

Universität Rostock

Fakultät für Ingenieurwissenschaften
Fachbereich Elektrotechnik und Informationstechnik
Institut für Nachrichtentechnik und Informationselektronik



Studienarbeit

Modellierung und Simulation von Ethernet-Netzwerkverkehr

Bearbeiter:

cand. ing. Christian Groth
cand. ing. Jens Kosubek

Betreuer:

Dr.-Ing. H.-D. Melzer
Dipl.-Ing. Th. Kessler

Datum:

05.03.2004

Gliederung

1. Motivation	10
2. Einführung.....	11
2.1 Aufgabenstellung und Herangehensweise	11
2.2 Grundlagen	12
2.2.1 Stationarität.....	12
2.2.2 Long Range Dependence (LRD)	12
2.2.3 Parameter	14
2.2.3.1 Hurst-Parameter	14
2.2.3.2 Fractal Onset Time Scale	15
2.2.4 Selbstähnlichkeit.....	16
2.3 Entstehung von Long Range Dependence	17
2.4. Spezielle Verteilungen	18
2.4.1 Die Exponentialverteilung.....	18
2.4.2 Die Weibullverteilung.....	20
2.4.3 Pareto-Verteilung.....	21
2.4.4 Kurzübersicht über die betrachteten Verteilungen	22
3. Methoden zur Schätzung des Hurst Parameters	23
3.1 Variance-Time-Plot (VTP).....	23
3.2 R/S-Statistik (Rescaled Adjusted Range)	25
3.3 Periodogramm	26
3.4 Whittle-Estimator	26
4 Analyse der Meßreihen.....	28
4.1 Berechnung der Bins	28
4.2 Bestimmung der Ober- und Untergrenzen	29
4.2.1 Mittels Mittelwertbildung.....	30
4.2.2 Mittels Wahrscheinlichkeitsdichtverteilung	31
4.2.3 Verwendungsmöglichkeiten der Methoden.....	32
4.3 Betrachtete Meßreihen	33
4.4 Logarithmische Optimierung (Log-Opt).....	34
4.5 Berechnung des Hurst Parameters aus Zwischenankunftszeiten.....	34
4.6 Untersuchung auf generelle Langzeitabhängigkeit.....	37
4.7 Bestimmung des Hurst-Parameters unter verschiedenen Belastungssituationen	42
4.7.1 Vollast.....	43
4.7.2 Niedriglast.....	44
4.8 Abhängigkeit des Hurst-Parameters von der Größe des Grundintervalls.....	46
4.9 Abweichungen in den Meßreihen	48
4.10 Simulationsvorbereitende Analysen	51
4.10.1 Analyse des Datensatzes uni_1h_21.02.03_11:52_in.....	51
4.10.2 Analyse der Datensätze uni_1h_22.02.03_01:43_in und	

uni_1h_22.02.03_01:43_out	52
4.11 Analyse der Proben mittels des Periodogramms	54
4.12 Anwendung des Whittle-Estimators	59
5. Simulation in OPNET	60
5.1 Fraktale PunktProzesse (FPP - Fractal Point Process).....	60
5.1.1 FRP (Fractal Renewal Process)	61
5.1.2 Sup-FRP	61
5.1.3 Fraktal Modulierter Poisson Prozeß (FMPP).....	62
5.1.4 On-Off FMPPs	62
5.2 Auswahl eines geeigneten Modells	63
5.3 Aufbau eines geeigneten Netzwerks	64
5.4 Simulation des eingehenden Verkehrs unter Vollast	65
5.4.1 Simulation mit einer Bingröße von 1000msec	66
5.4.2 Simulation mit 100msec	69
5.5 Simulation des eingehenden und des ausgehenden Verkehrs unter Niedriglast	71
6 Analyse der Zwischenankunftszeiten und Simulation.....	74
6.1 Gewinnung der Parameter für die Simulation in OPNET	75
6.3 Analyse der Proben mittels empirischer Bestimmung von Wahrscheinlich- keitsverteilung und Wahrscheinlichkeitsdichteverteilung	78
6.4 Prüfung der Simulationsergebnisse auf Selbstähnlichkeit.....	83
6.5 Erzeugung selbstähnlicher Folgen mittels der Paretoverteilung.....	84
6.6 Weitere Auswertung einer RPG-Simulation.....	88
7. Schlußbetrachtung	90
8. Quellen	91
Anhang	93

Abkürzungsverzeichnis

AFRP	Alternating Fractal Renewal Process
AKF	Autokorrelationsfunktion
ATM	Asynchronous Transfer Mode
CGI	Common Gateway Interface
CSMA/CD	Carrier Sense Multiple Access / Collision Detection
DSPP	Doubly Stochastic Poisson Process
FARIMA	Fractional Autoregressive Integrated Moving-Average
FBNDP	Fractal Binomial Noise Driven Poisson Process
FGN	Fractional Gaussian Noise
FFT	Fast Fourier Transform
FMPP	Fraktal Modulierter Poissonprozess
FOTS	Fractal Onset Time Scale
FPP	Fraktaler Punktprozess
FRP	Fractal Renewal Process
FSNDP	Fractal Shot Noise Driven Poisson Process
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
IP	Internet Protocol
IPX	Internetwork Packet Exchange
ISO	International Organization for Standardization
LRD	Long Range Dependence
MSE	Mean Square Error
MTU	Maximum Transmission Unit
OSI	Open Systems Interconnection
PSD	Power Spectral Density
RIP	Routing Information Protocol
RPG	Raw Packet Generator
SAR	Source Activity Ratio
SRD	Short Range Dependence
TCP	Transfer Control Protocol
VTP	Variance Time Plot
WkDV	Wahrscheinlichkeitsdichteverteilung
WkV	Wahrscheinlichkeitsverteilung

Abbildungsverzeichnis

Abbildung 2.1	Vergleich einer Poisson verteilten Größe mit einer gemessenen Größe bei größer werdenden Aggregationsstufen.....	13
Abbildung 2.2	Gegenüberstellung Weibullverteilung – Poissonverteilung.....	14
Abbildung 2.3	Abhängigkeit der Korrelationskoeffizienten von α	16
Abbildung 2.4	WkDV und WkV einer Exponentialverteilung.....	19
Abbildung 2.5	WkDV und WkV einer Weibullverteilung.....	20
Abbildung 2.6	WkDV und WkV einer Paretoverteilung.....	21
Abbildung 4.1	Darstellung der Originalfolge von Zwischenankunftszeiten.....	28
Abbildung 4.2	Einteilung der Zwischenankunftszeiten zum entsprechenden Intervall...	28
Abbildung 4.3	Darstellung der resultierenden Bins.....	29
Abbildung 4.4	Erklärung zur Schätzung der Bereichsgrenzen durch Bildung von Mehreren Mittelwerten.....	30
Abbildung 4.5	Wahrscheinlichkeitsdichteverteilung des Datensatzes uni_1h_21.02.03_11:52_in mit Grundintervall von 100msec.....	31
Abbildung 4.6	Markierung eines exemplarischen Schwellenwertes und des daraus resultierenden Bereichs.....	31
Abbildung 4.7	Vergleich zweier Wahrscheinlichkeitsdichteverteilungen.....	32
Abbildung 4.8	Blockberechnungszeiten in Abhängigkeit von der Blockgröße.....	35
Abbildung 4.9	Gesamtberechnungszeiten in Anhängigkeit von der Blockgröße.....	35
Abbildung 4.10	R/S-Statistik der Zwischenankunftszeiten (eine unvollständige Berechnung).....	36
Abbildung 4.11	R/S-Statistik der Bins, uni_1h_out.....	38
Abbildung 4.12	Variance Time Plot der Bins, uni_1h_out.....	39
Abbildung 4.13	R/S-Statistik der Bins, uni_1h_in.....	40
Abbildung 4.14	Variance Time Plot der Bins, uni_1h_in.....	41
Abbildung 4.15	resultierende Graphen aus der Analyse, Vollast-Situation.....	43
Abbildung 4.16	resultierende Graphen aus der Analyse, Niedriglast-Situation.....	45
Abbildung 4.17	R/S-Statistik mit min. Grundintervall, uni-1h-in.....	47
Abbildung 4.18	R/S-Statistik mit Grundintervall von 100.000 μ s.....	47
Abbildung 4.19	Zeitliche Verteilung der größten Zwischenankunftszeiten, uni_1h_in und uni_1h_out.....	48
Abbildung 4.20	Vergrößerung mit geschätzten Geraden, uni_1h_in.....	48
Abbildung 4.21	Zeitliche Verteilung der jeweils größten Zwischenankunftszeiten, Vollast und Niedriglast.....	50
Abbildung 4.22	Variance Time Plot bei verschiedenen Binsgrößen (100ms und 1000ms).....	51

Abbildung 4.23	Variance Time Plot für die Datensätze uni_1h_22.02.03_01:43_in und uni_1h_22.02.03_01:43_out bei einer Bingröße von 1000ms.....	53
Abbildung 4.24	Periodogramm der empfangenen Pakete (Probe 1, Bingröße 12481 μ s)...	54
Abbildung 4.25	Periodogramm der empfangenen Pakete (Probe 1, Bingröße 1s).....	55
Abbildung 4.26	Periodogramm der gesendeten Pakete (Probe 1, Bingröße 16760 μ s).....	55
Abbildung 4.27	Periodogramm der gesendeten Pakete (Probe 1, Bingröße 1s).....	56
Abbildung 5.1	Einteilung der FPPs.....	60
Abbildung 5.2	Zusammenhang zwischen FRP und AFRP.....	62
Abbildung 5.3	Schematischer Aufbau der Simulationsumgebung.....	64
Abbildung 5.4	Darstellung der Bins der zu simulierenden Folge (eingehender Verkehr unter Vollast, Bingröße 1s).....	66
Abbildung 5.5	Darstellung der Bins der simulierten Folge bei verschiedenen Werten für den Parameter T.....	67
Abbildung 5.6	Darstellung der ermittelten Bereichsgrenzen bei einer Bingröße von 1000ms.....	68
Abbildung 5.7	Darstellung der ermittelten Bereichsgrenzen ei einer Bingröße von 100ms.....	70
Abbildung 5.8	Vergleich der ermittelten Bereichsgrenzen (eingehend und ausgehend)..	72
Abbildung 5.9	Variance Time Plot einer Originalfolge und der zugehörigen simulierten Folge.....	73
Abbildung 6.1	Vergleich von eingehendem und ausgehendem Verkehr mit Weibull-WkV (Probe1).....	76
Abbildung 6.2	Vergleich von eingehendem und ausgehendem Verkehr mit Weibull-WkV (Probe2).....	76
Abbildung 6.3	Vergleich von eingehendem und ausgehendem Verkehr mit Weibull-WkV (Probe3).....	77
Abbildung 6.4	WkDV und WkV bei Aggregationgröße 1s, empfangseitig (Probe1).....	81
Abbildung 6.5	WkDV und WkV des simulierten Empfangsprozesses.....	81
Abbildung 6.6	WkDV und WkV der Probe 1 (sendeseitig).....	82
Abbildung 6.7	WkDV und WkV des simulierten Empfangsprozesses.....	82
Abbildung 6.8	Periodogramm des simulierten Empfangsprozesses (Probe 1).....	83
Abbildung 6.9	Vergleich empfangener(bzw. gesendeter) Verkehr mit Paretoapproximation.....	84
Abbildung 6.10	WkDV und WkV des simulierten Ankunftsprozesses (Probe 2).....	85
Abbildung 6.11	WkDV und WkV des simulierten Sendeprozesses (Probe 3).....	86
Abbildung 6.12	Periodogramm des simulierten Empfangsprozesses.....	87
Abbildung 6.13	Periodogramm des simulierten Sendeprozesses.....	87
Abbildung 6.14	WkDV und WkV des mittels des RPG simulierten Zufallsprozesses.....	89
Abbildung 6.15	Periodogramm der mittels des RPG simulierten Folge (FOTS = 0,05)....	89

Tabellenverzeichnis

Tabelle 2.1	Zusammenfassung der Verteilungen.....	22
Tabelle 4.1	einige Eckdaten der untersuchten Datensätze.....	33
Tabelle 4.2	Beispiel zur Erläuterung (log-opt).....	34
Tabelle 4.3	Eigenschaften der Datensätze uni_1h_in und uni_1h_out.....	37
Tabelle 4.4	Zusammenfassung der ermittelten Daten (uni_1h_in, uni_1h_out).....	42
Tabelle 4.5	Eigenschaften der Datensätze unter Niedriglast und unter Vollast.....	42
Tabelle 4.6	Hurst-Parameter ermittelt durch VTP, Vollast.....	44
Tabelle 4.7	Zusammenfassung der Ergebnisse, Vollast.....	44
Tabelle 4.8	Hurst-Parameter ermittelt durch VTP, Niedriglast.....	45
Tabelle 4.9	Zusammenfassung der Ergebnisse, Niedriglast.....	46
Tabelle 4.10	Zusammenfassung der gemessenen Zeiten und der geschätzten Hurst-Parameter.....	46
Tabelle 4.11	Zusammenfassung der Stufen und deren Abstände.....	49
Tabelle 4.12	Zusammenfassung der ermittelten Parameter (simulationsvorbereitende Analyse, eingehender Verkehr unter Vollast).....	52
Tabelle 4.13	Zusammenfassung der ermittelten Parameter (simulationsvorbereitende Analyse, eingehender und ausgehender Verkehr unter Niedriglast).....	53
Tabelle 4.14	wichtige Eigenschaften der Probe 1.....	54
Tabelle 4.15	Zusammenfassung der Ergebnisse aus der Schätzung mittels Periodogramm (Probe 1).....	56
Tabelle 4.16	wichtige Eigenschaften der Probe 2.....	57
Tabelle 4.17	Zusammenfassung der Ergebnisse aus der Schätzung mittels Periodogramm (Probe 2).....	57
Tabelle 4.18	wichtige Eigenschaften der Probe 2.....	58
Tabelle 4.19	Zusammenfassung der Ergebnisse aus der Schätzung mittels Periodogramm (Probe 2).....	58
Tabelle 4.20a	Zusammenfassung der Ergebnisse aus den Schätzungen mittels Whittle-Estimator (empfangsseitig).....	59
Tabelle 4.20b	Zusammenfassung der Ergebnisse aus den Schätzungen mittels Whittle-Estimator (sendeseitig).....	59
Tabelle 5.1	Benötigte Parameter der verschiedenen Prozesse.....	63
Tabelle 5.2	Zusammenfassung der Netzwerkeinstellungen.....	64
Tabelle 5.3	Festlegungen in der statischen Routingtabelle.....	65
Tabelle 5.4	Zusammenfassung der ermittelten Eingangsparameter (Bingröße = 1s).....	66
Tabelle 5.5	Zusammenfassung der Simulationsergebnisse (Bingröße = 1s, 3600 Bins)....	68

Tabelle 5.6	Zusammenfassung der ermittelten Eingangsparameter (Bingröße = 100ms)..	69
Tabelle 5.7	Zusammenfassung der Simulationsergebnisse (Bingröße = 100ms, 36000 Bins).....	70
Tabelle 5.8	Zusammenfassung der ermittelten Eingangsparameter (eingehender und ausgehender Verkehr unter Niedriglast, Bingröße = 1s).....	71
Tabelle 5.9	Zusammenfassung der Simulationsergebnisse (Niedriglast, Bingröße: 1000ms, eingehender Verkehr).....	71
Tabelle 5.10	Zusammenfassung der Simulationsergebnisse (Niedriglast, Bingröße: 1000ms, ausgehender Verkehr).....	72
Tabelle 6.1	Ergebnisse der simulationsvorbereitenden Analyse (empfangsseitig).....	75
Tabelle 6.2	Ergebnisse der simulationsvorbereitenden Analyse (sendeseitig).....	75
Tabelle 6.3	Einstellungen für die Simulation.....	77
Tabelle 6.4	Zusammenfassung der simulationsvorbereitenden Analysen (Proben 1-3, empfangsseitig).....	79
Tabelle 6.5	Zusammenfassung der Simulationsergebnisse (Proben 1-3, empfangsseitig).....	79
Tabelle 6.6	Zusammenfassung der simulationsvorbereitenden Analysen (Proben 1-3, sendeseitig).....	80
Tabelle 6.7	Zusammenfassung der Simulationsergebnisse (Proben 1-3, sendeseitig).....	80
Tabelle 6.8	Vergleich der ermittelten Hurst-Parameter.....	83
Tabelle 6.9	Zusammenfassung der Ergebnisse aus der Analyse.....	85
Tabelle 6.10	Einstellungen für die Simulation.....	85
Tabelle 6.11	Vergleich der Ergebnisse aus der Simulation und der Analyse.....	86
Tabelle 6.12	Auswertung einer mittels RPG durchgeführten Simulation.....	88

1. Motivation

Die Suche nach Modellen, die Netzwerkverkehr zuverlässig und genau beschreiben, gewinnt immer mehr an Bedeutung: zum einen, um Planungssicherheit für die zukünftig nötigen Kapazitäten zu erhalten; zum anderen, um die vorhandene Infrastruktur optimal auszulasten. Aufgrund ihrer Einfachheit und Zuverlässigkeit bei der Modellierung der Auslastung von Leitungen in öffentlichen Telefonnetzen stellen Poissonprozesse das beliebteste Werkzeug zur Modellierung leitungsvermittelter Netze dar. Die Gedächtnislosigkeit der zugrundeliegenden Exponentialverteilung ist bei der Betrachtung zellen- und paketvermittelter Netze jedoch nicht mehr gegeben; vielmehr deuten die Meßergebnisse von ATM- und Ethernetverkehr auf Langzeitabhängigkeit (sog. *Long Range Dependence*) hin. Hervorstechendstes Merkmal ist ihre fraktale Natur, die mittels zweier Größen - des *Hurst-Parameters* sowie der *Fractal Onset Time Scale* - beschreibbar ist. Im Gegensatz hierzu zeigen Prozesse, die der sog. *Short Range Dependence* unterliegen, kein fraktales Verhalten; sie ändern also ihr Erscheinungsbild über wechselnde Zeitskalen. Zu letzteren zählen exponentialverteilte und Poissonprozesse. Es hat sich gezeigt, daß die Ankunftsprozesse in paketvermittelten Netzen eher einer Weibullverteilung, welche eine Verallgemeinerung der Exponentialverteilung darstellt, folgen [F01]. Insbesondere bedeutet dies, daß die Wahrscheinlichkeit extremer Schwankungen der Ankunftszeiträume aufeinanderfolgender Pakete gegenüber Poissonprozessen bedeutend größer ist. Auf einen Zeitraum von Niedriglast kann also abrupt eine Überlastsituation folgen bzw. umgekehrt. Dies schlägt sich durch die sog. *Burstiness* in der graphischen Darstellung der Ankunftsprozesse in paketvermittelten Netzen nieder. Das Auftreten dieses höchst unerwünschten Phänomens, welches auch nach Multiplexierung mehrerer Quellen in paketvermittelten Netzen weiterbesteht, bringt deutliche Komplikationen bei der Gewährleistung von aus leitungsvermittelten Netzen bekannten und gewohnten Dienstmerkmalen mit sich und erfordert komplizierte Queuing-Strategien. Da es nicht möglich ist, *Burstiness* zu vermeiden, konzentrieren sich die Anstrengungen auf die Ausnutzung der *Long Range Dependence*, um die Auslastung der zur Verfügung stehenden Bandbreite optimal zu halten. Voraussetzung dafür ist die Reproduzierbarkeit des realen Verkehrsverhaltens. In dieser Arbeit werden drei unter verschiedenen Lastsituationen genommene Proben aus dem Backbone der Universität Rostock auf die Merkmale von Langzeitabhängigkeit und ihre statistische Verteilungsstruktur untersucht. Darüberhinaus werden die Ergebnisse der Simulationen, die mit den gewonnenen Erkenntnissen mittels des Programms OPNET durchgeführt wurden, vorgestellt.

2. Einführung

Kapitel 2 wird die Problemstellung dieser Studienarbeit erörtern sowie einen Überblick über die gängigen Termini der damit zusammenhängenden Thematik geben. Einige Gründe für Langzeitabhängigkeit in paketvermittelten Kommunikationsnetzen werden ebenfalls angegeben. Die für diese Arbeit relevanten statistischen Verteilungen werden abschließend kurz vorgestellt.

2.1 Aufgabenstellung und Herangehensweise

Die konkrete Aufgabe dieser Studienarbeit besteht in der Analyse von Timestamps, welche aus dem Backbone der Universität Rostock stammen. Das Ergebnis der drei Meßvorgänge bestand aus jeweils eine Stunde umfassenden Aufzeichnungen der Empfangs- bzw. Sendezeiten der externen Universitätsverbindung. Die Ethernet-Rahmen wurden über einen separaten Port des Switches abgegriffen und ihre Ankunftszeit im 64-Bit-Festkommaformat registriert. Die Auflösung des Meßgeräts beträgt 2^{-32} s.

Nähere Informationen zum verwendeten Messsystem sind in [KMV03] zu finden.

Ziel der Arbeit ist es, den Hurst-Parameter der so erhaltenen Folgen mit möglichst vielen Verfahren zu bestimmen sowie die Verteilungsstruktur der Proben zu analysieren, um dann diese Prozesse mit den gewonnenen Ergebnissen im Programm OPNET zu simulieren. Die gewählte Vorgehensweise stellt sich folgendermaßen dar:

- Aufbereitung der Timestamps zur Darstellung der Anzahl empfangener Pakete pro Zeiteinheit
- Verifikation der Merkmale von Langzeitabhängigkeit
- Ermittlung des Hurst-Parameters
- Bestimmung der Charakteristika der empirischen statistischen Verteilung und Berechnung der entsprechenden Parameter analytischer Verteilungen
- Simulation von ausgewählten Ankunftsprozessen mit den ermittelten Parametern im Programm OPNET
- Vergleich und Diskussion

2.2 Grundlagen

Dieser Abschnitt befaßt sich mit den grundlegenden Begriffen aus der Welt der Selbstähnlichkeit. Auf die zwei wichtigsten Parameter zur Beschreibung selbstähnlichen Verhaltens wird ebenfalls eingegangen.

2.2.1 Stationarität

Ein Zufallsprozeß ist stationär, wenn seine Wahrscheinlichkeitsverteilungsfunktion invariant gegenüber absoluten Zeitverschiebungen ist. Es wird zwischen Stationarität im engeren und weiteren Sinn unterschieden. Die mathematische Definition für *Stationarität im engeren Sinn* lautet:

$$F_t(x) = F_{t+h}(x) \quad \forall -\infty \leq h \leq \infty \quad . \quad [2.1]$$

Die Verteilungsfunktion bleibt also unabhängig von einer Verschiebung des Beobachtungsintervalls. Damit bleiben sowohl Trendfunktion (jeweilige Mittelwerte der beobachteten Größe in den n Abschnitten) als auch die Varianzen konstant.

$$\begin{aligned} \mu(t) &= E(X(t)) \equiv \mu \\ \sigma^2(X(t)) &= \text{const} \end{aligned} \quad [2.2]$$

Es ist zu beachten, daß die Umkehrung i. allg. nicht gilt.

In der Praxis ist es zumeist nicht möglich, den betrachteten Prozeß auf Stationarität im engeren Sinn zu überprüfen, deshalb schwächt der Begriff der *Stationarität im weiteren Sinn* die Bedingung [2.1] zu [2.2] ab. Stationäre Prozesse sind deutlich einfacher zu handhaben als instationäre; auch in dieser Studienarbeit wird deshalb grundsätzlich Stationarität vorausgesetzt, obwohl diese Prämisse nur annähernd gilt.

2.2.2 Long Range Dependence (LRD)

Die Long Range Dependence ist auch als *Langzeitabhängigkeit*, *Joseph-Effect* oder *Heavy-Tailedness* bekannt. Diese Erscheinung kennzeichnet das Verkehrsverhalten in paketvermittelten Datennetzen. Sie ist durch starke Schwankungen der Ankunftszeiten von Paketen gekennzeichnet, welche unter der Voraussetzung der Stationarität des Ankunftsprozesses der Exponentialverteilung nicht oder nur unzureichend folgen¹. Im Falle exponentialverteilter Ankunftszeiten ergibt die Zusammenfassung der eingetroffenen Pakete in festen Zeitintervallen (Bins, vgl. Kap. 4.1.) eine Poissonverteilung. Ein Zufallsprozeß ist langzeitabhängig, wenn k (k=2,3,...,n) Werte aus der Meßreihe durch Summation zu einer kürzeren Reihe zusammengefaßt werden, sich die gebildete Autokorrelationsfunktion (AKF) nicht ändert. Die Autokorrelationsfunktion eines langzeitabhängigen Prozesses ist - abhängig vom Hurst-Parameter (s. 2.2.3.1.) - durch eine lineare Abnahme unterschiedlicher Ausprägung gekennzeichnet (vgl. Graph 2.3). Die Summation unendlich vieler

Autokorrelationsfunktionen wird unendlich: $\sum_{k=1}^{\infty} r(kT_s) = \sum_{k=1}^{\infty} r(X_k) = \infty$.

Jedes X_k repräsentiert die Anzahl der Ereignisse (z.B. Anzahl der empfangenen Pakete) zwischen den Zeitpunkten k und k+1. Dies impliziert $r(k) \sim k^{-(\alpha-1)}$ für $k \rightarrow \infty$ und $0 < \alpha < 1$. Da in der Realität nur endlich viele Realisierungen existieren, wird die Autokorrelationsfunktion nur endliche Werte erreichen und mit fortschreitendem k abnehmen. Im Unterschied zu z.B. exponentialverteilten Zufallsprozessen, die sog. *Short Range Dependence* aufweisen (Summe der AKF nimmt exponentiell ab), beobachtet man hier nur

¹ Es ist möglich, die Exponentialverteilung mittels eines instationären Ansatzes besser an die empirischen Verteilungen der Zwischenankunftszeiten anzunähern. Das Resultat reicht in [F01] jedoch nicht an die Weibullverteilung heran.

einen polynomialen Rückgang, der entsprechend flacher ausfällt.

Im Gegensatz zu exponentialverteilten Ankunftszeiten bewirken bei langzeitabhängigen Prozessen weder eine Vergrößerung des Grundintervalls noch die Multiplexierung mehrerer gleichgearteter Quellen eine „Glättung“ des Verlaufs. Als Konsequenz hieraus ist eine Modellierung langzeitabhängiger stationärer Zufallsprozesse mittels der Exponentialverteilung nur über „kurze“ Zeiträume möglich. Abb. 2.1 veranschaulicht dieses Phänomen.

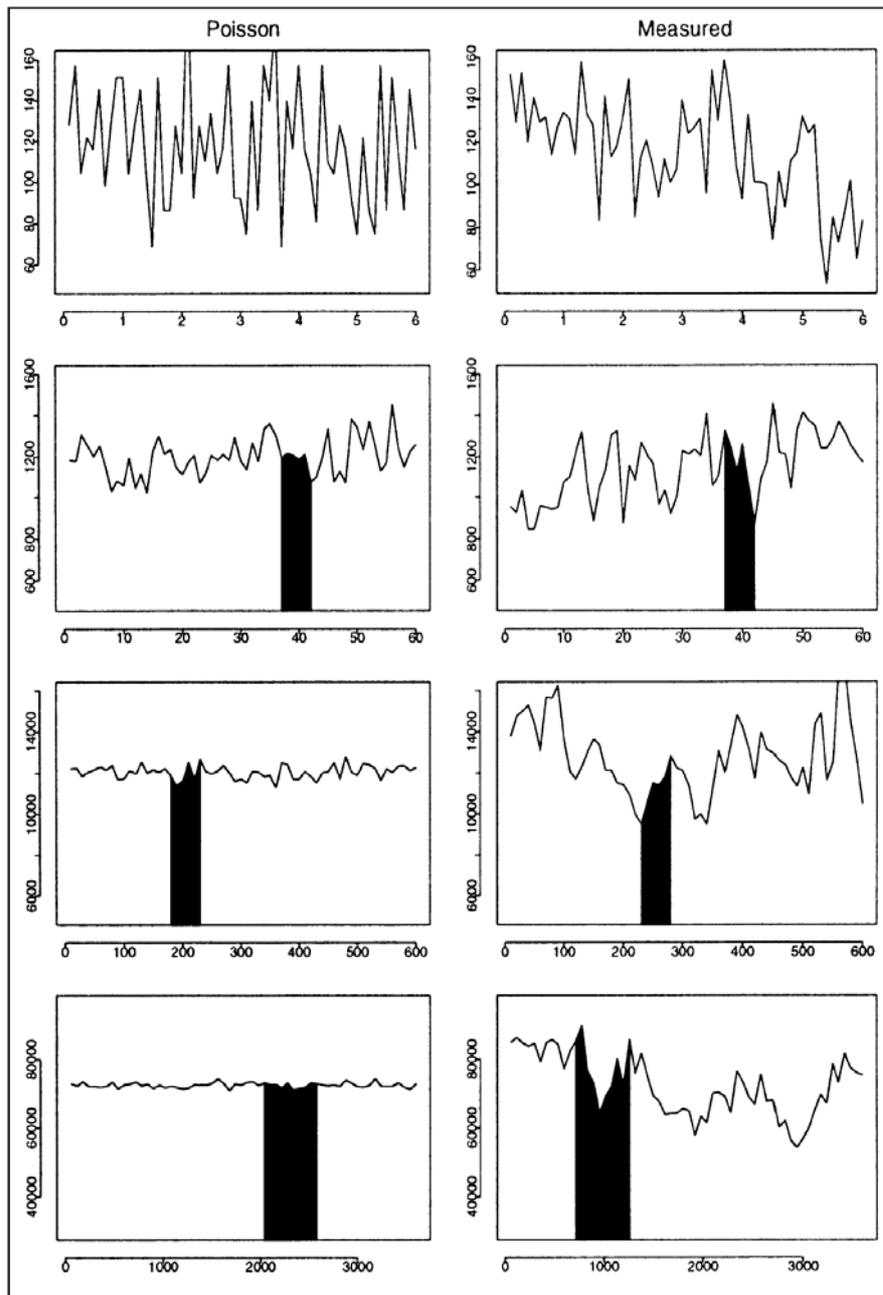


Abbildung 2.1: Im Fall der poissonverteilten Größe ist in höheren Aggregationsebenen eine immer stärkere Glättung zu registrieren. Die gemessene Datenfolge wird von der Aggregation kaum beeinflusst. (Quelle: [WP98])

Vielmehr ähneln die gewonnenen empirischen Verteilungsfunktionen solchen, die im englischsprachigen Schrifttum als *Heavy-Tailed* bezeichnet werden. Hierzu zählen u.a. die Weibull- und die Paretoverteilung. Im Graphen der Wahrscheinlichkeitsdichteverteilung einer weibullverteilten Größe (vgl. Abb. 2.2) ist der „Schwanz“ im Vergleich zu einer Poissonverteilung deutlich zu erkennen.

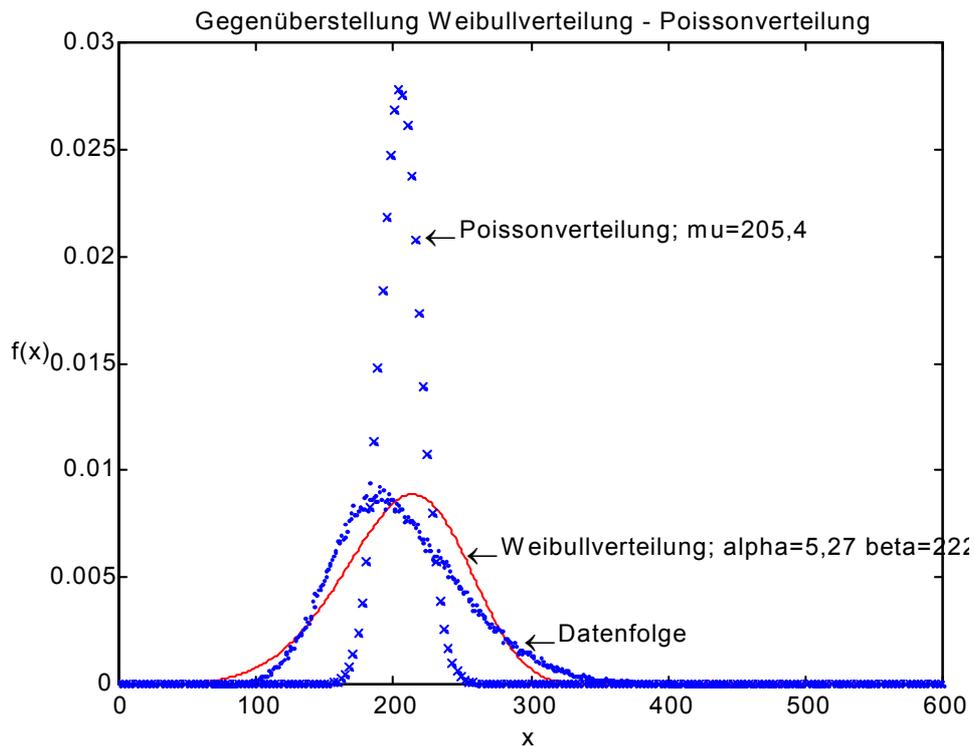


Abbildung 2.2: Dargestellt ist eine weibullverteilte Größe, die aus den untersuchten Messungen stammt. Eine Beschreibung mittels der Poissondichteverteilung weicht deutlich vom Originalgraphen ab.

2.2.3 Parameter

In diesem Kapitel werden für selbstähnliches Verhalten relevante Größen vorgestellt.

2.2.3.1 Hurst-Parameter

Der Grad der Selbstähnlichkeit wird maßgeblich vom sog. Hurst-Parameter H bestimmt, einem Zahlenwert, der zwischen Null (keine Selbstähnlichkeit) und Eins (je näher H an 1 liegt, desto höher ist der Grad der Selbstähnlichkeit) liegt. Insbesondere zu Simulationen zwecken ist seine Kenntnis erforderlich, da er die wichtigste Eingangsgröße für diverse Simulationswerkzeuge bildet, welche auf die in OPNET enthaltenen und in dieser Arbeit betrachteten fraktalen Erneuerungsprozesse (vgl. 5.1.1) aufsetzen. Zur Bestimmung von H existiert eine Vielzahl von graphischen und analytischen Methoden, z.B.:

- Variance-Time Plot (graphisch)
- RS-Plot (graphisch)
- Periodogramm (graphisch)
- Whitte-Estimator (analytisch)

Ziel der Bestimmung dieses Parameters war die Nutzung des Hurst-Parameters als Eingangsgröße für die Simulation selbstähnlichen Netzwerkverkehrs im Netzwerkplanungs- und -analyseprogramm OPNET. In jedem der zur Verfügung gestellten Simulationsmodelle ist die Angabe dieser Größe erforderlich.

Die o.g. Verfahren kamen bei den insgesamt sechs Messproben zum Einsatz. Eine ausführliche Vorstellung und Bewertung ist in den jeweiligen Abschnitten enthalten (vgl. Kap. 3 und 4).

2.2.3.2 Fractal Onset Time Scale (FOTS)

Für die Kovarianz eines fraktalen Prozesses gilt bei seiner Unterteilung in k Teilprozesse (Realisierungen, Ausschnitte) der Länge T :

$$\begin{aligned} \text{Cov}(k, T) &= \lambda T \cdot \left(1 + \left(\frac{T}{T_0} \right)^\alpha \right) \quad k = 0 \\ \text{Cov}(k, T) &= \lambda T \cdot \left(\frac{T}{T_0} \right)^\alpha \cdot \left(\frac{\nabla^2(k^{\alpha+1})}{2} \right) \end{aligned} \quad (2.1); [\text{RL97}]$$

mit $\nabla^2(k^{\alpha+1}) = ((k+1)^{\alpha+1} - 2k^{\alpha+1} + (k-1)^{\alpha+1})$

$$\forall k \in \mathbb{N}; 0 < \alpha < 1$$

λ – Mittelwert

T_0 – FOTS

Des weiteren gilt

$$\text{Cov}(k, T) = \text{Cov}(T, kT)$$

Die Herleitung aus der allgemeinen Form ist ebenfalls in [RL97] zu finden. Die Variable α ist der fraktale Exponent. Für den Korrelationskoeffizienten ergibt sich

$$\begin{aligned} r(k, T) &= \frac{\text{Cov}(k, T)}{\text{Cov}(0, T)} = \frac{\left(\frac{T}{T_0} \right)^\alpha \cdot \left(\frac{\nabla^2(k^{\alpha+1})}{2} \right)}{1 + \left(\frac{T}{T_0} \right)^\alpha} \\ &= \frac{\left(\frac{T}{T_0} \right)^\alpha \cdot \left(\frac{\nabla^2(k^{\alpha+1})}{2} \right)}{\left(\frac{T_0}{T_0} \right)^\alpha + \left(\frac{T}{T_0} \right)^\alpha} \quad (2.2); [\text{RL97}] \\ &= \frac{T^\alpha}{T^\alpha + T_0^\alpha} \cdot \left(\frac{\nabla^2(k^{\alpha+1})}{2} \right); \quad \alpha = 2H - 1 \end{aligned}$$

H ist der Hurst-Parameter, T der Skalierungsfaktor. Die fraktale Charakteristik steckt im zweiten Teil der Gleichung, deren Graphen halblogarithmisch für verschiedene α in Abb. 2.3 dargestellt sind. Mit wachsenden α fallen die Graphen immer flacher ab. Von Interesse für den Vorfaktor sind die Extremfälle $T \ll T_0$ und $T \gg T_0$. Im zweiten Fall wird der Vorfaktor ca. 1, so daß die Graphen die Situation zutreffend wiedergeben. Sinkt jedoch T unter T_0 , ist kein fraktales Verhalten mehr beobachtbar. T_0 wird deshalb *Fractal Onset Time Scale* (FOTS) genannt. Unterhalb dieser „Zeitgrenze“ (*Time Scale*) zeigen fraktale Prozesse keine Selbstähnlichkeit.

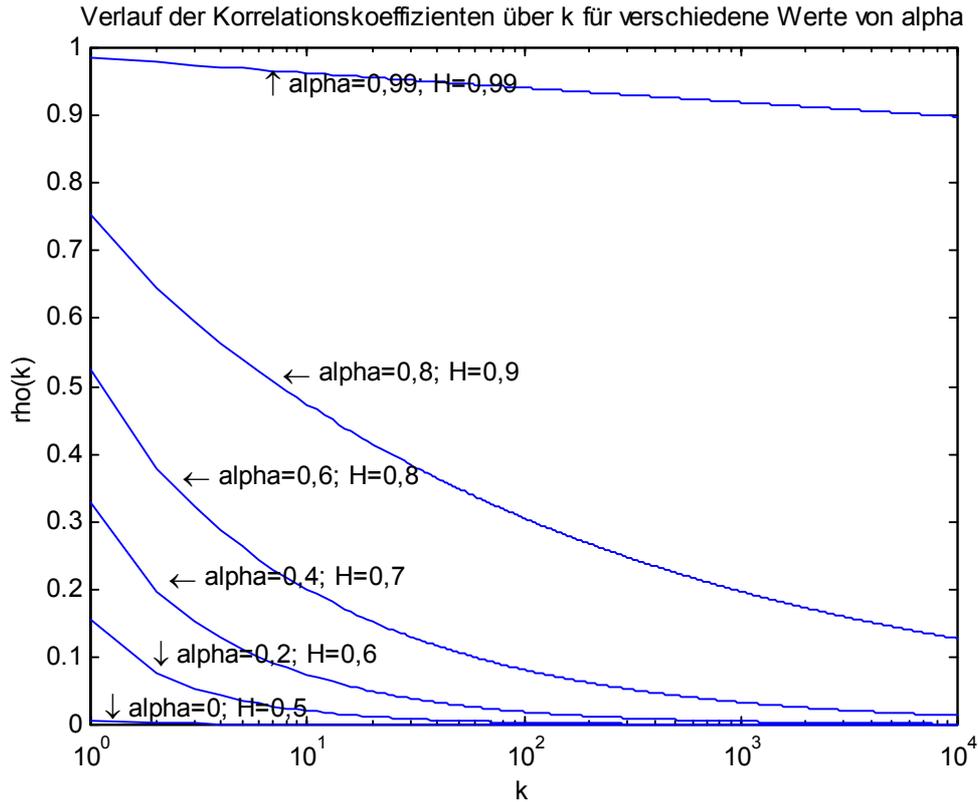


Abbildung 2.3: Abhängigkeit der Korrelationskoeffizienten von α

2.2.4 Selbstähnlichkeit

Selbstähnliche Zufallsprozesse bilden die einfachste Möglichkeit, langzeitabhängige Prozesse zu modellieren, da sie mittels nur einer Größe - des Hurst-Parameters - beschreibbar sind.

Es wird zwischen exakter und asymptotischer Selbstähnlichkeit unterschieden. Ein stochastischer Prozeß ist exakt selbstähnlich, wenn seine Autokorrelationsfunktion die Form $r^{(m)}(k, T) = r(k, mT)$ besitzt¹. Bei T handelt es sich um den Skalierungsfaktor aus vorigem Abschnitt. Einsetzen in Formel (2.2) liefert

$$r(k, mT) = \frac{(mT)^\alpha}{(mT)^\alpha + T_0^\alpha} \cdot \left(\frac{\nabla^2(k^{\alpha+1})}{2} \right) \quad (2.3)$$

Führt man den Grenzübergang für $m \rightarrow \infty$ durch, nimmt der Vorfaktor den Wert 1 an. Bei α handelt es sich um den fraktalen Exponenten aus Formel (2.1). Die Summe der AKF ergibt wieder einen Graphen, der polynomial fällt.

Hieraus wird eine lineare Abnahme (im Gegensatz zur exponentiellen bei nicht selbstähnlichen Prozessen) ersichtlich. Dabei darf die Varianz ebenfalls nur polynomial abnehmen: $Var(X^{(m)}) \sim m^{-(1-\alpha)}$. Dieser Zusammenhang wird in einem Verfahren zur Bestimmung des Hurst-Parameters - dem *Variance-Time-Plot* - ausgenutzt.

Für asymptotische Selbstähnlichkeit zweiter Ordnung (die Momente zweiter und damit auch erster Ordnung des Prozesses existieren) schwächt die Korrelationsbedingung sich für hinreichend große k ab zu $r^{(m)}(k) \rightarrow r(k)$ für $m \rightarrow \infty$, da m in Formel (2.3) für praktische Szenarien nur endliche Werte annehmen kann. Sie wird in paketvermittelten Weitverkehrs-

¹ Es wird also geprüft, ob bei einer Vergrößerung des Grundbeobachtungsintervalls T sich die AKF ändert. Dies entspricht verschiedenen Blockgrößen bei der Binbildung (vgl. Kap. 4.1).

netzen und lokalen Netzwerken unter geringer Last beobachtet. Für traditionelle Modelle gilt $Var(X^{(m)}) \sim am^{-1}$ mit $a \in R$, die Varianz nimmt deutlich „schneller“ ab.

2.3 Entstehung von Long Range Dependence

Die Netzwerkanalyse und -modellierung hat ihre Wurzeln in der klassischen Telefonie und geht demzufolge von einigen Prämissen aus, die für paketvermittelte Netze aus technisch-ökonomischen Gründen nicht erfüllbar sind. Während in der leitungsvermittelten Sichtweise Gesprächsdauer und Anzahl der belegten Leitungen - die feste Bandbreiten besitzen - zu bestimmten Zeitpunkten Gegenstand des Interesses sind und von gedächtnislosen Kanälen und exponentialverteilten Gesprächsdauern ausgegangen werden kann, zeigen paketvermittelte Netze mit durch statistisches Multiplexing ständig sich ändernden Lastsituationen ein völlig anderes Verhalten. Zwar sind die Eigenschaften der einzelnen Sitzungen weiterhin weitestgehend unabhängig voneinander, jedoch ist schon der einzelne Endknoten im Netz in der Lage, langzeitabhängige Verkehrsmuster zu bilden. Es hat sich insbesondere bei „neuen“ Applikationen wie HTTP, welches einen großen Anteil am Verkehrsaufkommen im Internet hat, die Überlegenheit von Verteilungen mit Berücksichtigung von Langzeitabhängigkeit gezeigt.

Einige Gründe für die Entstehung von *Long Range Dependence* sind sollen näher beleuchtet werden. Als eine Ursache gilt der Umstand, daß in HTML-Dokumente beliebig viele Multimedia-Objekte wie Bilder, Animationen und Audiodateien eingebettet werden können. Im HTTP-Protokoll 1.0 ist vorgesehen, daß für jede zu übertragende Datei eine eigene TCP-Verbindung benutzt wird. Der generelle Ablauf ist demnach durch folgende Schritte gekennzeichnet:

- Download der HTML-Datei, Schließen der Verbindung, Ermittlung der eingebetteten Objekte
- Download von Objekt 1, Schließen der Verbindung, Verarbeitung von Objekt 1
- ...
- Download von Objekt n, Schließen der Verbindung, Verarbeitung von Objekt n
- Benutzer verarbeitet Webpage
- Benutzer ruft neue Webpage auf
- ...

Durch die Unterschiede in Dateigröße und Verarbeitungsaufwand entstehen verschieden lange Pausen, welche eine erste Erklärung für die beobachtete *Burstiness* auf Hostebene liefern.

HTTP - Version 1.1 nutzt zwar das Konzept der *Persistent Connections* (Übertragung mehrerer Objekte über eine Verbindung), in Fällen, in denen die Größe der zu übertragenden Datei nicht im voraus bekannt ist (z.B. CGI), ist dieser Ansatz jedoch nicht möglich. Andere Anwendungen, wie z.B. *Video-on-Demand* oder *Streaming Audio* erzeugen durch variable Bitratencodierung des Materials schon auf Quellenebene ebenfalls eine zeitlich schwankende Last.

Neben applikationsspezifischen Charakteristika spielt auch der Adaptationsmechanismus des verwendeten Transportprotokolls (meist TCP) eine Rolle. Techniken wie *Slow Start* und *Exponential Backoff* bedingen starke Schwankungen der Auslastung der Bandbreiten. In [VKM00] wird auf die Rolle der Bandbreitenadaptation bei der Entstehung selbstähnlichen Netzwerkverkehrs näher eingegangen.

Neben den verwendeten Protokollen spielt auch die Art der Zugriffsteuerung des Mediums eine Rolle. Der mit großem Abstand meistverbreitete Medienzugriffsmechanismus CSMA/CD nutzt die gesamte vom Medium zur Verfügung gestellte Bandbreite für einen sehr kurzen Zeitraum aus und gibt das Medium danach wieder frei. Im Falle mehrerer sendebereiter Quellen im Segment kann es beim Senden zu Kollisionen kommen, wodurch ein *Backoff*-Mechanismus wirksam wird, so daß nach einem zufällig gewählten Zeitwert ein

neuer Medienzugriff erfolgt. Diese Vorgehensweise erzwingt in Segmenten mit vielen Stationen eine stark schwankende Auslastung des Mediums mit Zeiträumen von Null- (direkt nach einer Kollision), Voll- (Station sendet) und Überlast (mehrere Stationen senden gleichzeitig, Kollision) in zufälliger Reihenfolge. Durch die fortschreitende Mikrosegmentierung moderner Netzwerke verliert dieser Punkt an Einfluß.

Die Gründe für Langzeitabhängigkeit sind auf mehreren Ebenen des OSI-Referenzmodells zu finden und stellen den Preis für eine gegenüber leitungsvermittelten Netzen billigere Technologie dar. Die folgenden Betrachtungen setzen sich mit den in paketvermittelten Netzen beobachteten Wahrscheinlichkeitsverteilungen der Zeiträume zwischen den Ankünften aufeinanderfolgender Pakete auseinander. In Abschnitt 6 werden Proben aus dem Backbone-Netz der Universität Rostock u.a. auf ihre Verteilungsstruktur untersucht und simuliert.

2.4. Spezielle Verteilungen

Dieser Abschnitt stellt die für die Netzwerkverkehrsanalyse gebräuchliche Verteilungen vor und beurteilt sie bezüglich ihrer Eignung für die Aufgabenstellung.

2.4.1 Die Exponentialverteilung

In der klassischen Warteschlangentheorie, die insbesondere Problemstellungen der leitungsvermittelte Telefonie behandelt, sind Gesprächsdauer und die Anzahl der belegten Leitungen in einer Vermittlungsstelle in Abhängigkeit von der Zeit exponentialverteilt. Die dafür ausschlaggebende Eigenschaft ist die Gedächtnislosigkeit des Kanals bzw. der Verteilung, d.h. die Vergangenheit eines exponentialverteilten Prozesses wirkt sich nicht auf seine Zukunft aus. Dies äußert sich darin, daß seine Autokorrelationsfunktion exponentiell abnimmt. In paketvermittelten Netzen ist diese Voraussetzung, wie in (2.3) gesehen, nicht gegeben. Das gilt als Grund dafür, daß Modelle, die bei der Simulation von paketvermitteltem Verkehr eine exponentialverteilte Ankunfts-wahrscheinlichkeitsdichte annehmen, bei der Modellierung selbstähnlicher Prozesse keine realistischen Resultate über alle Zeitskalen erzielen.

Die Dichteverteilung ist gegeben mit $f(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$; $x \geq 0$ (vgl. Bild 2.4) , wobei x i.a. als

Zeit angenommen wird und der Parameter $\mu = \frac{1}{\lambda}$ als Intensität (z.B. mittlere Ankunftsrate)

bezeichnet wird. Sie läßt sich z.B. interpretieren als die Wahrscheinlichkeit, daß innerhalb eines bestimmten Zeitintervalls ein Ereignis stattfinden wird. Die wichtigsten Charakteristika der Exponentialverteilung sind der Erwartungswert $E(T) = \mu$ sowie die Varianz

$$\sigma^2 = E(T^2) - E^2(T) = \mu^2 .$$

Die weiteren Verteilungen unterscheiden sich von der Exponentialverteilung durch die sogenannte *Heavy-Tailedness*, d.h. die Fläche unter dem „Schwanz“ des Graphen der Wahrscheinlichkeitsdichtefunktion repräsentiert einen erheblichen Teil der Gesamtfläche. Eine Zufallsvariable, die einer solchen Verteilung folgt, kann extrem große Werte mit einer nicht vernachlässigbaren Wahrscheinlichkeit annehmen.

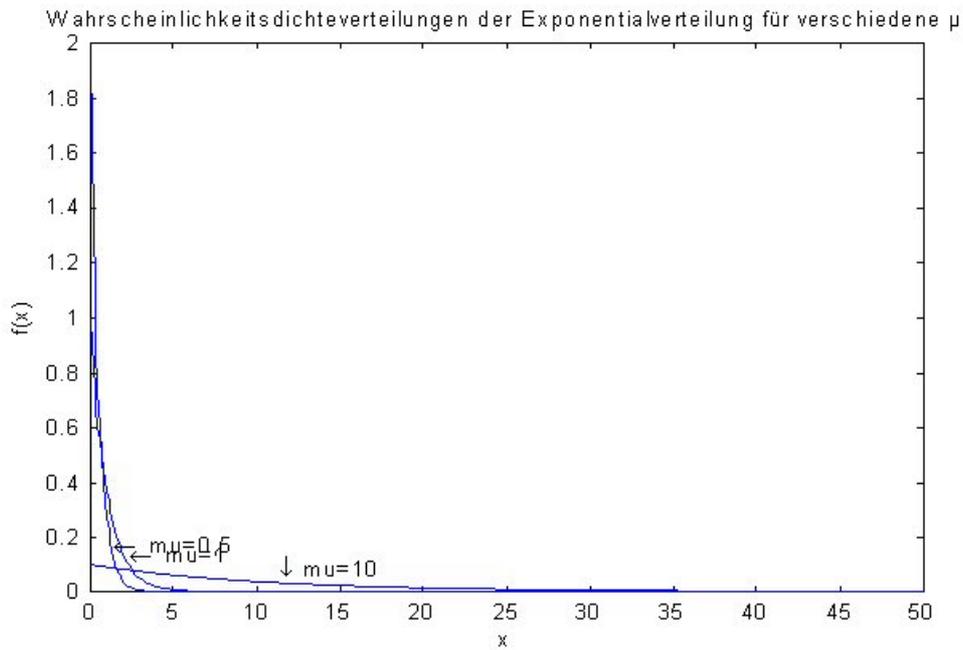
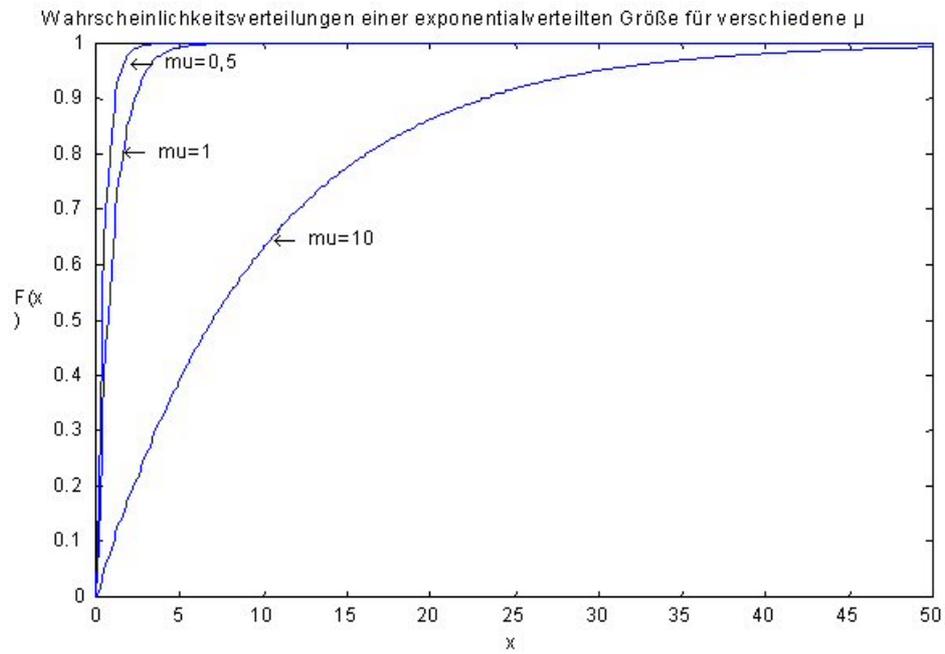


Abbildung 2.4: Wahrscheinlichkeitsverteilung (oben), Wahrscheinlichkeitsdichteverteilung (unten) der Exponentialverteilung

2.4.2 Die Weibullverteilung

Die Dichteverteilungsfunktion der zweiparametrischen Weibullverteilung ist gegeben durch

$$f(x) = \frac{\alpha}{\beta} \cdot \left(\frac{x}{\beta}\right)^{\alpha-1} \cdot \exp\left(-\frac{x}{\beta}\right)^{\alpha} \quad \text{mit } \alpha, \beta > 0; x \geq 0.$$

α wird als Form-, β als Vergrößerungsparameter bezeichnet. Für $\alpha=1$ gewinnt man die Exponentialverteilung mit dem Parameter $\beta = \mu$ zurück. Durch entsprechende Anpassung der Parameter ist in [Fe01] der Nachweis einer Weibullverteilung der gemessenen TCP-Ankunftsprozesse gelungen, während alle anderen dort betrachteten Verteilungen z.T. deutlich abwichen. In Abb. 2.5 sind die Verläufe der Wahrscheinlichkeitsdichteverteilung und der Wahrscheinlichkeitsverteilung dargestellt.

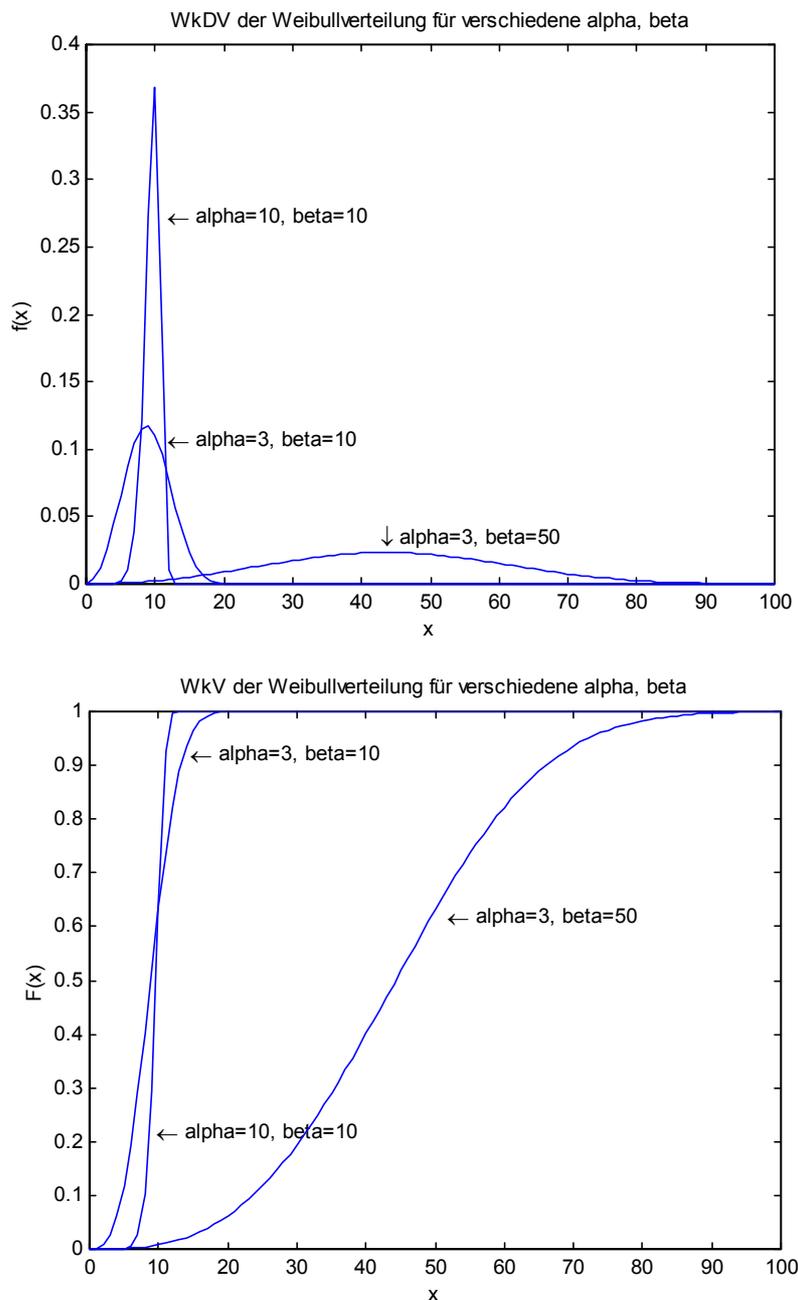


Abbildung 2.5: Wahrscheinlichkeitsdichteverteilung (oben), Wahrscheinlichkeitsverteilung (unten) der Weibullverteilung

2.4.3 Pareto-Verteilung

Paretoverteilte Zufallsprozesse spielen aufgrund ihrer selbstähnlichen Charakteristik in vielen Veröffentlichungen, so z.B. in [PF95] eine herausragende Rolle in der Erzeugung langzeitabhängiger Prozesse. Die Verteilungsdichtefunktion lautet $f(x) = \frac{\alpha}{t_0} \cdot \left(\frac{t_0}{x}\right)^{\alpha+1}$ mit dem Formparameter α und dem Lageparameter t_0 . Seit der Mitte der neunziger Jahre ist die Pareto-Verteilung am häufigsten zur Modellierung selbstähnlicher Ankunftsprozesse benutzt worden. Es ist jedoch zu beachten, daß Ankunftszeiten $< t_0$ nicht definiert sind!

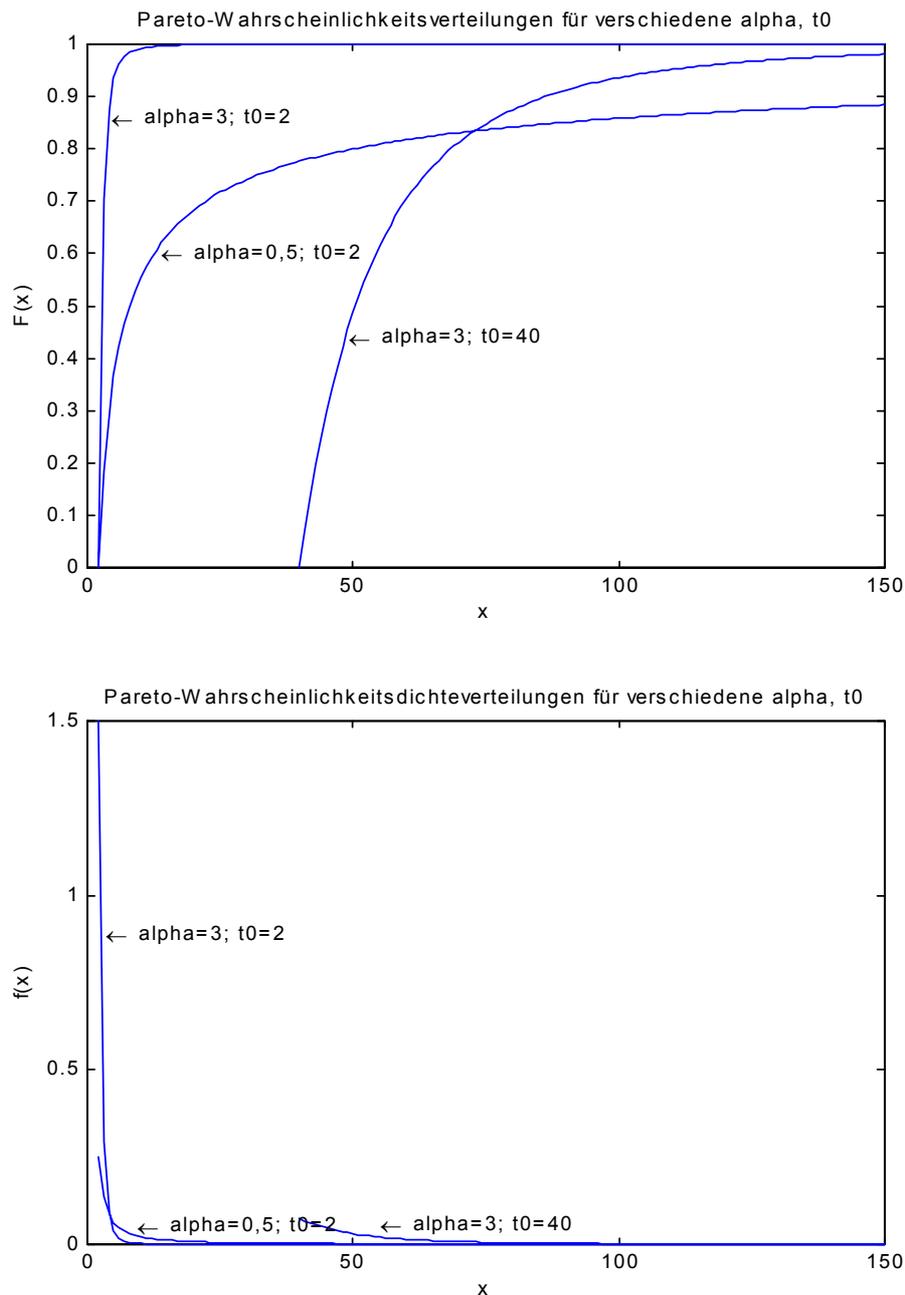


Abbildung 2.6: oben Wahrscheinlichkeitsverteilung, unten Wahrscheinlichkeitsdichteverteilung der Paretoverteilung

2.4.4 Kurzübersicht über die betrachteten Verteilungen

Name	Exponentialverteilung	Paretoverteilung	Weibullverteilung
Wahrscheinlichkeitsverteilung	$F_t(x) = 1 - e^{-\frac{x}{\mu}}$; $\mu > 0; x \geq 0$	$F(x) = 1 - \left(\frac{t_0}{x}\right)^\alpha$ $x > t_0$ $t_0 > 0$	$F(x) = 1 - \exp\left(-\frac{x}{\beta}\right)^\alpha$ $\alpha, \beta > 0; x \geq 0$
Wahrscheinlichkeitsdichteverteilung	$f(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$	$f(x) = \alpha t_0^\alpha \cdot x^{-\alpha-1}$	$f(x) = \frac{\alpha}{\beta} \cdot \left(\frac{x}{\beta}\right)^{\alpha-1} \cdot \exp\left(-\frac{x}{\beta}\right)^\alpha$
Erwartungswert	$E = \mu$	$E = \frac{\alpha t_0}{\alpha - 1}, \alpha > 1$	$E = \beta \Gamma\left(1 + \frac{1}{\alpha}\right)^*$
Varianz	$\sigma^2 = \mu^2$	$\sigma^2 = \frac{\alpha t_0^2}{(\alpha - 2)(\alpha - 1)^2}$	$\sigma^2 = \beta^2 \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right) \right]^*$

Tabelle 2.1: Zusammenfassung der Verteilungen

* $\Gamma(x) = \int_{x=0}^{\infty} t^{x-1} e^{-t} dt \quad x > 0$

3. Methoden zur Schätzung des Hurst Parameters

Der Hurst-Parameter gibt den Grad der Langzeitabhängigkeit in einer zu betrachtenden Meßreihe (z.B. Zwischenankunftszeiten) an. Somit kann der Hurst-Parameter (definiert von Harold Hurst 1965) auch als Grad der Selbstähnlichkeit angesehen werden. Gemeint ist damit die Korrelation von Zwischenankunftszeiten in verschiedenen Zeitbereichen oder wie schnell die „Burstiness“ mit dem Grad der Aggregation der Datenfolgen zunimmt.

Der Hurst-Parameter ist immer als Schätzparameter zu sehen, dessen exakter Wert nie genau ermittelt werden kann. Im folgenden werden Methoden vorgestellt, mit denen aufgrund selbstähnlicher Eigenschaften der Hurst-Parameter geschätzt wird.

Unter anderem nach [LTWW94] eignen sich diese Methoden ebenfalls, um Selbstähnlichkeit nachzuweisen. Näher betrachtet werden:

1. Variance-Time-Plot
2. R/S-Statistik
3. Periodogrammanalyse im Frequenzbereich
4. Whittle-Estimator

3.1 Variance-Time-Plot (VTP)

Der Variance-Time-Plot wurde bereits in einer Vielzahl von Veröffentlichungen ([LTWW94], [A99]) verwendet, um den Hurst-Parameter zu bestimmen.

Es wird von einer Reihe von Meßwerten ausgegangen, die auf Langzeitabhängigkeit geprüft werden soll. Im Zusammenhang mit dieser Arbeit wurden zum einen die einzelnen Zwischenankunftszeiten und zum anderen die Anzahl der Pakete innerhalb eines Grundintervalls (Bins) betrachtet. Zur weiteren Erläuterung des VTP soll von einer „Meßreihe“ ausgegangen werden. Diese könnte auch aus einer Folge von Paketgrößen bestehen. Eine Meßreihe sei definiert als Folge von Meßwerten $X = (X_t : t = 0,1,2,...)$ und wird beim VTP aufgeteilt in einzelne, sich nicht überlappende Blöcke der Größe m . Bei einer Blockgröße von $m=2$ werden jeweils immer zwei aufeinanderfolgende Meßwerte in einem Block zusammengefaßt.

Durch Bildung des Mittelwertes jedes einzelnen Blocks nach der Gleichung

$$X_k^{(m)} = \frac{1}{m} [X_{km-m+1} + \dots + X_{km}] \text{ für } k \geq 1 \quad (3.1)$$

entsteht eine neue Wertefolge $X^{(m)} = (X_k^{(m)} : k = 1,2,3,...)$. Für jede gewählte Blockgröße entsteht eine neue Wertefolge aus den jeweiligen Mittelwerten.

Es ergibt sich z.B. für eine Blockgröße von $m=10$ eine Folge von Werten mit

$$\begin{aligned} m=10, k=1: \quad X_1^{(10)} &= \frac{1}{10} [X_{1*10-10+1} + \dots + X_{1*10}] = \frac{1}{10} [X_1 + \dots + X_{10}] \\ m=10, k=2: \quad X_2^{(10)} &= \frac{1}{10} [X_{2*10-10+1} + \dots + X_{2*10}] = \frac{1}{10} [X_{11} + \dots + X_{20}] \end{aligned}$$

...

Beim VTP wird nun die Varianz der neuen Wertefolgen betrachtet ($\text{var}(X^{(m)})$). Die Betrachtung der Varianz bei einer Blockgröße von $m=1$ führt zu $\text{var}(X^{(1)}) = \text{var}(X)$, was die Varianz der Original-Meßreihe ist. MATLAB bietet die Funktion „var()“, was die Berechnung des VTP erleichtert. Implementiert wurde der Variance-Time-Plot in verschiedenen Varianten:

- Varianz_lin_delusec.m
- Varianz_log_opt_delusec.m
- Varianz_log_opt_bins.m

Hierbei bedeutet „delusec“, daß der Variance-Time-Plot aus Zwischenankunftszeiten berechnet wird. Diese müssen in bereits berechneter Form in Mikrosekunden vorliegen. „Log_opt“ weist auf eine Variante mit logarithmischer Optimierung hin (vgl. Abschnitt 4.4). Für eine Berechnung des VTP aus bereits berechneten Binvektoren ist das Skript „Varianz_log_opt_bins.m“ zu verwenden.

Zur Bestimmung des Hurst-Parameters ist die Varianz in Abhängigkeit von der Blockgröße mit jeweils logarithmisch geteilten Achsen graphisch darzustellen. Aus dem Anstieg β des resultierenden Graphen wird der Hurst-Parameter nach

$$H = 1 - \frac{|\beta|}{2} \quad (3.2)$$

ermittelt. Wenn Langzeitabhängigkeit vorliegt, muß der Anstieg des Graphen zwischen 0 und -1 liegen. Langzeitabhängigkeit gilt als nachgewiesen, wenn die Varianz langsamer als das Reziproke der Blockgröße m abnimmt (vgl. Abschn. 2.2.4).

$$\text{var}(X^{(m)}) \sim am^{-\beta}$$

Bei der Ermittlung des Anstiegs sind klare Grenzen zu definieren, in denen dieser zu betrachten ist. Es sind die jeweils kleinsten und größten Blockgrößen bei einer linearen Regression auszuschließen. Bei großen Blockgrößen entstehen starke Schwankungen, die mit Rauschen vergleichbar sind.

Die eigentliche Bestimmung des Hurst-Parameters durch lineare Regression wurde separat implementiert in „regress_lin_VTP.m“ und „regress_log_opt_VTP.m“. Dieses ermöglicht eine Veränderung der zu betrachtenden Grenzen, ohne den VTP neu zu berechnen.

3.2 R/S-Statistik (Rescaled Adjusted Range)

R/S-Statistics basieren auf einer Gesetzmäßigkeit, die bei Meßreihen mit Langzeitabhängigkeit nachgewiesen werden konnten. Der von Hurst (1955) beschriebene Zusammenhang wird auch als Hurst-Effekt oder Hurst's Law bezeichnet.

Obwohl in einer Vielzahl von Veröffentlichungen die R/S-Statistik näher erläutert wird und Verwendung findet, werden hier nur [HDLK95], [B01] und [LTWW93] erwähnt, da diese als Quellen für eine eigene Implementierung verwendet wurden.

Ausgangspunkt ist wiederum eine Meßreihe (bestehend aus Zwischenankunftszeiten oder Blöcken gemäß Kap. 4.1) $(X_k : k = 1, 2, 3, \dots, n)$, in der eine Langzeitabhängigkeit nachzuweisen ist. $\bar{X}(n)$ ist der arithmetische Mittelwert (MATLAB: *mean()*) und $S(n)$ die Standardabweichung (MATLAB: *std()*) der Meßreihe. Nach [LTWW94] ist die R/S-Statistik definiert als:

$$\frac{R(n)}{S(n)} = \frac{\max(0, W_1, W_2, \dots, W_n) - \min(0, W_1, W_2, \dots, W_n)}{S(n)}$$

mit

$$W_k = (X_1 + X_2 + \dots + X_k) - k\bar{X}(n) \quad \text{und} \quad k = 1, 2, 3, \dots, n \quad (3.3)$$

Hierbei sei zu erwähnen, daß W_n immer 0 ist. Hurst konnte nachweisen, daß viele „Zeitreihen“ (engl. time series) mit

$$E[R(n)/S(n)] \sim b * n^H$$

gut beschrieben werden können. Der Faktor b sei eine Konstante. H ist der Hurst-Parameter und nimmt bei Reihen mit Langzeitabhängigkeit einen Wert zwischen 0,5 und 1 an.

$$E[R(n)/S(n)] \sim b * n^{0,5}$$

weist jedoch auf eine Reihe mit Kurzzeitabhängigkeit (engl. Short-range dependence) hin.

Bei der Schätzung des Hurst-Parameters mit Hilfe der R/S-Statistik wird eine Meßreihe $(X_k : k = 1, 2, 3, \dots, N)$ in K nicht-überlappende Blöcke zerlegt. Berechnet wird die zugehörige R/S-Statistik $R(t_i, n)/S(t_i, n)$ für jeden Block. t_i ist hierbei der Startpunkt der zu betrachtenden Werte und n die Blockgröße. Der erste zu untersuchende Bereich startet mit $t_1=1$, dem ersten Meßwert. Jeder weitere Startpunkt wird berechnet werden nach:

$$t_2 = \frac{N}{K} + 1, \quad t_3 = \frac{2N}{K} + 1, \quad \dots$$

Die Bedingung $(t_i - 1) + n \leq N$ muß erfüllt sein. Bei der Aufteilung der Blöcke wird W_k aus Gleichung (3.3) ersetzt durch $W_{t_i+k} - W_{t_i}$.

Zur graphischen Bestimmung des Hurst-Parameters werden die R/S-Werte in Abhängigkeit von der Blockgröße aufgetragen mit jeweils logarithmisch geteilten Achsen. Der Anstieg des resultierenden Graphen ist durch lineare Regression zu ermitteln und ist der Hurst-Parameter der untersuchten Meßreihe.

Im MATLAB-Skript „rescaled_bins.m“ ist eine mögliche Implementierung der R/S-Statistik zu finden.

3.3 Periodogramm

Die Auswertung einer Meßreihe mittels des *Periodogramms*, einer Methode zur Abschätzung der spektralen Leistungsdichte unter Verwendung der diskreten Fourier(rück-)transformation, ist Gegenstand dieses Abschnitts. Das Periodogramm ist gegeben durch

$$I(\lambda_k) = \frac{1}{N} \cdot \left| \sum_{n=1}^N (X_n) e^{jn\lambda_k} \right|^2 \quad \text{mit } \lambda_k = \frac{2\pi k}{N}, \quad (3.4)$$

wobei k den Wertevorrat von 1 bis $N/2$ durchläuft. Im Fall von LRD ist das Periodogramm bei niedrigen Frequenzen proportional zu $|\lambda|^{1-2H}$, aufgrund des Zusammenhangs $\lg(f(\omega)) \sim (1-2H)\lg|\omega| + C$ [DJ96].

Eine Darstellung des logarithmierten Periodogramms gegen die logarithmierten Frequenzen liefert einen Schätzwert für den Hurst-Parameter. Hierzu werden die Ergebnisse als Punktwolke in den Graphen eingetragen; anschließend ist eine lineare Regression durchzuführen. Einige Quellen, z.B. [B01], empfehlen eine Beschränkung auf die unteren 10% des Frequenzbereichs, um den Einfluß der *Short Range Dependence* (tritt als „Rauschen“ bei hohen Frequenzen in Erscheinung) zu verringern. Dieser Ansatz wird an dieser Stelle nicht verfolgt, da hierdurch bei einer Aggregationsgröße von 1s die ohnehin schon wenigen Meßwerte weiter reduziert würden.

Matlab bietet mit der Funktion `psd()` (Power Spectral Density) der Signal-Toolbox eine komplette Implementation des Verfahrens. Als Eingangsgröße dienen der Binvektor, die Länge der FFT (hier die Länge des Binvektors) sowie optional der Vertrauensbereich. Der Aufruf erfolgt in der Form `[I,lambda] = psd(binvektor, length(binvektor), P)`. Nach der Logarithmierung der beiden Ergebnisvektoren kann unmittelbar die Berechnung der linearen Regressionsgeraden erfolgen. Hier bietet Matlab mit der Funktion `polyfit()` ebenfalls eine komfortable Lösung. Der Hurst-Parameter ergibt sich dann aus der Relation

$$H = -\frac{m-1}{2}.$$

3.4 Whittle-Estimator

Im Gegensatz zu den oben angewandten Methoden zur Bestimmung eines Näherungswertes für den Hurst-Parameter handelt es sich beim Whittle-Estimator um ein numerisches Verfahren, welches darüberhinaus zumindest für künstlich generierte selbstähnliche Folgen i.allg. die genauesten Ergebnisse liefert. Als Eingangsgröße dient das Periodogramm. Der Vorteil besteht darin, daß bei diesem Verfahren die Subjektivität des Betrachters ausgeschlossen ist; während insbesondere bei RS-Plot und Variance-Time-Plot zunächst am Graphen die nichtlinearen Bereiche von der weiteren Betrachtung ausgeschlossen werden müssen, ist hier ein solches Einwirken nicht nötig.

Beim Einsatz des Whittle-Estimators ist jedoch Vorsicht geboten: es handelt sich hier nicht um eine Methode, um Langzeitabhängigkeit nachzuweisen, sondern lediglich um ein Schätzverfahren für den Hurst-Parameter. Das Vorhandensein von LRD muß also durch Anwendung der anderen Verfahren sichergestellt sein, sonst können irreführende bzw. falsche Ergebnisse entstehen! Des weiteren muß der Anwender die Art des angenommenen zugrundeliegenden Zufallsprozesses vorgeben. Neben dem hier betrachteten Fall von *Fractional Gaussian Noise* (FGN) ist ebenfalls ein *FARIMA* (Fractional Autoregressive Integrated Moving-Average) - Ansatz möglich.

Bei der Anwendung dieser Methode ist ebenfalls die Empfindlichkeit gegenüber „Störungen“ zu beachten. Eine große Fehlerquelle für den Whittle-Estimator stellt gauß'sches Rauschen dar. In [KFR02] ist eine Betrachtung der Anfälligkeit der gebräuchlichsten Schätzalgorithmen

gegen Periodozität, Nonstationarität und Rauschen enthalten. Unter Umständen widersprechen sich die Estimatoren sogar bezüglich des Vorhandenseins von LRD.

Die nun folgende Beschreibung des Prinzips des Whittle-Estimators beruht auf einem Text von Paxson [P97].

Um H abzuschätzen, ist ein \hat{H} zu bestimmen, welches die Gleichung

$$g(\hat{H}) = \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda, \hat{H})} d\lambda \quad (3.5)$$

minimiert. Voraussetzung dieses Verfahrens ist die statistisch asymptotische Unabhängigkeit der Elemente des Periodogrammvektors.

Bei einer endlichen Anzahl von N Eingangswerten ändert sich die Gleichung zu einer Summation mit $\lambda = n \cdot \frac{2\pi}{N}$; $n = 1, 2, \dots, N-1$. Die neue Gleichung lautet demnach

$$g(\hat{H}) = \sum_{\lambda=\frac{2\pi}{N}}^{\frac{2\pi(N-1)}{N}} \frac{I(\lambda)}{f(\lambda, \hat{H})}. \quad (3.6)$$

Unter der hier betrachteten Voraussetzung, daß ein FGN-Prozeß zugrundeliegt, ergibt sich für den Nenner

$$f(\lambda, \hat{H}) = A(\lambda, \hat{H}) \cdot \left[|\lambda|^{-2\hat{H}-1} + B(\lambda, \hat{H}) \right] \quad \text{mit}$$

$$A(\lambda, \hat{H}) = 2 \sin(\pi \hat{H}) \Gamma(2\hat{H} + 1) (1 - \cos(\lambda)) \quad \text{und} \quad [\text{P97}]$$

$$B(\lambda, \hat{H}) = \sum_{n=1}^{\infty} \left[(2\pi n + \lambda)^{-2\hat{H}-1} + (2\pi n - \lambda)^{-2\hat{H}-1} \right]$$

Insbesondere die Bestimmung von B birgt die Frage, wie viele Glieder für eine hinreichende Genauigkeit aufsummiert werden müssen. Paxson löst dieses Problem durch eine Näherungsgleichung, die ebenfalls in [P97] hergeleitet wird:

$$B(\lambda, H) \approx a_1^d + b_1^d + a_2^d + b_2^d + a_3^d + b_3^d + \frac{a_3^{d'} + b_3^{d'} + a_4^{d'} + b_4^{d'}}{8\hat{H}\pi} \quad \text{mit}$$

$$d = -2\hat{H} - 1$$

$$d' = -2\hat{H} \quad [\text{P97}]$$

$$a_k = 2k\pi + \lambda$$

$$b_k = 2k\pi - \lambda$$

Durch diese Beziehung wird es erst technisch möglich, den Hurst-Parameter mittels eines Standard-PC zu berechnen.

4 Analyse der Meßreihen

In den folgenden Abschnitten sind die Ergebnisse aus der Analyse der Meßreihen dargestellt. Die hier vorgestellten Ergebnisse wurden mittels Matlab 5.1 mit den Toolboxen *Signal* und *Stats* berechnet; eine Aufstellung der Listings befindet sich im Anhang.

4.1 Berechnung der Bins

Bins können als „Container“ angesehen werden, in denen die Zwischenankunftszeiten zusammengefaßt sind. Die Berechnung der Bins setzt die Angabe eines Grundintervalls voraus.

Ein vereinfachtes Beispiel verdeutlicht das Grundprinzip. Gegeben sei eine Reihe von zeitlich aufeinander folgenden Werten

[1, 3, 5, 6, 4, 3, 2, 5, 1, 6, 4, 2] die in Abbildung 4.1 dargestellt ist.

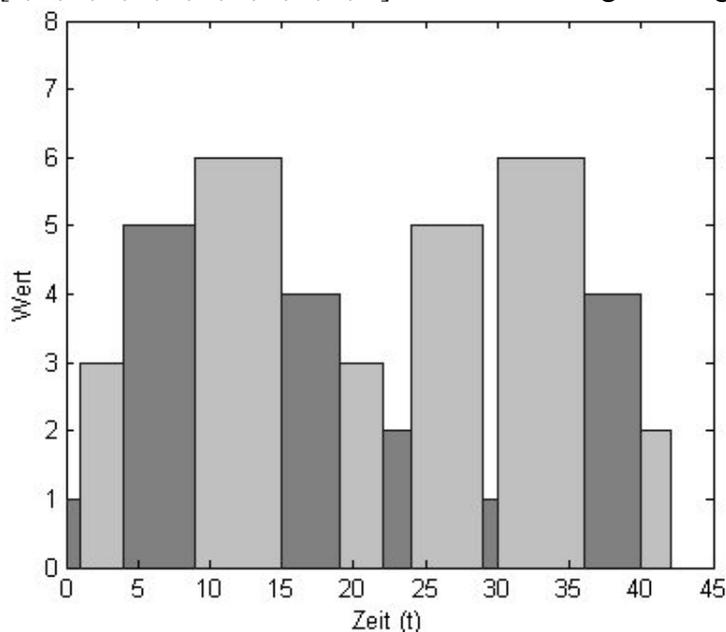
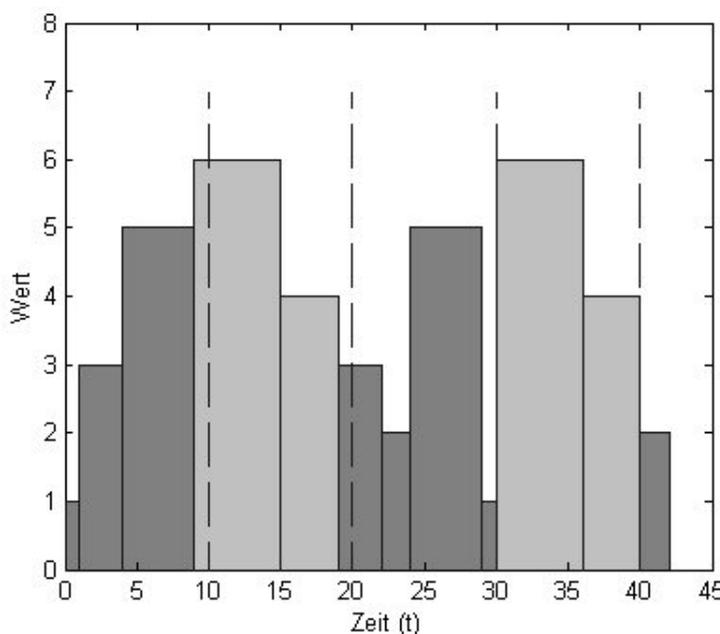
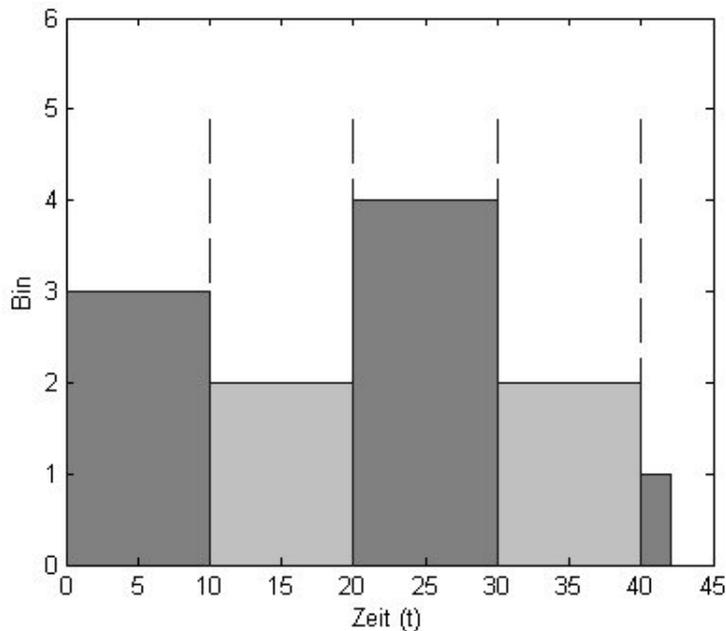


Abbildung 4.1:
Darstellung der Beispielfolge



Gewählt sei ein Grundintervall von 10 Zeiteinheiten. Diese Binsgröße ist in Abbildung 4.2 markiert.

Abbildung 4.2:
Darstellung des Grundintervalls und Markierung der zugehörigen Wert



Bins enthalten die Information, wie viele Pakete innerhalb des gewählten Intervalls eintreffen. Die Zuordnung der einzelnen Werte zu den resultierenden Bins ist ebenfalls in der obigen Abbildung zu erkennen. Die resultierenden Bins ergeben sich durch das Zählen der betreffenden Werte. In Abbildung 4.3 sind diese dargestellt.

Abbildung 4.3:
Darstellung der resultierenden Bins

Die Implementierung der Bin-Berechnung in MATLAB ist als unproblematisch anzusehen, da nur ein Intervall festzulegen ist und die zugehörigen Werte auszuzählen sind. Im Rahmen dieser Arbeit handelt es sich bei den Meßwerten um Zwischenankunftszeiten, die in MATLAB als Vektor geladen werden. Bei der Realisierung ist der Vektor vollständig zu durchlaufen, um alle Zwischenankunftszeiten zu erfassen. Eine Hilfsvariable gibt den „Füllstand“ des Bins in Bezug auf das Grundintervall an. Eine zu betrachtende Zwischenankunftszeit wird mit der Hilfsvariablen (vorläufig) addiert und mit dem Grundintervall verglichen. Ist die Summe kleiner als das (bzw. gleich dem) Grundintervall, wird der Wert des Bins um eins erhöht und die Hilfsvariable erhält ihren endgültigen Wert. Ist sie größer, wird ein neues Bin mit eins und die Hilfsvariable mit dem überlappenden Rest initialisiert.

Die Summe aller Bins ergibt die Anzahl der Zwischenankunftszeiten, was als sinnvoller Test anzusehen ist. Eine Realisierung ist im MATLAB-Skript „bin_berechnung.m“ zu finden.

4.2 Bestimmung der Ober- und Untergrenzen

Bei der Auswertung von Meßreihen ist in einigen Fällen die Darstellung der Bins genauer zu betrachten, da hier Eigenschaften auftreten, die sehr aussagekräftig sein können. So hängt die Größe des Bandes, in dem sich die Bins befinden, unmittelbar von der Fractal Onset Time Scale (kurz: FOTS oder T) ab. Dieser Zusammenhang soll jedoch erst in späteren Abschnitten (5.4 und 5.5) näher betrachtet werden. Es ist sinnvoll, Grenzen in der Bindarstellung festzulegen, in denen die Mehrzahl aller Bins auftreten. Die Festlegung der Grenzen über die maximalen bzw. minimalen Bins ist zu fehlerbehaftet, da diese Werte - abhängig vom Datensatz - extrem voneinander abweichen können. Weiterhin treten starke Abweichungen der Minimal- bzw. Maximalwerte in der Analyse und in den späteren Ergebnissen aus der Simulation auf. Dieses kann mit den zugrunde liegenden Wahrscheinlichkeitsverteilungen zusammenhängen, wenn sich diese bei Analyse und Simulation unterscheiden. In den folgenden Abschnitten werden Methoden vorgestellt, mit denen eine Schätzungen der Ober- und Untergrenzen des Bandes möglich ist. Diese sind weitgehend unabhängig von der Wahrscheinlichkeitsdichtverteilung. Abschließend werden Vor- und Nachteile sowie Anwendungsfälle untersucht.

4.2.1 Mittels Mittelwertbildung

In diesem Abschnitt wird eine Methode vorgestellt, mit der Ober- und Untergrenzen in einer Bindarstellung geschätzt werden können. Diese Methode beruht auf der Bildung mehrerer Mittelwerte. Hierbei ist zunächst der Mittelwert der Bins zu bilden. Abbildung 4.4 (oben links) zeigt die Bindarstellung einer bereits simulierten Meßreihe mit $T=0,1$ und den zugehörigen Mittelwert.

Auf die Modellierung und Simulation soll näher im Abschnitt 5 eingegangen werden.

Es ergeben sich nun zwei Bereiche. Der obere enthält alle Bins, die größer sind als der Mittelwert und der untere Bereich alle Bins, die kleiner sind als der Mittelwert.

Weiterhin kann nun ein Mittelwert des oberen Bereichs berechnet werden. Wird der Mittelwert des unteren Bereichs gebildet, ergeben sich daraus zwei Werte, die als vorläufige Grenzen angesehen werden können. Abbildung 4.4 (oben rechts) zeigt noch einmal die Bindarstellung mit den neu gewonnenen Werten. Es ist zu erkennen, daß ein bestimmter Bereich eingegrenzt wird, auf den nun ca. 57 % aller auftretenden Bins entfallen. Da dieser Wert für eine Aussage über die Ober- bzw. Untergrenze nicht ausreichend ist, ist der Bereich noch einmal zu vergrößern. Erreicht wird dieses durch die nochmalige Bildung des Mittelwertes der Werte, die die vorläufigen Grenzen über- bzw. unterschreiten. Durch diese Verbreiterung fallen nun etwa 83% aller auftretenden Bins in diesen Bereich. Dieses ist in Abbildung 4.4 (unten) zu erkennen.

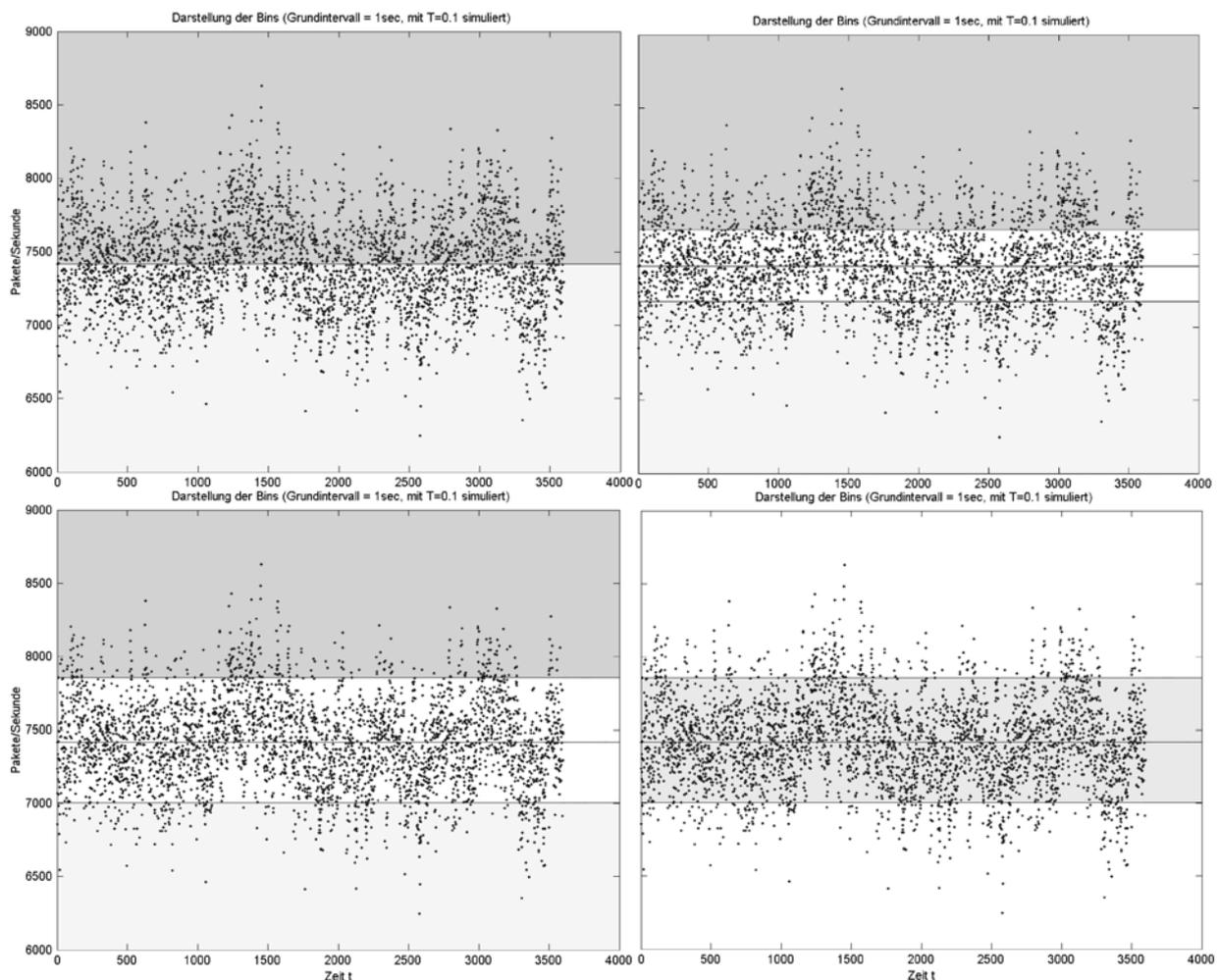


Abbildung 4.4: Erklärung zur Schätzung der Bereichsgrenzen durch Bildung von mehreren Mittelwerten.

Eine mögliche Implementierung stellt das MATLAB-Skript „*bandapprox_mittelwerte.m*“ dar. Hierbei wurde zur Bildung der Mittelwerte die Funktion *mean()* verwendet. Für eine Selektion aller Bins, die einen Wert über- bzw. unterschreiten, war die Funktion *find()* anzuwenden.

4.2.2 Mittels Wahrscheinlichkeitsdichtverteilung

In einigen Fällen ist die im vorherigen Abschnitt beschriebene Methode zur Schätzung der Ober- und Untergrenzen nicht sicher anwendbar und kann zu ungenauen Ergebnissen führen. Es sei auf Abschnitt 4.2.3 verwiesen, da hier beide Methoden verglichen und deren Vor- und Nachteile herausgestellt werden.

Mit Hilfe der relativen Häufigkeiten aus der Wahrscheinlichkeitsdichtverteilung kann ein Bereich festgelegt werden, der die Mehrzahl aller Bins einschließt. In Abb. 4.5 ist eine exemplarische WkDV dargestellt.

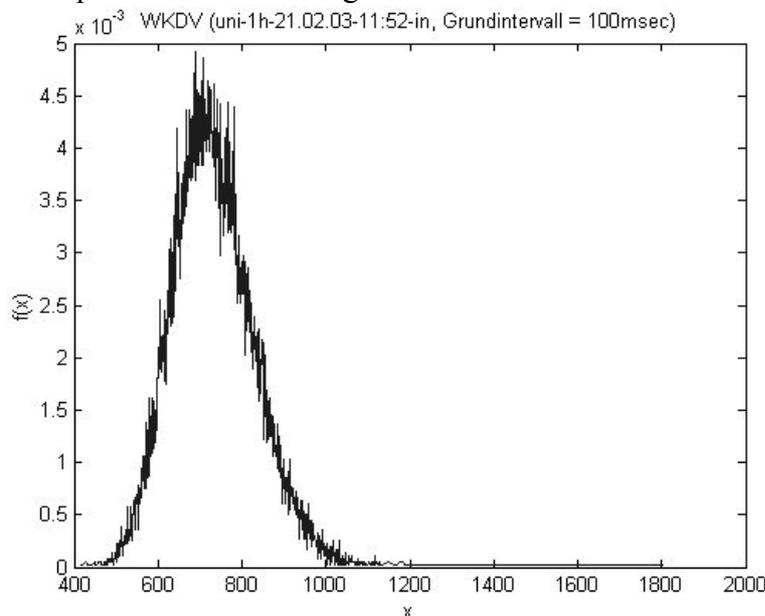


Abbildung 4.5:
WkDV des Datensatzes
uni_1h_21.02.03_11:52_in mit
Grundintervall von 100msec

Implementiert wurde ein Verfahren, das eine bestimmte relative Häufigkeit als Schwellenwert festlegt und alle Bins zusammenfaßt, die diesen Schwellenwert als rel. Häufigkeit besitzen bzw. überschreiten. Dabei wird ein Bereich in der WKDV definiert, der die am häufigsten auftretenden Bins enthält. In Abbildung 4.6 ist die obige WKDV mit einem Schwellenwert dargestellt, mit dem daraus resultierenden Bereich.

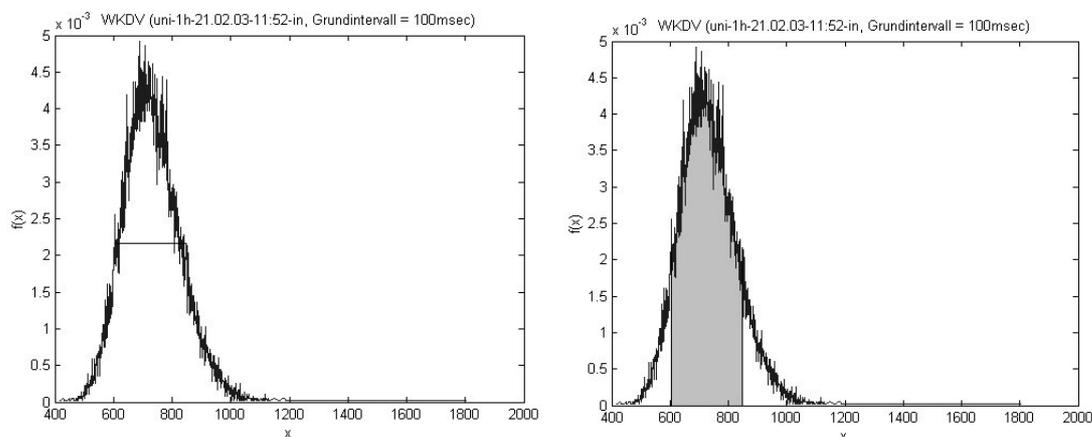


Abbildung 4.6: Markierung eines exemplarischen Schwellenwertes und des daraus resultierenden Bereiches

Die Grenzen des Bereichs sind gleichzeitig die gesuchten Ober- und Untergrenzen. Die Summe aller relativen Häufigkeiten zwischen den beiden Grenzen gibt wiederum prozentual an, wie viele Bins in diesem Bereich enthalten sind.

Es ist bei der Implementierung sinnvoll, zunächst diesen Prozentsatz festzulegen (hier: 80%) und den zugehörigen Schwellenwert zu suchen. Beginnend bei der maximal auftretenden relativen Häufigkeit als Anfangsschwellenwert, kann dieser in weiteren Schritten so lange gesenkt werden, bis der gewünschte Prozentsatz erreicht wird.

Die resultierenden Grenzen seien immer als Richtwerte zu sehen und stellen auf keinen Fall exakt ermittelte Werte dar. Das MATLAB-Skript „*bandapprox_wdv.m*“ enthält eine mögliche Implementation.

4.2.3 Verwendungsmöglichkeiten der Methoden

Die Entscheidung, welche Methode zur Schätzung der Ober- und Untergrenzen geeignet ist, hängt maßgeblich vom Verlauf der Wahrscheinlichkeitsdichteverteilung ab. Hierzu sind in Abbildung 4.7 zwei Wahrscheinlichkeitsdichteverteilungen vergleichend gegenübergestellt, mit zwei verschiedenen Bingrößen.

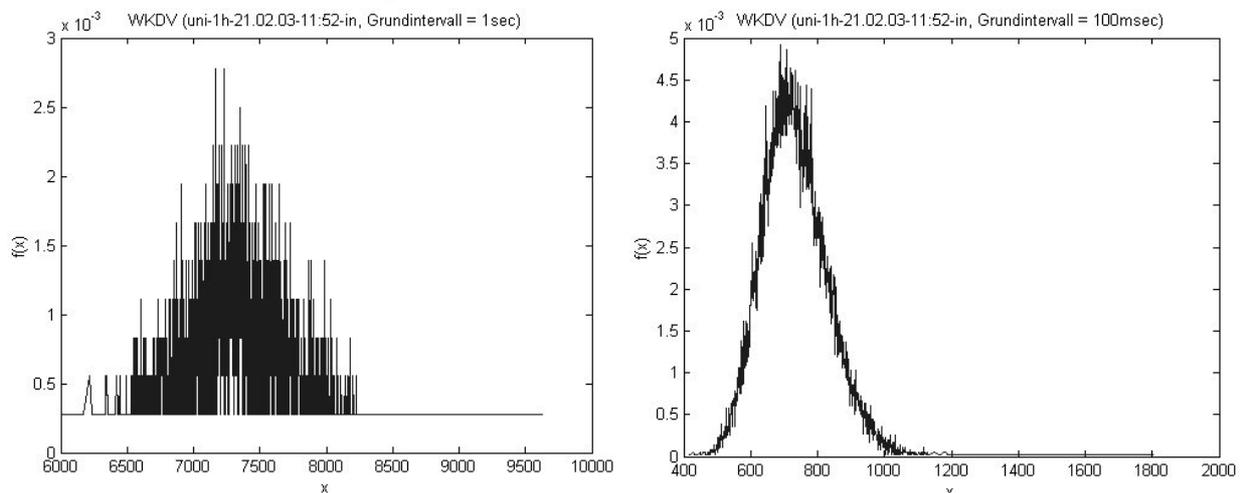


Abbildung 4.7: Vergleich zweier Wahrscheinlichkeitsdichteverteilungen

Zu erkennen ist die WKDV der Bins des Datensatzes uni_1h_21.02.03_11:52_in mit einem Grundintervall von 1 Sekunde bzw. 100 Millisekunden.

Bei einer Bingröße von 1 Sekunde ergeben sich 3600 Bins, deren Wert zwischen 6017 und 9633 liegt. Die Wahrscheinlichkeit daß mehrere Bins den gleichen Wert besitzen, ist aufgrund der Breite und der geringen Anzahl der Bins relativ gering. Es ist kein klar definierter Verlauf in der zugehörigen WKDV zu erkennen. Es ist somit fast unmöglich, die Grenzen über die WKDV zu schätzen. Hier ist die Methode der Bildung mehrerer Mittelwerte anzuwenden.

Bei einer Bingröße von 100msec ist die jedoch WKDV klar definiert, und die Grenzen können auch mit Hilfe dieser geschätzt werden. Da die durch die Ober- und Untergrenzen definierten Bereiche hier sehr viel kleiner sind, findet die Methode der Berechnung mehrerer Mittelwerte nur eingeschränkt Verwendung. Weiterhin ist zu beobachten, daß bei unterschiedlichen Werten für den Parameter T auch die Grenzen weitaus weniger stark schwanken.

Abschließend ist festzustellen, daß die Grenzen mit Hilfe der WKDV zu schätzen sind, wenn der Verlauf der WKDV dieses zuläßt.

4.3 Betrachtete Meßreihen

Analysiert wurden Zeitstempel, die Auskunft über die Zwischenankunftszeiten geben. Dabei lagen die Datensätze in zwei möglichen Formen vor. Zum einen als absolute Zeitstempel im UNIX-Format, bei denen jeder eine genaue Angabe über den Zeitpunkt der Aufnahme enthält. Zum anderen als Differenzen zweier aufeinanderfolgender Timestamps. Informationen über den Zeitpunkt der Aufnahme enthielt hier nur der erste Zeitstempel, da dieser in absoluter Form vorlag.

Der Zeitpunkt der Messung erlaubt eine Zuordnung der Datensätze zu verschiedenen Belastungssituationen. Eine mögliche Implementation der Umwandlung des UNIX-Timestamp-Formats in menschenlesbare Form bietet das MATLAB-Skript „timeconvert.m“.

Der erste Datensatz enthielt 91.135 Timestamps über einen Zeitraum von 12.898,353ms und wurde aufgenommen am 04.04.03 um 10:35:15 Uhr. Mit diesem Datensatz konnte die korrekte Funktionsweise der Approximationsmethoden geprüft werden. Da kleine Datensätze relativ schnell berechnet werden können, stellen Veränderungen im Algorithmus keinen großen zeitlichen Aufwand dar. Auf diesen Datensatz soll jedoch in der Analyse nicht weiter eingegangen werden.

Die nächsten Datensätze betrachten den Ethernetverkehr der Universität Rostock, getrennt nach eingehenden und ausgehendem Verkehr über jeweils eine Stunde. Für folgende Betrachtungen sollen jedoch nur die Bezeichnungen uni_in_1h und uni_out_1h verwendet werden. Diese Bezeichnungen sollen eine schnellere Zuordnung erlauben. Mit Hilfe dieser Datensätze wurde untersucht, ob überhaupt eine Langzeitanhängigkeit nachzuweisen ist.

Weiterhin wurden Datensätze untersucht, die unterschiedliche Netzauslastungen beschreiben. Die Messungen zur Hauptverkehrszeit wurden am 21.02.03 aufgenommen, beginnend um 11:52:17 Uhr. Auch hier wurden wieder der eingehende und der ausgehende Verkehr innerhalb einer Stunde (an der Universität Rostock) getrennt betrachtet. An dieser Stelle sei festgestellt, daß die aufgenommenen Datensätze repräsentativ sind für alle Volllastsituationen in diesem Netz. Der Datensatz des eingehenden Verkehrs zur Hauptverkehrszeit soll im weiteren als uni_1h_21.02.03_11:52_in und der des ausgehenden Verkehrs als uni_1h_21.02.03_11:52_out bezeichnet werden.

Die Messungen außerhalb der Hauptverkehrszeit wurden aufgenommen am 22.02.03 um 01:43:52 Uhr über einen Zeitraum von einer Stunde. Der Datensatz des eingehenden Verkehrs soll im folgenden als uni_1h_22.02.03_01:43_in und der des ausgehenden als uni_1h_22.02.03_01:43_out bezeichnet werden.

In Tabelle 4.1 sind einige wichtige Informationen über die einzelnen Datensätze zusammengefaßt.

Datensatz	Dauer	Anzahl	Max. Zwischenankunftszeit
uni_1h_in	≈ 1h	27.729.715	12.480,7µs
uni_1h_out	≈ 1h	23.079.982	4.190,47µs
uni_1h_21.02.03_11:52_in	≈ 1h	26.361.175	9.338,6µs
uni_1h_21.02.03_11:52_out	≈ 1h	22.369.975	3.910,5µs
uni_1h_22.02.03_01:43_in	≈ 1h	22.183.296	8.497,4 µs
uni_1h_22.02.03_01:43_out	≈ 1h	18.482.111	4.487,0 µs

Tabelle 4.1 einige Eckdaten der untersuchten Datensätze

Erst in den betreffenden Abschnitten wird näher auf die einzelnen Datensätze eingegangen.

4.4 Logarithmische Optimierung (Log-Opt)

Bei der Analyse der Datensätze war zu erkennen, daß die Berechnungszeit ein Faktor von nicht zu unterschätzender Bedeutung ist. Erste Versuche bei der Implementierung der einzelnen Approximationsmethoden ergaben zwar das korrekte Ergebnis, aber bei immer größer werdenden Datensätzen ist die Performanceoptimierung eine weitere zeitaufwendige Aufgabe.

Im Zuge einer nötigen Geschwindigkeitsoptimierung kann eine Begrenzung der Anzahl verschiedener Blockgrößen zu einem verbesserten Ergebnis führen. Die Anforderungen an eine solche Begrenzung ergeben sich aus den Eigenschaften von Graphen mit logarithmisch geteilten Achsen, wie sie beim Variance time plot und bei der RS-Statistik verwendet werden. Die ausgewählten Punkte (hier: Blockgrößen) sollen möglichst gleich verteilt sein und die Anzahl sei so gewählt, daß gute Aussagen getroffen werden können.

Realisiert wurde ein Verfahren, mit dem die Anzahl der zu untersuchenden Blockgrößen auf meist 100 pro Zehnerpotenz begrenzt wird. Die Anzahl der Werte wird durch die Angabe einer Schrittweite bestimmt, die hier 0,01 beträgt. Eine Auflösung von 0,1 entspricht demnach 10 zu untersuchende Blockgrößen pro Zehnerpotenz.

Wie sich die Blockgrößen ergeben, soll folgende Tabelle kurz verdeutlichen:

Auflösung	Berechnete Blockgröße					
0,1	$10^{3,0}$	$10^{3,1}$	$10^{3,2}$	$10^{3,3}$	$10^{3,4}$...
	1000	1258	1584	1995	2511	...
0,01	$10^{3,0}$	$10^{3,01}$	$10^{3,02}$	$10^{3,03}$	$10^{3,04}$...
	1000	1023	1047	1071	1096	...

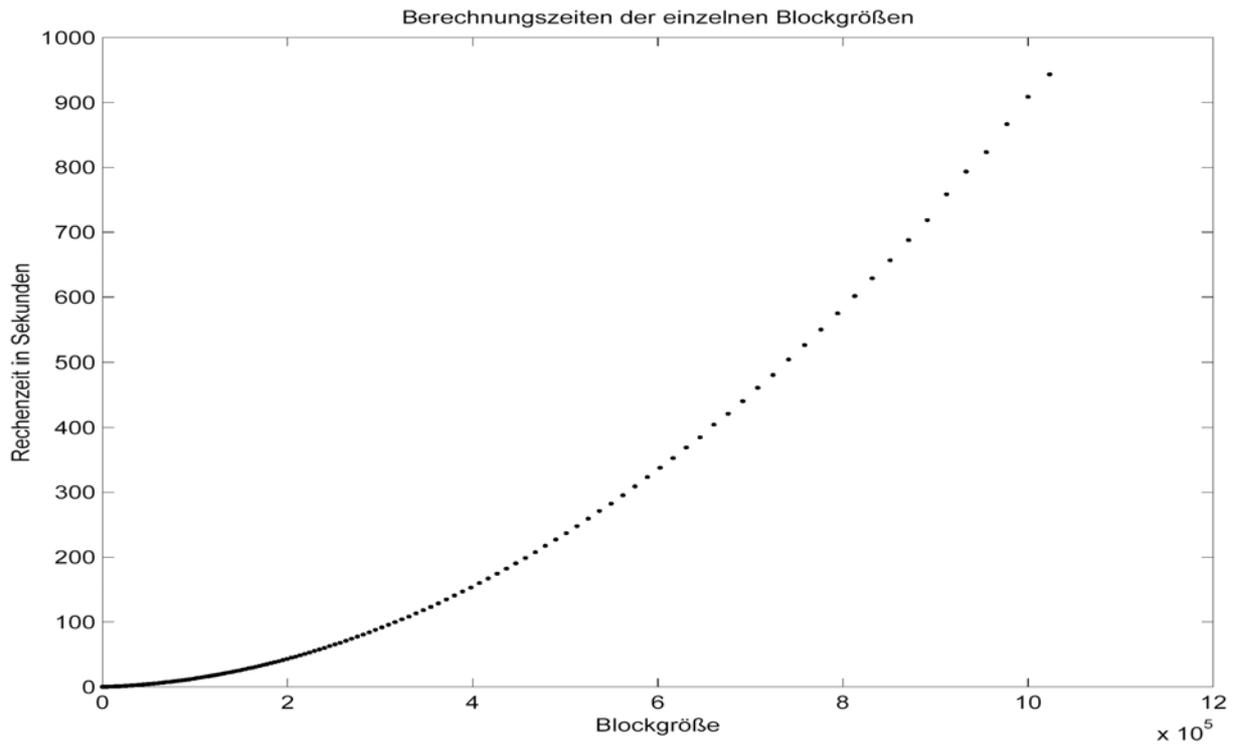
Tabelle 4.2 Beispiel zur Erläuterung

Denkbar ist eine Realisierung in einer for-Schleife bei der mit der angegebenen Schrittweite gezählt wird. Implementierungen sind in verschiedenen Varianten der R/S-Statistik und des Variance-Time-Plot zu finden. Diese sind durch „log_opt“ gekennzeichnet.

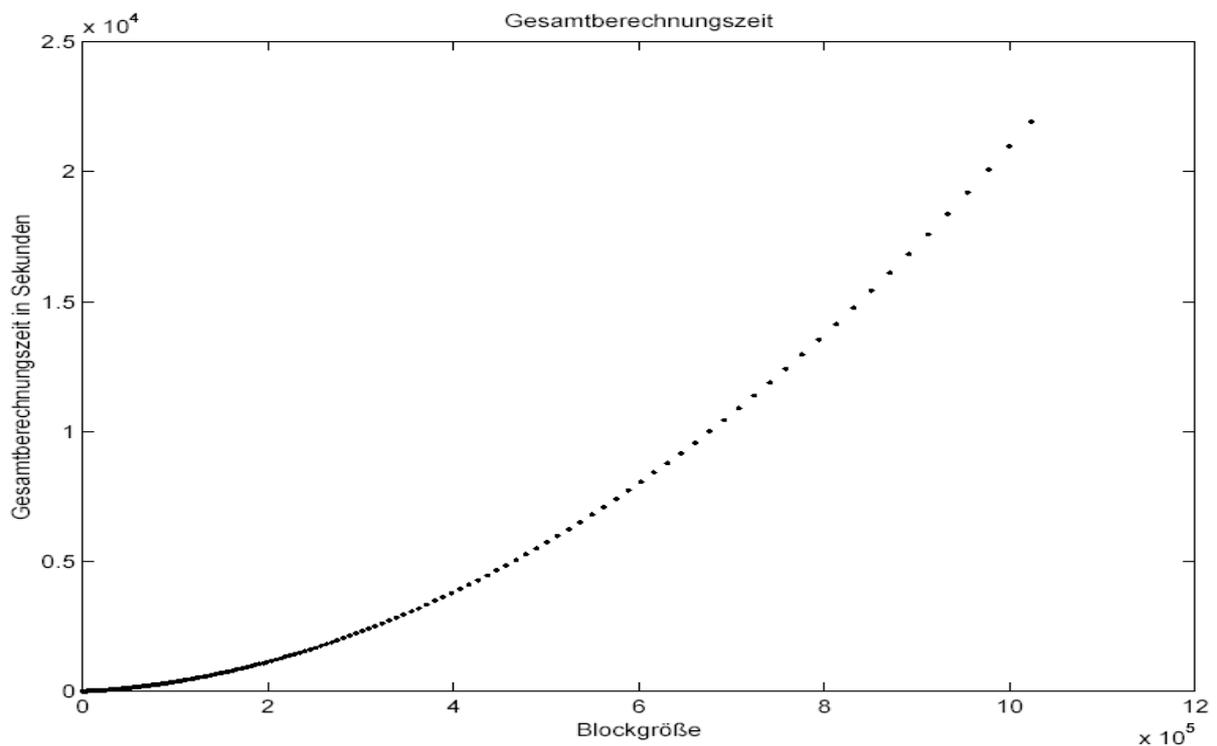
4.5 Berechnung des Hurst Parameters aus Zwischenankunftszeiten

Testberechnungen bei großen Datenmengen haben ergeben, daß eine Abschätzung der Berechnungszeit von Vorteil sein kann. In diesem Zusammenhang wurde versucht den Hurst-Parameter für den Datensatz uni_in_1h mittels R/S-Statistik (log-opt Variante) zu schätzen. Dabei wurden eine Auflösung von 0.01 (entspricht 100 Blockgrößen pro Zehnerpotenz) und acht verschiedene Startpunkte ($t_1..t_8$) verwendet. Die Berechnung wurde nach 21.921 Sekunden abgebrochen, was 6 Stunden 05 Minuten und 35 Sekunden entspricht.

Berechnet wurden 471 von 523 verschiedenen Blockgrößen. Auch wenn 52 verbliebene Blockgrößen im ersten Moment als recht überschaubar anzusehen sind, ist dennoch damit zu rechnen, daß hier die meiste Rechenzeit verbraucht wird. Darstellung 4.8 veranschaulicht den Anstieg der Berechnungszeit mit steigender Blockgröße, und Darstellung 4.9 zeigt den Anstieg der Gesamtberechnungszeit.



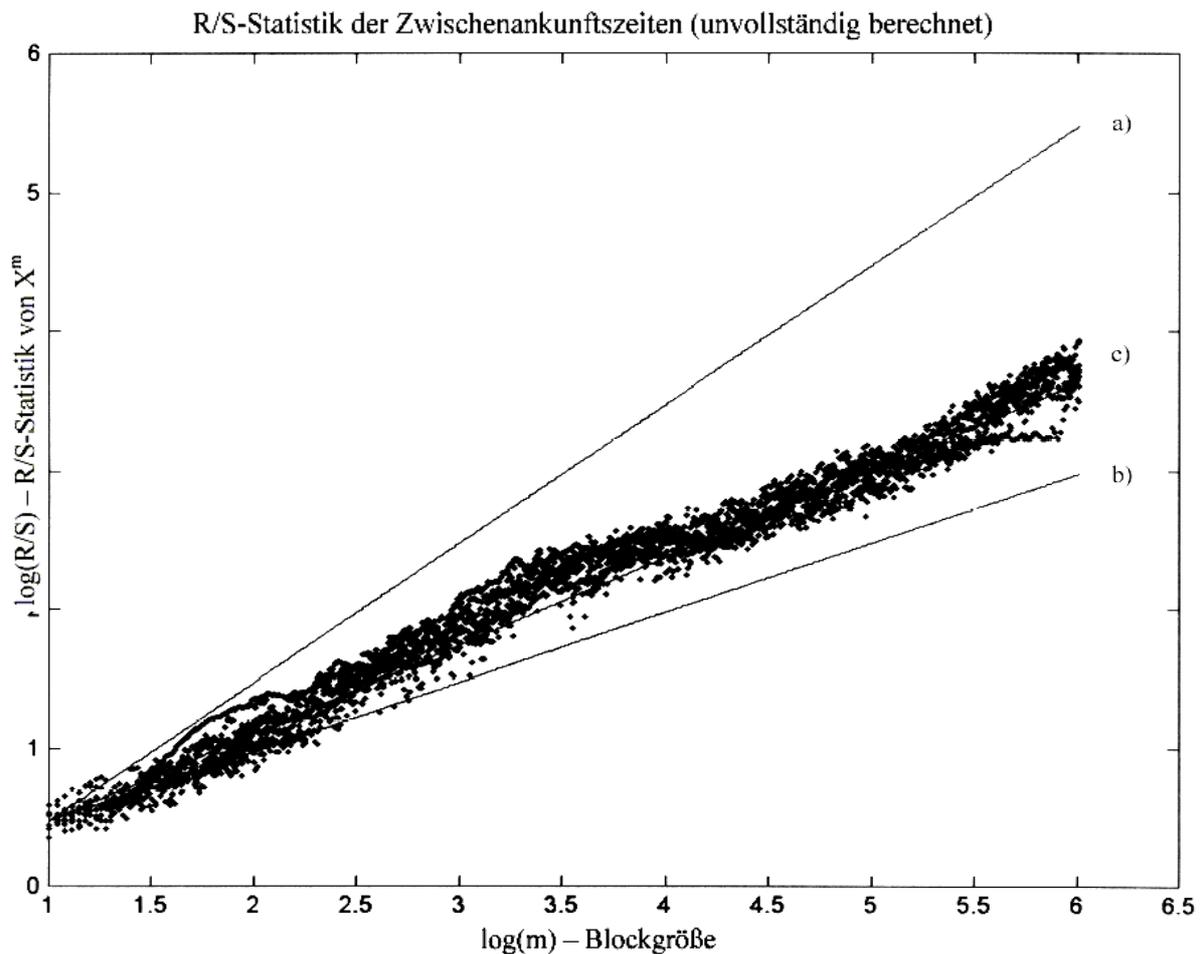
Darstellung 4.8: Blockberechnungszeiten in Abhängigkeit von der Blockgröße



Darstellung 4.9: Gesamtberechnungszeiten in Anhängigkeit von der Blockgröße

Die letzte berechnete Blockgröße enthielt 1.023.292 Zwischenankunftszeiten bei einer Berechnungszeit von 943,24 Sekunden. Die Gesamtberechnungszeit, um den Datensatz vollständig zu analysieren, kann nur geschätzt werden. In den letzten acht Blöcken würden jeweils 3.388.441 Zwischenankunftszeiten ($10^{6,53}$) betrachtet. Schätzungen ergaben, daß die Gesamtberechnungszeit bei 3 Tagen läge. Die Schätzung des Hurst-Parameters mittels Variance-Time-Plot ist noch zeitaufwendiger.

Obwohl die Analyse nicht abgeschlossen wurde, können dennoch einige Aussagen getroffen werden. Darstellung 4.10 zeigt das Ergebnis der „unvollständigen“ Analyse.



Darstellung 4.10: R/S-Statistik der Zwischenankunftszeiten (unvollständig berechnet)
a) Anstieg $\beta=1,0$ b) Anstieg $\beta=0,5$ c) Anstieg durch lineare Regression $\beta=0,6312$.

Der Anstieg des resultierenden Graphen ergibt sich durch lineare Regression. Der Hurst-Parameter beträgt hier $H = 0,6312$. Es ist nicht davon auszugehen, daß die Analyse der verbliebenen Blockgrößen das Gesamtergebnis gravierend verändern kann. Da die Bedingung $0,5 < H \leq 1,0$ erfüllt ist, ist festzustellen, daß die Zwischenankunftszeiten des betrachteten Datensatzes uni_1h_in eine Langzeitabhängigkeit aufweisen.

4.6 Untersuchung auf generelle Langzeitabhängigkeit

Da die Analyse der Datensätze anhand der Zwischenankunftszeiten mit einem extrem hohen Rechenaufwand verbunden war, wurde im weiteren die Datenmenge begrenzt. Hierzu wurden die Zwischenankunftszeiten innerhalb eines Grundintervalls gezählt. Es ergab sich also die Anzahl der Pakete, die innerhalb des gewählten Intervalls eingetroffen sind bzw. gesendet wurden. Dabei ist zu beachten, daß das Grundintervall mindestens die Größe der maximalen Zwischenankunftszeit haben muß. Weitere Informationen hierzu sind in Abschnitt 4.1 zu finden. Die Rechenzeit, um die Bins zu bilden, lag zwischen 1800 und 2100 Sekunden (30-35 Minuten) und war eine annehmbare Zeitspanne, da weitere Berechnungen extrem verkürzt wurden, abhängig von der Größe des Grundintervalls.

Im Abschnitt 4.8 soll jedoch noch die Abhängigkeit des Hurst-Parameters von der Größe des Grundintervalls herausgestellt werden. Es war weiterhin zu prüfen, inwiefern sich die Hurst-Parameter des eingehenden und des ausgehenden Verkehrs unterscheiden.

Zunächst wurden die Datensätze uni_1h_in und uni_1h_out auf Langzeitabhängigkeit geprüft. Als Bingrößen wurden die jeweils maximalen Zwischenankunftszeiten gewählt. Die maximalen Zwischenankunftszeiten und weitere wichtige Eckdaten sind in Tabelle 4.3 zusammengefaßt.

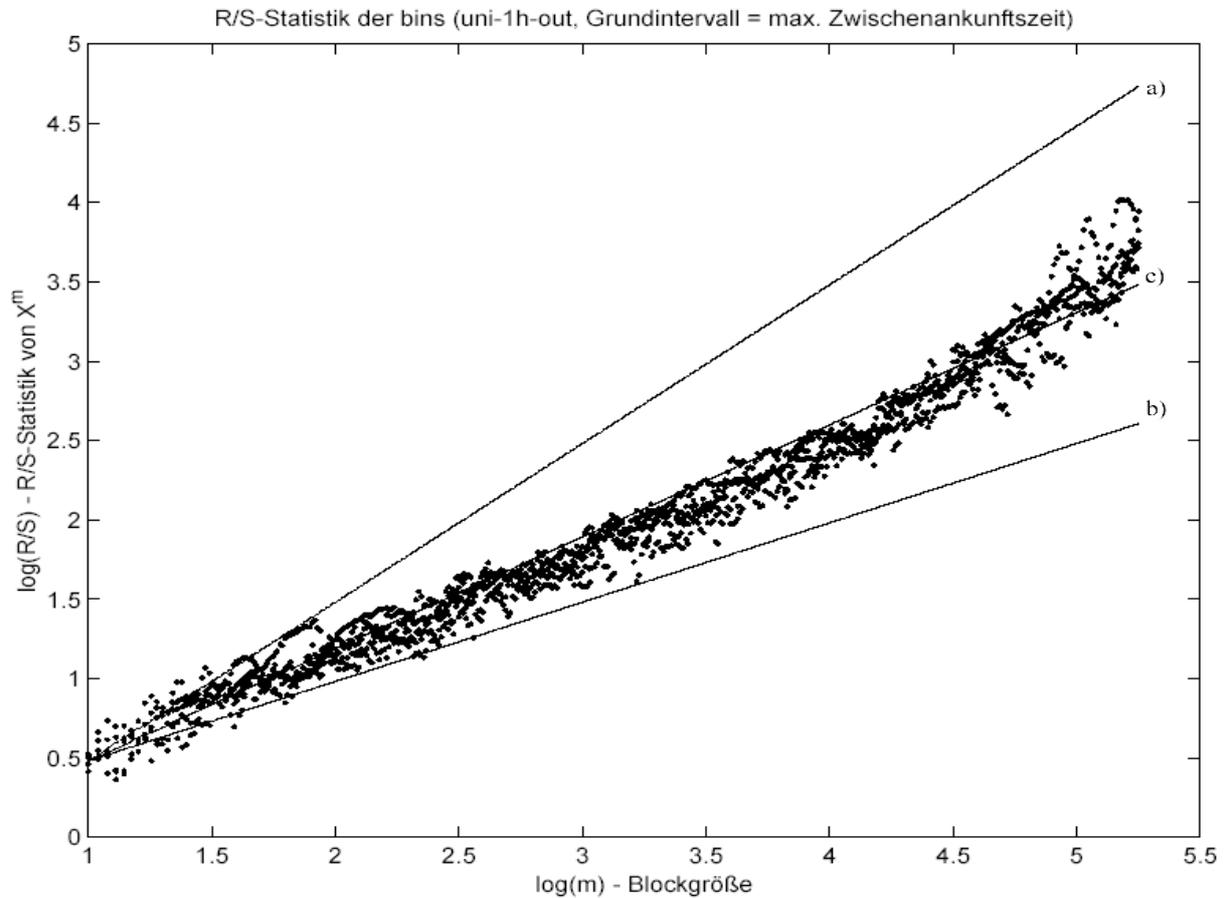
	Uni 1h in	Uni 1h out
Zwischenankunftszeitenanzahl	27.729.715	23.079.982
Arithmetisches Mittel	129,8246 μ s	155,9793 μ s
Bingröße	12.480,7 μ s	4.190,47 μ s
Bin-Anzahl	288.445	894.935
Berechnungszeit der Bins	1.911,5 s	16.646,0 s

Tabelle 4.3: Eigenschaften der Datensätze uni_1h_in und uni_1h_out

Wie aus der obigen Tabelle zu entnehmen ist, unterscheiden sich die maximalen Zwischenankunftszeiten und somit die Größe der Grundintervalle relativ stark. Dieses gilt auch für die Anzahl der Bins und für die Rechenzeit bei der Bin-Bildung.

Für jeden Datensatz wurden jeweils die R/S-Statistik und der Variance-Time-Plot verwendet, um eine Langzeitabhängigkeit nachzuweisen und den Hurst-Parameter zu schätzen.

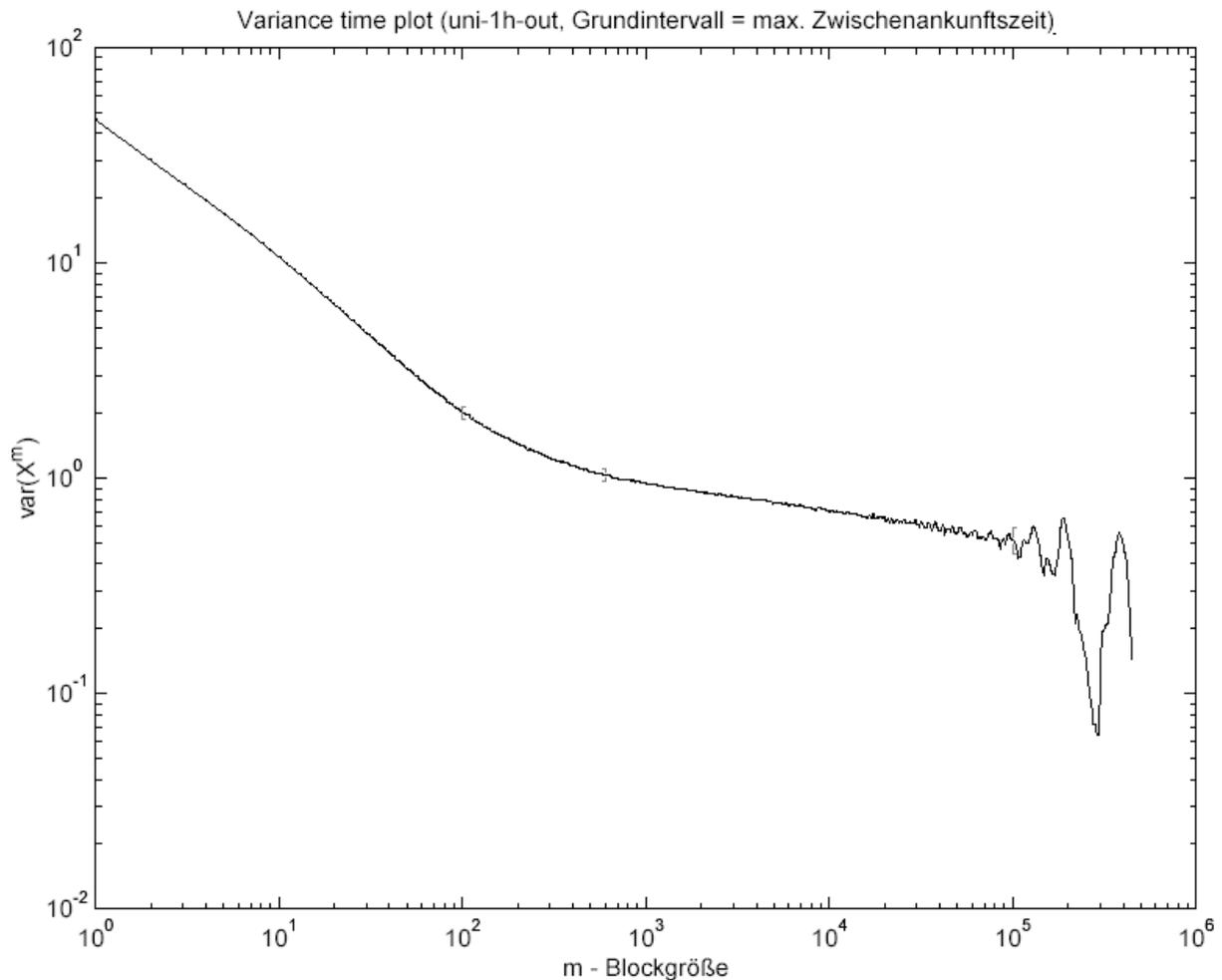
Zunächst sei näher auf die Ergebnisse der Analyse des ausgehenden Verkehrs eingegangen. Die R/S-Statistik für den Datensatz uni_1h_out ist in Darstellung 4.11 abgebildet.



Darstellung 4.11: R/S-Statistik der Bins a) Anstieg $\beta=1.0$ b) Anstieg $\beta=0,5$ c) Anstieg durch lineare Regression $\beta=0,7059$

Der Hurst-Parameter über den Gesamtbereich konnte mittels linearer Regression auf 0,7059 geschätzt werden und ist als sehr guter Wert anzusehen, um Langzeitabhängigkeit nachzuweisen. Die Berechnung der R/S-Statistik ist immer logarithmisch optimiert und resultiert in einer relativ kurzen Berechnungszeit von 842,33 Sekunden. Verwendet wurden dabei eine Auflösung von 0.01 (entspricht 100 Blockgrößen pro Zehnerpotenz) und fünf verschiedene Startpunkte ($t_1..t_5$).

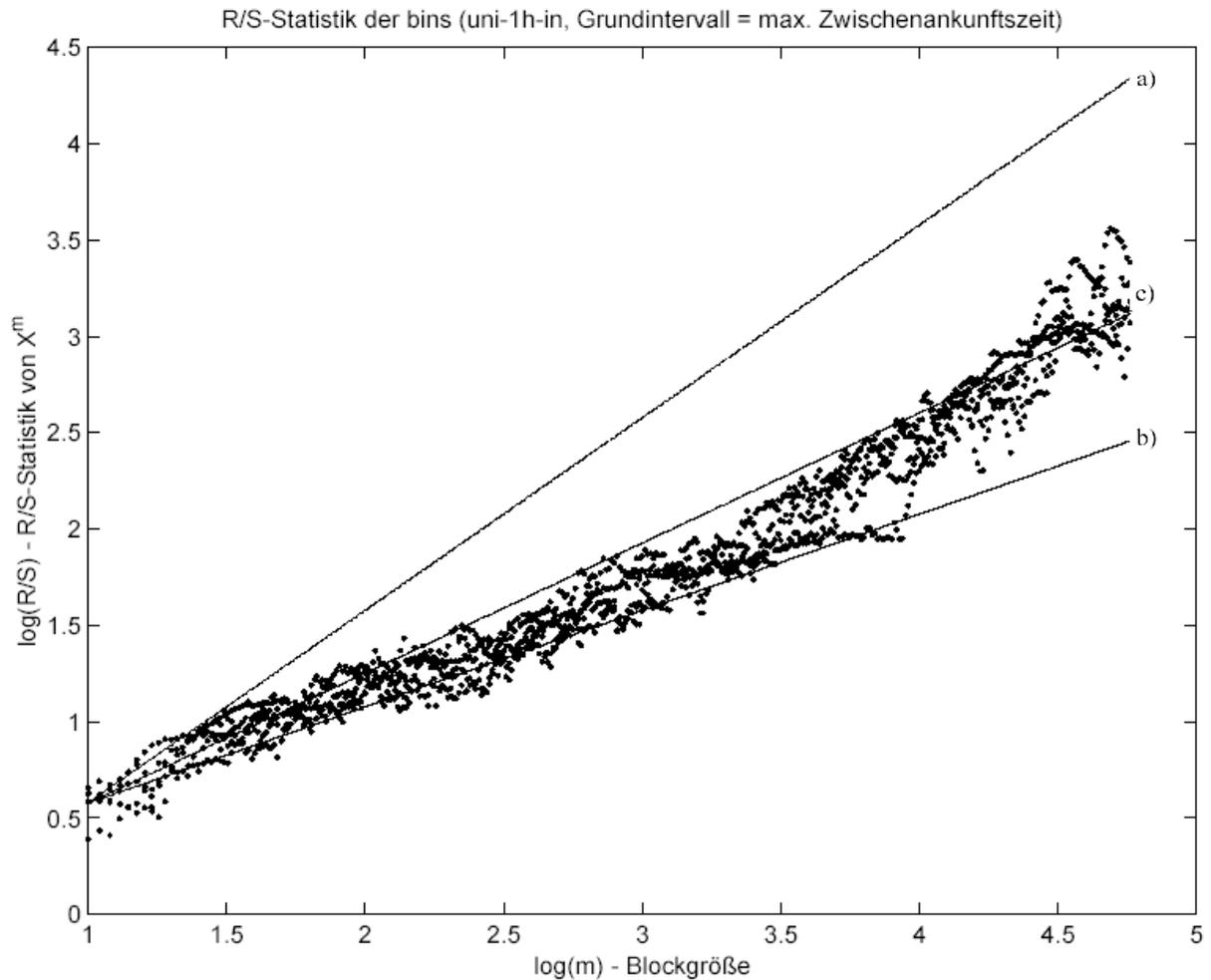
Der VTP für den Datensatz uni_1h_out ist in Darstellung 4.12 abgebildet.



Darstellung 4.12: Variance Time Plot der Bins, ausgehender Verkehr

Obwohl hier schon eine logarithmisch optimierte Version des VTPs verwendet wurde, betrug die Berechnungszeit 10.671,0 Sekunden, bei einer Auflösung von 0.01 (entspricht 100 Blockgrößen pro Zehnerpotenz). Der Verlauf entspricht keinesfalls dem Ideal. Hier ist eine Veränderung des Anstiegs bei den Blockgrößen um 10^2 zu erkennen. In diesem Zusammenhang wurde der Hurst-Parameter für zwei verschiedene Bereiche geschätzt. Der erste Bereich enthält die Blockgrößen von 1 bis 100 (entspricht 10^0 bis 10^2). Durch lineare Regression in diesem Bereich ergab sich der Anstieg zu $\beta = -0,7017$ und der Hurst-Parameter (vgl. Gleichung 3.2) zu $H = 0,6492$. Der zweite Bereich enthielt die Blockgrößen von 600 bis 100.000 (entspricht $10^{2,78}$ bis 10^5). Der Anstieg betrug hier $\beta = -0,1315$, und der Hurst-Parameter ergab sich zu $H = 0,9343$. Auch wenn H im ersten Bereich auf dem ersten Blick als guter Wert erscheint, hat dies jedoch den Nachteil, daß genau dieser Bereich nie verwendet werden sollte, um den Hurst-Parameter zu schätzen. Der Hurst-Parameter für den zweiten Bereich ist eindeutig als zu hoch anzusehen und wäre der zu betrachtende Bereich. Auch wenn über den Hurst-Parameter keine genaue Aussage getroffen werden kann, gilt eine Langzeitabhängigkeit des Datensatzes als nachgewiesen, da beide Bereiche getrennt auf diese hinweisen.

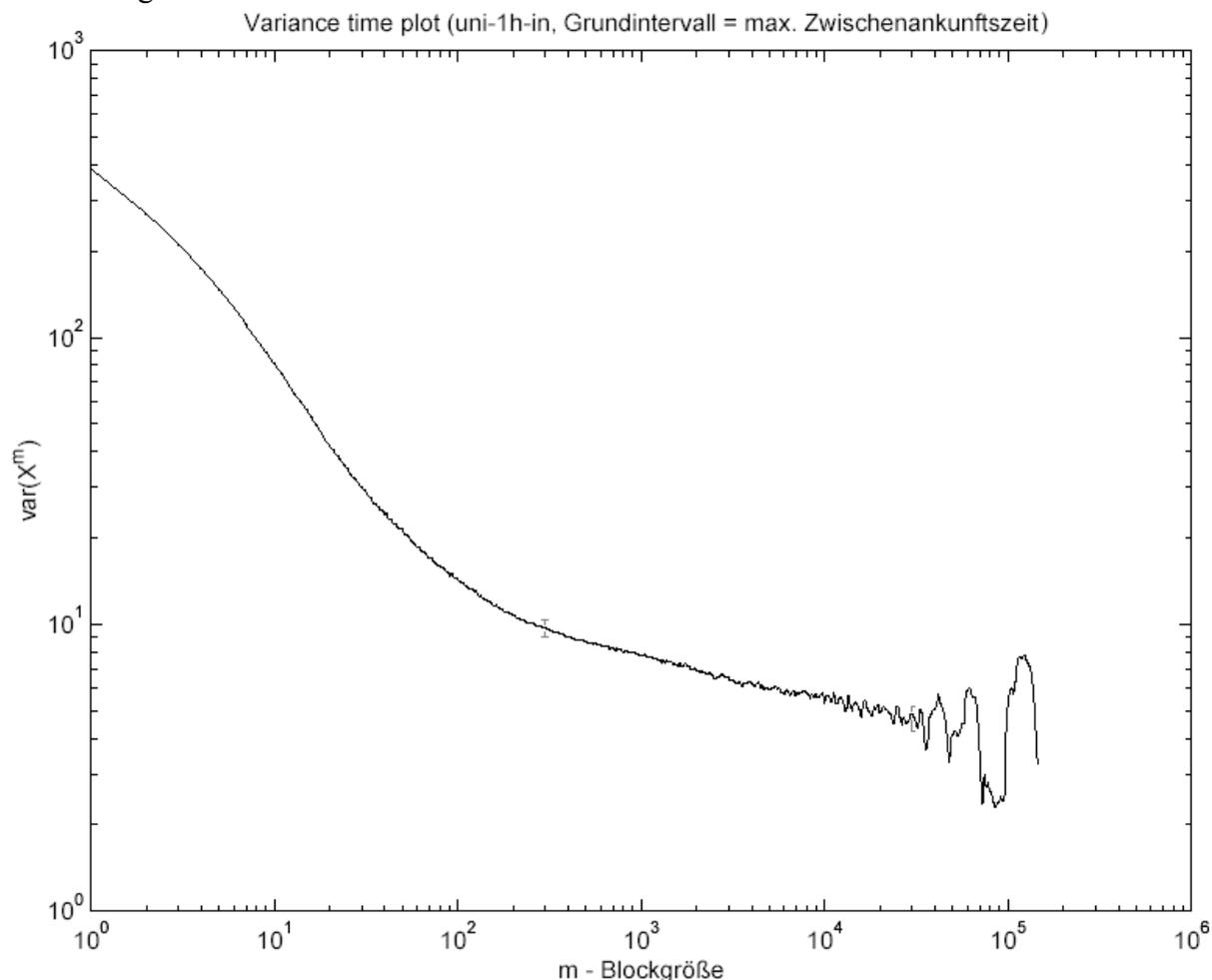
Im Folgenden wurde der Datensatz uni_1h_in analysiert. Darstellung 4.13 zeigt die zugehörige R/S-Statistik.



Darstellung 4.13: R/S-Statistik der Bins a) Anstieg $\beta=1,0$ b) Anstieg $\beta=0,5$ c) Anstieg durch lineare Regression $\beta=0,6742$

Verwendet wurden hier fünf verschiedene Startpunkte ($t_1..t_5$) und eine Auflösung von 0.01. Mittels linearer Regression über den gesamten Graphen konnte auch hier wieder der Hurst-Parameter auf $H = 0,6742$ geschätzt werden. Mit diesem Wert kann davon ausgegangen werden, daß eine Langezeitabhängigkeit existiert. Die Berechnungszeit ist mit 122,98 Sekunden sehr kurz, was mit der geringen Anzahl der berechneten Bins zusammenhängt.

Im nächsten Schritt wurde der VTP des Datensatzes berechnet. Der zugehörige Graph ist der Darstellung 4.14 zu entnehmen.



Darstellung 4.14: Variance Time Plot der Bins, eingehender Verkehr

In diesem Fall wurde sowohl die log-opt-Variante als auch die Variante mit linear ansteigender Blockgröße verwendet. Während die erste Variante nur 947,7 Sekunden benötigt, sind es bei der zweiten bereits 3.729,1. Dabei sind die resultierenden Graphen nahezu identisch. Auch hier entspricht der Verlauf des VTP mit seiner Veränderung des Anstiegs bei den Blockgrößen um 10^2 nicht dem Ideal. Dementsprechend wurde der Hurst-Parameter für zwei verschiedene Bereiche geschätzt. Für den ersten Bereich, der die Blockgrößen 1 bis 300 (entspricht 10^0 bis $10^{2,48}$) enthielt, konnte durch lineare Regression ein Anstieg von $\beta = -0,6533$ ermittelt werden. Somit ist $H = 0,6733$. Der zweite Bereich enthielt die Blockgrößen 300 bis 30.000 (entspricht $10^{2,48}$ bis $10^{4,48}$). Hier ergab sich der Anstieg zu $\beta = -0,1515$, was einem Hurst-Parameter von $H = 0,9243$ entspricht. Auch hier ergab sich wieder das gleiche Problem wie beim betrachteten ausgehenden Verkehr, und es kann nur festgestellt werden, daß eine ermittelte Langzeitabhängigkeit durch die R/S-Statistik sehr wahrscheinlich ist. In Tabelle 4.4 sind noch einmal die Ergebnisse zusammengefaßt.

Datensatz	RS		VTP			Bin-Anzahl
	Hurst	Zeit	Hurst		Zeit	
uni_1h_out	0,7059	842,33s	0,6492	0,9343	10.671,0s	894.935
uni_1h_in	0,6742	122,98s	0,6733	0,9243	947,7s	288.445

Tabelle 4.4: Zusammenfassung der Daten

Zu erwarten war, daß der Hurst-Parameter des eingehenden Verkehrs größer ist als der des ausgehenden. Diese Erwartung resultierte aus dem höheren Verkehrsaufkommen beim eingehenden Verkehr. Die Ergebnisse lassen jedoch keine genauen Rückschlüsse auf einen Zusammenhang zwischen Hurst-Parameter und dem Verkehrsaufkommen zu. Der obigen Tabelle ist zu entnehmen, daß die Hurst-Parameter des eingehenden und des ausgehenden Verkehrs nahezu identisch sind. Weiterhin widersprechen sich die Ergebnisse aus der R/S-Statistik und aus dem VTP sogar. Erschwerend kommt hinzu, daß es sich hierbei nur um Schätzwerte handelt.

4.7 Bestimmung des Hurst-Parameters unter verschiedenen Belastungssituationen

In diesem Abschnitt soll der Grad der Selbstähnlichkeit unter verschiedenen Belastungssituationen betrachtet werden. Zu erwarten ist hier, daß der Hurst-Parameter unter Niedriglast einen geringeren Wert annimmt als unter Vollast.

Untersucht wurden in diesem Zusammenhang die Datensätze uni_1h_21.02.03_11:52_in und uni_1h_21.02.03_11:52_out, die eine mögliche Vollastsituation darstellen, getrennt nach eingehendem und ausgehendem Verkehr. Auf der anderen Seite beschreiben die Datensätze uni_1h_22.02.03_01:43_in und uni_1h_22.02.03_01:43_out eine geringere Datenverkehrsdichte.

Um auch hier die Berechnungszeiten zu verkürzen, wurden Bins gebildet mit der jeweils minimalen Bingröße. Diese ergab sich aus der maximalen Zwischenankunftszeit. In Tabelle 4.5 sind die wichtigsten Eigenschaften der betrachteten Datensätze mit den jeweiligen Bin-Berechnungszeiten zusammengefaßt.

Vollast		
	uni_1h_21.02.03_11:52_in	uni_1h_21.02.03_11:52_out
Zwischenankunftszeitenanzahl	26.361.175	22.369.975
Arithmetisches Mittel	136,4667 µs	160,93 µs
Bingröße	9.338,6 µs	3.910,5 µs
Bin-Anzahl	385.497	920.589
Berechnungszeit der Bins	3.648,0 s	17.392,0 s
Niedriglast		
	uni_1h_22.02.03_01:43_in	uni_1h_22.02.03_01:43_out
Zwischenankunftszeitenanzahl	22.183.296	18.482.111
Arithmetisches Mittel	162,28 µs	194,78 µs
Bingröße	8.497,4 µs	4.487,0 µs
Bin-Anzahl	423.662	802.311
Berechnungszeit der Bins	3.579,5 s	13.067,0 s

Tabelle 4.5: weitere Eigenschaften der zu betrachtenden Datensätze

Wie aus den Tabellen 4.3 und 4.5 zu entnehmen ist, ist davon auszugehen, daß die Datensätze uni_1h_in und uni_1h_out ebenfalls eine Vollastsituation beschreiben. Deutlich wird dieses

durch die Anzahl der Pakete, die empfangen bzw. gesendet wurden. Weiterhin sind auch die arithmetischen Mittelwerte fast identisch.

Wie zu erwarten, unterscheidet sich die Anzahl der Zwischenankunftszeiten unter verschiedenen Belastungssituationen. Jedoch ist hier der Unterschied relativ gering. Bei einem Vergleich beider eingehenden Verkehrsdaten ist zu erkennen, daß zu Niedrigverkehrszeiten immer noch 84,2 % der Anzahl der Pakete der Hauptverkehrszeit übertragen werden.

4.7.1 Vollast

Zunächst wurden die Meßreihen des ausgehenden und des eingehenden Verkehrs unter Vollast analysiert. Die zugehörigen Graphen sind in Abbildung 4.15 dargestellt.

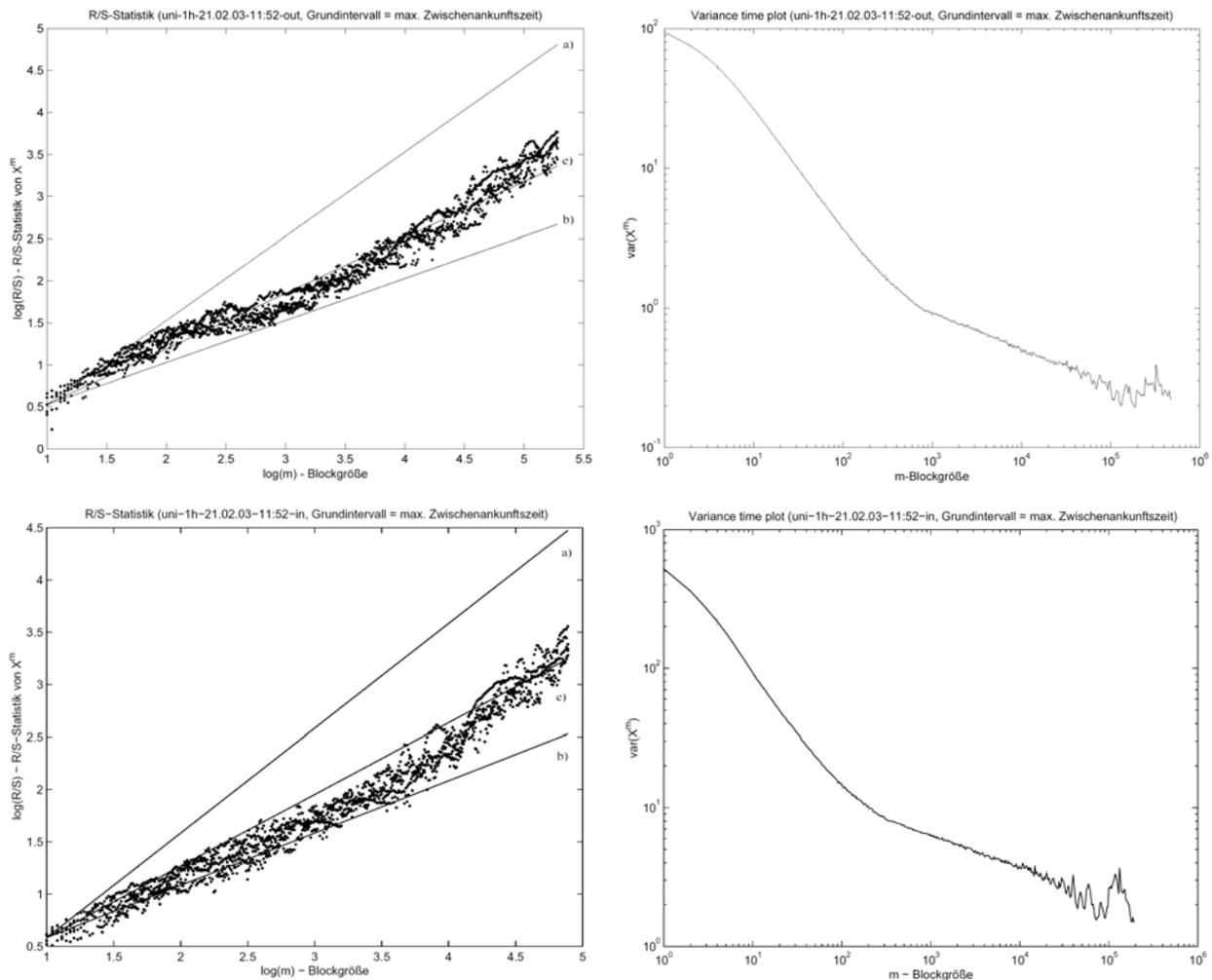


Abbildung 4.15: resultierende Graphen aus der Analyse, Vollast

Dargestellt sind jeweils die R/S-Statistiken (links) und die Variance time plots (rechts). Die oberen Darstellungen zeigen die Graphen des ausgehenden Verkehrs und die unteren die Graphen des eingehenden Verkehrs.

Für die R/S-Statistik beider Datensätze wurden fünf verschiedene Startpunkte ($t_1..t_5$) und eine Auflösung von 0.01 (entspricht maximal 100 Blockgrößen pro Zehnerpotenz) verwendet. Durch lineare Regression ergab sich der Hurst-Parameter für den ausgehenden Verkehr zu $H = 0,6628$ und für den empfangenen Verkehr zu $H = 0,6812$.

Bei der Auswertung der VTP ist zu erkennen, daß die Graphen wiederum keinen idealen Verlauf aufweisen. Bei einem solchen Verlauf sind zunächst zwei potentielle „Fehlerquellen“

denkbar. Entweder liegt ein Fehler in der Implementierung vor oder die analysierten Datensätze weisen dieses Verhalten auf.

Zur näheren Untersuchung wurden mehrere Bereiche gebildet und deren Anstiege berechnet. In Tabelle 4.6 sind die gewählten Bereiche und die zugehörigen Hurst-Parameter aufgelistet.

Datensatz		Bereich (Blockgröße)	Anstieg (β)	Hurst (H)
uni_1h_21.02.03_11:52_out	I	4 bis 300	-0,8195	0,5903
	II	1000 bis 50000	-0,2621	0,8690
uni_1h_21.02.03_11:52_in	I	4 bis 100	-0,8201	0,5900
	II	200 bis 20000	-0,2349	0,8825

Tabelle 4.6: Hurst-Parameter ermittelt durch VTP

Aus der Tabelle ist zu entnehmen, daß beide Graphen im zweiten Bereich einen zu hohen Wert aufweisen. Durch die Ergebnisse aus der R/S-Statistik, kann davon ausgegangen werden, daß der Hurst-Parameter beider Datensätze zwischen 0,6 und 0,7 liegt. Die Ergebnisse aus den Variance-Time-Plots bestätigen zumindest, daß in einer Vollastsituation Langzeitabhängigkeit existiert. In Tabelle 4.7 sind noch einmal alle Ergebnisse anschaulich mit den zugehörigen Berechnungszeiten zusammengestellt.

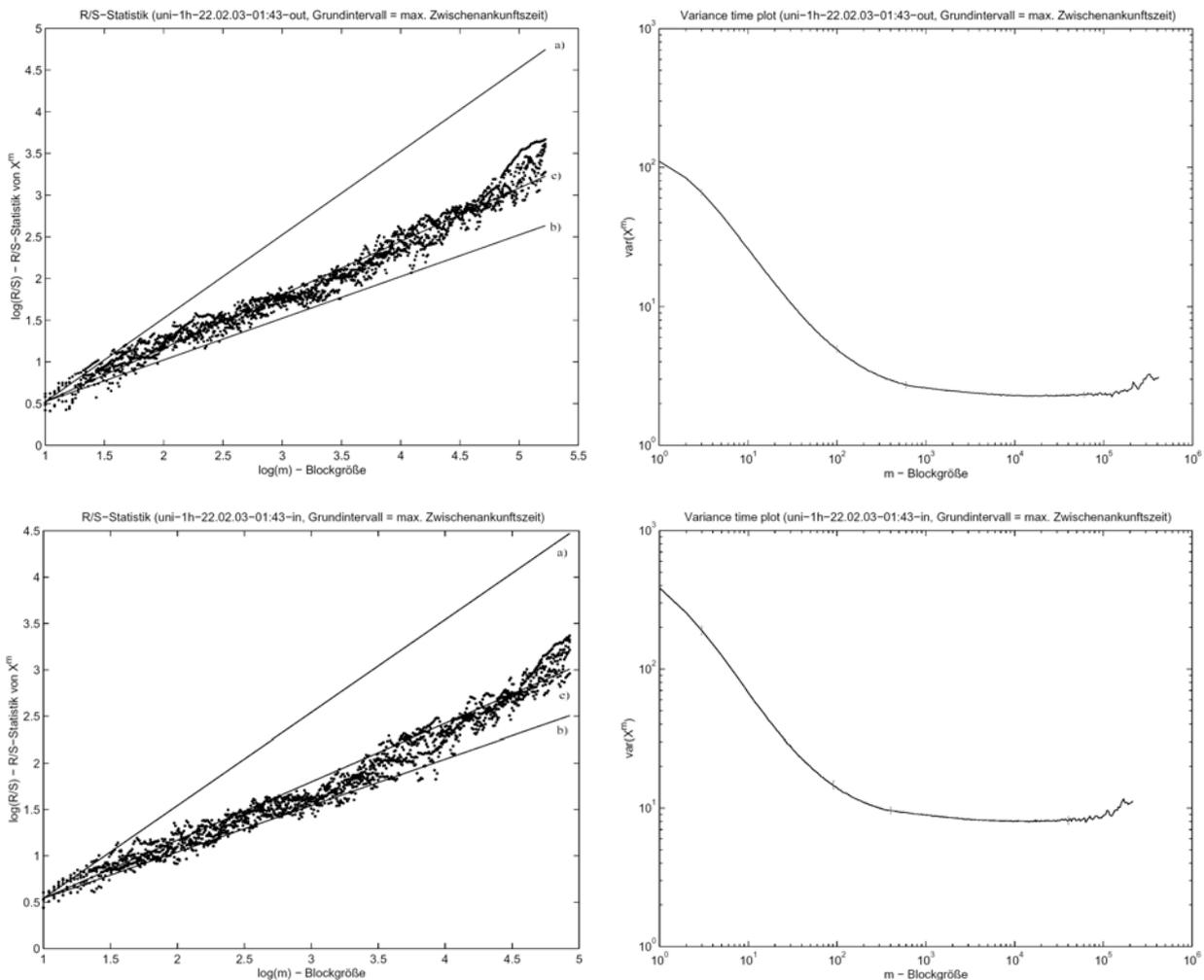
Datensatz	RS		VTP		Bin-Anzahl	
	Hurst	Zeit	Hurst	Zeit		
uni_1h_21.02.03_11:52_out	0,6628	891,67 s	0,5903	0,8690	11.191,0 s	920.589
uni_1h_21.02.03_11:52_in	0,6812	183,62 s	0,5900	0,8825	1.774,93 s	385.497

Tabelle 4.7: Zusammenfassung der Ergebnisse

Zu erkennen ist, daß die Berechnungszeiten für den ausgehenden Verkehr deutlich höher sind als beim eingehenden Verkehr. Dieses ist eindeutig abhängig von der Anzahl der Bins und somit von der maximalen Zwischenankunftszeit.

4.7.2 Niedriglast

Diese Meßreihen wurden außerhalb der Hauptverkehrszeit aufgenommen. Es ist sinnvoll, die Datensätze uni_1h_21.02.03_11:52_out und uni_1h_21.02.03_11:52_in auf die gleiche Weise auf Langzeitabhängigkeit zu untersuchen wie im vorangegangenen Abschnitt. Verwendet wurden wiederum die Methoden der R/S-Statistik und des Variance-Time-Plot. Darstellung 4.16 zeigt die zugehörigen Graphen. Jeweils links sind die R/S-Statistiken dargestellt und rechts die zugehörigen VTP. Die beiden oberen Graphen zeigen die Ergebnisse für den ausgehenden Verkehr und die unteren die Ergebnisse des eingehenden Verkehrs.



Darstellung 4.16: resultierende Graphen aus der Analyse, Niedriglast

Bei der Betrachtung der VTP beider Datensätze ist wiederum ein starke Veränderung im Anstieg zu erkennen. Mehrere Bereiche sind getrennt zu betrachten. In Tabelle 4.8 sind die einzelnen Bereiche mit den resultierenden Anstiegen und Hurst-Parametern dargestellt.

Datensatz		Bereich (Blockgröße)	Anstieg (β)	Hurst (H)
uni 1h 21.02.03 01:43 out	I	3 bis 100	-0,7402	0,6299
	II	400 bis 60000	-0,0412	0,9794
uni 1h 21.02.03 01:43 in	I	3 bis 90	-0,7276	0,6362
	II	400 bis 40000	-0,0337	0,9831

Tabelle 4.8 Hurst-Parameter ermittelt durch VTP

Besonders auffällig ist bei beiden Datensätzen der zweite Bereich. Hier liegt ein Anstieg von etwa 0 vor, was einem Hurst-Parameter von fast 1 entspricht. Dieser Wert ist erneut als zu hoch einzuschätzen.

Der Hurst-Parameter des ausgehenden Verkehrs, ermittelt durch die R/S-Statistik, beträgt $H = 0,6402$ und des eingehenden Verkehrs $H = 0,6327$. Beide Werte sind als realistisch anzusehen und deuten in Verbindung mit dem Variance-Time-Plot eindeutig auf Langzeitabhängigkeit in beiden Datensätzen hin.

Tabelle 4.9 faßt die Hurst-Parameter noch einmal zusammen.

Datensatz	RS		VTP			Bin-Anzahl
	Hurst	Zeit	Hurst		Zeit	
uni_1h_22.02.03_01:43_out	0,6402	664,15 s	0,6299	0,9794	7.973,1 s	802.311
uni_1h_22.02.03_01:43_in	0,6327	220,80 s	0,6362	0,9831	2.049,0 s	423.662

Tabelle 4.9: Zusammenfassung der Ergebnisse

Im Vergleich zu den Ergebnissen in Tabelle 4.7 ist zu erkennen, daß die durch die R/S-Statistik ermittelten Hurst-Parameter bei Vollast etwas höher ausfallen. Jedoch sind diese Unterschiede weniger gravierend als zu erwarten war. Der VTP wird den Erwartungen nicht gerecht.

4.8 Abhängigkeit des Hurst-Parameters von der Größe des Grundintervalls

Die Auswertung der Datensätze in den vorherigen Abschnitten lassen den Schluß zu, daß die Berechnungszeiten für die R/S-Statistik und den Variance-Time-Plot abhängig sind von der Anzahl der gebildeten Bins. Weiterhin kann davon ausgegangen werden, daß eine Abhängigkeit zwischen der Zeit zur Berechnung der Bins und der Größe des Grundintervalls existiert. Im Folgenden wurde nun geprüft, welche Vor- und Nachteile sich durch die Wahl des geeigneten Grundintervalls ergeben. Untersucht wurde in diesem Zusammenhang repräsentativ der Datensatz uni_1h_in. Weiterhin wurde hier zur Schätzung des Hurst-Parameters nur die R/S-Statistik verwendet, da mittels VTP keine genauen Aussagen über den Hurst-Parameter getroffen werden konnten (vgl. Abschnitt 4.6).

Betrachtet wurden verschiedene Grundintervalle, die den Bingrößen entsprechen. Weiterhin wurden die Berechnungszeiten dokumentiert und der Hurst-Parameter geschätzt für die jeweiligen Bins. Tabelle 4.10 zeigt die Ergebnisse der Meßreihen.

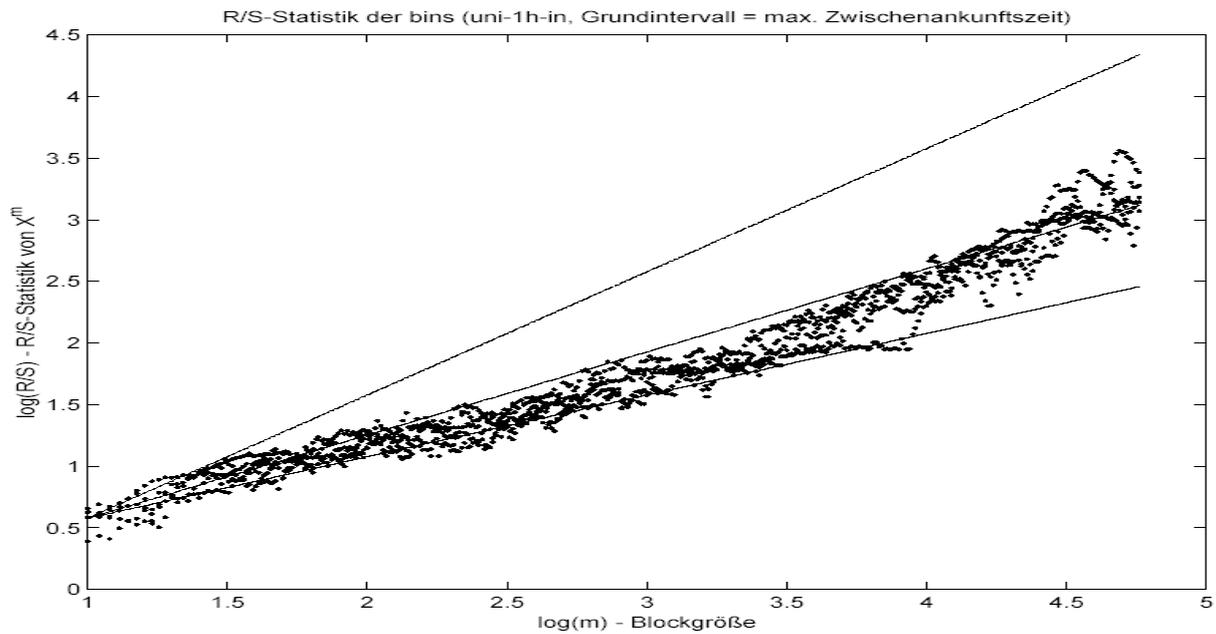
Bingröße	Bin-Anzahl	Hurst	Berechnungszeiten in Sekunden		
			bins	R/S	VTP
12.480,7 μ s	288.445	0,6742	1.911,5	123,0	947,7
15.000 μ s	240.000	0,6828	1.338,6	94,7	667,1
20.000 μ s	180.000	0,6919	841,1	60,4	369,2
25.000 μ s	144.000	0,7045	630,4	42,7	230,8
35.000 μ s	102.858	0,7228	418,0	27,1	109,2
50.000 μ s	72.000	0,7459	313,3	17,2	47,2
100.000 μ s	36.000	0,7730	246,5	8,6	14,7

Tabelle 4.10: Zusammenfassung der gemessenen Zeiten und der geschätzten Hurst-Parameter

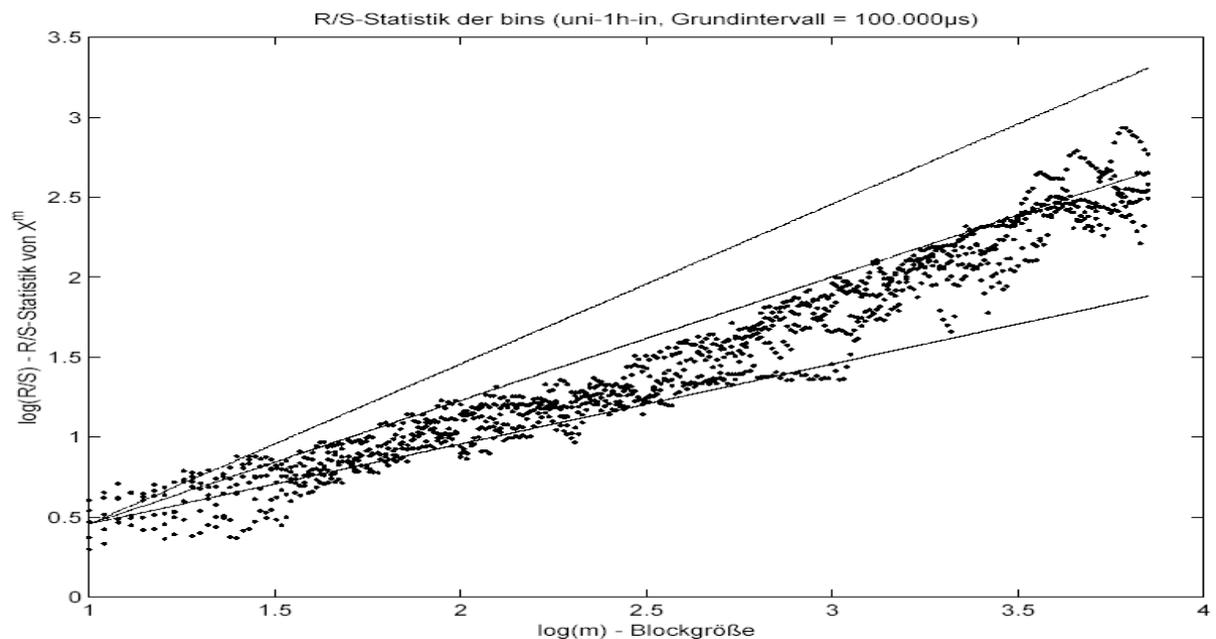
Die Hurst-Parameter ergaben sich durch lineare Regression der Graphen über den Gesamtbereich.

Aus der obigen Tabelle ist zu entnehmen, daß bei steigender Bingröße die Berechnungszeiten teilweise extrem verkürzt werden. Dieses kann bei großen Datenmengen mit sehr langen Netzwerkverkehrsobservationszeiten von großem Vorteil sein, da so die Berechnungszeiten in einem überschaubaren Rahmen bleiben. Nachteilig erweist sich jedoch der zunehmenden größer werdende Hurst-Parameter bei steigender Bingröße. Je größer das Grundintervall gewählt wird, desto ungenauer wird die Aussage über den Hurst-Parameter, auch wenn dieser niemals mit einem festen Wert angegeben werden kann. Eine Unabhängigkeit des Hurst-Parameters von der Bingröße kann hier zumindest nicht nachgewiesen werden.

Die Darstellungen 4.17 und 4.18 sollen den Unterschied noch einmal verdeutlichen.



Darstellung 4.17: R/S-Statistik mit min. Grundintervall

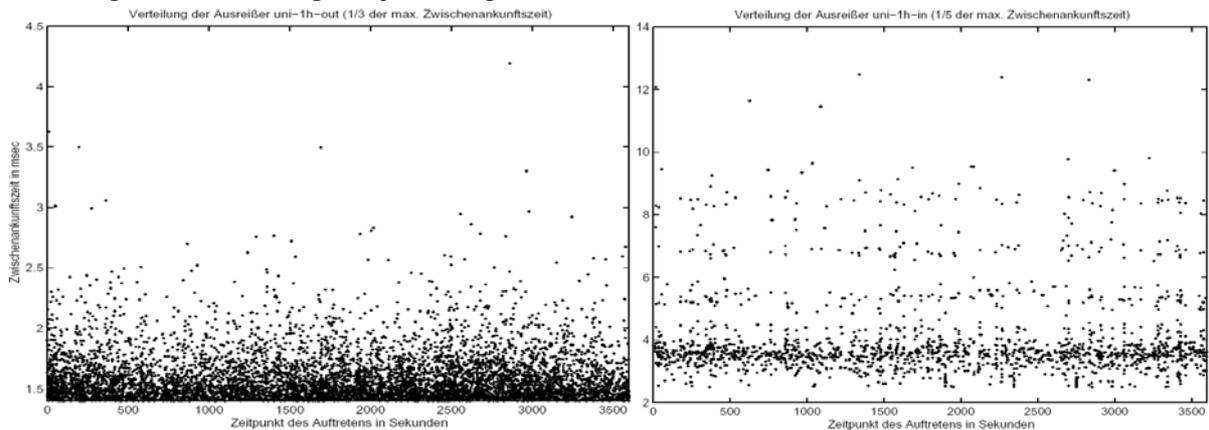


Darstellung 4.18: R/S-Statistik mit Grundintervall = 100.000µs

Mit der Erhöhung des Grundintervalls kommt es zu einer Skalierung des resultierenden Graphen sowohl auf der x- als auch auf der y-Achse. Beide Skalierungen sind jedoch unterschiedlich stark ausgeprägt und resultieren in verschieden großen Anstiegen. Die „unvollständige“ Analyse aus Abschnitt 4.5 ergab $\beta = H = 0,6312$ und deutet darauf hin, daß die Binsgröße so klein wie möglich zu halten ist, wenn Aussagen über die Langzeitabhängigkeit eines Datensatzes zu treffen sind.

4.9 Abweichungen in den Meßreihen

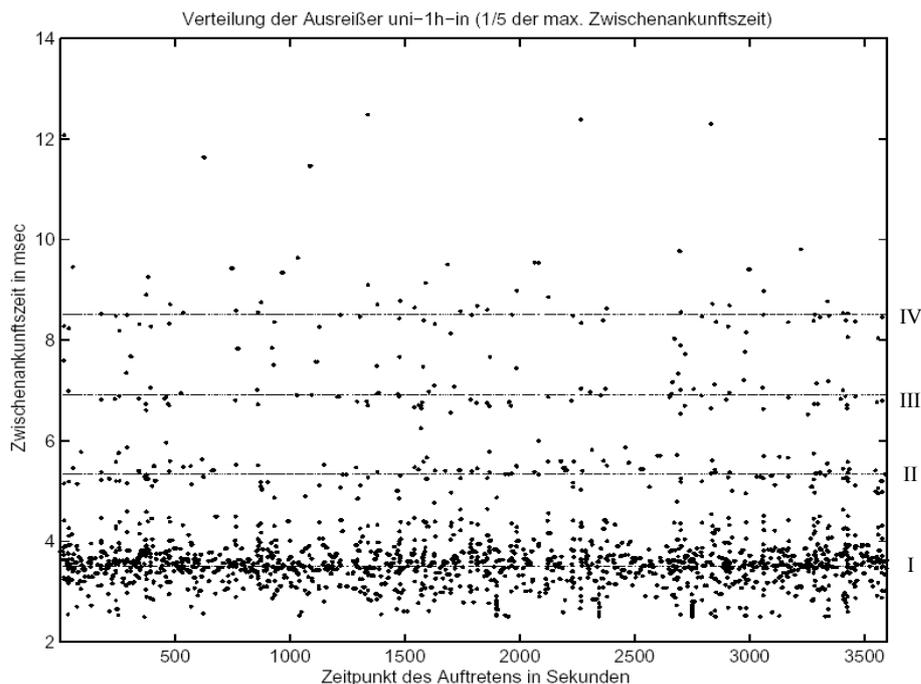
Im folgenden sollen die Abweichungen in den Meßreihen untersucht werden. Hierzu zählen Zwischenankunftszeiten, die einen auffallend hohen Wert aufweisen. Diese sind als „Ausreißer“ anzusehen. In den Datensätze uni_1h_out und uni_1h_in wurde zunächst nach einzelnen Zwischenankunftszeiten mit außergewöhnlich hohem Wert gesucht. Darstellung 4.19 zeigt die Verteilung der jeweils größten Zwischenankunftszeiten.



Darstellung 4.19 Bindarstellung der größten Werte (links: ausgehend, rechts: eingehend)

Betrachtet wurden hierbei nur Zwischenankunftszeiten, die größer sind als ein vorgegebener Schwellenwert. Für den Datensatz uni_1h_out wurden alle Zwischenankunftszeiten betrachtet, die größer sind als $1/3$ der maximalen Zwischenankunftszeit. Für den Datensatz uni_1h_in lag der Schwellenwert bei $1/5$ der maximalen Zwischenankunftszeit. Aus den Graphen ist zu entnehmen, daß keine der Zwischenankunftszeiten in den Datensätzen als „echte“ Ausreißer anzusehen sind, die während der Aufnahme der Meßreihen auf technische Fehler hinweisen könnten. Jedoch sind im eingehenden Verkehr klare Abstufungen zu erkennen.

Darstellung 4.20 zeigt eine Vergrößerung des betreffenden Graphen mit Geraden, die die Abstufungen verdeutlichen sollen.



Darstellung 4.20:
Vergrößerung mit
geschätzten Geraden

Aus den ermittelten Stufen lassen sich nun die zugehörigen Mittelwerte berechnen. Tabelle 4.11 zeigt die Mittelwerte und die Abstände zwischen den einzelnen Stufen.

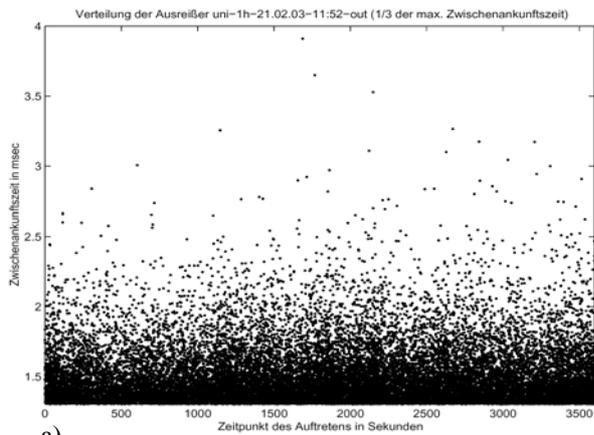
Stufe	Mittelwert	Bereich	Abstand
I	3.5085 ms		
II	5.3407 ms	I – II	1,8322 ms
III	6.9190 ms	II – III	1,5783 ms
IV	8.5173 ms	III – IV	1,5982 ms

Tabelle 4.11: Zusammenfassung der Stufen und deren Abstände

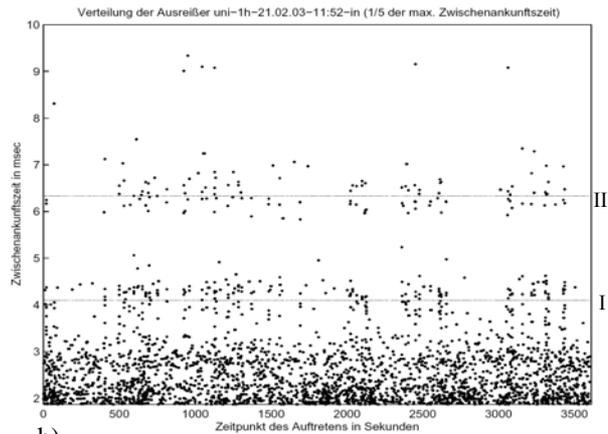
Bei der Betrachtung der Abstände ist zu erkennen, daß sie voneinander eindeutig abweichen. Dieses kann damit zusammenhängen, daß wahrscheinlich nicht jeder Punkt der richtigen Stufe zugeordnet werden kann. Auf der anderen Seite müssen die Abstände nicht unbedingt einem festen Schema folgen.

Es war naheliegend, weitere Datensätze nach Ausreißern und Abstufungen zu untersuchen. In diesem Zusammenhang wurden auch die Datensätze mit verschiedenen Belastungssituationen analysiert. An den maximalen Zwischenankunftszeiten in den Datensätzen ist zu erkennen, daß wohl keine (einzelne) Zwischenankunftszeit signifikant von den anderen abweicht und auf technische Probleme während der Aufnahme der Meßwerte hinweist.

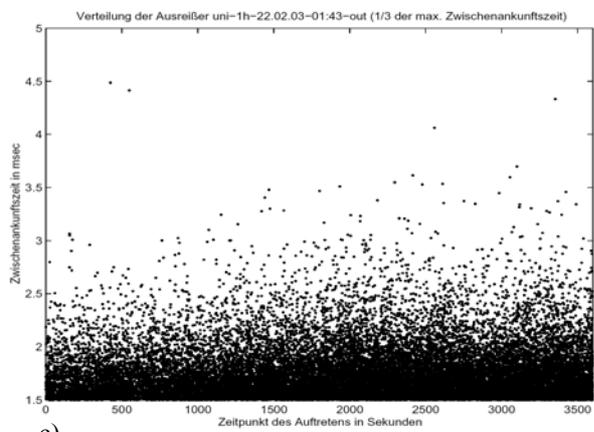
Um nun Abstufungen erkennen zu können, ist es nötig, die zeitliche Verteilung der jeweils größten Zwischenankunftszeiten darzustellen. Darstellung 4.21 zeigt diese Verteilung für den jeweiligen Datensatz.



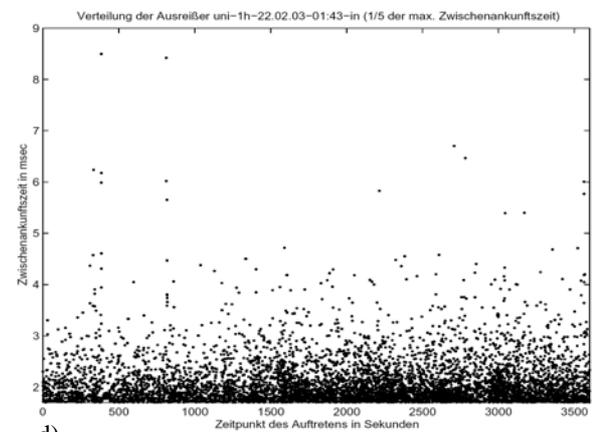
a)



b)



c)



d)

Darstellung 4.21: Zeitliche Verteilung der jeweils größten Zwischenankunftszeiten
oben: Vollast, unten: Niedriglast, links: ausgehender, rechts: eingehender Verkehr.

Anhand der Darstellung ist festzustellen, daß der ausgehende Verkehr keiner Abstufung unterliegt. Beim jeweils eingehenden Verkehr sind leichte Abstufungen zu erkennen. Selbst unter Niedriglast treten diese auf, auch wenn sie schwer zu identifizieren sind. Darstellung 4.21b zeigt zwei klar erkennbare Stufen. Die erste liegt bei 4,0976 ms und die zweite bei 6,3295 ms. Der Abstand beider Stufen beträgt 2,2319 ms. Dieser Wert paßt zu keinem der Abstände aus Tabelle 4.11, was die Vermutung bestätigt, daß die Abstände keinem bestimmten Schema folgen.

Alles deutet auf eine Überlastkontrolle im gemessenen System hin, da die Abstufungen besonders stark unter Vollbelastung auftreten. Zu erkennen ist dieses an der Anzahl der ermittelten Stufen.

4.10 Simulationsvorbereitende Analysen

In den nächsten Abschnitten werden einige Analysen betrachtet, die den Zweck erfüllen, Werte für eine anschließende Simulation zu gewinnen.

4.10.1 Analyse des Datensatzes uni_1h_21.02.03_11:52_in

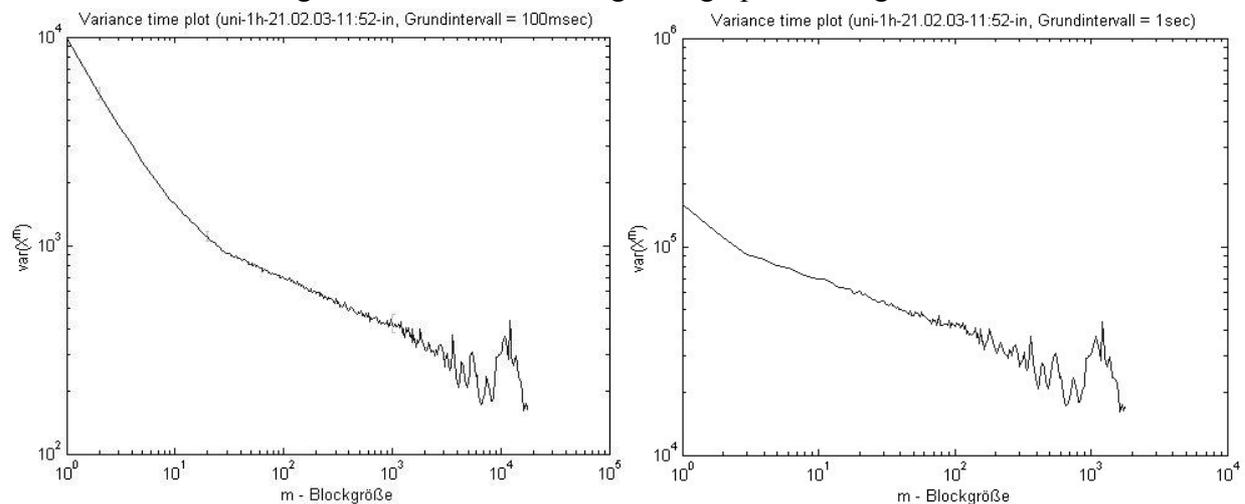
In diesem Abschnitt soll noch einmal der Datensatz uni_1h_21.02.03_11:52_in gesondert betrachtet werden mit den Bingrößen 100ms und 1sec, da beide Bingrößen in der späteren Simulation vorkommen. Folglich sind bei einer Meßdauer von einer Stunde aus den Zwischenankunftszeiten des Datensatzes 36000 bzw. 3600 Bins zu berechnen. Die Berechnung der Bins kann mit dem Skript „Binberechnung.m“ erfolgen. Nähere Informationen sind in Abschnitt 4.1 zu finden.

Aus den Ergebnissen dieser Analyse ergeben sich später die Eingangsparameter für die Simulation mittels des Raw Packet Generator in OPNET. Hierzu sei auf Abschnitt 5.4 verwiesen. Für die Simulation sind folgende Parameter erforderlich:

- die mittlere Ankunftsrate (μ)
- der Hurst-Parameter (H)
- die Fractal Onset Time Scale (kurz FOTS oder T)

Für die Bestimmung der mittleren Ankunftsrate ist der Mittelwert der Bins zu berechnen für die jeweilige Bingröße. Bei einem Grundintervall von 100msec beträgt die mittlere Ankunftsrate somit $\mu \approx 732.26 \frac{\text{Pakete}}{100\text{ms}}$ und bei 1sec $\mu \approx 7322.55 \frac{\text{Pakete}}{\text{s}}$.

Die Schätzung der Hurst-Parameter erfolgte wiederum mittels R/S-Statistik und Variance Time Plot. Es ist jedoch zu prüfen, ob die ermittelten Hurst-Parameter aus dem Variance Time Plot auch zu geeigneten Werten für die Simulation führt. Die Variance Time Plots für die verschiedenen Bingrößen sind in Abbildung 4.22 graphisch dargestellt.



Darstellung 4.22: Variance Time Plots bei verschiedenen Bingrößen (links 100ms, rechts 1000ms)

Es ist zu erkennen, daß bei einer Bingröße von 100ms der bereits aus vorherigen Abschnitten bekannte „Knick“ im Verlauf auftritt. Bei einer Bingröße von 1sec ist jedoch die Veränderung im Anstieg weitaus geringer. Bei einem Grundintervall von 100ms sind somit verschiedene Bereiche zu betrachten. Da der ermittelte Hurst-Parameter aus der R/S-Statistik keine weiteren Auffälligkeiten aufweist, sind in Tabelle 4.12 die Ergebnisse zur Schätzung des Hurstparameters zusammengefaßt. Auf eine graph. Darstellung sei verzichtet.

Bingröße [ms]	μ [Pakete/Bingröße]	R/S (Hurst)	VTP		\bar{H}
			Bereich	Hurst	
100	732,26	0,7806	2 bis 20	0,6851	0,7806
			20 bis 2000	0,8824	
			2 bis 2000	0,8527	
1000	7322,55	0,9177	2 bis 200	0,8841	0,9009

Tabelle 4.12: Zusammenfassung der ermittelten Parameter

Bei der Betrachtung des Variance-Time-Plot bei einer Bingröße von 100ms ist zu erkennen, daß kein Wert für den Hurst-Parameter festgelegt werden kann. Die Entscheidung für einen bestimmten Bereich ist zu fehlerbehaftet. Somit ist hier der endgültige Hurst-Parameter derjenige aus der R/S-Statistik, welcher als sehr guter Wert zu sehen ist.

Bei einer Bingröße von 1000ms ist nur ein Bereich im VTP zu betrachten. Dieses resultiert in einen Hurst-Parameter, der sehr gut mit dem aus der R/S-Statistik vereinbar ist. Allerdings ist es fraglich, ob H wirklich auf einen solchen Wert ansteigen kann. Die Abhängigkeit des Hurst-Parameters von der Bingröße ist, wie in Abschnitt 4.8 gezeigt, gut zu erkennen.

Weiterhin wurde nun versucht, einen Wert für den Parameter T zu bestimmen. Die hier durchgeführten Betrachtungen beruhen jedoch nicht auf den Aussagen aus Abschnitt 2.2.4.2, da ein praktischer Wert dieser Formeln im Zusammenhang mit dieser Arbeit kaum auszumachen ist. Verwendet wurden hier Erklärungsversuche in RL[00], XY[01] u.a. . In diesen Arbeiten wird die Fractal Onset Time Scale definiert als der „kleinste“ Zeitbereich, in der selbstähnliches Verhalten auftritt. Diese Definition ist jedoch sehr weit gefaßt. Somit ist die Schätzung des Parameters T (Fractal Onset Time Scale) nicht eindeutig zu bewerkstelligen, da nur 3600 bzw. 36000 Bins zu betrachten sind. Bei einer Bingröße von 1sec ergeben die ersten 360 Bins noch einen auf Selbstähnlichkeit hinweisenden Hurst-Parameter (durch R/S-Statistik). Werden jedoch nur die ersten 36 Bins auf Selbstähnlichkeit untersucht, so kann hier kein Hinweis darauf festgestellt werden. Somit ist es naheliegend, daß der Parameter T entweder 0,1 beträgt oder zwischen 0,01 und 0,1 liegt. Genauere Ergebnisse ergeben sich erst aus der Simulation.

4.10.2 Analyse der Datensätze uni_1h_22.02.03_01:43_in und uni_1h_22.02.03_01:43_out

Um eine gleichzeitige Simulation des eingehenden und des ausgehenden Verkehrs zu ermöglichen, wurden noch einmal die Datensätze uni_1h_22.02.03_01:43_in und uni_1h_22.02.03_01:43_out mit einer Bingröße von 1000ms betrachtet. Bei der Berechnung der Bins mittels „Binberechnung.m“ ergeben sich bei einer Meßdauer von einer Stunde genau 3600 Bins.

Aus den ermittelten Bins sind wiederum Werte für

- die mittlere Ankunftsrate (μ)
- den Hurst-Parameter (H)
- die Fractal Onset Time Scale (T bzw. FOTS)

festzulegen.

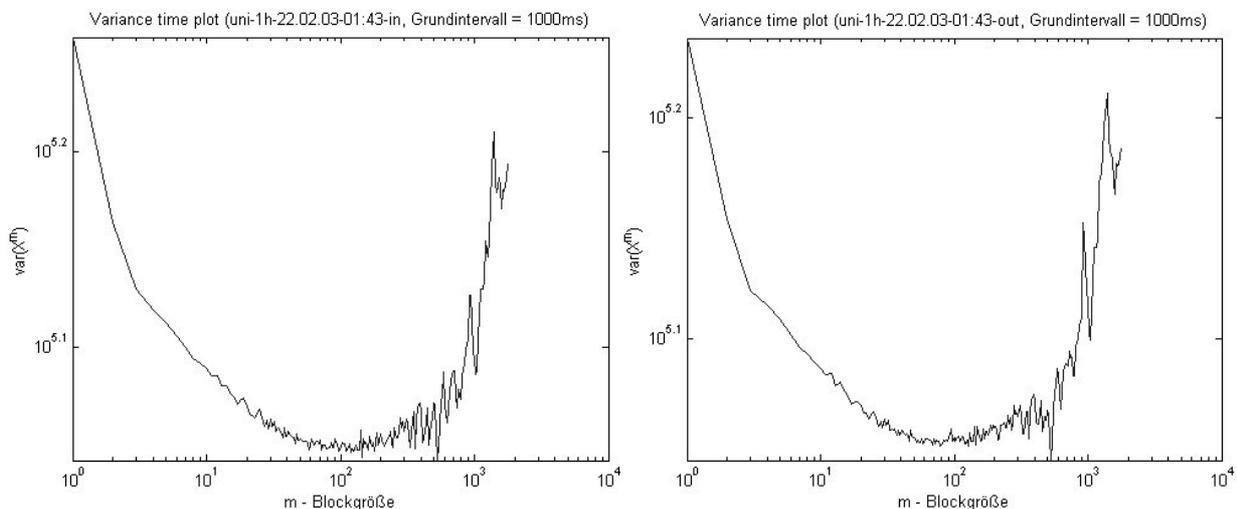
Die mittleren Ankunftsraten sind die Mittelwerte der berechneten Bins und ergeben sich zu :

$$\mu \approx 6162,0 \frac{\text{Pakete}}{\text{s}} \quad \text{eingehender Verkehr}$$

$$\mu \approx 5133,9 \frac{\text{Pakete}}{\text{s}} \quad \text{ausgehender Verkehr}$$

Die Schätzung der entsprechenden Hurst-Parameter erfolgte mittels R/S-Statistik und Variance Time Plot. Der Hurst-parameter für den eingehenden Verkehr unter Vollast betrug bei einer Bingröße von 1000ms $H = 0,9009$ (vgl. Abschnitt 4.10.1). Es ist zu erwarten, daß

der Hurst-Parameter hier einen geringeren Wert annimmt, da es sich hier um eine Niedriglastsituation handelt. Weiterhin war in Abschnitt 4.7.2 bereits zu erkennen, daß der Verlauf des Variance-Time-Plot dieser Datensätze mit am stärksten vom idealen Verlauf abweicht. Es ist zu prüfen, ob dieser Verlauf durch Veränderung der Bingröße positiv beeinflusst werden kann. In diesem Zusammenhang sollen wiederum die VTP für den eingehenden und den ausgehenden Verkehr in Abbildung 4.23 dargestellt werden.



Darstellung 4.23: Variance Time Plots, links eingehender, rechts ausgehender Verkehr

In allen vorherigen Betrachtungen führte eine Vergrößerung des Grundintervalls zu einem schwächeren „Knick“ im Verlauf des Variance Time Plot. In diesem Fall jedoch wurde die Krümmung eher verstärkt, was den Schluß zuläßt, daß der VTP nicht unmittelbar abhängig ist von der Bingröße und somit von der Anzahl der Bins. Verwertbare Ergebnisse sind aus den resultierenden Variance-Time-Plots nicht zu ermitteln.

Bei der Ermittlung des Hurst-Parameters mittels R/S-Statistik traten keine weiteren Auffälligkeiten auf, deshalb sei auf die jeweiligen Darstellungen verzichtet. In Tabelle 4.13 sind die geschätzten Hurstparameter und die mittleren Ankunftsfraten zusammengefaßt.

Datensatz	μ [Pakete/s]	R/S (Hurst)
uni 1h 22.02.03 01:43 in	6162,0	0,8207
uni 1h 22.02.03 01:43 out	5133,9	0,8187

Tabelle 4.13: Zusammenfassung der ermittelten Parameter

Da die geschätzten Hurst-Parameter mit den erwarteten Werten gut übereinstimmen, sollen diese gleichzeitig als entgültige Hurst-Parameter für eine spätere Simulation gelten.

Auf eine Aussage über den Wert des Parameters T (Fractal Onset Time Scale) sei hier verzichtet, da diese nicht eindeutig möglich ist. In bezug auf den vorherigen Abschnitt ist es jedoch naheliegend, daß der Parameter T einen Wert zwischen 0,001 und 0,1 annimmt. Diese Vermutung bestätigte sich später in den entsprechenden Simulationen (vgl. Abschn. 5.5).

4.11 Analyse der Proben mittels des Periodogramms

Aufgrund des großen Umfangs der Proben war es bei dieser Methode ebenfalls nötig, die Delta-Timestamps (Differenzen der aufeinanderfolgenden Timestamps) in Binvektoren zusammenzufassen. Für die Größen der Blöcke wurde empfangsseitig zunächst der Maximalwert der Delta-Timestamps - sendeseitig das Dreifache - gewählt, sodann die Berechnung mit der Aggregationsgröße von 1s wiederholt. Je Probe entstanden so vier Ergebnisse.

- Probe 1, 17.02.2003 17:12-18:12

Probe 1	Empfangen	Gesendet
Minimalwert	0,48µs	0,48µs
Bingröße l	12481µs	16760µs
Mittelwert in Ticks (2^{-32} s)	130µs $5,6 \cdot 10^5$	156µs $6,7 \cdot 10^5$
Varianz mittl. Abweichung	$3,25 \cdot 10^{11}$ $3,91 \cdot 10^5$	$5,19 \cdot 10^{11}$ $5,26 \cdot 10^5$
Anzahl der Elemente	27.729.715	23.079.982

Tabelle 4.14: Die mittlere Abweichung wurde bestimmt, indem der Mittelwert eines neuen Vektors gebildet wurde, der den Betrag der Differenzen von Probenvektor und Mittelwert des Probenvektors enthält. Das im Vergleich zum Erwartungswert sehr große Resultat ist ein weiterer Indikator für Burstiness.

Diese Probe wurde zu einem Zeitpunkt hoher Netzauslastung genommen. Es ist also anzunehmen, daß, wenn Selbstähnlichkeit vorherrscht, der zu bestimmende Hurst-Parameter einen relativ hohen Wert annimmt.

Unter Zuhilfenahme der von Matlab bereitgestellten o.g. Funktionen (vgl. Abschn. 3.3) ergaben sich folgende Plots:

a) Empfangener Traffic bei Blockgröße 12481µs, komplettes Intervall

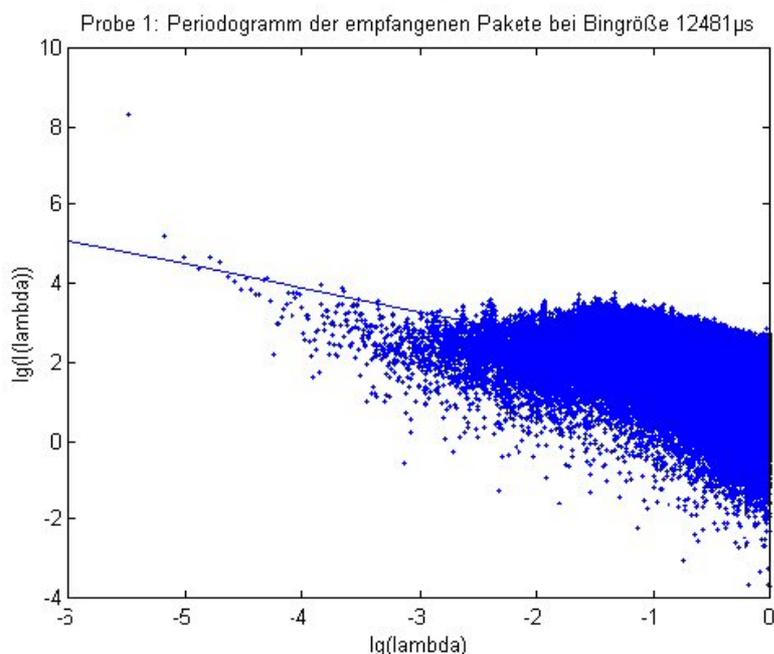


Abbildung 4.24: Die Regressionsgerade besitzt einen Anstieg von $m = -0,603$. Einsetzen in die Bestimmungsgleichung, $H = -\frac{m-1}{2}$, liefert $H \approx 0,8$. Dieses Schätzverfahren deutet also auf LRD hin.

b) Empfangener Traffic bei Blockgröße 1s

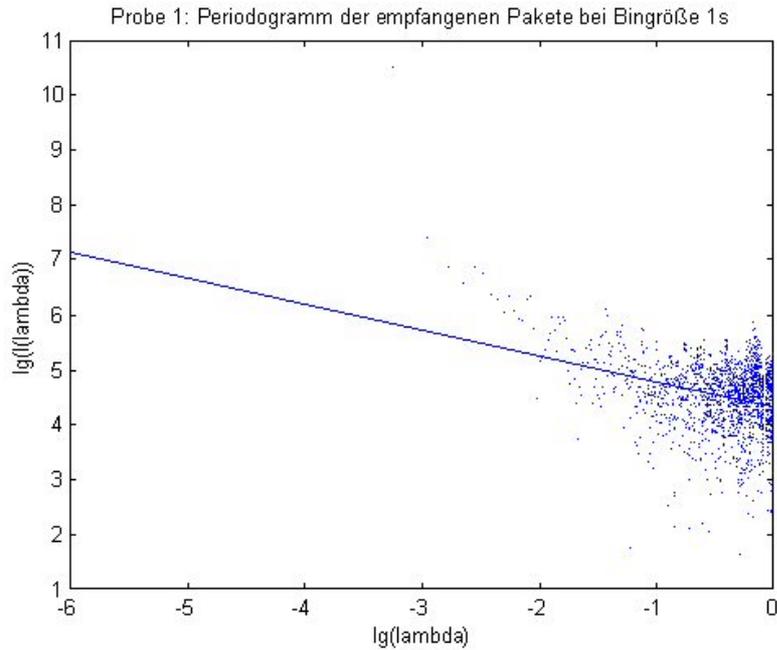


Abbildung 4.25: Die Steigung der Regressionsgeraden beträgt $m=-0,465$ und unterstützt damit die Aussage von a). Der entsprechende Hurst-Parameter berechnet sich zu $H \approx 0,733$.

c) Gesendeter Traffic bei Blockgröße 16760 μ s, gesamte Probe

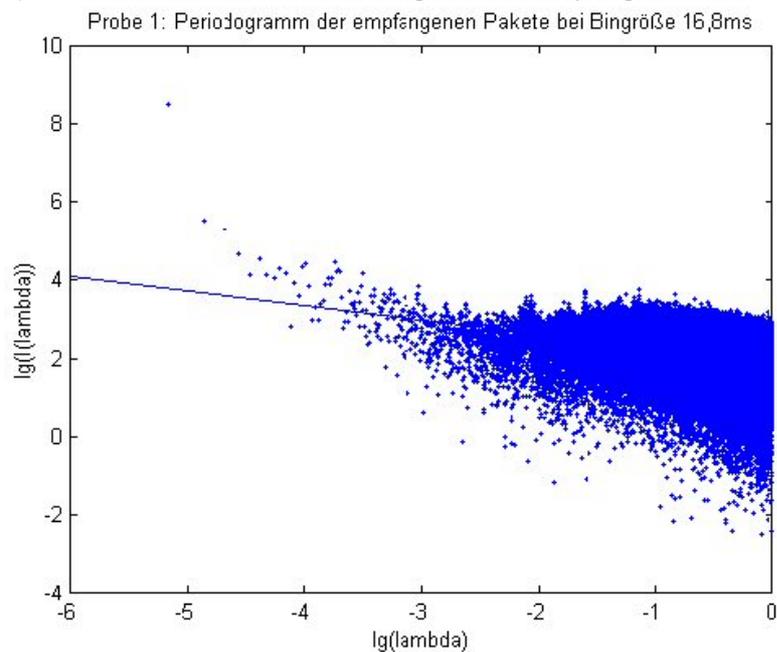


Abbildung 4.26: Die Anwendung der linearen Regression liefert einen Anstieg von $-0,372$. Der resultierende Hurst-Parameter läßt sich mit $0,686$ angeben.

d) Empfangener Traffic bei Blockgröße 1s

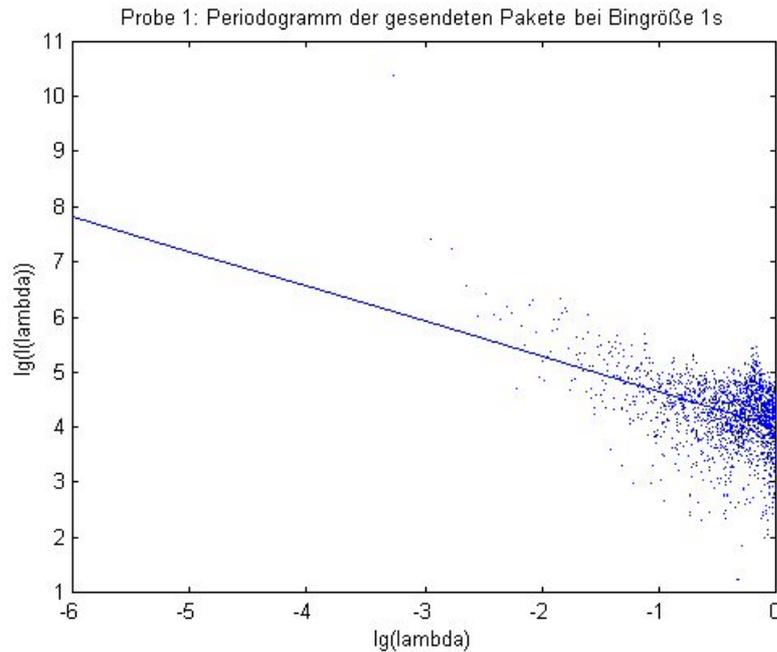


Abbildung 4.27: Der Anstieg ergibt mit $m=-0,642$ einen Hurst-Parameter von $0,821$ und liegt somit weitab der Aussage von c). Ein Grund für diese starke Abweichung ist in der geringen Anzahl der Meßwerte zu suchen.

Im Fall der gesendeten Pakete sind auf der Empfängerseite ungefähr übereinstimmende Ergebnisse erzielt worden, auch ist die Spannweite der Schätzungen für H mit $0,067$ in einem akzeptablen Rahmen. Insgesamt ist H nach Anwendung dieses Verfahrens im Bereich $0,7...0,8$ zu vermuten. Sendeseitig beträgt die Abweichung $0,135$, was für eine genauere Aussage bezüglich des Zahlenwerts des Hurst-Parameters zu ungenau ist, jedoch eindeutig auf Langzeitabhängigkeit hinweist. Die zusammengefaßten Schätzwerte für H sind in Tab. 4.15 zusammengefaßt.

Richtung	empfangen		gesendet	
	a	b	c	d
Plot				
Binggröße	12481 μ s	1s	16760 μ s	1s
Anstieg	-0,603	-0,465	-0,372	-0,642
res. H	0,801	0,733	0,686	0,821
\bar{H}	0,767		0,753	

Tabelle 4.15

- Probe 2: 21.02.2003, 11:52 – 12:52

Probe 2	Empfangen	Gesendet
Minimalwert	0,48 μ s	0,48 μ s
Bingröße 1	9338 μ s	11732 μ s
Mittelwert in Ticks (2^{-32} s)	136 μ s	161 μ s
	$5,86 \cdot 10^5$	$6,9 \cdot 10^5$
Varianz	$5,26 \cdot 10^{11}$	$6 \cdot 10^{11}$
mittl. Abweichung	$4,88 \cdot 10^5$	$5,49 \cdot 10^5$
Anzahl der Elemente	26.361.175	22.369.974

Tabelle 4.16

Diese während der Mittagspause genommene Probe läßt aufgrund der großen Anzahl von Timestamps ebenfalls einen hohen Hurst-Parameter vermuten.

Zur Wahrung einer besseren Lesbarkeit fiel die Entscheidung, nur die Tabelle mit den erhaltenen Hurst-Parametern anzuführen. Auf eine graphische Darstellung wird verzichtet.

Richtung	empfangen		gesendet	
Bingröße	9338 μ s	1s	11732 μ s	1s
Anstieg	-0,401	-0,490	-0,278	-0,423
Res. H	0,700	0,745	0,639	0,712
\bar{H}	0,723		0,675	

Tabelle 4.17

Auch hier zeigt sich empfangsseitig eine größere Übereinstimmung bezüglich des Hurst-Parameters als sendeseitig.

- Probe 3: 22.02.2003, 01:43 – 02:43

Diese Meßreihe unterscheidet sich von den vorangegangenen insofern, als daß sie aus einer Niedriglastsituation stammt. Laut Definition müßte der gewonnene Hurst-Parameter demnach niedriger als in den Proben 1 und 2 ausfallen.

Probe 3	Empfangen	Gesendet
Minimalwert	0,48µs	0,48µs
Maximalwert (Bingröße 1)	8497µs	13461µs
Mittelwert in Ticks (2^{-32} s)	162µs $6,97 \cdot 10^5$	195µs $8,37 \cdot 10^5$
Varianz mittl. Abweichung	$7,59 \cdot 10^{11}$ $6,07 \cdot 10^5$	$9,00 \cdot 10^{11}$ $6,75 \cdot 10^5$
Anzahl der Elemente	22.183.295	18.482.110

Tabelle 4.18

Richtung	Empfangen		gesendet	
Bingröße	8497µs	1s	13461µs	1s
Anstieg	-0,320	-0,316	-0,168	-0,354
res. H	0,660	0,658	0,584	0,677
\bar{H}	0,659		0,62	

Tabelle 4.19

Tatsächlich sind an dieser Stelle etwas niedrigere Schätzungen für H sowohl sende- als auch empfangsseitig vorhanden. Bezüglich der Konsistenz zeigt sich auch hier die größere Abweichung für den Schätzwert auf seiten des Senders.

Das Periodogramm ergibt für alle untersuchten Fälle einen deutlichen Hinweis auf Langzeitabhängigkeit und bestätigt damit die Ergebnisse aus Variance-Time-Plot und RS-Plot. Die in Abhängigkeit von der Blockgröße festgestellten Schwankungen des Hurst-Parameters sind sendeseitig besonders stark ausgeprägt.

4.12 Anwendung des Whittle-Estimators

Der Whittle-Estimator wurde auf jeden Datensatz zweimal angewandt, bei einer Aggregationsschwelle des Maximalwerts des jeweiligen Datensatzes sowie bei dessen Dreifachem. Ein Berechnungsversuch bei einer Aggregationsgröße von 1s scheiterte an der dann zu geringen Anzahl von Meßwerten. Die Ergebnisse sind ebenfalls tabellarisch zusammengefaßt. Für das Skript zur Berechnung sei auf die beigegefügteten Datenträger verwiesen.

Probe	1	1	2	2	3	3
Bingröße [μ s]	12481	37442	9338	28016	8497	25492
res. H	0,690	0,735	0,690	0,695	0,661	0,688
\bar{H}	0,713		0,693		0,675	
H _{Periodogramm} *	0,767		0,723		0,659	

Tabelle 4.20a: empfangsseitige Schätzungen

Probe	1	1	2	2	3	3
Bingröße [μ s]	4190	12571	3911	11732	4487	13461
res. H	0,680	0,714	0,690	0,667	0,641	0,662
\bar{H}	0,692		0,678		0,652	
H _{Periodogramm} *	0,753		0,675		0,620	

Tabelle 4.20b: sendeseitige Schätzungen

* - Zu beachten sind die unterschiedlichen Bingrößen.

Die Schätzungen nach Whittle liegen durchweg in der Nähe der Periodogramm-Schätzungen. Angesichts der hohen Rechenzeit im Vergleich zu den graphischen Verfahren erscheint der Einsatz an dieser Stelle fragwürdig, zumal der Whittle-Estimator nicht geeignet ist, Selbstähnlichkeit nachzuweisen. Es zeigt sich auch hier eine Abhängigkeit von der gewählten Blockgröße.

5. Simulation in OPNET

5.1 Fraktale Punktprozesse (FPP - Fractal Point Process)

Dieser Abschnitt befaßt sich mit einigen fraktalen Verkehrsmodellen, die auf Fraktalen Punktprozessen (FPP) basieren. In [RL98] und [RL00] werden zwei Methoden vorgestellt, wie Fraktale Punkt Prozesse konstruiert werden können. Auf der einen Seite sind dies die die *Renewal Point Process* - Methode, mit der fraktale Erneuerungsprozesse erzeugt werden können; auf der anderen Seite eignet sich die *Doubly Stochastic Poisson Process* (DSPP) - Methode, um *Fraktal Modulierte Poissonprozesse* (FMPP) zu erzeugen.

Abbildung 5.1 aus [RL00] stellt eine übersichtliche Einteilung der FPPs dar. Hier ist die Vielzahl der bisher entwickelten FPPs zu erkennen.

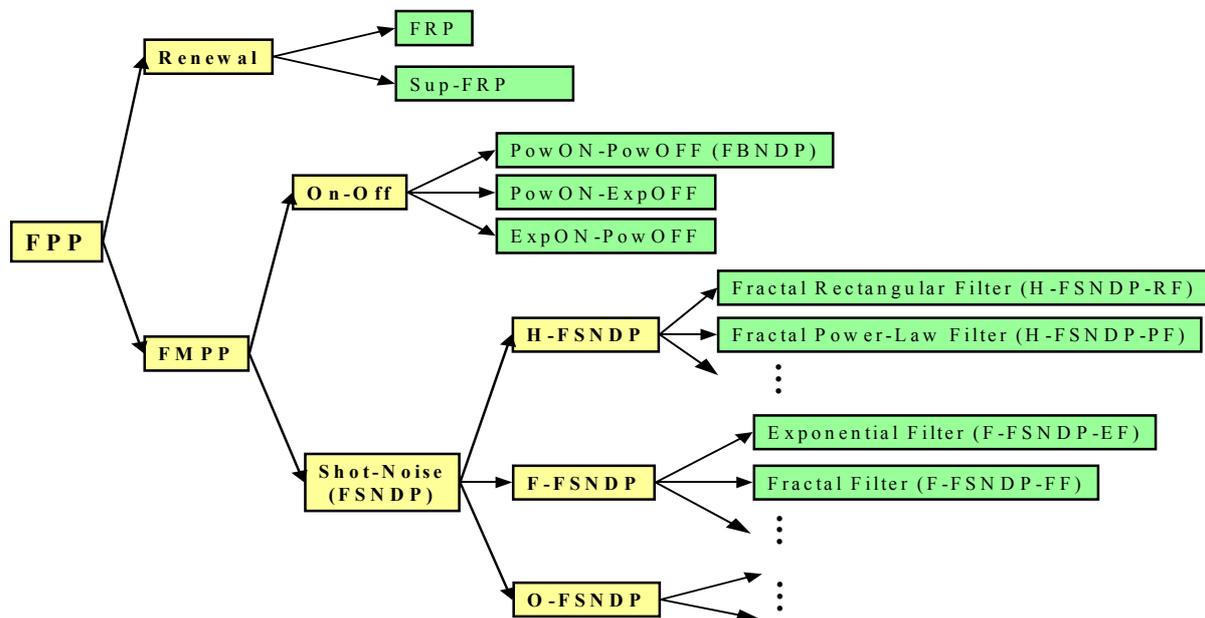


Abbildung 5.1: Einteilung der FPPs

Diese Übersicht erhebt keinen Anspruch auf Vollständigkeit, da sich weitere FPPs in der Entwicklung befinden.

Der in OPNET zur Verfügung stehende Raw Packet Generator stellt die Implementierung einzelner FPPs aus [RL98] und [RL00] dar. Zur Verfügung stehen:

- Sup-FRP
- POWON-POW OFF, EXPON-POW OFF, POWON-EXP OFF
- H-FSDNP
- F-FSDNP-EF und F-FSDNP-FF

In den folgenden Abschnitten sind nur der Sup-FRP und die ON-OFF-FMPPs vorgestellt, da FSNDPs im Zusammenhang mit dieser Arbeit keine Anwendung finden. Für weiterführende Informationen zu den einzelnen FPPs sei hier wiederum auf [RL98] und [RL00] verwiesen.

5.1.1 FRP (Fractal Renewal Process)

Der Standard-Erneuerungspunktprozeß ist der einfachste fraktale Punktprozeß und Grundlage für die Sup-FRP.

Es wird davon ausgegangen, daß die Zwischenankunftszeiten voneinander unabhängig und unabhängig verteilt sind. Die Wahrscheinlichkeitsdichteverteilung der Zwischenankunftszeiten beschreibt den Prozeß vollständig. Ist die Wahrscheinlichkeitsdichteverteilung heavy-tailed, führt dies zu einem fraktalen Erneuerungspunktprozeß (FRP). Da alle Ereignisse von derselben Wahrscheinlichkeitsdichteverteilung abhängen, ist die Anwendbarkeit auf Netzwerkverkehrsmodelle sehr begrenzt, wie später auch gezeigt wird. Es besteht eine ausgeprägte Korrelation zwischen den Ankunftszeiten der Pakete.

In [RL98] wird die Wahrscheinlichkeitsdichteverteilung angegeben mit:

$$p(x) = \begin{cases} 0 & \text{für } x \leq A \\ \gamma A^\gamma x^{-(\gamma+1)} & \text{für } x > A \end{cases} \quad (5.1)$$

Die obige Gleichung kann umgeformt werden zu:

$$p(x) = \gamma A^\gamma x^{-(\gamma+1)} = \gamma \frac{A^{\gamma+1}}{A} x^{-(\gamma+1)} = \frac{\gamma}{A} \left(\frac{A}{x} \right)^{(\gamma+1)}$$

mit $\gamma = \alpha$ und $A = t_0$ ergibt sich die Wahrscheinlichkeitsdichteverteilung zu:

$$p(x) = \frac{\alpha}{t_0} * \left(\frac{t_0}{x} \right)^{\alpha+1}$$

In Verbindung mit der Dokumentation des Raw Packet Generators und Tabelle 2.1 in Abschnitt 2.4.4 ist zu erkennen, daß es sich hierbei um eine Pareto-Verteilung handelt. In [RL98] und [RL00] werden weitere Wahrscheinlichkeitsdichteverteilungen angegeben, bei denen es sich jedoch nur um andere Varianten der obigen handelt.

Der fraktale Exponent γ beeinflusst die Eigenschaften des FRP besonders stark.

Wird der Exponent $\gamma > 2$ gewählt, zeigt der resultierende Prozeß keine Eigenschaften von Selbstähnlichkeit 2. Ordnung. Obwohl für $0 < \gamma < 1$ der Prozeß vollständig fraktal ist, führt das zu Simulationsergebnissen, die sich extrem „bursty“ verhalten. Dies deckt sich nicht mehr mit realen Verkehrsdaten. In der Praxis hat sich gezeigt, daß zur Modellierung von Netzwerkverkehr der Bereich $1 < \gamma < 2$ von Bedeutung ist.

5.1.2 Sup-FRP

Sup-FRP ist eine Verknüpfung von mehreren zwar identischen aber unabhängigen FRPs. Der daraus resultierende Prozeß ist nicht mehr zu den Erneuerungsprozessen zu zählen, soll aber dennoch in diesem Abschnitt Erwähnung finden, da er in OPNET implementiert ist. Die Randverteilungen der Zwischenankunftszeiten weisen trotz der Verknüpfung ein langzeitabhängiges Verhalten auf. Das bereichsübergreifende (scaling) Verhalten der einzelnen FRPs ist weiterhin gegeben, jedoch mit einer leichten Reduzierung der Zeitbereiche.

In OPNET ist dieser Prozeß bereits im Raw Packet Generator implementiert. Angegeben werden als Argumente nur noch der Hurst-Parameter, die mittlere Ankunftsrate und die

Fractal Onset Time Scale. Sup-FRP wird auch als Eingangsprozeß für die F-FSNDPs verwendet, die im RPG verfügbar sind. Erfolgreich eingesetzt wurde dieses Modell als flexibler, faktaler Hochgeschwindigkeits-ATM-Verkehrsgenerator an der Columbia University [R96].

5.1.3 Fraktal Modulierter Poisson Prozeß (FMPP)

Die DSPP-Methode führt auf eine Vielzahl fraktal modulierter Poisson-Prozesse (FMPP), vergleichbar mit markovmodulierten Poisson-Prozessen. Um die gewünschten selbstähnlichen Eigenschaften zu erzielen, wird ein stationärer stochastischer Prozeß $I(t)$ mit einer Autokorrelationsfunktion $R_I(\tau)$ benötigt. Diese AKF muß eine polynomiale Form besitzen nach:

$$R_I(\tau) \sim \tau^{-\alpha} \text{ und } 0 < \alpha < 1 \quad (5.2)$$

Nach der Art, wie der Prozeß $I(t)$ gebildet wird, läßt sich eine weitere Unterteilung treffen:

- FMPPs vom ON/OFF-Typ
- FMPPs vom Shot-Noise-Typ (FSNDPs)

Der erste Typ wird im nächsten Abschnitt vorgestellt.

5.1.4 On-Off FMPPs

In diesem Abschnitt werden ON-OFF FMPPs näher betrachtet. Hierzu ist zunächst noch einmal auf den Standard-FRP einzugehen, bei dem eine verteilte Menge von Punkten (bzw. Markierungen) auf einer Zeitebene erzeugt wird, die polynomial verteilt ist. Die Punkte seien nun als Ereignisse zu sehen, die zu einer Zustandsänderung führen. Weiterhin sollen nun „reale“ Werte (z.B. 0 und $R(>0)$) verwendet werden. Da mit jedem Eintreten eines Ereignisses am FRP zwischen 0 und R gewechselt wird, entsteht ein alternierender fraktaler Erneuerungsprozeß (AFRP – Alternating Fractal Renewal Process). Verdeutlicht wird dies in Abbildung 5.2 aus [RL98].

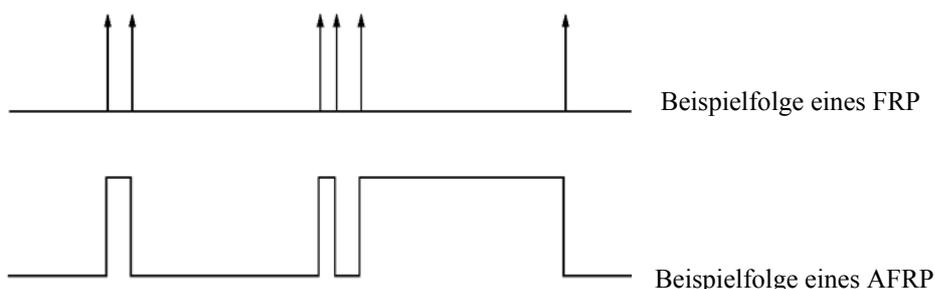


Abbildung 5.2: Zusammenhang zwischen FRP und AFRP

Die Verteilung der ON/OFF-Perioden weist somit das gleiche heavy-tailed Verhalten auf wie die beim ursprünglichen FRP. Der Prozeß $I(t)$ wird nun gebildet durch die Überlagerung von mehreren unabhängigen AFRPs und wird in einem folgenden Poisson-Prozeß als stationärer, stochastischer Eingangsprozeß (oder Bedienprozeß) verwendet. Es entsteht somit ein FMPP. Die Wahrscheinlichkeitsdichteverteilungen der Verweildauer in den ON- und OFF-Zuständen besitzen hier eine polynomiale Form. Somit handelt es sich um einen PowON/PowOFF – FMPP, auch bekannt als FBNDP (Fractal Binomial Noise Driven Poisson Process). Weitere Formen ergeben sich, wenn die Wahrscheinlichkeitsdichteverteilungen der Verweildauer in den ON- und OFF- Zuständen unterschiedlich sind.

So kann die Verweildauer der ON- Perioden exponentiell verteilt und die der OFF-Perioden polynomial verteilt sein (Exp-ON / Pow-OFF). Auf der anderen Seite ist es auch möglich, daß die Verweildauer der ON-Perioden polynomial verteilt und die der OFF-Perioden exponentiell verteilt ist (Pow-ON/Exp-OFF).

Es ergeben sich folgende drei Formen von ON/OFF-FMPPs, die auch im RPG in OPNET verfügbar sind:

- PowON-PowOFF
- ExpON-PowOFF
- PowON-ExpOFF

In [RL98] wurde dieses Modell verwendet, um den Verkehr in einem Ethernetnetzwerk zu simulieren.

5.2 Auswahl eines geeigneten Modells

Da die analysierten Datensätze ausschließlich Zwischenankunftszeiten beinhalten, ist die Anzahl der anwendbaren FRPs begrenzt. Betrachtet wurden nur einzelne Pakete, die keine Aussage über die einzelnen Verbindungen bzw. Datenströme (flows) beinhalten. Da FSNDPs solche Datenströme beschreiben, finden sie hier keine Verwendung. Somit ist die Auswahl beschränkt auf Sup-FRP bzw. auf einen der ON-OFF Prozesse. Hierbei sollen die ON-OFF Prozesse nur eine untergeordnete Rolle spielen, da keine Aussagen über die einzelnen ON- bzw. OFF-Perioden getroffen werden können. Es besteht also keine Möglichkeit, aus den Zwischenankunftszeiten den passenden ON-OFF Prozeß gezielt auszuwählen.

Auch die benötigten Parameter sind für die Auswahl eines geeigneten Modells von Bedeutung. Es besteht immer die Frage, ob für einen in Betracht gezogenen Prozeß auch die benötigten Parameter angegeben werden können. Aus diesem Grund sind in Tabelle 5.1 alle Parameter für den zugehörigen Prozeß zusammengefaßt.

	Sup-FRP	Pow ON Pow OFF	Exp ON Pow OFF	Pow ON Exp OFF	H-FSNDP	F-FSNDP EF	F-FSNDP FF
H	x	x	x	x	x	x	x
μ	x	x	x	x	x	x	x
FOTS	x	x	x	x	x	x	x
max/mean			x	x			
SAR		x	x	x			

Tabelle 5.1: Benötigte Parameter der verschiedenen Prozesse

Eine ähnliche Zusammenstellung ist in [R02] zu finden. In der obigen Tabelle ist zu erkennen, daß für jeden im RPG zur Verfügung stehenden Prozeß die drei Hauptparameter:

- H (Hurst-Parameter)
- μ (mittlere Ankunftsrate)
- FOTS (Fractal Onset Time Scale, auch T)

anzugeben sind.

Die Abkürzung max/mean beschreibt das Verhältnis zwischen dem Maximalwert und dem arithmetischen Mittel der zu betrachtenden Wertefolge. SAR steht hier für „Source Activity Ratio“ und gibt den Prozentsatz der Zeit an, in der mindestens ein Sender aktiv ist.

Es ist von Vorteil für den Raw Packet Generator, daß alle zur Verfügung stehenden Prozesse mit sehr wenigen Parametern auskommen.

5.3 Aufbau eines geeigneten Netzwerks

Es wird nun die allgemeine Simulationsumgebung näher betrachtet. Das Modell so einfach wie möglich zu halten ist hierbei von äußerster Wichtigkeit, um die Ergebnisse im Überblick behalten zu können. Weiterhin kann ein zu komplexes Modell auch eine Verminderung der Leistungsfähigkeit darstellen, was sich besonders auf die Simulationszeit auswirkt.

Der Raw Packet Generator ermöglicht eine Modellierung auf den Schichten 2 und 3 im ISO/OSI Referenzmodell. Auf Schicht 3 wird nur IP als Protokoll unterstützt. Möglichkeiten zur Simulationen mit Protokollen wie AppleTalk oder IPX sind in Version 9.1 nicht berücksichtigt.

Vorausgesetzt sei, daß es sich bei den analysierten Datensätzen in Kapitel 4 um reine IP-Pakete handelt, und so sollen im zu verwendenden Modell nur reine IP-Pakete verwendet werden, ohne nähere Angabe von Protokollen in höheren Schichten. Dieses wäre ohnehin nicht möglich, da die analysierten Datensätze nur Zwischenankunftszeiten enthalten.

Da der eingehende und ausgehende Verkehr zunächst getrennt betrachtet werden sollen, ist ein Modell zu entwickeln, welches dem Quelle – Senke – Prinzip entspricht. In Abbildung 5.3 ist dieses schematisch dargestellt. Gleichzeitig sind die verwendeten IP-Adressen angegeben.



Darstellung 5.3: Schematischer Aufbau

Somit erfüllt nur der Sender die eigentliche Aufgabe der Generierung selbstständlichen Verkehrs. Der Empfänger hat nur die Aufgabe, die Pakete zu empfangen. Es gilt einen Weg zu finden, den Empfänger vom Senden von Daten abzuhalten.

Verwendet wurden nur Elemente aus der RPG-Objektpalette. Als Sender und Empfänger dienen jeweils eine „ethernet_rpg_wkstn“, die mit einem 100BaseT-Link verbunden sind. Neben den IP-Adressen werden noch weitere Einstellungen benötigt, welche als „Interface Information“ unter „IP-Host-Parameter“ einzugeben sind. In Tabelle 5.2 sind die wichtigsten kurz zusammengefaßt.

	Sender	Empfänger
IP-Adresse	172.16.20.1	172.16.20.2
Maske	Class B (natural)	
MTU(bytes)	ethernet	
Passive RIP Routing	disabled	
Default Route	172.16.20.2	172.16.20.1

Tabelle 5.2: Zusammenfassung der Netzwerkeinstellungen

Um die Prämisse der geringen Komplexität aufrecht zu erhalten, ist jede Art von Routingprotokoll zu deaktivieren. So kann auch vermieden werden, daß zusätzliche Pakete gesendet werden, die das Simulationsergebnis verfälschen würden. Eine statische Routingtabelle ist somit für dieses Netzwerk vollkommen ausreichend; die nötigen Einstellung sind in Tabelle 5.3 zusammengefaßt. Diese können wiederum als einzelner Eintrag in „Static Routing Table“ unter „IP-Host-Parameter“ eingegeben werden.

	Sender	Empfänger
Destination	172.16.20.2	172.16.20.1
Subnet Mask	Class Based	
Next Hop	172.16.20.2	172.16.20.1

Tabelle 5.3: Festlegen der statischen Routingtabelle

Um der Realität näherzukommen, ist zu prüfen, ob der eingehende und der ausgehende Verkehr nicht auch in der Simulation zusammengefaßt werden können. Interessant ist dabei, welche Auswirkungen auf die Leistungsfähigkeit zu erkennen sind. Hierzu wäre der Empfänger lediglich auch zum Sender umzufunktionieren, was einem Vollduplexbetrieb entspräche.

Weitere Einstellungen werden in den entsprechenden Abschnitten gesondert erläutert. Dies gilt vor allem für den betreffenden Ankunftsprozeß in der Simulation.

5.4 Simulation des eingehenden Verkehrs unter Vollast

In den folgenden Abschnitten wurden die ermittelten Werte aus Abschnitt 4.10.1 in mehreren Simulationen verwendet. Die für den Sup-FRP benötigten Parameter beschränken sich auf die drei Hauptparameter (H, μ und T). Obwohl in OPNET die mittlere Ankunftsrate nur in Paketen/Sekunde (bei FSNDP Flows/s) angegeben werden kann, wurden zwei verschiedene Bingrößen verwendet. Aus der Angabe der mittleren Ankunftsrate wird ersichtlich, daß streng genommen nur eine Bingröße von 1000ms in Frage kommt. Dieses führt bei einer Aufzeichnungsdauer von einer Stunde zu 3600 Bins. Bei der Analyse der verschiedenen Datensätze wurden jedoch weitaus kleinere Bingrößen verwendet, was dann in weitaus mehr Bins resultierte. Es ist fragwürdig, wie aussagekräftig 3600 Bins sein können, wenn der Ausgangsdatensatz mehr als 20 Millionen Zwischenankunftszeiten beinhaltet. Um die durch OPNET gesetzten Grenzen zu erweitern, wurde hierzu die simulierte Zeit auf 10 Stunden erhöht. So wurde es möglich, Simulationen mit einer „virtuellen“ Bingröße von 100ms durchzuführen. Dieses bedeutet eine Vergrößerung des Bereichs auf 36000 Bins. Es sei darauf hingewiesen, daß dieses nicht unbedingt als „elegante“ Lösung für dieses Problem anzusehen ist.

Die Eingabebeschränkung für die mittlere Ankunftsrate läßt den Schluß zu, daß der Raw Packet Generator für größere Zeitbereiche (Stunden, Tage, Wochen) in der simulierten Zeit konzipiert wurde.

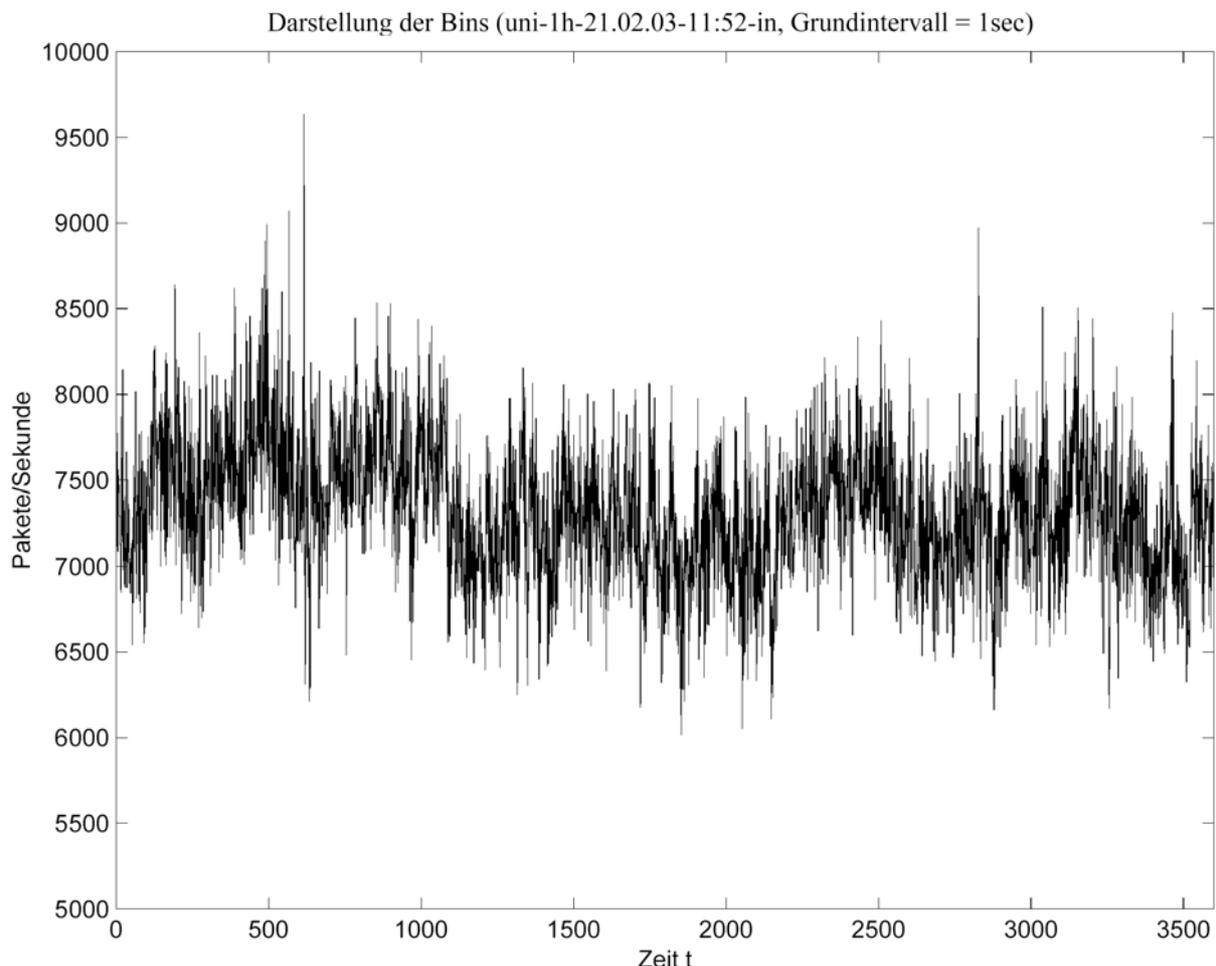
5.4.1 Simulation mit einer Bingröße von 1000msec

Zunächst wurden Simulationen mit einer Bingröße von 1000 ms durchgeführt. In Abschnitt 4.10.1 wurden bereits einige Werte für die Simulation aus dem zu betrachtenden Datensatz uni_1h_21.02.03_11:52_in ermittelt. Die Ergebnisse sind noch einmal in Tabelle 5.4 zusammenfassend dargestellt.

Bingröße [msec]	μ [Pakete/s]	Hurst		\bar{H}
		R/S	VTP	
1000	7322,55	2 bis 200	0,8841	0,9009

Tabelle 5.4: Zusammenfassung der ermittelten Eingangsparameter

Mit diesen Daten aus der Analyse konnte nun der Ankunftsprozeß in OPNET konfiguriert werden. Neben den Daten aus der Analyse war weiterhin die Fractal Onset Time Scale anzugeben. Aus der Analyse der Daten war für diesen Parameter in der Analyse kein eindeutiger Wert zu schätzen. Somit war zu prüfen, welche Auswirkung der Parameter auf das Simulationsergebnis hat, indem mit mehreren möglichen Werten simuliert wurde. Es war davon auszugehen, daß der Verlauf der Wahrscheinlichkeits- bzw. Wahrscheinlichkeitsdichteverteilungen unmittelbar vom Wert des Parameters T abhängig sind. Veränderungen in der Darstellung der Bins waren ebenfalls als wahrscheinlich anzusehen. Um die Darstellung der Bins auch vergleichen zu können, sind in Abbildung 5.4 die erstellten Bins aus dem Datensatz über der Zeit grafisch dargestellt.

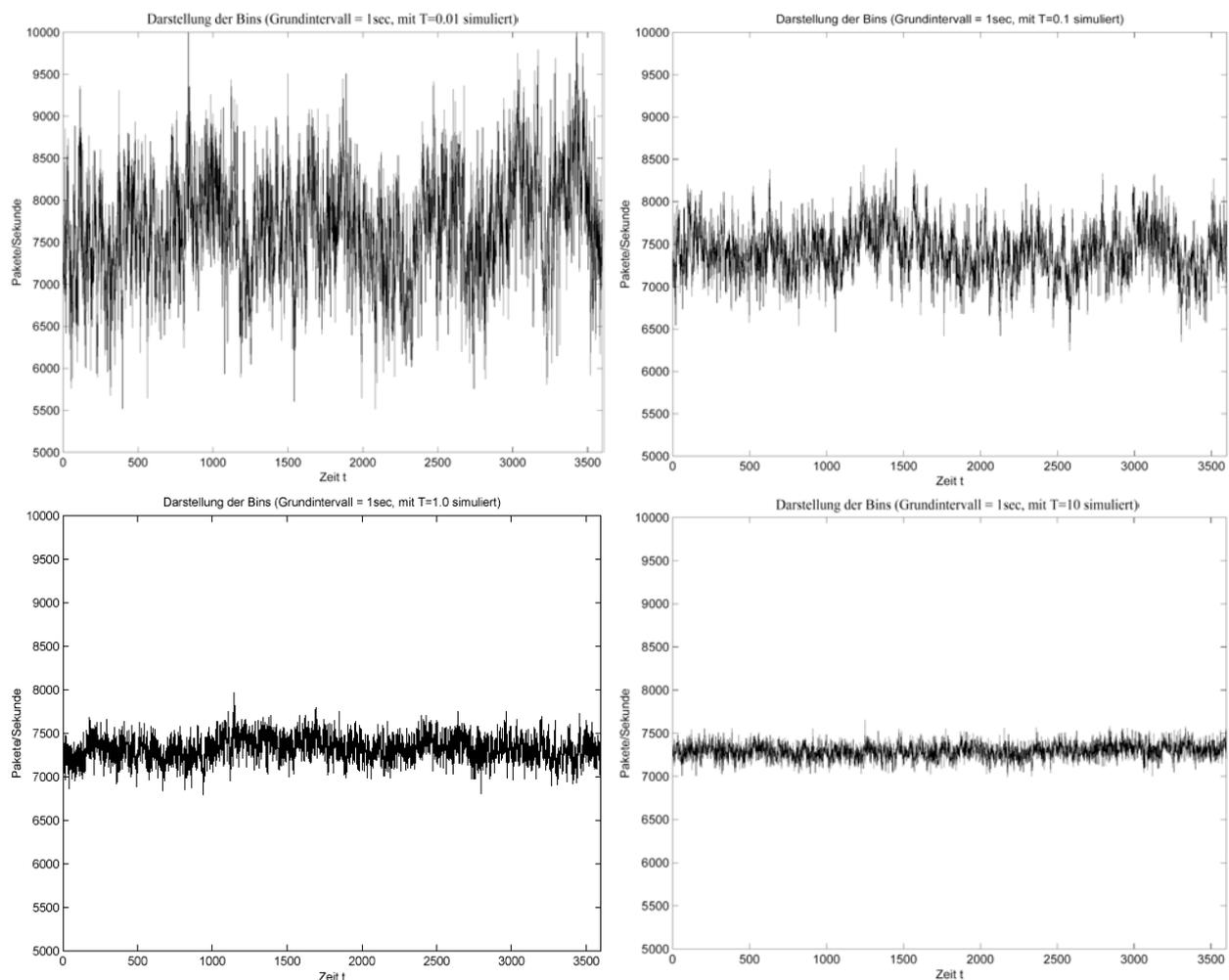


Darstellung 5.4: Darstellung der Bins aus der zu simulierenden Folge

Wie zu erwarten war, traten relativ starke Schwankungen um den Mittelwert bzw. um die mittlere Ankunftsrate auf. Mit Hilfe der Methode zur Schätzung der Ober- und Untergrenzen aus Abschnitt 4.2.1 konnte ein Bereich ermittelt werden, in dem die Mehrzahl aller Bins auftraten. Dieser Bereich lag zwischen 6793 und 7887 Paketen/s.

Mit den Erkenntnissen aus der Analyse und mit Hilfe der in Abschnitt 5.3 beschriebenen Simulationsumgebung, konnten nun Simulationen durchgeführt werden. Zu erwähnen sei hierbei, daß der Empfänger direkt ausgeschaltet wurde (in OPNET mit „Fail This Node“), um ihn selbst am Senden von Paketen sicher zu hindern. In späteren Betrachtungen wird sich diese Vorgehensweise als nicht ideal herausstellen. Im Ankunftsprozeß des Senders war weiterhin die Startzeit anzugeben. Nach Ablauf dieser Zeit beginnt der Raw Packet Generator mit dem Erzeugen von Paketen. Für den sofortigen Beginn ist ein Wert von 0 Sekunden anzugeben.

Verwendet wurden weiterhin sechs verschiedene Werte für den Parameter T zwischen 0,001 und 10, um seine Auswirkungen vergleichen zu können. Jede der sechs durchgeführten Simulationen benötigte eine Simulationszeit von mehr als 20 Stunden. In Abbildung 5.5 sind einige der resultierenden Bindarstellungen dargestellt.



Darstellung 5.5: Darstellung der Bins bei verschiedenen Werten für den Parameter T

Das Übernehmen der Darstellungsbereiche auf der y-Achse aus dem Original ermöglicht einen besseren Vergleich mit der Darstellung 5.4. Es sei darauf hingewiesen, daß bei einer FOTS von $T=0,001$ und $T=0,01$ Bins berechnet wurden, die größer waren als der Darstellungsbereich.

Im folgenden konnten nun wiederum die Ober- und die Untergrenzen festgelegt werden. Da jedoch die Schätzung der Grenzen aus der Bindarstellung nicht frei vom subjektiven Einfluß

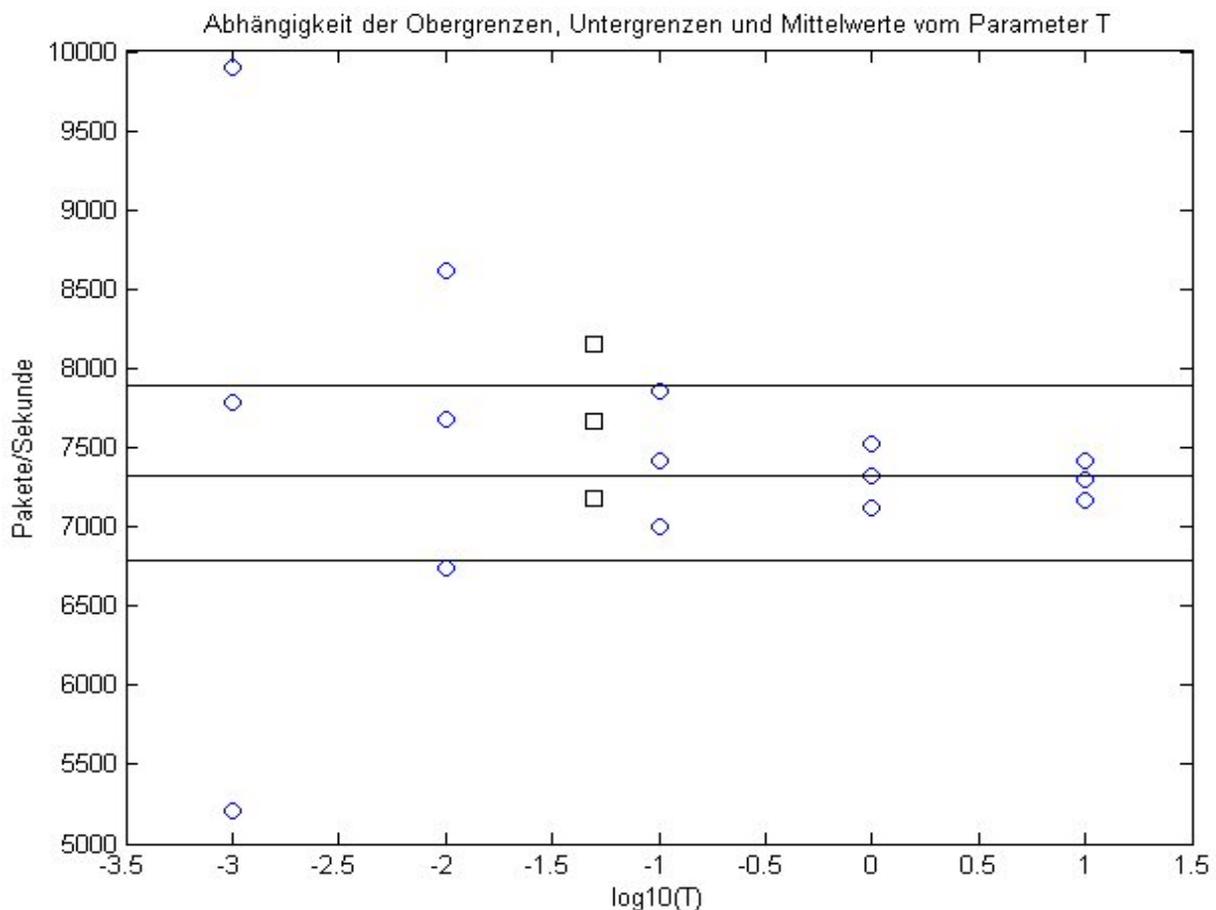
des Betrachters ist, sind die folgenden Bereichsschätzungen wiederum mit Hilfe der in Abschnitt 4.2.1 beschriebenen Methode durchzuführen. In Tabelle 5.5 sind die Ergebnisse aus der Bereichsschätzung und die resultierenden Hurst-Parameter zusammengefaßt.

Der Hurst-Parameter wurde mittels R/S-Statistik und VTP (im Bereich 2 bis 200) geschätzt.

	λ	Grenze		Bereich		Hurst			
		untere	obere		in %	R/S	VTP	Peri	\bar{H}
Original	7322,5	6793	7887	1094	0,9177	0,8841	0,9177	-	0,9009
T	0,001	7788,0	9894	4682	0,7812	0,8758	0,7812	-	0,8285
	0,01	7683,5	8614	1869	0,7808	0,8207	0,7808	0,89	0,8008
	0,1	7417,4	7855	849	0,7803	0,8033	0,7803	0,89	0,7918
	1	7322,0	7519	401	0,8438	0,8370	0,8438	-	0,8404
	10	7293,3	7421	256	0,7248	0,7466	0,7248	-	0,7357
	0,05	7664,1	8148	965	0,7994	0,8143	0,7994	0,86	0,8069

Tabelle 5.5: Binggröße = 1sec, 3600 Bins

Die Ergebnisse untermauern die Vermutung, daß sich die FOTS direkt auf die Ober- und Untergrenzen im Verlauf der Bindarstellung auswirken, was zu einem breiteren Verlauf in der Wahrscheinlichkeitsdichteverteilung führt. Zur Verdeutlichung sind in Abhängigkeit vom Parameter T die mittleren Ankunftsrate, die Ober- und die Untergrenzen aus Tabelle 5.5 graphisch dargestellt.



Darstellung 5.6: Darstellung der ermittelten Bereichsgrenzen bei einer Binggröße von 1000ms

— Werte des Datensatzes
 ○ Werte aus den Simulationen

Aus der obigen Darstellung und der vorangegangenen Tabelle wird ersichtlich, daß der Parameter T=0,05 das Verhalten am besten beschreibt. Die mittleren Ankunftsrate wurden nach

den Vorgaben in der Simulation im groben eingehalten. Hier kam es im Höchstfall zu einer Abweichung von 1,7 %.

Aus Tabelle 5.5 geht weiterhin hervor, daß der resultierende Hurst-Parameter relativ stark zwischen 0,7357 und 0,8404 schwankt. Auf die mittels Periodogramm geschätzten Werte sei hier verzichtet, da dieser Schätzer als zu „optimistisch“ gilt. Dabei wird jedoch bei der Schätzung mittels R/S-Statistik und Variance Time Plot nie der Hurst-Parameter von 0,9009 des berechneten Datensatzes erreicht. Dies deckt sich mit den Erfahrungen aus [XY01]. Hier wurden mehrere Simulationen mit lediglich veränderten Hurst-Parametern durchgeführt. Die besten Ergebnisse erzielten Simulationen mit einem Hurst-Parameter zwischen 0,7 und 0,8. Es bleibt zu prüfen, ob dies als Schwäche des Raw Packet Generators auszulegen ist.

5.4.2 Simulation mit 100msec

Wie bereits erwähnt, ist die Erhöhung Binanzahl durch Erhöhung der simulierten Zeit nicht als geeignete Methode anzusehen. Da jedoch die Ergebnisse aus einer Simulation mit mehr als 3600 simulierten Bins von Interesse sind, wurde dennoch eine Erhöhung der simulierten Zeit vorgenommen.

In diesem Zusammenhang sind nochmal die Ergebnisse aus der Analyse für eine Binggröße von 100ms aus Abschnitt 4.10.1 in Tabelle 5.6 zusammengefaßt.

Binggröße [msec]	μ [Pakete/100ms]	R/S (Hurst)	VTP		\bar{H}
			Bereich	Hurst	
100	732,26	0,7806	2 bis 20	0,6851	0,7806
			20 bis 2000	0,8824	
			2 bis 2000	0,8527	

Tabelle 5.6: Zusammenfassung der ermittelten Parameter

Wie bereits bei der betreffenden Analyse erwähnt, ist für den Hurst-Parameter nur der Wert aus der R/S-Statistik zu verwenden, da der Variance Time Plot kein brauchbares Ergebnis lieferte.

Eine Veränderung in der simulierten Zeit bedeutet eine Skalierung auf der Zeitebene.

Somit war zu erwarten, daß in der Simulation das Verhalten am besten beschrieben wird, wenn der Parameter T einen Wert von 0,01 bzw. zwischen 0,01 und 0,001 annimmt. Diese Erwartung begründet sich aus den Ergebnissen der vorherigen Simulationen (Binggröße 1000ms), bei denen als am besten passender Wert $T=0,05$ ermittelt wurde.

Auch hier wurde wiederum mit verschiedenen Parametern T und gleichbleibendem Hurst-Parameter und gleichbleibender mittlerer Ankunftsrate simuliert, um die Ergebnisse vergleichend darstellen zu können.

Die Simulation erfolgte ebenfalls in der allgemeinen Simulationsumgebung, wie in Abschnitt 5.3 dargestellt. Auf das Ausschalten des Empfängers wurde in diesem Fall verzichtet. Um den Empfänger dennoch am Senden von Paketen zu hindern, wurde die Startzeit des RPGs im Ankunftsprozeß des Empfängers auf einen Wert gesetzt, der höher war als die zu simulierende Zeit. Bei einer simulierten Zeit von 36000 Sekunden ist eine Startzeit von z.B. 40000 Sekunden geeignet.

Die Startzeit des Senders war wieder auf 0 zu setzen, um die sofortige Generierung von Paketen zu ermöglichen. Mit diesen Einstellungen wurden vier Simulationen durchgeführt mit Werten für den Parameter T zwischen 0,001 und 1.

Die Simulationszeit betrug hier jedoch nur jeweils etwa 2 Stunden 15 Minuten. Wird der Empfänger mittels „Fail This Node“ ausgeschaltet, wird während der Simulation eine etwa 4 Gigabyte große Log-Datei erstellt. Es ist anzunehmen, daß bei jeder Generierung eines

Paketes ein Eintrag in der Log-Datei erstellt wird, der anzeigt, daß der Empfänger nicht zu erreichen ist. Obwohl diese Vermutung nicht eindeutig bestätigt werden konnte, wäre hier durchaus ein Grund für die langen Berechnungszeiten aus dem vorherigen Abschnitt zu finden.

Auf die Bindarstellung der Simulationsfolgen sei hier verzichtet, wobei hier die Ober- und Untergrenzen weitaus kleinere Bereiche eingrenzen. So war hier die Methode aus Abschnitt 4.2.2 zur Schätzung der Grenzen anzuwenden.

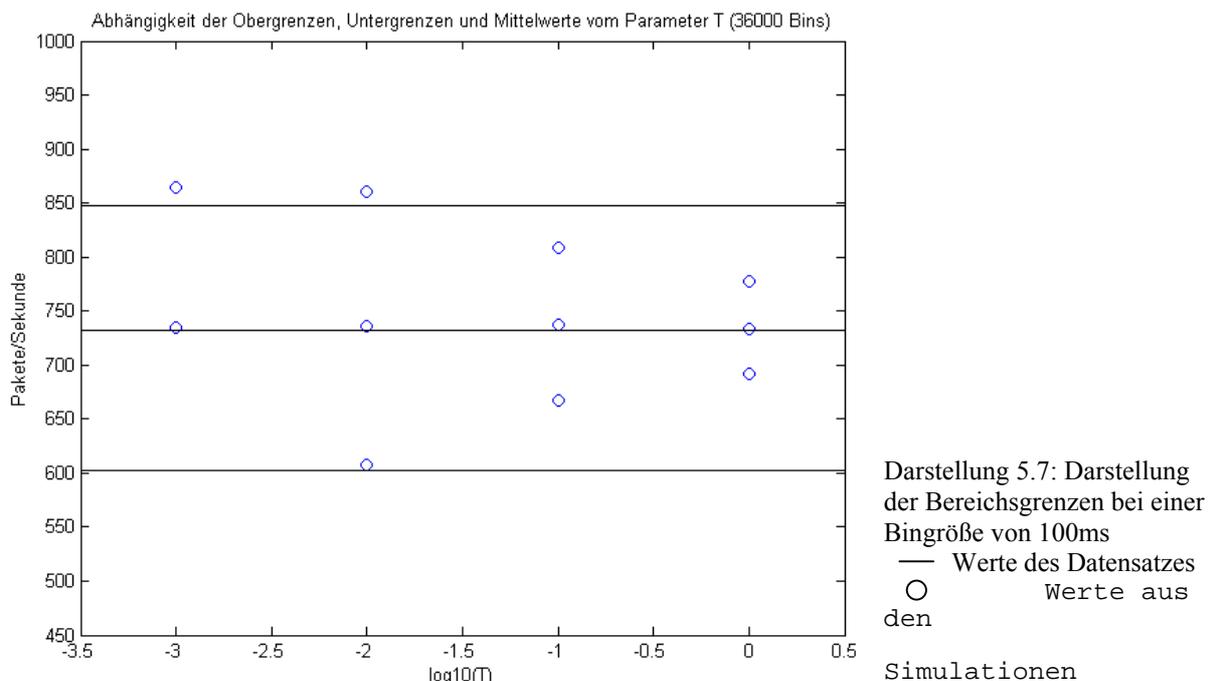
In Tabelle 5.7 sind die Ergebnisse aus der Bereichsschätzung und die resultierenden Hurst-Parameter zusammengefaßt. Der Hurst-Parameter wurde wiederum mittels R/S-Statistik und VTP geschätzt.

	μ	Grenze		Bereich		Hurst		
		untere	obere		in %	R/S	VTP	\bar{H}
Original	732,3	603	847	244	80,05	0,7806		0,7806
0,001	735,3	0	864	k.A.	82,89	0,7367	0,5838	0,6603
0,01	736,2	608	861	253	79,52	0,8140	0,7255	0,7698
0,1	736,8	667	809	142	81,05	0,7853	0,7467	0,7660
1	733,6	692	777	85	81,34	0,7726	0,7311	0,7519

Tabelle 5.7: Binggröße = 100msec, 36000 Bins

In der obigen Tabelle ist zu erkennen, daß bei $T=0,001$ die untere Grenze bei 0 zu finden ist, obwohl die zugehörige mittlere Ankunftsrate kaum abweicht. Dies hängt damit zusammen, daß in der Simulation durch OPNET relativ lange Nullfolgen erzeugt werden. Auch bei einer Wiederholung der Simulation konnte keine Verbesserung im Ergebnis erzielt werden. Es ist zu vermuten, daß hier ein unterer Grenzwert unterschritten wird, welcher mit der zugrundeliegenden Pareto-Verteilung zusammenhängt.

In Darstellung 5.7 sind noch einmal Mittelwert sowie Ober- und Untergrenzen aus der obigen Tabelle dargestellt.



Der Tabelle ist weiterhin zu entnehmen, daß die Hurst-Parameter hier weitaus weniger stark schwanken und gut mit dem Hurst-Parameter aus dem Original übereinstimmen. Die

Simulation mit dem Parameter $T=0,01$ gibt das Verhalten des Originals am besten wieder. Der Hurst-Parameter unterscheidet sich bei diesem Wert für den Parameter T noch um $0,018$. Auch hier ist eine deutliche Übereinstimmung mit den Beobachtungen in [XY01] zu erkennen.

5.5 Simulation des eingehenden und des ausgehenden Verkehrs unter Niedriglast

Anschließend wurde der eingehende und der ausgehende Verkehr einer Verbindung gleichzeitig simuliert, um dem realen Verhalten näherzukommen. Hierzu war eine Veränderung des Modells aus Abschnitt 5.3 vorzunehmen. Der Empfänger hatte nun nicht mehr nur die Aufgabe Pakete zu empfangen, sondern wurde ebenfalls zu einem Sender. Da der eingehende Verkehr unter Vollast bereits in einer Simulation verwendet wurde, war es sinnvoll, eine andere Belastungssituation zu wählen und mit den Ergebnissen aus der Analyse zu vergleichen. Die Wahl fiel auf die Datensätze uni_1h_22.02.03_01:43_in und uni_1h_22.02.03_01:43_out. In Abschnitt 4.7 wurden bereits die benötigten Parameter für diese Simulation bestimmt und sind in Tabelle 5.8 noch einmal zusammengefaßt.

Datensatz	μ [Pakete/s]	R/S (Hurst)
uni_1h_22.02.03_01:43_in	6162,0	0,8207
uni_1h_22.02.03_01:43_out	5133,9	0,8187

Tabelle 5.8: Zusammenfassung der ermittelten Werte

Als Hurst-Parameter kamen hier nur die Ergebnisse aus der R/S-Statistik in Betracht. Mit diesen Parametern konnten nun wiederum mehrere Simulationen durchgeführt werden, welche bei den beiden Stationen getrennt anzugeben waren. Somit war die allgemeine Simulationsumgebung aus Abschnitt 5.3 zu modifizieren. Simuliert wurde mit fünf verschiedenen Werten für den Parameter T zwischen $0,001$ und 1 , bei einer Simulationszeit von jeweils etwa 4 Stunden und 45 Minuten. Da hier der eingehende und ausgehende Verkehr gleichzeitig simuliert wurden, war eine Verdopplung der Berechnungszeit zu beobachten. Neben der zu simulierenden Zeit beeinflusst somit die Anzahl der Pakete pro Zeiteinheit die Simulationszeit maßgeblich. Allerdings ist fraglich, wie groß ein Netzwerk werden darf, wenn hier für zwei Stationen die Berechnungszeit bei über vier Stunden liegt.

Mit der in Abschnitt 4.2.1 beschriebenen Methode wurden wiederum die Bereiche, in denen die Mehrzahl aller Bins anzutreffen waren, geschätzt. Die Ober- und Untergrenzen für den jeweils eingehenden und ausgehenden Verkehr sind in Tabelle 5.9 und 5.10 zu finden. Weiterhin sollen in diesen Tabellen auch die geschätzten Hurst-Parameter aufgeführt werden.

		μ	Grenze		Bereich		Hurst		
			untere	obere		in %	R/S	VTP	\bar{H}
	Original	6162,0	5656	6768	1112	81,00	0,8207	k.A.	0,8207
T	0,001	6241,7	5087	6993	1906	81,58	0,7243	0,6615	0,6929
	0,01	6267,0	5765	6749	984	82,69	0,7524	0,7524	0,7524
	0,1	6188,7	5938	6423	485	83,33	0,7966	0,7379	0,7673
	1	6150,0	6012	6286	274	82,53	0,8043	0,6781	0,7412
	0,005	6259,6	5629	6838	1209	82,69	0,7210	0,7689	0,7450

Tabelle 5.9: Niedriglast, Binsgröße: 1000ms, eingehender Verkehr

		μ	Grenze		Bereich		Hurst		
			untere	obere		in %	R/S	VTP	\bar{H}
T	Original	5133,9	4656	5780	1124	81,53	0,8187	k.A.	0,8187
	0,001	5147,4	4028	5848	1820	81,53	0,7414	0,6846	0,7130
	0,01	5195,9	4731	5620	889	83,06	0,6955	0,7075	0,7015
	0,1	5073,2	4856	5294	438	82,47	0,7625	0,7421	0,7523
	1	5146,4	5023	5270	247	82,47	0,7111	0,7318	0,7215
	0,005	5221,9	4673	5712	1039	83,03	0,7517	0,7124	0,7321

Tabelle 5.10: Niedriglast, Bingröße: 1000ms, ausgehender Verkehr

Bei Betrachtung der Grenzen der entsprechenden Datensätze fällt auf, daß der eingehende und ausgehende Verkehr ähnlich große Bereiche eingrenzen, bei ähnlichen Werten für den Hurst-Parameter. Um nun die Bereichsgrenzen aus den Simulationen besser bewerten zu können, sind diese in Abbildung 5.8 noch einmal vergleichend dargestellt.

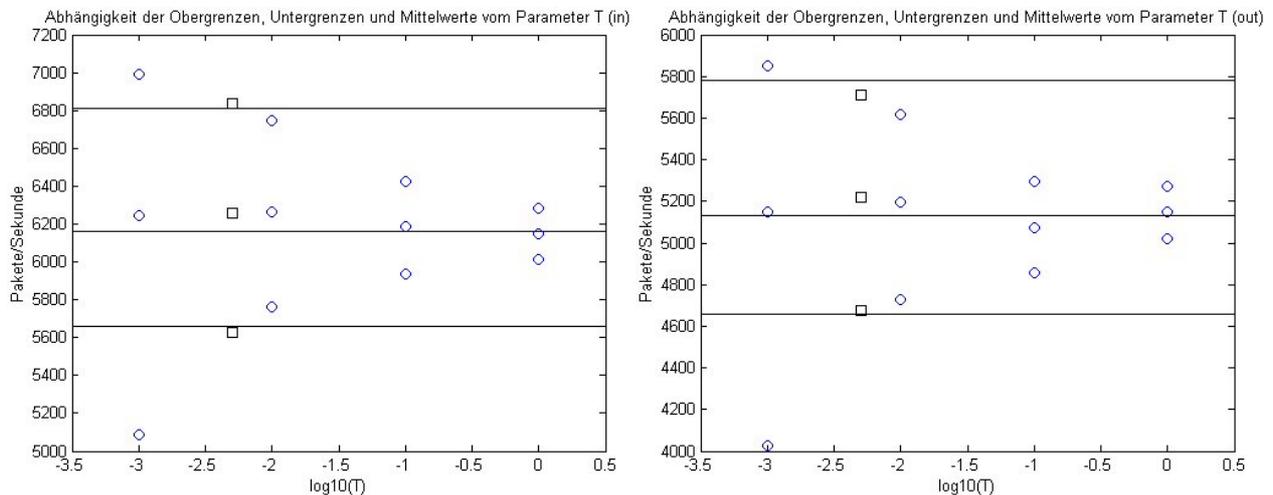


Abbildung 5.8: Vergleich der ermittelten Bereichsgrenzen, links eingehend, rechts ausgehend

Aus den Simulationsergebnissen in den obigen Tabellen und der Abbildung geht hervor, daß mit $T=0,005$ sowohl der eingehende als auch der ausgehende Verkehr am besten beschrieben wird. Beim eingehenden Verkehr (mit $T=0,005$) ist der Bereich um 97 Pakete/s kleiner als der des betrachteten Originals. Beim ausgehenden Verkehr hingegen ist der Bereich bei $T=0,005$ um 85 Pakete/s größer. Es besteht somit die Möglichkeit, daß der Parameter T für den eingehenden und ausgehenden Verkehr verschiedene Werte besitzt.

Weiterhin ist aus den Tabellen 5.9 und 5.10 zu entnehmen, daß der Hurst-Parameter der Simulationsfolgen nicht an den Wert der simulationsvorbereitenden Analyse herankommen. Hier ist jedoch der Unterschied nicht so gravierend wie in Abschnitt 5.4.1.

Hierbei sei zu erwähnen, daß bei der Schätzung des Hurst-Parameters aus den Simulationsfolgen mittels VTP keine Auffälligkeiten im Verlauf festzustellen waren.

Um dies zu verdeutlichen, sind in Abbildung die Variance Time Plots der zu simulierenden Folge und einer der Simulationsfolge vergleichend dargestellt.

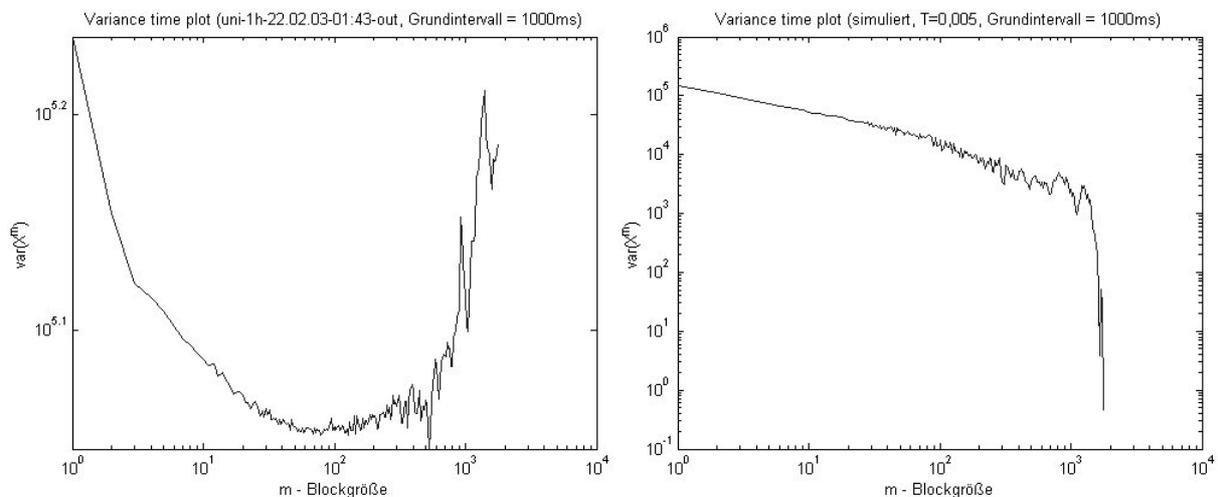


Abbildung 5.9: Variance Time Plot einer Originalfolge und der zugehörigen simulierten Folge

Der Variance-Time-Plot aus der Simulationsfolge folgt dem gewünschten Verlauf. Dies konnte auch bei der Analyse in den vorherigen Abschnitten beobachtet werden. Somit wird ein Fehler in der Implementierung des VTPs eher unwahrscheinlich. Dieses läßt den Schluß zu, daß der „Knick“ im Verlauf durch die Eigenschaften der Datensätze an sich verursacht wird. Das wiederum weist auf eine allgemeine Einschränkung der Verwendungsmöglichkeiten des VTP hin.

Wie auch in den anderen Simulation ist abschließend festzustellen, daß der RPG in OPNET für die Simulation von selbstähnlichem Netzwerkverkehr geeignet ist.

6 Analyse der Zwischenankunftszeiten und Simulation

OPNET ist ein sehr umfangreiches graphisches Netzwerkanalyse- und -Simulationsprogramm. Die Aufgabenstellung dieser Studienarbeit beinhaltet die Simulation einer Punkt-zu-Punkt-Verbindung zweier Netzknoten mittels der aus den Proben gewonnenen Ergebnisse. (vgl. Bild 5.3). Hierzu stellt OPNET für den Knotentyp „Ethernet_Wkstn“ der Toolbox Ethernet sog. *Traffic Generation Parameters* zur Verfügung. Der zugrundeliegende Zufalls-generator beherrscht u.a. die Bildung pareto- und weibullverteilter Größen. Als Hauptvorteil gegenüber dem ebenfalls in dieser Arbeit benutzten *Raw Packet Generator* würde die sehr viel kürzere Simulationszeit gelten. Das zur Verfügung gestellte Computersystem berechnete für 1h zu simulierende Zeit die Weibull-Folge innerhalb (10...15)min, während der RPG hierfür ca. 2h benötigte. Es ergäbe sich also ein Zeitvorteil von 8.

Das Szenario beinhaltet zwei Ethernet-Workstations, von denen eine aus Gründen der Einfachheit (es werden nur die Zwischenankunftszeiten betrachtet) Pakete konstanter Größe aussendet, die vom anderen Rechner empfangen werden sollen. Um die maximal benötigte Bandbreite des Transportmediums zwischen Sender und Empfänger zu bestimmen, dividiert man die Anzahl der Timestamps durch die Beobachtungszeit. Als Maximum kommen 27 Mio. Stamps vor; die Beobachtungszeit betrug immer genau 1h. Das Ergebnis bewegt sich demnach um ca. 7500 Pakete/s. Unter der Voraussetzung, daß nur Pakete der Maximalgröße verschickt wurden, erhält man $B_{\max} = 7500/s \cdot 1536 \cdot 8\text{Bit} = 87,9\text{MBit}/s$. Es reicht also in jedem Fall ein 100Mbit/s-Link für die Simulation aus. Dieses Ergebnis ist konsistent mit der Tatsache, daß der VPN-Server der Universität über eine 100-Mbit/s-Schnittstelle an das DFN angeschlossen ist.

Da der Simulator von OPNET lediglich die Eingabe von Parametern, die sich aus der Auswertung der Zwischenankunftszeiten, und nicht - wie in den vorhergegangenen Abschnitten praktiziert - von Binanalysen unterstützt, ergaben sich Komplikationen bezüglich der Genauigkeit und Speicher-/Rechenzeitanforderungen der errechneten Ergebnisse. Es war nicht möglich, die Traces in ihrer Gesamtgröße von 1h zu verarbeiten. Vielmehr erfolgte eine Aufteilung jeder Probe in je sechs bis acht etwa gleichgroße Anteile, die getrennt analysiert wurden. Die für diesen Fall zwingend erforderliche Voraussetzung von Stationarität ist wegen des kurzen Beobachtungszeitraums näherungsweise erfüllt. Da in OPNET die Verteilungsparameter nur auf der Basis von Sekunden eingegeben werden können, ergaben sich zwangsläufig insbesondere für den Maßstabsparameter sehr kleine Werte, wodurch eine genaue Simulation weiter erschwert wurde. Weiterhin zeigte sich, daß bei dieser Darstellung die Exponentialverteilung überraschend gute Näherungen für die empirischen Verteilungsfunktionen lieferte, was den α - Parameter nahe bei 1 vermuten läßt.

Die Schätzung der Weibull-Parameter verlief nicht problemlos. Die in Matlab integrierte WEIBFIT-Funktion lieferte derart unpräzise Ergebnisse, daß ein „manuelles Herantasten“ an die empirische Verteilungsfunktionen über ein selbstgeschriebenes Skript erforderlich wurde. Weiterhin erscheint der „Heavy Tail“ der integrierten Weibullverteilung zu einflußreich, da auch der äußerst unrealistische Fall von 0 empfangenen/gesendeten Paketen pro Sekunde in jedem Simulationslauf mehrfach auftrat. Demgegenüber ist die große Mehrheit der Paketanzahl pro Sekunde über einen viel kleineren Bereich verteilt, als dies bei den durchgeführten Messungen der Fall ist.

6.1 Gewinnung der Parameter für die Simulation in OPNET

Zunächst erfolgte eine Unterteilung der Traces in Ausschnitte zu jeweils ca. 3Mio. Timestamps. Diese kleineren Proben liefern aufgrund der annähernden Stationarität des zugrundeliegenden Ankunftsprozesses annähernd identische Ergebnisse, so daß im weiteren je Trace nur ein zufällig ausgewählter Abschnitt exemplarisch analysiert wird.

Aus den Tabellen und Abbildungen wird deutlich, daß die Exponentialverteilung als Annäherung der Ankunftsprozesse Ergebnisse liefert, die von der Weibullverteilung teilweise nur geringfügig genauer berechnet werden. Deren Parameter β ist gut durch den Mittelwert der Eingangsfolge approximiert, und α liegt erwartungsgemäß nahe 1. Als Gütemaß für die Genauigkeit der Parameterschätzungen dient der mittlere quadratische Fehler (mean square error, MSE).

Probe	1	2	3
α	1,08	0,86	0,78
$\beta=\mu$	$1,28 \cdot 10^{-4}$	$1,32 \cdot 10^{-4}$	$1,56 \cdot 10^{-4}$
MSE – Weibullverteilung	35,783	3,761	6,289
MSE – Exponentialverteilung	39,045	5,176	10,575

Tabelle 6.1: empfangen

Probe	1	2	3
α	0,91	0,91	0,89
$\beta=\mu$	$1,61 \cdot 10^{-4}$	$1,59 \cdot 10^{-4}$	$1,76 \cdot 10^{-4}$
MSE – Weibullverteilung	6,178	1,215	1,415
MSE – Exponentialverteilung	12,004	2,059	2,626

Tabelle 6.2: gesendet

Wie aus den Vergleichen der mittleren quadratischen Fehler und den folgenden Graphen zu ersehen, gelingt die Approximation durch die Weibullverteilung im Fall der Sendeprozesse z.T. deutlich besser als bei den empfangenen Paketen.

1. Probe 1, empfangener und gesendeter Traffic

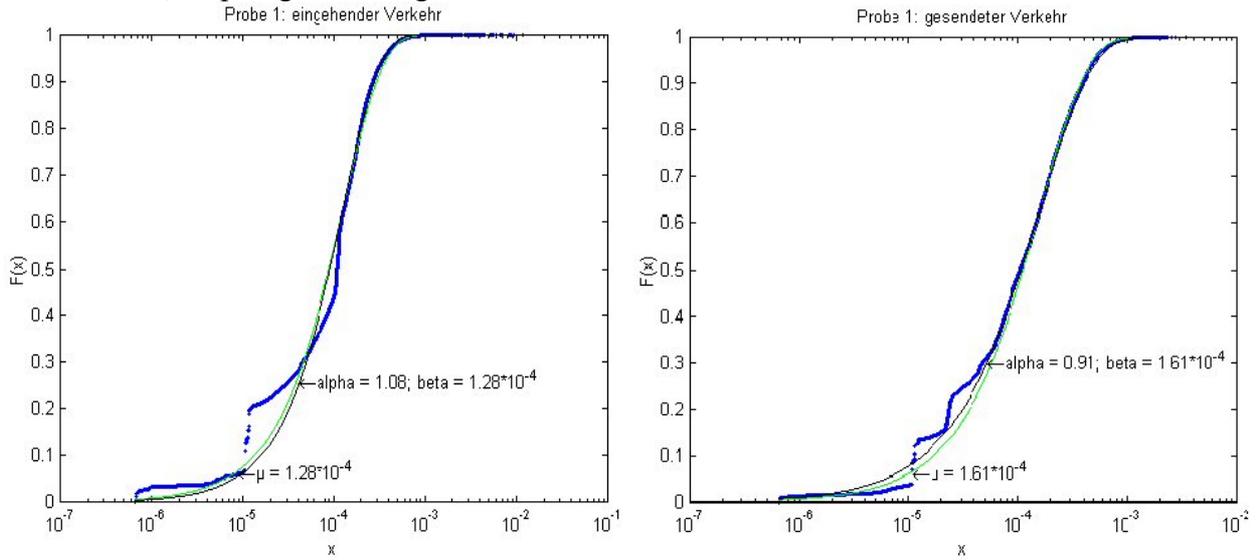


Abb.6.1: Die Weibull-Wahrscheinlichkeitsverteilungsfunktion (schwarz) paßt sich dem Verlauf der empirischen Verteilungsfunktion gut an. Der grün gezeichnete Graph gibt den Verlauf der Exponential-WkV mit dem Parameter μ (Mittelwert) wieder.

2. Probe 2, empfangener und gesendeter Traffic

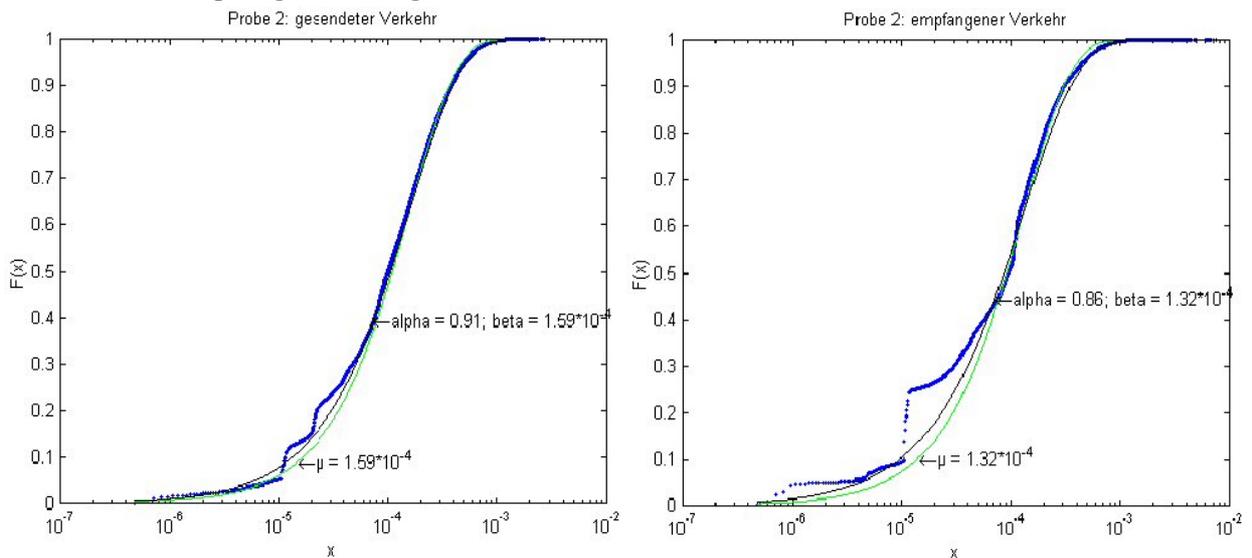


Bild 6.2: Annäherung von Probe 2

3. Probe 3, empfangener und gesendeter Traffic

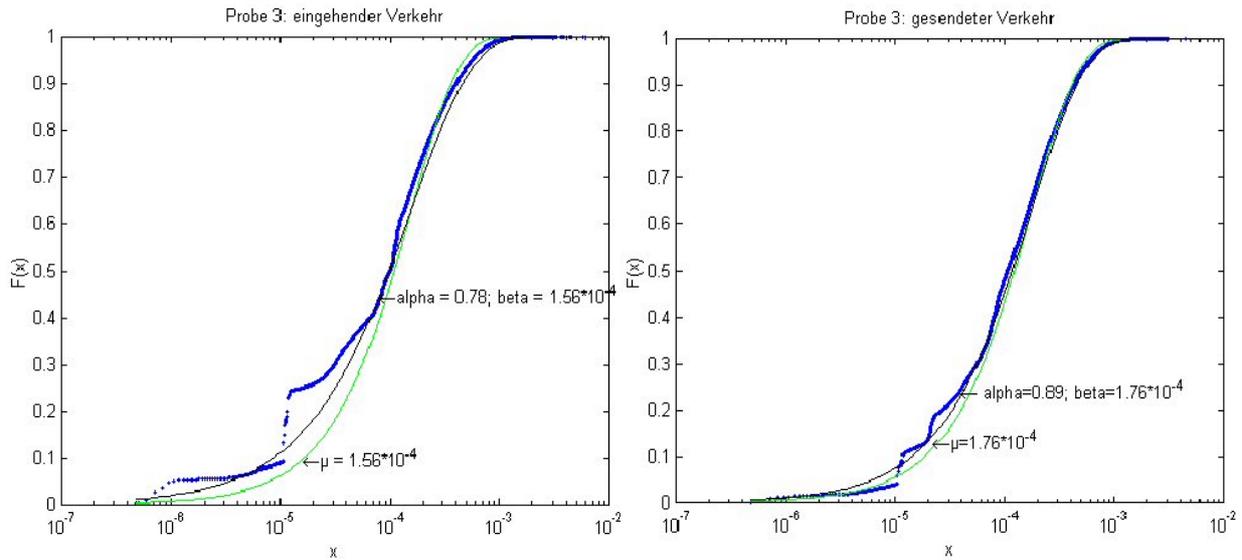


Abbildung 6.3

Es fällt auf, daß in jedem Graphen „Stufen“, deren deutlichste Ausprägung jeweils bei ca. $10^{-5} s$ auftritt, vorkommen, die sich in dieser Form nicht nachbilden lassen. Diese sind auf der Empfängerseite deutlich stärker ausgeprägt, was zwangsläufig zu einem Genauigkeitsverlust gegenüber dem gesendeten Verkehr führt.

Die erhaltenen Parameter wurden in ein vorbereitetes Simulationsszenario von OPNET übernommen. Hierin sind 2 Computer des Typs „Ethernet_Wkstn“ aus der Ethernet-Palette über eine vollduplexfähige 100-Mbit/s-Leitung miteinander verbunden. Die in der Simulation verwendeten Größen sind in der folgenden Tabelle aufgeführt.

Argument – Name	Wert
start time	constant ($(5 \cdot 10^{-5})$)
on state time	exponential (200)
off state time	exponential (1)
stop time	never
packet size	uniform (46,1500)
segmentation	no
interarrival time	Weibull (...,...)

Tabelle 6.3

Es wurde zunächst festgestellt, daß die On/Off State Times keinen Einfluß auf das Simulationsergebnis haben. Die Simulationszeit wurde nun analog zu den Meßproben auf eine Stunde begrenzt. Der Simulator zählte die empfangenen Pakete pro Sekunde. Es ist in der vorliegenden Version 9.1 nicht möglich, das Zählintervall z.B. auf 100ms zu variieren, so daß die Analysen mit der für verlässliche Ergebnisse unzureichenden Anzahl von 3600 Werten durchzuführen waren.

6.3 Analyse der Proben mittels empirischer Bestimmung von Wahrscheinlichkeitsverteilung und Wahrscheinlichkeitsdichteverteilung

Nach [Fe01] ist TCP-Traffic weibullverteilt. Im Rahmen eines alternativen Ansatzes zur Modellierung von Ethernetverkehr wurden die aufgenommenen Proben der Binbildung unterzogen. Um die empirische Wahrscheinlichkeitsdichteverteilung zu ermitteln, wurden die Häufigkeiten der in den Bins enthaltenen Zahlenwerte gegen diese Werte aufgetragen. Die Wahrscheinlichkeitsverteilung konnte durch eine Summation dieses Ergebnisses über die Abzisse gewonnen werden.

Die zu lösende Aufgabe besteht in der Ermittlung der Parameter der Weibullverteilung. Zu diesem Zweck kam auch ein selbstgeschriebenes Skript zum Einsatz, da die in der Statistics-Toolbox vorhandene Weibfit-Funktion nicht immer verwendbare Ergebnisse lieferte. Die Parameter α und β lassen sich folgendermaßen bestimmen. Für die Wahrscheinlichkeitsverteilungsfunktion gilt bei einer dreiparametrischen Weibull-Verteilung

$$F(x) = 1 - \exp\left(-\left(\frac{x-\gamma}{\beta}\right)^\alpha\right). \quad [6.1] \text{ aus [BSMM99]}$$

Im zweiparametrischen Standardfall wird $\gamma=0$ gesetzt. Umstellung und Logarithmierung ergibt

$$\ln(1 - F(x)) = -\left(\frac{x}{\beta}\right)^\alpha. \quad [6.2]$$

Über die Formel zur Berechnung des Mittelwerts,

$$\mu = \beta\Gamma\left(1 + \frac{1}{\alpha}\right), \quad [6.3]$$

läßt sich durch Substitution von β folgender Zusammenhang herstellen:

$$\ln(1 - F(x)) = -\left(\frac{x\Gamma\left(1 + \frac{1}{\alpha}\right)}{\mu}\right)^\alpha; \quad [6.4]$$

$$\Gamma(x) = \int_{x=0}^{\infty} t^{x-1} e^{-t} dt \quad x > 0. \quad \text{aus [BSMM99]}$$

Nach Festlegung eines Startwerts für α wird dieser in frei wählbaren Schritten iterativ erhöht, bis ein Minimum gefunden ist. β folgt sodann direkt aus [6.3]. Das entsprechende Skript, welches im Anhang zu finden ist, wurde mittels eines in Matlab durch die Funktion „Weibrnd“ erzeugten Arrays von 10.000 Elementen erfolgreich getestet.

Da jedoch in den vorliegenden Proben nicht jede Bingröße repräsentiert ist, entstehen in den empirischen Wahrscheinlichkeitsverteilungsfunktionen „Lücken“. Diese werden zunächst mittels linearer Approximation geschlossen. Der Nachteil dieses Ansatzes besteht darin, daß diese nachträglich eingefügten Funktionswerte sich nicht im Originalbinvektor wiederfinden, so daß dessen Mittelwert je nach Größe der Lücken verfälscht wird. Aus diesem Grunde wird α nach Bestimmung von β in einem zweiten Approximationsschritt nochmals berechnet. Diese Vorgehensweise wird durch den Umstand, daß β trotz der erwähnten Ungenauigkeiten bei der Ermittlung von α mit hoher Genauigkeit berechnet wird, ermöglicht. Anschließend wird – wenn erforderlich - der Verschiebeparameter berechnet.

An dieser Stelle sei darauf hingewiesen, daß die empirische Wahrscheinlichkeitsverteilung (WkV) sich keinesfalls exakt durch eine Weibull-WkV modellieren läßt, wie auch aus den folgenden und vorhergehenden Abbildungen hervorgeht. Die Ergebnisse aus [F01] und dieser

Arbeit zeigen jedoch, daß die Weibullverteilung zur annähernden Beschreibung von TCP-Ankunfts- und Sendeprozessen besser geeignet ist als alle anderen bisher untersuchten Verteilungen.

Nachfolgend abgebildete Wahrscheinlichkeitsverteilungsfunktionen sind gegen die Anzahl der gesendeten/ empfangenen Pakete in einem Block aufgetragen. Die Punkte der empirischen WKV sind nicht miteinander verbunden, während die Graphen der parametrisierten Verteilungsfunktionen durchgezogen sind. Zur Bestimmung der Parameter wurden genauigkeitsabhängig die Weibfit() – Funktion bzw. das eigene Skript verwendet.

Probe	1	2*	3
α	27,1	23,1	16,1
β	7820	7498	6326
γ	0	0	0
Minimalwert (Pakete/s)	6612	6048	5110
Maximalwert Pakete/s)	9350	10134	7958
Mittelwert (Pakete/s)	7703	7323	6162
MSE – Weibullverteilung	0,948	2,238	2,449
MSE – Poissonverteilung	33,166	46,957	60,208

Tabelle 6.4: empfangen

Probe	1	2	3
α	109,7	84,0	65,9
β	7902	6922	5483
γ	194	144	124
Minimalwert (Pakete/s)	0	0	0
Maximalwert Pakete/s)	8313	7342	5884
Mittelwert (Pakete/s)	7993	6992	5521
MSE – Weibullverteilung	1,225	1,017	0,990
MSE - Poissonverteilung	9,272	1,895	4,485

Tabelle 6.5: OPNET-Folge, simulierter empfangener Verkehr

Probe	1	2*	3
α	25,3	22,5	13,5
β	6524	6366	5315
γ	0	0	0
Minimalwert (Pakete/s)	5549	5161	4184
Maximalwert (Pakete/s)	7850	9008	6904
Mittelwert (Pakete/s)	6411	6214	5134
MSE – Weibullverteilung	1,327	2,234	3,512
MSE - Poissonverteilung	28,557	39,494	62,558

Tabelle 6.6: gesendet

Probe	1	2	3
α	78,4	83,3	69,7
β	5863	5944	4531
γ	126	116	90
Minimalwert (Pakete/s)	0	0	0
Maximalwert (Pakete/s)	6267	6315	4850
Mittelwert (Pakete/s)	5914	6002	4566
MSE – Weibullverteilung	1,110	0,847	0,824
MSE – Poissonverteilung	2,403	0,635	0,874

Tabelle 6.7 : OPNET-Folge, simulierter Sendeprozess

* - Parameter mit Weibfit() bestimmt

Es zeigen sich große Unterschiede zwischen den simulierten und den tatsächlichen Ankunftsprozessen. Daß die simulierten Prozesse allesamt annähernd poissonverteilt sind, deutet darauf hin, daß der in OPNET integrierte Zufallsgenerator in der vorliegenden Form nicht zur Simulation selbständlicher Vorgänge geeignet ist. Weiterhin werfen die Ergebnisse die Frage auf, ob die Weibullverteilung die gemessenen Ankunftsprozesse entweder nicht genau genug wiedergibt oder ob weibullverteilte Ankunftszeiten notwendige, aber nicht hinreichende Bedingung für Selbstähnlichkeit sind.

Probe 1: Situation empfangsseitig

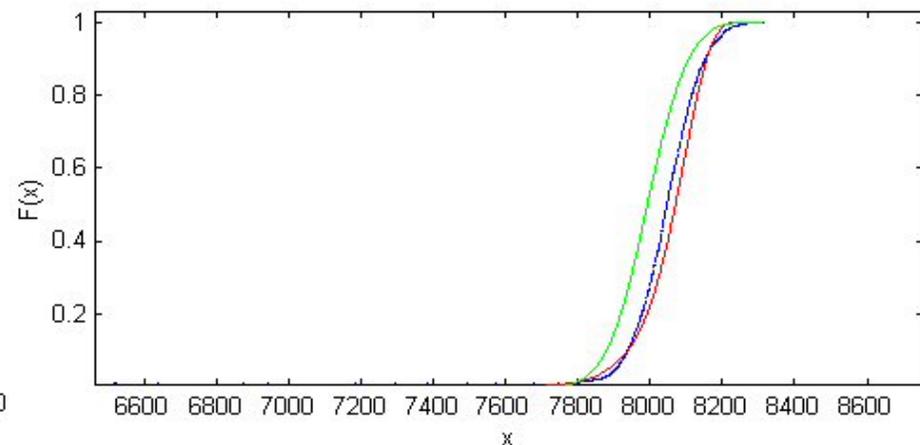
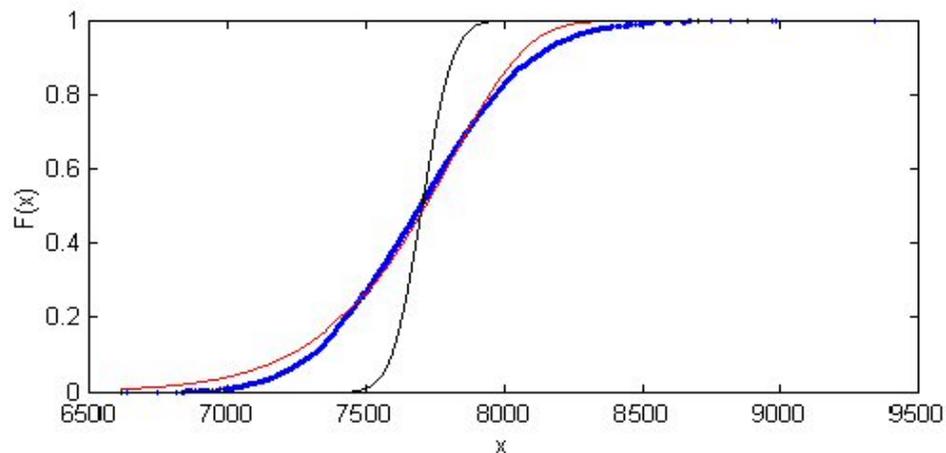
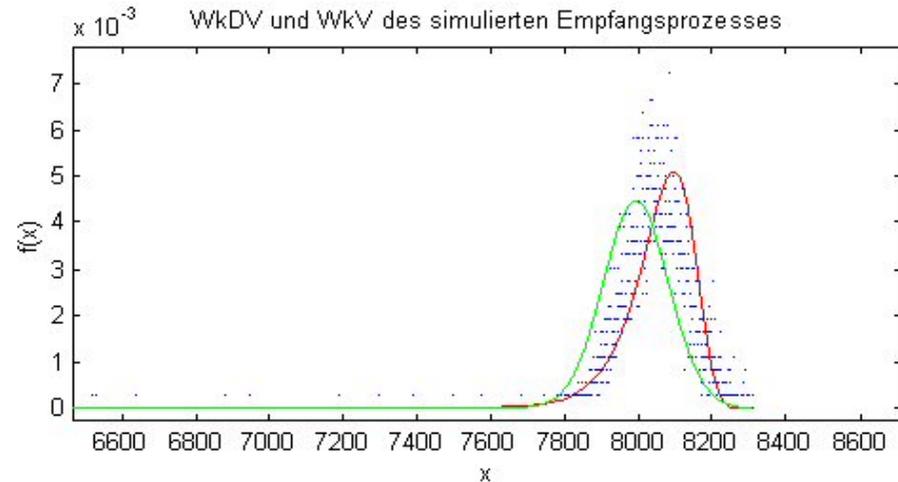
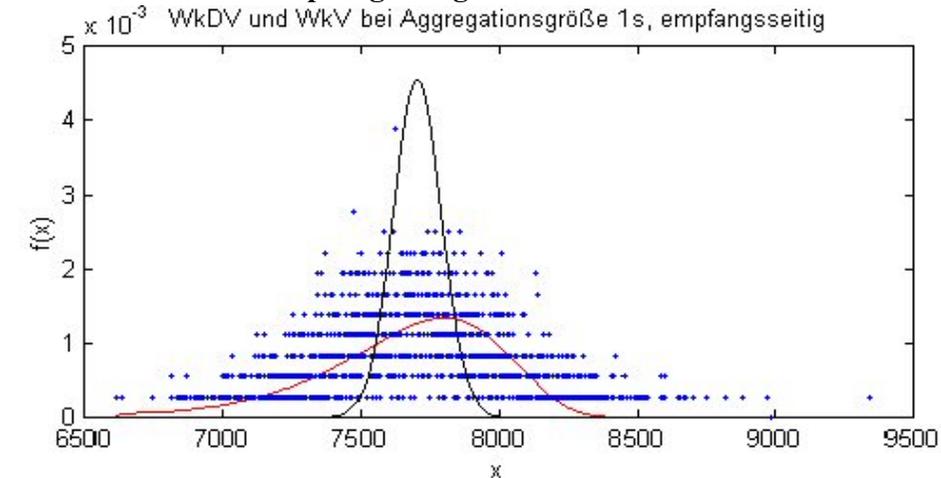


Abbildung 6.4: Wie zu erkennen, wird der aggregierte Empfangsprozess gut durch eine Weibullverteilung mit den entsprechenden Parametern (rot) approximiert. Im Gegensatz hierzu versagt das Poisson-Modell (schwarz) wie auch in den anderen gemessenen Realisierungen, obwohl die Zwischenankunftszeiten annähernd exponentialverteilt sind. Im weiteren wird es nicht mehr berücksichtigt.

Abbildung 6.5: Der in OPNET simulierte Empfangsprozess zeigt, wie schon erwähnt, völlig andere Eigenschaften. Obwohl die Weibullverteilung (rot) weiterhin die genauere Approximation ermöglicht, kommt die Poissonverteilung (grün) der simulierten Eingangsfolge viel näher als in der Originalfolge; der Heavy Tail ist nahezu verschwunden. Unter diesen Umständen ist kein selbstähnliches Verhalten zu erwarten.

Probe 1: Situation sendeseitig

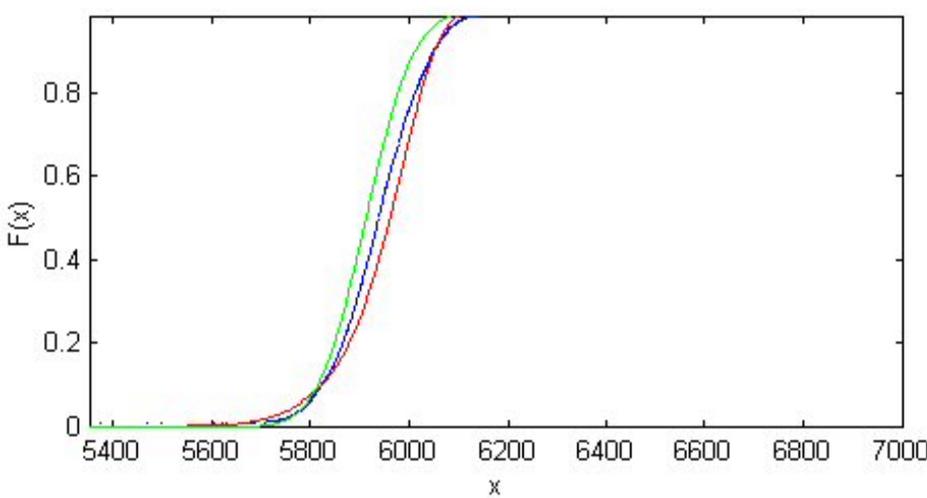
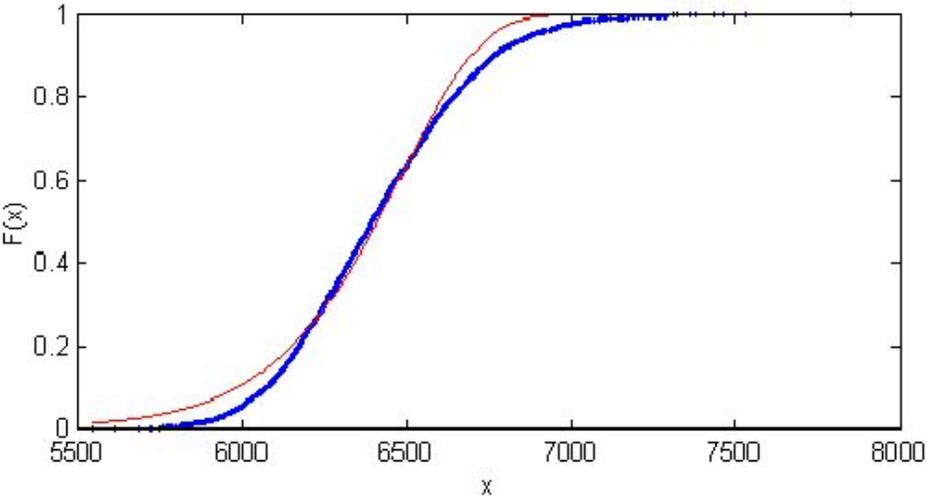
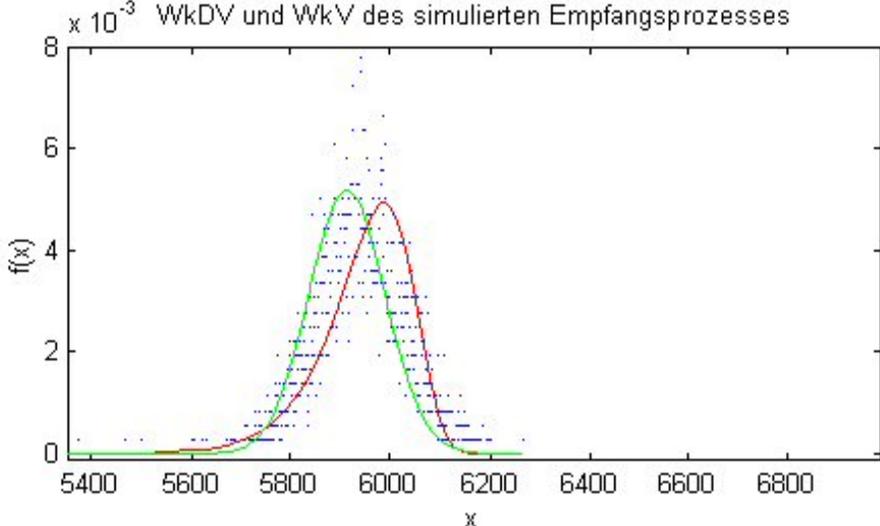
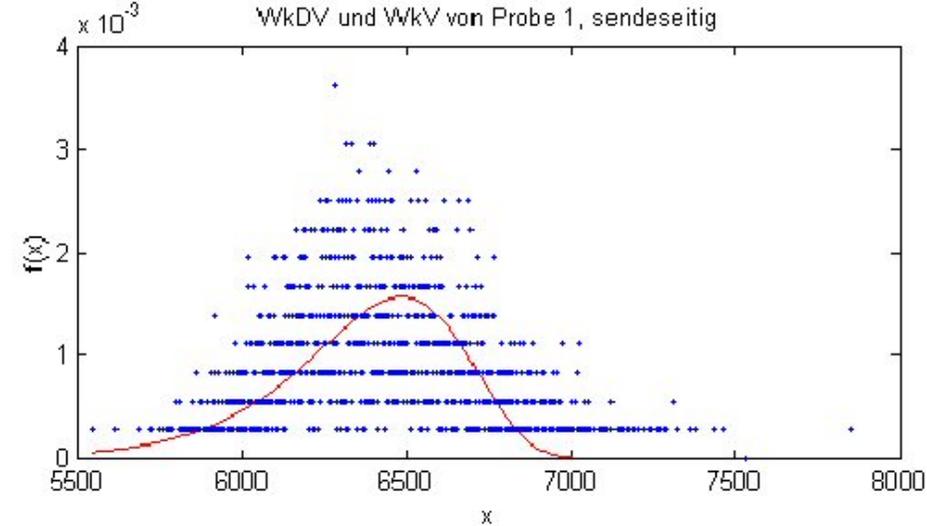


Abbildung 6.6

Abbildung 6.7

6.4 Prüfung der Simulationsergebnisse auf Selbstähnlichkeit

Nach der verteilungsbasierten Analyse folgte ein Test der Proben auf Selbstähnlichkeit. Dieser Teil der Untersuchung der simulierten Proben konzentriert sich auf die Bestimmung des Hurst-Parameters mittels des Periodogramms, des RS-Plots sowie des Variance-Time-Plots. Der Whittle-Estimator lieferte einen Wert von $H > 1$, was auf die geringe Anzahl von nur 3600 Meßwerten zurückgeführt wird. Diese Vermutung stützt sich darauf, daß auch die Berechnung des Hurst-Parameters der Originalfolgen mit einer Blockgröße von 1s diesen Wert ergibt.

Aus Gründen der besseren Übersicht ist lediglich ein Graph abgebildet. Die Wahl fiel auf den empfangenen Verkehr aus Probe 1. Die Hurst-Parameter der anderen Proben sind in Tab. 6.8 zusammengefaßt.

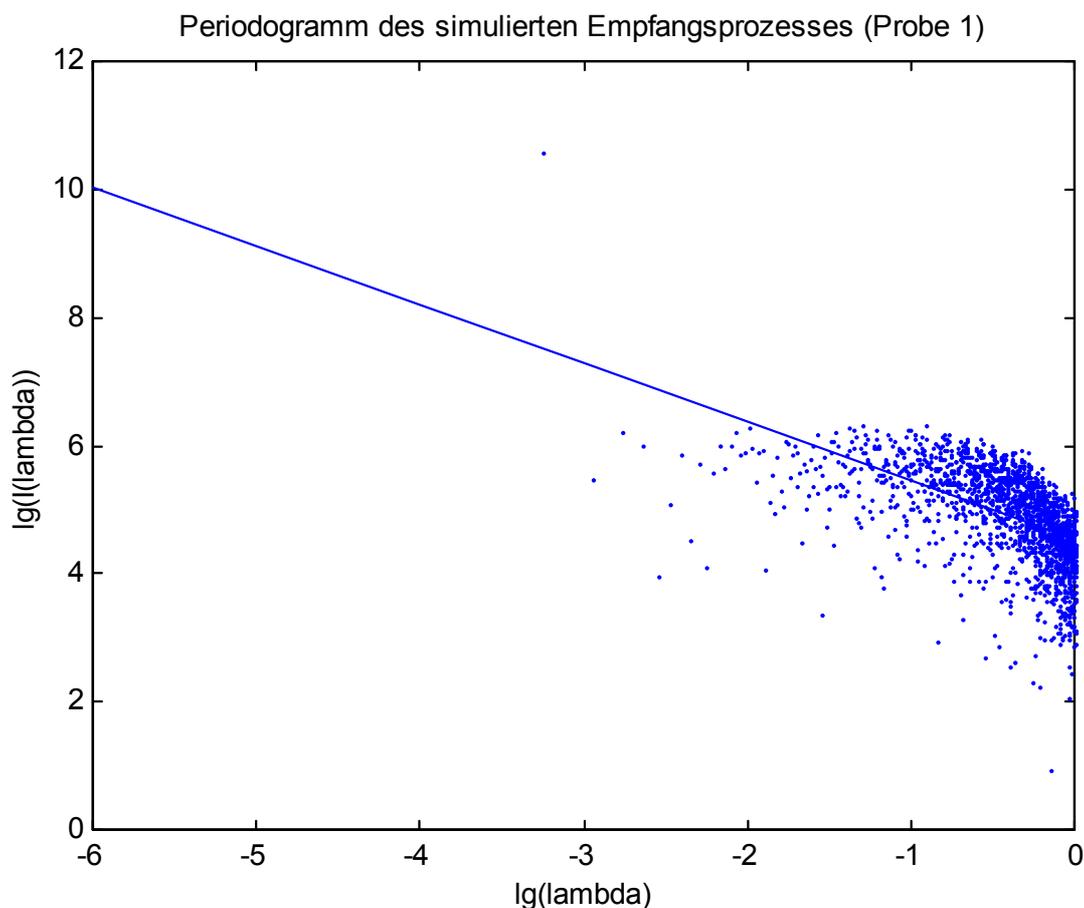


Abb. 6.8: Der Anstieg der Regressionsgeraden liegt bei $-0,918$. Der hieraus berechnete Hurst-Parameter ergibt sich zu ca. $0,96$. Dieser Wert erscheint für ein realistisches Szenario als zu hoch.

Probe	1 in	1 out	2 in	2 out	3 in	3 out
$H_{\text{Periodogramm}}$	0,96	0,91	0,77	1,02	0,91	0,68
H_{VTP}	0,43	1,01	0,45	1,01	1,01	1,01
H_{RS}	0,63	0,54	0,55	0,54	0,54	0,52

Tabelle 6.8

Während das Periodogramm in einigen Fällen durchaus auf Selbstähnlichkeit hindeutet, zeigen Variance-Time-Plot und RS-Plot meist das Gegenteil und bestätigen damit die Ergebnisse aus den Verteilungsuntersuchungen. Es bestätigt sich auch hier die Unzuverlässigkeit des Variance-Time-Plot. Da das Periodogramm in vielen Veröffent-

lichungen als zu „optimistisch“ bezüglich seiner Aussagen zur Langzeitabhängigkeit gilt, sind die Ergebnisse mit realistischen H-Schätzungen als „Ausreißer“ anzusehen. Bestärkt wird diese Schlußfolgerung durch eine Gegenprobe mit dem Zufallsgenerator aus MATLAB. Es wurde ein Vektor mit 3 Mio. Elementen unter Benutzung der ermittelten Weibull-Parameter gebildet und der Binbildung unterzogen. Die Ergebnisse unterschieden sich gegenüber den OPNET-Folgen nicht.

6.5 Erzeugung selbstähnlicher Folgen mittels der Paretoverteilung

Aufgrund der Tatsache, daß der in OPNET realisierte Zufallsgenerator des Raw Packet Generator auf die Paretoverteilung zurückgreift, fiel der Entschluß, für jeweils eine gesendete und eine empfangene Folge die Verteilungsparameter zu bestimmen und in OPNET zu simulieren. Die approximierten Graphen für die Zwischenankunftszeiten sind in Bild 6.9 zu finden. Es fällt sofort auf, daß sowohl Exponential- als auch Weibullverteilung weitaus bessere Annäherungen an die empirischen Verläufe ermöglichen.

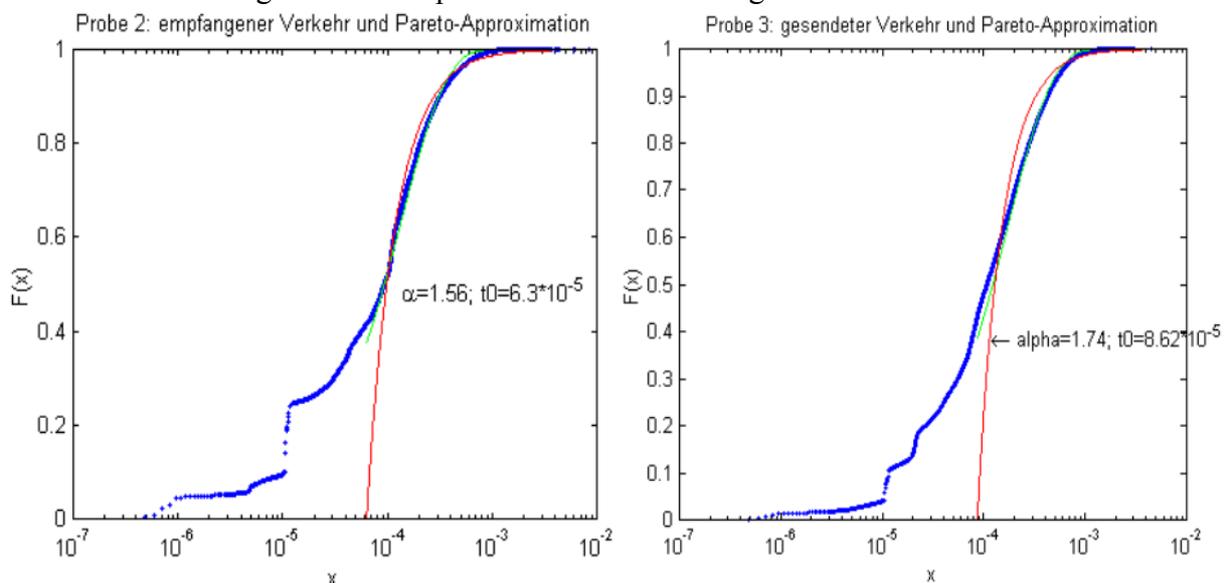


Abbildung 6.9

Auf die Berechnung der Parameter sei kurz eingegangen. Die in OPNET implementierte auf der Paretoverteilung beruhende „rohe“¹ Formel des RPG lautet [RL00]

$$H = \frac{3 - \alpha}{2} \quad \text{mit } H - \text{Hurst-Parameter, } \alpha - \text{Formparameter der Paretoverteilung}$$

Einfache Umstellung ergibt $\alpha = 3 - 2H$.

Für die Simulation kann mit Hilfe dieser Formel aus dem ermittelten Hurst-Parameter direkt der Formparameter der entsprechenden Paretoverteilung berechnet werden. Hierin verbirgt sich auch gleichzeitig die Schwäche des Verfahrens, denn für eine exakte Simulation muß der Hurst-Parameter genau bekannt sein, was mit Schätzmethoden nicht zu bewerkstelligen ist. Zur Verringerung des Rechenaufwands erfolgte auch hier eine Begrenzung auf 3Mio. Feldelemente.

Die Abschätzung des Lageparameters t_0 , der gleichzeitig auch die Form beeinflusst, kann man auf die Methode des kleinsten quadratischen Fehlers zurückgreifen, wie es im entsprechenden Matlab-Skript getan wird. Für die untersuchten Proben sind die Werte in Tab. 6.9 aufgelistet.

¹ Das gilt nur für $t > t_0$. Es werden auch „geglättete“ Prozesse angeboten, die Ankunftszeiten $< t_0$ erlauben.

Probe	21.02.2003, 11:52	22.02.2003, 01:43
Richtung	empfangen	gesendet
Hurst-Parameter	0,72	0,63
α	1,56	1,74
t_0	$6,3 \cdot 10^{-5}$	$8,62 \cdot 10^{-5}$
MSE	14,628	28,972

Tabelle 6.9

Die ermittelten Formparameter liegen zwischen 1 und 2 und belegen damit die Aussage aus Kap. 4. Die weiteren zur Simulation gebrauchten Einstellungen sind in Tab. 6.10 aufgeführt.

Argument – Name	Wert
Start time	constant ($(5 \cdot 10^{-5})$)
On state time	exponential (200)
Off state time	exponential (1)
stop time	never
Packet size	constant (46)
Segmentation	no
Interarrival time	Pareto (...,...)

Tabelle 6.10

Analog zur Simulation der weibullverteilten Zwischenankunftszeiten zählte der Simulator auch hier die empfangenen Pakete pro Sekunde. Die empirischen Verteilungsfunktionen ähneln der Weibullverteilung am besten. Die Poissonverteilung erweist sich hingegen als ähnlich ungeeignet wie in den Fällen der gemessenen Ankunftszeiten. Im Mittel werden jedoch nach wie vor zu wenige Pakete empfangen.

Es läßt sich abschließend festhalten, daß, obwohl die Weibullverteilung die Interarrival-Times deutlich besser nachbildet (vgl. MSE Tab. 6.1, 6.2, 6.9), die Paretoverteilung im Ergebnis überlegen ist, vgl. Bild 6.10.

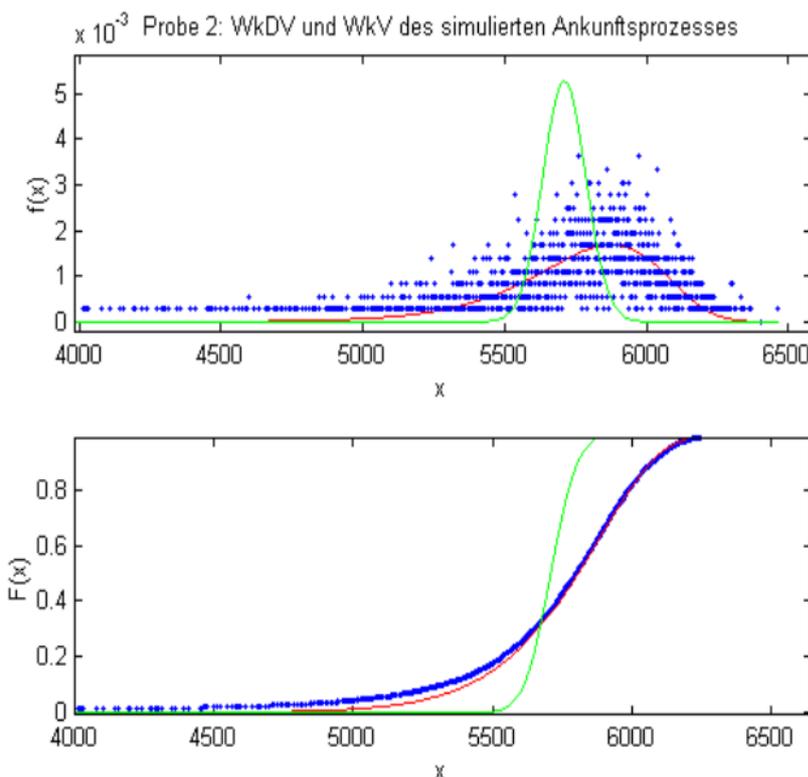


Abb.6.10: Im Vergleich zu weibullverteilten Ankunftszeiten sind hier große Unterschiede zwischen empirischer und Poisson-Wahrscheinlichkeitsfunktion (grün) zu erkennen.

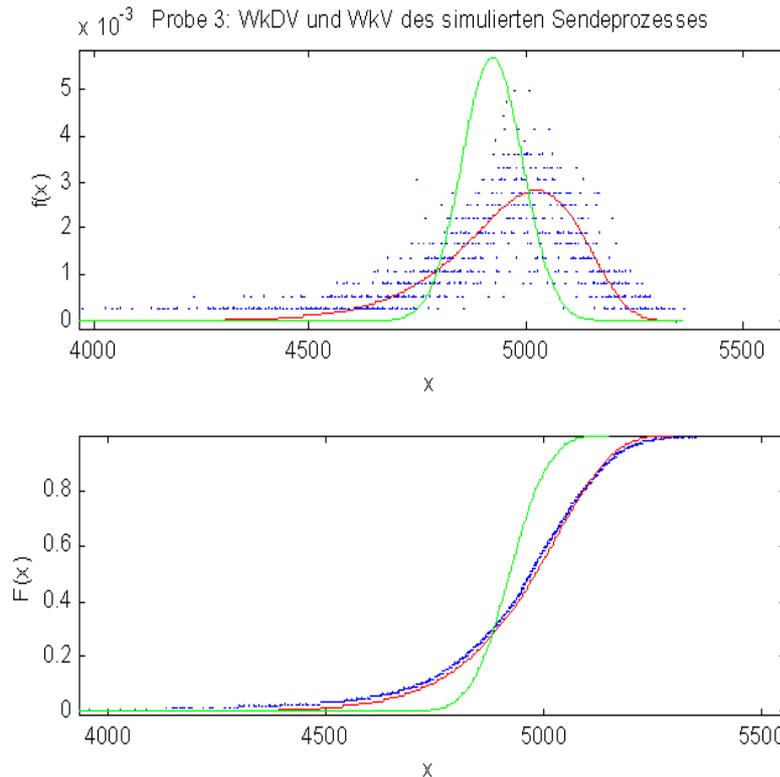


Abbildung 6.11: Bei niedrigerer Auslastung ist im simulierten Fall die Abweichung zwischen empirischer und Poissonverteilung geringer. Das steht im Gegensatz zur gemessenen Situation.

Die Schlußfolgerung aus diesen Ergebnissen lautet, daß die Weibullverteilung über keine fraktalen Eigenschaften, welche für die Erzeugung selbstähnlicher Folgen notwendig sind, verfügt.

Abschließend ist aus Gründen der Konsistenz noch festzustellen, wie genau der zur Berechnung des Formparameters der Paretoverteilung eingesetzte Hurst-Parameter zurückgewonnen werden kann. Hierzu kamen das Periodogramm, der Variance-Time-Plot sowie der RS-Plot zum Einsatz. In der folgenden Tabelle 6.11 sind alle ermittelten Werte von Relevanz zusammengefaßt. Zum Vergleich sind auch die Bezugsgrößen der Originalfolgen eingetragen.

Probe	OPNET-2 in	2 in (21.02.2003, 11:52)	OPNET-3 out	3 out (22.02.2003, 01:43)
α	26,3	23,1	37,6	13,5
β	5737,5	7498	4915,5	5315
γ	149	0	113	0
Minimalwert (Pakete/s)	0	6048	0	4184
Maximalwert (Pakete/s)	6461	10134	5366	6904
Mittelwert (Pakete/s)	5710	7323	4923	5134
MSE – Weibullverteilung	0,497	2,238	0,350	3,512
MSE - Poissonverteilung	38,587	46,957	14,123	62,558
$H_{\text{Periodogramm}}$	0,692	0,723	0,750	0,630
H_{VTP} (Blockgrößen 4 – 400)	0,491	0,884	0,538	-
H_{RS}	0,574	0,917	0,598	0,819

Tabelle 6.11: Da der VTP sich in der Vergangenheit als nicht sehr aussagekräftig erwiesen hat, sollte seine Aussage nicht überbewertet werden. Der RS-Plot gibt noch geringe Anhaltspunkte für Langzeitabhängigkeit.

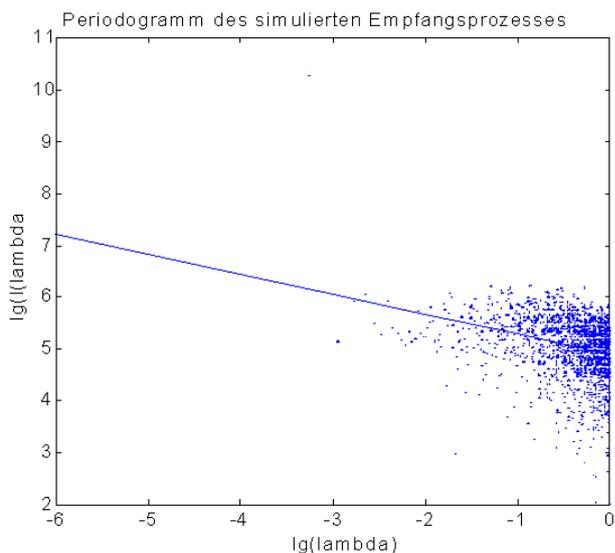


Abbildung 6.12: Der Anstieg beträgt hier $-0,384$. Der Hurst-Parameter ergibt sich demnach zu ca. $0,69$.

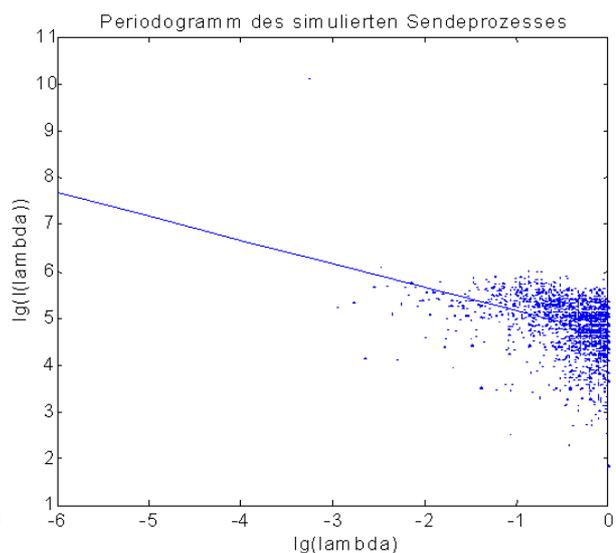


Abbildung 6.13: Obwohl die Vorgabe für H sendeseitig kleiner ist, fällt der Anstieg mit $-0,501$ aggressiver aus. H berechnet sich zu rd. $0,75$.

Mittels der Pareto-Verteilung scheint es nun zwar begrenzt möglich, selbstähnliche Folgen zu erzeugen, jedoch sind beim Maßstabsparameter der Weibullverteilung (der von der mittleren Anzahl empfangener Pakete beeinflusst wird), große Unterschiede feststellbar. Das ist der Tatsache geschuldet, daß in der Simulation 1000 bis 1500 Pakete je Sekunde „zu wenig“ erzeugt werden. Dieser Umstand ist durch einen nochmaligen Blick auf die Graphen der angenäherten Zwischenankunftszeiten schnell erklärt: die approximierte Paretoverteilung läßt keine Ankunftszeiten $< t_0$ zu. Im Fall der empfangenen Pakete liegen von den 3Mio. untersuchten Elementen über 1,24Mio. „links“ des Graphen der Paretoverteilung, sendeseitig sind es sogar ca. 1,29Mio. Fast die Hälfte der Zwischenankunftszeiten wird demzufolge in der Simulation als zu hoch angenommen. Ein Vergleich mit Kap. (5.1.1) verdeutlicht, daß es sich bei dieser Simulation um einen fraktalen Erneuerungsprozeß handelt.

Durch den stark ausgeprägten *Heavy Tail* der reinen Paretoverteilung kommt es in der Simulation zu Fällen, in denen 0 Pakete/s generiert werden, so daß für eine genaue Annäherung der Weibullverteilung zusätzlich der Lageparameter benutzt werden muß. Im *Raw Packet Generator* ist aus diesem Grund eine geglättete Version der Paretoverteilung implementiert, welche diesen äußerst unrealistischen Fall eliminiert und Zwischenankunftszeiten $< t_0$ erlaubt [s. Kap. 5.1, vgl. Kap. 5.4, 5.5].

Während im Fall hoher Netzauslastung (Probe 2) die Formparameter zwischen simulierten und gemessenen Ankunftszeiten nicht signifikant voneinander abweichen, zeigt sich bei Probe 3 in der Simulation eine Annäherung an die Poissonverteilung. Dies steht im klaren Gegensatz zur gemessenen Folge, die gegenüber der Poissonverteilung einen noch größeren quadratischen Fehler aufweist als die Proben unter Vollast. Bei der Simulation von Niedriglast erweist sich die reine Paretoverteilung als ungeeignet, während die Frage der Simulation von Vollastsituationen mittels der Paretoverteilung aufgrund des geringeren Zeitaufwands für grobe, schnelle Abschätzungen offenbleibt.

6.6 Weitere Auswertung einer RPG-Simulation

Ein zweiter Ansatz zur Erzeugung selbstähnlichen Ethernetverkehrs besteht in der Zuhilfenahme des OPNET-eigenen *Raw Packet Generator* (RPG, vgl. Kap. 5.3). Drei der so entstandenen Folgen werden auf Verteilungsfunktion sowie Hurst-Parameter untersucht. Diese Untersuchungen beziehen sich auf Simulationen der eingehenden Verkehr unter Vollast bei einer Bingröße von 1000 Millisekunden. Nähere Informationen hierzu sind in Abschnitt 5.4.1 zu finden.

Probe	1	2*	3*
Fractal Onset Time Scale**	0,01	0,05	0,10
α	13,3	27,0	30,0
β	7955	7822	7555
γ	0	0	0
Minimalwert (Pakete/s)	5517	6474	6248
Maximalwert (Pakete/s)	10403	9066	8629
Mittelwert (Pakete/s)	7684	7664	7417
MSE – Weibullverteilung	1,111	1,579	1,800
MSE – Poissonverteilung	97,160	40,694	30,020
$H_{\text{Periodogramm}}$	0,89	0,86	0,89

Tabelle 6.12: OPNET-Folge

* - Parameter mit Weibfit() bestimmt

** - vgl. Kap. 2.2.4.2

Ein Vergleich mit den im vorigen Abschnitt bestimmten Parametern (vgl. Tab. 6.8, 6.11) impliziert eine deutlich bessere Übereinstimmung mit den Originalfolgen. Eine Gegenüberstellung der in der obigen Tabelle aufgeführten Größen mit denen der Originalfolgen ergibt für Probe 2 die besten Resultate. Ein Blick auf die entsprechenden Graphen (in Abb. 6.14, 6.15 für Probe 2) bestätigt diese Schlußfolgerung. Es ist weiterhin festzustellen, daß mit steigender Fractal Onset Time Scale sich die Prozesse immer weiter einer Poissonverteilung annähern. Der mittlere quadratische Fehler gegenüber der Poissonverteilung sinkt entsprechend, wie in Tab. 6.12 verdeutlicht.

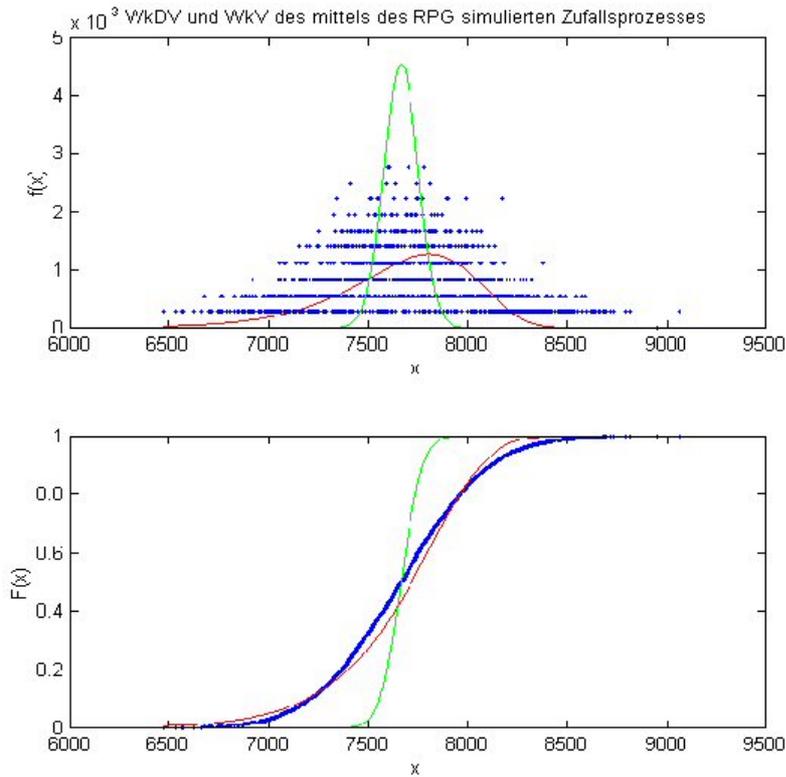


Abb. 6.14: Eine weitaus größere Übereinstimmung mit den empirischen Empfangsprozessen ist feststellbar. Zum Vergleich ist die Poissonverteilung ebenfalls dargestellt (grün).

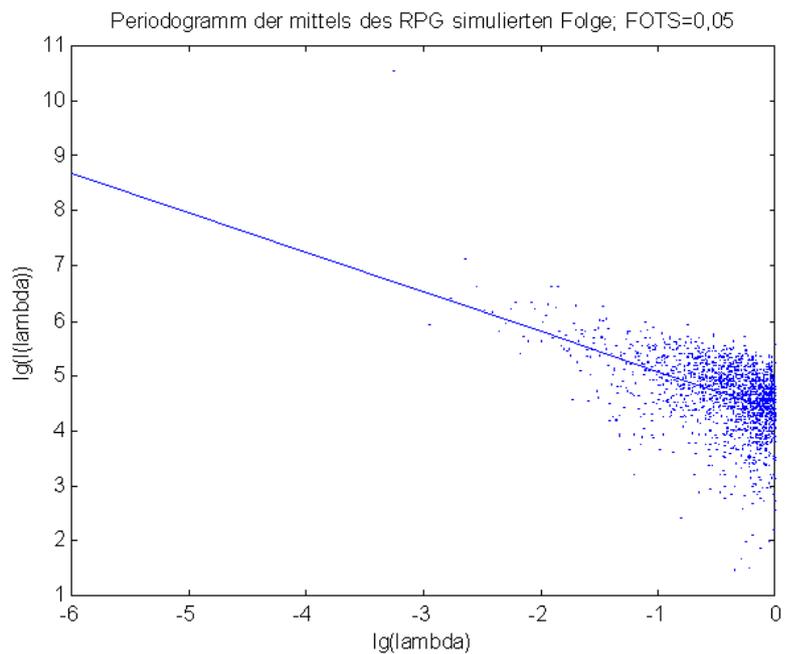


Abb. 6.15: Die Regressionsgerade hat einen Anstieg von ca. 0,719. Das entspricht einem Hurst-Parameter von 0,86.

Diese Schätzung gibt einen starken Hinweis auf selbstähnliches Verhalten des Simulators.

7. Schlußbetrachtung

Es folgt eine Zusammenfassung der in dieser Arbeit gewonnenen Kernaussagen.

In den in dieser Arbeit untersuchten Proben erweist sich nach der Binbildung der Poisson-Prozeß als Modellierungswerkzeug für die aggregierten Ankunftsprozesse in Übereinstimmung mit [F01], [LTWW94], [PF95] u.a. als ungeeignet. Die durch einen zusätzlichen Parameter bedingte höhere Flexibilität der Weibullverteilung ermöglicht eine weitaus bessere Annäherung besonders in Niederlastsituationen. Demgegenüber sind die Zwischenankunftszeiten der untersuchten Proben annähernd exponentialverteilt, wobei die Weibullverteilung eine genauere Approximation ermöglicht (vgl. Kap. 6.1).

Jede Analyse der im Backbone gemessenen Proben ergab eindeutige Anzeichen für Selbstähnlichkeit [vgl. Kap. 4.6ff] und belegt damit die in der einschlägigen Literatur vorgestellten Ergebnisse. Da ein großer Teil des gemessenen Verkehrsaufkommens aus Peer-To-Peer-Quellen stammt, ist als gesichert anzusehen, daß auch diese Verkehrsart Selbstähnlichkeit hervorruft. Als zuverlässig bei der Schätzung des Hurst-Parameters erwiesen sich die R/S-Statistik, das Periodogramm sowie der Whittle-Estimator, sofern eine ausreichende Zahl von Meßwerten vorlag. Der Variance-Time-Plot berechnete für Niederlastsituationen unbrauchbare Ergebnisse. Dieser Fakt widerspricht zusammen mit der ermittelten Abhängigkeit des Hurst-Parameters von der Blockgröße (vgl. Kap. 4.8, 4.11f) der Annahme, daß es sich bei den analysierten Sende- bzw. Ankunftsprozessen um exakt selbstähnliche Prozesse handelt.

Nach der Analyse der erfolgte die Simulation im Programm OPNET. Die Annahme, selbstähnliche Folgen lediglich mittels der jeweiligen Parameter der Weibullverteilung auf stationärer Basis zu erzeugen, führte nicht zum Ziel (vgl. Kap. 6.3f). Dies wäre vor allem wegen des geringeren Zeitaufwandes vorteilhaft gewesen. Ein zweiter Anlauf, welcher die Zwischenankunftszeiten als paretoverteilt annahm, führte trotz geringerer Genauigkeit der Näherung näher ans Ziel. Die vorgegebenen Hurst-Parameter sind allerdings nur für eine Schätzmethode - das Periodogramm - in den jeweiligen Simulationsergebnissen „eindeutig“ nachzuweisen. Auf der anderen Seite weichen die Mittelwerte der korrespondierenden gemessenen und simulierten Folgen sehr stark voneinander ab; die Mittelwerte der simulierten Prozesse liegen weit unter denen der gemessenen (vgl. Kap. 6.5). Für realistische Simulationsergebnisse ist eine Glättung des Verlaufs der Paretoverteilung notwendig; insbesondere, um Zwischenankunftszeiten kleiner als t_0 nachzubilden.

Der ebenfalls in OPNET implementierte Raw Packet Generator verfügt über geglättete Wahrscheinlichkeitsfunktionen, die Ankunftszeiten, welche kleiner bzw. größer als der Verschiebeparameter sind, getrennt berücksichtigen. Hier konnte mit den richtigen Einstellungen ein im Vergleich zur Originalfolge sehr gutes Resultat erzielt werden (vgl. Kap. 5.4f, 6.6). Trotz dieses Beispiels kommt es bei Vorgaben für den Hurst-Parameter $H \geq 0,9$ zu Unzulänglichkeiten in der Simulation, die dazu führen, daß niedrigere Schätzwerte bei der Analyse entstehen.

Als weiterer Nachteil erweist sich der höhere Zeitaufwand gegenüber der rein verteilungsbasierten Methode.

Die in dieser Arbeit gewonnenen Ergebnisse lassen einige Fragen offen.

Eine Vergrößerung der Beobachtungszeit ist für aussagekräftigere Ergebnisse unbedingt notwendig. Durch die damit verbundenen Lastschwankungen bleibt allerdings die Frage offen, inwiefern dann noch von Stationarität ausgegangen werden kann.

Die Abhängigkeiten zwischen Zwischenankunftszeiten und entsprechenden Paketgrößen wurden nicht untersucht. Insbesondere hängen die Zwischenankunftszeiten direkt von der Größe der Pakete ab.

8. Quellen

- [A99] H. Abrahamsson *Traffic measurement and analysis*
www.sics.se/~henrik/t9905.pdf, September 1999
- [BSMM99] Bronstein, Semendjajew, Musiol, Mühlig: *Taschenbuch der Mathematik*
4. Auflage, Verlag Harri Deutsch, Frankfurt a. M.
- [B01] Michael S. Borella: *On Estimating Long Range Dependence of Network Delay*,
International Journal of Chaos Theory and Applications, Vol. 6 (2001), No. 4
- [CL99] H. K. Choi, John O. Limb *A Behavioral Model of Web Traffic* Proc. of
International Conference of Networking Protocol, 1999
- [F00] Anja Feldmann: *Characteristics of TCP Connection Arrivals*, Self-Similar
Network Traffic and Performance Evaluation, John Wiley & Sons, Inc., 2001
ISBN 0-471-31974-0
- [F01] A. Feldmann *Characteristics of TCP connection arrivals*. In: Park, K. (Hrsg.),
Willinger, W. (Hrsg.): *Self-similar network traffic and performance evaluation*.
New York, Chichester, Weinheim [u. a.]: Wiley, 2000
- [HDLK95] C. Huang, M. Devetsikiotis, I. Lambadaris, A.R. Kaye, *Modeling and*
Simulation of Self-Similar Variable Bit Rate Compressed Video: A Unified
Approach Proc. of the SIGCOMM '95 Conference, ff. 114-125, 1995
- [KMV03] T. Kessler, H.-D. Melzer, T. Vergin: *Analyse des Verkehrsaufkommens in*
hochbitratigen Zugangnetzen. In: 4. IuK-Tage Mecklenburg-Vorpommern,
Rostock, 18.-20. Juni 2003
- [KFR02] T. Karagiannis, M. Faloutsos, R. H. Riedi *Long-Range Dependence: Now you*
see it, now you don't! Rice University 2002 <http://cmc.rice.edu/docs/list.pl>
- [LTWW93] W. Leland, M. Taqqu, W. Willinger, D. Wilson, *On the Self-Similar Nature of*
Ethernet Traffic ACM SIGCOMM Computer Communication Review
Volume 23 , Issue 4 (October 1993)
- [LTWW94] W. Leland, M. Taqqu, W. Willinger, D. Wilson, *On the Self-Similar Nature of*
Ethernet Traffic (Extended Version) IEEE/ACM Transactions on Networking,
Februar 1994
- [M95] P. R. Morin *The Impact of Self-Similarity on Network Performance Analysis*
Carleton University 1995
- [P97] Vern Paxson: *Fast, Approximate Generation of Fractional Gaussian Noise for*
Generating Self-Similar Network Traffic
Computer Communications Review, V. 27 N. 5, October 1997, pp. 5-18

- [PF95] V. Paxson, S. Floyd *Wide-Area Traffic - The Failure of Poisson Modeling* University of California, Berkley, 1995
- [R96] B. K. Ryu *Fractal Network Traffic: From Understanding to Implications* PhD thesis, Columbia University, 1996
- [R00] B. K. Ryu, *A Tutorial on Fractal Traffic Generators in OPNET for Internet Simulation*, OPNETWORK 2000.
- [R02] B. K. Ryu *Fractal Traffic Models for Internet Simulation*, Präsentation 2002 <http://www.koren21.net/workshop/2002/pdf/524-9.pdf>
- [RL98] B. K. Ryu und S. Lowen *Point Process models for self-similar network traffic, with applications* Stochastic Models, 14(3):731-761, 1998
- [RL00] B. K. Ryu und S. Lowen *Fractal Traffic Models for Internet Simulation* IEEE Int'l Symposium on Computer Communications, July 2000
- [RO96] O. Rose *Estimation of the Hurst Parameter of Long-Range Dependent Time Series* Universität Würzburg 1996
- [TP98] T. Tuan, K. Park *Congestion Control for Self-Similar Network Traffic* Purdue University, West Lafayette 1998
- [VKM00] A. Veres, Zs. Kenesi, S. Molnár, G. Vattay: *On The Propagation of Long-Range Dependence in the Internet*, Ericsson Research, Budapest 2000
- [VM02] G. Varatkar, R. Marculescu *Traffic Analysis for On-chip Networks Design of Multimedia Applications* Proc. 39 Design Automation Conference, 2002 http://www.gigascale.org/pubs/198/49_4.pdf
- [WP98] W. Willinger, V. Paxson, *Where Mathematics meets the Internet*, Notices of the American Mathematical Society, vol. 45, no. 8, ff. 961-70, 1998.
- [WPRT01] W. Willinger, V. Paxson, R. H. Riedi, M. Taqqu *Long-Range Dependence and Data Network Traffic*. In: P. Doukhan, G. Oppenheim and M. S. Taqqu *Long-range Dependence: Theory and Applications*, Birkhauser, 2001.
- [XY01] Fei Xue, S. J. Ben Yoo, *The Effect Of Aggregation on Self-similar Traffic*, OPNETWORK'01 (2001) (http://sierra.ece.ucdavis.edu/documents/xf_opnetwork_2001.pdf)

Anhang

Verwendete Hard- und Software

	System 1	System 2
Prozessor	AMD Athlon XP +2400 (2000MHz)	Intel Celeron Coppermine 600 (900MHz)
Speicher	512 MB DDR-RAM	256 MB SD-RAM
Betriebssystem	MS Windows ME	MS Windows 2000 pro
MATLAB-Version	5.1/6.1	5.1

Tabelle A.1: Analyse der Datensätze

	System 3
Prozessor	AMD Athlon XP 2100+
Speicher	1536 MB RAM
Betriebssystem	MS Windows 2000 Server
OPNET-Version	Modeler 9.1

Tabelle A.2: Simulation in OPNET