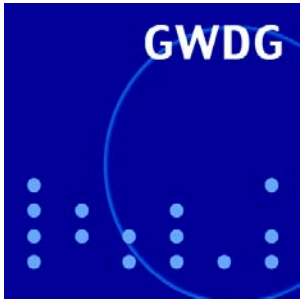


GWDG-Bericht Nr. 57

Helmut Hayd, Rainer Kleinrensing
(Hrsg.)

17. und 18. DV-Treffen der Max-Planck-Institute

22. - 24. November 2000
21. - 23. November 2001
in Göttingen

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Helmut Hayd, Rainer Kleinrensing (Hrsg.)

17. und 18. DV-Treffen der Max-Planck-Institute

22. - 24. November 2000
21. - 23. November 2001
in Göttingen

Helmut Hayd, Rainer Kleinrensing (Hrsg.)

17. und 18. DV-Treffen der Max-Planck-Institute

22. - 24. November 2000
21. - 23. November 2001
in Göttingen

GWDG-Bericht Nr. 57
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
Am Faßberg
D-37077 Göttingen
Telefon: 0551-201-1510
Telefax: 0551-21119
E-Mail: gwdg@gwdg.de

Satz: Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Druck: Offset- und Dissertations-Druck Jürgen Kinzel, Göttingen-Weende

ISSN 0176-2516

Inhalt

Vorwort

Vom 22. - 24. November 2000 und vom 21. - 23. November 2001 fanden bei der GWDG in Göttingen das 17. und 18. DV-Treffen der Max-Planck-Institute statt, auf dem die mit Betrieb und Planung von Instituts-EDV befassten Mitarbeiterinnen und Mitarbeiter Gelegenheit hatten, sich über aktuelle Entwicklungen innerhalb der MPG auszutauschen bzw. vorgestellte Problemlösungen für das eigene Institut zu adaptieren.

Der vorliegende Band enthält ausführliche Artikel zu einigen der gehaltenen Vorträge; die (elektronischen) Folien fast aller Beiträge des 18. Treffens sind zusätzlich unter http://www.gwdg.de/dv-treffen2001/programm.html verfügbar. Für einige der Vorträge konnten externe Vortragende gewonnen werden, so z. B. 2001 mit Herrn Siering ein Redakteur der Zeitschrift c't sowie 2000 mit Herrn Krägelin der für Sicherheitstechniken Verantwortliche der FHG. Weiterhin haben wir versucht, durch die Vorträge von Herrn Bastian - Leiter Einkauf bei der GV - eine Brücke zu den auch für EDV-Schaffende immer wichtiger werdenden Verwaltungsvorschriften, z. B. bei Ausschreibungen, zu schlagen. Durch den Vortrag von Frau Velden wurden Ausblicke auf die interessanten Fragen elektronischer Dokumente und Zeitschriften sowie virtueller Bibliotheken gegeben.

Zum Schluss möchten wir uns bei den Teilnehmern und natürlich bei allen Vortragenden für das rege Interesse an dieser Veranstaltung bedanken. Ganz besonderer Dank gebührt - last, but not least - Herrn Otto und seinen Kolle-

ginnen und Kollegen bei der GWDG, die durch die exzellente lokale Organisation und durch viel Detailarbeit wesentlich zum Gelingen der Tagung und zur Entstehung dieses Berichtes beigetragen haben.

Leipzig, im Oktober 2002 Helmut Hayd, Rainer Kleinrensing

Teil 1: Beiträge vom 17. DV-Treffen

Lotus Notes bei der GWDG

Wilfried Grieger

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Einleitung

Die Verwendung von Groupware-Lösungen in den wissenschaftlichen Instituten und Abteilungen wird auf Grund der zunehmenden Datenvielfalt und Datenkomplexität immer wichtiger. Arbeitsabläufe müssen nicht nur in kommerziellen Einrichtungen sondern auch in der Wissenschaft optimiert werden. Dies wird häufig schon von Gutachtern gefordert.

1. Was ist Groupware?

Was man unter dem Begriff „Groupware“ verstehen sollte, darüber gibt es unterschiedliche Meinungen. Eine genormte Definition ist leider noch nicht erarbeitet worden. Eine zutreffende Beschreibung ist die folgende:

> ***Groupware*** ist eine Software, deren Technologie die **Kommunikation**, **Kollaboration** und **Koordination** einer Gruppe von Benutzern erleichtert.

Diese „Technologie der drei K“ wurde einer Diplomarbeit[1] entnommen, die einen möglichen Übergang eines herkömmlichen WWW-Servers in eine Groupware-Umgebung beschreibt.

2. Groupware-Systeme

Mittlerweile gibt es eine ganze Reihe von Groupware-Systemen, also Software, die die oben erwähnten Kriterien erfüllt. Bei der GWDG wurden bis zum Jahr 2000 die folgenden drei Systeme getestet und miteinander verglichen:

1. **Netscape Professional**

   Netscape Professional bot zusammen mit dem integrierten Netscape WWW-Browser eine bequeme Weise an, Groupware-Lösungen über das WWW zu realisieren. Inzwischen ist die Groupware-Variante von Netscape nicht mehr unter diesem Namen verfügbar.

2. **Microsoft Outlook/Exchange**

   Die Kombination von Microsoft Outlook mit Exchange ist sicherlich die bekannteste verfügbare Groupware-Lösung. Andere Microsoft-Produkte lassen sich einfach in dieses System integrieren bzw. sind automatisch integriert.

3. **Lotus Notes/Domino**

   Die mittlerweile von der Firma IBM übernommene Lotus-Software zählt zu den ältesten und am weitesten verbreiteten Groupware-Systemen. Laut Lotus Deutschland gab es Anfang 1998 weltweit über 28 Millionen Einzelplatzlizenzen von Lotus Notes; der nächste Mitbewerber kam zu der Zeit nur auf 13,5 Millionen verkaufter Lizenzen. „Notes stellt damit im Sektor Workgroup Computing Platform den Quasi-Standard der Industrie dar.“[2]

3. Anforderungen an ein Groupware-System

Da die heute verfügbare Groupware in der Regel extrem komplex aufgebaut ist und damit natürlich auch in ein allumfassendes System ausgeweitet werden kann, ist es wichtig, vor der Entscheidung für eine spezielle Software einen Anforderungskatalog zu erstellen.

Für die GWDG sind die folgenden vier Kriterien entscheidend für den Einsatz einer Groupware-Lösung:

1. Erik Rolshausen, Migration dateibasierter Webserver in eine Groupware-Umgebung. Diplomarbeit der Fachhochschule Karlsruhe - Hochschule für Technik, Fachbereich Wirtschaftsingenieurwesen, 1998, p. 4 ff.
2. Erik Rolshausen, a. a. O., p. 9

1. Unter dem Groupware-System muss ein umfangreiches Kalender-Management geführt werden können. Sowohl Einzel- als auch Gruppenkalender müssen verfügbar sein. Zugriffsrechte müssen detailliert festgelegt werden können.
2. Die Einzelkalender müssen mit sogenannten Personal Digital Assistents (PDAs), vorzugsweise Palm Pilot, synchronisiert werden können.
3. Das Groupware-System muss ein Datenbank-System beinhalten, dessen Zugriffsrechte flexibel gesetzt werden können. Die Datenbanken müssen auf einfache Art und Weise erstellt und zentral gehalten werden können.
4. Auf die Kalender und die Datenbanken muss auch über das WWW gesichert zugegriffen werden können.

4. Lotus Notes/Domino

Da die Synchronisation von PDAs mit Netscape Professional nicht zufrieden stellend durchgeführt werden konnte, blieben nur die beiden Systeme Microsoft Outlook/Exchange und Lotus Notes/Domino als Alternativen übrig. Diese beiden erfüllen im Funktionsumfang die oben aufgeführten Kriterien. Um nun an Hand objektiver Kriterien die Entscheidung für eines der beiden Systeme herbeiführen zu können, hätten beide Systeme intensiv getestet werden müssen. Dafür standen jedoch die personellen Ressourcen nicht zur Verfügung.

Die GWDG hat sich nun für Louts Notes/Domino entschieden, weil ein Verlassen eines Microsoft-Systems schwieriger zu sein scheint als eines Lotus-Systems, falls die Entscheidung zu Gunsten des einen revidiert werden müsste.

In der Industrie ist immer noch die Version 4.6 am weitesten verbreitet, weil der Übergang zur Version 5 einen erheblichen Aufwand bedeutet. Die GWDG hat von Anfang an gleich die Version 5 von Lotus Notes/Domino eingesetzt und damit einen großen Migrationsaufwand umgangen.

Um nun auch die Verbreitung von Lotus Notes/Domino innerhalb des Forschung-und Lehre-Bereichs zu fördern, ist die GWDG Mitglied in der Deutschen Notes User group e. V. (DNUG).

Lotus Notes/Domino wird in den Hochschulen und anderen wissenschaftlichen Einrichtung noch sehr wenig genutzt.

5. Was bietet Lotus Notes/Domino?

Lotus Notes/Domino ist ein Client-Server-System, bei dem die Clients „Lotus Notes“ und die Server „Domino“ genannt werden. Die folgenden Eingeschaften und Funktionalitäten zeichnen das System aus:

- Terminplanung
- Gruppenkalender
- Aufgabenverwaltung
- Adressverwaltung
- Synchronisation mit PDAs
- Memos (Mails)
- Datenbanken
- Dokumentverwaltung

In der Regel erfolgt der Zugriff auf die Terminkalender, auf die Aufgaben, Adressen und Datenbanken über den Lotus-Notes-Client, und zwar verschlüsselt. Die Verschlüsselung wird mit einer persönlichen ID-Datei des Anwenders gesteuert.

Da der Domino-Server ein eigenständiger WWW-Server ist, lässt sich der Zugriff auf die Terminkalender, auf die Aufgaben, Adressen und Datenbanken auch über einen WWW-Browser realisieren, und zwar bei Bedarf auch verschlüsselt.

Die Zugriffsrechte lassen sich sowohl über den Lotus-Notes-Client als auch über das WWW individuell festlegen.

6. Lotus-Notes/Domino-Lizenzen

Sowohl eine Lotus-Notes-Client-Lizenz als auch eine Domino-Server-Lizenz kostet innerhalb des auch für die Max-Planck-Gesellschaft gültigen Lizenzprogramms Lotus-Academic-Solution (LAS) für Forschung und Lehre **9 DM** zuzüglich Mehrwertsteuer.[3] In der Lizenz enthalten ist der kostenlose Update innerhalb von zwei Jahren.

Die Lizenzen können sowohl bei der Firma Steckenborn e-com in Gießen als auch über den Software-Shop der GWDG

`https://gwdg.asknet.de`

erworben werden, der von der Firma ASKnet in Karslruhe bereit gestellt wird.

7. Einsatz bei der GWDG

Das Lotus-Notes/Domino-System wird bei der GWDG hautpsächlich als Kalender-Management-System eingesetzt. Einzelkalender werden dabei mit den vorhandenen PDAs synchronisiert. Für die Belegung der Kursräume und die Entleihe von Geräten werden ebenfalls Kalender verwendet, die von verschiedenen Mitarbeiterinnen und Mitarbeitern ausgefüllt werden. Lesend kann von einer größeren Gruppe auf diese Kalender zugegriffen werden.

Besprechungsprotokolle werden in einer Datenbank niedergelegt. Dabei sorgt ein ausgeklügeltes integriertes Wiedervorlagesystem dafür, dass Beschlüsse auch durchgeführt werden und keinesfalls vergessen werden können. Über die Ergebnisse kann innerhalb einer eigenen Datenbank sogar im Mitarbeiterkreis diskutiert werden.

8. Einsatz bei der GWDG/MPG geplant

Zur Zeit sind die folgenden Einsatzgebiete des Lotus-Notes/Domino-Systems innerhalb der GWDG und der MPG geplant:

Ein Projektverfolgungssystem soll die Einzelarbeiten an Projekten transparenter und effizienter gestalten.

---

3. Leider hat die Firma IBM im Mai 2001 das Lizenzprogramm LAS für Forschung und Lehre ohne Begründung gekündigt. Lotus hat jedoch versichert, dass ein ähnliches Programm mit ähnlichen Preisen demnächst neu aufgesetzt werden wird.

Die von der GWDG angebotenen Kurse sollen in eine Datenbank integriert werden, so dass Anmeldungen über das WWW ermöglicht werden und die Kursplanung und -organisation weitestgehend automatisiert wird.

Der von der GWDG bereit gestellte Dienstleistungskatalog, der bisher ausschließlich in gedruckter Form und über WWW-Dokumente verfügbar war, soll ebenfalls als Datenbank erfasst und komplett in das WWW übernommen werden.

Der elektronische BAR, der bisher für die Mitglieder die Anträge und Sitzungsprotokolle als WWW-Dokumente zur Verfügung stellte, soll ebenfalls in eine Lotus-Datenbank integriert werden. Damit ist dann auch eine Volltextsuche automatisch gewährleistet.

Adressenlisten von Institutsmitarbeiterinnen und -mitarbeitern lassen sich auf eine bequeme Weise erzeugen und nach den Wünschen der Eingetragenen in das WWW stellen.

9. Kurse zu Lotus Notes/Domino

Die GWDG bietet nun auch Kurse für Anfänger zu Lotus Notes/Domino an. Zunächst wird dabei das gesamte Kalender-Management, einschließlich der Aurfgaben- und Adressverwaltung vorgestellt. Danach erfolgt die Einführung in das Erstellen von eigenen Datenbanken mit dem Domino Designer, der ebenfalls in jeder Lizenz enthalten ist.

Parallelrechnen bei der GWDG: Erfahrungen mit IBM RS/6000 SP

Oswald Haan

Gesellschaft für wissenschaftliche Datenverabeitung mbH Göttingen

1. Einleitung

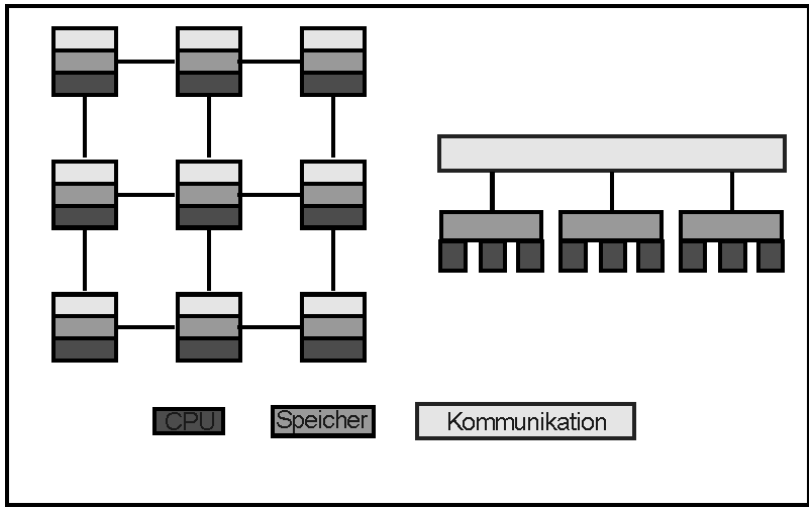
Zu Anfang des Jahres 2000 wurde bei der GWDG ein Parallelrechner der Firma IBM vom Typ RS/6000 SP in Betrieb genommen. Mit 160 Power3-Prozessoren, einer Peakperformance von 240 Gflop/s und einem Hauptspeicher-Ausbau von 120 GB rangierte dieses System auf Platz 110 der TOP500-Liste vom November 2000. In der neuesten TOP500 Liste vom Juli 2001 ist das System auf den 137. Platz zurückgefallen. Berücksichtig man jedoch die Aufstockung des Systems um 64 Prozessoren, die zu Anfang des Jahres 2001 erfolgte, so rückt unsere SP in dieser Liste auf Platz 107 vor. Unter Berücksichtigung dieses Ausbaustandes nimmt das GWGD-System in Deutschland den Platz 9 ein.

Für die GWDG ist die SP die vierte Generation von Parallelrechnern, nach der KSR1 (1993), der SGI Power Challenge (1995) und der Cray T3E (1997). In diesem Bericht soll ein Überblick über die Rechnerarchitektur und Leistungsfähigkeit des SP-Systems gegeben werden, das Betriebsmodell beschrieben werden und erste Nutzungserfahrungen dargestellt werden.

2. Architektur

Die SP hat gegenüber den Parallelrechnern der vorigen Generation, für die die Cray T3E den Prototypen darstellt, einen Architekturwechsel vollzogen, der auch in den Systemen aller anderen Hersteller von Parallelrechnern zu beobachten ist. Der Rechner ist nicht mehr ein Multiprozessorsystem mit verteiltem Speicher, sondern ein Multi-SMP-System mit verteiltem Speicher (s. Abb. 1).

Abb. 1: Vergleich Multiprozessor-System Cray T3E und Multi-SMP-System IBM SP

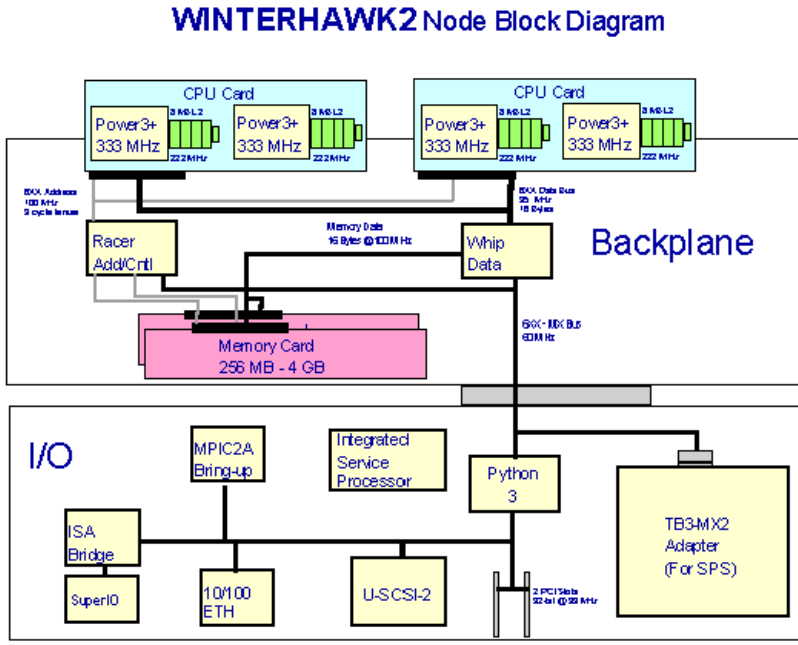


Das Charakteristische der neuen Parallelrechnerarchitektur ist ihr modularer Aufbau, bei dem Rechnerhardware und Kommunikationshardware voneinander unabhängige Komponenten sind. Bei der Cray T3E waren spezielle Register und Netzwerk-Router für die Kommunikation auf dem Rechnerknoten integriert. Für die Rechnerknoten der SP hingegen werden Standard-Rechnerkomponenten verwendet, die auch in kommerziellen Servern eingesetzt werden. Die Kommunikation ist in einen externen Switch ausgelagert worden, der mit jedem Rechnerknoten über ein Interface kommuniziert. Damit kann in der SP die jeweils neueste und leistungsstärkste Rechnerhardware eingesetzt werden, deren Entwicklung durch den großen kommerziellen Server-Markt getragen wird. Die leistungsfähigsten Server sind heute SMP-Systeme, die deshalb als Knoten im SP-System dienen. Die Entwicklung der Switch-Technologie wird hingegen im wesentlichen durch den sehr

viel kleineren technisch-wissenschaftlichen Markt getragen und fällt deshalb gegenüber der Rechnertechnologie zurück.

Diese Entwicklung führt also zu Parallelrechnern mit einer maximalen CPU-Leistung, aber einer durch die ökonomischen Randbedingungen, nicht durch den Stand der Technologie, beschränkten Kommunikationsleistung.

Abb. 2: Blockbild des Winterhawk2-Knotens



Der einzelne Rechnerknoten der SP vom Typ Winterhawk2 besteht aus vier Prozessoren der Power3-Linie, mit 375 MHz getaktet, von denen jeweils zwei auf einem Prozessorboard zusammengefasst sind (s. Abb. 2). Im SP-System der GWDG ist jeder Knoten mit 3 GB Hauptspeicher versehen. Auf diesen Speicher greifen alle vier Prozessoren über einen mit 375/4 = 93,75 MHz getakteten seriellen Bus mit einer Bandbreite von 1,5 GB/sec zu. Diese Speicherbandbreite ist bei speicherintensiven Anwendungen für einen Prozessor gerade ausreichend, müssen sich aber vier Prozessoren diese Bandbreite teilen, wird der Speicherzugriff zum Engpass. Die später dargestellten Leistungsmessungen werden dies bestätigen.

Der Power3-Prozessor hat eine superskalare RISC-Architektur mit out-of-order Befehlsausführung, Verzweigungsvorhersage und SMP-Unterstützung

durch ein 4-Zustands Protokoll zur Aufrechterhaltung der Cache Kohärenz. Jeder Prozessor besitzt einen eigenen Level1-Cache mit 64KB für Daten und 32 KB für Instruktionen mit 128-fach assoziativer Organisation, sowie einen 4-fach assoziativen Level2-Cache für Daten und Instruktionen der Größe 8 GB. Der Prozessor hat 2 Floating-Point-Einheiten für verkettete Multiplikation-Addition mit einer Latenz von3-4 Zyklen. Da jede Einheit pro Zyklus ein Ergebnis der verketteten Mult-Add-Operation liefern kann, liegt die Spitzenleistung des Prozessors bei 4 Flop/Takt oder 1,5 Gflop/s. Die Floating-Point Einheiten führen eine Division in 18-25 Zyklen, das Ziehen einer Wurzel in 22-31 Zyklen aus.

Für Festkomma-Operationen stehen 3 Rechen-Einheiten zur Verfügung. Zwei Load/Store-Einheiten bewegen Daten zwischen Level1 Cache und Registern, wobei gleichzeitig zwei Lade-, eine Lade- und eine Speicher- oder eine Speicher-Operation durchgeführt werden kann. Die Zugriffszeiten -und Bandbreiten zwischen den verschiedenen Stufen der Speicherhierarchie sind in der Abb. 3 zusammengefasst.

Abb. 3: Speicherhierarchie des Power3-Prozessors

Zugriff	Latenz [Zyklen]	Breite [bits]	Taktrate [MHz]	Bandbreite [GB/sec]
Laden von L1 Cache	1	128	375	6
Speichern nach L1 Cache	1	64	375	3
L1 <-> L2	9	256	250	8
L1 <-> Speicher	70	128	93.75	1.5

Der Zugriff zum Hauptspeicher wird durch einen von der Hardware unterstützten Prefetch-Mechanismus beschleunigt. Sobald Cache-Misses auf aufeinanderfolgende Cache-Zeilen registriert werden, setzt die Hardware vorsorglich das Laden der folgenden Cache-Zeilen in Gang, um so die hohe Latenz des Speicherzugriffs zu verbergen. Insgesamt kann die Hardware bis zu vier solcher Prefetch-Ströme verwalten.

3. Leistungsmessungen

3.1 Einzelprozessor

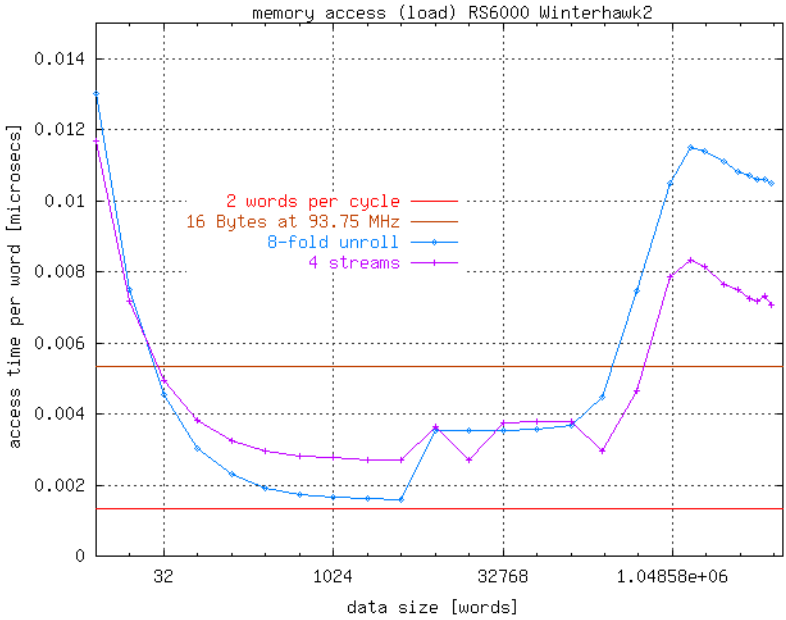
Diese Hardware-Beschreibung wird nun durch die Messung der Leistung beim Laden eines Feldes und bei der Matrix-Vektor-Multiplikation illustriert. Die folgende Fortran-Schleife bewirkt das Laden des Feldes a :

```
do 1 = 1 , n
     as = as + a(i)
end do
```

Je nach der Größe n des Feldes werden bei wiederholter Ausführung der Schleife die Daten aus dem Level1 Cache, dem Level2 Cache oder dem Speicher in die Register geladen. Die erwarteten Ladegeschwindigkeiten pro Datenelement mit 8B sind:

- Level1 Cache: 2 pro Takt : 0,5 /375 *e-6 = 1,33 ns
- Speicher: eff. Bandbreite ist 738 MB/s 8/738 * e-6 = 10,7 ns
- Speicher mit Prefetch: Busbandbreite ist 1,5 GB/s 8/1,5 * e-9 = 5,3 ns

Abb. 4: Zeitmessung für das Laden eines Feldes



Aus der Abb. 4 geht hervor, dass für Feldgrössen unter 64 KB der theoretisch erreichbare Wert von 1,33ns für das Laden eines Feldelementes nahezu erreicht wird. Dazu muss allerdings die Fortran-Schleife 8-fach abgerollt werden, da dann die Pipelines der Recheneinheiten gefüllt werden und ohne Verzögerung ihre Ergebnisse ausliefern können. Bei größeren Feldern treten Level1 Cache-Misses auf, die aus dem Level2 Cache bedient werden müssen und dadurch zu Verzögerungen führen. Übersteigt die Feldgröße auch die Größe des Level2 Caches, werden die Daten direkt vom Speicher geholt und die Lade-Zeit ist durch die Zugriffslatenz und die Bus-Bandbreite bestimmt. Der Effekt des Prefetch-Mechanismus mit vier Strömen ist deutlich zu sehen.

Auch bei der Matrix-Vektor-Multiplikation spielt der Zugriffsgeschwindigkeit auf die Daten die entscheidende Rolle. Die folgende Fortran-Schleife, bei der die äußere Schleife 8-fach abgerollt ist, erzeugt den optimalen Kompromiss zwischen Pipelinefüllung und Nutzung der beschränkten Zahl von Registern:

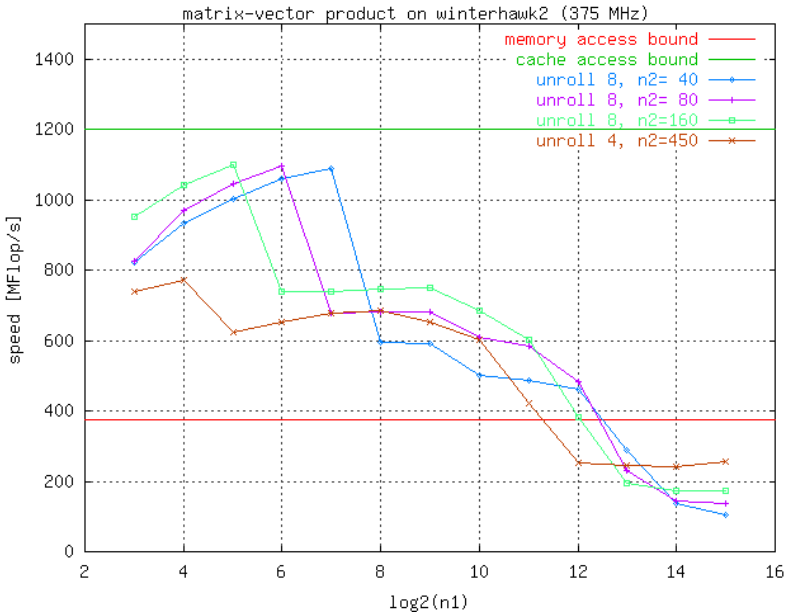
```
do i = 1 , n1 , 8
     s0 = y(i+0)
        ...
     s7 = y(i+7)
     do j = 1 , n2
          s0 = s0 + a(j,i+0)*x(j)
             ...
          s0 = s0 + a(j,i+7)*x(j)
     end do
     y(i+0) = s0
        ...
     y(i+7) = s7
end do
```

Eine Iteration der inneren Schleife enthält 9 Ladeoperationen : a(j,i+0) ,..., a(j,i+7) und x(j) und 8 verkettete Multiplikation-Addition-Operationen. Liegen die Daten im Level1 Cache, so können pro Zyklus 2 Daten geladen werden, insgesamt sind also 5 Zyklen notwendig. Die 8 verketteten Operationen erfordern dank der beiden parallelen Floatingpoint-Einheiten nur 4 Zyklen. Als maximale Rechengeschwindigkeit für diese Schleife ist also ein Wert von 16 Floating-Point-Oprerationen pro 5 Zyklen = 16 / 5*375 = 1,2 Gflop/s zu erwarten. Dabei ist nicht berücksichtigt, dass die Pipelines jeweils nur mit 4 verketteten Operationen gefüttert werden und deshalb ihre start-up-Zeit noch verzögernd wirkt.

Bei großen Matrizen, die aus dem Speicher geladen werden müssen, ist wieder die Bandbreite des Busses der begrenzende Faktor. Pro Multiplikation und Addition ist ein Element der Matrix zu laden und damit eine maximale Rechengeschwindigkeit von 2 * 1500 / 8 *e+6 = 375 Mflop/s zu erwarten.

Aus der Abb. 5 ist zu ersehen, dass die reale Rechengeschwindigkeit des Prozessors bei Datenzugriff aus dem Cache nahezu die Erwartung erfüllt. Bei Datenzugriff aus dem Speicher bleibt die Rechenleistung jedoch selbst unter Nutzung des Prefetch-Mechnismus noch ca. 30% hinter den Erwartungen zurück.

Abb. 5: Rechenleistung bei Matrix-Vektor-Multiplikation



3.2 Multiprozessor

Als nächstes wird die Parallelisierungseffizienz bei Nutzung aller vier Prozessoren eines Winterhawk2-Knotens untersucht. Ein Beispiel für eine "embarrisingly" parallele Anwendung ist die gleichzeitige Ausführung der Matrix-Vektormultiplikation auf allen vier Prozessoren. In Abb. 6 ist zu erkennen, dass sich die Prozessoren nicht gegenseitig behindern, solange sie ihre Daten aus ihren jeweils eigenen Level1- oder Level2-Caches nutzen können. Bei sehr großen Matrizen müssen aber alle Prozessoren über denselben Bus auf den gemeinsamen Speicher zugreifen. Wie aus der Abb. 6 zu entnehmen ist, sinkt dann die für den einzelnen Prozessor verfügbare Bandbreite mit der Anzahl gleichzeitig arbeitender Prozessoren.

Das gleiche Verhalten ist in Abb. 7 zu sehen, die die Rechengeschwindigkeit für eine auf mehrere Prozessoren verteilte Matrix-Vektor-Multiplikation zeigt. Für große Matrizen steigt die Rechenleistung nur unwesentlich mit der Zahl der beteiligten Prozessoren. Bei Matrizen mittlerer Größe, die im Level2-Cache Platz finden, steigt die Rechenleistung in etwa linear mit der Zahl der Prozessoren. Dieses ist das bei der Parallelverarbeitung angestrebte Leistungsverhalten. Bei kleinen Matrizen sinkt die Parallelisierungs-Effizi-

enz wieder ab, da dann die für Kommunikation und Synchronisation benötigte Zeit nicht gegen die Rechenzeit vernachlässigbar ist. Diese Abb. zeigt zugleich die verbesserte Parallelisierung unter der neuen Fortran-Compilerversion 7.1.

Abb. 6: Eine Matrix-Vektor-Multiplikation pro Prozessor

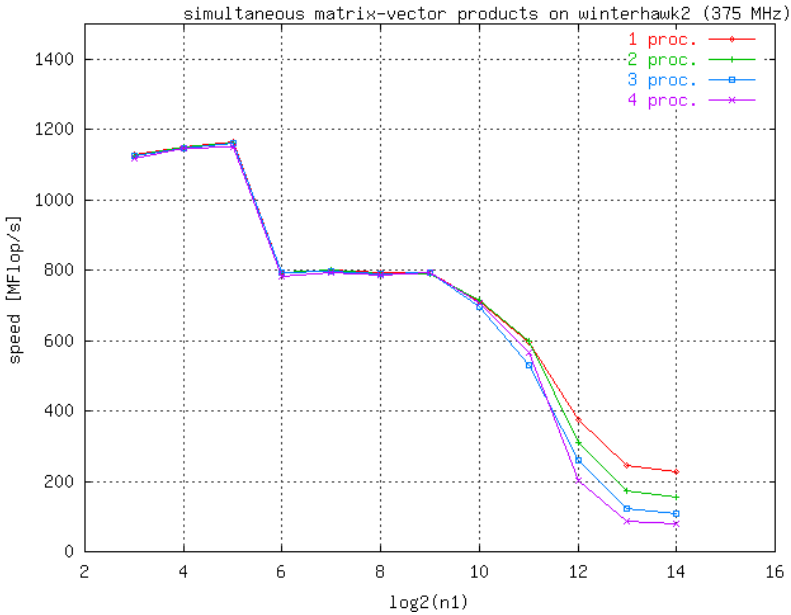
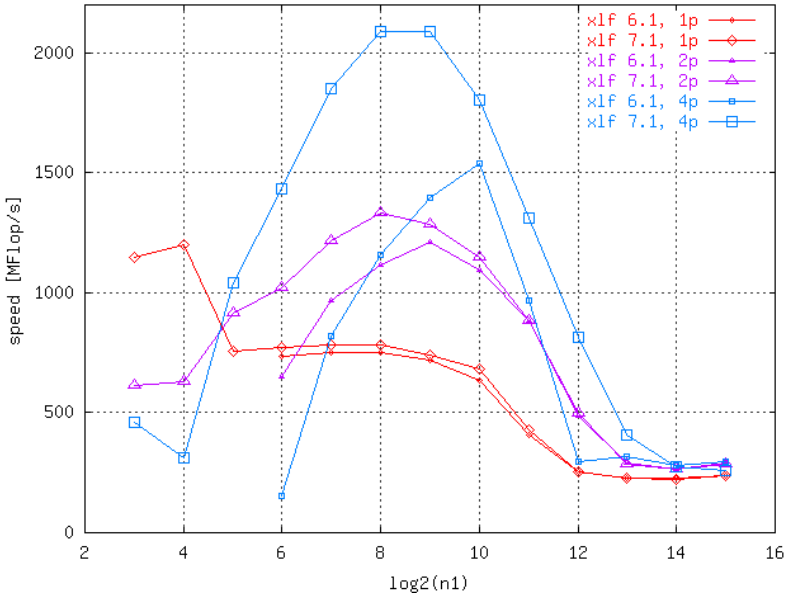


Abb. 7: Eine Matrix-Vektor-Multiplikation auf mehrere Prozessoren verteilt



3.3 Kommunikationsnetz

Die 4-Prozessorknoten der SP können untereinander über zwei verschiedene Netzwerke miteinander kommunizieren: Über die normale Fast-Erthernet Verbindung, die auch den Zugang der einzelnen Knoten zur Außenwelt herstellt, und über den High-Performance-Switch, der eine speziell für die Parallelverarbeitung vorgesehene Kopplung bereitstellt.

Mit einem MPI Programm, das die Zeit für die Übermittlung eines Datenpaketes zwischen zwei MPI-Tasks misst, können die charakteristischen Parameter der Netze bestimmt werden. Dies ist einmal die Latenzzeit, d.h. die Zeit, die vom Start der Übermittlung durch den Aufruf von MPI_SEND auf dem sendenden Prozessor bis zur Ankunft des ersten Datenbits auf dem empfangenden Prozessor vergeht. Praktisch wird diese Zeit aus der Dauer der Übertragung eines Datenpaketes von minimaler Länge ermittelt. Der andere wichtige Parameter ist die Datenrate, d.h. die Geschwindigkeit, mit der die Daten übertragen werden, wenn die Übertragungsverbindung aufgebaut ist. Aus der Abb.8 sind diese Größen für drei verschiedene Situationen angegeben. Im ersten Fall wird die Bandbreite durch die Geschwindigkeit

der Fast-Ethernet-Verbindung, 100 Mbit, bestimmt. Die Latenzzeit ist durch die Verwendung des Standard IP-Protokolls relativ hoch. Die zweite Zeile gibt die Möglichkeiten des Heigh-Performance-Switches bei der Kommunikation zwischen zwei Knoten wieder. Die dritte Zeile beschreibt die Leistung der Kommunikation mittels Speicherkopplung zwischen zwei Prozessoren innerhalb eines Rechnerknotens. Hierzu hat IBM eine MPI-Implementation bereitgestellt, die über den gemeinsamen Speicher kommuniziert. Die Verwendung dieser MPI-Implementierung wird angesteuert durch das Setzen der Umgebungsvariablen MP_SHARED_MEMORY=yes. Zum Vergleich sind in der letzten Zeile noch die auf der Cray T3E der GWDG gemessenen Werte angegeben. Hier zeigt sich die Überlegenheit der im Prozessor integrierten Kommunikationshardware.

Abb. 8: Kommunikationsleistung für MPI-Nachrichtenaustausch

Kommunikationsmodus	Latenz [micros]	Bandbreite [MB/s]
Internode Fast Ethernet, IP-Protokoll	108	12
Internode Switch, US-Protokoll	21	139
Intranode, SMP MPI-Implementierung	9	455
Cray T3E	5	330

4. Das Betriebsmodell

Die GWDG hat seit zwei Jahren ein Betriebsmodell für das wissenschaftliche Rechnen mit hohen Leistungsbedarf eingeführt, bei dem Forschungsgruppen und Institute ihre Mittel zur Beschaffung von Rechenleistung in ein von der GWDG betriebenes zentrales Hochleistungsrechnersystem einbringen können. Eigentums- und Nutzungsrechte an den von den dezentralen Einrichtungen finanzierten Anteilen des Systems verbleiben vollständig bei diesen. Für Beschaffung und Betrieb ist die GWDG zuständig. Diese Zusammenführung von zentralen und dezentralen Mitteln erlaubt die Ausschöpfung von Synergieeffekten in der Beschaffung, dem Betrieb und der Nutzung von Hochleistungsrechnerkapazität.

Bei der Beschaffung kommt die langjährige intensive Erfahrung der GWDG mit den Rechnerherstellern und die genaue Kenntnis ihrer Produkte für die

Ausschreibung, Angebotsprüfung und Auswahl zur Geltung. Damit reduziert sich dieser beträchtliche Aufwand auf ein einziges zentrales System und muss nicht für viele dezentralen Systeme immer wieder von anderen Personen neu geleistet werden. Für die Preisgestaltung spielt natürlich der Gesamtumfang eine entscheidende Rolle: je größer das Volumen desto größer auch der Spielraum und die Bereitschaft des Herstellers zur Gewährung günstiger Konditionen.

Mit der Beschaffung eines Rechnersystems ist es nicht getan, seine Nutzung erfordert eine Peripherie von Speichermedien für Benutzerdaten, von Backupsystemen für die Datensicherung und von Netzverbindungen für einen schnellen Zugang für die Nutzer und die Übertragung von Nutzerdaten. In einem Rechenzentrum wie der GWDG ist diese Peripherie vorhanden und muss für den Betrieb des Hochleistungsrechners eventuell erweitert werden. Weiterhin erfordert die Installation und Pflege der Systemsoftware für das Rechnersystem einen ständigen Arbeitseinsatz von Mitarbeitern mit hoher Qualifikation. Durch die finanzielle Beteiligung eines Institutes am Hochleistungsrechner der GWDG hat dieses Zugang zu eigener Rechenleistung ohne die Aufwendungen für die Einrichtung von Peripherie und das Personal für die Administration tragen zu müssen.

Der gemeinsame Betrieb eines Hochleistungsrechners bietet vor allem Vorteile bei der Auslastung der Rechnerressourcen. Die typische Nutzungsform der Rechner durch eine einzelne Forschungsgruppe ist variabel: nach Zeiten intensiver Beanspruchung der Rechnerressourcen durch intensive Simulationsrechnungen kommen Phasen der Auswertung der Resultate und der Aufbereitung der Ergebnisse für die wissenschaftliche Veröffentlichung. In diesen Zeiten ist der Bedarf an Rechenleistung gering. Ein gruppeneigener Rechner würde also phasenweise wenig genutzt und damit keine optimale Gesamtauslastung erreichen. In dem Betriebsmodell der GWDG mit gemeinsamer Nutzung eines zentralen Hochleistungsrechners durch die Nutzerschaft der GWDG und einzelne mitfinanzierende Forschungsgruppen ist durch Ressourcenübertragung eine sehr hohe Auslastung zu erreichen.

Jede Gruppe hat das volle Eigentumsrecht und das alleinige Nutzungsrecht für den durch sie finanzierten Teil des Rechnersystems. Es besteht aber die Möglichkeit, bei mangelnder Auslastung der eigenen Anteile am Gesamtsystem diese an die anderen Nutzergruppen auszuleihen. Damit erwirbt die ausleihende Gruppe ein Guthaben an Rechenzeit, das sie in Zeiten hohen Bedarfs durch Mitnutzung von Systemanteilen anderer Gruppen wieder einlösen kann. Das zentrale Hochleistungsrechnersystem dient als "Zeitsparkasse", die eine bedarfsgerechte Verteilung von Rechenzeit steuert und damit eine hohe Auslastung garantiert. Die Grundlage für diesen Rechenzeitaus-

gleich ist der GWDG-eigene Anteil des Systems, das die Rechnerressourcen für eine Vielzahl von Forschergruppen bereitstellt, die keine eigenen Mittel für die Erweiterung des Systems beisteuern können. Diese haben insgesamt einen Rechenzeitbedarf, der die im Rahmen des Investitionsvolumen der GWDG realisierbare Kapazität bei weitem überschreitet. Der "Zeitsparkassen"-Effekt wirkt sich für diese Gruppe so aus, dass ihr eine variable Kapazität an Rechenleistung zur Verfügung steht: bei Einspeisung von Zeit mehr, bei Rückforderung weniger. Insgesamt profitiert auch diese Nutzerschaft von dem Betriebsmodell, da durch einen Zeitentwertungsfaktor die externen Gruppen auf einen Teil des länger zurückliegenden Zeitkapitals verzichten müssen. Um in der Finanzsprache zu bleiben: Das Zeitkapital mit wird mit einem negativen Zinssatz belegt.

5. Die Realisierung

In einer ersten Stufe wurde das Betriebsmodell eines gemeinsamen Hochleistungsrechners Anfang 1988 in einer Kooperation der GWDG mit dem Institut für Geophysik der Universität Göttingen realisiert. Beide Partner finanzierten zu gleichen Teilen die Beschaffung eines Parallelrechners der Firma IBM vom Typ RS/6000 SP mit einem Ausbau von 12 Prozessoren Power2 (160MHz) und insgesamt 7 GB Hauptspeicher. Das SP-(Scalable Parallel)-System ist ein Cluster von Rechenknoten, die mittels eines Mehrstufenschalters mit hoher Bandbreite und geringer Latenz vernetzt sind. Der modulare Aufbau macht diese Architektur geeignet für den gemeinsamen Betrieb: jeder Partner nutzt die von ihm bezahlten Rechenknoten. Die Systemsoftware von IBM integriert alle Knoten in einer Verwaltungsumgebung, die den Aufwand für Administration, Zugangsregelung und Ressourcenverteilung gering hält.

Als bei der GWDG die nächste Investition zum Ausbau der Hochleistungsrechenkapazität anstand und zugleich im Institut für Geophysik aus dem Leibniz-Preis von Prof. Christensen weitere Mittel zur Rechnerbeschaffung bereitstanden, bewogen die positiven Erfahrungen mit dem Betriebskonzept die beiden Partner, das gemeinsame Betriebsmodell auch in einer größeren Ausbaustufe zu realisieren. Bei der gemeinsamen Ausschreibung kam wiederum von IBM das günstigste Angebot. So wurde Ende 1999 der Parallelrechner RS/6000 SP um 36 Rechenknoten mit insgesamt 144 Power3 (375 MHz) Prozessoren und 108 GB Speicher erweitert. Die Attraktivität eines gemeinsam betriebenen zentralen Hochleistungsrechner bewog bald darauf auch die Sternwarte der Universität Göttingen und das MPI für Aeronomie in Lindau, Investitionsmittel für Rechenleistungen in das zentrale System einzubringen. Nach einer weiteren Aufstockung der Rechenleistung durch

die GWDG steht für die Göttinger Wissenschaftler jetzt ein Hochleistungsrechnersystem mit insgesamt 236 Prozessoren, 171 GB Hauptspeicher und einer Spitzenleistung von 343 Gflop/s zur Verfügung (s. Abb.9).

Die Rechenknoten der SP sind durch drei Netzwerken untereinander und mit der Außenwelt verbunden: Der High-Performance-Switch dient für die bei der Paralelverarbeitung notwendige schnelle Kommunikation zwischen den Knoten und verbindet die Knoten mit den Servern des parallelen Filesystems GPFS. Ein Verwaltungsnetz schafft die Verbindung zur Kontroll-Workstation, von der aus das System verwaltet wird. Schließlich dient ein Switch-System mit Fast-Ethernet-Verbindung zu den Knoten und Gigbit Uplink-Leitungen zur Anbindung an das GWDG-Filesystem und die Außenwelt (s. Abb. 10).

Abb. 9: Aufteilung des SP-Systems nach Nutzergruppen

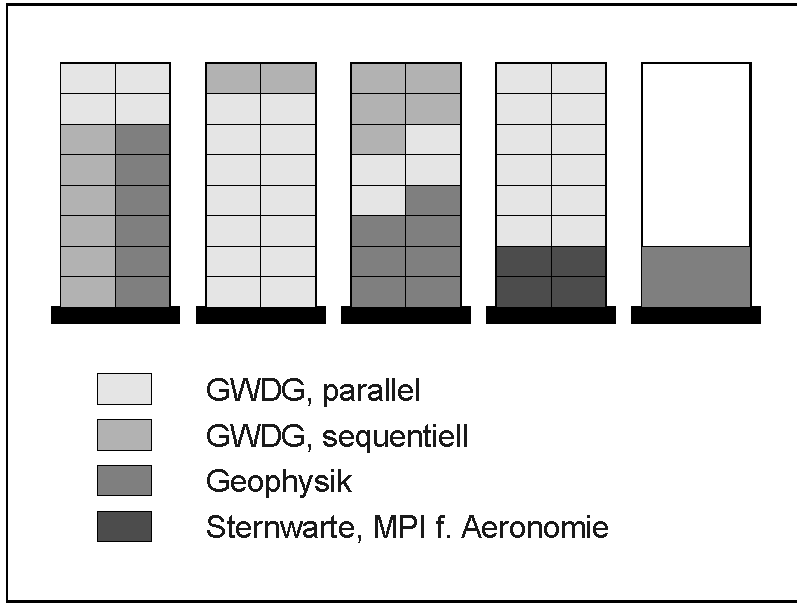
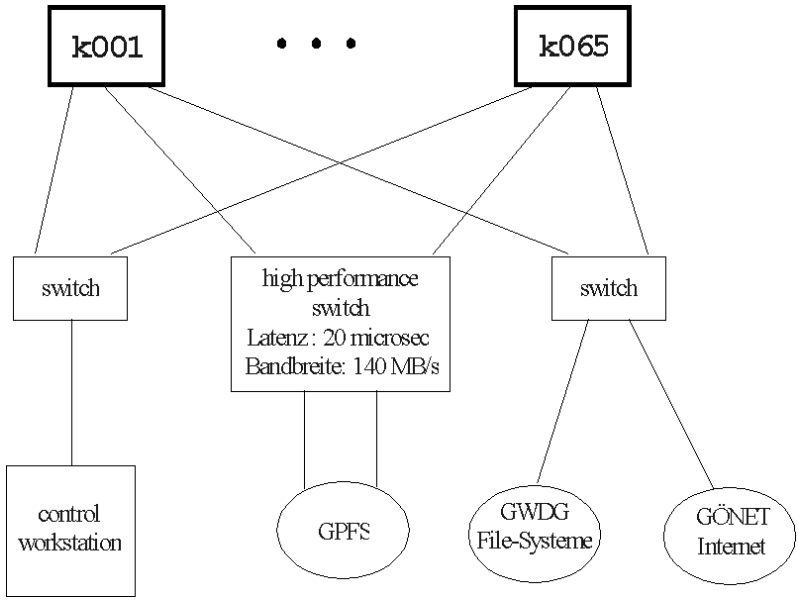


Abb. 10:Vernetzung des SP-Systems der GWDG



Für die Administration stellt sich die SP als ein relativ einheitliches System dar, dessen Betreuung mit nicht zu hohem Aufwand möglich ist. Von der Benutzerseite aus jedoch dient das System unterschiedlichen Gruppen für unterschiedliche Anwendungsarten. Die Minimierung von Personalkosten für den Rechnerbetrieb durch die Integration von verschieden Eigentümern, Benutzergruppen und Nutzungsarten in einem einzigen zentralen System ist wohl der bemerkenswerteste Aspekt des von der GWDG entwickelten Betriebsmodells.

6. Nutzungserfahrungen

Das SP-System wurde in seiner ersten Ausbaustufe im März 2000 für den Benutzerbetrieb freigegeben. Das System lief nach der Aufbauphase, die mit einem Austausch aller Prozessoren Ende Mai beendet war, sehr stabil, solange keine Eingriffe von außen erfolgten. Diese wurden erzwungen bei der Erweiterung des System, bei Einführung neuer Versionen von Systemsoftware und bei Notfall-Stromabschaltungen.

Die SP kann von allen Nutzern der GWDG, also den Wissenschaftlern der Universität Göttingen und der Institute der MPG, genutzt werden. Diese Nut-

zung ist natürlich auf den von der GWDG beschafften Teil des Systems beschränkt.

Die Nutzung wird über Batch gesteuert, wobei für sequentielle Jobs das auch auf den anderen Batchservern eingesetzte CODINE-System verwendet wird. Die parallelen Batchjobs werden über das IBM-proprietäre LoadLeveler-System verwaltet, das mit dem Backfill-Scheduler über die Möglichkeit zur Optimierung des Durchsatzes auch für parallele Anwendungen sorgt. In jüngster Zeit (September 2001) wurde zusätzlich der externe Scheduler Maui eingesetzt, der eine Priorisierung von Jobs gemäß der in der Vergangenheit verbrauchten Rechenzeit erlaubt. Dies vereinfacht eine gerechte Verteilung der Parallelrechnerressourcen.

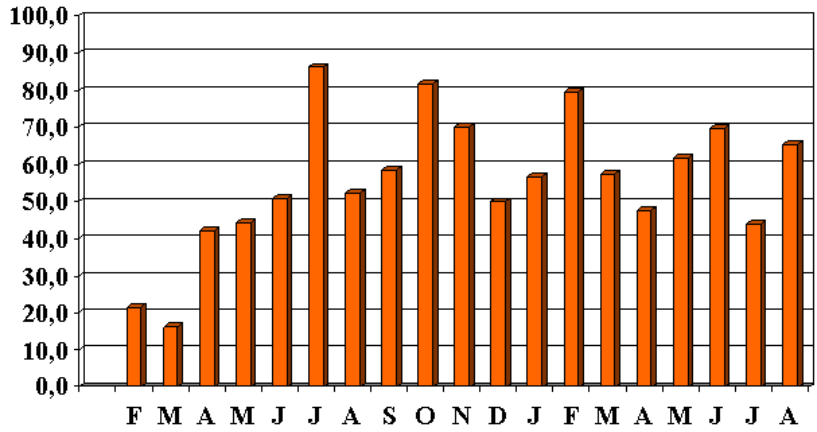
Zur Zeit sind im GWDG-Teil des Systems 5 Knoten für die sequentielle und 32 Knoten für die parallele Verarbeitung reserviert. Die parallele Nutzung erfolgt sowohl im Shared-Memory Programmiermodel, wobei maximal die vier Prozessoren eines Knotens genutzt werden können wie im reinen Message-Passing Programmiermodell, bei dem auf mehreren Knoten alle Prozessoren über Nachrichtenaustausch miteinander kommunizieren. Höchste Parallelisierungseffizienz lässt sich mit einem hybriden Programmiermodel erreichen, bei dem auf jedem Knoten ein MPI-Task läuft, der die Kommunikation zwischen den Knoten durchführt. Die vier Prozessoren innerhalb des Knotens werden im MPI-Task über OpenMP Direktiven zur Parallelverarbeitung im Shared-Memory-Modus integriert. Auf diese Weise wird die Granularität der Internode-Kommunikation erhöht und damit der Nachteil der relativ hohen Latenz vermindert.

Die Prozessorzahl der parallelen Anwendungen geht von 4 Prozessoren auf einem Knoten bis zu 64 Prozessoren auf 16 Knoten, mit einem Häufigkeitsmaximum bei 16 - 32 Prozessoren.

Die Zahl der Nutzer liegt beim sequentiellen Batch bei ca. 30, beim parallelen Batch bei ca. 10.

Die Nutzungsstatistik in Abb. 11 spiegelt mit ihren großen Schwankungen zwei verschiedene Einflussgrößen wieder. Zum einen die Betriebsunterbrechungen aufgrund von Hard- und Softwarearbeiten. Die durch die Erweiterung bedingten Ausfälle des Systems reduzierten die Auslastung im Dezember 2000 und Januar 2001, die Behebung von Schäden, die durch eine Notstromabschaltung verursacht wurden führten zu in der Juliauslastung sichtbaren Ausfallzeiten. Zum anderen hängt die Nutzung des Systems stark von der Aktivität der ein Projekt bearbeitenden Wissenschaftler ab. In vorlesungsfreien Zeiten mit hoher Reisetätigkeit, wie April oder Juli, sinkt dadurch die Auslastung ab.

Abb. 11:Nutzungsstatistik Feb. 2000 - Juli 2001



7. Ausblick

Die hohe Akzeptanz des SP-Systems durch Nutzer, Institute und Administratoren machen eine Fortentwicklung des von der GWDG entwickelten Betreibermodells auf der Basis hochintegrierterer SMP-Cluster interessant. Die GWDG will deshalb in einer weiteren Investitionsstufe 2002 das SP-Parallelrechnersystem mit der dann verfügbaren neuesten Technologie erweitern. Bei der Finanzierung werden sich wieder externe Institute beteiligen.

Dabei wird in Zukunft genau zu untersuchen sein, inwieweit Parallelrechner auf Basis der immer leistungsfähiger werdenden PC-Hardware und offener Software die Anforderungen der Nutzer für weniger Geld erfüllen können, ohne den Betreibern zu großen Arbeitsaufwand aufzubürden, der den Kostenvorteil bei der Beschaffung wieder zunichte machen würde.

Teil 2: Beiträge vom 18. DV-Treffen

Der neue IBM-Hochleistungsrechner der MPG

Hermann Lederer

RechenZentrum Garching der Max-Planck-Gesellschaft
Max-Planck-Institut für Plasmaphysik, Garching bei München

1. Zur Beschaffung

Im Feb. 2000 erfolgte die Vorankündigung des Beschaffungsvorhabens im Supplement zum Amtsblatt der EU, und damit begann die Phase der Markterkundung unter Einbeziehung jener Firmen, die daraufhin ihr Interesse anmeldeten. Im Sept. 2000 wurde von 12 MPIen ein Beschaffungsantrag beim Beratenden Ausschuss für Rechenanlagen (BAR) der MPG mit der Bitte um Begutachtung gestellt. Der Antrag wurde befürwortet. Daraufhin erfolgte im Okt. 2000 die Vergabebekanntmachung im Supplement zum Amtsblatt der EU durch die Generalverwaltung der MPG. Für ausgewiesene Firmen, die die Teilnahme am Teilnahmewettbewerb beantragten, wurden die Verdingungsunterlagen am 1. Dez. 2000 versandt. Am 29. Jan. 2001war Angebotsabgabe, am 31. Jan. 2001 Angebotseröffnung. Es gab vier gültige Angebote und zwei Absagen wegen im vorgegebenen Zeitrahmen nicht passender Produktpalette. Am 2. Feb. 2001 wurde die Unterkommission (UK) Hochleistungsrechner des Wissenschaftlichen Beirats über das Ergebnis der Ausschreibung informiert. Am 22. Feb. 2001 lag das Votum der UK vor. Am 9.3. 2001 wurde im Beratenden Ausschuss für Rechenanlagen (BAR)

der MPG die Entscheidung zugunsten des IBM-Angebots in Übereinstimmung mit dem Votum der UK getroffen. Am 12. März 2001 erfolgten die schriftlichen Absagen an jene Firmen, die nicht zum Zuge kamen. Eine nach neuem EU-Recht gültige Einspruchsfrist von 14 Tagen verstrich ohne Einsprüche. Am 27. April 2001 erfolgte der Vertragsabschluss zwischen der Fa. IBM und der MPG.

2. Installationsphasen

Es wurde gemäß den Anforderungen im Leistungsverzeichnis eine Erstinstallation für Dez. 2001, die Hauptinstallation für Nov. 2002 vereinbart.

Für die Erstinstallation sind mindestens zwei Regatta H Server (IBM eServer p690) mit je 32-way Power4 Prozessoren mit einer Taktrate von 1,3 GHz vorgesehen, mit einem Gesamthauptspeicher von mindestens 256 GB.

Für die Hauptinstallation im Nov. 2002 wurde ein System mit 3,8 TFlop/s Peakleistung vorgesehen, basierend auf 32-way Regatta-Knoten mit Power4-Prozessoren, einem SP-Switch für Knoten-Interconnect, einem Gesamthauptspeicher von 1,8 TB, sowie 20 TB Disks.

Anmerkung: Für das Jahr 2002 ist der Weiterbetrieb des Cray-T3E-Systems neben dem IBM-Erstsystem vorgesehen. Nach Inbetriebnahme des IBM Hauptsystems soll das T3E-System abgeschaltet werden.

3. Technische Beschreibung

Wesentlicher Baustein des neuen Systems ist ein Mehrprozessor-Rechenknoten mit großem, gemeinsamem Hauptspeicher. Eingesetzt werden sog. Power4-Prozessoren, von denen sich zwei auf einem Chip befinden. Der Chip basiert auf 0.18 µ Technologie, hat 170 Mio. Transistoren, zwei 64-bit-CPUs mit einer Taktrate von mind. 1,1 GHz mit 5-fach superskalaren Cores, dreifach Level-Cache-Hierarchie, 10 GB/s Main-Memory-Interface und 35 GB/s Multiprozessor -Interface.

Abb. 1: Power4 Multi-Chip Module (MCM)
Vier Chips pro MCM ergeben ein Acht-Prozessor-SMP-System
(aus: IBM pSeries 690, IBM, Okt. 2001)

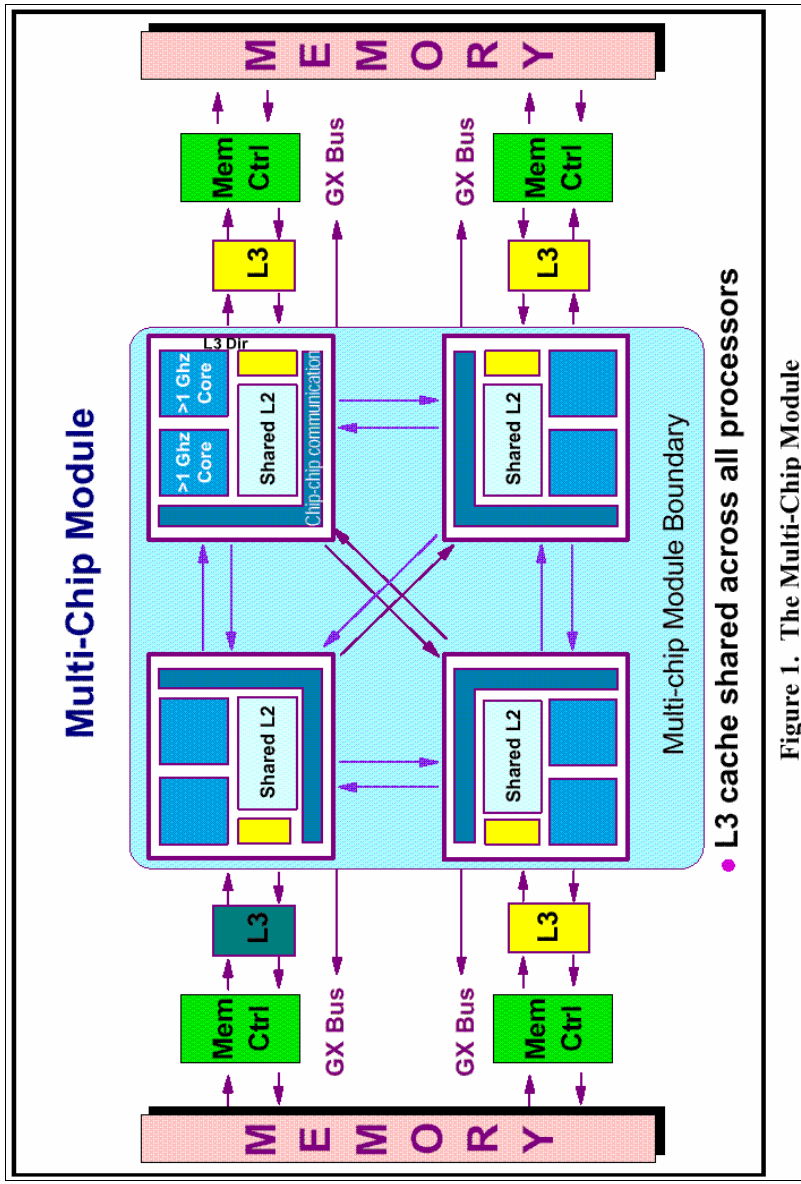


Abb. 2: Power4 32-Prozessor-SMP-System
durch Interconnect von vier MCMs in Ringtopologie
(aus: IBM pSeries 690, IBM, Okt. 2001)

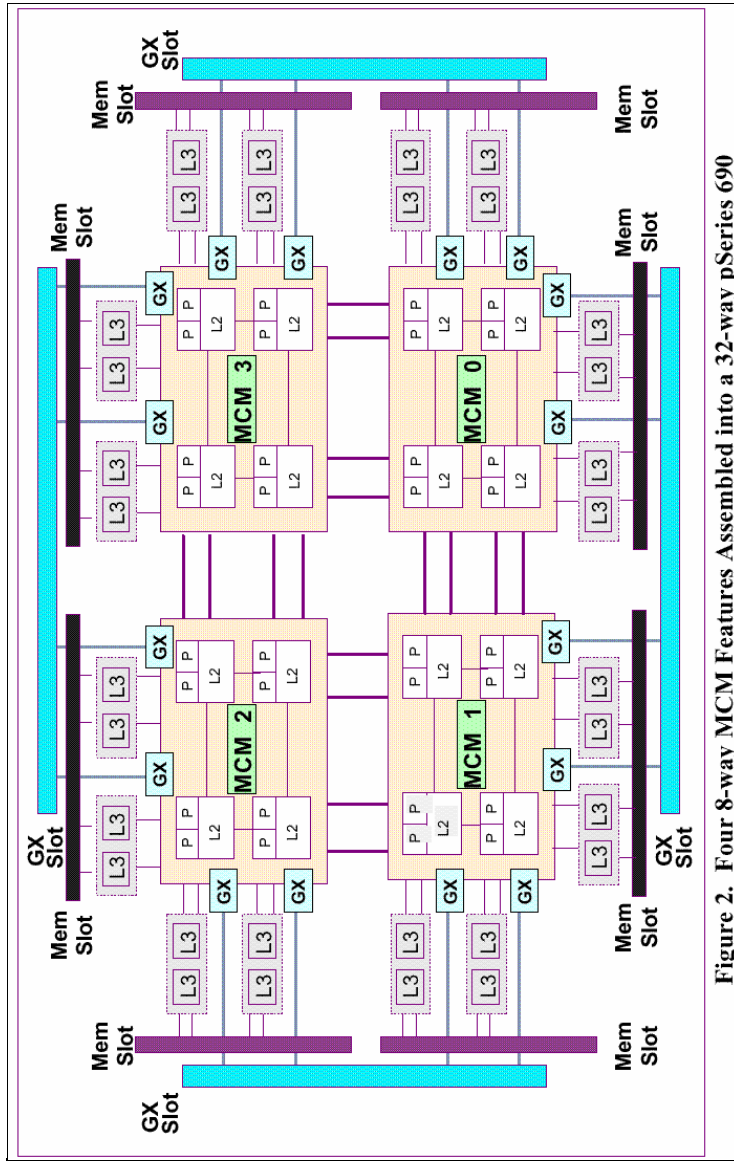


Figure 2. Four 8-way MCM Features Assembled into a 32-way pSeries 690

Regatta-Knoten basieren auf der in Abb. 2 gezeigten Konfiguration mit zusätzlichen I/O-Einschüben für externe Adapter.

Zum Erstsystem:
An die Knoten des Erstsystems werden 5 TB Benutzerplatten auf der Basis von SSA-Disks angeschlossen werden. Es kommt das knotenübergreifende, parallele GPFS-Filesystem zum Einsatz. Ein Knoten wird teilweise oder vollständig für interaktive Entwicklung eingesetzt werden, ein weiterer Knoten im Batchbetrieb mittels Loadleveler. Als Unix-Betriebssystem wird AIX 5.1 eingesetzt. Es stehen Fortran 90 und C/C++ Compiler zur Verfügung sowie die numerischen Bibliotheken ESSL und PESSL.

Zum Hauptsystem:
Eine Grundkonfiguration für die zugesagte Leistung liegt vor. Die Optimierungen von einzelnen Komponenten, insbesondere zur I/O-Konfiguration, wird im Laufe des Jahres 2002 erfolgen. Für das Hauptsystem wird eine Batch-Queue-Struktur analog zur Struktur auf dem T3E-System geplant:

- 512 Proz-Queue bis 1 TB Hauptspeicher
- 256 Proz-Queue bis 0,5 TB Hauptspeicher
- 128 Proz-Queue bis 0,25 TB Hauptspeicher
- 64 Proz-Queue bis 128 GB Hauptspeicher
- 32 Proz-Queue bis 64 GB Hauptspeicher

Zum Vergleich: derzeit sind auf T3E/512 max. 60 GB RAM verfügbar.

4. Leistungsfähigkeit

Die Leistungsfähigkeit des Endsystems soll mindestens die 10-fache Cray-T3E-600/512-Leistung für die MPG-Benchmark-Suite 2000 betragen, im Mittel über alle Codes.

Für die Benchmark-Suite waren 7 wichtige Anwendungen aus den Materialwissenschaften, der Fusionsforschung und der Astrophysik zusammengestellt worden. Sie enthält sechs hoch-parallele T3E-Anwendungen sowie eine moderat-parallele OpenMP-Shared-Memory-Anwendung. Die Suite enthält sechs Fortran90-Codes, einen C-Code. Dies spiegelt das derzeitige Anwendungsspektrum auf der T3E wider und zeigt die noch vorhandene Dominanz von Fortran90-Produktionscodes. Bei Neuentwicklungen wird inzwischen auch C/C++ stärker eingesetzt.

Die MPG-Benchmark-Suite-2000 besteht aus folgenden Anwendungen:

Code 1: GENE (Gyrokinetic Electromagnetic Numerical Experiment)
Fusionsforschung
Prof. Lackner, MPI für Plasmaphysik, Garching

Code 2: PROMETHEUS (Combustion processes in novae)
Astrophysik
Prof. Hillebrandt, MPI für Astrophysik, Garching

Code 3: CPMD (Ab-initio molecular dynamics)
Materialwissenschaften
Prof. Parrinello, ehemals MPI für Festkörperforschung, Stuttgart

Code 4: LAPW1 (Full-potential linearized augmented plane-wave)
Materialwissenschaften
Prof. Scheffler, Fritz-Haber-Institut, Berlin

Code 5: TORB (Global gyrokinetic non-linear calculations of ion temperature gradient modes (ITG)
Fusionsforschung
Prof. Nuehrenberg, MPI für Plasmaphysik, Greifswald

Code 6: POLY (Polymer simulations for polymer chains):
Materialwissenschaften
Prof. Kremer, MPI für Polymerforschung, Mainz

Code 7: YASP (Atomistic md calculations for polymer simulations)
Materialwissenschaften
Prof. Kremer, MPI für Polymerforschung, Mainz

5. Konsequenzen für Anwendungen

Zu numerischen Bibliotheken:

Der Cray/SGI-proprietären Scilib mit für die Architektur optimierten mathematischen Routinen stehen die numerischen IBM-Bibliotheken ESSL und pESSL gegenüber. Bei manchen Routinen ist die Umstellung einfach, bei anderen wie Lapack- und Scalapack-Routinen eventuell kompliziert, da erstens eine unterschiedliche Aufrufsyntax vorliegt und zweitens nur ein Subset der Routinen in ESSL/PESSL implementiert ist. Hierzu werden Änderungswünsche an IBM herangetragen werden.

Zur Kommunikation bei Parallelverarbeitung:

Die Cray-T3E-Netzwerkleistung muss nach wie als hervorragend bzgl. Latenzzeit und Bandbreite pro Prozessor bezeichnet werden. Zugriffszeiten auf lokales Memory und auf *remote* Memory sind nahezu gleichwertig. Der

allgemeine Trend bei neuen Supercomputer-Architekturen, hierarchische Systeme durch Cluster von SMPs zu bilden, trifft auch für das IBM-System zu. Damit stehen guten Kommunikationsverhältnissen im SMP schlechtere zwischen den SMPs gegenüber. Damit wird insbesondere die knotenübergreifende Kommunikation bei MPP-Codes wesentlich „teurer“ relativ zum Rechenanteil.

Deshalb wird für dieses T3E-Nachfolgesystem eine Änderung der Programmierkonzepte nötig, schon im Hinblick auf Portabilität. Max-Planck-Codes sind weltweit auf verschiedenen Hochleistungsrechnerplattformen im Einsatz.

Konzeptionelle Änderungen wären gekennzeichnet durch:

- Verzicht auf Einfachheit und Effizienz von einseitiger Kommunikation mit SHMEM
- Vermeidung häufiger Kommunikation zugunsten von „Bündelung“ von Kommunikation und „Rechnen statt Kommunizieren“, wo alternativ möglich
- Einführung „gemischter“ Programmierung:
  Shared-Memory-Programmierung innerhalb des SMPs (mit OpenMP oder *autotasking* mittels Compileroption), nur noch ein oder wenige MPI-*threads* pro SMP

Bei Verzicht auf einseitige Kommunikation mit der SHMEM-Bibliothek von Cray/SGI entsteht ein Umstrukturierungsaufwand, da aus Effizienzgründen meist keine *straight-forward*-Abbildung auf einseitige Kommunikation mittels MPI-2 erfolgen kann. Die „Bündelung“ von Kommunikation und „Rechnen statt Kommunizieren“ kann von Fall zu Fall einen sehr hohen Umstrukturierungsaufwand bedeuten. Die „gemischte“ Programmierung mit OpenMP und MPI wird nötig werden, nicht nur, um generelle Skalierungsprobleme von „flat“ MPI“ zu mildern, sondern insbesondere die Auswirkungen der schwächeren Kommunikationsleistung zwischen SMPs zu relativieren. Dies bedeutet oft Handarbeit, da *Autotasking* im SMP durch Compileroptionen oft in komplexeren Programmen nicht gut funktioniert. Dies impliziert eine Einfügung von OMP-Direktiven von Hand an geeigneter Stelle.

6. Erste Erfahrungen

Die MPG wurde von IBM als einzige europäische Forschungseinrichtung für das *Early Shipment Program* (ESP) für Regatta ausgewählt. Ziel des Programms sind Tests von Konfiguration und Software im Vorfeld einer regulären Installation.

Im Rahmen dieses Programms wurde am 24.10.2001 ein Acht-Prozessor-Power4-System mit 64 GB Hauptspeicher und einer Prozessortaktrate von 1 GHz installiert.

Abb. 3: Regatta-System (oben), zusammen mit 3 TB SSA Disksystem (unten) (1 Knoten mit 8 Proz. Power 4 (1 MCM), 1 GHz, 64 GB RAM)

Mit diesem Vorab-System wurden bereits verschiedene Tests zur Systeminstallationsprozedur, zur Ausfallsicherheit von Hardwarekomponenten und zum Einsatz eines breiten Anwendungsspektrums durchgeführt. Als Betriebssystem wird AIX 5.1 eingesetzt. Es stehen die Fortran90-Compiler XLF 7.1.1 und der VisualAge C++ Compiler XLC 5.0.2 zur Verfügung, für Parallelverarbeitung das Parallel Environment 3.2. Als Batchsystem wird Loadleveler 3.1 eingesetzt. Die numerischen IBM-Bibliotheken ESSL und PESSL sind ebenfalls bereits verfügbar. Für Diskplatz wurden 3 TB SSA-Platten aus dem Vertragskontingent eingesetzt.

Die AIX-Systeminstallation von CD-ROM funktionierte reibungslos, wie gewohnt. Die Installation von einem NIM-Server funktioniert bislang nur über ein Fast Ethernet Interface, nicht über das Gigabit Ethernet Interface.

Der Bootvorgang war problemlos mit dem 32-Bit-Kernel wie mit dem 64-Bit-Kernel möglich. Die Bootdauer lag bei jeweils ca. 20 Min. Von den prinzipiell zur Verfügung stehenden Filesystemen jfs (Journaled File System, lokal, bis 1 TB), jfs2 (Extended Journaled File System, lokal, prinzipiell bis 4 PB) und gpfs (General Parallel File System, global) wurde auf dem vorhandenen Disksystem ein 1,5 TB großes jfs2-Filesystem eingerichtet, das sich in der Testphase befindet. Ziel ist jedoch die Verwendung des knotenübergreifenden GPFS-Filesystems. Test hierzu werden stattfinden, sobald ein zweiter Knoten installiert sein wird.

Das Batchsystem Loadleveler und die Software Work Load Manager als Resource Manager wurden installiert, Tests zur Erprobung der Funktionalität haben begonnen.

Anwendungen wurden bereits mit dem 32-Bit-Kernel wie auch dem 64-Bit-Kernel getestet. Mit dem 32-Bit-Kernel sind prinzipiell bis zu 96 GB insgesamt im Shared Memory adressierbar, der 64-Bit-Kernel ist nötig für die Überwindung der 32-Bit-Grenzen. Vorgesehen sind 128 GB RAM für die beiden Regatta-Systeme der Erstinstallation.

Unter dem 32-Bit-Kernel von AIX 5.1 liefen Power3 Executables (erzeugt unter AIX 4.3) problemlos auf Power4, wie von IBM zugesagt. Benchmark-Codes waren, soweit bisher getestet, nach Neukompilation lauffähig. Dazu zählen GENE (IPP), TORB (IPP), YASP(MPIP), PROMETHEUS (MPA), LAPW (FHI). Weitere lauffähige, parallele Codes sind: TRIDYN (IPP), DYN5D (FHI), NRLMOL (MFK, MF), ECHAM5 (Meteorologie).

Unter dem 64-Bit-Kernel von AIX 5 konnten verschiedene Anwendungen bislang ebenfalls problemlos kompiliert werden. Beim Ablauf konnten aber bereits Problemzonen identifiziert werden, die bereits im Labor analysiert,

und die darauf hinweisen, dass derzeit für parallele Anwendungen nach Möglichkeit der 32-Bit-Kernel verwendet werden sollte.

Bzgl. des Problems bisher nicht-möglicher, performanter einseitiger Kommunikation auf IBM-Systemen wurde vom *Advanced Computing Technology Center* (ACTC) am *IBM Watson Research Center* unter Einsatz des *Low-level Application Programming Interface* (LAPI) eine sog. Turboshmem-Bibliothek zur Verfügung gestellt, die es erlaubt, Cray/SGI shmem-Calls direkt abzubilden. Funktionalität und Performance dieser Bibliothek werden in Zusammenarbeit mit ACTC getestet.

7. Dokumentation

Eine Dokumentationsseite zum neuen Hochleistungsrechner, insbesondere zur Software- und Programmierumgebung, ist in der RZG-Hompepage im Aufbau begriffen. Sie ist zu finden unter: http://www.rzg.mpg.de/computing/IBM_P

8. Erfahrungsaustausch

Mit dem Potsdam Institut für Klimafolgenforschung besteht bereits ein reger Informationsaustausch über große IBM-Power3-Systeme und ihre Softwareumgebung. Auf dem PIK-System konnten Programmtests und Code-Umstellungsarbeiten durchgeführt werden. Das PIK erhält ein Power4-System im Jahre 2002.

Das Oak Ridge National Lab (ORNL) hat dem RZG eine Kooperation sowie Erfahrungsaustausch auf System- und Anwendungsebene angeboten. ORNL erhält ein 4-Tflop/s-IBM -Zwillingssystem zur MPG und hat bereits auch ein ESP-System installiert.

Das Scientific Computer Center Finland (CSC) hat einen exzellenten IBM SP User Guide für Power 3 SP erstellt und dem RZG die Source für Anpassungen an das MPG-System überlassen. Der Austausch des Power3- zu einem Power4-System erfolgt bei CSC im 1. HJ 2002. Mit CSC wurde ebenfalls ein Erfahrungsaustausch vereinbart.

Das Göttinger Funk-LAN „GoeMobile“

Andreas Ißleiber

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Vorwort

Drahtlose Netze (WLANs) gibt es schon seit einiger Zeit. Eine größere Verbreitung erfahren sie aber erst seit 1999, nachdem die IEEE den Standard für WLANs (802.11b) verabschiedete, der den Aufbau von WLANs bei Übertragungsraten von bis zu 11 MBit/s erlaubt.

Seit Dezember 2000 betreibt die GWDG, insbesondere im Nahbereich von Gebäuden der Universität Göttingen, ein flächendeckendes Funk-LAN nach diesem Standard. Das Funknetz (GoeMobile) wurde zu gleichen Teilen aus

Forschungsmitteln des Wissenschaftsministeriums Niedersachsen und des Bundesforschungsministeriums finanziert und durch die GWDG geplant und aufgebaut.

Mit der wachsenden Verbreitung sinken natürlich auch die Preise, so dass ein Einsatz der Funk-LAN-Technologie keinen großen finanziellen Aufwand mehr darstellt.

1. Motive für den Einsatz von Funk-LAN-Technologien

Funk-LAN-Technologien sollen in erster Linie dem mobilen Benutzer den permanenten Zugang zum Netz ermöglichen.

Dabei entstehen folgende Motive für ein WLAN:

- Das Internet soll flächendeckend verfügbar sein.
- Laptop ist als mobiler Arbeitsplatz nutzbar.Lokale Netzdienste (z. B. GWDG) sind (mobil) erreichbar.
- Online-Recherchen in Bibliotheken wie z. B. SUB sind ohne Festnetzzugang möglich.
- Netzanbindung der via Kabel unzugänglichen Gebäude realisierbar.
- Funk-LAN kann als Medium für weitere Dienste wie Telefonie (VoIP) dienen.

2. Zahlen, Fakten und Standards zum Funk-LAN

Grundsätzlich erlaubt der Standard IEEE 802.11b zwei verschiedene Arten für den Aufbau eines WLANs:

- *Ad-Hoc-Netzwerk:* Entsteht einfach durch die Anwesenheit mehrerer Endgeräte mit WLAN-Adapter in dem Reichweitenbereich. Diese können dann ein Netzwerk ohne feste Basisstation bilden.
- *Funk-LAN mit Basisstationen (AP):* Eine zentrale Basisstation (AP, Access-Point) dient quasi als Verteiler für alle WLAN-Endgeräte innerhalb ihres Reichweitenbereiches. Mehrere Basisstationen können über Funk oder Ethernet zusammengeschlossen werden und arbeiten dann in einem Verteilsystem. Da die Hauptanwendung für WLANs im Bereich mobiler Endgeräte liegt, sind die meisten Funknetzwerkkarten als PC-Cards für Notebooks zu bekommen. Es gibt aber auch zunehmend Steckkarten für PCs.

Der Standard IEEE 802.11b entstand zunächst aus dem älteren IEEE 802.11, welcher mit 2 MBit/s Übertragungskapazität noch einige Wünsche bezüglich Geschwindigkeit offen ließ. IEEE 802.11b ermöglicht reichweitenabhänge Übertragungsraten von 1, 2 ,5 oder 11 MBit/s.

Als Frequenzbereich wird das 13-cm-Band verwendet, welches dem Bereich um 2,4 GHz entspricht. Es ist im übrigen exakt die Frequenz, auf der auch der heimische Mikrowellenofen arbeitet.

Die Sendeleistung darf bei diesem Standard eine Leistung von 100 mW an der Antenne nicht überschreiten. Das ist, im Vergleich zu anderen Funksystemen, eine geringe Sendeleistung, vergleicht man diese mit einem handelsüblichen Handy, welches mit 2 W immerhin das 20-Fache an Leistung abgibt.

Die tatsächlich an der Antenne abgegebene Sendeleistung ist ganz wesentlich von der Art und Qualität der Antenne abhängig. In der Regel senden die WLAN-Karten der Hersteller mit Leistungen von 7 - 50 mW. Erst die Antenne bewirkt mit ihrem Antennengewinn einer Art Verstärkung und bringt damit die Leistung auf 100 mW. Dennoch ist der Betrieb nicht jeder Kombination aus WLAN-Karte und Antenne erlaubt, da bei „guten“ Antennen (über 10 dbi Gewinn) die abgegebene Leistung rasch über die erlaubte 100-mW-Grenze kommt und damit den Vorgaben der RegTP nicht genügt.

2.1 Das Modulationsverfahren

Im Vergleich zu älteren WLAN-Standards wird bei 802.11b das DSSS (Direct Sequence Spread Spectrum) als Modulationsverfahren benutzt. Es ist relativ robust und unempfindlich gegenüber „schmalbandigen“ Störungen.

2.2 Die Reichweiten

Die Reichweite ist stark abhängig von der Umgebung und der verwendeten Antenne.

Typische Werte [m] sind:

1000 - 10000	bei freier Sicht und bei guten, externen Antennen auf Sender- und Empfängerseite
200 - 300	bei freier Sicht (freies Gebäude) ohne externe Antenne
15 - 30	innerhalb von Gebäuden

Bei diesen Zahlen wird klar, dass die Reichweite innerhalb von Gebäuden deutlich geringer ausfällt. Das liegt wesentlich an der Gebäudebeschaffen-

heit und der Tatsache, dass dort nicht immer eine direkte Sichtverbindung zwischen den Funkteilnehmern existiert. Weitere reichweitenreduzierende Faktoren sind:

- Antennen, Sendeleistung an der Karte
- Antennenkabel/Länge, Stecker etc.
- Interferenzen mit anderen Funksystemen (Funk-LAN/BlueTooth)

2.3 Übertragungsraten

Der 802.11b-Standard sieht eine max. Übertragungsrate von 11 MBit/s (brutto) vor. In der Praxis wird eine effektive Übertragungsleistung von etwa 5 - 6 MBit/s erreicht.

Dieser Wert ist ausreichend für die klassischen Aufgaben eines mobilen Benutzers. Jedoch muss berücksichtigt werden, dass ein AP (Access Point) häufig von mehreren Anwendern genutzt wird und sich dadurch die zur Verfügung stehende Bandbreite auf die Anzahl der Benutzer aufteilt.

Modernere Funksysteme versprechen größere Bandbreiten, wobei zwei Funkverbindungen parallel betrieben werden und damit eine Gesamtübertragungsrate von bis zu 22 MBit/s erreicht werden kann. Dieses wird aber meistens für eine Punkt-zu-Punkt-Verbindung genutzt, da für die Clients keine 22-MBit/s-Systeme (Funkkarten) zur Verfügung stehen.

Ein zukünftiger Standard (IEEE 802.11h) soll mit einer Bandbreite von bis zu 54 MBit/s eine deutliche Geschwindigkeitssteigerung bringen. Diese Systeme werden im Frequenzbereich von 5 GHz betrieben und stellen damit eine deutliche Verbesserung der Übertragungsleistung dar. Allerdings können bereits für das 2,4-GHz-Band installierte Antennen nicht mehr benutzt werden. Darüber hinaus ist aufgrund der höheren Frequenz eine große Reichweite nicht zu erwarten, wenn die Sendeleistung ebenfalls bei 100 mW beschränkt bleibt. Eine Sichtverbindung zwischen Sender und Empfänger ist dabei, wie auch schon bei 2,4 GHz, unerlässlich.

Die folgende Tabelle gibt einen Überblick, über die erreichbaren Entfernungen. Vorausgesetzt werden gleiche Antennen auf Sende- wie Empfangsseite.

Tab. 1: Reichweiten in Abhängigkeit der verwendeten Antennen

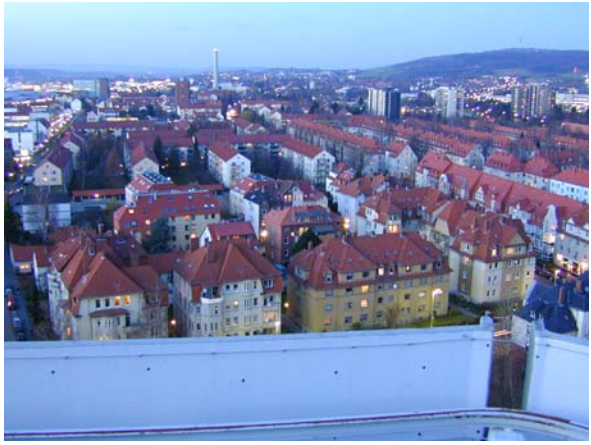
Datenrate [MBit/s]	14 dbi	12 dbi	10 dbi	7 dbi
1	6,4 km	6,3 km	5,2 km	3,7 km
2	4,8 km	4,7 km	3,7 km	2,6 km
5,5	3,4 km	3,3 km	2,6 km	1,9 km
11	2,4 km	2,3 km	1,9 km	1,2 km

3. Die Standorte im Göttinger Funk-LAN „GoeMobile“

Primäres Ziel des GoeMobile ist der mobile Benutzer. Deshalb wurde zunächst der stark frequentierte Bereich des Göttinger Campus mit Funk versorgt. Gerade dort ist „mobiles“ Internet gewünscht. Zusätzlich muss durch Verfahren wie „Roaming“ eine nahezu gleichbleibende Funkempfangsqualität bei wechselndem Standort erreicht werden. Dieses ist mit nahezu allen heute erhältlichen Access-Points möglich.

Um auch entfernt gelegene Bereiche zu erreichen, wurde das GoeMobile mit einigen Access-Points auf besonders exponierte, hohe Standorte erweitert. Dieses sind in Göttingen das Rathaus in einer Kooperation mit der Stadt Göttingen, das zentrale Hörsaalgebäude der Universität sowie ein 84 m hoher Schornstein.

Abb. 1 - 3: Exponierte Standorte für Access-Points



Eine flächendeckende Erreichbarkeit des GoeMobile ist allerdings allein schon aufgrund der geringen Sendeleistung und des finanziellen Aufwands in einer Stadt wie Göttingen nicht denkbar.

Die GWDG hat für die Außenbereiche des GoeMobile eine sog. Funk-LAN-Box entwickelt, die der Witterung standhält und darüber hinaus alle für das Funk-LAN erforderlichen Komponenten in einer kompakten Box vereinigt.

3.1 Funk-LAN-Box der GWDG

Abb. 4: Innenansicht einer Funk-LAN-Box der GWDG

In dieser Box sind zwei APs mit jeweils vier Karten und dem Anschluss von vier externen Antennen enthalten. Darüber hinaus sind auch ein interner Ethernet-Switch, ein Glasfaser-Medienwandler, ein Blitzschutz sowie eine Steckdosenleiste enthalten.

3.2 GoeMobile in Zahlen

Aufgrund der vorangegangenen Tests von Funk-LAN-Systemen diverser Hersteller haben wir uns im GoeMobile für die Systeme der Firma Lucent/ Orinoco entschieden (Stand: 11/2000).

Im GoeMobile sind bereits ca. 100 Access-Points installiert: davon 70 Accesspoints mit zwei Antennenanschlüssen und 30 Access-Points mit einem Antennenanschluss. Etwa 65 Access-Points besitzen externe Antennen, die restlichen werden mit der in der Funkkarte integrierten Antenne betrieben.

Als externe Antennen haben sich die Sektorantennen als geeignet herausgestellt, da diese für Außenmontage gedacht sind, einen breiten Öffnungswinkel (120 Grad) besitzen und einen guten Antennengewinn aufweisen (12 dbi). Dieser Antennentyp wird im GoeMobile am häufigsten eingesetzt, gefolgt von 7-dbi-Rundstrahlantennen, welche meist in Innenräumen verwendet werden.

3.3 Test der Funkabdeckungsqualität

Eine korrekte Planung der Standorte ist essentiell für eine gute Erreichbarkeit des GoeMobile.

Ein spezielles Programm, eine Eigenentwicklung der GWDG, ermöglicht durch zusätzliche Nutzung eines GPS-Empfängers eine Erfassung der Funkabdeckungsqualität. Dabei werden Ortsinformationen des GPS und Verbindungsqualitätsdaten mit Zeitstempel in einer Tabelle festgehalten, die im Anschluss durch eine Datenbank ausgewertet werden kann. Dadurch ist eine qualitative und vor allen Dingen rasche Aussage über die Funkausleuchtung möglich.

3.4 Zentrales Management des GoeMobile

Durch die zentrale Nutzung einer Management-Software und der Integration in ein bestehendes Netzmanagement-System (HP-Openview) kann auf etwaige Fehlersituationen schnell reagiert werden. Zusätzlich erlauben eigene Scripts, welche über SNMP direkten Zugriff auf die Access-Points haben, die schnelle Zustandsabfrage sowie eine zentrale Konfiguration der APs und dessen nachträgliche Anpassung auf sich verändernde Gegebenheiten.

3.5 Benutzerverwaltung und Zugang

Nutzungsberechtigt sind neben den Studierenden und Beschäftigten der Universität Göttingen auch weitere lokale wissenschaftliche Eirichtungen in Göttingen, wie z. B. die vier Max-Planck-Institute. Um Zugang zum Funk-LAN zu bekommen, ist lediglich ein gültiger Studierenden- oder GWDG-Account erforderlich und natürlich eine Funkkarte.

Jeder Benutzer hat die Möglichkeit sein eigenes Profil auf einer Web-Seite zu gestalten. Dort können Funkkarten, die der Benutzer betreiben möchte, eingetragen werden, woraufhin dann automatisch seine Karten für den Zugang zum GoeMobile zugelassen werden.

4. Sicherheit im Funk-LAN „GoeMobile“

Sicherheit ist das entscheidende Thema beim Betrieb eines Funk-LANs. Dies wird deutlich, da sämtliche, via Funk übertragenen Daten potentiell von Unberechtigten abgehört werden können. Der Zugang zum Funk-LAN ist, ohne Verwendung besonderer Sicherheitsmechanismen, für Unberechtigte erheblich einfacher, als dieses bei kabelgebundenen Netzen der Fall ist.

Der Einsatz von Sicherheitsmechanismen und die klare Abgrenzung der Benutzergruppen ist ein Muss für den Betrieb eines sicheren Funk-LANs.

Bereits der Standard für Funk-LAN IEEE 802.11b enthält einige, allerdings noch nicht nicht ausreichende Verfahren, das Funk-LAN sicherer zu machen. Einige dieser Verfahren werden im Folgenden näher betrachtet.

4.1 WEP (Wired Equivalent Privacy)

WEP ist ein Mechanismus, die Funkübertragung mit festen, vordefinierten „Keys“ zu verschlüsseln. Der Aufwand ist dabei recht groß, da auf Seiten der APs die Schlüssel vorgehalten werden und dem gesamten Benutzerkreis diese Schlüssel zur Verfügung gestellt werden müssen. Das ist de facto ein offenes Geheimnis und dadurch nicht besonders für einen sicheren Funk-LAN Zugang geeignet. Zudem ist der Verwaltungsaufwand sehr groß. Überdies ist es für Eindringlinge möglich, durch rein passive Attacken (einfaches Mithören) die Schlüssel nach einer bestimmten Anzahl empfangener Bytes zu ermitteln. Die Benutzer merken davon nichts.

WEP ist Bestandteil des Standards 802.11b und benutzt den RC4-Algorithmus von RSA Security Inc. Er ist mit Schlüsselstärken von 40 Bit (Standard) sowie 104 Bit implementiert, bei einem 24-Bit-Initialisierungsvektor.

Vorteile:

- In jedem 802.11b Gerät verfügbar
- Hardware-unterstützt
- Software-unabhängig

Nachteile:

- Manuelle Schlüsselverwaltung
- Keine Benutzerauthentifizierung
- 40-Bit-Schlüssel gelten als nicht sicher
- RC4-Algorithmus hat Designschwächen
- Bereits passive Attacken können Schlüssel ermitteln

Fazit:

WEP ist für eine Verschlüsselung zwar geeignet und besser als eine unverschlüsselte Übertragung, bietet aber allein schon aufgrund seiner Designschwächen keinen ausreichenden Schutz.

4.2 Service Set Identifier (SSID)

SSID ist ein sehr einfaches Verfahren, ein Funk-LAN vor dem Zugriff Dritter zu schützen. SSID ist quasi der geheim zu haltende Name eines Funk-LANs. Allerdings muss dieser Name auch den Benutzern des Funk-LANs zugänglich sein, damit die Benutzersysteme mit dem richtigen SSID am Netz teilnehmen können. Deshalb kann die SSID meist in der Praxis nicht lange geheim gehalten werden, wenn die „legale“ Benutzergruppe entsprechend groß ist.

Vorteile:

- Software-unabhängig
- Schnell und einfach einzurichten

Nachteile:

- Muss jedem Teilnehmer bekannt sein
- Nur ein SSID pro AP
- Lässt sich in großen Netzen nicht wirklich geheim halten

Fazit:

SSID ist lediglich als Zusatz zu sehen. Die Verwendung der SSID allein bietet keinen ausreichenden Schutz.

4.3 Media Access Control (MAC) Address Filtering

Jede Funkkarte besitzt eine eindeutige MAC-Adresse. Beim MAC-Filtering wird lediglich der Zugang für eingetragene Funkkarten (MAC-Adressen) ermöglicht. Dabei wird ein im Netz integrierter RADIUS-Server von den Access-Points befragt, ob eine Funkkarte mit der MAC-Adresse berechtigt ist am Funk-LAN teilzunehmen.

Vorteile:

- Software- & Client-unabhängig
- Keine Aktion des Benutzers notwendig

Nachteile:

- Jede berechtigte Netzwerkkarte muss erfasst werden
- MAC-Adressen lassen sich leicht fälschen
- MAC-Adresslisten auf den APs lassen sich schwer pflegen

Fazit:

Das alleinige Filtern von MAC-Adressen ist für einen sicheren Funk-LAN-Betrieb nicht ausreichend. Zum einen kann die MAC-Adresse auf Benutzerseite bei vielen Funkkarten leicht verändert werden und zum anderen ist diese nicht benutzerbezogen. Das Filtern von MAC-Adressen ist lediglich als Zusatz zu weiteren Sicherungsmechanismen zu sehen und schützt das Funk-LAN lediglich vor zufällig am Funk-LAN teilnehmenden Personen, die eigentlich keinen Zugriff bekommen sollen.

4.4 MS Point-to-Point Tunneling Protocol (MS-PPTP)

PPTP ist eine Microsoft-spezifische Implementierung des PPTP. Dieses Protokoll erlaubt das Tunneln von Point-to-Point-Protocol (PPP)-Verbindungen über TCP/IP. Es gehört in die Gruppe der VPN-Protokolle, die eine verschlüsselte Verbindung zwischen zwei Partnern erlauben. PPTP existiert in zwei Versionen: MS-CHAPv1 und MS-CHAPv2. Es erlaubt überdies eine Authentifizierung des Benutzers, was für einen Einsatz im GoeMobile ein wesentlicher Vorteil ist. PPTP benutzt den RC4-Algorithmus von RSA Security und besitzt 40-Bit- oder 104-Bit-Schlüssellängen.

In der Anfangsphase des GoeMobile wurde PPTP eingesetzt, da es bereits Bestandteil der meisten Betriebssysteme von Microsoft ist und auch Implementationen bei Linux und OpenBSD existieren.

Allerdings hat auch das in MS-PPTP verwendete MS-CHAPv1 schwere Sicherheitslücken, was für einen dauerhaften Einsatz im GoeMobile nicht geeignet ist. Deshalb sind wir gezwungen, nach Alternativen zu suchen.

Vorteile:

- Auf allen gängigen MS-Betriebssystemen verfügbar
- Bietet Verschlüsselung und Benutzerauthentifizierung

Nachteile:

- 40-Bit-Schlüssel gelten als nicht sicher
- MS-CHAPv1 hat schwere Sicherheitslücken
- Protokoll hat Designschwächen

Fazit:

MS-PPTP ist besser als keine Verschlüsselung. Es ist anfällig gegen Kryptoanalyse und gilt als unsicher. MS-PPTP ist darüber hinaus nicht zukunftssicher und für einen dauerhaften Einsatz in WLANs ungeeignet.

4.5 Internet Protocol Security (IPSec)

Auch IPSec gehört zu den VPN-Protokollen. Es ist eine Erweiterung der TCP/IP-Protokollsuite. Darüber hinaus genießt es eine weite Verbreitung und ist integraler Bestandteil von IPv6 (IPnG) und damit zukunftssicherer als andere, vergleichbare VPN-Protokolle.

Im IPSec werden zwei unterschiedliche Übertragungsmodi unterschieden:

1. Transportmodus

Dabei wird nur Datenteil verschlüsselt und der IP-Kopf bleibt erhalten.

2. Tunnelmodus

Dabei wird das komplette IP-Paket verschlüsselt. Es bietet aber die Möglichkeit eines Tunnels zwischen zwei Netzen.

Vorteile:

- Standard und auf vielen Plattformen verfügbar
- Keine festgelegten Algorithmen
- Keine bekannten Designschwächen
- Hoher Verschlüsselungsgrad

Nachteile:

- Keine Benutzerauthentifikation
- Clients müssen korrekt konfiguriert werden
- Oft zusätzliche Software auf Client erforderlich

Fazit:

IPSec ist besser als keine Verschlüsselung. Es unterstützt als sicher geltende Algorithmen wie Blowfish, IDEA, MD5 oder SHA und ist dadurch flexibel. Es hat eine weite Verbreitung in findet Anwendungen in vielen Komponenten und Geräten.

Aufgrund der Flexibilität ist es ein geeignetes Verschlüsselungsverfahren und insbesondere für den Einsatz im GoeMobile sinnvoll, wenn es mit einer zusätzlichen Benutzerauthentifizierung kombiniert eingesetzt wird.

5. Das Sicherheitsmodell im GoeMobile

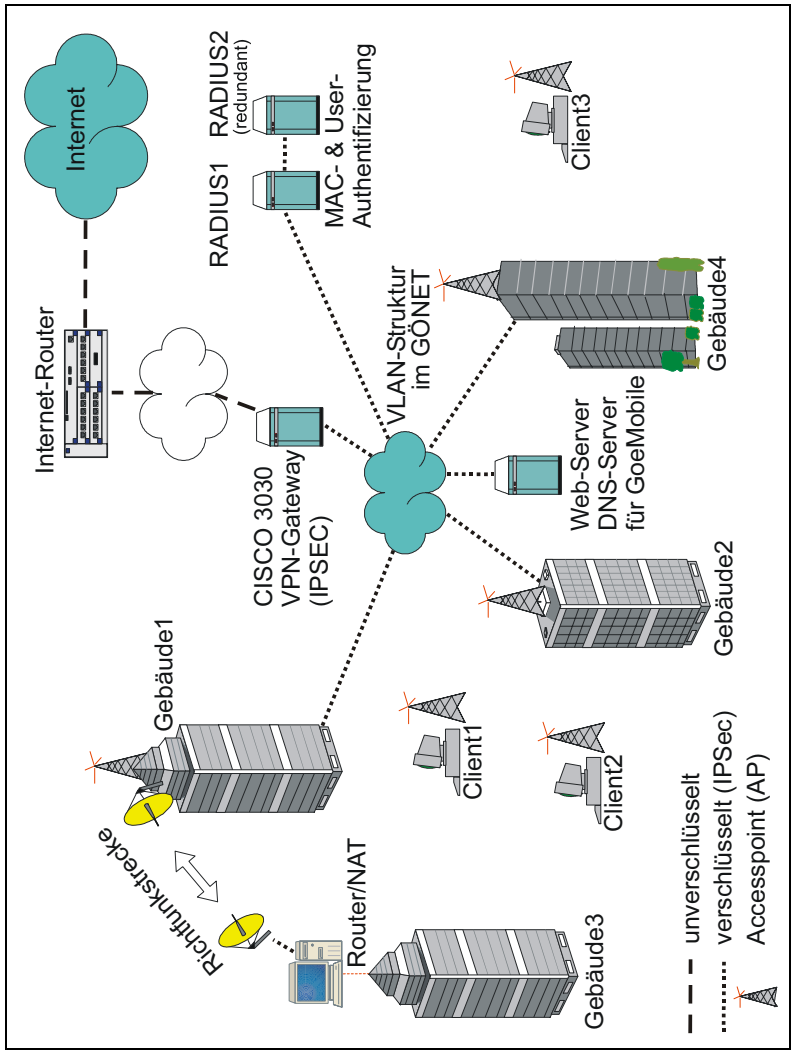
Betrachtet man die aufgeführten Verfahren zur Sicherung eines Funk-LANs, so wird klar, das ein Verfahren allein meist keinen ausreichenden Schutz bzw. Komfort bietet. Erst die Kombination diverser Verfahren macht ein Netz für eine Zeit lang sicher.

5.1 Die Sicherheitsbausteine des GoeMobile

Im Folgenden werden die im GoeMobile eingesetzten Sicherheitskomponenten aufgeführt. Erst die Kombination mehrerer Verfahren sichert ein Funk-LAN wie das GoeMobile ausreichend ab.

1. VLAN: Betrieb aller Access-Points in einem eigenen VLAN. Dadurch wird eine Gruppierung erreicht, wobei die Benutzer ohne weitere Authentifizierung keinen Zugriff auf andere Netze haben.
2. MAC-Adressen filtern: Hierbei werden nur eingetragene Funkkarten im GoeMobile zugelassen. Die MAC-Adressen dieser Karten werden gegen einen RADIUS-Server im Netz gegengeprüft.
3. IPSec: IPSec als Verschlüsselung ist z. Zt. die sicherste Methode für den Zugang der Benutzer zum GoeMobile. Überdies werden Benutzername/ Passwort von der VPN-Software auf dem Benutzerrechner abgefragt und zentral auf einem weiteren RADIUS-Server geprüft.
4. Zentrale Benutzerverwaltung: Die Verwendung der bereits existierenden, regulären Benutzerkonten ermöglicht eine rasche Implementation in bestehende Umgebungen. Eine eigene, für das GoeMobile eingerichtete Benutzerdatenbank ist nicht erforderlich.

Abb. 5: Komponenten des GoeMobile



6. Fazit

Funktechnik nach dem derzeitigen Standard IEEE 802.11b bietet eine deutliche Verbesserung der Netzqualität und ist aufgrund der ausreichenden Bandbreite auch schnell genug, um für die meisten Anwendungen gewappnet zu

sein. Entscheidend ist , dass ein Verständnis und die Sensibiliät beim Betreiber sowie Benutzer für das Thema Sicherheit geweckt werden. Gerade hier offenbaren sich neue Angriffspunkte für Unberechtigte, die einigen Aufwand auf Betreiberseite erfordern. Ein einmalig installiertes Sicherheitskonzept muss ständig den neuen Entwicklungen angepasst werden.

Die Funktechnik darf jedoch nicht darüber hinwegtäuschen, dass ein Funk-LAN, allein schon aufgrund seiner Beschaffenheit und Bandbreite, den schnellen kabelgebundenen Netzzugang nicht ersetzen kann. Ein Funk-LAN ist lediglich als Erweiterung zu verstehen. Wer als Betreiber, wie die GWDG, versteht, die Risiken und Gefahren des Funk-LANs in den Griff zu bekommen, stellt seinen Benutzern ein geeignetes, mobiles und sicheres Zugangsmedium zur Verfügung.

7. Weiterführende Informationen und Quellen

1. Das Göttinger Funk-LAN GoeMobile
   `http://www.goemobile.de`
2. Einfluß von BlueTooth und WLAN
   `http://www.teltarif.de/arch/2000/kw46/s3570.html`
3. Sicherheit in drahtlosen Netzen
   `http://www.networkworld.de/artikel/index.cfm?id=65705&pageid=400&pageart=detail`
4. Hersteller von Funk-LAN-Geräten
   `http://wiss.informatik.uni-rostock.de/hersteller/`
5. 5-GHz-Standards und Hiperlan/2
   `http://www.mez.ruhr-uni-bochum.de/projekte/wlan/mecki_standards.html`
6. 54-MBit-Chips
   `http://www.intersil.com/pressroom/20010619_PRISM_Indigo_German.asp`

KIT - Kompetenzzentrum für Informationstechnologie und IT-Management in der MPG

Andreas Oberreuter

Max-Planck-Institut für Radioastronomie, Bonn

Vorwort

Wenn man heutzutage in der IT-Branche oder besser IKT-Landschaft arbeitet, dann wird man zweifelsohne vom Tempo der Entwicklungen geradezu überfahren, solange man nicht mit ausreichendem Abstand seinem Arbeitswerkzeug „Computer“ begegnet und präventiv alle damit zusammenhängenden Faktoren gesellschaftlicher, ethischer wie natürlich technischer Art abwägt, ehe man sich darauf überhaupt einläßt.

Darum muß immer gelten: Der Computer ist ein Werkzeug, das man beherrschen muß, ansonsten Finger weg. Denn, wenn das Werkzeug uns beherrscht, dann muß etwas falsch gelaufen sein.

Kompetenz bedeutet, daß man zu einer Aussage, Tätigkeit befähigt oder befugt ist, weil man sich über die zugehörige Sachlage, für die man zuständig ist, ein Urteil bilden kann. Kompetenz wird erworben und nicht verliehen, setzt also Erfahrung bzw. Lernen voraus, und ist somit nie statisch, sondern ein fortlaufender Prozeß, der Zeit und Mühen bedeuten kann.

Wenn also über ein Kompetenzzentrum[1] gesprochen wird, das im wesentlichen die Beherrschung eines einzigen Arbeitswerkzeuges mit all seinen Optionen ermöglichen soll, dann könnte man leicht zu dem Eindruck kommen, daß das alles viel zu aufgeblasen ist, zumal doch heute schon Kinder dieses Werkzeug in die Hand gedrückt bekommen und damit scheinbar umgehen können. Wozu brauchen Wissenschaftler und Ingenieure, Techniker und Verwaltungsangestellte dann noch kompetente Unterstützung in ihren tagtäglichen Arbeitsabläufen?

Daß hier noch lange nicht alles „Plug and Play" ist, wie es uns manchmal seitens der IT-Hersteller suggeriert wird, und es darüber hinaus noch eine Vielzahl von Aspekten zu beachten gilt, wenn man sich als IKT-Verantwortlicher wirklich ernsthaft, wirtschaftlich und langfristig nutzbringend für seine Benutzer mit IT beschäftigt, soll der nachfolgende Bericht aufzeigen.

1. IT-Kompetenz in der MPG

Derzeit sind in der MPG ca. 11000 Mitarbeiter, verteilt auf ca. 80 Einrichtungen, in den unterschiedlichsten wissenschaftlichen Fachgebieten tätig, um weltweit anerkannte Spitzenforschung zu betreiben. Dabei wird auf vielfältige technische wie handwerkliche Grundlagen aufgebaut. Eine dieser Grundlagen bildet die IKT, also die Informations- und Kommunikationstechnologie.

In den meisten Fällen wird jede Einrichtung durch eine mehr oder wenige große IT-Abteilung oder IT-Verantwortliche unterstützt. Aufgrund der unterschiedlichen Anforderungen und Gegebenheiten bildet jede lokale IT-Einheit eine IT-Kompetenz-Zelle für sich selbst. Das ist wichtig und notwendig, um die Unabhängigkeit der lokalen Einrichtung in diesen Basisdiensten zu gewährleisten.

1. Im folgenden oftmals KIT, gemäß dem Titel dieses Berichtes, genannt. Dabei kann aber auch ein beliebiges Kompetenzzentrum gemeint sein, welches die angesprochenen Themengebiete abdeckt.

Schon früh erkannte man bei der MPG, daß es aber Sinn macht, außergewöhnliche Ressourcenanforderungen einzelner Einrichtungen durch zentrale Anlaufstationen (z.B. Kompetenzzentren für Hochleistungsrechnen) wirtschaftlich wie zweckmäßig zu bedienen. Aus diesen Bestrebungen sind die bekannten Kompetenzzentren RZG, GWDG und DKRZ entstanden. Mancherorts sind weitere Zusammenführungen der IT-Bedürfnisse vorgenommen worden (FHI Berlin), um durch Bündelung IT-Wissen und Ressourcen großzügiger aufbauen zu können. Damit hat der wissenschaftliche Effort einen deutlichen Schub erhalten.

Der IT-Sektor hat inzwischen mehrere Generationen/Phasen durchlebt, die von absolut zentralistischer bis zu völlig dezentralistischer Sichtweise mancherlei Umbrüche mit sich gebracht haben. Aus den Vor- und Nachteilen der beiden Extreme als auch der dazwischen befindlichen Betrachtungsweisen ist gelernt worden, so daß sich ein gewisses Gleichgewicht entwickelt hat.

Allerdings hat sich die Komplexität der Themen, mit denen sich heute die IT-Belegschaft auseinandersetzen muß, erheblich ausgedehnt und so findet man ein Ungleichgewicht an Wissen bzw. an Zeit, Mittel oder Planstellen, allen inzwischen als üblich geltenden Ansprüchen noch gerecht zu werden.

Hier sind es vor allem die Standardaufgaben, die den IT-Alltag derart belasten, so daß außergewöhnliche Ereignisse (neue Hard-/Software, Hacker, Systemausfall, ...) schnell alle Planungen über den Haufen werfen und für zukunftsweisende Konzepte, Visionen, aber auch optimalen Support der Benutzer kaum mehr Zeit bleibt.

Wie eingangs erwähnt ist aber gerade in den Standardbereichen viel erworbene Kompetenz vorhanden, die richtig abgerufen und weitervermittelt, reichlich Freiräume für o.g. Nicht-Standard-Situationen lassen sollte. Es gilt also die Standardverfahren zu optimieren, damit diese in den Hintergrund treten und beherrschbar bleiben, um den Ausnahmen mehr Konzentration schenken zu können.

Im folgenden sollen daher Kernthemen der IT aufgelistet werden, die leicht optimiert werden können, um als Standards sowohl die Arbeitsbelastung als auch den nicht minderwichtigen Etat weitestgehendst zu schonen.

2. Optimierbare Kernthemen

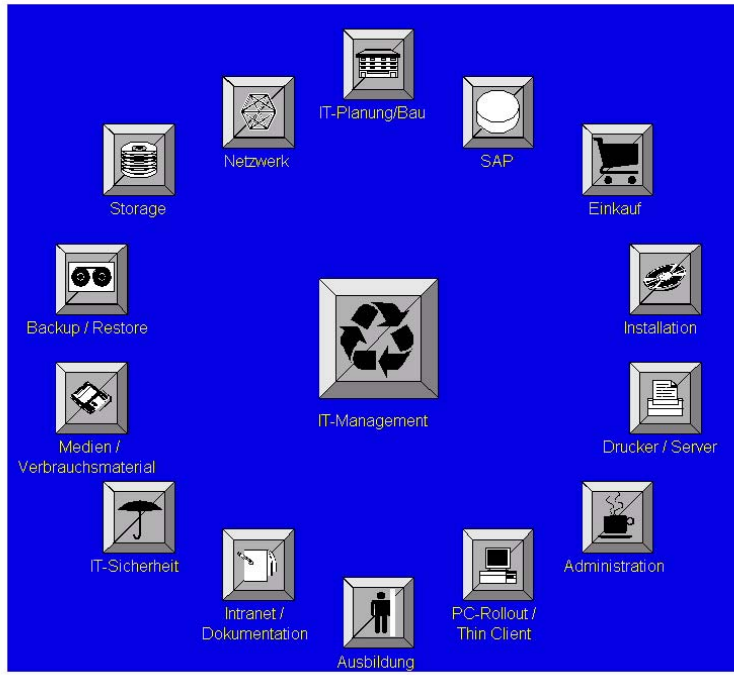
Wenn hier von Kernthemen gesprochen wird, dann sind i.d.R. immer IT-Themen gemeint, die von mehr als 80% aller Einrichtungen tagtäglich in Anspruch genommen werden dürften. So interessant auch spezielle Schwerpunkte in der IT sein können, müssen diese doch oftmals sehr individuell angegangen werden und interessieren daher nicht die Gesamtheit aller Einrichtungen.

Die Auswahl und spätere Begründung der Kernthemen zeigt, daß es nicht ausreicht, diese mit Scheuklappen losgelöst voneinander zu betrachten, sondern deren tw. enge Verzahnung aus einiger Distanz genau zu beachten, um wirklichen Nutzen aus der Behandlung der einzelnen Themen ziehen zu können.

Generalisten und Spezialisten müssen daher in einem ausgewogenen Verhältnis miteinander zusammenarbeiten, um die höchste Effizienz zu erzielen.

Folgende Themenkomplexe stehen zur Diskussion:

- IT-Planung / Bau
- Netzwerke
- Serverdienste
- Installation und Administration
- Storage und Backup / Restore
- Drucken und Präsentieren
- Medien, Verbrauchskosten
- Clients, Thin Clients, Rollouts
- Sicherheit, best setups
- Einkauf, e-Commerce, e-Procurement, SAP und IT-Management
- Dokumentation und Vermittlung
- Ausbildung

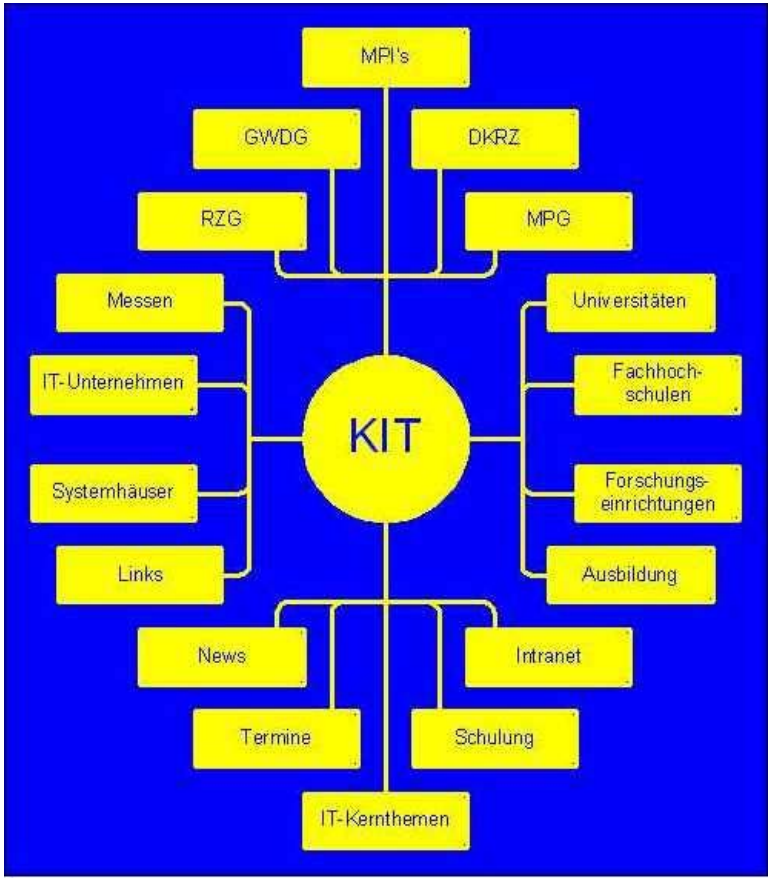


Diese Bereiche muß man natürlich differenzieren, um hier und dort Schwerpunkte zu setzen, die zu sinnvollen Standards führen, um bspw. in einer Community wie der MPG mit ca. 20.000 Computern (geschätzt), ähnlich einem Großunternehmen, zum Tragen zu kommen.

Vor allem muß sich jemand für den einen oder anderen Bereich verantwortlich fühlen, um diesen stets zu aktualisieren und Lösungen anzubieten.

3. Die Aufgaben eines Kompetenzzentrums

Kernthemen wie gerade angeführt lassen sich nur dann effektiv angehen, wenn zum einen auf die eigene, schon vorhandene Kompetenz im Unternehmen/Gesellschaft zurückgegriffen wird, und zum anderen fehlende Kompetenz durch Einholen von externer Kompetenz kompensiert wird.



In der Forschung und Lehre empfiehlt es sich, die Kooperation mit anderen F+L-Einrichtungen als auch mit den IT-Herstellern direkt zu suchen. Aus diesem Grunde ergibt sich zwangsläufig die abgebildete Vernetzung von Kompetenzquellen.

Das Wissen der IT-Hersteller und spezialisierter Systemhäuser, inklusive vergleichbarer Anwender in Universitäten, Fachhochschulen und Forschungseinrichtungen muß zentral zusammengeführt, aufgearbeitet und dann in die eigenen Einrichtungen eingebracht werden.

Dabei sollen die bereits existierenden Verbindungen der MPG-Einrichtungen zu den anderen Netzenden sinnvoll akkumuliert werden, so daß für alle der größtmögliche Nutzen entsteht. Es ist leicht ersichtlich, daß es hierzu einer koordinierenden Stelle bedarf, die mit dem nötigen Weitblick und entsprechender „Master"-Kompetenz Vorschub leistet.

4. Begründung der ausgewählten Kernthemen

Die Auswahl der Kernthemen mag dem einen oder anderen recht subjektiv erscheinen, ist aber bei genauerer Betrachtung recht sorgfältig abgewogen worden. Wer sich nun, wo auch immer, dieser Themen annimmt, steht hier nicht zur Diskussion, sondern allein die Tatsache, daß man möglichst bald die folgenden Fragestellungen aufgreift, weil hier derzeit innerhalb der MPG enormer Nachholbedarf besteht.

- IT-Planung / Bau
- Netzwerke
- Serverdienste
- Installation und Administration
- Storage und Backup / Restore
- Drucken und Präsentieren
- Medien, Verbrauchskosten
- Clients, Thin Clients, Rollouts
- Sicherheit, best setups
- Einkauf, e-Commerce, eProcurement, SAP und IT-Management
- Dokumentation und Vermittlung
- Ausbildung

4.1 IT-Planung und Bau

Im wesentlichen tritt dieses Thema bei Neubauten oder Altbausanierungen ans Licht und erfordert eine enge Zusammenarbeit von Bauabteilung, Architekten, ausführenden Betrieben, Direktoren, Abteilungsleitern und letztlich der IT-Abteilung.

Dabei sind strategisch wichtige Aspekte, die sorgfältige und langfristige Planung von Brandschutz, Kabeltrassen, -kanälen, Verkabelungsstrategien (Etage, Zimmer), Netzwerktopologien (z.B. wireless, optisch, ...), Stromversorgungen (Rechenzentrum, Zimmer, Labors), Klimaanlagen und Serverräumen, um nur einige zu nennen.

Wenn die o.g. Zusammenarbeit nur partiell verläuft, insbesondere die IT-Abteilung außen vor bleibt oder zu spät informiert wird, sind skalierbare Vorschläge und praktikable Lösungen meist nicht mehr realisierbar. Darum erfordert dieser Themenkomplex frühestmögliche Absprache und weitsichtige Umsetzung mit allen Beteiligten.

Sehr hilfreich wäre die Erstellung von Checklisten für die Planung und ggfs. Durchführung, die durch eine zentrale Arbeitsgruppe mit den o.g. Teams erstellt und mehr oder weniger als Standard ausgearbeitet wird. Das Mitspracherecht der Institute muß gewährleistet sein, denn diese sind schließlich die Endnutzer.

Für den IT-Bereich sollten unbedingt unabhängige externe Firmen zur Begutachtung und grundlegenden Planung mit einbezogen werden, die diese gerade genannten Themen tagtäglich bei Kunden installieren.

Aufgrund der hier typischerweise hohen Auf-/Um-Bau-Kosten sollte eine Arbeitsgruppe für dieses grundlegende Kernthema geschaffen werden, um die Institute vor Fehlplanungen und Einschränkungen im Betrieb bewahren. Ansätze sind in der MPG vorhanden, aber bislang vollkommen unzureichend in der Praxis umgesetzt.

Neben der reinen Bauplanung gehört die Rechenzentrumsplanung zu einem der grundlegendsten Kapitel eines IT-Kompetenzzentrums, weil hier die Basis jedes weiteren Vorgehens gelegt wird.

Hier kann sicher auf reichlich Erfahrung in den einzelnen Instituten bzw. bei externen Dienstleistern (die tw. durch ehemalige MPI-DV-Leiter geführt werden) gebaut werden. Es gilt dieses Wissen transparent an alle Interessierten weiterzugeben, damit nicht jeder das Rad neu erfinden muß. Ein Themenkatalog und Aufbau-Pläne in Form von Checklisten sollten da Abhilfe

schaffen und langfristig zu ähnlichen ausgerichteten IT-Abteilungen in der MPG führen, was die zukünftige Entwicklung derselben leichter macht.

Es ist also eine große Herausforderung, die bekannten Defizite, Reibungspunkte und Unzufriedenheit in dem einen bzw. anderen Bereich anzugehen und durch weitsichtiges Vorgehen abzubauen, so daß auch weiterhin die Basis für Spitzenforschung gehalten werden kann.

4.2 Netzwerke

Unmittelbar an die Bauinfrastruktur schließt sich die Frage der Vernetzung eines Institutes an (LAN, WAN). Aufgrund der unterschiedlichen Erfordernisse der Institute kann hier nicht alles vereinheitlicht werden, aber gewisse LAN-Topologien, WAN-Zugänge und Sicherheitsansätze können, ja müssen als Konzeptionsmuster den Einrichtungen frei zugänglich sein.

Insbesondere typische Konfigurationsdaten (Router, Firewall, ...) von aktiven Komponenten (u.a. Switches, Einwahlrouter, ...) bauen im wesentlichen auf den gleichen Sicherheitsstandards auf, vielfach auch auf den gleichen Betriebssystemen oder Softwarelösungen.

Einigt man sich auf ein Set von Hardwareprodukten, die in einem gesonderten „Lab“ sämtliche LAN/WAN-Strukturen eines Institutes nachbilden können, dann lassen sich Standardlösungen erstellen, die leicht an die Bedürfnisse und Gegebenheiten eines Institutes angepaßt werden können. Das hat den Vorteil, daß nicht überall vor Ort Netzwerkexpertise bis ins Detail vorhanden sein muß, um Netzwerkkomponenten zu kaufen und einzubinden.

Solche Referenzinstallationen würden namhafte Netzwerkkomponenten-Hersteller einrichten, in der MPG bekannte Netzwerkpartner einige wenige Male konfigurieren und ließen sich dann „hausintern“ beliebig oft klonen. Eine MPG-interne Arbeitsgruppe sondiert regelmäßig produkt- und herstellerneutral den Markt und empfiehlt dann den Einrichtungen schlüsselfertige Lösungen. Statt lokal einige Mitarbeiter durch Schulungen im Netzwerkbereich zu zertifizieren, könnte das zentral geschehen, so daß MPG-interne Netzwerk-„Consultants“ allen zur Verfügung ständen.

4.3 Serverdienste

Ausgehend vom erstmaligen Aufbau einer IT-Abteilung, ergibt sich aber auch in größeren Zeitabständen immer wieder mal der Wunsch, seine Serverlandschaft zu straffen, auszubauen und/oder zu homogenisieren.

Dabei sind grundlegende Dienste, wie bspw. NIS/LDAP-, DNS-, WWW-, FTP-, Mail-, Print- und Fileserver, ... aufzusetzen. Vielfach existieren sogenannte out-of-the-box-Lösungen, die sich hier anbieten würden. Um zu vermeiden, daß Institute einen recht großen zeitlichen Aufwand betreiben, alle Dienste selber zu konfigurieren bzw. aufeinander abzustimmen, könnten hier Standards zu einem wesentlich schnelleren Aufbau der Grundfunktionalitäten führen.

Es gilt daher in diesem Bereich, den aktuellen Markt nach verallgemeinerbaren Lösungen zu sondieren und diese für die Einrichtungen aufzuarbeiten. „Plug and Play“ im Bereich von Serverdiensten statt mühevollem Studium von Handbüchern und halb intakter Hardware.

Sind diese Dienste dann einmal implementiert, ist eine nachfolgende Schulung effektiv und vertieft das, was man zuvor per Checkliste oder Auto-Installation gestartet hat.

Frühe IT-Planung und weitsichtiger Entwurf von Hierarchien, z.B. in heterogenen Unix/Windows/Mac-Umgebungen für das File-System, ersparen dann später aufwendige Korrekturen. Wie in allen anderen hier genannten Kernthemen wird die dokumentierte Verallgemeinerung von Konzepten zum Schlüssel für IT-Management. Jeder Sonderwunsch an Hard- oder Software, der in einem Institut anfallen kann, ist dann schnell integriert.

4.4 Installation und Administration

Tagtäglich müssen Rechner neu installiert werden. Inzwischen existieren recht brauchbare Installationsserver (jeweils eines Herstellers natürlich), die dabei einen großen Teil der Arbeit abnehmen können und einheitliche Server/Client-Landschaften erlauben.

Die hier bereits in der MPG eingesetzten Mechanismen sollen zusammengetragen und nach Prüfung der Einsatzfähigkeit an die Institute auf Anfrage weitergegeben werden. Daneben sollten auch neue Verfahren ausführlich getestet werden. Damit sparen die lokalen Betreuer Zeit, um Routinearbeiten besser angehen zu können.

Nach der Installation nimmt die Administration den wesentlichen Teil der Arbeitszeit der IT-Abteilung in Anspruch, insbesondere, wenn man über keine geeigneten Werkzeuge verfügt, die Arbeiten abnehmen.

Oftmals ist nicht einmal bekannt, daß es für Standardaufgaben, wie Benutzer verwalten, Verzeichnisse erzeugen, Software zu verteilen, ..., schon längst gute Tools gibt. Oder aber es fehlt die Zeit, aktuelle Tools objektiv bewerten zu können.

Da das KIT im Idealfall nicht mit Alltagsarbeiten belastet ist, wie die lokalen IT-Abteilungen, sollte sich eine Abteilung mit genau dieser Problematik in Ruhe beschäftigen und marktübliche Werkzeuge auf die Leistungsfähigkeit im Umfeld einer Forschungseinrichtung testen.

Im Falle von kleinen Instituten kann evtl. sogar eine Fernwartung/-administration druch das Kompetenzzentrum sinnvoll sein. Gängige Beispiele findet man bereits bei der GWDG. Der Aufbau eines Helpdesks ist sehr wünschenswert, insbesondere wenn hier zertifizierte und erfahrene Administratoren Auskunft geben.

Themen wie Benutzerverwaltung, Geräteverwaltung (Rechner, Peripherie, Netzwerk), Fernwartung von Serverdiensten, Clients und aktiven Komponenten, das Update, Upgrade, Patch- und Softwareverteilung sowie ein Helpdesk sind in allen Einrichtungen noch zu optimierende Arbeitsgebiete.

4.5 Storage und Backup / Restore

Steht einmal das Grundgerüst einer IT-Abteilung und der angeschlossenen Einrichtungen, sind bald Fragen nach der Skalierbarkeit von Plattenplatz und zugehöriger Datensicherung da. Dieses Marktsegment explodiert geradezu und erfordert stets neueste und zumeist recht teure Technologien (NAS, SAN, iSCSI, ...).

Den tatsächlichen Nutzen der einen oder anderen Technologie einzuschätzen, muß die Aufgabe eines Kompetenzzentrums sein und hier dann klare (herstellerunabhängige) Empfehlungen an die Institute weiterzureichen.

Namhafte Hersteller haben auch hier angeboten, Teststellungen für ein zentrales Storage/Backup/Restore-„Lab“ bereitzustellen, in dem die verschiedenen Anforderungen in Ruhe getestet werden können und den IT-Abteilungen der angeschlossenen Einrichtungen die Möglichkeit gegeben wird, Hard- und Software in einem mehr oder weniger realen Umfeld zu sehen. Neben reiner Hardware ist nämlich insbesondere die Administration komplexer Storagelösungen eine Herausforderung an sich, deren Funktionalität man nur in einem Testszenario richtig beurteilen kann.

Neben den reinen Herstellern haben diverse Systemhäuser ausgezeichnete Kenntnisse auf diesem Gebiet, so daß eine Zusammenarbeit mit einem Systemhaus Sinn macht, zumal im Storage-Bereich leicht sechs- bis siebenstellige Kosten entstehen können, die man unbedingt langfristig anlegen sollte.

Natürlich haben auch unsere Großrechenzentren (GWDG, RZG, DKRZ) hier schon seit Jahren viel Praxiserfahrung angesammelt, auf die zurückgegriffen

werden muß. Das KIT sieht sich hier als Testcenter und Informationssammelstelle, das alle Links für die IT-Abteilungen bereitstellen soll, sich optimal auf dieses zukunftsträchtige Gebiet vorzubereiten, Fehler bei der Anschaffung zu vermeiden und Konzepte effektiv umzusetzen.

Auch können im Bereich von RAID-Systemen, Tape-Libraries (Serverless Backup), ... durch Sammelbestellungen oder Rahmenverträge enorme Kosten eingespart werden, die dann in andere infrastrukturelle Maßnahmen oder eben wissenschaftliche Projekte fließen können.

Wer viel Storage einsetzt, muß sich noch viel mehr um das richtige Backup/Restore kümmern. In den letzten Jahren haben Tape-Libraries die unterschiedlichen Einzelbandlaufwerke immer mehr abgelöst und aus kleinen Shellscripten wurden mächtige Softwaretools, die das Backup/Restore nun handeln.

Aufgrund der zuvor genannten Problematik des Storage sind die Strategien und die Hard- und Software, die man einsetzt, nicht ganz unbedeutend und ebensowenig billig. Ein Kompetenzzentrum muß hier einfache Lösungspakete in Form von Hard- und Software bilden, die zum einen eine mehr oder weniger homogene Backup-Landschaft in der MPG erlauben und zum anderen leicht auf viele Institute zu adaptieren sind.

Ebenso gibt es reichlich Bedarf an Schulung, um diese doch sehr umfangreichen Werkzeuge optimal einzusetzen (Installation, Administration, Backup schedules, ..). Zudem ist die frühzeitige Bewertung neuer Technologien unerlässlich, um nicht ins Hintertreffen mit der im Institut angebotenen Hardware zu geraten (bspw. DAT, DLT, LTO, SDLT, ..., was kommt danach?).

Hilfreich ist sicherlich immer eine aktuelle Liste der in der MPG zum Einsatz kommenden Hard- und Software für Storage/Backup/Restore, damit der informelle Austausch besser gelingt. Dazu später beim Thema „IT-Portal“ mehr.

Eine Konsolidierung bei der Anschaffung wird sowohl aus rein wirtschaftlichen als auch administrativen Gründen sehr nützlich sein. Eine Backup-Arbeitsgruppe wird daher auch sehr eng mit einer Storage-Arbeitsgruppe zusammenarbeiten müssen und das o.g. „Lab“ gemeinsam betreiben. Ferner ist eine Zusammenarbeit mit der Arbeitsgruppe (Medien, Verbrauchskosten) ratsam, da die Folgekosten bei den eingesetzten Medien oder ein Medienwechsel auch nicht zu vernachlässigen sind.

4.6 Drucken und Präsentieren

Drucker spielen in der MPG eine große Rolle, insofern diese die wissenschaftliche Arbeit zu Papier bringen und bspw. bei Postern die Öffentlichkeitswirksamkeit beeinflussen.

Doch in einem komplexen LAN sind Drucker nicht immer leicht einzubinden, besonders wenn heterogene Rechnerplattformen und Betriebssysteme zum Einsatz kommen, wie in der MPG meist der Fall. Einen Druckerserver zu installieren, der (fast) allen Bedürfnissen entgegenkommt, ist daher für viele EDV-Abteilungen immer noch mit viel Aufwand verbunden.

Neue Druckertypen, -treiber und Verbrauchsmaterialien wirken sich auf die Administration, die Kosten und die Benutzerfreundlichkeit aus.

Das KIT soll universelle Druckerserver aufbauen helfen, die mit den aktuellen Druckern (einiger ausgewählter Hersteller) optimal zusammenarbeiten, sowohl unter Windows/MAC wie auch unter UNIX.

Auch hier sind kostenfreie Teststellungen durch IT-Hersteller möglich, um solche Server aufbauen und testen zu können. Der Aspekt der nachfolgenden Vermarktung ist ein nicht zu unerheblicher Faktor bzw. Anreiz bei all diesen „Test-Labs“ und sollte unbedingt in Anspruch genommen werden.

Interessant könnten langfristige Wartungs-/Leasing-/Kaufmodelle sein, die durch die IT-Management-Abteilung des Kompetenzzentrums erarbeitet werden sollten. Hier werden ordentliche Einsparpotentiale erwartet.

Neben den Papierausgaben haben elektronische Präsentationen enormen Zuwachs gezeigt. Die entsprechende technische Ausstattung für den stationären bzw. mobilen Einsatz ist preiswerter, aber sicherlich noch nicht billig geworden, und Beratungshilfen sind gewünscht. Dieser Bereich könnte zukünftig mit dem Segment „Videokonferenz“ einen eigenen Arbeitskreis erfordern, der über die bisherigen Einzelinitiativen hinausgeht.

4.7 Medien, Verbrauchskosten

Täglich fallen in jedem Institut nicht unerhebliche Kosten für Verbrauchsmaterialien an. Seien es Druckermaterialien, wie Toner, Tinte, Papier und Folien (oder Ersatzteile wie Fuser, ...) oder aber Datenmedien, wie Disketten, ZIP/JAZ, Bänder jeder Art (DAT, DLT, ...), CD/DVD oder optische Datenträger, um nur einige zu nennen.

Die Arbeitsgruppe Medien im KIT soll sich zum einen in Zusammenarbeit mit dem eProcurement-Team mit der Beschaffung der Materialien befassen und aktuelle Preis-Übersichten von typischen IT-Verbrauchsmaterialien

(dazu kann man auch Kabel, Monitore, Drucker zählen) auf einer Webseite („IT-Portal“, s.u.) zum Abruf bereithalten, aber auch mit dem Test und der Erprobung neuer Medien auseinandersetzen, wie bspw. der neuen DVD-Technologie.

Derzeit brennen viele ihre eigenen CD's, bald werden es DVD's sein. Die derzeit erhältlichen Geräte sind weder einheitlich genormt noch preiswert. Darum macht es Sinn, zentral zu testen, was davon sich lohnt wirklich anzuschaffen. Der gleiche Medienwechsel vollzieht sich gerade im Bereich DLT zu Ultrium und weiteren Formaten. Eine herstellerunabhängige und frühzeitige Bewertung soll Kosten sparen helfen, aber auch einen Medienwechsel vorbereiten helfen, z.B. wenn ein ganzes Archiv auf ein neues Speichermedium umgesetzt werden soll.

Hier sollen Erfahrungen aus den Instituten und anderen Rechenzentren gesammelt und bewertet werden. Für die Hardware sollen die geeigneten Treiber auf diversen Plattformen getestet und mit den entsprechenden Scripten oder Konfigurationsdateien den Instituten zur Verfügung gestellt werden. Langes Probieren der örtlichen Gruppen sollte dann der Vergangenheit angehören.

Die Auswahl an Folien, Papier und Toner, Tinte für die Drucker erfolgt in Zusammenarbeit mit der Arbeitsgruppe Drucker, denn die Firmen werfen originale und kompatible Verbrauchsmaterialien auf den Markt und manch einer hat sich bei der Auswahl in Punkto Preis und Kompatibilität und Garantie vergriffen.

Bezüglich der Preisübersichten arbeiten die Gruppen IT-Management und Medien zusammen, bezüglich der praktischen Anwendung ist auch noch die Backup-Gruppe involviert.

4.8 Clients, Thin Clients, Rollouts

In der MPG gibt es schätzungsweise 20.000 Bildschirmarbeitsplätze. Man mag ca. 10% im Bereich Server ansiedeln, so daß immer noch eine immens hohe Zahl an sogenannten Clients übrigbleibt.

Bei der Anschaffung, Installation und Administration dieser Menge wäre ein homogenes Hard- und Softwareumfeld die idealste Lösung, allerdings treffen wir diese schon wegen der unterschiedlichen Anforderungen kaum an. Das Gerätespektrum ist breit, was Hersteller, Alter und Einsatzgebiet angeht.

In manchen Fällen ließe sich aber durch langfristige Konzepte dem Zoo von Geräten dadurch entgegenwirken, indem man Thin Clients einsetzt oder durch regelmäßige Rollouts von Rechnern, diese a) homogenisiert bzw. b)

leistungsmäßig aktualisiert, um state-of-the-art am Arbeitsplatz anbieten zu können, und damit die wissenschaftliche Leistungsfähigkeit stets auf hohem Niveau halten zu können.

Hierzu sind neue Ansätze notwendig, die Miet- oder Leasingmodelle ebenso umfassen, wie Rolloutkonzepte im Unix- und Windows-Umfeld, ohne den Betrieb zu stören, ja vielmehr zu verbessern.

Wie frühere Ansätze, angeregt seitens des BAR, gezeigt haben, reichen die typischen 08/15-Lösungen der Systemhäuser oder Hersteller nicht aus, um das Segment Forschung und Lehre sinnvoll abzudecken. Maßgeschneiderte Standards müssen her.

Was fehlt sind Rolloutpläne, um diesen Hardwarewechsel durchzuführen, der auch immer ein Software-Update/Neueinspielen/Überspielen erfordert. Hier sollen in Zusammenarbeit mit erfahrenen Herstellern geeignete Modelle für die MPI's gefunden und erprobt werden. Ferner soll eine klare Kosten-Leistungsrechnung aufgestellt werden, die es den Instituten erlaubt, ihren Haushalt zu planen und den personellen Aufwand beim Rollout abzuschätzen.

Sollte sich so ein Verfahren verallgemeinern lassen und für eine gewisse Zahl von MPI's lohnen, dann können allein durch gezielte Sammelbestellungen und einmalige Ausschreibungen enorme Summen an Investitions-, Administrationskosten und Arbeitszeit erspart werden. Als Nebenprodukt entstehen mehr oder weniger homogene Landschaften, deren Bestand leicht überwacht und mit flächendeckenden Softwareupdates ausgestattet werden können.

Auch hier haben sich in Vorgesprächen einige Hersteller sehr interessiert gezeigt, Pilotprojekte durchzuführen und Unterstützung anzubieten. Aufgrund der Komplexität der Aufgabe soll sich auch hier eine eigene Arbeitsgruppe konstituieren, die sowohl ein komplettes Pflichtenheft erstellen als auch die Durchführung realisieren kann.

Der zuvor genannte PC-Rollout bzw. die Vorarbeiten in den Pilotprojekten werden sicher schnell ergeben, in welchen Bereichen es sich lohnt auf Thin-Client-Konzepte zu wechseln. Mehrere Institute in der MPG setzen bereits Terminal-Server von Microsoft und Citrix ein. Leider ist die Umsetzung nicht so leicht, wie in den Prospekten versprochen, wie manche Erfahrungen zeigen.

Darum sollten folgende Gesichtspunkte in dieser Arbeitsgruppe vorrangig behandelt werden:

- Objektive Bewertung von technischen Möglichkeiten für a) Verwaltung b) techn. Abt. c) wiss. Anwendungen
- Implementation wieder im eigenen „Test-Lab“ (Server, Load-Balance, Clients) und Pilotprojekten
- Kosten, Lizenzmodelle, Wirtschaftlichkeitsprüfung (TCO, ROI)
- WTS, Metaframe, Alternativen untersuchen
- Erfahrungen in der MPG zusammenführen
- Spezial-Hardware, soweit notwendig, festlegen
- Software testen, die in ThinClient-Umgebungen zum Einsatz kommen kann

Auch in diesem Bereich werden Schulungen zur Zertifizierung angeboten, so daß mindestens ein Mitarbeiter diese Qualifikation ablegen sollte.

4.9 Sicherheit, best setups

Dieses in beinahe alle Unterbereiche der IKT hineinreichende Thema ist nach den Erfahrungen der letzten Jahre inzwischen auch für Forschungseinrichtungen nicht mehr wegzureden. Passende Strategien zu entwickeln, den Markt nach Lösungen zu erkunden und diese dann im eigenen Hause umzusetzen ist ein Muß geworden, erfordern aber meist mehr Man Power als angenommen und somit vorhanden.

Übergreifende Konzepte für sichere Mail-, Web-, SAP- und Logserver im Rahmen einer Router/Firewall/DMZ-Landschaft müssen erarbeitet bzw. weiterentwickelt werden, so daß es sich wiederum leicht auf einen Großteil der Institute adaptieren läßt. Das würde den Administrationsaufwand, aber auch die Anschaffungskosten reduzieren. Außerdem tun sich Betriebsräte und Direktoren leichter, solch heikle Bereiche an eine MPG-interne Einrichtung abzudelegieren, als eine externe Firma.

Manche Softwareprodukte sind bereits durch den Datenschutzbeauftragten Herrn Gerling für die Einrichtungen der MPG frei zugänglich geworden, aber nicht alle kennen die optimalsten Einstellungen, um diese Software auch voll zu nutzen. Bereits das automatische Update mancher Virenscanner bereitet immer wieder Probleme. Der Auslieferungszustand von Browsern und Mailclients ist i.d.R. immer einem GAU gleich, dabei könnte ein zentral angesiedeltes Team für die gängigsten Produkte gleich nach deren Erscheinen die besten (sichersten) Initialisierungen ermitteln und allen zugänglich machen.

Das KIT sollte die wichtigsten und wirkungsvollsten Produkte im Bereich Sicherheit sichten, prüfen und mit ihren Konfigurationseinstellungen zentral bereitstellen und ggfs. Warnmeldungen aus der Presse kompetent beurteilen und so HOAX-Meldungen in der MPG vermeiden helfen.

Störfälle können hier dem *„Information-Security-Manager“* gemeldet und an die entsprechenden Stellen weitergeleitet werden. Im Extremfall soll eine Einsatzgruppe gebildet werden, die vor Ort Hilfestellung geben kann. Der Aufbau eines *IT-Security Helpdesks* (z.B. wie funktioniert VPN?, wie stelle ich Virenscanner optimal ein?, wer hat Filterregeln für meinen Router?) sollte ein weiterer Schritt sein, aber auch Notfallpläne, Benutzerordnungen u.v.m. gehören in den Verantwortungsbereich des o.g. Managers in enger Zusammenarbeit mit dem Datenschutzbeauftragten und dem Betriebsrat und vor allem der MPG-Leitung.

Umso mehr man zentral an einer Stelle abrufen kann, umso leichter läßt sich Sicherheit in den Instituten auch implementieren.

4.10 Einkauf, e-Commerce, eProcurement, SAP und IT-Management

Dieses Themengebiet soll im Gegensatz zu den vorangegangenen aus hoffentlich ersichtlichen Gründen etwas ausführlicher betrachtet werden, obgleich es auf den ersten Blick mit der Leistungsfähigkeit der IT in einem wissenschaftlichen Sachverhalt nichts zu tun haben scheint. Jedoch haben Fehlplanungen in diesem Bereich unmittelbar oder mittelfristig Auswirkungen auf die wissenschaftliche Leistungsfähigkeit bzw. Ausstattung in einem Institut.

Wie mehrfach erwähnt, muß heutzutage ein Institut über ein *IT-Management* verfügen, daß breit und weitsichtig denkt und plant. Das ist mit dem aktuellen Personalbestand in den meisten MPI’s sicher nicht immer möglich.

Solch ein IT-Management muß in laufende wissenschaftliche, technische und verwaltungstechnische Projekte involviert sein, um rechtzeitig die IKT-Voraussetzungen zu schaffen, daß solch ein Projekt erfolgreich wird.

Andererseits sind es oft die alltäglichen Dinge, die die IT-Leitung von ihrer Arbeit stark abhalten. Sie müssen sich um den Einkauf, die Inventarisierung, Wartung und Pflege, Lizensierung und personelle Angelegenheiten kümmern, die sie i.d.R. nie gelernt haben, weil sie weder Betriebs- noch Volkswirt sind.

Darum wäre es sehr hilfreich für ihn, wo immer es möglich ist, auf Standards zurückzugreifen, die erprobt und praktikabel sind.

An einem Beispiel sei dies kurz demonstriert:

Es ist im Hause die Anschaffung einer bestimmten Hard- oder Software geplant. Der Bedarf muß mit den Anwendern genau ermittelt werden und sollte in die bestehende IT-Landschaft auch passen. Nun muß das Produkt ggfs. ausgeschrieben werden und danach eine Bewertung der vorliegenden Angebote vorgenommen werden. Bei der Bestellung sind gesetzliche und interne Regeln zu beachten. Ist die Ware angekommen, ist sie ordentlich zu inventarisieren und in Betrieb zu nehmen. Je nach Art der Ware sind ein Wartungsvertrag zu schliessen und Lizenzen zu registrieren. Im Falle eines Defektes sind Unterlagen über die Art und den Umfang der ursprünglichen Lieferung und Garantielaufzeit bereitzuhalten. Kleinere Defekte sind selbständig zu beheben oder mit dem Kundendienst abzuwickeln. Nach Alterung der Ware ist diese wieder zu deinventarisieren und ggfs. noch brauchbare Hard- oder Software anderweitig einzusetzen.

Das sind eine Menge Arbeitsschritte, die u.U. bei jedem Teil, das man beschafft, beachtet werden müssen.

Aus eigener Erfahrung wäre es gut, wenn man bereits vom ersten Schritt an, Arbeitsvorgänge rationalisieren und damit beschleunigen könnte. Z.B. durch webbasierende Formulare, die die Benutzer ausfüllen, die einen Bedarf melden. Kontextsensitive Formulare helfen dabei, daß ein Minimum an Informationen immer vorhanden ist. Bei Überschreiten der gesetzlichen Grenzen für nationale oder EU-weite Ausschreibungen sollten dann bedarfsgerechte Muster aus anderen MPG-Ausschreibungen vorliegen, die der IT-Verantwortliche leicht für die aktuelle Bestellung anpassen und auf den Weg bringen kann. Bei einer einfachen Bestellung kann er selber eine elektronische Bedarfsmeldung ausfüllen (die dann auch jeder lesen kann und bei der Buchhaltung nicht nochmals eingetippt, sondern nur noch überprüft und gegengezeichnet werden muß), die er sich ausdruckt, unterschreibt und weiterleitet. Anfragen bei Händlern sollten mit Standardvorlagen per Email oder Fax erfolgen, die so wichtige Kleinigkeiten wie ausführliche Beschreibungen des Produktes, Garantie, Lieferfristen und richtige Preiskennzeichnung verlangen, ohne die ein Angebot nicht mehr akzeptiert wird. Vergleichsangebote lassen sich dann auch wirklich vergleichen. Evtl. führt das dazu, daß man sich auf einige wenige, aber zuverlässige Händler konzentriert. Die Inventarisierung kann über webbasierende Datenbanken erfolgen, die Techniker, wie Administratoren und IT-Leiter gleichermaßen einsehen und ggfs. bearbeiten können. Eine jährliche Inventur ist dann schnell erstellt und mit geeigneten Facility-Management-Tools kann man schnell herausfinden, wer im Hause welche Software auf welcher Architektur in welchem Zimmer ein-

setzt und wo die gerade wieder einmal einkommende Werbepost am besten gebraucht werden kann.

Für die Benutzer könnten Checklisten erstellt werden, die ihnen dabei helfen, ihre Produkte nicht wahllos billig um die Ecke zu kaufen, sondern langfristig gut angelegt und in ein institutsinternes IT-Konzept passend zu erwerben, so daß sie auch damit rechnen können, daß sie Support von der örtlichen IT-Abteilung dafür erhalten.

In manchen Fällen lohnen sich angepaßte Wartungs-, Miet- oder Leasingmodelle, um immer eine IT-Ausstattung auf dem aktuellsten Stand zu haben, die auch bezahlbar bleibt. Im Bereich der Druckerwartung und des PC-Rollouts, die in zwei weiteren Abteilungen des KIT bearbeitet werden, sind evtl. langfristig größere Kosteneinsparungen und technisch anspruchsvolle Geräte möglich.

Das KIT könnte solche unternehmensinternen Abläufe überprüfen und allgemein adaptierbare Lösungen erstellen, die allen Beteiligten mehr Zeit für wissenschaftliches Arbeiten erlauben.

Als Partner ist ganz besonders die GV und das IKT zu sehen, aber auch die Arbeitsgruppe SAP am KIT. Da u.a. die Kosten für Verbrauchsmaterialien in der EDV nicht unerheblich sind und bei möglichst homogener Ausstattung der einzelnen MPI's in Punkto Drucker bspw. oftmals ähnliche Produkte eingekauft werden, ist auch angestrebt, auf einer Webseite den Instituten neben den aktuellen Rahmen- und Abrufverträgen die tagesaktuellen Preise für Verbrauchsmaterialien wie Druckerzubehör (Toner, Tinte, Papier, Folien, ...) als auch Medien (Bänder, Disketten, CD's, ...) von einer Auswahl von Firmen zu liefern. Somit können die Bestellungen aufgrund dieser Preise umgehend erfolgen, da die Vergleichsangebote gleich mitgeliefert werden. Auch Preise für die typischen Zubehörteile, wie Speicher, Festplatten, Monitore, ... können dort tagesaktuell (oder wochenaktuell) zusammengestellt werden. Dabei achtet das KIT bereits darauf, daß gute und zudem preiswerte Handelspartner auf diesen Listen erscheinen und die angebotenen Gerätschaften hochwertig sind. Bestellungen können ggfs. auch über ein Webinterface erfolgen, das sollte aber alles noch im Detail diskutiert werden.

Grundsätzlich ist dieser Kompetenz-Bereich für den gesamten täglichen Ablauf in einem Institut hinsichtlich IKT-Verwaltung zuständig und soll allgemeine, wie individuelle Module, Formulare und Datenbanken entwickeln, die allen kostenfrei zur Verfügung gestellt werden. Alle notwendigen Verfahren hinsichtlich Ausschreibungen werden hier im Intranet ausführlich beschrieben, so daß die GV letztlich nur noch die Anträge selber zu kontrollieren hat.

Erhofft wird aber auch, daß sich die Abläufe an einem MPI vereinfachen, die Einkäufe mehr zentralisieren lassen und damit evtl. wirtschaftlichen Wildwüchsen (jeder kauft, wie er will) entgegengearbeitet werden kann.

Nachrichtlich: Seit Mitte letzten Jahres hat die MPG das Projekt *eProcurement* gestartet, das unterstützt durch ein externes Consulting-Unternehmen einen Großteil der hier angesprochenen Defizite und Vorschläge bestätigt hat und nun gemeinsam mit GV, IKT, Betriebsräten und MPG-Leitung entsprechende Lösungsmodelle erarbeitet. Eine intensivere Zusammenarbeit mit den IT-Abteilungen der Institute bei der Erstellung der u.a. o.g. Einkaufsmöglichkeiten per Warenkorb wäre wünschenswert.

Das Kompetenzzentrum könnte ferner bei Abschluß von *Rahmenverträgen* mitwirken, indem es a) die Wünsche der Anwender sammelt und artikuliert und b) technischen Input bei der Auswahl der Anbieter der GV anbietet, so daß ergänzt um den immer notwendiger werdenden Rechtsbeistand der GV bei Vertragsabschlüssen, Rahmenverträge optimiert und ein hoher Qualitätsstandard der IT-Produkte gewährleistet wird.

Zuletzt muß der Einsatz von *SAP* in Kooperation mit Herstellern, GV, Anwendern und örtlicher IT-Abteilung verbessert werden. Das Kompetenzzentrum sollte um die optimalste Installation hinsichtlich Datenschutz, Sicherheit und Funktionalität (bessere Anwendermasken) informiert sein, Auskunft geben und ggfs. Einfluß nehmen können.

Es empfiehlt sich, sofern nicht in der GV vorhanden, eine(n) Fachmann/frau für diesen Bereich zu haben bzw. darin ausbilden zu lassen, um die Interna dieser Standardsoftware besser beleuchten zu können. Das KIT soll hier Kompetenz aufbauen, die der GV, den Betriebsräten, dem Datenschutzbeauftragten, den Verwaltungsmitarbeitern und den IT-Leitern helfen kann, dieses System besser und sicherer als bisher betreiben zu können. Insbesondere technisches Know-How gegenüber dem Hersteller und seinem beauftragten Systemhaus entgegenzubringen, um sinnvolle Hard- und Softwarelösungen MPG weit aufzubauen.

4.11 Dokumentation (Intranet) und Vermittlung (Schulung)

Ein ganz wesentlicher Punkt für die Effizienz des KIT innerhalb der MPG wird die Informationspolitik sein, mit der gewonnene Erfahrung in die Institute gebracht werden kann.

Hier soll an erster Stelle ein MPG-IT-Intranet aufgebaut werden, das die verschiedenen Ansätze der GV (IKT), des DSB, des Virtuellen Instituts (Buss-

mann) und anderer zusammenfassen soll. Der Zugriff sollte über IP- und Passwort-Authentifizierung laufen, solange keine andere Methode gewünscht wird, damit diese doch teilweise streng vertraulichen Informationen den richtigen Partner erreichen.

Alle Arbeitsgruppen (des Kompetenzzentrums) werden angehalten sein, sämtliche abgeschlossenen, laufenden und geplanten Projekte zu dokumentieren und transparent im Intranet einsichtig zu machen. Neben dem Intranet werden KIT-intern durch Seminare und Arbeitstreffen Informationen zwischen den Gruppen ausgetauscht, an Auszubildende, Diplomanden und Doktoranden weitergegeben, aber letztlich in kleinen MPG-Workshops zu den einzelnen Themen den Einrichtungen der MPG vorgestellt (dies gilt alles nur dann, wenn die Aufgaben des Kompetenzzentrums in einer Hand liegen).

Das alljährliche DV-Treffen soll sowohl einen aktuellen Überblick über die Arbeiten des KIT erhalten, aber auch möglicherweise neue, wegweisende und alle betreffende Themen bestimmen, die in die Arbeit des KIT mit aufgenommen werden sollen.

Nicht geplant sind Randgebiete, wie Höchstleistungsrechnen oder PalmPilot-Programmierung bspw., da sie nicht auf alle Institute anwendbar und oft nicht den Alltagsbetrieb betreffen. Außerdem stehen dazu genügend Experten in den großen Rechenzentren, aber auch in der MPG-INFO zur Verfügung, die hier adhoc Fragen beantworten können.

Alle Arbeiten des KIT sollen für alle Institute zugänglich sein, das heißt Lösungen können jederzeit, sofern es personell und technisch machbar ist, abgerufen werden, um lokal installiert zu werden. Das können PIX-Firewall-Konfigurationen, Apache-Webserver-Einstellungen, Eudora-Mail-Optionen oder Formblätter zur Bestellung von PC's sein. Ein Helpdesk mit FAQs und Foren zur Diskussion von Themen soll eingerichtet werden.

Links zu anderen Lösungen in der MPG und bei Herstellern sollen die Informationslücken in den genannten Gebieten schließen helfen. Der Besuch von Messen, Kursen und das Sichten der üblichen Fachzeitschriften auf diesem Gebiet soll zu einem brauchbaren Extrakt für die MPI-DV's verarbeitet werden.

Partner für das Webdesign und die inhaltliche Gestaltung sind zwar im Gespräch, aber zunächst muß der Informationsgehalt über dem Layout stehen, um schnell zur Lösung seiner Fragen zu kommen.

Nachrichtlich: Nach dem ersten Treffen eines kleinen Arbeitskreises (Jan 2002) aufgrund meines Vortrages zum KIT beim DV-Treffen (Nov 2001) ergab sich ein erstes Testportal bei der GWDG in Verantwortlich-

keit von Herrn Zite-Ferenczy (u.a. Protokollführer des BAR). Sobald hier die Betaphase abgeschlossen ist, wird die entsprechende URL über mpg-info bekanntgegeben.

4.12 Ausbildung

Ein Kompetenzzentrum wie das hier vorgestellte dürfte eine ideale Plattform für die Ausbildung von Fachinformatikern bzw. Diplomanden und Doktoranden der Informatik darstellen, da es viele Gelegenheiten bietet, auf Kernthemen der IT zu hospitieren.

Gerade im Falle der Fachinformatiker, einem recht jungen Ausbildungsweg, könnten auch Musterlehrpläne erstellt werden, die den ebenfalls ausbildenden Instituten als Hilfestellung an die Hand gegeben werden könnten. Zumindest wäre der Austausch mit diesen Instituten hier leicht zu koordinieren.

Grundsätzlich darf die Zukunft des IT-Personals nicht vernachlässigt werden, evtl. bildet das Kompetenzzentrum sogar zukünftige Mitarbeiter für die Institute (mit) aus.

Fazit:

- Homogenisieren wo immer machbar und sinnvoll
- Durch Bündelung von Aktivitäten Kosten einsparen bei man power und Anschaffungen
- Effizienz steigern, durch Abruf fertiger Lösungen, Standards, Wissensbasen

Generell kann man zur Umsetzung des vorgestellten Modells zwei Ansätze wählen:

- ein zentrales Kompetenzzentrum

  Vorteile: zielgerichtetes Erarbeiten von Lösungen, die auf die meisten Institute zutreffen. Ein Ansprechpartner für Interne und Externe. Zeit für Lösungen.

  Nachteile: erhöhter Man-power Bedarf gegenüber jetzigen Lösungen. Der Praxisbezug darf nicht verlorengehen.

- verteilte Kompetenzzentren in den einzelnen Instituten

  Vorteile: evtl. praxisnähere Lösungen in Pilotprojekten. Man power muß

nicht unbedingt erhöht werden.

Nachteile: Lösungen evtl. nicht adaptierbar für andere Institute. Kaum Zeit für ein Angebot an Dokumentation und Implementation an andere Institute.

Persönlich würde ich dem BAR empfehlen:

- entweder die GWDG personell zu verstärken, um die bisherigen Kernthemen zu erweitern und besser angehen zu können.
- oder die übrigen Themen an Gruppen von MPI's zu verteilen, die dafür die Verantwortlichkeit übernehmen möchten und eine finanzielle Unterstützung bei Pilotprojekten (auch personeller Art) zu geben.
- oder aber einer neuen Einrichtung für 2-3 Jahre die Gelegenheit geben, unabhängig vom üblichen Geschäftsverkehr all die genannten Punkte konsequent anzugehen und den Nutzen am Ende praktisch wie wirtschaftlich unter Beweis zu stellen. Hier ist dann Anschubfinanzierung seitens der MPG notwendig.

Eine absolute Notwendigkeit muß das gemeinsame Wollen einer Lösung sein, die neue Wege beschreitet, neue Möglichkeiten eröffnet und Unterstützung auf allen Ebenen der MPG findet.

Das Ziel ist klar: eine Leistungs- und Qualitätssteigerung der wissenschaftlichen Arbeit in der MPG durch professionellere IT-Unterstützung, so daß die Investition in die Basis IT bzw. IKT neue Dimensionen aufstößt.

Bemerkung: Auf die Nennung von Firmennamen wurde in dieser Zusammenfassung bewußt verzichtet, obgleich sich einige Dutzend namhafte Hersteller, Softwarehäuser oder Distributoren an einer Zusammenarbeit mit einem zentralen Kompetenzzentrum sehr interessiert gezeigt und teilweise die technische Ausstattung in Aussicht gestellt haben.

Anfragen oder Anregungen können gerne an den Autor (ao@mpifr-bonn.mpg.de) bzw. an den MPG-Arbeitskreis kit_workshop@mpiv-hd.mpg.de weitergeleitet werden. Dieser Arbeitskreis hat sich aufgrund einer Einladung „zur Mitarbeit an möglichen Kooperationen im Sektor IKT der MPIs" in der mpg-info-Mailingliste Mitte Januar 2002 zusammengefunden und diskutiert Probleme und Lösungsvorschläge, die seit meinem Vortrag auf dem 18. DV-Treffen 2001 an uns herangetragen wurden.

Linux auf einem Mainframe IBM S390

Dirk von Suchodoletz

Mathematisches Institut der Universität Göttingen

Vorwort

Im Herbst vergangenen Jahres wurden im Rahmen eines gemeinsamen kleinen Forschungsprojektes zwischen IBM und der GWDG die Einsatzmöglichkeiten von Linux auf dem Mainframe IBM S390 getestet. In den GWDG-Nachrichten 8/2001 wurde zu Beginn des Tests bereits kurz darüber berichtet, insbesondere wurde die für den Test zur Verfügung gestellte Hardware genauer beschrieben. Der vorliegende Artikel liefert nun einen ausführlichen Abschluss- bzw. Erfahrungsbericht zu diesem Projekt und basiert auf einem Vortrag [1], der auf dem 18. DV-Treffen der Max-Planck-Institute im November 2001 in Göttingen gehalten wurde.

1. Überblick

Es ist schon etwas ungewöhnlich, sich Linux auf einer S390 anzusehen. Das PC-UNIX, welches vor zehn Jahren seine Geschichte auf preisgünstigen, auch für Studenten finanzierbaren Rechnern begann und als Betriebssystem inzwischen auf sehr viele verschiedene Plattformen portiert wurde, ist nun auch für die Aufnahme in die „Profi-Liga“ für wert und wichtig befunden

worden, in der schon kürzeste Systemausfälle Millionenschäden anrichten können.

Hardware aus einem Geschäftsbereich der Firma IBM, welcher sich an Großkunden wie Rechenzentren, Banken oder Telekommunikationsunternehmen wendet und in dem die Anschaffung bereits sechs- bis siebenstellige Summen erfordert, bekommt man als Student bzw. Endanwender eher selten zu fühlen bzw. zu sehen. IBM hatte in Zusammenarbeit mit dem damaligen Geschäftsführer der GWDG, Herrn Prof. Schneider, die Teststellung eines IBM-Mainframes für drei Monate vereinbart, um den Einsatz im Umfeld eines wissenschaftlichen Rechenzentrums zu evaluieren. Am Projekt beteiligt waren neben mir (wissenschaftliche Hilfskraft am Mathematischen Institut der Universität Göttingen) noch Alexander Eickhoff (Entwickler in der Automatisierungs-Technik im „Measurement Valley“ in Göttingen), Bernhard Kaindl (S390-Linux-Entwickler bei SuSE), Eberhard Mönkeberg und Manfred Röhrig (Programmierer bei der GWDG), Stefan Teusch (Leiter der dezentralen Netze beim Studentenwerk Göttingen) und weitere.

Im Folgenden soll es deshalb nach einem kurzen geschichtlichen Abriss um die Darstellung der Erfahrungen und Eindrücke unserer Arbeitsgruppe gehen, wie Linux auf einer solchen Maschine aussieht, wie es als Instanz des Host-Betriebssystems VM (Virtual Machine) eingerichtet wird und welche Möglichkeiten und Besonderheiten es gibt. Zum Ende dieses Berichts werden einige Einsatzmöglichkeiten und Einschätzungen einer S390 dargestellt, wie sie sich aus unserer Sicht ergeben könnten.

2. Geschichte Linux S390

Die Geschichte von Linux auf S390-Hardware reicht noch nicht lange zurück und beginnt im Jahre 1999 mit der Portierung von Linux auf diese Plattform durch das Marist College in Poughkeepsie (NY), das IBM-Labor in Böblingen und einige andere. Die erste verfügbare Distribution war die des Marist College (siehe hierzu [9]), wobei inzwischen eine komplette SuSE- und Turbo-Linux-Distribution für diese Plattform vorliegt. Die im Folgenden geschilderten Experimente und Erfahrungen beziehen sich dabei auf das SuSE-Linux. Zum einen gibt es eine enge Zusammenarbeit von IBM und diesem Linux-Anbieter im Bereich der S390, zum anderen liegen bei der GWDG und der Universität Göttingen langjährige Erfahrungen mit der SuSE-PC-Linux-Distribution vor. So ließ sich ein sehr guter Vergleich zwischen Linux auf den beiden sehr unterschiedlichen Architekturen ziehen.

3. Das Setup der S390

Es bestehen grob drei Optionen, Linux auf einer solchen Maschine einzurichten. Das üblicherweise im Zusammenhang mit Linux verwendete Basisbetriebssystem VM erlaubt es, als Gastnutzer andere Betriebssysteme als einzelne Prozesse ablaufen zu lassen, wobei diese Gäste wiederum VM-Instanzen sein können, die virtuelle Rechner in Form anderer OS zulassen. Diese Software-Abstraktion erlaubt es, sehr viele verschiedene Setups anzulegen und komplette Entwicklungsumgebungen auf einer einzigen Maschine zu schaffen.

Ein klassisches Gastbetriebssystem ist das Conversational Monitoring System (CMS), welches eine Reihe von Standardapplikationen wie die interpretierte Skriptsprache REXX und den Editor XEDIT zur Verfügung stellt. OS390 ist ein UNIX-Derivat für die S390-Plattform und kann genau wie CMS als VM-Gast in vielfacher Ausführung auf der Maschine laufen.

Neben diesen Varianten bietet sich die Option an, „logische Partitionen“ einzurichten, wobei Teile der Hardware an einzelne Partitionen dediziert werden. Dieses stellt zwar zum einen sicher, dass immer eine bestimmte Rechenpower in einer Partition zur Verfügung steht, verhindert aber zum anderen ein übergreifendes Ressourcenmanagement, worin eine der Stärken von VM auf S390 liegt.

Als weitere Option bietet sich an, Linux allein auf einer solchen Maschine zu installieren. Eine nähere Beschreibung dieser Setup-Varianten findet sich in [2] und im IBM-Redbook [3] zum Thema, so dass ich hier nicht ausführlicher darauf eingeben möchte, sondern mich gleich auf die im beschriebenen Fall interessanteste Lösung beziehen möchte.

Im vorgestellten Projekt haben wir uns für eine Reihe von Linux-VM-Benutzern in einer VM-Umgebung entschieden, da wir in der Zusammenarbeit von Linux und VM die größten Vorteile der Architektur sahen und dieses eine der Besonderheiten von Linux/S390 gegenüber klassischen Installationen ausmacht.

Dieses Setup kann man mit dem Betreiben einer Linux-Instanz unter VMware vergleichen (ein Teil der VMware-Entwickler stammt von IBM), wobei weit mehr Instanzen nebeneinander existieren können, da die Maschine über ein sehr intelligentes Ressourcenmanagement verfügt. Der Vergleich von VM mit seinem PC-Pendant kann zum Verständnis der Zusammenarbeit der beiden Betriebssysteme beitragen.

4. Eingesetzte Hardware

Die beschriebene Maschine ist eine zSeries, eine S390 der sechsten Generation, eine 31bit-Architektur mit drei von zwölf für Betriebssysteme freigeschalteten Prozessoren. Maximal können 16 Prozessoren auf dem Board untergebracht werden. Eingebaut waren weiterhin 2 GByte Hauptspeicher und 6 GByte Erweiterungsspeicher, 16 ESCON-Kanäle mit je 17 MByte Bandbreite zum Anschluss der Festspeichereinheit. Der Festplattenstapel selbst war ein Shark Enterprise Storage Server (ESS) in einem eigenen Schrank. Dieser war in seinem Minimalausbau mit 420 GByte Speicherkapazität bestückt.

Wo nun schon PCs mit 2 GByte und als Dual-Prozessormaschinen verkauft werden, klingen die Werte vielleicht nicht besonders aufregend, jedoch sollte man sich vergegenwärtigen, dass alle Hardware komplett redundant ausgelegt ist, angefangen von den zwei mit 24 A abgesicherten Drehstromanschlüssen und entsprechenden Netzteilen über die flüssigkeitsgekühlten Prozessoren und Speicherblöcke bis zu den beiden als Konsole zur Hardware-Steuerung eingebauten PIII-Thinkpads mit OS2/Warp.

Die Testmaschine verfügte über zwei Ethernet-Interfaces: ein Fast-Ethernet und ein Gigabit-OSA-Adapter, welche für den Testbetrieb mit unterschiedlich gerouteten IP-Netzen verbunden waren.

Die Maschine selbst (ohne die Festplatten-Speicherblöcke) ist in einem 19"-Schrank untergebracht und füllt diesen bis auf einige Erweiterungssteckplätze ziemlich aus.

Abb. 1: IBM S390 im Maschinenraum der GWDG

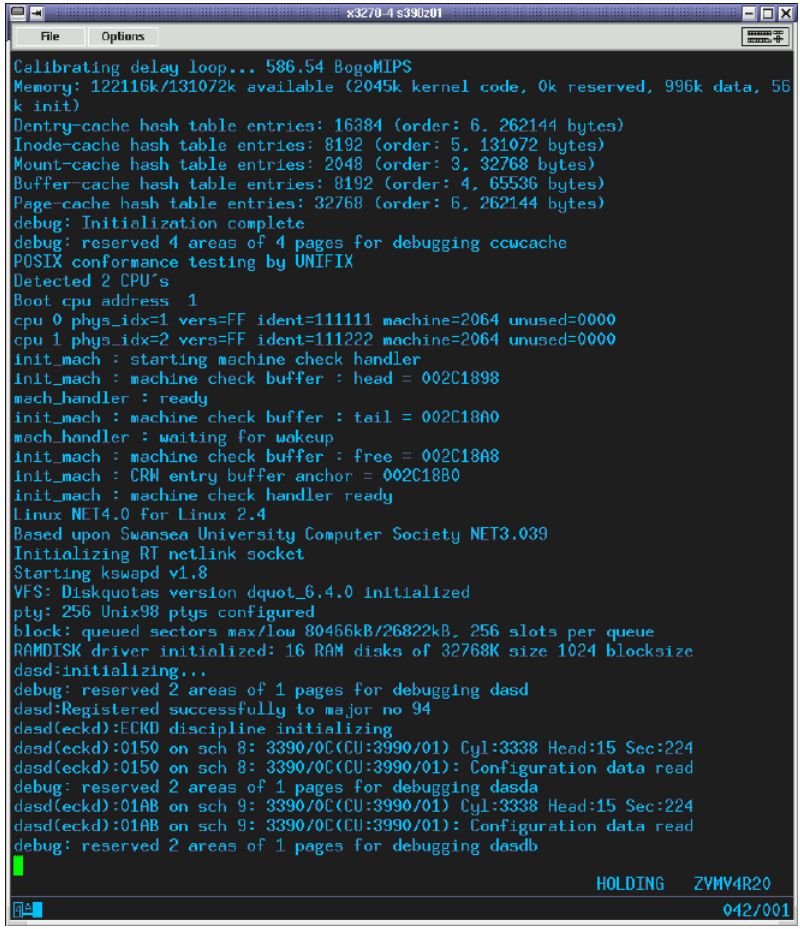
5. Einrichten der VM

Zum Glück oblag es uns nicht, die Maschine selbst von Grund auf frisch einzurichten, sondern es genügte zuzusehen, wie eine vorbereitete VM-Installation per Band (was schon ein größerer Aufwand war, da das Bandlaufwerk nicht direkt an einem der ESCON-Kanäle, sondern mittels Workstation an die S390 angeschlossen wurde) eingespielt wurde. Nach der grundlegenden Einrichtung der Hardware, in erster Linie der Zuordnung der Speichereinheit, einer Sharc, ging es dann an die Einrichtung bzw. Konfiguration einiger wichtiger UserIDs. Deshalb benötigt auch eine Linux-Installation auf einer S390 zumindest eine Person, welche einige Hintergründe des VM-Betriebssystems kennt. Das betrifft besonders spätere Optimierungen des Betriebs und die Feinabstimmung zwischen verschiedenen Konfigurationsoptionen wie CPU-Dedizierung, Minidisk-Cache-Optimierung und Festplattenaufteilungen. Einführungen und Erklärungen zum Betriebssystem VM findet man z. B. in GWDG-Schulungsunterlagen [4] oder auch in der Literatur [5].

Die meisten Administrationsaufgaben werden mittels des X3270-Telnet-Emulators erledigt, welcher eine klassische Textkonsole einer solchen Maschine inkl. einer ganzen Reihe von Spezialtasten über ein angehängtes Keyboard auf dem Bildschirm darstellt. Dieses Tool gab im vorgestellten Setup jedem Linux-Nutzer die Möglichkeit, an die Konsole seiner Maschine zu kommen, was notwendig wurde, wenn das System neu installiert, beobachtet oder neu gestartet werden sollte oder es Probleme mit dem Netzwerk-

Setup des einzelnen Gast-Linux gab. Die speziellen Funktionstasten, welche Einigen sicherlich nicht von den IBM-Terminals bekannt sein dürften, können über ein Extra-Fenster mit einer „Spezialtastatur“ mittels des X3270-Telnet-Emulators dargestellt werden. Hierzu gehört z. B. die zum „Blättern“ bekannte „PF“-Taste oder auch der „Reset“-Key, wenn eine Eingabe innerhalb der Konsole aus dem gültigen Bereich herausfiel.

Abb. 2: Screenshot eines typischen Startvorgangs einer Linux-Maschine auf der S390-Architektur (links oben ein kleines Icon, welches eine erweiterte Tastatur für die benötigten Sondertasten darstellen kann)



Der Benutzer „maint“ ist der Systemadministrator mit den notwendigen Rechten zur Erstellung neuer UserIDs und Zuordnung der Maschinenressourcen. Dieses geschieht in der so genannten Directory-Datei, der Hauptkonfigurationsdatei der VM. Neben diesem übernimmt der User „tcpmaint“ die Konfiguration des TCP/IP-Stacks der VM-Maschine, was bedeutet, dass er die Konfigurationsdateien der TCP-Benutzer in deren Home-Minidisk modifizieren darf. Das TCP/IP-Netzwerk läuft unter einer eigenständigen

UserID in der VM, wie auch jeder Dienst basierend auf TCP/IP eine eigene UserID beansprucht (welche aber hier nicht konfiguriert werden brauchten, da Linux diese Aufgaben übernehmen sollte). Da in der gestellten Maschine zwei Ethernet-Interfaces zur Verfügung standen, wurde die Aufgabe auf zwei VM-User aufgeteilt, hier „tcpipf“ und „tcpipg“, um eine bessere Redundanz beim Neustart des TCP/IP zu erreichen.

Ein weiterer zentraler VM-User ist „fconx“, welcher das gleichnamige Monitoring-Tool verwaltet, mit dem sich die Auslastung und Performance der S390 bis ins tiefste Detail analysieren ließ. In der aktuellsten Fassung dieses Tools gibt es bereits Schnittstellen zur Analyse einzelner Linux-Maschinen. Neben dem reinen Monitoring lassen sich mit diesem Tool weitere System-Tasks wie das Aufräumen des Spoolbereichs erledigen. Dieser kommt z. B. beim später beschriebenen Installations-Booten der Linux-User zum Einsatz und verhindert bei Verstopfung das Hochfahren. Darüber hinaus lassen sich die Meldungen zur Fehleranalyse heranziehen. Dabei kommt es natürlich auf das globale Setup der Maschine an, welche weiteren Tasks noch anfallen könnten.

Da die Linux-User als Gäste im VM neu angelegt werden müssen (die eben beschriebenen Accounts werden/müssen in den meisten Fällen bereits existieren), will ich im Folgenden näher darauf eingehen. Am besten lässt sich die Einrichtung eines Benutzers in der Systemdatei (auf der zentralen Minidisk des Benutzers „maint“) an einem Beispiel erläutern. Die Zahlen vor den Einträgen geben nur die Zeilennummer wieder, so wie man sie auch bei der Benutzung des „xedit“ zu sehen bekommt. Groß- und Kleinschreibung spielt keine Rolle.

```
000 USER VMLINUX1 PASSWORD 64M 256M G
001 MACHINE ESA 4
002 OPTION QUICKDSP LNKNOPAS
003 CONSOLE 0009 3215
004 SPOOL 000C 2540 READER *
005 SPOOL 000D 2540 PUNCH A
006 SPOOL 000E 1403 A
007 LINK MAINT 0190 0190 RR
008 LINK MAINT 019D 019D RR
009 LINK MAINT 019E 019E RR
010 LINK MAINT 019F 019F RR
011 LINK TCPMAINT 0192 0592 RR
012 MDISK 0191 3390 5201 0050 VM411A MR READ WRITE MULTIPLE
013 MDISK 0401 3390 0001 2000 VMLINUX1 MR READ WRITE MULTIPLE
014 MDISK 0402 3390 0001 2000 VMLINUX1 MR READ WRITE MULTIPLE
015 MDISK 0403 3390 0001 1000 VMLINUX1 MR READ WRITE MULTIPLE
016 SPECIAL 1010 CTCA VMLINUX
017 SPECIAL 1011 CTCA VMLINUX2
018 DEDICATE 580A 580A
019 DEDICATE 580B 580B
020 DEDICATE 580C 580C
021 IPL 200
```

In Zeile 000 wird der Benutzer (VMLINUX1) mit dem Keyword „USER“ definiert und ein Klartextpaßwort vergeben. Dahinter wird die Menge des (virtuellen) Speichers angegeben, die ihm default-mäßig und maximal zur Verfügung steht, und zum Schluss der Benutzerlevel (G) festgelegt. Dieses ist der niedrigste User-Level, genügt aber für die Bedürfnisse des Gastsystems. Der höchste Level (A) ist dem Systemadministrator zugeordnet, dazwischen gibt es weitere Abstufungen, welche im Regelbetrieb interessant sind. Linux-Maschinen können maximal 1920 MByte ansprechen, mindestens 12 MByte werden zum Betrieb eines Minimalsystems benötigt.

Die anschließende Zeile 001 legt den Prozessortyp (ESA: simuliert die ESA/370- bzw. ESA/390-Architektur) und die Anzahl der CPUs (dieses sind keine realen Prozessoren) fest. Das Maximum für einen Linux-Gast liegt bei vier, generell bei 64 CPUs, der Default-Wert bei einem Prozessor. Die Option QUICKDSP in Zeile 002 stellt sicher, dass die virtuelle Maschine sofort und ohne Verzögerung in die Bearbeitungsliste von VM aufgenommen wird, wenn etwas zu tun ist. LNKNOPAS erlaubt das Linken von System-Minidisks, ohne ein Paßwort angeben zu müssen. Es ist auch möglich, einem Nutzer eine CPU zu dedizieren, welche dann aber ausschließlich von diesem verwendet werden kann.

Die nächsten Zeilen (003 - 006) finden sich bei den meisten VM-Nutzern; sie legen den verwendeten Konsolentyp (3215/3270) und die unterschiedlichen Spoolgeräte (virtueller Lochkartenleser, -stanzer und Drucker) fest.

Im Folgenden werden Links auf die VM-CMS-Platte (007) und auf CMS-Erweiterungen (008 - 010) sowie den Benutzer „tcpmaint“ angelegt, um Standardapplikationen und Netzwerkprogramme (ping, ftp ...) aufrufen zu können. Die nächsten Zeilen dedizieren dem Benutzer verschiedene Minidisks, wobei die erste Zeile quasi das „Homeverzeichnis“ (0191 ist immer die Minidisk des entsprechenden Nutzers, im Beispiel vom Disktyp 3390 mit Startzylinder 5201 und 50 Blöcken) definiert. Die Größe wird in Blöcken angegeben. Eine virtuelle Platte vom Typ 3390 hat ca. 9 GByte Kapazität und ist die größte derzeit unter VM verwaltbare Einheit. Es gibt 10017 Blöcke insgesamt, wobei einer für das DiskLabel reserviert ist. (Ein Block entspricht also 850 KByte.) Die weiteren Minidisks werden unsere Linux-Partitionen für das Root-Filesystem, die Home-Partition und Swap werden. Hinter den genannten Daten steht der Zugriffsmodus für die jeweilige Minidisk, wobei MR für Multiple-Write-Access steht. Am Ende der jeweiligen Zeile werden die Paßwörter für den Lese- (Paßwort READ), Schreib- (WRITE) und gleichzeitigen Zugriff durch mehrere User (MULTIPLE) festgelegt.

Die Zeilen 016 - 017 definieren die virtuellen Channel-to-Channel-Verbindungen (eine Art Point-to-Point-Verbindung, welche durch den Command Processor (CP) von VM komplett in Software realisiert werden und die Daten mittels Speicherpuffer austauschen), im Beispiel zu einem zweiten Gast-Linux (VMLINUX2). Auch die nächsten Zeilen (018 - 020) definieren eine Netzwerkverbindung, wobei hier der exklusive Zugriff auf eine Hardware-Ressource erlaubt wird. Der Gigabit-OSA-Adapter erlaubt es, bis zu acht verschiedenen Benutzern unabhängigen Zugriff zu gestatten (einer dieser Nutzer ist üblicherweise das VM). Dieses besondere Feature einiger S390-Hardware erlaubt damit einige interessante Netzwerk-Setups inkl. verschiedener Backup-Settings. Neben diesen Netzwerkschnittstellen wären noch IUCV-Devices und seit kurzem Hipersockets zu nennen. Letztere sollen die erreichbaren Bandbreiten gegenüber CTC und IUCV erhöhen und die Latenzzeit nochmals senken. Erfahrungen konnten wir zum Zeitpunkt unseres Tests jedoch damit noch nicht sammeln.

Mit dem Befehl in Zeile 021 kann direkt ein Initial Program Load (IPL) angestoßen werden, was in diesem Fall von der Minidisk 200 erfolgt, die z. B. die Linux-Installation enthält. Auf diese Weise würde beim Login des Nutzers sofort das Gast-OS gestartet werden. In vielen Fällen findet man hier ein IPL CMS, womit das Conversational Monitoring System aufgerufen

wird. Der gleiche Effekt ließe sich auch über die PROFILE.EXEC auf der User-Minidisk erzielen, was den Vorteil hat, dass diese direkt vom jeweiligen Nutzer editiert werden kann. In dieser Datei kann z. B. auch die Kopplung der CTC-Interfaces erfolgen.

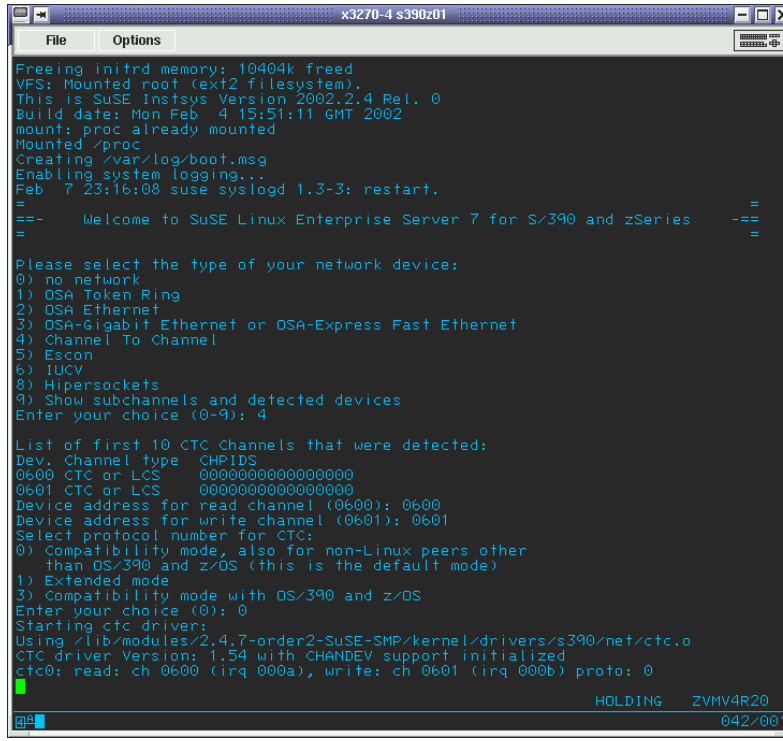
6. Installation von Linux

Um Linux auf einer S390 zu installieren, sind einige Arbeitsschritte notwendig, welche von den bekannten Verfahren doch etwas abweichen. Einen CD-Schlitz sucht man an dieser Maschine erstmal vergeblich, da es üblicherweise auch nicht vorgesehen ist, direkt auf die Hardware zugreifen zu können. Für die Einrichtung in einer logischen Partition soll es jedoch möglich sein, über ein CD-Laufwerk der Hardware-Konsole eine Installation vorzunehmen. Dieses auszuprobieren, war in dem gegebenen Setup jedoch nicht möglich.

Für die hier gewählte Maschineninstallation unter dem VM ist es deshalb notwendig, für die Einrichtung unter VM speziell angepasste Installations-Images auf die Maschine zu kopieren. Dieses kann z. B. mittels des VM-eigenen FTP-Kommandos geschehen, welches auf der Minidisk des „tcpmaint“ installiert ist. Die benötigten Images könnten in das jeweilige Homeverzeichnis des entsprechenden Linux-Users kopiert oder zentral zur Verfügung gestellt werden, da nur in einigen Fällen Anpassungen in der Parameterdatei notwendig sind.

Die SuSE-Linux-Distribution bringt drei Installationsdateien mit: ein spezieller Bootkernel, ein Ramdisk-Image und eine Parameterdatei. In der Parameterdatei können Command-Line-Options für die Übergabe an den Kernel, z. B. zu den zu verwendenden Platten, den Interfaces und eine optionale Ramdisk übergeben werden. Damit der Kernel und das Ramdisk-Image von der S390-Architektur verarbeitet werden können, müssen die Images zu einem IPL-fähigen Format zusammengefügt werden. Dieses geschieht durch das Senden der Dateien auf den virtuellen Lochkartenstanzer und das anschließende Einlesen durch den virtuellen Lochkartenleser. Dieser Vorgang, der ungefähr 130.000 virtuelle Lochkarten umfasst, erinnert an die Ursprünge der Architektur.

Abb. 3: Aussehen eines typischen Startbildschirms in einem X3270-Terminal zur Installation eines SuSE-Linux in einer Virtuellen Maschine



Wenn alles glatt ging, kann man nun die Kernel-Meldungen in seinem 3270-Terminal verfolgen, die aufgrund der etwas vom Gewohnten abweichenden Funktionalität dieses Terminals etwas verschoben aussehen. Es erscheint ein Login-Prompt, an dem man sich als „root“ anmeldet und sodann aufgefordert wird, Netzwerkeinstellungen vorzunehmen: Zuerst ist das gewünschte Netzwerk-Interface (CTC, IUCV, OSA-Adapter) aus einer Liste auszuwählen, welches dann mittels der eingegebenen Netzwerkparameter versucht wird zu initialisieren. Konnte auch dieser Vorgang erfolgreich abgeschlossen werden, steht einem Login mittels SSH über das Netz nichts mehr im Wege. Da man sich nun von jeder gewünschten Maschine einloggen kann, muss man sich nicht mehr mit den Beschränkungen des 3270-Protokolls auseinandersetzen, welches das Arbeiten mit interaktiven Programmen (Editoren, Pager, YaST, ...) erschwert.

Nach dem Login über das Netz kann nun „yast“ gestartet werden und die Maschine fast wie gewohnt konfiguriert werden. Als Installationsquellen kommen natürlich nur Netzwerkressourcen in Frage, welche mittels FTP oder NFS erreicht werden können. Ein weiterer, etwas ungewohnter Punkt ist das Einrichten der „Festplatten“, welches durch „yast“ jedoch stark erleichtert wird. Zuallererst sind die zur Verfügung stehenden DASDs (Direct Attached Storage Devices) lowlevel zu formatieren, welches beim direkten Ausführen des „dasdfmt“-Kommandos immer mit dem Hinweis geschieht, dass es viel Zeit konsumiert. Anschließend können diese mittels „fdasd“ partitioniert (in früheren Kernel- und Utility-Versionen war nur eine Partition erlaubt, jedoch soll es inzwischen möglich sein, mehrere Partitionen in einer DASD anzulegen, welches dann auch nicht mehr das Aufsplitten auf VM-User-Ebene durch den Administrator „maint“ erfordert) und wie gewohnt den Mountpoints zugeordnet oder als Swap eingerichtet werden.

Die Auswahl der Software-Pakete erfolgt wie gewohnt, wobei jedoch nicht die komplette Vielfalt wie für die Intel-Plattform zur Verfügung steht; dazu jedoch weiter unten mehr. Der anschließende Installations- und Konfigurationsvorgang seitens „yast“ hält bis auf die Auswahl des Kernels und dessen Installation mittels „zipl“ keine weiteren Überraschungen bereit. Natürlich wird die S390-Architektur einen anderen Bootloader als „lilo“ benötigen, welcher mit „zipl“ zur Verfügung gestellt wird und den Kernel mit den notwendigen Kommandozeilenoptionen so installiert, dass er vom IPL des VM von einer Minidisk gestartet werden kann. Die Kernel- und Parameterdateien finden sich gewohnt unterhalb des Verzeichnisses „/boot“, die Konfigurationsdatei unter „/etc/zipl.conf“. Mittels „zipl“ wird der Kernel mit den notwendigen Kommandozeilenoptionen so installiert, dass er vom IPL des VM von einer Minidisk gestartet werden kann. Ältere Kernelversionen (2.2.X) konnten mittels „zilo“ VM-bootfähig gemacht werden, wobei „zipl“ einen weitaus besser handhab- und durchschaubaren Eindruck macht.

Die Netzwerkinterfaces sind zumeist Punkt-zu-Punkt-Verbindungen zum Hostbetriebssystem bzw. anderen Gast-Rechnern und werden wie klassische PtP-Interfaces unter Linux eingetragen. Der Direktzugriff auf einen Slot eines OSA-Express bietet den bekannten Anblick eines Ethernet-Adapters auf der PC-Workstation mit den üblichen Konfigurationen.

Etwas absurd erscheint das Einrichten sehr großer Festspeicherbereiche: Da die VM-Architektur nur virtuelle Festplatten bis zu 9 GByte Kapazität (vom Typ 3390) verwalten kann, wird der Festplattenplatz, welcher üblicherweise bereits auf einem Raid-Array zur Verfügung steht, aufgespalten. Unter Linux wird er dann mittels LVM wieder zusammengefügt, welches mit „yast“-Unterstützung am leichtesten geschieht. Hier hatten wir einige Schwierigkei-

ten zu überwinden, welche sich aber mit neueren Kernel- und Utility-Versionen verringerten.

7. Wie sieht Linux auf S390 aus?

Linux/S390 sieht aus wie Linux. Wenn man z. B. das Remote-Login per XDMCP erlaubt, kann man sich auf der Maschine mit der üblichen KDE-GUI anmelden und wird kaum einen Unterschied bemerken. Es gibt natürlich Tools, welche auf bestimmte Hardware bzw. Software-Schnittstellen wie das /proc-Interface geschrieben sind und nun nicht wie gewohnt funktionieren oder nicht zur Verfügung stehen. Aber alles, was ausreichend von der Hardware abstrahiert und in Sourcecode vorliegt, wird man in gewohnter Form vorfinden.

Hier zeigt sich wieder ein überragender Vorteil von Open-Source-Software, speziell von Linux: Nachdem Kernel und C-Compiler auf eine Architektur portiert sind, steht dem Einsatz der Software nichts mehr im Wege. Deshalb kann man mit dem Mozilla oder dem Konqueror surfen, jedoch nicht Netscape 4.XY installieren, da man hier auf ein Kompilat des Herstellers warten muss. Gleiches gilt für die Anpassung anderer kommerzieller Software, wo man sich in die direkte Abhängigkeit von den Vorstellungen des jeweiligen Anbieters begibt. Inzwischen berichtet IBM jedoch von einer steigenden Anzahl von Softwareproduzenten, welche ihre Produkte auch für S390-Architektur anpassen.

Da aber alle wesentlichen Netzwerk-Tools wie WWW-, FTP-, NFS-, DNS- oder Sambaserver und, wie schon erwähnt, die grafischen Oberflächen zur Verfügung stehen, kann jedoch kaum von Einschränkungen die Rede sein, so dass sich die Maschine für die im nächsten Abschnitt beschriebenen Anwendungen problemlos eignet. Die getroffenen Aussagen zur Software betreffen natürlich auch die Kernel-Module: Leider sind zumindest die direkten Hardware-Treiber, z. B. für die OSA-Netzwerk-Adapter, proprietär, so dass man auf bestimmte Kernel-Versionen, die von IBM geliefert werden, festgelegt ist und für einige Anwendungen die gewohnten Freiheiten, den Kernel komplett selbst zu übersetzen, wegfallen. Hier wird man zukünftige Entwicklungen beobachten wollen.

Die Performance von Linux auf S390 lässt sich aus Sicht unseres Projektes nicht so leicht einschätzen. Die installierten Standard-Services funktionierten klaglos und in der gewohnten Geschwindigkeit, wobei wir dabei selten an die Grenzen des Systems gingen. Darüber hinaus haben wir die Auswirkungen verschiedener Stellschrauben zur Veränderung der Performance kaum genutzt: So wäre es jederzeit möglich, die Priorität einzelner VM-

Benutzer hervorzuheben oder zurückzusetzen, Prozessoren in bestimmter Anzahl zu dedizieren oder mit SMP-Systemen zu experimentieren. Dieses setzt jedoch ein gutes Monitoring und Kenntnis der Maschine und ihrer Parameter z. B. mittels „fconx“ voraus. Weiterhin ließ sich schwer abschätzen, wie optimal der mittels „gcc“ erzeugte Code für die Maschine arbeitet. Dies betrifft zum einen die generelle Effizienz von Kernel-Treibern für einzelne Geräte (z. B. DASD-Zugriff) und zum anderen die Güte des Binär-Codes der Bibliotheken und Applikationen.

8. Erfahrungen und Probleme im Betrieb

Einige Probleme lagen darin, dass nicht sämtliche bekannte Software zur Verfügung stand: Das lag zum einen an proprietären Lizenzmodellen (Netscape 4.78 baut nicht für Linux/S390, StarOffice war auch nicht so schnell zu beschaffen) auf Seiten der Applikationen und zum anderen auf der Kernel-Seite. So ist es etwas mühsamer, einen speziellen Kernel mit den gewünschten Features zu bauen, wenn man auf die Object-Files für die Ethernet-Schnittstelle zurückgreifen muss, die es gibt (Problem der Abhängigkeiten). Allgemein war jedoch der Eindruck durchaus positiv: eine IBM-Hotline, die auch freitags nach 16.00 Uhr reagierte, gute Einführung in die Materie und etliche IBM-Unterlagen. Die aktuellen Dokumentationen und Kernel findet man unter [6].

Mit der aktuellen SuSE-Distribution für die S390-Plattform lag eine vernünftige Zusammenstellung an Software vor. Mittels Yast lassen sich die Standardaufgaben der Einrichtung eines S390-Systems auch ohne tiefere Kenntnis der Unterschiede (Lowlevel-Format der DASDs, Einrichten von LVM, Konfiguration und Installation des Bootloaders) zur PC-Architektur recht bequem erledigen.

9. Einsatzgebiete von Linux auf S390

9.1 Server-Konsolidierung

Nachdem die Installation von Linux, von einigen ungewohnten Punkten einmal abgesehen, reibungslos verlaufen ist, wird man sich nun dafür interessieren, wofür die Kombination Linux und S390 sinnvoll einsetzbar ist. Die Verknüpfung von Linux mit ausfallsicherer Hardware bietet sich gerade für eine Reihe von Standard-Services eines Rechenzentrums an. Da es nicht so sehr auf reine Rechenpower ankommt und viele Services nicht exakt zeitgleich nachgefragt werden und VM für eine sehr intelligente Lastverteilung zwischen seinen Nutzern sorgt, bietet es sich an, z. B. Mailserver, Domain Name Service, FTP- bzw. Samba-Fileserver oder Datenbanken (so sie nicht

in eigenen OS390-, VM-native-Installationen auf der S390 abgedeckt werden) in die Maschine zu verlagern. Man kann sich neben den Standard-Services eine solche Maschine als Bootserver für Diskless-Clients auf PC-Basis vorstellen, womit sich die Stabilität der Maschine und Linux mit der Wartungsarmut einer großen Zahl von PC-Arbeitsplätzen verbinden lassen. Innerhalb dieser Maschine lassen sich komplexe Netzwerkstrukturen abbilden, ohne dafür externe Hardware bereitstellen zu müssen.

So wie bereits der Telekommunikationsanbieter Telia seinen Kunden eine „eigene“ Maschine in einem solchen Mainframe einräumt, hat man den Vorteil, dass der Wartungsaufwand für das Server-Housing gering ausfällt und dem Kunden trotzdem eine eigene Maschine mit allen Optionen zur Verfügung steht, die Linux zu bieten hat. Ähnliches könnte man sich auch in einem wissenschaftlichen Rechenzentrum bzw. im universitären Bereich vorstellen: Da heute noch viele Fakultäten, Institute und Wissenschaftsbereiche ihre Server für die Außendarstellung selbst und damit dezentral betreiben, ließen sich mit dem beschriebenen Konzept diese Server in einer Maschine zusammenfassen, ohne dass die beteiligten Einrichtungen, wie sonst üblich, auf einen normalen Benutzerzugang eingeschränkt wären, sondern jede weiterhin „root“ für ihre „eigene Maschine“ bleiben würde.

Vorstellbar wären auch verschiedene Fail-Over-Szenarien: Durch das gemeinsame Nutzen von Platten ließe sich z. B. Ausfällen von Fileservern vorbeugen, indem eine „Ersatzmaschine“ für den Zugriff auf denselben Datenbestand vorgehalten wird. Dieses kostet im Leerlauf des Systems fast keine Ressourcen. Das gilt auch für Testinstallationen oder das Update der Services. Diese lassen sich sofort auf der Ziel-Hardware einrichten, ohne dafür diese duplizieren zu müssen. Die Aussicht, auf diese Weise das „Chaos im Server-Raum“ in den Griff zu bekommen, könnte einige RZ-Betreiber durchaus von dieser Maschine überzeugen.

9.2 Ausbildung

Möchte man im Zuge der universitären oder firmeninternen Aus- und Weiterbildung den Kursteilnehmern eine einheitliche Übungsumgebung zur Verfügung stellen, so kann vermieden werden, dass öffentlich zugängliche Kursrechner dafür abgestellt werden müssen und Sicherheitslücken bieten. Eine solche Kursumgebung lässt sich mittels Firewall bequem abschirmen; dazu lassen sich über die diversen Netzwerk-Setups in der S390 verschiedene Szenarien ausprobieren. Selbst wenn einzelne Maschinen nicht mehr erreichbar sein sollten, steht immer noch die 3270-Konsole für den virtuellen Reset oder zur Rekonfiguration zur Verfügung. Normale Hardware in einem Server-Raum ist bei weitem nicht so komfortabel zu erreichen und im Feh-

lerfall zu bedienen. Da auch im VM wieder ein VM installiert werden kann, lassen sich darüber hinaus Übungssysteme zur Zusammenarbeit von Linux/VM denken.

9.3 Firewall

Im Linux-Magazin (siehe hierzu [2]) wird beschrieben, wie eine Firewall-Lösung auf Linux S390 portiert wurde. Linux kann als Sicherheitsmodul auf einer S390 installiert werden, um deren Services im Intra- oder Internet anzubieten. So ließe sich das virtuelle interne Netz der Maschine in einem privaten IP-Bereich installieren, wobei nur bestimmte Services über den Linux-Firewall zugänglich gemacht werden. Im hier vorgestellten Projekt wurden alle VM-Schnittstellen nur dem internen IP-Bereich zugänglich gemacht, nur die Linux-Clients erhielten eine weltweit gültige IP-Nummer, wodurch sich bereits eine einfache Abstimmung erreichen ließ.

10. Fazit: Ein Saurier holt auf

Linux auf S390 bietet eine Ergänzung des Funktionsumfangs und der Anwendungsmöglichkeiten der Mainframe-Architektur S390. Mit einer großen Palette freier Software lassen sich viele zusätzliche Aufgaben in eine solche Maschine verlagern und damit Hardware, Platzbedarf und Administrationsaufwand einsparen. Attraktiv für bisherige Nutzer dieser Architektur dürfte neben der Vielfalt der hinzukommenden Software auch deren Lizenzmodell sein, gerade wenn eine größere Zahl von Maschinen installiert werden soll. Darüber hinaus lässt sich nun auch die eine oder andere herkömmliche Anwendung mit geringerem Kostenaufwand installieren. In der März-Ausgabe der Zeitschrift iX (siehe [10]) findet sich unter dem Aspekt Total Cost of Ownership eine Reihe weiterer Argumente für die Kombination von S390 und Linux.

Die geringen Unterschiede eines Linux auf PC-Basis und S390 sparen Umlernaufwand und bringen Effizienzvorteile der Rechnerverwaltung mit sich. In vielen Rechenzentren ist das Knowhow zur Bedienung des VM-Betriebssystems noch vorhanden und kann entsprechend für die neueren Generationen des Maschinentyps reaktiviert werden.

Die Standardprogramme und -dienste sind unter Linux/S390 genauso zu benutzen und zu konfigurieren, wie man dieses gewohnt ist. Mit der vorliegenden SuSE/S390-Distribution liegen bereits die meisten Software-Pakete angepasst vor, oder es lässt sich die mitgelieferte Entwicklungsumgebung dazu nutzen, noch nicht für S390 kompilierte Komponenten nachzuinstallieren. Proprietäre Software stellt unter der etwas exotischen Architektur natür-

lich ein gewisses Problem dar, obwohl nach Aussagen von IBM die Zahl der auf die S390-Architektur portierten Linux-Applikationen steigt.

Eine Weiterentwicklung der Linux-VM-Tools wird zukünftig den Zugriff auf Minidisks aus Linux-Sicht erlauben und damit helfen, den Administrationsaufwand von Linux-Clients aus VM-Sicht zu senken.

Die Architektur der Maschine erlaubt den langsamen Einstieg und die Migration zu Linux aus Sicht der bisherigen Nutzer dieser Architektur. Die S390 wird damit als Plattform auch für andere Nutzerkreise interessanter.

Literatur

[1] **http://www.stud.uni-goettingen.de/~dsuchod/s390**

[2] Frank Bernard u. Simon Fischer: Firewall auf S/390 portieren. In: Linux-Magazin 2/2002, S. 60

[3] IBM-Redbook „Linux for S390“

[4] Jürgen Hattenbach: CMS-Fibel - Eine Einführung in die Arbeit unter dem IBM-Betriebssystem VM/CMS. GWDG, Göttingen 1992

[5] Kolacki (Hrsg.): VM/CMS - Virtuelle Maschinen, Praxis und Faszination eines Betriebssystems. Braunschweig 1991

[6] Die „offiziellen“ IBM-Kernels:
http://oss.software.ibm.com/developerworks/opensource/linux390/current2_4.shtml

[7] Ulrich Wolf: Pinguin im Mainframe-Land. In: Linux-Magazin 6/2000, S. 48

[8] Presseinformation der Firma IBM zu „Linux öffnet dem Mainframe die Tür zu Universitäten“ vom 18.12.2001:
http://www.gwdg.de/aktuell/presse/ibm.011218.html

[9] Die Marist-Linux-Distribution für IBM S390:
http://linux390.marist.edu

[10] Martin Arndt: Unerwarteter Rivale - Linux gräbt Unix das Wasser ab. In: iX 3/2002, S. 80

repositorium - Multimediales Redaktions- und Publikationssystem für die Geisteswissenschaften

Dagmar Ullrich

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Zusammenfassung

Mit *repositorium*[1] entsteht eine Digitale Bibliothek für die Geisteswissenschaften. Multimediale Lehrmaterialen und digitalisierte Originale können verwaltet und recherchiert werden. *repositorium* ermöglicht die online-Publikation wissenschaftlicher Arbeiten. Als Backend wird ein Content Management System mit integrierter Datenbank eingesetzt. Der Zugriff kann mittels gängiger Browsersoftware erfolgen, so dass eine ort- und zeitunabhängige Nutzung der Daten möglich ist.

1. Einleitung

repositorium ist eine internetbasierte Arbeitsumgebung für zwei zentrale Bereiche wissenschaftlichen Arbeitens. Zum einen können Lehrmaterialien aller Art zentral gesammelt und verfügbar gemacht werden. Zum anderen

1. http://www.repositorium.net

bietet *repositorium* die Möglichkeit, wissenschaftliche Publikationen zu erstellen und zu recherchieren. Je nach Bedarf können die unterschiedlichen Dokumente kurzfristig z. B. für Lehrveranstaltungen genutzt oder langfristig archiviert werden.

In der Entwicklungsphase richtet sich *repositorium* speziell an Frühneuzeit-Historiker, die mit digitalen Dokumenten arbeiten wollen - sei es in der Forschung, zur Projektdokumentation oder zur Unterstützung eigener Seminare und Vorlesungen. Eine spätere Ausweitung des Nutzerkreises auf andere Fachbereiche der Geisteswissenschaften ist geplant. Das Projekt wurde im Oktober 2000 begonnen. Die Entwicklungsphase wird voraussichtlich bis Ende 2002 dauern. Danach soll das System in den Regelbetrieb übernommen werden.

Beteiligte Institutionen[2]

repositorium wird im Rahmen des DFG Projekts MELISSA entwickelt. MELISSA wird von der Deutschen Forschungsgemeinschaft innerhalb des Programms zur Förderung des wissenschaftlichen Bibliothekswesens unterstützt. Das Projekt entsteht in Kooperation der Bayerischen Staatsbibliothek München, der Abteilung für Geschichte der Frühen Neuzeit der Ludwig-Maximilians-Universität München und der Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG). Die Arbeitsteilung sieht vor, dass die Bayerische Staatsbibliothek und die Abteilung für Geschichte der Frühen Neuzeit vorwiegend für die Auswahl der Inhalte zuständig sind. Die technische Umsetzung und Implementierung dagegen liegt bei der Gesellschaft für wissenschaftliche Datenverarbeitung in Göttingen.

2. Einsatzmöglichkeiten

2.1 Bereitstellung von Lehrmaterialien

Mit *repositorium* können Materialsammlungen für Lehrveranstaltungen erstellt werden, die jederzeit von jedem Ort über einen Internetbrowser verfügbar sind. Solche Sammlungen können Reader, Protokolle oder Übungsunterlagen enthalten. Eigenes Material kann um bereits vorhandenes öffentliches Material ergänzt werden. Ziel hierbei ist nicht nur, die Fachbe-

2. http://www.dfg.de
   http://www.bsb-muenchen.de
   http://www.lmu.de
   http://www.gwdg.de

reiche vor Ort zu unterstützen, sondern auch eine überregionale Zusammenarbeit im Sinne eines virtuellen Fachbereichs zu fördern.

2.2 Veröffentlichen digitalisierter Originale

repositorium soll die Aufgaben einer Digitalen Bibliothek für die Geisteswissenschaften übernehmen. Dazu gehört auch die Verwaltung und Archivierung digitalisierter Originale. Für die Geisteswissenschaften handelt es sich hierbei vorwiegend um alte Handschriften, aber z. B. auch um Bildbestände oder Tonaufnahmen. Diese Dokumente könne inklusive ihrer Metadaten verwaltet und archiviert werden. Der Verwaltung digitalisierten Quellmaterials kommt besondere Bedeutung zu, denn sie gewährleistet die ortsunabhängige Verfügbarkeit dieser Quellen. Gleichzeitig werden die oft empfindlichen Originale geschont. Speziell für Bilddateien steht eine besondere Zoomfunktion zur Verfügung, die mehrfache Ausschnittvergrößerungen von digitalisierten Karten, Autographen, Flugschriften und anderem Material ermöglicht.

2.3 Recherchieren im Dokumentenbestand und angeschlossenen Archiv- oder Bibliotheksbeständen

Umfangreiche Suchfunktionen u.a. nach Titel, Stichwort, Autor oder Erscheinungsjahr unterstützen die Dokumentenrecherche sowohl im eigenen Bestand als auch in daran angeschlossenen Archiven oder Bibliotheken. Lokale Unterrichtsmaterialien standen bisher nicht für eine Recherche dieser Art zur Verfügung, sondern konnten nur im engen Rahmen der jeweiligen Lehrveranstaltung vor Ort genutzt werden.

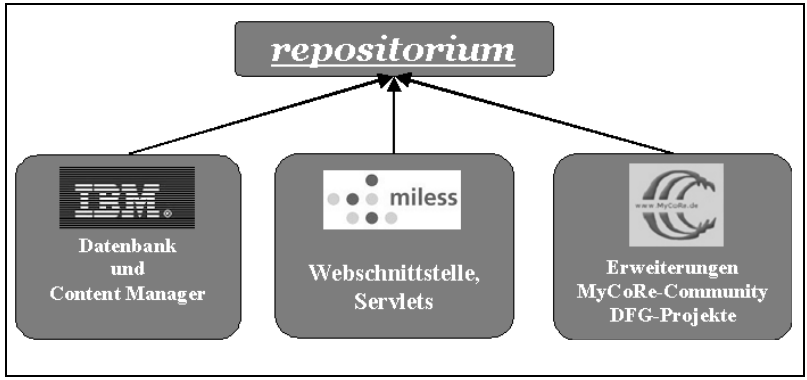
2.4 Online-Publikation wissenschaftlicher Arbeiten

Sowohl Einzelpersonen als auch Autorenteams können *repositorium* nutzen, um Dokumente zu erstellen und zu veröffentlichen. Es können z. B. Dissertationen, Quellensammlungen, Dokumentationen, Monografien oder Lexika auf diesem Wege publiziert werden. Den Autoren bleiben alle Urheberrechte einschließlich der Option zur Publikation der Inhalte auch an anderer Stelle voll erhalten. Besonders interessant ist *repositorium* für räumlich verteilte Autorenteams. Der ortsunabhängige Zugriff und die Möglichkeit differenzierter Schreib- und Leserechte sind ideale Voraussetzungen für diese Arbeitsform. Speziell für Autorenteams wird an einem Workflow-Modul gearbeitet.

3. Anforderungen an *repositorium*

Bei der Entwicklung von *repositorium* wird besonderer Wert auf eine einfache, intuitive Bedienbarkeit, solide Langzeitarchivierung sowie die Einhaltung bibliothekarischer Standards gelegt. Um *repositorium* zu nutzen, wird lediglich ein Internetzugang und eine gängige Browsersoftware (z. B. Internet Explorer, Netscape oder Opera) vorausgesetzt. Alle eingestellten Dokumente werden über ein Backup-System regelmäßig und langfristig gesichert. Eingestellte Dokumente unterliegen einer redaktionellen Qualitätskontrolle durch die jeweils zuständigen Fachbereiche. Es werden Dokumente unterschiedlichster Formate verwaltet. Dazu gehören Textdokumente, Bild-, Video- und Tonaufnahmen als auch kleinere Programme (z. B. zur Simulation). Der Zugriff auf die Dokumente kann über eine Rechteverwaltung gesteuert werden.

4. Komponenten



4.1 IBM Content Manager[3]

repositorium nutzt den Content Manager von IBM (ehemals Digital Library), der als Datenbank die DB2 einbindet. Der Content Manager bietet eine Reihe von Funktionalitäten, die für *repositorium* nützlich sind. Er ermöglicht die Verteilung der Datenbestände auf unterschiedliche Standorte bzw. Server, mit zum Teil unterschiedlichen Aufgaben. Der Content Manager besteht aus zwei Hauptkomponenten, dem Library-Server, der die Anfragen entgegennimmt und (ggf. mehreren) Objektservern, die in der Regel die

3. http://www-4.ibm.com/software/data/cm

gewünschten Dokumente ausliefern. Diese Architektur erlaubt es, sehr flexibel auf unterschiedliche Hardwarebedingungen und Standortvorgaben einzugehen. Der Content Manager verfügt weiter über eine differenzierte Rechteverwaltung und ein Workflow-Management. Beide Funktionalitäten sollen von *repositorium* genutzt und z. T. erweitert werden. Auch der angeschlossene Text- und Bildsuchserver, die Möglichkeit des Video-Streaming, die Speicherverwaltung sowie das Backup-System sind für *repositorium* von Interesse.

4.2 Miless[4]

Miless ist ein Projekt der Universität Essen. Dort wurde ein multimedialer Lehr- und Lernserver für die dortigen lokalen Bedürfnisse entwickelt. Miless baut ebenfalls auf den Content Manager auf und stellt u.a. eine Webschnittstelle in Form von Servlets und Applets zur Verfügung. Sie bietet Suchmasken, ein Autoren-GUI zum Einstellen, Ändern oder Löschen von Dokumenten sowie eine Darstellungsfunktion für die Dokumente mittels entsprechender Plugins. Neben dieser Webschnittstelle bietet Miless eine Metadatenverwaltung, die auf dem Dublin Core[5] Standard basiert und damit eine wichtige Voraussetzung zur Nutzung als Digitale Bibliothek schafft. Ebenfalls in wissenschaftlichen Kontexten wichtige Fachsystematiken wurden durch Miless integriert. Von kleineren Anpassungen abgesehen wurde der Miless-Code von *repositorium* übernommen und stellt die Grundlage des weiteren Ausbaus dar.

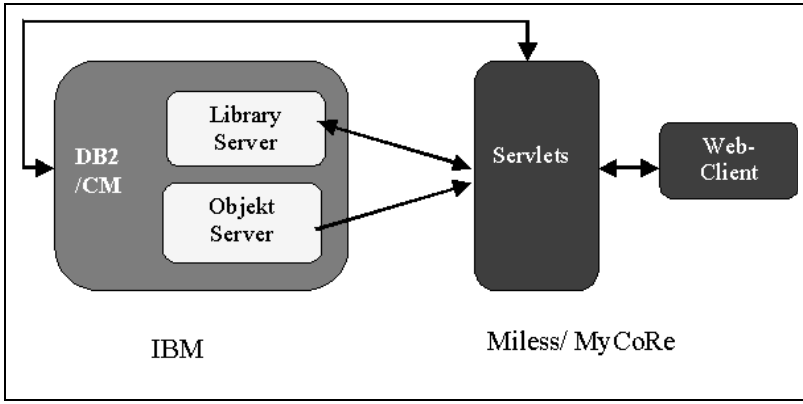
4.3 MyCoRe[6]

Nicht nur *repositorium* arbeitet an einem System auf der Grundlage der oben genannten Komponenten Content Manager und Miless. An einer Reihe von Universitäten entstehen verwandte Projekte auf dieser Basis. Die Beteiligten haben sich in einer Gruppe zusammengefunden und arbeiten an verschiedenen Weiterentwicklungen des Miless-Codes als Open Source Produkt. *repositorium* ist Mitglied dieser MyCoRe-Entwicklerguppe.

---

4. http://miless.uni-essen.de
5. http://dublincore.org
6. http://www.mycore.de

5. Architektur



Diese Skizze zeigt das Zusammenspiel der jeweiligen Komponenten. Der Web-Client kommuniziert mit dem Server mittels Servlets. Diese nehmen die Requests entgegen und leiten die entsprechenden Daten an den Content Manager weiter. Der Library-Server nimmt diese Daten entgegen, bearbeitet sie und gibt ggf. Antworten zurück. Die Auslieferung von gesuchten Dokumenten erfolgt in der Regel über den Objekt-Server. Der Content Manager nutzt die DB2 zur Verwaltung der Daten. Auf Client-Seite wird zum Einstellen, Bearbeiten und Löschen von Dokumenten ein Applet verwendet, dem ein spezielles Servlet auf der Serverseite gegenübersteht. Die Dokumentbearbeitung umfasst u.a. die Angabe von Metatdaten, Angaben zu Personen und Körperschaften und das Einstellen verschiedener Dokument-Derivate (z. B. ein Text in unterschiedlichen Dateiformaten).

6. Weiterentwicklungen durch *repositorium*

Aufbauend auf den durch den Content Manager und den Miless-Code zur Verfügung gestellten Funktionalitäten finden im Rahmen von *repositorium* folgende Weiterentwicklungen statt:
Das Dublin Core basierte **Datenmodell** von Miless wird an die speziellen Erfordernisse der Geisteswissenschaften angepasst. Die **Binnenstruktur einzelner Dokumente** soll aufgelöst werden können, um so z. B. Suchanfragen auch unterhalb der Dokumentebene zu bearbeiten. D. h., es soll möglich sein, auf Suchanfrage hin, ein spezielles Kapitel oder eine eingebundene Grafik zu erhalten, nicht notwendig das gesamte Dokument. Hierdurch wird eine größere Genauigkeit der Suchergebnisse und eine Mehrfachnutzung von Teildokumenten angestrebt. Weiter soll die **Verlinkung von Dokumen-**

ten untereinander, optimalerweise auch unterhalb der Dokumentebene, möglich sein. Geplant ist hierbei sowohl ein explizites Linking, das durch den Autor eines Dokumentes aktiv veranlasst wird, als auch ein implizites Linking auf der Basis einer Verschlagwortung. Die letztere Variante erfordert besondere Aufmerksamkeit hinsichtlich der Autorenrechte. Diese Links sollen gegenüber herkömmlichen die Möglichkeit der Bidirektionalität bieten. Um Lehrveranstaltungen und das Online-Publizieren zu unterstützen, wird ein **Workflow-Modul** erstellt, das eng mit der Rechteverwaltung des Systems verbunden sein wird. Die **Rechteverwaltung** wird den für die Qualitätssicherung erforderlichen redaktionellen Ablauf unterstützen. Es kann jederzeit kontrolliert werden, wer auf welche Dokumente wie zugreifen kann. Es wird in Lese-, Schreib- und Veröffentlichungsrechte unterschieden werden. Die Rechtervergabe soll z. T. dezentral möglich sein, d.h., hinsichtlich bestimmter Arbeitsbereiche werden unterschiedliche Personen die Rechtevergabe steuern können. Da der Zugriff auf *repositorium* über einen Webbrowser erfolgt, werden entsprechende **Webseiten** erstellt werden.

7. Ansprechpartner

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Thorsten Agemar tagemar@gwdg.de
Tel.: 0551 / 201 1831

Dagmar Ullrich dullric@gwdg.de
Tel.: 0551 / 201 1827

Bayerische Staatsbibliothek München

Gregor Horstkemper horstkemper@bsb-muenchen.de
Tel.: 089 / 28638 2914

Dr. Marianne Dörr doerr@vd17.bsb.badw-muenchen.de
Tel.: 089 / 28638 2600

In der Reihe GWDG-Berichte sind zuletzt erschienen:

Nähere Informationen finden Sie im Internet unter
`http://www.gwdg.de/forschung/publikationen/gwdg-berichte`

Nr. 40 *Plesser, Theo und Peter Wittenburg* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 1994
1995

Nr. 41 *Brinkmeier, Fritz* (Hrsg.):
Rechner, Netze, Spezialisten. Vom Maschinenzentrum zum Kompetenzzentrum - Vorträge des Kolloquiums zum 25jährigen Bestehen der GWDG
1996

Nr. 42 *Plesser, Theo und Peter Wittenburg* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 1995
1996

Nr. 43 *Wall, Dieter* (Hrsg.):
Kostenrechnung im wissenschaftlichen Rechenzentrum - Das Göttinger Modell
1996

Nr. 44 *Plesser, Theo und Peter Wittenburg* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 1996
1997

Nr. 45 *Koke, Hartmut und Engelbert Ziegler* (Hrsg.):
13. DV-Treffen der Max-Planck-Institute - 21.-22. November 1996 in Göttingen
1997

Nr. 46 **Jahresberichte 1994 bis 1996**
1997

Nr. 47 *Heuer, Konrad, Eberhard Mönkeberg und Ulrich Schwardmann:*
Server-Betrieb mit Standard-PC-Hardware unter freien UNIX-Betriebssystemen
1998

Nr. 48 *Haan, Oswald* (Hrsg.):
Göttinger Informatik Kolloquium - Vorträge aus den Jahren 1996/97
1998

Nr. 49 *Koke, Hartmut und Engelbert Ziegler* (Hrsg.):
IT-Infrastruktur im wissenschaftlichen Umfeld - 14. DV-Treffen der Max-Planck-Institute, 20. - 21. November 1997 in Göttingen
1998

Nr. 50 *Gerling, Rainer W.* (Hrsg.):
Datenschutz und neue Medien - Datenschutzschulung am 25./26. Mai 1998
1998

Nr. 51 *Plesser, Theo und Peter Wittenburg* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 1997
1998

Nr. 52 *Heinzel, Stefan und Theo Plesser* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 1998
1999

Nr. 53 *Kaspar, Friedbert und Hans-Ulrich Zimmermann* (Hrsg.):
Internet- und Intranet-Technologien in der wissenschaftlichen Datenverarbeitung - 15. DV-Treffen der Max-Planck-Institute, 18. - 20. November 1998 in Göttingen
1999

Nr. 54 *Hayd, Helmut und Theo Plesser* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 1999
2000

Nr. 55 *Kaspar, Friedbert und Hans-Ulrich Zimmermann* (Hrsg.):
Neue Technologien zur Nutzung von Netzdiensten - 16. DV-Treffen der Max-Planck-Institute, 17. - 19. November 1999 in Göttingen
2000

Nr. 56 *Plesser, Theo und Helmut Hayd* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 2000
2001

Nr. 57 *Hayd, Helmut und Rainer Kleinrensing* (Hrsg.):
17. und 18. DV-Treffen der Max-Planck-Institute
22. - 24. November 2000, 21. - 23. November 2001 in Göttingen
2002

Nr. 58 *Macho, Volker und Theo Plesser* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 2001
2002