

# 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](#)

[制定网络运营管理概念](#)

[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](#)

[为基线创建变量列表](#)

[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

性能管理包括优化网络服务响应时间以及管理单个网络服务和整体网络服务的一致性和质量。最重要的服务是测量用户/应用响应时间的需要。对于大多数用户而言，响应时间是关键的性能成功因素。此变量决定了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。在容量规划领域，网络基线（CPU、内存、缓冲区、输入/输出八位组等）可能影响响应时间。因此，请记住，性能问题通常与容量相关。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，虽然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，但长期而言，他们倾向于寻求短期经济收益，并面临（通常无法预见的）更高成本的风险。在每个预算周期中，网络管理员和项目人员都在努力在性能和快速实施之间找到平衡。此外，网络管理员还面临一些挑战，包括快速开发产品以满足狭窄的市场窗口、复杂的技术、业务整合、竞争市场、计划外停机、缺乏专业知识，以及通常工具不足。

鉴于这些挑战，性能在网络管理框架中如何适应？理想的网络管理系统的主要功能是优化网络的运

行能力。一旦您将此作为网络管理的最终目标，网络管理的重点就是保持网络以最佳性能运行。

理想的网络管理系统包括以下主要操作：

- 通知操作员性能即将恶化。
- 在性能恶化或故障发生时，提供简单的替代路由和解决方法。
- 提供工具，以查明性能恶化或故障的原因。
- 充当网络恢复力和生存能力的主站。
- 实时传达性能。

根据理想系统的这一定义，性能管理对网络管理至关重要。以下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

必须注意的是，对于语音和视频等较新的应用，性能是成功的关键变量，如果无法实现一致的性能，则服务会被视为价值低且失败。在其他情况下，用户会因应用程序间歇性超时而遭受性能变化的影响，从而降低工作效率和用户满意度。

本文档详细介绍了最关键的性能管理问题，包括关键成功因素、关键性能指标和性能管理的高级流程图。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

关键成功因素确定实施最佳实践的要求。为了获得关键成功因素的资格，流程或程序必须提高可用性，否则程序的缺失必须降低可用性。此外，关键成功因素应可衡量，以便组织能够确定其成功程度。

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：

- 收集网络和应用数据的基线。
- 对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。
- 定期查看网络 and 应用的容量信息，以及基准和异常。
- 设置升级或调整程序，以便在被动和长期的基础上处理容量问题。

## 绩效管理指标

绩效指标提供了组织衡量关键成功因素的机制。绩效规划的绩效指标包括：

- 记录网络管理业务目标。这可能是网络管理运营的正式概念，也可能是有关所需功能和目标的不太正式的声明。
- 创建详细而可衡量的服务级别目标。

- 提供服务级别协议的文档，其中包含图表或图，这些图表显示了这些协议在一段时间内的履行方式是否成功。
- 收集基线的变量列表，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器以及针对每个变量使用的趋势分析。
- 定期召开会议，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）。
- 当阈值超过时，请制作有关用于增加网络资源的方法的文档。要记录的一项是增加WAN带宽和成本表所需的时间行。

## 性能管理流程

以下步骤为性能管理提供了高级流程：

1. [制定网络运营管理概念定义所需功能：服务、可扩展性和可用性目标定义可用性和网络管理目标定义性能SLA和指标定义SLA](#)
2. [衡量绩效收集网络基线数据评估可用性测量响应时间测量精度衡量利用率容量规划](#)
3. [执行主动故障分析使用阈值进行主动故障管理网络管理实施网络运营指标](#)

## 制定网络运营管理概念

在定义网络的详细性能和容量变量之前，必须先了解组织内网络管理的整体运行概念。当您定义这一整体概念时，它为您构建网络所需功能的精确定义提供了业务基础。如果您未能制定网络管理的运营概念，可能会导致缺乏目标或目标，这些目标或目标会因客户需求而不断变化。

通常，在网络管理程序的系统定义阶段，您会将网络管理操作概念作为第一步。其目的是从操作角度描述总体期望的系统特征。本文档的用途是协调网络运营、工程、设计、其他业务单位和最终用户的整体业务（非定量）目标。本文档的重点是为网络管理和运营组织远程运营规划活动。它还为开发所有后续定义文档（如服务级别协议）提供指导。这组初始定义显然不能过于狭隘地关注特定网络问题的管理，而是强调对整体组织的重要性以及与必须管理的成本相关的项目。目标包括：

- 确定高效使用网络基础设施所必须具备的特征。
- 确定网络支持的服务/应用。
- 启动端到端服务管理。
- 启动基于绩效的指标以改进整体服务。
- 收集并分发性能管理信息。
- 通过用户反馈支持网络战略评估。

换句话说，网络管理运营理念应侧重于整体组织目标和您实现这些目标的理念。主要要素包括对任务、任务目标、系统目标、组织参与和总体运营理念的更高级别定义。

作为网络经理，您可以统一用户通常不一致的性能期望。例如，如果网络的主要要求是从一个位置向另一个位置传输大型文件，则您希望将重点放在高吞吐量上，而较少关注交互式用户的响应时间。请注意，除非考虑各种问题，否则不要限制您对性能的看法。例如，在测试网络时，请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的数据包的吞吐量。这些性能测试中的任一项都可能产生非常积极的效果，但根据您的网络流量负载，这些测试可能无法提供真实的性能图景。研究尽可能多的工作负载条件下的网络性能，并记录性能。

此外，虽然许多网络管理组织都有有效的警报技术来通知技术人员设备故障，但为端到端应用性能定义和实施评估流程却更加困难。因此，虽然网络运营中心(NOC)可以快速响应路由器或交换机故障，但可能影响网络性能和影响用户感知的网络条件可能很容易被忽略，直到这种感觉变为负面为

止。无论多么困难，第二个流程都可以为业务组织和网络管理带来巨大的好处。

最后，确保不要对网络性能产生不切实际的期望。如果您误解了网络协议或应用的详细信息，通常会产生不切实际的期望。通常，性能不佳不是网络故障，而是应用设计不佳的结果。记录和衡量应用性能的唯一方法是在安装应用之前确定网络性能基线。

## 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。此信息的首次使用是确定应用对组织目标的重要性。您还可以应用此信息创建知识库以用于逻辑设计，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。

创建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，并确保建议的设计不会因网络增长或扩展而受到资源限制。资源限制包括：

- 总流量
- 音量
- 路由数
- 虚电路数
- 邻居计数
- 广播域
- 设备吞吐量
- 介质容量

网络规划人员应确定设计所需的寿命、设计所需的扩展时间或站点、新用户数量以及预期流量或更改。此计划有助于确保建议的解决方案在设计的预计寿命内满足增长需求。

如果您不调查解决方案的可扩展性，可能会被迫实施重大的被动式设计更改。此设计更改可能包括其他层次结构、介质升级或硬件升级。在依赖相当精确的预算周期进行主要硬件采购的组织中，这些变化可能是整体成功的主要制约因素。在可用性方面，网络可能遇到意外的资源限制，导致出现非可用性和被动措施。

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。互操作性可以指不同的硬件供应商，或在网络实施期间或之后必须相互结合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发送硬件信令，以解决路由或传输问题。在网络解决方案迁移之前、期间或之后，可能会发生互操作性问题。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发生的拓扑问题。

解决方案比较是一种实践，在这种实践中，您可以比较与其他解决方案要求实践相关的不同潜在设计。此实践有助于确保解决方案最适合特定环境，并且个人偏见不推动设计流程。比较可能包括不同的因素，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。实施设计后，所有这些都可能对网络整体可用性产生重大影响。您还可以比较介质、层次结构、冗余、路由协议和类似功能。创建一个图表，其中包含X轴上的因素和Y轴帮助上的潜在解决方案，以总结解决方案比较。在实验室环境中进行详细的解决方案比较还有助于客观地研究与不同比较因素相关的新解决方案和功能。

作为网络管理运营概念的一部分，以所有用户都能理解的方式定义网络和支持服务的目标至关重要。操作概念发展后的活动受到该文件质量的很大影响。

以下是标准绩效目标：

- 响应时间

- 使用
- 吞吐量
- 容量 ( 最大吞吐率 )

虽然这些测量对于简单的LAN而言可能微不足道，但在交换园区网络或多供应商企业网络中，这些测量非常困难。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，应用“x”在高峰工作时间的最短响应时间为500毫秒或更短。这定义了用于识别变量的信息、度量变量的方式以及网络管理应用程序应关注的一天中的某个时间段。

## [定义可用性和网络管理目标](#)

可用性目标定义了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，并详细说明每个类别的网络要求，以满足可用性要求。网络的不同区域可能还需要不同的可用性级别。更高的可用性目标可能需要增加冗余和支持程序。当您定义特定网络服务的可用性目标并衡量可用性时，您的网络组织可以了解实现预计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。为了设置可管理性目标，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应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实现可用性目标至关重要。

可管理性目标应揭示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、针对新服务的人员配备模式以及任何其他支持要求。通常，在部署前未发现此信息，而且由于分配了用于支持新网络设计的资源不足，整体可用性受到影响。

## [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。建议的解决方案的性能可以通过性能监控工具或通过建议的网络基础设施的简单ping来衡量。性能SLA应包括平均预期流量、峰值流量、平均响应时间和允许的最大响应时间。然后，这些信息可以稍后在解决方案验证部分中使用，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您为用户或客户定义服务时。企业称这些服务级别协议为服务级别管理，而服务提供商称其为服务级别管理。服务级别管理通常包括问题类型和严重性的定义以及帮助台职责，例如上报路径和每个层支持级别上报前的时间、开始处理问题的时间以及根据优先级关闭目标的时间。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。

当组织不预先定义服务级别时，很难改进或获得以后确定的资源需求。此外，很难理解为了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后才应用。

## [衡量绩效](#)

绩效管理是一个综合术语，它包含不同绩效领域的配置和衡量。本节介绍以下六个绩效管理概念：

- [收集网络基线数据](#)
- [评估可用性](#)
- [测量响应时间](#)

- [测量精度](#)
- [衡量利用率](#)
- [容量规划](#)

## [收集网络基线数据](#)

大多数企业内部网具有足够的带宽。但是，如果数据不足，您可能无法排除网络拥塞是导致应用性能下降的因素。拥塞或错误的一个线索是性能不佳是间歇性的还是取决于时间的。这种情况的一个例子是，在深夜时表现足够，但在早上和高峰时段表现非常缓慢。

定义网络管理运营概念并定义所需的实施数据后，就需要随着时间的推移收集这些数据。此类收集是网络基线的基础。

在部署新解决方案（应用或IOS更改）之前和部署之后，执行当前网络的基线，以衡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。根据运营概念中定义的目标，您可能还包括其它类型的基线数据。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基线，在您对应用网络需求趋势化时，它非常有价值。此信息可用于升级周期中的计费 and/或预算。在应用可用性方面，应用基准对于每个应用的首选服务或服务质量也很重要。应用基线信息主要包括应用每个时间段使用的带宽。某些网络管理应用程序还可以确定应用程序性能的基线。流量类型（Telnet或FTP）的细分对规划也很重要。在某些组织中，网络中资源受限的区域会针对最大流量生成者进行监控。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，可以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的时间。从网络角度来看，可用性表示网络中各个组件的可靠性。

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

网络冗余是衡量可用性时要考虑的另一个因素。冗余丢失表示服务降级，而非网络故障。结果可能是响应时间较慢，以及由于丢包而丢失数据。结果也可能出现在其它性能测量领域，如利用率和响应时间。

最后，如果您根据SLA提供服务，您应考虑计划内停机。这些中断可能是移动、添加和更改、工厂停工或您可能不想报告的其他事件的结果。这不仅是一项艰巨的任务，而且可能是一项手动任务。

## [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。响应时间比正常时间慢（通过基线比较或超过阈值）可能表示拥塞或网络故障。

响应时间是衡量客户网络使用情况的最佳指标，可帮助您评估网络的有效性。无论响应缓慢的原因是什么，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
- 到达目的地的路由少于期望路由（或根本没有路由）

- 网络设备供电不足
- 网络故障，例如广播风暴
- 噪音或CRC错误

在采用QoS相关排队的网络中，响应时间测量对于确定正确类型的流量是否按预期通过网络非常重要。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以恒定的速率传输，以保持好的语音质量。您可以生成分类为语音流量的流量，以测量流量对用户显示时的响应时间。

您可以测量响应时间，以帮助解决应用程序服务器和网络管理器之间的冲突。当应用程序或服务器速度缓慢时，通常会假定网络管理员有罪。网络管理员必须证明网络不是问题。响应时间数据收集为证明或否定网络是应用故障的根源提供了无可争辩的手段。

您应尽可能测量对用户显示的响应时间。用户将响应视为从按Enter键或单击按钮到屏幕显示的时间。此时间包括处理流量所需的每个网络设备、用户工作站和目的服务器的时间。

遗憾的是，由于用户数量和工具的缺乏，在此级别进行测量几乎是不可能的。此外，当您用户和服务器响应时间纳入考虑范围时，在确定未来网络增长或排除网络故障时，它几乎不提供任何价值。

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

在简单化的层面，您可以将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的响应时间定为一定时间，以测量响应时间。这种方法的问题在于它不能准确反映用户对机器和目的机器之间响应时间的感知。它只从网络管理站的角度收集信息并报告响应时间。此方法还可以逐跳地屏蔽整个网络中的响应时间问题。

替代以服务器为中心的轮询的方法是将工作分布在更接近您希望模拟的测量源和目标的位置。使用分布式网络管理轮询器并实施Cisco IOS服务保障代理(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（如服务器或其他路由器）之间的响应时间。您还可以指定TCP或UDP端口，该端口强制以与模拟流量相同的方式转发和定向流量。

随着语音、视频和数据在多服务网络上的集成，客户在其网络中实施QoS优先级。由于不同应用程序接收不同的优先级，因此简单的ICMP或UDP测量不能准确反映响应时间。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可能会在每台路由器处理优先级的方式上收到不同的优先级，并且可能会收到不同、效率较低的路由。

在这种情况下，衡量响应时间的唯一方法是生成类似于特定应用或相关技术的流量。这会迫使网络设备像处理实际流量一样处理流量。您可能能够通过SAA或使用第三方应用感知探测器来达到此级别。

## 测量精度

准确性是不导致错误的接口流量的度量，可以用百分比表示，该百分比将成功率与一段时间内的总数据包速率进行比较。您必须首先测量错误率。例如，如果每100个数据包中有2个数据包出错，则错误率为2%，准确率为98%。

在早期的网络技术中，特别是在广域网中，可以接受一定程度的错误。但是，在高速网络和当今的WAN服务中，传输要精确得多，而且错误率接近于零，除非存在实际问题。接口错误的一些常见原因包括：

- 不符合规格的布线
- 电子干扰

- 硬件或软件故障

使用降低的准确率来触发更深入的调查。您可能会发现特定接口出现问题，并确定错误是可接受的。在这种情况下，您应调整此接口的精度阈值，以反映错误率不可接受的位置。不可接受的错误率可能已在较早的基线中报告。

下表中描述的变量用于精度和错误率公式：

记法	描述
$\Delta$ ifInErrors	收集snmp ifInErrors对象的两个轮询周期之间的增量（或差值），该对象表示出错的进站数据包的计数。
$\Delta$ ifInUcastPkts	收集snmp ifInUcastPkts对象（表示进站单播数据包的计数）的两个轮询周期之间的增量。
$\Delta$ ifInNUcastPkts	收集snmp ifInNUcastPkts对象的两个轮询周期之间的增量，该对象表示进站非单播数据包（组播和广播）的计数。

错误率的公式通常以百分比表示：

$$\text{错误率} = (\Delta \text{ifInErrors}) * 100$$

—

$$(\Delta \text{ifInUcastPkts} + \Delta \text{ifInNUcastPkts})$$

请注意，在错误率和准确性公式中未考虑出站错误。这是因为设备永远不应该有意将带有错误的数据包放在网络上，出站接口错误率也不应该增加。因此，进站流量和错误是接口错误和准确性唯一关注的指标。

精度公式采用错误率，并从100中减去（同样，以百分比形式）：

$$\text{精度} = 100 - (\Delta \text{ifInErrors}) * 100$$

—

$$(\Delta \text{ifInUcastPkts} + \Delta \text{ifInNUcastPkts})$$

这些公式反映了MIB II接口(RFC 2233)通用计数器的错误和准确性。结果以将错误与查看和发送的数据包总数进行比较的百分比表示。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，因此您必须采用两个轮询周期并计算两个轮询周期之间的差异（因此在等式中使用Delta）。

## 衡量利用率

利用率衡量特定资源在一段时间内的使用情况。该度量通常以资源的使用与其最大操作能力进行比较的百分比形式表示。通过利用率措施，您可以识别整个网络的拥塞（或潜在的拥塞）。您还可以找出未充分利用的资源。

利用率是确定网络管道（链路）满度的主要指标。测量CPU、接口、队列和其他与系统相关的容量

测量，以确定网络系统资源消耗的程度。

高利用率并不一定是坏的。低利用率可能表示流量在意外位置传输。随着行的过度使用，其影响将变得显著。当排队通过接口的流量超过其可处理的流量时，会发生过度利用。资源利用率的突然跳变可能表示出现故障。

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丢弃该数据包。如果路由器尝试将数据包存储在完整队列中，则会丢弃该数据包。当流量从快速接口转发到较慢接口时，会导致丢包。这在公式  $Q = u / (1-u)$  中表示，其中  $u$  是利用率， $Q$  是平均队列深度（假设随机流量）。因此，链路上的高利用率级别会导致较高的平均队列深度，这在您知道数据包大小时是可预测的延迟。一些网络报告供应商表示，您可以订购更少的带宽，而且WAN的费用更低。但是，当您以95%的利用率运行WAN链路时，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其策略并以大约50%的利用率运行广域网链路。

当丢弃数据包时，较高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。这种类型的反应可能导致设备在线路更下方进行备份。要解决此问题，可以设置不同的阈值。

用于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
$\Delta ifInOctets$	收集snmp ifInOctets对象（表示流量的进站八位组计数）的两个轮询周期之间的增量（或差异）。
$\Delta ifOutOctets$	收集snmp ifOutOctets对象（表示流量的出站八位组计数）的两个轮询周期之间的增量。
ifSpeed	snmp ifSpeed对象中报告的接口速度。请注意ifSpeed可能不能准确反映WAN接口的速度。

共享LAN连接通常为半双工，这主要是因为争用检测要求设备在传输之前先侦听。WAN连接通常为全双工，因为连接是点对点的；两台设备可以同时发送和接收，因为它们知道只有另一台设备共享连接。

由于MIB II变量存储为计数器，因此您必须采用两个轮询周期并计算两个轮询周期之间的差异（因此在等式中使用Delta）。

对于半双工介质，使用以下公式计算接口利用率：

$$(\Delta ifInOctets + \Delta ifOutOctets) * 8 * 100$$

—

$$(\text{秒数(以}\Delta\text{表示)} * ifSpeed$$

对于全双工介质，利用率计算更为复杂。例如，使用完整的T-1串行连接时，线速为1.544 Mbps。这意味着T-1接口可以接收和传输1.544 Mbps的带宽，总带宽为3.088 Mbps。

当计算全双工连接的接口带宽时，可以使用以下公式，在此公式中，您将取较大的输入和输出值并生成利用率百分比：

$$\max(\Delta ifInOctets, (\Delta ifOutOctets) * 8 * 100$$

---

(秒数(以Δ表示) \* 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。一种更精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 =  $\Delta \text{InOctets} * 8 * 100$

---

(秒数(以Δ表示) \* ifSpeed

和

输出利用率 =  $\Delta \text{OutOctets} * 8 * 100$

---

(秒数(以Δ表示) \* ifSpeed

虽然这些公式稍加简化，但它们不考虑与特定协议相关的开销。使用更精确的公式来处理每个协议的独特方面。例如，RFC 1757包含以太网使用公式，该公式考虑了数据包开销。但是，高可用性团队发现，在大多数情况下，此处介绍的通用公式可在LAN和WAN接口之间可靠地使用。

## 容量规划

如前所述，容量规划是确定未来可能的网络资源需求以防止性能或可用性对业务关键型应用的影响的过程。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。为性能管理而收集的相同类型的数据可用于主动故障分析。但是，主动故障管理和性能管理之间，此数据的时间和使用有所不同。

主动故障管理是理想的网络管理系统实现您确定目标的方式。绩效管理的关系是通过您使用的基线和数据变量。主动式故障管理将定制事件、事件关联引擎、故障通知单和基线数据的统计分析集成在一起，以便将故障、性能和变更管理结合到一个理想、有效的网络管理系统中。

如果性能数据轮询通常每10、15甚至30分钟完成一次，则对故障状况的识别必须以更短的时间间隔进行。主动故障管理的一种方法是使用RMON警报和事件组。您可以在未由外部设备轮询的设备上设置阈值，以便阈值更短。本文档未介绍的另一种方法是使用分布式管理系统，该系统支持在本地级别进行轮询，并在管理器管理器中聚合数据。

## 使用阈值进行主动故障管理

阈值是在触发阈值时定义特定数据流中关注点并生成事件的过程。使用网络性能数据设置这些阈值。

阈值有几种不同的类型，其中有些更适用于某些类型的数据。阈值仅适用于数字数据，因此可将任何文本数据转换为离散的数值。即使您不知道对象的所有可能文本字符串，您仍然可以枚举“感兴趣

的”字符串并将所有其他字符串分配给设置值。

两类数字数据有两类阈值：连续和离散。连续阈值适用于连续或时间系列数据，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。布尔对象是具有两个值的枚举值：正确或错误。离散数据也可称为事件数据，因为事件标记从一个值到下一个值的转换。

连续阈值可在时间序列对象超过阈值的指定值时触发事件。对象值要么高于阈值，要么低于阈值。另外，设置不同的上升和下降阈值也很有用。此技术称为滞后机制，有助于减少从此类数据生成的事件数。滞后机制用于减小由快速变化的时间序列数据的阈值产生的事件量。该机制可以与任何阈值技术一起用于时间序列数据。

通过产生用于跟踪对象值的警报来减少事件量。此警报将分配上升和下降阈值。仅当超出上升阈值时才会触发警报。一旦超过该阈值，在超过该下降阈值之前不会再次生成上升警报。同样的机制防止产生下降警报，直到再次超过上升阈值。此机制可以显著减少事件数量，并且不会消除确定是否存在故障所需的信息。

时间序列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中，也可以表示为量规，其中数据在时间间隔内以速率表示。每种数据类型有两种不同形式的连续阈值：绝对连续阈值和相对连续阈值。使用带量规的绝对连续阈值和带计数器的相对连续阈值。

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）上使用什么阈值，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，您已经完成了大量工作。此外，NSA和高可用性服务(HAS)计划可以提出帮助您设置对象和创建范围的建议。但是，您必须针对您的特定网络定制这些建议。

当您收集了网络的性能数据时，HAS程序建议您按类别对接口进行分组。这简化了设置阈值，因为您可能需要为每个类别的介质类型而不是该设备和对象确定阈值。例如，您需要为以太网和FDDI网络设置不同的阈值。通常认为，与共享以太网网段相比，您可以以接近100%的利用率运行FDDI网络。但是，全双工以太网可以运行更接近100%的利用率，因为它们不会发生冲突。您可能希望将全双工链路的冲突阈值设置得非常低，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。使用这些因素来设置事件的优先级，从而确定事件的重要性及其受到网络运营人员的关注。

对网络设备和接口进行分组和分类的强调不过分。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](#)，了解详细信息。

## 网络管理实施

组织应该有一个实施的网络管理系统，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，该系统可以将阈值消息存档到日志文件中，以供日常查看，或使用更完整的数据库解决方案来搜索给定参数的阈值异常。网络运营人员和经理应持续了解相关信息。网络管理实施应包括检测软件/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的能力。

## [网络运营指标](#)

与性能管理重叠的主动故障管理的最后一个领域是网络运营指标。这些指标为故障管理流程改进提供了有价值的信息。至少，这些指标应包括给定时间段内发生的所有问题的明细。细目应包括以下信息：

- 按呼叫优先级发生的问题数
- 在每个优先级中关闭的最短、最长和平均时间
- 按问题类型（硬件、软件崩溃、配置、电源、用户错误）划分问题
- 每种问题类型的关闭时间细分
- 按可用性组或SLA的可用性
- 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。总体指标应按月提供。应实施基于讨论的流程改进，以改进未满足的服务级别协议要求或改进处理某些问题类型的方式。

## [绩效管理指标](#)

绩效指标为组织衡量关键成功因素提供了机制。

## [记录网络管理业务目标](#)

本文档可能是网络管理操作的正式概念，也可能是有关所需功能和目标的不太正式的陈述。但是，文档应帮助网络经理衡量成功。

本文档是组织网络管理策略，应协调网络运营、工程、设计、其他业务单位和最终用户的整体业务（非定量）目标。此重点使组织能够形成网络管理和运营的长期规划活动，包括预算流程。它还为获取工具和实现网络管理目标（如SLA）所需的集成路径提供指导。

本战略文档不能过于狭隘地关注特定网络问题的管理，而只关注对整个组织而言非常重要的项目，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绩效的指标。
- 规划性能指标数据的收集和分发。
- 确定网络评估和用户反馈所需的支持。
- 具有记录、详细和可衡量的服务级别目标。

## [记录服务级别协议](#)

要正确记录SLA，您必须完全定义服务级别目标指标。此文档应供用户评估。它提供反馈环路，以确保网络管理组织继续测量维护服务协议级别所需的变量。

SLA是“活”文档，因为业务环境和网络本质上是动态的。今天用来衡量SLA的功能可能会在明天过时。只有当他们从用户那里建立反馈环路并对该信息采取行动时，网络运营才能保持组织所需的高可用性编号。

## [为基线创建变量列表](#)

此列表包括以下项：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。

这些变量不限于上述服务级别目标所需的指标。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、技术特定数据、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操作和规划人员：

- 通过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进行衡量。

## [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。这可提供额外的反馈，并主动解决网络中的潜在问题。

这些会议应包括操作人员和规划人员。这为规划人员提供了接收基线和趋势数据运营分析的机会。它还使运营人员“处于循环状态”，以便进行一些规划分析。

这些会议中要包括的另一种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预防措施防止目标丢失，在某些情况下，此数据可用作部分预算理由。数据可以显示如果不采取适当措施，将在何处违反服务级别目标。此外，由于这些目标已通过业务服务和应用确定，因此在财务上更容易证明其合理性。

每两周进行一次审核，每六至十二周举行一次更彻底的分析会议。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进行建模和验证。在向网络添加新解决方案（新应用或Cisco IOS版本的更改）之前，请记录一些备选方案。

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，假设分析是一种实验，其他人应该能够使用文档中提供的信息重新创建。

## [记录用于提高网络性能的方法](#)

本文档包括额外的WAN带宽和有助于增加特定类型链路带宽的成本表。此信息有助于组织了解增加带宽需要花费多少时间和金钱。正式文档使性能和容量专家能够了解如何和何时提高性能，以及此类工作的时间线和成本。

定期查看此文档（可能是季度绩效评估的一部分），以确保文档保持最新。

## [摘要](#)

实现理想网络管理系统目标的唯一途径就是将性能管理的组件主动集成到系统中。此目标应包括在超出阈值时使用与通知系统关联的可用性和响应时间指标。它必须包括使用容量规划的基线，该基线将具有指向调配和异常报告的启发式模型的链接。它可以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的规划和故障排除级别。

虽然此系统中的许多组件看起来都是不可能实现的理想，但每个组件目前都可用。此外，集成这些组件的工具也存在于MicroMuse等程序中。我们应继续朝着这个理想努力，因为它现在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](#)