

# 容量和性能管理：最佳实践白皮书

## 目录

[简介](#)

[容量和性能管理概述](#)

[CPU](#)

[底板或I/O](#)

[内存](#)

[接口和管道尺寸](#)

[队列、延迟和抖动](#)

[速度和距离](#)

[应用特征](#)

[容量和性能管理最佳实践](#)

[服务级别管理](#)

[网络和应用假设分析](#)

[基线化和趋势](#)

[例外管理](#)

[QoS管理](#)

[收集和报告容量信息](#)

[确定您的需求](#)

[定义流程](#)

[定义容量区域](#)

[定义容量变量](#)

[解释数据](#)

[相关信息](#)

## 简介

在大型企业和服务提供商网络中，高网络可用性是一项关键任务要求。网络经理在提供更高可用性方面面临着日益严峻的挑战，包括计划外的停机时间、缺乏专业知识、工具不足、技术复杂、业务整合以及市场竞争。容量和性能管理可帮助网络经理实现新的全球业务目标，并实现一致的网络可用性和性能。

本文档探讨以下主题：

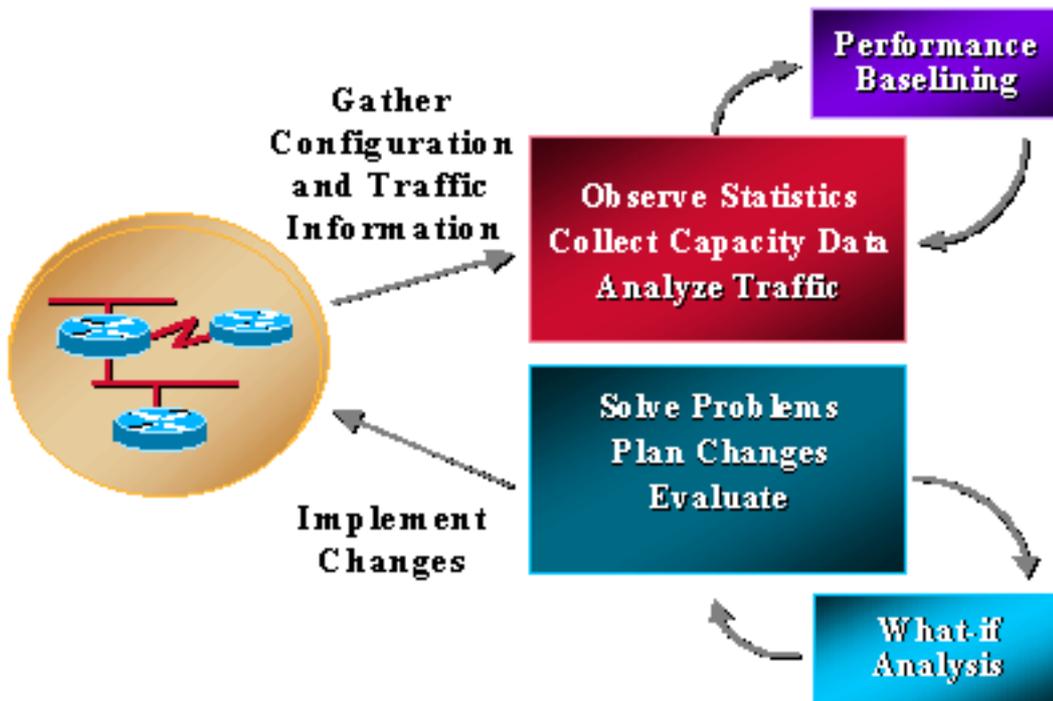
- 一般容量和性能问题，包括网络中的风险和潜在的容量问题。
- 容量和性能管理最佳实践，包括假设分析、基线建立、趋势分析、异常管理和QoS管理。
- 如何制定容量规划策略，包括容量规划中使用的常见技术、工具、MIB变量和阈值。

## 容量和性能管理概述

容量规划是确定网络资源所需的过程，以防止对关键业务应用的性能或可用性产生影响。性能管理是对单个服务和整体服务的网络服务响应时间、一致性和质量进行管理的实践。

**注意：**性能问题通常与容量有关。应用速度较慢，因为带宽和数据在通过网络传输之前必须在队列中等待。在语音应用中，延迟和抖动等问题直接影响语音呼叫的质量。

大多数组织已经收集了一些与容量相关的信息，并始终如一地解决问题、规划更改和实施新的容量和性能功能。但是，组织不会定期执行趋势分析和假设分析。假设分析是确定网络变化影响的过程。趋势分析是执行由网络容量和性能问题基线组成的过程，并查看网络趋势基线，以了解未来的升级需求。容量和性能管理还应包括异常管理（即用户呼入之前识别并解决问题）和QoS管理（即网络管理员规划、管理和识别各个服务性能问题）。下图说明了容量和性能管理流程。



容量和性能管理也有其局限性，通常与CPU和内存有关。以下是一些潜在的关注领域：

- CPU
- 底板或I/O
- 内存和缓冲区
- 接口和管道尺寸
- 队列、延迟和抖动
- 速度和距离
- 应用特征

一些对容量规划和性能管理的引用也提及了所谓的“数据平面”和“控制平面”。数据平面只是通过网络传输的数据涉及的容量和性能问题，而控制平面意味着维持数据平面正常运行所需的资源。控制平面功能包括服务开销，例如路由、生成树、接口保持活动以及设备的SNMP管理。这些控制平面要求使用CPU、内存、缓冲、队列和带宽，就像流经网络的流量一样。许多控制平面要求对于系统的整体功能也至关重要。如果他们没有所需的资源，网络就会发生故障。

## CPU

CPU通常由任何网络设备上的控制平面和数据平面使用。在容量和性能管理中，必须确保设备和网络具有足够的CPU来始终正常工作。CPU不足通常会造成网络崩溃，因为一台设备上的资源不足可能会影响整个网络。CPU不足也会增加延迟，因为在没有主CPU的硬件交换时，数据必须等待处理。

。

## 底板或I/O

背板或I/O是指设备可以处理的总流量，通常以总线大小或背板容量来描述。背板不足通常会导致数据包丢失，从而导致重新传输和额外的流量。

## 内存

内存是另一种具有数据平面和控制平面要求的资源。路由表、ARP表和其他数据结构等信息需要内存。当设备内存不足时，设备上的某些操作可能会失败。该操作可能会根据情况影响控制平面进程或数据平面进程。如果控制平面进程失败，整个网络可能会降级。例如，当路由收敛需要额外内存时，可能会出现这种情况。

## 接口和管道尺寸

接口和管道大小是指可在任意一个连接上同时发送的数据量。这通常被错误地称为连接速度，但数据实际上不会以不同的速度从一个设备传输到另一个设备。硅速度和硬件功能有助于根据介质确定可用带宽。此外，软件机制可以“限制”数据以符合服务的特定带宽分配。在帧中继或ATM的服务提供商网络中，您通常可以看到这一点，这些网络本身具有速度为1.54kpbs到155mb及更高的功能。当存在带宽限制时，数据会在传输队列中排队。发送队列可以具有不同的软件机制以优先化队列中的数据；但是，当队列中有数据时，它必须等待现有数据，然后才能将数据从接口转发出去。

## 队列、延迟和抖动

队列、延迟和抖动也会影响性能。可以调整传输队列以通过不同方式影响性能。例如，如果队列很大，则数据等待时间更长。当队列较小时，数据将被丢弃。这称为taildrop，对于TCP应用程序是可接受的，因为数据将被重新传输。但是，由于队列丢弃甚至队列延迟过长，语音和视频的表现不佳，需要特别注意带宽或管道大小。如果设备没有足够的资源来立即转发数据包，输入队列也可能发生队列延迟。这可能是由于CPU、内存或缓冲区。

延迟是指从收到数据包到转发数据包的正常处理时间。正常现代数据交换机和路由器在正常情况下具有极低的延迟(< 1ms)，不受资源限制。采用数字信号处理器转换和压缩模拟语音数据包的现代设备可能需要更长的时间，甚至高达20毫秒。

抖动描述了流应用（包括语音和视频）的数据包间间隙。如果数据包到达的时间不同，且数据包间间隔时间不同，则抖动很高，语音质量下降。抖动主要是排队延迟的因素。

## 速度和距离

速度和距离也是影响网络性能的一个因素。根据光速，数据网络具有一致的数据转发速度。这大约每毫秒100英里。如果组织在国际上运行客户端—服务器应用程序，则可能会出现相应的数据包转发延迟。当应用未针对网络性能进行优化时，速度和距离可能会对应用性能产生巨大影响。

## 应用特征

应用特征是影响容量和性能的最后—个区域。小窗口大小、应用程序keepalive以及通过网络发送的数据量与所需的数据量等问题可能会影响应用程序在许多环境（尤其是WAN）中的性能。

## 容量和性能管理最佳实践

本节详细讨论五个主要的容量和性能管理最佳实践：

- [服务级别管理](#)
- [网络和应用假设分析](#)
- [基线化和趋势化](#)
- [例外管理](#)
- [QoS管理](#)

## 服务级别管理

服务级别管理定义并管理其他所需的容量和性能管理流程。网络经理知道他们需要容量规划，但他们面临着预算和人员限制，这会妨碍完整的解决方案。服务级别管理是一种行之有效的方法，通过定义交付项并为与该交付项相关的服务创建双向责任来帮助解决资源问题。您可以通过两种方式完成此操作：

- 为包括容量和性能管理的服务创建用户和网络组织之间的服务级别协议。该服务将包括保持服务质量的报告和建议。但是，用户必须准备好为服务和任何所需的升级提供资金。
- 网络组织定义其容量和性能管理服务，然后逐一尝试为该服务和升级提供资金。

无论如何，网络组织都应当从定义容量规划和性能管理服务开始，该服务应当包括他们目前可以提供的服务的哪些方面以及未来的规划内容。完整的服务包括网络更改和应用更改的假设分析、已定义性能变量的基线化和趋势分析、已定义容量和性能变量的异常管理以及QoS管理。

## 网络和应用假设分析

执行网络和应用假设分析以确定计划更改的结果。如果没有假设分析，组织会冒极大的风险，来改变成功情况和整体网络可用性。在许多情况下，网络更改会导致拥塞崩溃，导致大量生产中断。此外，数量惊人的应用引入会失败，并会对其他用户和应用造成影响。这些故障在许多网络组织中都存在，但只需一些工具和一些额外的规划步骤，它们就能完全避免。

通常需要一些新的流程来执行质量假设分析。第一步是确定所有更改的风险级别，然后针对风险较高的更改进行更深入的假设分析。风险级别可以是所有提交变更的必填字段。风险等级越高，就越需要对变更进行明确的“假设分析”。网络假设分析确定网络更改对网络利用率和网络控制平面资源问题的影响。应用假设分析将确定项目应用成功、带宽要求以及任何网络资源问题。下表是风险级别分配和相应测试要求的示例：

| 风险级别 | 定义  | 变更规划建议  |
|------|---|---|
| 1    | <ul style="list-style-type: none"><li>• 由于新产品、软件、拓扑或功能的引入，对大量用户（500多名）或关键业务服务产生巨大潜在影响。</li><li>• 更改涉及预期的网络停机时间。</li></ul> | <ul style="list-style-type: none"><li>• 验证新解决方案的实验室。实验室验证包括已记录的解决方案测试和验证，以及显示对现有基础设施影响的假设分析。我们建议试用解决方案。新的解决方案需要完成运营支持文档。</li><li>• 执行思科NSA设计审核。</li><li>• 创建回退计划。</li><li>• 制定实施计划。</li></ul> |

|   |   |   |
|---|---|---|
|   |   | <ul style="list-style-type: none"> <li>• 创建变更进程。</li> </ul>   |
| 2 | <ul style="list-style-type: none"> <li>• 由于流量或用户的大幅增加、主干更改或路由更改，对大量用户（超过500名）或关键业务服务的潜在影响较大。</li> <li>• 更改可能需要一些停机时间。</li> </ul>  | <ul style="list-style-type: none"> <li>• 执行假设分析以确定对现有环境的影响（应在实验室环境中进行）。</li> <li>• 测试并检查路由更改的功能。</li> <li>• 创建回退计划。</li> <li>• 对主要路由或主干更改进行设计审核。</li> <li>• 制定实施计划。</li> <li>• 创建变更进程。</li> </ul> |
| 3 | <ul style="list-style-type: none"> <li>• 由于任何非标准更改，对较少的用户或业务服务产生的潜在影响不大。</li> <li>• 包括新产品、软件、拓扑、新增功能或新用户、增加的流量或非标准拓扑。</li> <li>• 更改可能需要一些停机时间。</li> </ul>   | <ul style="list-style-type: none"> <li>• 对新解决方案进行工程分析（可能需要实验室验证）。</li> <li>• 制定实施计划。</li> <li>• 创建变更进程。</li> </ul>  |
| 4 | <ul style="list-style-type: none"> <li>• 降低潜在的服务或用户影响。</li> <li>• 包括添加新的标准模板网络模块，例如在路由器上构建或服务器交换机/集线器。</li> <li>• 包括启用新的WAN站点或额外的经验证的接入服务。</li> <li>• 所有风险级别3的变化都已在生产环境中经过技术验证。</li> <li>• 更改可能需要一些停机时间。</li> </ul> | <ul style="list-style-type: none"> <li>• 制定实施计划。</li> <li>• 创建变更进程。</li> </ul>  |
| 5 | <ul style="list-style-type: none"> <li>• 不影响用户或服务。</li> <li>• 包括将个人用户添加到网络和标准配置更改，例如密码、横幅、</li> </ul>   | <ul style="list-style-type: none"> <li>• 更改进程是可选的。</li> </ul>   |

|                             |  |
|-----------------------------|--|
| SNMP或其他标准配置参数。<br>• 没有停机时间。 |  |
|-----------------------------|--|

一旦定义了需要假设分析的位置，就可以定义服务。

您可以使用建模工具或通过模拟生产环境的实验室执行网络假设分析。建模工具受应用程序对设备资源问题的了解程度限制，而且由于大多数网络更改都是新设备，应用程序可能无法理解更改的影响。最佳方法是在实验室中构建一些生产网络的表示形式，并使用流量发生器测试负载下所需的软件、功能、硬件或配置。将生产网络的路由（或其他控制信息）泄漏到实验室中也会增强实验环境。测试不同流量类型（包括SNMP、广播、组播、加密或压缩流量）的其他资源要求。使用所有这些不同的方法，分析潜在压力情况（例如路由收敛、链路抖动和设备重新启动）下的设备资源需求。资源利用率问题包括CPU、内存、背板利用率、缓冲区和队列等正常容量资源区域。

新应用还应执行假设分析，以确定应用成功和带宽要求。通常要在实验室环境中使用协议分析器和WAN延迟模拟器执行此分析，以了解距离的影响。您只需要将PC、集线器、WAN延迟设备和实验路由器连接到生产网络。通过在测试路由器上使用通用流量整形或速率限制来限制流量，可以在实验室中模拟带宽。网络管理员可以与应用组合作，了解在LAN和WAN环境中应用的带宽要求、窗口问题和潜在性能问题。

在部署任何业务应用之前执行应用假设分析。如果不这样做，应用组会将性能不佳归咎于网络。如果您能够通过变更管理流程要求对新部署进行应用假设分析，则有助于防止部署失败，并更好地了解客户端 — 服务器和批处理要求的带宽消耗的突然增加。

## 基线化和趋势

通过基线化和趋势化，网络管理员可以规划和完成网络升级，避免出现容量问题导致网络中断或性能问题。比较连续时段内的资源利用率或提取数据库中一段时间内递减的信息，使规划者能够查看上一小时、天、周、月和年的资源利用率参数。无论是哪种情况，都必须有人每周、每两周或每月复查信息。基线化和趋势分析的问题在于，在大型网络中，需要查看大量的信息。

可以通过多种方法解决此问题：

- 构建足够的容量并切换到LAN环境，因此容量不是问题。
- 将趋势信息划分为不同的组，并集中关注网络的高可用性或关键区域，如关键WAN站点或数据中心LAN。
- 报告机制可以突出超过特定门槛的领域以引起特别关注。如果您首先实施关键可用性领域，则可以显着减少审核所需的信息量。

使用上述所有方法，您仍需要定期查看信息。基线化和趋势分析是一项主动工作，如果组织仅拥有被动支持资源，个人不会阅读报告。

许多网络管理解决方案提供有关容量资源变量的信息和图形。不幸的是，大多数人只使用这些工具被动支持现有问题；这会破坏基线化和趋势化的目的。两种有效提供思科网络容量趋势信息的工具是Concord Network Health产品和INS EnterprisePRO产品。在许多情况下，网络组织运行简单的脚本语言来收集容量信息。以下是一些通过脚本收集的有关链路利用率、CPU利用率和ping性能的示例报告。其他重要的资源变量可能包括内存、队列深度、广播量、缓冲区、帧中继拥塞通知和背板利用率。有关链路利用率和CPU利用率的信息，请参阅下表：

### 链路利用率

| 资源 | 地址 | 网段 | 平均利用率 | 峰值利用率 |
|----|----|----|-------|-------|
|----|----|----|-------|-------|

|            |          |          | (%)  | (%)   |
|------------|----------|----------|------|-------|
| JTKR01S2   | 10.2.6.1 | 128 Kbps | 66.3 | 97.6  |
| JYKR01S0   | 10.2.6.2 | 128 Kbps | 66.3 | 97.8  |
| FMCR18S4/4 | 10.2.5.1 | 384 Kbps | 51.3 | 109.7 |
| PACR01S3/1 | 10.2.5.2 | 384 Kbps | 51.1 | 98.4  |

### CPU 利用率

| 资源     | 轮询地址        | 平均利用率(%) | 峰值利用率(%) |
|--------|-------------|----------|----------|
| FSTR01 | 10.28.142.1 | 60.4     | 80       |
| NERT06 | 10.170.2.1  | 47       | 86       |
| NORR01 | 10.73.200.1 | 47       | 99       |
| RTCR01 | 10.49.136.1 | 42       | 98       |

### 链路利用率

| 资源        | 地址            | AvResT(mS)09-09-98 | AvResT(mS)09-09-98 | AvResT(mS)09-09-98 | AvResT(mS)10-01-98 |
|-----------|---------------|--------------------|--------------------|--------------------|--------------------|
| AA DR01   | 10.19.0.56.1  | 469.1              | 852.4              | 461.1              | 873.2              |
| AB NR01   | 10.19.0.52.1  | 486.1              | 869.2              | 489.5              | 880.2              |
| AP RR01   | 10.19.0.54.1  | 490.7              | 883.4              | 485.2              | 892.5              |
| ASAR01    | 10.19.6.170.1 | 619.6              | 912.3              | 613.5              | 902.2              |
| ASR01     | 10.19.6.178.1 | 667.7              | 976.4              | 655.5              | 948.6              |
| ASYR01S   |               |                    |                    |                    | 503.4              |
| AZ WR T01 | 10.17.7.32.1  | 460.1              |                    | 444.7              |                    |

|            |                     |        |        |      |        |
|------------|---------------------|--------|--------|------|--------|
| BEJ<br>R01 | 10.19<br>5.18.<br>1 | 1023.7 | 1064.6 | 1184 | 1021.9 |
|------------|---------------------|--------|--------|------|--------|

## 例外管理

例外管理是识别和解决容量和性能问题的重要方法。其理念是接收容量和性能阈值违规通知，以便立即调查和解决问题。例如，网络管理员可能收到路由器上CPU使用率较高的警报。网络管理员可以登录路由器来确定CPU使用率这么高的原因。然后，她可以执行一些补救配置，以减少CPU或创建访问列表，防止导致问题的流量，尤其是在流量似乎不是业务关键型流量时。

只需在路由器上使用RMON配置命令，或者使用更高级的工具（例如Netsys服务级别管理器与SNMP、RMON或Netflow数据），即可为更重要的问题配置例外管理。大多数网络管理工具能够设置违规的阈值和警报。例外管理流程的重要方面是提供近实时问题通知。否则，问题会在有人注意到通知收到之前消失。如果组织具有一致的监控功能，则可在NOC中完成此操作。否则，我们建议使用寻呼机通知。

以下配置示例为路由器CPU提供针对日志文件的上升和下降阈值通知，该日志文件可以持续查看。您可以为严重链路利用率阈值违规或其他SNMP阈值设置类似的RMON命令。

```
rmon event 1 trap CPUtrap description
"CPU Util >75%"rmon event 2 trap CPUtrap description
"CPU Util <75%"rmon event 3 trap CPUtrap description
"CPU Util >90%"rmon event 4 trap CPUtrap description
"CPU Util <90%"rmon alarm 75 lsystem.56.0 10 absolute rising-threshold
75 1 falling-threshold 75 2rmon alarm 90 lsystem.56.0 10 absolute rising-threshold
90 3 falling-threshold 90 4
```

## QoS管理

服务质量管理包括创建和监控网络内的特定流量类。流量为特定应用组（在流量类中定义）提供更一致的性能。流量整形参数为特定流量类的优先级划分和流量整形提供了极大的灵活性。这些功能包括承诺接入速率(CAR)、加权随机早期检测(WRED)和基于类的公平加权队列等功能。流量类别通常基于性能SLA创建，用于更多关键业务应用和特定应用要求，例如语音。对非关键或非业务流量的控制方式也不会影响优先级较高的应用和服务。

创建流量类需要基本了解网络利用率、特定应用要求以及业务应用优先级。应用要求包括数据包大小、超时问题、抖动要求、突发要求、批处理要求以及整体性能问题的知识。有了这些知识，网络管理员可以创建流量整形计划和配置，从而在各种LAN/WAN拓扑中提供更一致的应用性能。

例如，一个组织在两个主要站点之间有一个10兆位ATM连接。链路有时因大型文件传输而堵塞，从而导致在线事务处理的性能下降，语音质量较差或不可用。

组织设置了四种不同的流量类别。语音被赋予最高优先级，并且即使突发超过估计流量速率，也允许保持该优先级。关键应用类具有下一个最高优先级，但不允许在总链路大小减去估计的语音带宽要求后进行突发。当它发生突发时，将被丢弃。文件传输流量被赋予较低的优先级，而所有其他流量则位于中间的某个位置。

现在，组织必须对此链接执行QoS管理，以确定每个类所接收的流量大小，并衡量每个类内的性能。如果组织未能做到这一点，则某些类可能会发生耗竭，或者某个特定类可能无法满足性能SLA。

由于缺乏工具，管理QOS配置仍是一项艰巨的任务。一种方法是使用思科的Internet性能管理器(IPM)通过属于每个流量类的链路发送不同的流量。然后，您可以监控每个类的性能，IPM提供趋势

分析、实时分析和逐跳分析以查明问题区域。其他方法可能仍然依赖比较手动的方法，例如根据接口统计信息检查每个流量类中的排队和丢弃的数据包。在某些组织中，可以通过SNMP收集此数据，或者将数据解析到数据库中以获取基线和趋势。市场上也存在一些工具，它们通过网络发送特定流量类型，以确定特定服务或应用的性能。

## 收集和报告容量信息

收集和报告能力信息应与建议的三个能力管理领域联系起来：

- 假设分析：围绕网络变化以及变化对环境有何影响
- 基线化和趋势化
- 例外管理

在上述每个领域内，制定信息收集计划。在网络或应用假设分析中，您需要一些工具来模拟网络环境，并了解相对于设备控制平面或数据平面内潜在资源问题的变化所造成的影响。在基线化和趋势分析的情况下，您需要显示当前资源利用率的设备和链路的快照。然后查看随时间变化的数据，了解潜在的升级要求。这样，网络管理员可以在出现容量或性能问题之前正确规划升级。出现问题时，您需要进行异常管理来提醒网络管理员，以便他们调整网络或解决问题。

此过程可分为以下步骤：

1. 确定您的需求。
2. 定义流程。
3. 定义容量区域。
4. 定义能力变量。
5. 解释数据。

### 确定您的需求

制定容量和绩效管理计划需要了解您需要的信息以及信息的用途。将计划分为三个必需区域：分别用于假设分析、基线建立/趋势分析和例外管理。在每个区域内，了解可用的资源和工具以及需要什么。许多组织在使用工具部署时都会失败，因为他们考虑工具的技术和功能，但不考虑管理工具所需的人员和专业技能。在计划中纳入所需人员和专业技能，并改进流程。这些人可能包括系统管理网络管理站的系统管理员、帮助进行数据库管理的数据库管理员、使用和监控工具的训练有素的管理员，以及确定策略、阈值和信息收集要求的高级别网络管理员。

### 定义流程

您还需要一个流程来确保成功一致地使用工具。可能需要改进流程，以定义发生阈值违规时网络管理员应该执行什么操作，或者执行什么流程进行网络基线、趋势分析和升级。确定成功容量规划的需求和资源后，可以考虑方法。许多组织选择将此类功能外包给网络服务组织（如INS），或构建内部专业知识，因为他们认为服务是一项核心能力。

### 定义容量区域

能力规划计划还应包括能力领域的定义。以下是可以共享通用容量规划策略的网络区域：例如，企业LAN、WAN现场办公室、关键WAN站点和拨号接入。定义不同区域有帮助，原因如下：

- 不同区域可能有不同的阈值。例如，LAN带宽比WAN带宽便宜得多，因此利用率阈值应更低。
- 不同区域可能需要监控不同的MIB变量。例如，帧中继中的FECN和BECN计数器对了解帧中继

容量问题至关重要。

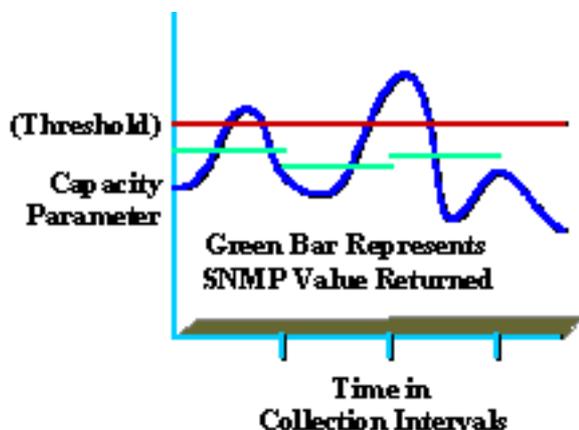
- 升级某些网络区域可能更困难或更耗时。例如，国际电路可以具有更长的提前期，并且需要相应的更高水平的规划。

## 定义容量变量

下一个重要领域是定义要监控的变量和需要操作的阈值。容量变量的定义在很大程度上取决于网络中使用的设备和介质。通常，CPU、内存和链路利用率等参数很有价值。但是，对于特定技术或要求，其他领域可能很重要。这些可能包括队列深度、性能、帧中继拥塞通知、底板利用率、缓冲区利用率、netflow统计信息、广播量和RMON数据。请牢记您的长期计划，但首先从几个关键方面帮助确保成功。

## 解释数据

了解收集的数据也是提供高质量服务的关键。例如，许多组织并不完全了解峰值和平均利用率级别。下图显示基于5分钟SNMP收集间隔的容量参数峰值（显示为绿色）。



即使报告的值小于阈值（显示为红色），峰值仍然可以在高于阈值（显示为蓝色）的收集间隔内出现。这一点非常重要，因为在收集间隔期间，组织可能遇到了影响网络性能或容量的峰值。请注意选择有意义的收集间隔，该间隔既有用，又不会产生过多开销。

另一个示例是平均利用率。如果员工在办公室的时间从8到5不等，但平均利用率是7X24，则信息可能会产生误导。

## 相关信息

- [技术支持 - 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）。