

Applikationen zum Laufen bringen

Autor(en): **Pupers, Arnold**

Objektyp: **Article**

Zeitschrift: **Comtec : Informations- und Telekommunikationstechnologie = information and telecommunication technology**

Band (Jahr): **80 (2002)**

Heft 5

PDF erstellt am: **28.06.2024**

Persistenter Link: <https://doi.org/10.5169/seals-877200>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Performance

Applikationen zum Laufen bringen

Wenn ein zunehmender Anteil des Unternehmensumsatzes von der Online-Performance abhängt, wenn die internetfähigen Applikationen von heute mehr als je zuvor kosten und wenn die Produktivität der Mitarbeiter von den Response-Zeiten abhängt, dann muss man sich fragen: Warum plagen Sie sich mit einer schlechten Performance? Dieser Artikel beschreibt effektive Lösungen zur Applikations-Performance und entwickelt Konzepte für unterschiedliche Szenarios.

Im Kampf um die Bandbreite überlasteter Access-Verbindungen können mächtige Applikationen wie Napster oder grosse Attachments an E-Mails die Kapazität so in Anspruch nehmen, dass wichtiger Datenverkehr vollständig un-terlaufen wird.

ARNOLD PIJPERS

Die Performance von webbasierenden Applikationen und E-Commerce muss mit weiteren Störfaktoren, wie unterschiedliche Access-Geschwindigkeiten von Wählverbindungs- und Breitbandteilnehmern, oder dem steilen Leistungsabfall mit weniger Verbindungen pro Sekunde fertig werden, sobald die Online-Transaktionen um Security-Massnahmen erweitert werden.

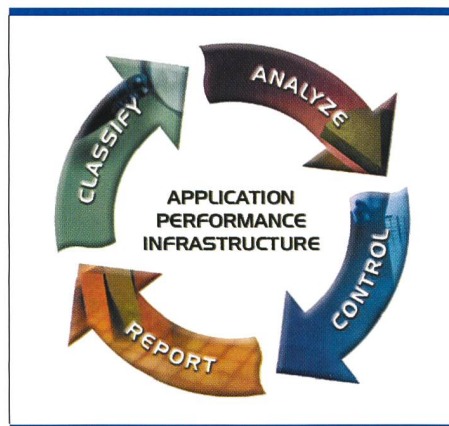


Bild 1. Analyse-Controlling-Reporting-Klassifizierung.

Prioritätsebene	Applikationen
7	Client/Server, SAP, PeopleSoft, Oracle, TB3270, Web-Zugang der Anwender
5	MS Exchange, Applikationen zur Gruppenarbeit
4	Intranet, Netzwerk-Management
3	Kalender, E-Mail, Druckaufträge
1	Gelegentliches Surfen in Web
0	Napster, Gnutella, Internet-Radio

Tabelle 1

Bei der Performance können sich folgende Fälle ereignen:

- Eine Universität findet heraus, dass ihr Netzwerk durch ständige Flows von Musik-Downloads der Studenten im Wohnheim missbraucht wird.
- Das Team eines Service-Providers vergeudet mehrere Diagnosestunden mit Performance-Problemen, die aus dem Netz eines Teilnehmers statt aus dem eigenen stammen.
- Eine Versicherung stellt ihr Schadensfallsystem vom Mainframe auf das Internet um. Bei der Freigabe sinkt die Produktivität der mit niedrigen Geschwindigkeiten angeschlossenen Ausstellen auf ein Minimum.
- Ein grosses Produktionsunternehmen führt eine wichtige ERP-Applikation ein. Die Planer wollen sie vor unwichtigem Webtraffic schützen und die Response-Zeiten als Erfolgsanzeige messen.
- Nachdem ein Service Provider die Zusammenarbeit mit einem Teilnehmer um SSL-basierende Sicherheitsmassnahmen erweitert hat, sinkt die Performance der Website. SSL ist so prozessorintensiv, dass die Zahl der Ver-

bindungen pro Sekunde am Webserver von 300 auf vier sinkt.

- Ein Lebensmittel- und Getränkegrosshändler vertreibt seine Produkte an grosse Lebensmittelketten und an kleine Restaurants. Die Frustration des Grosshändlers steigt mit den Kosten

für die Entwicklung und Wartung einer doppelten Auftragsverarbeitungs-Website für Abnehmer mit niedriger und hoher Geschwindigkeit.

Diese Szenarios erfordern unterschiedliche Lösungen. Dennoch sind alle auf eine schnelle, gleichmässige und vorher-sagbare Performance der Applikationen angewiesen.

Die Lösung: ein Vorgang in vier Schritten

Die herkömmliche Priorisierung von Datenpaketen ist keine effektive Lösung zur Applikations-Performance mehr. Die Antworten von heute sollten Aspekte des Traffic Management, Beschleunigung, Service Level Management und Policy Management kombinieren. Eine umfassende Lösung muss die Schwierigkeiten bei der Content-Übertragung auf der letzten Meile überwinden, Hilfen bei einem Leistungsabfall aus Sicherheitsgründen bieten, Strategien zum Schutz umsetzen oder die Nutzung von Applikationen unterwegs ermöglichen und die Grenzen setzen, sobald Applikationen als Services übertragen werden.

Der folgende Management-Zyklus aus vier Stufen verhilft Applikations-Managern zu einer performanten Übertragung und unterstützt sie bei der Auswahl von Produkten zur Applikations-Performance (Bild 1):

- Identifizieren Sie den Netzwerk-Traffic und klassifizieren Sie ihn nach den Applikationen, die über das Netzwerk übertragen werden.
- Analysieren Sie das Netzwerk und das Verhalten der Applikationen und teilen Sie die Performance in entsprechende Anteile auf: Netzwerk gegen Server, Provider gegen Teilnehmer usw.
- Steuern Sie die Verteilung der Ressourcen und beschleunigen Sie die Übertragung der Applikation.
- Bewerten Sie das Geschehen im Netz durch Reports über die Qualität der Ergebnisse.

Schritt 1:

Der Datenverkehr wird klassifiziert

Das Aufspüren und Identifizieren von Typen des Netzwerkdatenverkehrs, die um begrenzte Ressourcen konkurrieren, ist der erste Schritt bei der Beseitigung von Überlastungen. Die umfassende Klassifizierung des Datenverkehrs ist entscheidend: Die Steuerung der Ressourcen nutzt nur dann, wenn sie sich genau den Vorstellungen über den Datenverkehr anpassen lässt. Normalerweise sind Administratoren überrascht, wenn sie nach einer automatischen Klassifizierung die Verschiedenheit des Netzwerkverkehrs erkennen.

Eine spezifische Klassifizierung des Datenverkehrs bringt bessere Ergebnisse. Isolierte Traffic-Kategorien lassen sich durch Policies unterschiedlich behandeln, welche die Verteilung der Ressourcen und dadurch die Performance regeln. Sich auf Protokolle niedriger Schichten, IP-Adressen oder statische Port-Nummern zu verlassen, um den Datenverkehr zu klassifizieren, schliesst die Entdeckung echter Traffic-Trends aus und unterläuft die Policies.

Die robuste Klassifizierung setzt die Möglichkeit voraus, den Traffic nach einer Fülle von Kriterien zu differenzieren, darunter die Applikation, Protokolle höherer Schichten, Subnet, Website, Browser, MIME-Art, Anwender, Richtung des Datenverkehrs, LDAP Host List sowie Ursprung und Ziel des Datenverkehrs. Optimale Lösungen sollten den Datenverkehr nach dynamischen und migrierenden Port-Nummern aufzeichnen und

zwischen Applikationen unterscheiden, die mit dem unter TCP so gängigen Port 80 arbeiten. Die Lösungen müssen Peer-to-Peer-Applikationen (z. B. Napster) unterscheiden, um sie zu kontrollieren. Auch sollte passives FTP verfolgt werden, sogar wenn es dynamisch die Ports wechselt. TN3270 und damit verbundener Druck-Datenverkehr dürfen sich unter Telnet nicht verbergen können. Auch sollte SAP-Datenverkehr trotz der unvorhersagbaren Ports identifiziert werden. erinnert man sich an die oben beschriebenen Szenarios, benötigt die Universität, die unter den Musik-Downloads leidet, eine Lösung zum Erkennen von Napster-Datenverkehr. Das Produktionsunternehmen, das Bedenken über die Auswirkungen des Webbrowsing auf die

Nützliche Analyse-Tools, die die Effizienz des Netzwerkes und andere Informationen darstellen, liefern einen schnellen und einfachen Zugriff auf Grafiken und Tabellen mit den durchschnittlichen und Spitzenlasten, Response-Zeiten, Top-Anwendern und Applikationen. Oft bringen diese Grafiken aber nicht alles ans Tageslicht. So tarnen sich Trends, bis eine Grafik das Augenmerk auf besondere Applikationen oder Server lenkt. Ein Beispiel: Nehmen unnötige Rückübertragungen nur 5% der Kapazität einer WAN-Verbindung in Anspruch, fällt das nicht weiter auf. Steigt dieser Wert jedoch auf 45% des Datenverkehrs einer bestimmten Applikation oder eines Servers, ist das Problem offensichtlich. Zeitgemässe Analyse-Tools müssen daher einen weiten

Applikation	Vorgabe
SAP	40 % der Bandbreite sind für SAP zu reservieren, das auch beim Zugriff auf zusätzliche Bandbreite eine hohe Priorität erhält. Ist SAP nicht aktiv und benötigt weniger als 40% der Bandbreite, können andere Applikationen darauf zugreifen.
Napster	Jeder Napster-Anwender erhält eine verringerte Bandbreite mit dem Limit von 1 k/bit. Sollte der Napster-Datenverkehr mehr als 2 % der Kapazität beanspruchen, ist er dort zu knapp.
Web	Das Web sollte eine niedrige bis mittlere Priorität mit bestmöglicher Performance erhalten. Mit TCP Rate Control können die vier Applikationen hohen Spitzen-Traffic auffangen.
E-Mail	TCP Rate Control verhindert, dass E-Mails mit sehr grossen Attachements von jemand anderem Bandbreite abziehen. Da es sich bei der E-Mail um eine Business-Funktion handelt, sollte sie eine mittlere bis hohe Priorität erhalten.

Tabelle 2

Applikation	Vorgabe
Geschäftskritisches WEB	Erhält eine hohe Priorität und TCP Rate Control. Bei Bedarf schützt ein reservierter Bandbreitenbereich diese Website, was wahrscheinlich jedoch überflüssig ist.
Gelegentliche Web-Zugriffe	Die ursprüngliche Vorgabe soll weiteren gelegentlichen Web-Datenverkehr zulassen.

Tabelle 3

ERP-Performance hegt, benötigt eine Lösung zur Unterscheidung dieser beiden Applikationen. Zur Erinnerung: Nur wenn sich der Datenverkehr differenzieren lässt, ist es möglich, eine Applikation zu schützen.

**Schritt 2:
Das Verhalten analysieren**

Während die Erforschung des Datenverkehrs Aufschluss über das «Was» im Netz gibt, zeigt die Analyse, «wie» es sich verhält (Bild 2).

oder engen Blickwinkel aufweisen. Die Analyse der Response-Zeit ist ein wichtiger Schritt bei der Verwaltung der Applikations-Performance. Mit den Response-Zeiten lassen sich die Ausgangswerte der ursprünglichen Performance festlegen und Performance-Probleme erkennen, bevor sie sich auf das Business auswirken. Auch kann man die Ergebnisse von Veränderungen bewerten und Service Level Agreements überwachen. Die Lösungen sollten die End-to-End-Latenz messen und zwischen einem

langsamen Netzwerk und langsamen Servern unterscheiden, die langsamsten Clients oder Server identifizieren und die Applikationen auf dem Performance-Standard halten.

Unterschiedliche Zielgruppen benötigen verschiedene Analysen, denen die Performance-Lösungen untergeordnet sein sollten. So sind Endanwender an der gesamten End-to-End-Response-Zeit interessiert, IT-Mitarbeiter hingegen wünschen eine Response-Gesamtzeit, die nach Netzwerk- und Server-Komponenten aufgeteilt ist. Die Anbieter von applikationsbasierenden Services wiederum müssen die Response-Zeit in den Anteil unter Verantwortung des Providers und der des Teilnehmers aufteilen.

Die Analyselösungen sollten keine Veränderungen auf Desktops, Applikationen, Server und Router bewirken und niemals das Netz mit noch mehr Datenverkehr belasten. Die Lösung darf nicht Teil des Problems sein.

Die Szenarios am Beginn des Beitrags erfordern beispielsweise folgende Analysen:

- Die Universität möchte sehen, welche Applikationen den Löwenanteil der Bandbreite in Anspruch nehmen. Welches sind die zehn wichtigsten Applikationen? Welchen Prozentsatz nimmt Napster für sich in Anspruch? Wer sind die 20 Top-Anwender mit Napster?
- Der Service Provider muss in der Lage sein, Beschwerden über die langsame Performance bestimmter Applikationen überprüfen. Existiert wirklich eine langsame Schwachstelle? Liegt sie im eigenen Netzwerk oder beim Teilnehmer?
- Die Versicherung muss die Response-Zeiten ihrer webbasierenden Schadensfall-Applikation analysieren. Wie lang sind die durchschnittlichen Response-Zeiten für Mitarbeiter in der Zentrale? Und wie in den Aussenstellen?
- Das Produktionsunternehmen benötigt eine Ausgangsbasis für die ERP-Performance vor dem allgemeinen Rollout. Ist sie langsamer oder schneller als die Performance-Vorgaben? Welchen Anteil hat sie an der Gesamtzeit? Hängt sie von der Tageszeit oder dem Standort ab?

**Schritt 3:
Die Performance steuern**

Die Steuerung ist der Mittelpunkt der Applikations-Performance. Während sich die anderen Schritte dem Problem wid-

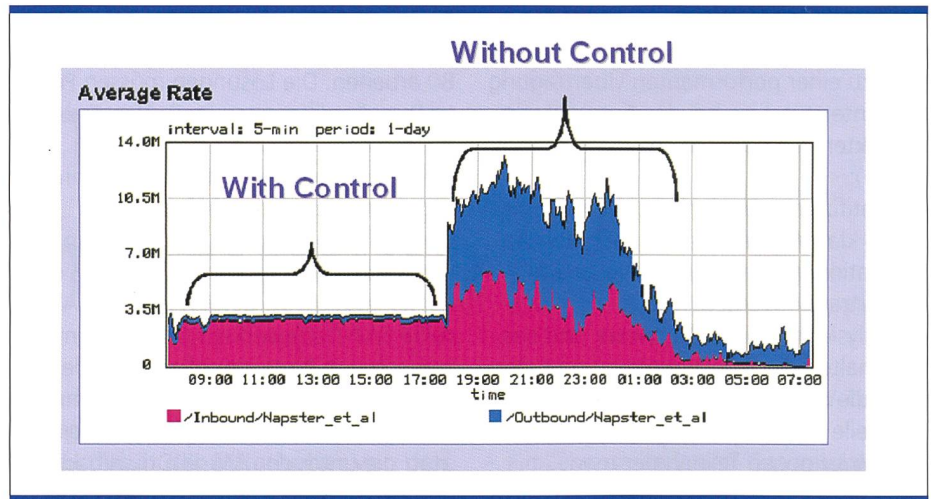


Bild 2. Analyse.

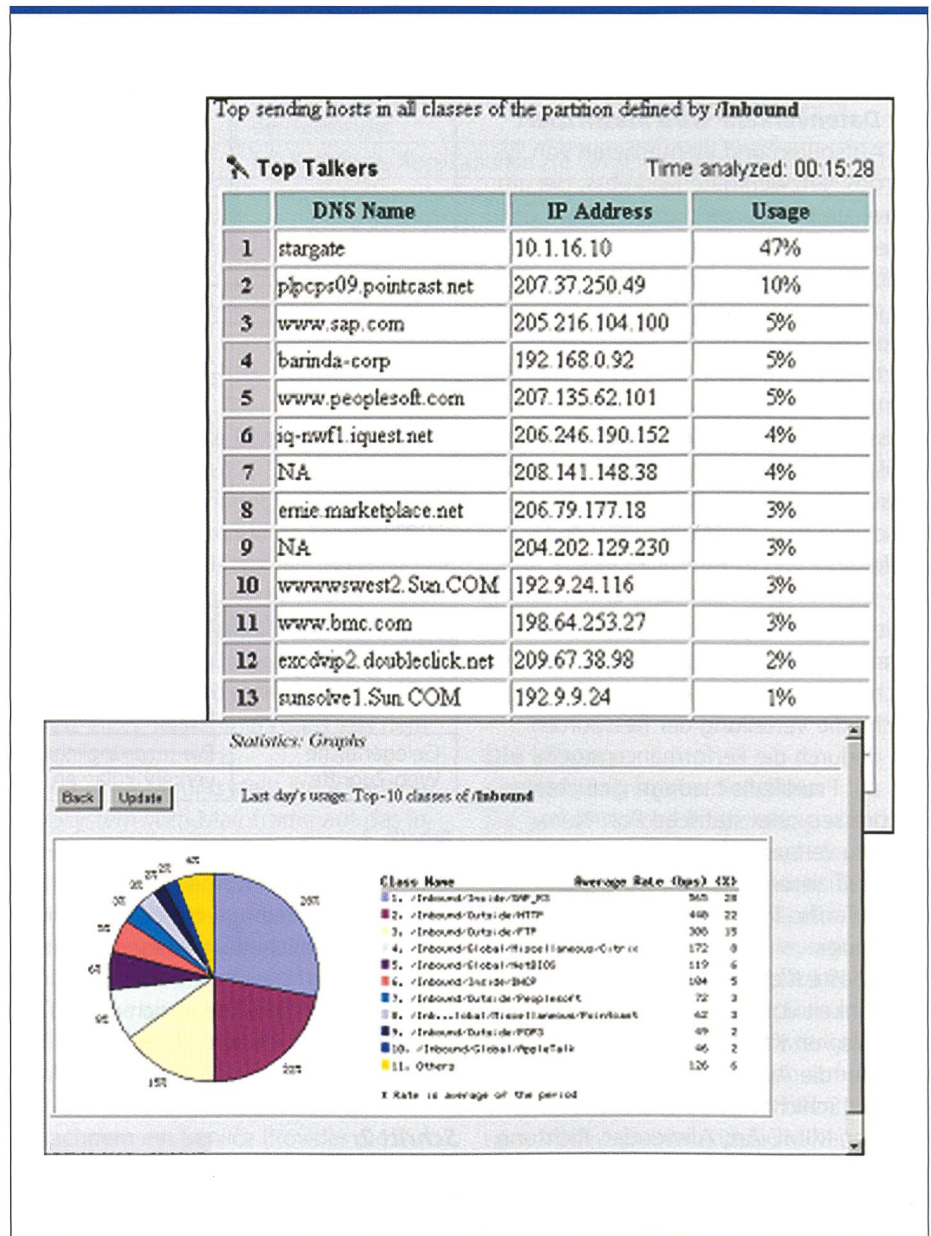


Bild 3. Ein besonderes Bits-pro-Sekunde-Bandbreiten-Minimum pro Applikation.

men, bietet dieser Schritt die Möglichkeit, die Probleme zu beheben. Zu einer leistungsfähigen Lösung gehören daher der Schutz kritischer Applikationen, der Erhalt nichtkritischer Applikationen, das Tempo des gesamten Traffic für optimale Effizienz und die schnelle Übertragung. Die folgenden Listen beschreiben Features, die diesen Zielen entsprechen, und wie sie in den oben beschriebenen Szenarios und in anderen Situationen greifen (Bild 3).

Schutz

- Das Produktionsunternehmen sollte einen Teil der Bandbreitenkapazität nur für seine ERP-Applikation frei halten, ebenso die Versicherung für die Schadensfall-Applikation in den Aussenstellen. Flexible Lösungen vergeuden keine Bandbreite, falls die wichtigen Applikationen sie nicht benötigen - sie stellen sie zeitweise anderen zur Verfügung.
- Ein besonderes Bits-pro-Sekunde-Bandbreiten-Minimum pro Session: Die Verteilung eines Minimums pro Session für schützenswerte Audio- und Video Streams (z. B. VoIP) zur Vermeidung von Stottern und Hängern. Mit Zugangskontroll-Features lässt sich vermeiden, dass eine Flut von spät ankommenden Datenpaketen die Performance verzögert.
- Prioritäten zur Einstufung des Zugriffs bestimmter Applikationen auf der verfügbare Bandbreite: Diese Prioritäten bestimmen, wie die vorher festgelegte Bandbreitenverteilung skaliert wird. Tabelle 1 beschreibt eine mögliche Prioritätszuweisung mit Prioritäten von 0 bis 7. Einfache Prioritäten sind ungenügend, aber sie ergänzen andere, leistungsfähigere Features.
- Das Offloading hoch ausgelasteter Server und Load Balancing: Der Service

Provider, der unter einem Leistungsabfall von 400 auf drei Transaktionen pro Sekunden litt, könnte ein spezielles Gerät zur Verschlüsselung und Entschlüsselung der Zusatzlast verwenden, die durch sichere Transaktionen entsteht. So sollten die Server-Lasten automatisch verteilt werden, vor allem für Server mit kritischen Applikationen. Sollte eine prozessorintensive Aufgabe die Effizienz des Servers verringern, empfehlen sich dedizierte, effiziente Lösungen.

Begrenzung

- Ein besonderes Bits-pro-Sekunde-Bandbreiten-Maximum pro Applikation: Eine gängige Begrenzungsstrategie, die auch für das Produktionsunternehmen in Frage käme, ist das Kappen des Zugriffs auf die Bandbreite für Dateiübertragungen und gelegentliches Webbrowsing. So könnte die Universität den Napster-Datenverkehr auf 5% der Gesamtnetzkapazität oder weniger begrenzen. Ein anderes Verfahren ist die Kombination von Minimum und Maximum, um wichtige Applikationen zu balancieren, die zum Anschwellen tendieren. So lassen sich beispielsweise 20% der Kapazität für Microsoft Exchange reservieren, aber bei 60% beschneiden.
- Ein besonderes Bits-pro-Sekunde-Maximum pro Session: Eine gängige Begrenzungsstrategie pro Session verhindert, dass ein FTP-Anwender die gesamte Bandbreite mit seinem FTP-Traffic belegt. Ein ähnliches System könnte beim gelegentlichem Webbrowsing sogar vermeiden, dass ein Napster-Download andere Musikfreunde blockiert.
- Blockaden: Um einen reibungslosen Datenverkehr und saubere Applikatio-

nen zu erhalten, sollte Traffic, der nicht klassifiziert werden kann, nicht übertragen werden. Eine entsprechende Meldung kann dem so blockierten Anwender gesendet werden oder auch nicht. So könnte man beispielsweise alle, die eine beliebige Website mit Streaming-Audio besuchen möchten, auf eine andere Website umleiten, die erläutert, dass Streaming-Audio nur zwischen 22.00 Uhr und 06.00 Uhr erlaubt ist.

Pacing

- TCP Rate Control: Die Engpässe von TCP lassen sich überwinden, indem man Überlastungen durch ankommende und ausgehende Applikations-Flows proaktiv vermeidet. Werden die Endgeräte angewiesen, langsamer zu reagieren, können die Pakete nicht schneller übertragen werden, als sie beim Empfänger angenommen werden. Statt die Datenpakete in einer überlasteten Warteschlange fallen zu lassen, legt TCP Rate Control die Geschwindigkeit der Pakete fest und verhindert Überlastungen. So wird vermieden, dass ein Anwender mit hoher Kapazität Bandbreite von allen anderen abzieht, und man erhält eine sanfte und gleichmässige Flow-Rate zum maximalen Durchsatz. TCP Rate Control misst die Netzwerk-Latenz, sieht die Zeit zwischen der Ankunft von Datenpaketen voraus, stellt die Fenstergrösse gemäss der Latenz und dem Forecast ein und misst den so definierten Stream, um die Just-in-time-Übertragung sicherzustellen.

Beschleunigung

- Kompression der Inhalte: Viele Websites lassen sich um 80% komprimieren, was kürzere Wartezeiten und

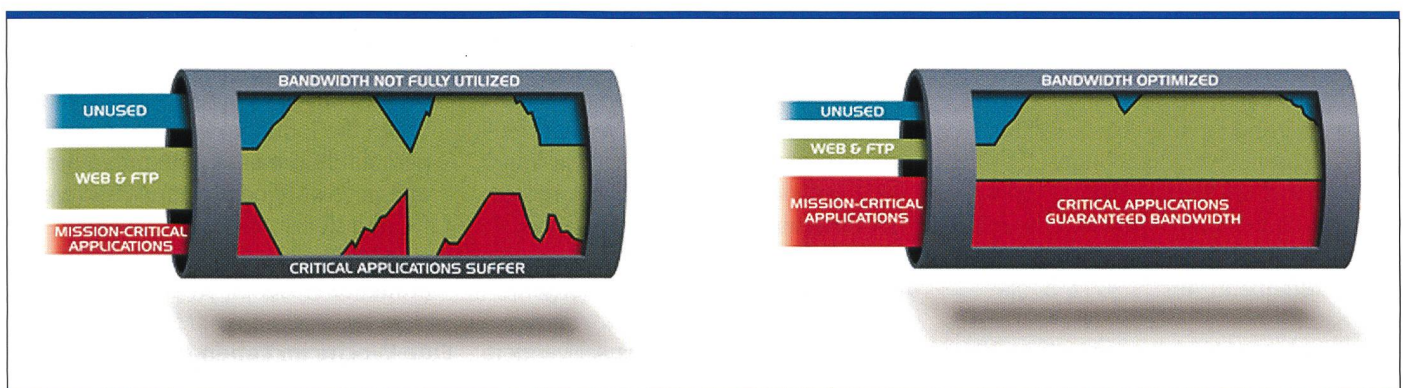


Bild 4. Controlling.

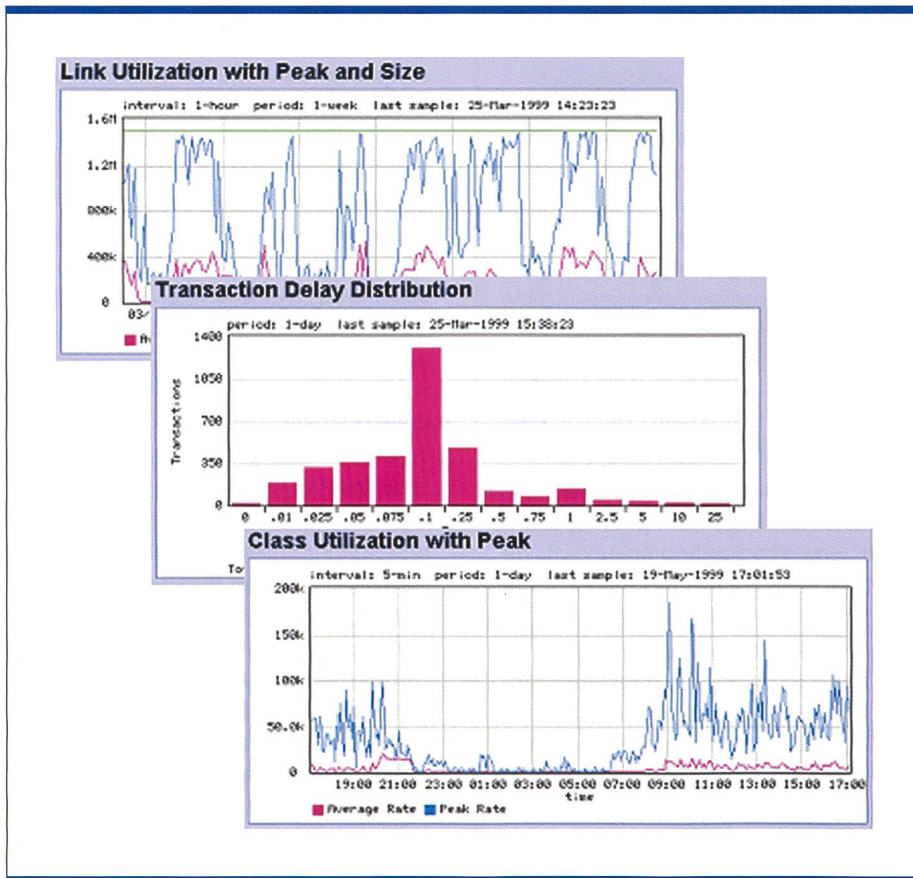


Bild 5. Reporting.

niedrigere Bandbreiten zur Übertragung zur Folge hat. Falls die Kompressionstechniken bereits im Browser enthalten sind, wird beim Client keine Software zur Dekomprimierung benötigt. In diesem Fall könnte die oben genannte Versicherung den Webcontent für ihre Mitarbeiter in den Aussenstellen komprimieren, um bei der Umstellung auf eine webfähige Lösung einen Geschwindigkeitsverlust zu vermeiden.

- Dynamische Erkennung von Geschwindigkeit und Browser: Das Erkennen der Charakteristiken eines Anwenders in Echtzeit, darunter Verbindungsgeschwindigkeit und Art des Browsers, bildet eine ausgezeichnete Basis für die Entscheidungen, in welchem Fall Beschleunigungsverfahren eingesetzt werden sollen. So könnte der Lebensmittel- und Getränkegrosshändler erkennen, ob ein aktiver Abnehmer über einen Breitband-Anschluss (z. B. die grosse Lebensmittelkette) oder langsame Wählverbindungen (z. B. das kleine Restaurant) verfügt. Die entsprechenden Inhalte lassen sich automatisch erzeugen.

- Inhaltsbasierte Umwandlung: Veränderbare Web-Objekte wie Grafiken bereiten einigen Anwendern Probleme, anderen hingegen nicht. Die Transformationsverfahren lassen sich so anpassen, dass sie die Access-Geschwindigkeit des Teilnehmers in Echtzeit ermitteln und dann entscheiden, ob Grafiken verändert werden müssen oder nicht. Der Lebensmittel- und Getränkegrosshändler könnte komplexe Grafiken für die Anwender mit niedriger Geschwindigkeit automatisch verkleinern. Diese Anwender bleiben dennoch produktiv, ohne lange auf anspruchsvoll gestaltete Seiten warten zu müssen. Das doppelte Design und Authoring von Websites kann gespeichert werden, um unterschiedliche Geschwindigkeit aufzufangen.
- Browserbasierte Umwandlung: Jeder Browser, manchmal sogar jede Version, hat eigene Auflösungsmuster. Die Inhalte lassen sich so organisieren, dass die Auflösungsgeschwindigkeit für den Browser jedes Anwenders optimiert wird. Bei der minimalen Darstellung komplexer Websites erhalten die Anwender kurze Response-

Zeiten, was die Wartezeit deutlich verkürzt.

- Caching: Herkömmliche Caching-Lösungen, die statische Inhalte nahe an den Edge des Netzwerks verlagern, verkürzen die Übertragungszeiten und vermeiden wiederholte Zugriffe auf die Server. Zudem lassen sich die unterschiedlichen Versionen des Content, die komprimiert und transformiert wurden, zur späteren Verwendung zwischenspeichern. Die nachfolgenden Anfragen sind noch schneller als das bereits beschleunigte Original.

Schritt 4: Bericht der Ergebnisse

Mit umfangreichen Report-Funktionen können Netzwerk-Manager die Management-Vorgaben an die Performance verfeinern, indem sie die Effizienz und Einhaltung von Service Level überwachen, nach historischen Trends suchen und die Kapazität planen. Grafiken und Reports beschreiben das aktuelle oder zurückliegende Netzverhalten und konzentrieren sich auf einzelne Applikationen, Anwender, URLs oder andere relevante Zusammenhänge (Bild 5).

Dieser Beitrag zeigt die Ergebnisse der Leistungsauslastung und Effizienz bei der Implementierung einiger Techniken, die in Schritt 3 (Steuerung) beschrieben wurden. Die Reports beantworten Fragen wie:

- Welche Response-Zeit hatte meine Oracle-Applikation im vergangenen Monat?
- Wie oft überschritt ein Client den ihm zugestandenen Anteil der Bandbreite des Service-Providers?
- Was verursachte gestern die langsamen Response-Zeiten, ein langsames Netzwerk oder ein langsamer Server? Von allen vier Schritten zum Bandbreiten-Management bietet das Reporting vermutlich den meisten Raum zur Problemlösung. Die meisten Tools zum Netzwerk-Management und Performance-Management ermöglichen mindestens einige Reports und Statistiken. Wenn die Tools ihre Statistiken in einem mit Kommas abgetrennten Format mit SNMP oder einer API zur Verfügung stellen, lässt sich mit jedem standardisierten Reporting-Paket eine attraktive Präsentation erstellen.

Eine Woche im Leben des Performance-Tunings

Das im obigen Szenario beschriebene Produktionsunternehmen plant den Ein-

satz einer SAP-Applikation und möchte deren Performance gewährleisten. Wie können die in diesem Artikel beschriebenen Verfahren und Features die IT-Abteilung bei einer erfolgreichen Umsetzung unterstützen?

1. Nach Installation und Test der Applikation misst die IT-Abteilung die Ausgangsbasis der Performance. Die Response-Zeit ist schnell und gleichmässig. Zu den besonderen Ergebnissen gehört:
 - End-to-End-Response-Zeit: 900 Millisekunden (etwas weniger als eine Sekunde)
 - Server-Komponente: 700 ms
 - Netzwerk-Komponente: 200 ms
 Die Testumgebung weist jedoch nur wenig konkurrierenden Datenverkehr auf und alle Anwender im LAN verfügen über einen Breitband-Anschluss ans WAN. Kein fairer Test.
2. Das Team nimmt mehrere zusätzliche Applikationen ins Netzwerk, darunter eine Menge Webtraffic, mehrere Napster-Downloads und E-Mails mit grossen Attachments. Zudem kommen mehrere Anwender in der Testumgebung hinzu, davon einige ohne Breitbandanschluss. Das Team arbeitet mit Klassifizierungsverfahren, um den gesamten Traffic jeder Applikation zu isolieren, sodass sich separate Performance-Reports erstellen lassen.

Eine zweiter Performance-Test für SAP bringt schlechtere Ergebnisse:

 - End-to-End-Response-Zeit: 7400 Millisekunden (über sieben Sekunden)
 - Server-Komponente: 800 ms
 - Netzwerk-Komponente: 6600 ms
 Wie die Statistiken belegen, verursacht die Verzögerung im Netz die unakzeptable Performance. Würde ein Service Provider die Applikationen des Herstellers vorhalten, könnten entsprechende Messungen dies ergeben:
 - End-to-End-Response-Zeit: 7400 ms
 - Server-Komponente: 4600 ms
 - Netzwerk-Komponente des Teilnehmers: 2000 ms
3. Einige Vorgaben zur Bandbreitenverteilung können die Performance wieder verbessern (Tabelle 2).
4. Ein weiteres Problem taucht auf: eine geschäftskritische Funktion auf der Website für Mitarbeiter und Lieferanten. Wird der Webtraffic mit der bestmöglichen Vorgabe von oben begrenzt, fällt die Performance für diese Applikation steil ab.

5. Mit Hilfe von Klassifizierungstechniken erkennt das IT-Team eine webbasierte Applikation mit einigen URLs als Unterscheidungskriterium. Danach werden die Vorgaben zur Bandbreitenverteilung gemäss Tabelle 3 verändert.
6. Die Performance der kritischen Website scheint nun korrekt zu sein. Bis ein Mitarbeiter in einer Aussenstelle darauf hinweist, dass der Inhalt der Website für die limitierte und von mehreren Mitarbeitern genutzte Verbindung zu gross ist; das Problem der letzten Meile.
7. Eine Beschleunigungslösung vor dem Webserver löst das Problem. Das Gerät übernimmt wahlweise die Komprimierung, Verwandlung und das Caching, um die Übertragung und Auflösung der Inhalte auf der Website für Anwender mit unterschiedlichen Access-Geschwindigkeiten vorzunehmen. Diese Vorgehensweise ist gewiss einfacher, als vom Technikerteam mehrere Website-Versionen für unterschiedliche Geschwindigkeiten entwickeln zu lassen. Und sie löst mehr Probleme für mehr Anwender.
8. Das Team definiert ein SAP Service Level Agreement mit einer Response-Zeit von 1,5 s. Die Anwender in der Zentrale erhalten eine Vorgabe von zwei Sekunden beim Zugriff auf die kritische Website, die Anwender in den Aussenstellen vier Sekunden.
9. Ein Report über die Einhaltung der Service Levels ergibt, dass diese durchgängig eingehalten werden. Sollten sich die Bedingungen verändern, verfügt das Team über ein leistungsfähiges Tool, um die Performance-Probleme zu lösen.

Die Lösung zur Applikations-Performance

Für eine Organisation, die eigene Applikationen vorhält, ist die Applikations-Performance eine wichtige Voraussetzung für Produktivität und Gewinn. Service-Providern ermöglicht die Applikations-Performance attraktive, differenzierte Dienstleistungen, die Kontrolle der Kosten und Möglichkeiten für weitere Umsätze. Gut integrierte Rahmenbedingungen gewährleisten die Performance von Applikationen und verfügen über die Informationen und Tools, um Probleme zu diagnostizieren, eine Strategie zu entwickeln, das Verhalten zu regeln und die Performance sicherzustellen. Quality of Service und Service Level Agreements

können sich nun vom Marketing-Verprechen zu echten, messbaren Bemühungen weiterentwickeln. 5

Arnold Pijpers, Vice President, EMEA Operations, Packeteer

Info:

Packeteer GmbH Deutschland
Peter Kirst
Obere Hauptstrasse 10b
D-85386 Eching bei München
Tel. +49 (0)89 370629-25
E-Mail: pkirst@packeteer.com

Summary

How to get Applications Running Efficiently

Competition for bandwidths on overburdened access routes can mean that large-scale applications such as Napster or large e-mail attachments take up so much capacity that important data traffic is interrupted. The performance of Web-based applications and e-commerce is further impeded by factors such as dial-up and broadband users with varying access speeds or a marked fall in performance with fewer connections per second once online transactions are supplemented by security measures. This article describes effective solutions for application performance and develops concepts for the outlined scenarios.