

Synthetische Lasttests auf dem 100-Gigabit-Testbed zwischen der TU Dresden und der TU Bergakademie Freiberg

Andy Georgi, Thomas William, Prof. Dr. Wolfgang E. Nagel

Andy.Georgi@tu-dresden.de
Thomas.William@zih.tu-dresden.de
Wolfgang.Nagel@tu-dresden.de

Abstract: Das Hochleistungsrechenzentrum der TU Dresden wurde in Kooperation mit Alcatel-Lucent und T-Systems über eine 100-Gigabit-Strecke mit dem Rechenzentrum der TU Bergakademie Freiberg verbunden. Die Umsetzung erfolgt erstmals mit kommerzieller Hardware. Eine weitere Besonderheit ist die Übertragung der Daten über eine einzige Wellenlänge. In dieser Arbeit werden nach der Beschreibung des 100-Gigabit-Testbeds erste Ergebnisse synthetischer Lasttests auf verschiedenen Ebenen des OSI-Referenzmodells vorgestellt und diskutiert. Die Vorgaben und der spezielle Aufbau des Testbeds machen eine Überarbeitung vorhandener Messmethoden notwendig. Die Entwicklung der Burst-Tests wird ebenfalls in dieser Arbeit beschrieben. Zusätzlich wird ein Ausblick auf weitere geplante Teilprojekte, sowie Einsatzmöglichkeiten der 100-Gigabit-Technik in Forschung und Industrie gegeben.

1 Einführung

Die verfügbare Bandbreite der Datennetze wird durch aktuelle Trends wie Grid- und Cloud-Computing, aber auch durch die stetig steigende Menge an zu verteilenden Daten vielfach zum bestimmenden Faktor. Für Wissenschaft und Forschung werden daher Gigabit-Netze in Zukunft unerlässlich. Die Erprobung der neuen 100-Gigabit-Technologie zwischen Dresden und Freiberg unter realen Bedingungen ist dafür ein wichtiger Meilenstein.

Alcatel-Lucent und T-Systems haben das Hochleistungsrechenzentrum der TU Dresden und das Rechenzentrum der TU Bergakademie Freiberg im Rahmen eines Testbeds miteinander vernetzt. Es kommt erstmals kommerzielle 100-Gigabit-Technik mit aufeinander abgestimmten IP-Routern und optischen Übertragungssystemen zum Einsatz. Diese Kombination ermöglicht eine besonders hohe Skalierbarkeit, Kapazität und Kompatibilität in Core-, Edge- und Metro-Netzen. Die Daten auf der 100-Gigabit-Strecke zwischen den beiden Technischen Universitäten werden in Echtzeit verarbeitet und über eine einzige Wellenlänge übertragen.

Dank ihrer hohen Bandbreite ermöglicht die 100-Gigabit-Technik den Forschungsinstituten eine neue Qualität der Zusammenarbeit. Sie eignet sich für HPC- und Cloud-basierte

Dienste sowie für Multimedia-Anwendungen, welche das gemeinsame Arbeiten von Nutzern unterstützt (Kollaboration). Die neue Verbindung bietet eine Umgebung für die Evaluierung der Eigenschaften von bandbreitenintensiven Anwendungen und Diensten unter realen Bedingungen. Wissenschaftler der beiden Universitäten untersuchen dabei den Ressourcenbedarf von HPC-, Multimedia- und verteilten Server-Anwendungen.

In dieser Arbeit werden nach der Beschreibung des 100-Gigabit-Testbeds (Kapitel 2) erste Ergebnisse synthetischer Lasttests auf verschiedenen Ebenen des OSI-Referenzmodells vorgestellt und diskutiert (Kapitel 4). In Zusammenarbeit mit den Herstellern wurden konkrete Testszenarien definiert und implementiert, welche in Kapitel 3 erläutert werden. Dabei soll untersucht werden, welche Effekte der Burst-Tests auf TCP/IP Ebene bei Verwendung burst-artiger I/O Muster auf der Anwendungsebene noch sichtbar sind. Zum Abschluss erfolgt in Kapitel 5 eine Zusammenfassung sowie ein Ausblick auf weitere geplante Teilprojekte auf dem 100-Gigabit-Testbed und Einsatzmöglichkeiten in der Forschung und Industrie.

2 Aufbau der Teststellung

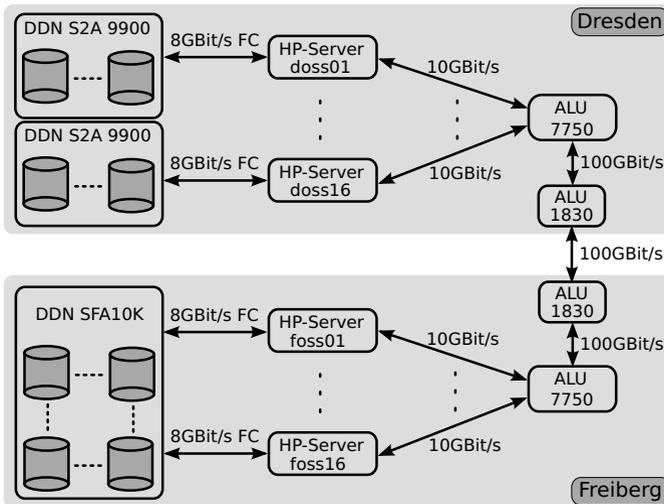


Abbildung 1: Aufbau der Teststrecke

Die Abbildung 1 zeigt den Aufbau des 100-Gigabit-Testbeds. Für die näherungsweise 60 km lange Dark-Fibre-Verbindung zwischen Dresden und Freiberg kamen die bereits vorhandenen Leitungen der T-Systems zum Einsatz.

Der Alcatel-Lucent 1830 Photonic Service Switch [Alc10a] ist in der Lage 100 Gbit/s auf einer Wellenlänge mittels eines einzelnen Trägers zu übertragen. Die Schnittstelle zwischen der optischen Verbindung und den Netzwerkadaptern bildet ein 7750 SR-12 Ser-

vice Router [Alc10b] von Alcatel-Lucent. Aktuell enthält der Router an beiden Standorten einen Media Dependent Adapter (MDA) mit 1x100 Gbit/s, zwei Adapter mit 5x10 Gbit/s sowie fünf MDAs mit 2x10 Gbit/s. Zusätzlich stehen noch 20x1 Gbit/s Ports zur Verfügung.

Über die verfügbaren 10 Gbit/s Ports sind 18 ProLiant DL160 G6 Server [Hew10] der Firma Hewlett-Packard angeschlossen. Enthalten ist jeweils ein Intel Xeon mit einer Taktfrequenz von 2.67 GHz, 24 GByte Hauptspeicher, ein HP NC550SFP 10GbE Server Adapter, ein HP 4x QDR ConnectX-2 InfiniBand Host Channel Adapter, sowie ein HP StorageWorks 81Q 8Gb Fibre Channel Host Bus Adapter. Zwei der HP Server sind für andere Teilprojekte reserviert, so dass wie in Abbildung 1 dargestellt, in Dresden und Freiberg 16 Knoten für die Durchführung der synthetischen Lasttests zur Verfügung stehen.

Die Speicherlösung in Freiberg ist eine SFA10K [Dat10b] der Firma Direct Data Networks (DDN) mit einer Leistung von 10 GByte/s (Herstellerangabe). Das System in Dresden besteht aus zwei S2A 9900 [Dat10a] mit jeweils 6 GByte/s. Diese Systeme sind über 8Gb Fibre Channel an die HP-Server angebunden. Insgesamt sind 1,5 TByte Speicherplatz verfügbar.

Die InfiniBand Karten der HP-Server (nicht in der Abbildung 1 enthalten) dienen zur Anbindung der lokalen Cluster. Diese wurden als Clients für die I/O-Tests zur Durchsatzoptimierung und maximalen Auslastung des 100-Gigabit-Links verwendet (siehe [Uni10]), wobei die HP-Server als Lustre-Router agieren. Außerdem kommunizieren die HP-Server als Object Storage Targets (OST) des Lustre Dateisystems über die InfiniBand Karten mit dem Metadaten Server (MDS), welcher auf einem extra System installiert ist. In der aktuellen Arbeit wurden die Experimente direkt auf den ProLiant Servern ohne Nutzung der Cluster ausgeführt.

3 Beschreibung der durchgeführten Tests

Um das Potenzial der neu eingeführten 100-Gigabit-Technologie zu quantifizieren werden synthetische Lasttests auf dem Testbed durchgeführt. Der Verwaltungsaufwand beeinflusst dabei die Ergebnisse. Da dieser auf Anwendungsebene häufig nicht exakt definiert werden kann, werden außerdem Messungen auf der Transportebene durchgeführt. In diesem Kapitel wird nach der Definition der Projektziele auf die verschiedenen Testscenarien eingegangen.

3.1 Definition der Projektziele

Neben den klassischen Latenzzeitmessungen und der Bestimmung des Datendurchsatzes mit Hilfe von kontinuierlichen Datenströmen wurden in Zusammenarbeit mit Alcatel-Lucent und T-Systems weitere Projektziele definiert. Im Vordergrund stand dabei das Nutzungsverhalten innerhalb einer realen Umgebung auf das 100-Gigabit-Testbed abzubilden. Dazu gehört das Auftreten von unregelmäßigem Datenverkehr, sogenannten Daten-Bursts.

3.2 Burst-Tests

Auf TCP/IP-Ebene steht, soweit bekannt, kein adäquater Benchmark für die Durchführung von Burst-Tests zur Verfügung. Aus diesem Grund war eine Neuimplementierung notwendig. Eingesetzt wurde hierfür BenchIT [GJ03], ein frei verfügbares Framework für Performance-Messungen. Somit war lediglich die Integration der Funktionalität und die Verknüpfung mit den vorhandenen Schnittstellen notwendig. BenchIT verwaltet die gesammelten Leistungsdaten und bietet eine Oberfläche um diese graphisch darzustellen.

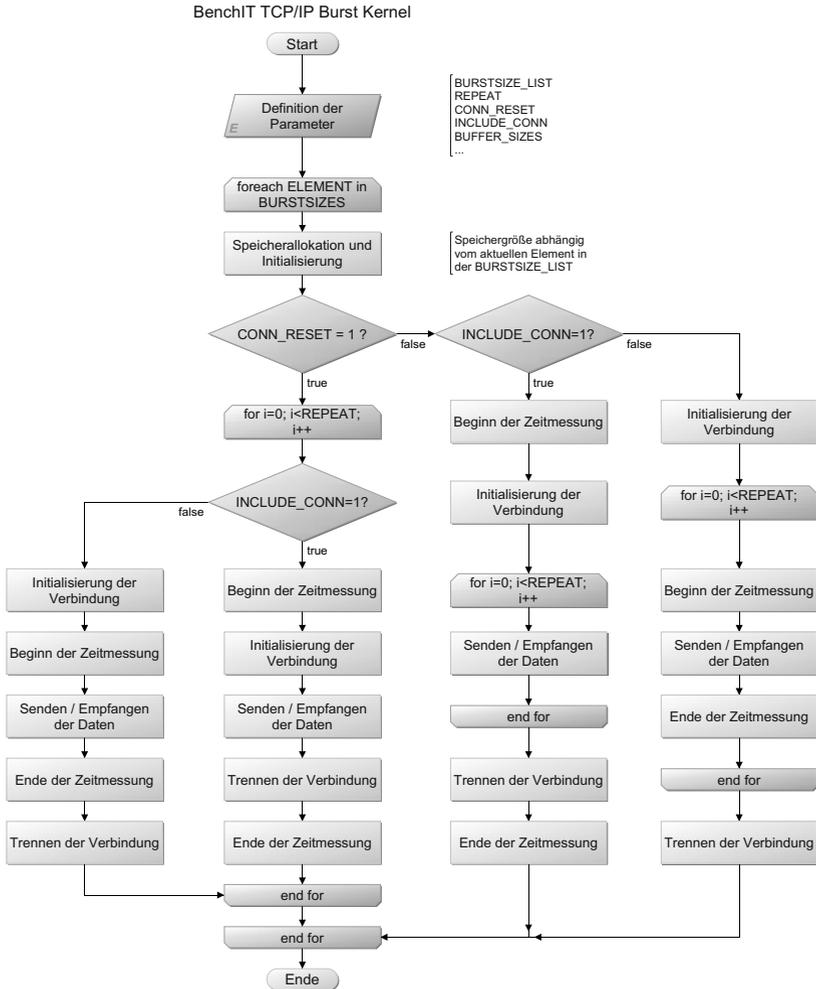


Abbildung 2: Programmablauf des BenchIT TCP/IP Burst Microbenchmarks

Die Abbildung 2 zeigt einen Programmablaufplan des BenchIT-Microbenchmarks. Dabei handelt es sich um ein Client-/Server-Modell wobei der Client zuerst Daten an den Server sendet, welcher diese im Anschluss zurück schickt. Vom Nutzer wird über eine Liste die Größe der Daten-Bursts spezifiziert und wie häufig diese übertragen werden. Zusätzlich können Systemparameter, wie bspw. die Empfangs- und Sendepuffergrößen definiert werden. Diese dienen zur Initialisierung der Verbindung. Weiterhin kann der Nutzer festlegen, ob die Initialisierung der Verbindung zeitlich erfasst wird oder lediglich die Datenübertragung gemessen werden soll. Auch eine Trennung der Verbindung nach jeder Übertragung ist möglich. Unabhängig von den definierten Parametern wird nach jeder Iteration die Differenz aus den beiden Zeitstempeln berechnet und gesichert. Wird die vorgegebene Anzahl an Wiederholungen erreicht, erfolgt die Berechnung des minimalen, maximalen und durchschnittlichen Zeitverbrauchs. Diese Werte werden an BenchIT übergeben und am Ende zusammen mit den Parametern und Umgebungsvariablen in eine Ergebnisdatei geschrieben. Diese wiederum kann im Anschluss mit Hilfe der graphischen Oberfläche von BenchIT visualisiert werden.

Ausgehend von den Burst-Tests auf TCP/IP-Ebene wurden zusätzlich synthetische I/O-Tests auf der Anwendungsebene mit den selben Größen für die Datenbursts und gleicher Anzahl der Prozesse durchgeführt. Dabei kam der Benchmark IOR 2.10.3¹ zum Einsatz. Mit diesem Programm der Universität Kalifornien ist es möglich die Leistung paralleler Dateisysteme zu ermitteln [NLC08]. Zur Synchronisation der Prozesse dient MPI wobei im vorliegenden Testszenario jeder Prozess in eine eigene Datei schreibt. In den Tests wurde ausschließlich POSIX I/O mit dem Flag `O_DIRECT` zur Umgehung der Caches verwendet. Zwischen den einzelnen Tests wurde eine `MPI_Barrier` zur Synchronisation der Prozesse verwendet, um so die Synchronität der Bursts zu gewährleisten. Wie bei den Burst-Tests der Transportschicht wird jeder Test zehnmal durchgeführt. Da die kleinste übertragbare Datenmenge in IOR 8 Byte beträgt, wurden die Größen der Bursts jeweils zum nächst größeren Vielfachen von acht erhöht. IOR unterscheidet zwei für die hier präsentierten Messungen wichtige Werte, die `blockSize` und die `transferSize`. Erstere beschreibt die Größe der Datei und Letztere die Menge der durch einen I/O-Aufruf übertragenen Daten. Um die Vergleichbarkeit der Messungen zu gewährleisten wurden beide Variablen immer auf den gleichen Wert gesetzt wobei allerdings zu beachten ist, dass IOR nur eine Transfergröße von maximal 2 GByte erlaubt. Alle Tests über 2 GByte wurden daher von IOR automatisch in zwei I/O-Operation aufgeteilt. Wie im TCP/IP-Test wurden Daten nur in eine Richtung übertragen. Dies wurde durch reine Schreiboperationen umgesetzt.

Für die IOR Messungen wurde in Freiberg ein Lustre Dateisystem auf den 16 HP-Servern aufgesetzt. Jeder Server verfügt über drei Laufwerke aus dem DDN-System. Jedes Laufwerk wird über InfiniBand als Object Storage Target (OST) an den auf einem eigenen Knoten laufenden Meta Data Server (MDS) exportiert. Somit stehen insgesamt 48 OSTs mit einer theoretischen Bandbreite von 16*8 Gbit/s zur Verfügung. Auf jedem der 16 Server in Dresden wird das Lustre Dateisystem eingebunden. Auf einem weiteren Rechner liegen die Konfigurationsdateien und der IOR Benchmark in einem, über NFS auf allen Servern verfügbaren, Ordner.

¹URL: <http://ior-sio.sourceforge.net>

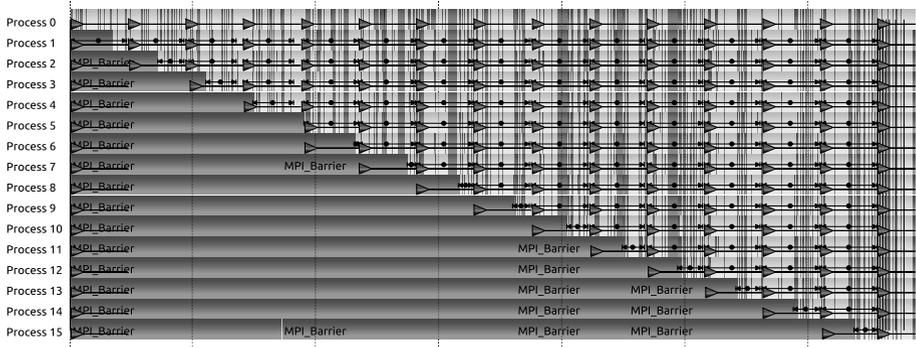


Abbildung 3: Programmspur des Messlaufs visualisiert mit Vampir

Wie im TCP/IP Test werden die Messungen nacheinander mit steigender Anzahl von Rechnern durchgeführt, wobei IOR `MPI_Barrier` zur Blockierung aller inaktiven Knoten einsetzt. Mit Hilfe des Analysewerkzeugs Vampir [WMM10] wurde eine Programmspur des Testlaufs aufgezeichnet, in welche zusätzlich die Statistikwerte des 7750 Service Routers eingefügt wurden. Der gesamte Messlauf, wie er in Vampir dargestellt wird, ist in Abbildung 3 zu sehen. Dabei ist mit der Farbe rot jegliche Kommunikation mit MPI markiert, gelb stellt I/O-Operationen dar und grün sind Funktionsaufrufe des IOR Benchmarks. Diese dienen hauptsächlich zum Erstellen, Füllen und Leeren der Puffer, welche dann durch die I/O Funktionen auf das Lustre Dateisystem geschrieben werden.

4 Ergebnisse

Im Folgenden werden die in Kapitel 3 beschriebenen Messläufe auf Transport- und Anwendungsebene durchgeführt und diskutiert. Dabei wurde in beiden Fällen über die Anzahl der Knoten und ähnliche Burst-Größen iteriert um die Vergleichbarkeit zu gewährleisten.

4.1 Latenzzeitmessung und Stream-Tests auf der Transportebene

Im ersten Schritt wurde eine Punkt-zu-Punkt-Verbindung zwischen einem Knoten in Dresden und einem Knoten in Freiberg untersucht. Dabei belief sich die Round-Trip-Time (RTT), bestimmt durch die Messung der Zeit für die Hin- und Rückübertragung einer 64 Byte Nachricht, auf 0.72 ms. Diese setzt sich wie folgt zusammen:

- TCP/IP-Stack bzw. Netzwerkkarte: 0.1 ms (Direktverbindung zweier Knoten)
- Alcatel-Lucent 7750 Service Router: 0.08 ms (Messungen an einem Standort)

- Alcatel-Lucent 1830 Photonic Service Switch sowie eine 60 km Glasfaserverbindung: 0.54 ms

Darüber hinaus konnte mit Hilfe einer Path Maximum Transmission Unit (PMTU) von 9000 Byte für die o.g. Verbindung ein Datendurchsatz von 9.92 Gbit/s gemessen werden. Dabei wurde die PMTU durch die in Kapitel 2 beschriebenen Netzwerkkarten definiert. Die Messung erfolgte dabei mit Hilfe des in Kapitel 3 beschriebenen BenchIT-Kernels, wobei anstelle einer Liste von Bursts lediglich ein Wert von 2 GByte übergeben wurde und weder die Initialisierung noch die Trennung der Verbindung gemessen wurden. Die Verwendung höherer Datenmengen zeigte keine besseren Resultate.

Die Ursachen sowohl für die niedrige Latenzzeit, als auch für den hohen Datendurchsatz liegen in der kurzen Übertragungsstrecke von ca. 60 km und in der Tatsache begründet, dass sich alle Knoten in einem Subnetz befinden, wodurch der Routing-Aufwand reduziert wird.

Im nächsten Schritt wurde die Anzahl der Streams, die sich gleichzeitig auf der 100-Gigabit-Strecke befinden, erhöht, indem der für die Punkt-zu-Punkt-Verbindung verwendete Messlauf auf mehrere Knoten verteilt wurde. Dabei konnten mit Hilfe der verfügbaren Monitoring-Funktion des Service Routers die Auslastung des 100-Gigabit-Ports und über BenchIT der Datendurchsatz jeder einzelnen Verbindung aufgezeichnet werden.

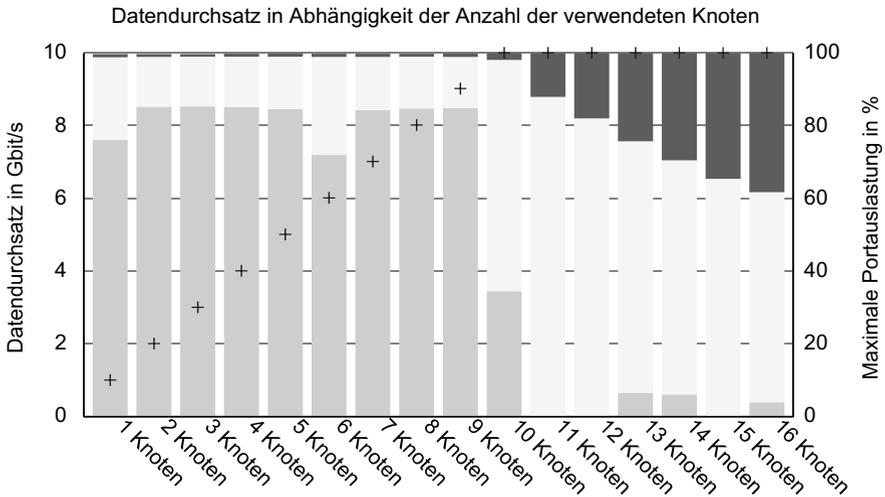


Abbildung 4: Minimaler (grün), durchschnittlicher (gelb) und maximaler (rot) Datendurchsatz, sowie die maximale Auslastung des 100-Gigabit-Ports (blau)

Das Ergebnis in Abbildung 4 zeigt wie die Auslastung des 100-Gigabit-Ports nahezu linear ansteigt, bis sich zehn parallele Streams auf der Leitung befinden. Ab diesem Punkt befindet sich die Strecke unter Überlast, was sich in der steigenden Varianz und dem sinkenden mittleren Datendurchsatz widerspiegelt. Die bereits zuvor sichtbare Differenz zwischen

minimalem und maximalen Datendurchsatz lässt sich mit Hilfe der Initialisierungsphase erklären, in der noch keine volle Datenrate erreicht werden kann.

4.2 Burst-Tests auf der Transportebene

Neben den klassischen Netzwerkparametern wurden, wie in Kapitel 3 beschrieben, auch Burst-Tests durchgeführt. Dabei erfolgte eine nichtlineare, aber reproduzierbare, Iteration über eine Liste, welche Nachrichtengrößen von 1 Byte bis zu 2 GByte enthielt. Die Zeitmessung erfolgte ohne Verbindungsauf- und -abbau und es wurden insgesamt zehn Wiederholung für jede Nachrichtengröße durchgeführt. Aus den daraus generierten Zeitstempeln wurde die minimale, mittlere und maximale Kommunikationszeit bestimmt. Diese Werte wurden zusammengetragen und am Ende des gesamten Messlaufs sortiert ausgegeben. Einen Ausschnitt von 10 MByte bis 2 GByte über 13 Knoten zeigt die Abbildung 5.

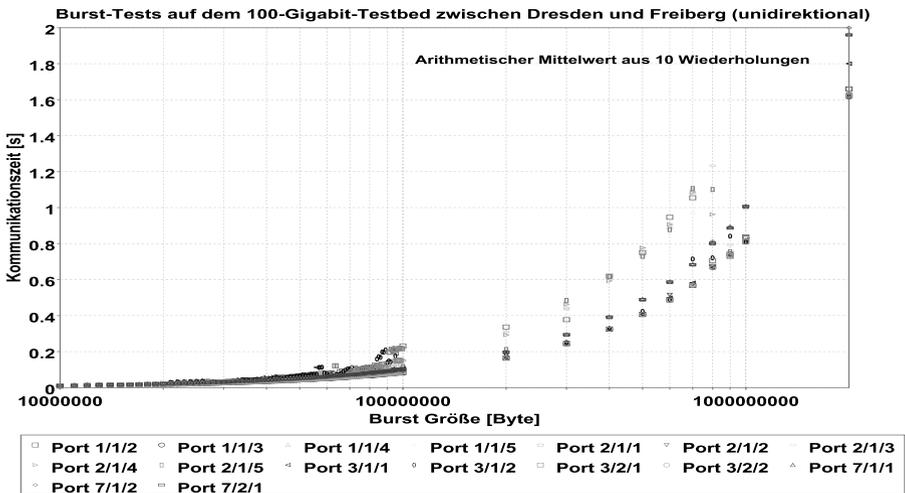


Abbildung 5: Durchschnittliche Kommunikationszeit in Abhängigkeit der übertragenen Datenmenge

Daraus wird ersichtlich, dass bis zu Burst-Größen von 30 MByte kaum Unterschiede in der Kommunikationszeit festzustellen sind, obwohl in Abschnitt 4.1 gezeigt wurde das unter Überlast einzelne Verbindungen benachteiligt werden. Die Erklärung hierfür liegt in der kurzen Übertragungsdauer, so dass die Kapazitäten der Leitung nahezu umgehend wieder zur Verfügung stehen. Bei größeren Datenmengen erhöht sich, wie bereits in Abbildung 4 gezeigt, ebenfalls die Varianz zwischen den einzelnen Verbindungen.

4.3 Burst-Tests auf der Anwendungsebene

Mit IOR wurden insgesamt 2672 Messungen durchgeführt. Die Burst-Größe variierte dabei von 8 Byte bis 4 GByte in unterschiedlichen Schrittgrößen. Es wurde nacheinander für 1-16 Prozesse die verschiedenen Blockgrößen getestet. Dabei wurden 26.173 GiByte an Nutz- und Metadaten von Dresden nach Freiberg gesendet und 75 GiByte an Metadaten in Dresden empfangen. Mit einem Prozess wurden im Durchschnitt 386 MiByte/s übertragen (maximal 468 MiByte/s). Mit vier Prozessen sank die durchschnittliche Rate über alle Burst-Größen auf 350 MiByte/s. Mit allen 16 Prozessen wurden durchschnittliche Datenraten von 280 MiByte/s (Prozess 1) bis 318 MiByte/s (Prozess 12) erzielt. Dass in den Messungen mit allen 16 Servern gerade jene Prozesse, welche in allen Messungen beteiligt waren, die schlechtesten Werte erzielten, lässt sich mit dem *fairness control* Modell von Lustre erklären ([FW09]) welches den Knoten mit weniger Transfervolumen eine höhere Priorität einräumt und somit für eine gleichbleibende Gesamtbandbreite sorgt. Über den kompletten Messzeitraum ergeben sich die in den Abbildungen 6 und 7 dargestellten Werte. In einer Kollaboration mit der Indiana University konnte gezeigt werden,

240 MiB/s	160 MiB/s	80 MiB/s	0 MiB/s
339.457 MiB/s	/lustre_freiberg/...ior/data.00000000		
334.564 MiB/s	/lustre_freiberg/...ior/data.00000001		
325.524 MiB/s	/lustre_freiberg/...ior/data.00000002		
319.548 MiB/s	/lustre_freiberg/...ior/data.00000003		
316.651 MiB/s	/lustre_freiberg/...ior/data.00000008		
315.753 MiB/s	/lustre_freiberg/...ior/data.00000005		
315.415 MiB/s	/lustre_freiberg/...ior/data.00000004		
315.251 MiB/s	/lustre_freiberg/...ior/data.00000012		
315.199 MiB/s	/lustre_freiberg/...ior/data.00000006		
314.598 MiB/s	/lustre_freiberg/...ior/data.00000011		
314.113 MiB/s	/lustre_freiberg/...ior/data.00000009		
313.697 MiB/s	/lustre_freiberg/...ior/data.00000007		
313.672 MiB/s	/lustre_freiberg/...ior/data.00000010		
310.866 MiB/s	/lustre_freiberg/...ior/data.00000013		
308.489 MiB/s	/lustre_freiberg/...ior/data.00000014		
304.513 MiB/s	/lustre_freiberg/...ior/data.00000015		

Abbildung 6: Durchschnittliche Datenraten für den gesamten Messzeitraum, entspricht 5 GiByte/s am 100-Gigabit-Port

320 MiB/s	160 MiB/s	0 MiB/s
468.295 MiB/s	/lustre_freiberg/...ior/data.00000000	
466.774 MiB/s	/lustre_freiberg/...ior/data.00000007	
466.189 MiB/s	/lustre_freiberg/...ior/data.00000001	
465.698 MiB/s	/lustre_freiberg/...ior/data.00000008	
465.43 MiB/s	/lustre_freiberg/...ior/data.00000013	
465.042 MiB/s	/lustre_freiberg/...ior/data.00000006	
464.269 MiB/s	/lustre_freiberg/...ior/data.00000005	
462.722 MiB/s	/lustre_freiberg/...ior/data.00000011	
462.704 MiB/s	/lustre_freiberg/...ior/data.00000004	
462.591 MiB/s	/lustre_freiberg/...ior/data.00000009	
462.507 MiB/s	/lustre_freiberg/...ior/data.00000002	
462.435 MiB/s	/lustre_freiberg/...ior/data.00000003	
462.406 MiB/s	/lustre_freiberg/...ior/data.00000012	
461.167 MiB/s	/lustre_freiberg/...ior/data.00000014	
460.017 MiB/s	/lustre_freiberg/...ior/data.00000010	
459.827 MiB/s	/lustre_freiberg/...ior/data.00000015	

Abbildung 7: Maximale Datenraten für den gesamten Messzeitraum, entspricht 7,2 GiByte/s am 100-Gigabit-Port

dass auf der Lustre Netzwerk Schicht, d.h. ohne I/O-Subsystem, *lnet* 94.4% der 100Gbit erreichbar sind [Uni10], dies entspricht 11,79 GByte/s. In einer hoch optimierten Messung konnte mit dem IOR Benchmark beim Schreiben von Dresden auf das Freiburger SFA10K im Spitzenwert 10,8 GByte/s übertragen werden. Dabei wurden allerdings die Lustre Parameter optimiert, ein passendes LFS Striping über alle 3*16 OST's verwendet und der Benchmark selbst lief auf einem über 40Gbit QDR InfiniBand Links an die lokalen 16 HP-Server angeschlossenen Cluster mit bis zu 720 IOR Clients. Die 16 HP Server wurden in dieser Konfiguration als reine *lnet* Router eingesetzt.

5 Zusammenfassung & Ausblick

Die hier vorgestellten Ergebnisse zeigen bereits das Potential der neu eingeführten 100-Gigabit-Technologie. Daraus ergeben sich eine Vielzahl von Anwendungsmöglichkeiten, zum Beispiel die Durchführung dezentraler Backups über große Entfernungen oder die Verbindung von Rechenzentren zur gemeinsamen Nutzung. Dieses Prinzip ist bereits beim Norddeutschen Verbund für Hoch- und Höchstleistungsrechnen (HLRN) mit einer geringeren Übertragungsrates in Anwendung.

Anfang 2011 fand eine Streckenverlängerung auf 200 und 400 km statt. Die dadurch steigenden Latenzzeiten machten eine weitere TCP-Parameteroptimierung erforderlich, um nach wie vor eine volle Linkauslastung zu erreichen. Auch bei den I/O-Messungen waren die bei 60 km im Vergleich zur IO-Hardware vernachlässigbaren Netzwerklatenzen nun ein signifikanter Einfluss. Diese Arbeit diente zur Vorbereitung der Tests auf den längeren Strecken und soll einen Vergleich zwischen den verschiedenen Konfigurationen ermöglichen.

Literatur

- [Alc10a] Alcatel-Lucent. *Alcatel-Lucent 1830 PSS (PSS-32 and PSS-16) - Photonic Service Switch*, 2010.
- [Alc10b] Alcatel-Lucent. *Alcatel-Lucent 7750 SR - Service Router*, 2010.
- [Dat10a] Data Direct Networks, <http://ddn.com/9900>. *Data Direct Networks Silicon Storage Architecture S2A9900*, 2010.
- [Dat10b] Data Direct Networks, <http://ddn.com/10000>. *Data Direct Networks Storage Fusion Architecture SFA10000*, 2010.
- [FW09] Galen Shipman Feiyi Wang, Sarp Oral. Understanding Lustre Filesystem Internals. *Oak Ridge National Laboratory Reporter*, TM-2009(117), 2009.
- [GJ03] Stefan Börner et al. Guido Juckeland. BenchIT - Performance Measurement and Comparison for Scientific Applications. In *PARCO2003 Proceedings*, 2003.
- [Hew10] Hewlett-Packard. *HP ProLiant DL160 G6 Server - Datasheet*, Feb 2010.
- [NLC08] Arifa Nisar, Wei-keng Liao und Alok Choudhary. Scaling parallel I/O performance through I/O delegate and caching system. In *Proceedings of the 2008 ACM/IEEE conference on Supercomputing*, SC '08, pages 9:1–9:12, Piscataway, NJ, USA, 2008. IEEE Press.
- [Uni10] Indiana University. Indiana University Announces Saturation of First Commercial 100 Gigabit Link. <http://www.hpcwire.com/offthewire/Indiana-University-Announces-Saturation-of-First-Commercial-100-Gigabit-Link-108251309.html>, 11 2010.
- [WMM10] Thomas William, Hartmut Mix und Bernd Mohr. Enhanced Performance Analysis of Multi-Core Applications with an Integrated Tool-Chain. In B. Chapman et al., Herausgeber, *Parallel Computing: From Multicores and GPU's to Petascale*, volume 19, pages 701–708. IOS Press, 2010.