Auswirkungen der Virtualisierung auf Transparenz und Fehlerdiagnose in lokalen Netzen

Andreas Binzenhöfer, Kurt Tutschku

Lehrstuhl für Informatik III, Universität Würzburg
Am Hubland
97074 Würzburg
Germany
E-mail: [binzenhoefer,tutschku]@informatik.uni-wuerzburg.de

Die Virtualisierung im IT-Betrieb zielt darauf ab, vorhandene IT-Ressourcen zu einem gemeinsamen Pool zusammenzufassen und dabei den mit der Verwaltung der Ressourcen verbundenen Aufwand, sowohl administrativer als auch technischer Art, zu reduzieren und so insgesamt Kosten einzusparen. In unserem Beitrag zeigen wir die hiermit verbundenen Probleme und Gefahren auf und stellen einen ersten Lösungsansatz vor. Anschließend geben wir einen Überblick über bereits vorhandene Möglichkeiten, verdeutlicht am Beispiel des Freeware Pakets Nagios. Die Ausarbeitung wird schließlich mit einem kurzen Blick in die Zukunft abgerundet.

1 Einleitung

In den letzten zwei Jahrzehnten hat sich die IT-Landschaft rasant verändert. Verteilte Computersysteme werden immer komplizierter und komplexer. Der Trend geht eindeutig in Richtung immer leistungsfähigerer Komponenten. Grosse Teile der Forschung konzentrieren sich zudem darauf, bestehende Komponenten bis an ihre Grenzen hin optimal auszunutzen. Als Folge verlieren schon heute zahlreiche Firmen die Übersicht über Ihre internen Firmennetze. Die Komplexität der hochgradig vernetzten IT-Komponenten ist nur noch schwer überschaubar und von menschlichen Administratoren beinahe nicht mehr in den Griff zu bekommen. Das Zusammenspiel von Netzwerkkomponenten verschiedener Hersteller, sowie fehlende Standards verstärken diese Effekte zusätzlich. Als direkte Folge sind Ursachen für Systemausfälle und sonstige Fehler im Betrieb nur schwer zu lokalisieren und erinnern an die berühmte Nadel im Heuhaufen.

Im Folgenden wollen wir die Virtualisierung von IT-Ressourcen als eine der zurzeit wohl meist diskutierten Gegenmaßnahmen kurz aus unserer Sicht definieren und vorstellen. Der hier vorliegende Beitrag versteht sich als Konzeptstudie und möchte die offenen Fragen in Zusammenhang mit einer verstärkten Virtualisierung in Rechnersystemen und Rechnernetzen diskutieren. Neben den bekannten Vorteilen einer Virtualisierung möchten wir vor allem aber die Nachteile und Gefahren aufzeigen. Um diese Gesichtspunkte zu verdeutlichen, skizzieren wir dazu die Problemsituation in aktuellen Netzen am Beispiel einer Direktion eines großen deutschen Versicherungsunternehmens (ca. 14000 Anwender unternehmensweit und ca. 150 Einzelplatz- und Serverrechnern pro Direktion).

Als ersten möglichen Lösungsansatz umreißen wir kurz das neue Forschungsgebiet „Autonomic Networks“ [5] am Lehrstuhl III für Informatik der Universität Würzburg. Abschließend geben wir einen Überblick über bereits vorhandene Möglichkeiten, die Transparenz von abstrakten IT-Systemen aufrecht zu erhalten, verdeutlicht am Beispiel des Freeware Pakets Nagios [6].

2 Auswirkungen der Virtualisierung

Der Begriff Virtualisierung von IT-Ressourcen scheint derzeit allgegenwärtig zu sein. Er dient gleichermaßen als Thema für zahlreiche wissenschaftliche Veröffentlichungen und als Zugpferd großer IT-Firmen wie IBM, HP und SUN. Bisher gibt es jedoch keine zufrieden stellende einheitliche und allgemein akzeptierte Definition des Begriffes der Virtualisierung. Wir bemühen uns daher im nächsten Abschnitt um eine Definition der Virtualisierung von IT-Ressourcen. Weiterhin wird erläutert, was sich hinter dem Begriff der Netzwerk-Transparenz verbirgt und wie diese mit der Virtualisierung in Beziehung gebracht werden kann.

2.1 Definition Virtualisierung und Transparenz

Der Begriff Virtualisierung lässt sich als ein Zusammenfassen vorhandener IT-Ressourcen zu einem gemeinsamen virtuellen Pool beschreiben. Die beiden Hauptziele der Virtualisierung sind dabei die Reduktion des mit der Verwaltung der Ressourcen verbundenen Aufwands, sowie die Senkung der globalen Betriebskosten. Im Zuge der Virtualisierung wird eine weitere Abstraktionsebene in das bestehende IT-System eingefügt. Die neue virtuelle Komponente erscheint dem Benutzer als eigenständige, leistungsfähigere Einheit (siehe Abbildung 1).

Abb.1: IT-Ressourcen werden zu einem virtuellen Massenspeicher zusammengefasst

Die Virtualisierung der Ressourcen ist eng verbunden mit dem so genannten Grid Computing, bei dem die Ressourcen mehrerer vernetzter Computer zur gleichen Zeit zur Lösung eines bestimmten Problems eingesetzt werden. Diese Technik findet ihre Anwendung meist in der Forschung, wie z.B. bei der verteilten Suche nach Außerirdischen im Projekt seti@home der Universität Berkeley. Auch zahlreiche kommerzielle Firmen wie HP,

IBM und SUN bieten bereits erste Lösungen für auf mehrere Computer verteilte Massenspeicher an. Die virtuellen Speichersysteme erscheinen dem Administrator dabei als einheitlicher Massenspeicher, der in etwa mit einer virtuellen Festplatte vergleichbar ist. Die Speicherkapazität lässt sich von zentraler Stelle aus verwalten und bspw. verschiedenen Servern zuweisen. Schon durch diese gemeinsame Nutzung der Speicherressourcen lässt sich der physikalische Speicher wesentlich effizienter nutzen. Alles in allem beschreibt die Virtualisierung also die Emulation einer physikalischen Komponente mittels verschiedener Architekturen, Produkte und Geräte.

Im Gegensatz zur Virtualisierung zielt die Transparenz darauf ab, zu jeder Zeit die direkte Sicht auf die einzelnen Komponenten und deren Zustand zu gewährleisten. Das Ziel der transparenten Sicht auf das lokale Netzwerk besteht darin, die wichtigen Leistungsmerkmale mit angemessenem Aufwand überprüfen zu können und im Bedarfsfall schnell an der richtigen Stelle reagieren zu können. Ein transparentes Netzwerk kennt die Antworten auf Fragen der Art: „Welche Ressourcen sind zurzeit überlastet und stellen somit den Engpass dar?“, „Welche Komponenten müssen aufgerüstet bzw. erneuert werden?“ und „Wo entstehen die eigentlichen Kosten in meinem Netzwerk?“. Um weiterhin die Übersicht in komplexen IT-Systemen zu gewährleisten, werden meist verschiedene Sichten auf die gleichen Daten zur Verfügung gestellt. So benötigt ein Vorstandsmitglied bspw. andere Merkmale zur Entscheidungsfindung, als etwa ein Systemadministrator. Mit der zusätzlichen Abstraktionsebene der Virtualisierung läuft man nun aber Gefahr die Transparenz des eigenen Netzwerkes zu riskieren.

Aufgrund der oben angeführten Argumente kommen wir zu dem Schluss, dass ein Gleichgewicht zwischen Virtualisierung und Transparenz in lokalen Netzwerken erreicht werden muss, sodass eine effiziente Fehlerdiagnose ermöglicht wird. Das bedeutet, die wichtigen Leistungsmerkmale müssen vom Anwender beobachtbar bleiben, während die Virtualisierung weiterhin die einfache Nutzung der Ressourcen gewährleistet. Somit erscheint es unabdingbar, dass für komplexe IT-Systeme im Zuge der Virtualisierung parallel ein unabhängiges Überwachungs- und Diagnosesystem entwickelt wird. Zunächst erscheint es aber sinnvoll, die Vor- und Nachteile einer Virtualisierung im IT-Bereich gegeneinander abzuwägen.

2.2 Positive und negative Auswirkungen der Virtualisierung

Die positiven Auswirkungen der Virtualisierung auf den IT-Bereich liegen auf der Hand. Gerade in kommerziellen Systemen geht es zunächst primär darum Kosten einzusparen und den Return on Invest zu verbessern. Dies wird hauptsächlich durch die effizientere Ausnutzung vorhandener Ressourcen erreicht. Während bei einer herkömmlichen Server-Architektur für gewöhnlich mindestens 15% des Festplattenspeichers als Sicherheitsreserve einkalkuliert werden, lassen sich zusätzliche Festplatten und Speichersysteme ohne Betriebsunterbrechung in einen virtuellen Speicherpool integrieren. Durch die Möglichkeit im laufenden Betrieb unbemerkt in das System einzugreifen, werden bisher diffizile Datenspiegelungen, Datenduplizierungen und vor allem Datensicherungen deutlich vereinfacht.

Diese Ressourcenaustauschbarkeit garantiert zudem eine erhöhte Flexibilität und Verfügbarkeit. Weiterhin ermöglicht die Virtualisierung den Zusammenschluss einer heterogenen Serverlandschaft und verhindert somit die Abhängigkeit von einzelnen Herstellern. Unternehmen erhalten hierdurch einen größeren Spielraum und mehr Freiheit in der Gestaltung ihrer Hardware-Ausstattung.

Auf der anderen Seite darf man allerdings nicht die möglichen Gefahren und Nachteile, die mit der Einführung einer weiteren Abstraktionsebene unweigerlich einhergehen, vernachlässigen. Der Grund der bisher am stärksten gegen eine Virtualisierung von IT-Ressourcen spricht, ist das Fehlen allgemein akzeptierter Standards in diesem Bereich. Die Open Grid Service Architecture (OGSA) stellt einen ersten Schritt in die richtige Richtung dar, von einem marktreifen Konzept ist man allerdings noch weit entfernt. Erste Erfahrungen in lokalen Netzwerken zeigen zudem, dass die Virtualisierung ab einer gewissen Größenordnung nicht unbedingt die angestrebte Skalierbarkeit mit sich bringt. Alleine die Organisation und Einteilung der einzelnen Ressourcen birgt einen nicht zu vernachlässigenden Overhead in Form eines zusätzlichen Verwaltungsaufwands in sich. Im Falle kritischer Anwendungen und sensibler Daten kommen zusätzlich sicherheitsrelevante Aspekte mit ins Spiel. So stellt etwa die gemeinsame Nutzung einer gemeinsamen Ressource durch zwei konkurrierende Prozesse ein weiteres Sicherheitsrisiko dar.

Die aus unserer Sicht größte Gefahr einer weiteren Abstraktion durch Virtualisierung liegt jedoch in dem Verlust der Transparenz in lokalen Netzwerken. Eine virtualisierte Komponente verhält sich zunächst für den Benutzer wie eine Art Black Box, über die er nur beschränkte Kontrolle besitzt. Fehlerdiagnose und Lokalisierung von Schwachstellen und Ursachen werden hierdurch zusätzlich erschwert. Die Virtualisierung der Ressourcen reduziert nun also als Nebeneffekt die direkte Sicht auf die einzelnen Komponenten und deren Zustand. Aus diesem Grund ist es notwendig, einen Kompromiss zwischen Virtualisierung, im Sinn der Vereinfachung der Ressourcenaustauschbarkeit und Ressourcenverwaltung, und der Transparenz der Systemzustände zu finden. Das bedeutet, die wichtigen Leistungsmerkmale bleiben vom Anwender beobachtbar, während die Virtualisierung weiterhin die einfache Nutzung der Ressourcen gewährleistet. Im nächsten Abschnitt umreißen wir daher kurz die Problemsituation in aktuellen Netzen, um dieses Argument am konkreten Beispiel zu verdeutlichen.

2.3 Problemsituation in aktuellen Netzen

Bestehende Netzwerke enthalten, selbst wenn sie eine ausgeprägte Client-Server Architektur aufweisen, zahlreiche Fehlerquellen und sind von zentraler Stelle aus nur schwer zu überschauen. Die Einführung einer weiteren Abstraktionsebene durch die Virtualisierung einer gewissen Anzahl von Einzelressourcen maskiert nun diese Fehlerquellen und fügt damit eine weitere Sichtbarriere bei der Fehlerdiagnose in lokalen Netzwerken ein.

Um die aktuelle Problemsituation zu dokumentieren und in die Entwicklung neuer Administrationsverfahren mit einbeziehen zu können, wurden in einer Direktion eines großen deutschen Versicherungsunternehmens (ca. 14000 Anwender mit ca. 150 Einzelplatz- und Serverrechnern pro Direktion) Messungen durchgeführt und Logfiles ausgewertet. Dabei wurde festgestellt, dass sich typische Fehler und Leistungsengpässe in heutigen lokalen

Netzen durch unpräzise Fehlersymptome auszeichnen. Diese werden weiterhin vom Benutzer meist nur vage zum Beispiel als ein „nicht reagierendes Netzlaufwerk“, eine „langsame Web-Site“, oder eine „geringe Dateiübertragungsleistung“ beschrieben. Die ungenaue Beschreibung erschwert die Diagnose und Behebung der Fehler. Bei genauer Betrachtung stellen sich diese Fehler häufig jedoch als ein Konglomerat verschiedener einzelner Ursachen dar. Alleine das Zusammenspiel von Netzwerkkomponenten verschiedener Hersteller kann unvorhersehbare Folgen auf das Leistungsniveau eines IT-Systems haben.

Eine der Hauptursachen für Probleme in lokalen Netzwerken findet sich in den zahlreichen Einstellungsmöglichkeiten und Konfigurationsdateien, die über das komplette Netzwerk verteilt liegen und nicht automatisch verwaltet werden. Überlegungen und Ideen hinter der aktuellen Konfiguration sind dem System meist nicht bekannt und spiegeln sich ausschließlich in den Köpfen weniger Mitarbeiter wider. Darüber hinaus führt das System keine eigenständigen Plausibilitätschecks der aktuellen Konfiguration durch. Ein DNS-Server in einem nicht erreichbaren Subnetz fällt somit erst im laufenden Betrieb auf. Die bisherige Fehlerdiagnose findet häufig redundant und dezentral statt. Mehrere Administratoren suchen parallel die Ursache für ein und dasselbe Problem. In unserem Forschungsgebiet „Autonomic Networks“ arbeiten wir daher an Lösungen, die die Fehlerdiagnose und -behebung automatisieren, Netzwerke an den richtigen Stellen transparent halten und sich mit der Virtualisierung von IT-Ressourcen in Einklang bringen lassen.

3 Unsere Vision: Autonomic Networks

Das Forschungsgebiet Autonomic Networks [5] beschäftigt sich mit der Entwicklung selbstorganisierender Algorithmen. Das erklärte Ziel sind Computernetzwerke, die sich selbstständig konfigurieren, verwalten und reparieren.

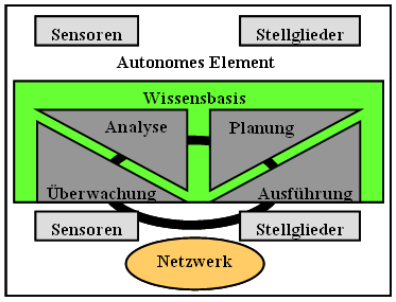

Abb.3: Ein Autonomes Netzwerk-Element nach dem Vorbild von IBM

Abbildung 3 zeigt den Regelkreis einer autonomen Netzwerkkomponente nach dem Vorbild der Firma IBM [4]. Jede Netzwerkkomponente ist mit Sensoren ausgestattet, über die von einer Überwachungseinheit aktuelle Systemzustände erfragt werden können. Die Analyseeinheit interpretiert die gewonnen Daten und vergleicht die Ergebnisse mit einer Wissensbasis. Im gegebenen Fall wird die Planungseinheit verständigt, die eine entspre-

chende Lösung erarbeitet. Die nötigen Gegenmaßnahmen werden letztendlich von der Ausführungseinheit über die Stellglieder der Netzwerkkomponente eingeleitet. Dabei ist zu beachten, dass es sich nicht um einen strengen Regelkreis handelt, d.h. die einzelnen Komponenten können bei Bedarf bspw. neue Messdaten von der Überwachungseinheit anfordern.

Durch diese autonome Verwaltung der Netzwerkkomponenten wird die Ausfallsicherheit des gesamten Systems verbessert. Administratoren erhalten im Optimalfall ausschließlich Fehlermeldungen, die einen manuellen Eingriff zwingend erfordern, wie etwa im Falle einer defekte Hardwarekomponente. Im Hinblick auf die Virtualisierung lassen sich die gefundenen Ergebnisse zur Selbstverwaltung des virtuellen Ressourcenpools einsetzen. Da sich Ausfälle von Komponenten und Softwarefehler nicht gänzlich vermeiden lassen, kommen zusätzlich Algorithmen wie etwa das Tool Pinpoint der Universität Berkely [3] zum Einsatz, um Fehlerursachen genauer lokalisieren zu können. Im nächsten Abschnitt geben wir am Beispiel Nagios [6] einen kurzen Überblick über die aktuellen Möglichkeiten zur transparenten Fehlerdiagnose.

4 Der erste Schritt in Richtung Autonomic Networks

Die Fehlerdiagnose in aktuellen Netzwerken geschieht im Normalfall mit Hilfe von Softwarepaketen wie z.B. Tivoli oder HP Openview. All diesen Tools ist gemeinsam, dass sie sich auf messbare Daten stützen müssen. Angefangen mit einfachen pings, über snmp-Anfragen bis hin zu port-Tests wird alles ausgenutzt, was sich aktiv überprüfen lässt. In diesem Abschnitt möchten wir am Beispiel des Freeware-Pakets Nagios aufzeigen, was mit den herkömmlichen Mitteln möglich ist und wie sich die Fehlerdiagnose in Richtung Selbstverwaltung ausdehnen lässt.

Wie in Abbildung 3 zu sehen, stellt Nagios das lokale Netzwerk übersichtlich auf einen Blick dar. Problemzonen werden dabei rot gekennzeichnet. Neben den üblichen ping-Anfragen, die überprüfen, ob ein Computer noch ansprechbar ist, besteht die Möglichkeit aktiv Services auf externen Rechnern zu überprüfen. Nagios sendet hierzu bspw. eine http-Anfrage an den dedizierten Webserver und bereitet die Antwort grafisch für den Administrator auf. In Fehlersituationen (z.B. Papierstau im lokalen Drucker) benachrichtigt der Nagios-Prozess die zuständige Kontakt-Gruppe per eMail, ICQ oder SMS. Externe Messgrößen, wie z.B. der belegte Plattenplatz, können über einen lokal installierten Daemon, der periodisch aktiviert wird, abgefragt und an den Nagios-Server übermittelt werden. Dieses Beispiel beschreibt, wie Nagios das so genannte „Push“-Konzept umsetzt. Beim „Push“-Konzept werden die Daten nicht erst auf Anfrage des Managementsystems übermittelt. Das überwachte System führt vielmehr eine Selbstdiagnose durch, sammelt die Daten, bereitet sie auf und übermittelt sie selbstständig an die Managementstation. Das „Push“-Konzept steht somit im Kontrast zum „Pull“-Konzept, das in den heute üblichen Management-Architekturen, wie zum Beispiel SNMP, verwendet wird. Da beim „Push“-Modus keine Anfragen versendet werden, wird der Kommunikationsaufwand im Netz zusätzlich reduziert. Die Vorverarbeitung der Daten vermindert die zusätzliche Belastung der Managementstation und erlaubt eine genauere Analyse der Daten. Das „Push“-

Konzept reduziert somit die Fehleranfälligkeit der Managementstation und erhöht ihre Leistungsfähigkeit.

Der Hauptvorteil des Open Source Projektes Nagios liegt nun in seiner Konfigurierbarkeit und Erweiterbarkeit. Über externe Skripten lässt sich die Fehlerdiagnose an individuelle Bedürfnisse anpassen. Auf diese Art lassen sich gezielt die für das jeweilige Netzwerk entscheidenden Größen überwachen.

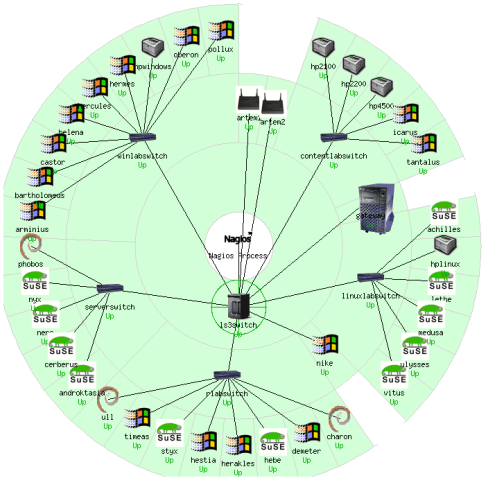

Abb.3: Das Lehrstuhl-Netz in der übersichtlichen Nagios Statusmap

Weiterhin erlaubt eine einfache Benutzerverwaltung verschiedene Sichten auf das Netzwerk. Administratoren können sich somit gezielt auf den für sie interessanten Bereich konzentrieren, wodurch eine effizientere Fehlerdiagnose ermöglicht wird.

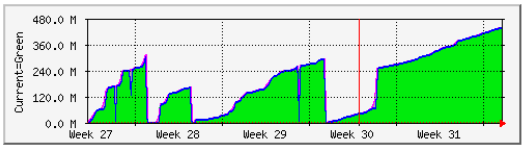

Abb.4: Zahlreiche Statistiken erzeugen die gewünschte Transparenz

Durch systematische Messungen lassen sich außerdem umfangreiche Statistiken (siehe z.B. Abbildung 4) erzeugen, die das lokale Netzwerk an den gewünschten Stellen transparent gestalten.

Die so genannten Service-Escalations stellen den ersten Schritt in Richtung „Autonomic Networks“ dar. Hier lassen sich automatische Gegenmaßnahmen für mögliche Fehlerszenarien im Vorhinein festlegen. Eine mögliche Anwendung wäre etwa den Webserver im

Fehlerfall ohne menschlichen Eingriff automatisch neu zu starten. Eine Meldung an den zuständigen Kontakt findet erst statt, falls der Neustart das ursprüngliche Problem nicht beheben konnte. Wie aber lässt sich dieses Überwachungs- und Diagnosesystem nun zur Lösung des Konfliktes zwischen der Virtualisierung und der Transparenz in lokalen Netzwerken einsetzen? Mit Hilfe der oben erwähnten Verfahren besteht die Möglichkeit einzelne Teil-Ressourcen eines virtualisierten Ressourcenpools individuell zu überwachen. Je nach Abstraktionsebene und gewünschter Transparenz, kann ein Fehler in einer Teilkomponente als Ausfall der gesamten virtuellen Black Box oder transparent mit detaillierter Beschreibung gemeldet werden. Mit Hilfe verschiedener Sichten auf das Netzwerk, lassen sich also die Vorteile einer Virtualisierung ausnutzen, ohne dabei die Kenntnis über die internen Systemzustände zu verlieren.

5 Ein Blick in die Zukunft

Abschließend lässt sich zusammenfassen, dass die Virtualisierung von IT-Ressourcen ein sinnvoller Schritt ist, um Benutzern unnötig wiederkehrende Aufgaben abzunehmen, Systeme effizienter ausnutzen und Kosten einzusparen. Im Gegenzug droht die Gefahr, die Transparenz, in dem Sinn, dass die wichtigen Leistungsmerkmale beobachtbar bleiben, zu verlieren.

Aktuelle Fehlerdiagnosesysteme verstehen es bereits sehr gut, die zur Verfügung stehenden Messdaten effizient auszuwerten. Um jedoch auch in Zeiten der Virtualisierung eine transparente Selbstverwaltung zu ermöglichen, sind neue Messtechniken erforderlich. Zum einen werden Informationen benötigt, die mit bisherigen Mitteln nicht erfasst werden können, angefangen mit der Temperatur einzelner kritischer Hardwarekomponenten bis hin zu quantitativen Aussagen über die Qualität verlegter Kabel. Zum anderen müssen überflüssige manuelle Prozesse, wie etwa eine redundante Konfiguration ähnlicher Komponenten, automatisiert werden. Mit den entsprechenden Grundvorrausetzungen und dem nötigen Weitblick kann eine transparente Virtualisierung also Realität werden.

Literatur

[1] D. Ruzicka, *Network Management with Nagios, Netsaint's Successor: What's Going On?*, Linux Magazine, Ausgabe 29, Seite 62 (April 2003)

[2] R. C. Harlan, *Aus der Praxis: Netzwerkmanagement, Wer sucht, der findet*, Linux Administrator, Ausgabe Juli 2003, Seite 1 (Juli 2003)

[3] D. A. Patterson et al, *Recovery-Oriented Computing (ROC): Motivation, Definition, Techniques, and Case Studies*, UC Berkeley Computer Science Technical Report UCB//CSD-02-1175 (März 2002)

[4] IBM, *An architectural blueprint for autonomic computing*, http://www-3.ibm.com/autonomic/r/downloads/blueprint/index.html (April 2003)

[5] A. Binzenhoefer, K. Tutschku, *Autonomic Networks*, http://www3.informatik.uni-wuerzburg.de/research/autonomic.shtml

[6] Nagios, http://www.nagios.org