量（最大吞吐率）

虽然这些测量值对于简单的LAN而言可能微不足道，但在交换式园区网络或多供应商企业网络中可能非常困难。当您使用深思熟虑的运营计划概念时，每个绩效目标都以可衡量的方式定义。例如，在高峰工作时间，应用“x”的最小响应时间为500毫秒或更短。定义用于标识变量的信息、度量变量的方法以及网络管理应用程序应在一天中重点研究的时间。

定义可用性和网络管理目标

可用性目标定义网络服务的服务级别或服务级别要求。这有助于确保解决方案满足最终可用性要求。为特定组织定义不同的服务类别，并详述适合可用性需求的每个类别的网络要求。网络的不同区域可能还需要不同级别的可用性。要实现更高的可用性目标，可能需要增加冗余和支持流程。在为特定网络服务定义可用性目标并测量可用性时，您的网络组织可以了解实现计划的SLA所需的组件和服务级别。

定义可管理性目标以确保整体网络管理不缺少管理功能。要设置可管理性目标，您必须了解组织的支持流程和相关网络管理工具。可管理性目标应包括了解新解决方案如何适合当前支持和工具模型，并参考任何潜在差异或新要求。这对网络可用性至关重要，因为支持新解决方案的能力对于部署成功和达到可用性目标至关重要。

可管理性目标应涵盖支持潜在网络所需的所有重要MIB或网络工具信息、支持新网络服务所需的培训、新服务的人员配备模式以及任何其他支持要求。通常，这些信息在部署之前未被发现，由于缺乏分配用于支持新网络设计的资源，整体可用性会受到影响。

定义性能SLA和指标

性能SLA和指标有助于定义和衡量新网络解决方案的性能，以确保它们满足性能要求。推荐的解决方案的性能可以通过性能监控工具测量，也可以通过简单ping来测量推荐的网络基础设施性能。性能SLA应该包括平均预期流量、峰值流量、平均响应时间和允许的最大响应时间。这些信息稍后可以在解决方案验证部分使用，最终帮助确定所需的网络性能和可用性。

定义SLA

网络设计的一个重要方面是为用户或客户定义服务。企业称之为服务级别协议，而服务提供商称之为服务级别管理。服务级别管理通常包括对问题类型和严重程度以及帮助台职责的定义，例如上报

路径和每个级别支持级别上呈报之前的时间、开始处理问题的时间以及根据优先级结束目标的时间。其他重要因素包括在容量规划、主动故障管理、变更管理通知、阈值、升级标准和硬件更换方面提供的服务。

如果组织不预先定义服务级别，以后就很难改进或获得确定的资源需求。此外，也很难理解为帮助支持网络而需要添加哪些资源。在许多情况下，这些资源仅在发现问题之后才会应用。

衡量绩效

性能管理是一个包含不同性能区域的配置和测量的总称。本节介绍绩效管理的以下六个概念：

- 收集网络基线数据
- 评估可用性
- 测量响应时间
- 测量精度
- 测量利用率
- 容量规划

收集网络基线数据

大多数企业内部网都有足够的带宽。但是，如果没有足够的数据，您可能无法排除网络拥塞是应用程序性能不佳的一个因素。拥塞或错误的一个线索是性能不佳是间歇性的还是时间相关的。例如，晚上较晚时性能良好，但早上和工作高峰时性能非常缓慢。

一旦定义了网络管理操作的概念并定义了所需的实施数据，就有必要随着时间的推移收集这些数据。此类数据收集是网络基线的基础。

在新解决方案（应用程序或IOS更改）部署之前和部署之后执行当前网络的基线，以测量为新解决方案设置的期望值。此基准有助于确定解决方案是否满足性能和可用性目标以及基准容量。典型的路由器/交换机基线报告包括与CPU、内存、缓冲区管理、链路/介质利用率和吞吐量相关的容量问题。根据操作概念中定义的目标，还可以包括其他类型的基线数据。例如，可用性基线表明网络环境的稳定性/可用性得到了提高。在旧环境和新环境之间执行基线比较，以验证解决方案要求。

另一个专用基线是应用程序基线，当您确定应用程序网络需求趋势时，此基线很有价值。此信息可用于升级周期中的计费和/或预算目的。与每个应用的首选服务或服务质量相关的应用可用性领域，应用基线也很重要。应用程序基线信息主要由应用程序每时间段使用的带宽组成。一些网络管理应用程序还可以将应用程序性能作为基准。流量类型（Telnet或FTP）的细分对规划也很重要。在一些组织中，对网络更重要的资源受限区域进行监控，以监控最大流量生成者。网络管理员可以使用此信息来预算、规划或调整网络。调整网络时，可能会修改网络服务或应用的服务质量或队列参数。

评估可用性

网络管理员使用的主要指标之一是可用性。可用性是衡量网络系统或应用程序对用户可用时间的指

标。从网络的角度来看，可用性代表网络中各个组件的可靠性。

例如，为了测量可用性，您可以将帮助台电话呼叫与从受管设备收集的统计数据进行协调。但是，可用性工具无法确定故障的所有原因。

当您衡量可用性时，网络冗余是另一个要考虑的因素。冗余丢失表示服务降级，而不是网络故障总数。结果可能是响应时间较慢，并且由于数据包被丢弃而导致数据丢失。结果也可能显示在性能测量的其他领域，例如利用率和响应时间。

最后，如果根据SLA交付，您应该考虑计划停机。这些中断可能是由于移动、添加和更改、工厂关闭或您可能不希望报告的其它事件所导致的。这不仅是一项困难的任务，而且可能也是一项手动任务。

测量响应时间

网络响应时间是流量在两个点之间传输所需的时间。通过基线比较或超过阈值发现响应时间比正常稍慢，可能表示存在拥塞或网络故障。

响应时间是衡量客户网络使用的最佳指标，可以帮助您衡量网络的有效性。无论缓慢响应的来源是什么，用户都会因流量延迟而感到沮丧。在分布式网络中，影响响应时间的因素很多，例如：

- 网络拥塞
- 少于目标路由的路由（或根本就没有路由）
- 网络设备供电不足
- 广播风暴等网络故障
- 噪音或CRC错误

在采用QoS相关排队的网络中，响应时间测量非常重要，可以确定正确的流量类型是否如预期一样通过网络传输。例如，当您通过IP网络实施语音流量时，语音数据包必须按时按恒定速率传送，以保持良好的语音质量。您可以生成分类为语音流量的流量，以测量对用户显示的流量的响应时间。

您可以测量响应时间，以帮助解决应用服务器和网络管理器之间的争斗。当应用或服务器速度较慢时，网络管理员通常会被推定为有罪。网络管理员必须证明网络不是问题。响应时间数据收集提供了无可争议的方法来证明或证明网络是应用程序故障的根源。

只要可能，您就应该测量对用户显示的响应时间。用户将响应视为从按下Enter键或单击按钮到屏幕显示为止的时间。此经过时间包括每台网络设备、用户工作站和目标服务器处理流量所需的时间。

遗憾的是，由于用户数量和工具匮乏，这一级别的测量几乎是不可能的。此外，当您结合使用用户和服务器响应时间时，当您确定未来的网络增长或排除网络故障时，它几乎没有什么价值。

您可以使用网络设备和服务器来测量响应时间。您还可以使用ICMP等工具来测量事务，虽然在上层处理事务时不会考虑引入系统的任何延迟。此方法解决了网络性能知识的问题。

在简单层面上，您可以对从网络管理站到网络关键点（如大型机接口、服务提供商连接的端点或关键用户IP地址）的ping响应进行时间安排，以测量响应时间。这种方法的问题是它不能准确反映用

户对其机器和目的机器之间的响应时间的感知。它只是从网络管理站的角度收集信息并报告响应时间。此方法还逐跳地屏蔽了整个网络中的响应时间问题。

以服务器为中心的轮询的一种替代方法是将工作分布到您想要模拟进行测量的源和目标附近。使用分布式网络管理策略并实施Cisco IOS服务保证代理(SAA)功能。您可以在路由器上启用SAA，以测量路由器与目标设备（如服务器或另一台路由器）之间的响应时间。您还可以指定TCP或UDP端口，强制以与模拟流量相同的方式转发和定向流量。

通过在多服务网络上集成语音、视频和数据，客户可以在其网络中实施QoS优先级。简单的ICMP或UDP测量无法准确反映响应时间，因为不同的应用程序接收不同的优先级。此外，使用标记交换时，流量路由可能因特定数据包中包含的应用类型而异。因此，ICMP ping在每台路由器处理它的方式上可能具有不同的优先级，也可能收到不同的、效率较低的路由。

在这种情况下，测量响应时间的唯一方法是生成类似特定应用或相关技术的流量。这会强制网络设备按照处理实际流量的方式处理流量。您可以使用SAA或使用第三方应用感知探测器达到此级别。

测量精度

准确性是衡量不会导致错误的接口流量的指标，可用将一段时间内成功率与总数据包率进行比较的百分比来表示。您必须首先测量错误率。例如，如果每100个数据包中有两个导致错误，则错误率将为2%，准确率将为98%。

在早期的网络技术中，特别是在广域网中，一定程度的错误是可以接受的。但是，对于高速网络和当今的WAN服务，传输要准确得多，除非出现实际问题，否则错误率接近于零。接口错误的一些常见原因包括：

- 超出规格的布线
- 电子干扰
- 有故障的硬件或软件

使用更低的准确率来触发更详细的研究。您可能会发现特定接口出现问题，并决定这些错误是可以接受的。在这种情况下，应调整此接口的准确度阈值，以反映错误率无法接受的位置。在较早的基线中可能已报告了无法接受的错误率。

此表描述的变量用于精确度和错误率公式：

记法	描述
ΔifInErrors	收集snmp ifInErrors对象的两个轮询周期之间的差值（或差值），它表示出错的入站数据包计数。
ΔifInUcastPkts	收集snmp ifInUcastPkts对象的两个轮询周期之间的增量，该对象表示入站单播数据包的计数。
ΔifInNUcastPkts	收集snmp ifInNUcastPkts对象的两个轮询周期之间的增量，它表示入站非单播

	数据包（组播和广播）的计数。

错误率的公式通常以百分比表示：

错误率=(ΔifInErrors)*100

(ΔifInUcastPkts +(ΔifInNUcastPkts)

请注意，错误率和准确度公式中不考虑出站错误。这是因为设备绝不应故意将存在错误的数据包放到网络上，出站接口错误率也绝不应增加。因此，入站流量和错误是衡量接口错误和准确性的唯一标准。

精确度公式采用错误率，并从100中减去它（同样以百分比形式）：

准确性= 100 -(ΔifInErrors)*100

(ΔifInUcastPkts +(ΔifInNUcastPkts)

这些公式反映MIB II接口(RFC 2233)通用计数器的错误和准确性。结果以一个百分比表示，该百分比将错误与看到和发送的数据包总数进行比较。从100中减去结果的错误率，从而产生准确率。100%的准确率是完美的。

由于MIB II变量存储为计数器，因此您必须执行两个轮询周期并计算两者之间的差值（因此公式中使用了Delta）。

测量利用率

利用率可衡量一段时间内特定资源的使用情况。衡量标准通常以百分比的形式表示，其中资源使用量与其最大运营能力进行比较。通过利用措施，您可以确定整个网络中的拥塞（或潜在拥塞）。您还可以确定未充分利用的资源。

利用率是确定网络管道（链路）的完整程度的主要衡量标准。 测量CPU、接口、队列和其他与系统相关的容量测量值，以确定网络系统资源消耗的程度。

高利用率不一定很差。低利用率可能表示流量流入了意外的位置。随着线路过度使用，效果会变得显着。当排队通过某个接口传输的流量超过该接口可处理的流量时，就会发生过度利用。资源利用率的突然跳变可能表明存在故障情况。

当接口变得拥塞时，网络设备必须将数据包存储在队列中或将其丢弃。如果路由器尝试将数据包存储在全队列中，则数据包将被丢弃。当流量从快速接口转发到较慢的接口时，将会导致数据包丢失。这显示在公式Q = u /(1-u)中，其中u是利用率，而Q是平均队列深度（假设的随机流量）。 因此，链路上的高利用率级别会导致较高的平均队列深度，如果您知道数据包大小，这是可预测的延迟。一些网络报告供应商表示，您可以为广域网订购更少的带宽，并且支付更少的费用。但是，当您以95%的利用率运行WAN链路时，将会出现延迟影响。此外，随着网络迁移到VoIP，网络管理员可能需要更改策略并以大约50%的利用率运行WAN链路。

当数据包被丢弃时，上层协议可能会强制重新传输数据包。如果丢弃了多个数据包，则可能会产生过多的重试流量。这种反应会导致设备沿线路进一步备份。为了解决此问题，您可以设置不同程度的阈值。

网络利用率的主要衡量标准是接口利用率。根据您测量的连接是半双工还是全双工，使用下表中描述的公式：

记法	描述
ΔifInOctets	收集snmp ifInOctets对象的两个轮询周期之间的差值（或差值），它表示流量的入站八位组计数。
ΔifOutOctets	收集snmp ifOutOctets对象的两个轮询周期之间的增量，该对象表示出站八位组流量的计数。
ifSpeed	snmp ifSpeed对象中报告的接口速度。请注意，ifSpeed可能无法准确反映WAN接口的速度。

共享LAN连接倾向于半双工，主要是因为争用检测要求设备在传输之前侦听。WAN连接通常为全双工，因为连接是点对点连接；两台设备可以同时发送和接收，因为它们知道只有一台其它设备共享连接。

由于MIB II变量存储为计数器，因此您必须执行两个轮询周期并计算两者之间的差值（因此公式中使用了Delta）。

对于半双工介质，使用以下公式计算接口利用率：

(ΔifInOctets + ΔifOutOctets)* 8 * 100

（Δ中的秒数）* ifSpeed

对于全双工介质，利用率计算更为复杂。例如，在完全T-1串行连接的情况下，线路速度为1.544 Mbps。这意味着T-1接口可以同时接收和传输1.544 Mbps带宽，总带宽为3.088 Mbps。

当您计算全双工连接的接口带宽时，可以使用此公式，在该公式中，您可以取较大的in和out值，并生成利用率百分比：

max(ΔifInOctets，(ΔifOutOctets)* 8 * 100

（Δ中的秒数）* ifSpeed

然而，这种方法隐藏了指令的使用率，它的价值和准确度较低。更准确的方法是分别测量输入利用率和输出利用率，例如：

输入利用率= ΔifInOctets *8 * 100

（Δ中的秒数）* ifSpeed

和

输出利用率= ΔifOutOctets *8 * 100

（Δ中的秒数）* ifSpeed

虽然这些公式稍加简化，但是没有考虑与特定协议相关的开销。存在更精确的公式来处理每个协议的独特方面。例如，RFC 1757包含考虑数据包开销的以太网使用公式。但是，高可用性团队发现，在大多数情况下，此处提供的通用公式在LAN和WAN接口上都可以可靠地使用。

容量规划

如前所述，容量规划是确定可能的未来网络资源需求以防止对业务关键型应用的性能或可用性影响的过程。请参阅容量和性能管理：最佳实践白皮书，了解有关此主题的更多详细信息。

执行主动故障分析

主动故障分析对性能管理至关重要。为性能管理收集的数据类型同样可用于主动故障分析。但是，主动式故障管理和性能管理之间的时间安排和使用情况有所不同。

主动故障管理是理想的网络管理系统实现您确定的目标的方式。与性能管理的关系是通过基线和使用的数据变量。主动式故障管理集成了自定义事件、事件关联引擎、故障通知单以及基线数据的统计分析，以便将故障、性能和变更管理结合到一个理想而有效的网络管理系统中。

如果性能数据轮询通常每10、15或甚至30分钟完成一次，则故障条件识别的时间间隔必须短得多。主动故障管理的方法之一是使用RMON警报和事件组。可以在未由外部设备轮询的设备上设置阈值，从而使阈值更短。本文档未介绍的另一种方法是，使用分布式管理系统在本地级别启用轮询，并在管理器的管理器中聚合数据。

使用阈值进行主动故障管理

阈值是在特定数据流中定义关注点并在触发阈值时生成事件的过程。使用网络性能数据设置这些阈值。

有几种不同类型的阈值，其中一些更适用于特定类型的数据。阈值仅适用于数字数据，因此可以将任何文本数据转换为离散的数字值。即使您不知道一个对象的所有可能的文本字符串，您仍然可以枚举“有趣的”字符串并将所有其他字符串分配到一个设置值。

两类数字数据有两种类型的阈值：连续和离散。连续阈值适用于连续数据或时序数据，例如存储在SNMP计数器或量规中的数据。离散阈值适用于枚举对象或任何离散数字数据。布尔对象是用两个值枚举的值：正确或错误。离散数据也称为事件数据，因为事件标记从一个值到下一个值的转换。

连续阈值可以在时序对象超过阈值指定值时触发事件。对象值高于阈值或低于阈值。设置单独的上升阈值和下降阈值也非常有用。这种技术称为迟滞机制，可帮助减少此类数据生成的事件数。迟滞机制用于减小由快速变化的时序数据上的阈值生成的事件量。此机制可用于时间序列数据的任何阈值技术。

事件数量通过生成以跟踪对象值的警报来减少。此警报具有上升和下降阈值。仅当超过上升阈值时才会触发警报。一旦超过此阈值，在超过下降阈值之前，不会再次生成上升警报。并且相同的机制可以防止产生下降阈值，直到再次超过上升阈值。此机制可以显着减少事件数量，而且无法消除确定是否存在故障所需的信息。

时序数据可以表示为计数器，其中每个新数据点被添加到先前数据点的总和，或者可以表示为量规，其中数据被表示为一段时间间隔的速率。适用于每种数据类型的连续阈值有两种不同的形式：绝对连续阈值和相对连续阈值。对仪表使用绝对连续阈值，对计数器使用相对连续阈值。

要确定网络的阈值，请完成以下步骤：

1. 选择对象。
2. 选择设备和接口。
3. 确定每个对象或对象/接口类型的阈值。
4. 确定每个阈值生成的事件的严重性。

要确定在哪些对象（以及哪些设备和接口）上使用哪些阈值，需要进行大量工作。 幸运的是，如果您收集了性能数据的基线，您已经完成大量该工作。此外，NSA和高可用性服务(HAS)程序可以提出一些建议，帮助您设置对象和创建范围。但是，您必须为您的特定网络定制这些建议。

当您收集了网络的性能数据后，HAS程序建议您按类别对接口进行分组。这样可以简化阈值设置，因为您可能需要确定每个类别的媒体类型的阈值，而不是确定该设备上的每个设备和对象的阈值。例如，您想要为以太网和FDDI网络设置不同的阈值。一般认为，与共享以太网网段相比，运行FDDI网络的利用率更接近100%。但是，由于全双工以太网不会发生冲突，因此其运行速度可能接近100%的利用率。您可能需要将全双工链路的冲突阈值设置为非常低的值，因为您永远不会看到冲突。

您还可以考虑接口重要性和阈值类型的类别/严重性的组合。使用这些因素设置事件的优先级，从而设置事件的重要性以及网络操作人员的关注度。

对网络设备和接口进行分组和分类的重要性怎么强调都不为过。对阈值事件进行分组和分类的能力越强，就越容易将阈值事件集成到网络管理平台中。使用基线作为此信息的主要资源。请参阅容量和性能管理：最佳实践白皮书，了解更多信息。

网络管理实施

组织应实施网络管理系统，能够检测定义的阈值并报告指定时间段的值。使用RMON网络管理系统，可以将阈值消息存档到日志文件中，以便每天查看，也可以使用更完整的数据库解决方案，允许搜索给定参数的阈值异常。该信息应该能够持续地提供给网络操作人员和经理。网络管理实施应能够检测软件/硬件崩溃或回溯、接口可靠性、CPU、链路利用率、队列或缓冲区未命中、广播量、载波转换和接口重置。

网络运营指标

与性能管理重叠的最后一个主动故障管理领域是网络运营指标。这些指标为改进故障管理流程提供了宝贵的数据。至少，这些指标应该包括给定时间段内发生的所有问题的明细。细目应包括如下信息：

- 按呼叫优先级划分的问题数
- 每个优先级中的最小、最大和平均关闭时间
- 按问题类型（硬件、软件崩溃、配置、电源、用户错误）细分问题
- 每种问题类型的关闭时间细分
- 可用性按可用性组或SLA
- 您达到或错过SLA要求的频率

帮助台通常拥有能够生成指标或报告的报告系统。收集此数据的另一种方法是使用可用性监控工具。总体指标应每月提供。应根据讨论实施流程改进，以改进未达到服务级别协议要求或改进某些问题类型的处理方式。

绩效管理指标

绩效指标是组织用来衡量关键成功因素的机制。

记录网络管理业务目标

本文档可能是网络管理的正式操作概念，也可能是所需功能和目标的非正式说明。但是，在衡量成功与否时，本文档应该为网络管理员提供帮助。

本文档为组织网络管理策略，应协调网络运营、工程、设计、其他业务部门和最终用户的总体业务（非定量）目标。此重点使组织能够形成网络管理和运营的远程规划活动，其中包括预算流程。它还指导您如何获得实现网络管理目标（例如SLA）所需的工具和集成路径。

本战略文件不能过于狭隘地侧重于特定网络问题的管理，而是侧重于对整个组织而言重要的项目，其中包括预算问题。例如：

- 确定具有可实现目标的综合计划。
- 确定需要网络支持的每项业务服务/应用。
- 确定衡量服务所需的基于性能的指标。
- 规划性能度量数据的收集和分发。
- 确定网络评估和用户反馈所需的支持。
- 有文档记录、详细且可衡量的服务级别目标。

记录服务级别协议

为了正确记录SLA，您必须完全定义服务级别目标指标。用户应可使用本文档进行评估。它提供反馈环路，确保网络管理组织继续测量保持服务协议级别所需的变量。

SLA是“活的”文档，因为业务环境和网络在本质上是动态的。今天用来衡量SLA的方法可能在明天就过时了。只有用户形成反馈环路，并根据该信息采取行动，网络运营才能保持组织所需的高可用性数量。

创建基线的变量列表

此列表包括诸如轮询间隔、产生的网络管理开销、可能的触发阈值、变量是否用作陷阱的触发器以及针对每个变量使用的趋势分析等项目。

这些变量不限于上述服务级别目标所需的指标。它们至少应包括以下变量：路由器运行状况、交换机运行状况、路由信息、特定于技术的数据、利用率和延迟。这些变量会定期轮询并存储在数据库中。然后可以根据此数据生成报告。这些报告可通过以下方式帮助网络管理运营和规划人员：

- 使用历史数据库通常可以更快地解决被动问题。
- 性能报告和容量规划需要此类数据。
- 服务级别目标可以根据它来衡量。

查看基线和趋势分析

网络管理人员应召开会议，定期查看具体报告。这样可以提供更多反馈，并主动解决网络中的潜在问题。

这些会议应包括业务人员和规划人员。这为规划人员提供了接收基准数据和趋势数据的操作分析的机会。它还使操作人员进入环路，进行一些规划分析。

这些会议中要包括的另一个项目类型是服务级别目标。在接近目标阈值时，网络管理人员可以采取措施防止错过目标，而且在某些情况下，这些数据可以作为部分预算理由使用。数据可以显示，如果不采取适当的措施，服务级别目标将遭到哪些破坏。此外，由于业务服务和应用程序已确定了这些目标，因此更容易在财务上证明其合理性。

每两周进行一次审核，每六至十二周举行一次更全面的分析会议。通过这些会议，您可以解决短期和长期问题。

记录假设分析方法

假设分析涉及对解决方案的建模和验证。在向网络添加新解决方案（新应用或Cisco IOS版本中的更改）之前，请记录一些替代方案。

此分析的文档包括主要问题、方法、数据集和配置文件。要点是，假设分析是一个实验，其他人应该能使用文档提供的信息重新创建它。

记录用于提高网络性能的方法

本文档包括额外的WAN带宽和成本表，可帮助增加特定链路类型的带宽。此信息可帮助组织了解增加带宽所需的时间和成本。正式文档使性能和容量专家能够了解提高性能的方式和时间，以及提高性能的时间表和成本。

定期查看此文档，可能作为季度绩效审核的一部分，以确保其保持最新。

摘要

实现理想的网络管理系统的唯一途径就是将性能管理的各个组件主动地集成到系统中。此目标应该包括使用在阈值超出时绑定到通知系统的可用性和响应时间指标。它必须包括容量规划基准的使用，该基准将具有到调配和异常报告启发式模型的链接。它可能有一个内置的建模或模拟引擎，使模型能够实时更新，并通过软件模拟提供一定级别的规划和故障排除。

虽然这个系统的大部分看似不可能实现的理想，但其中的每个组件目前都已上市。此外，用于集成这些组件的工具也存在于MicroMuse等程序中。我们应该继续朝着这个理想努力，因为今天这个理想比以往任何时候都更加现实。

相关信息

- 高可用性技术白皮书
- 容量和性能管理白皮书
- 技术支持和文档 - Cisco Systems

关于此翻译
思科采用人工翻译与机器翻译相结合的方式将此文档翻译成不同语言，希望全球的用户都能通过各自的语言得到支持性的内容。

请注意：即使是最好的机器翻译，其准确度也不及专业翻译人员的水平。

Cisco Systems, Inc. 对于翻译的准确性不承担任何责任，并建议您总是参考英文原始文档（已提供链接）。

关于此翻译
思科采用人工翻译与机器翻译相结合的方式将此文档翻译成不同语言，希望全球的用户都能通过各自的语言得到支持性的内容。

请注意：即使是最好的机器翻译，其准确度也不及专业翻译人员的水平。

Cisco Systems, Inc. 对于翻译的准确性不承担任何责任，并建议您总是参考英文原始文档（已提供链接）。