

Nicht-asymptotische Leistungsbewertung und abtafbasierte Parameterschätzung für Kommunikationsnetzwerke mit langzeitkorreliertem Datenverkehr*

Amr Rizk

Institut für Kommunikationstechnik

Leibniz Universität Hannover

amr.rizk@ikt.uni-hannover.de

Abstract: Die Entdeckung der statistischen Eigenschaften der Selbstähnlichkeit und Langzeitkorrelation im Internetverkehr hat die Annahmen und somit die Anwendbarkeit traditioneller Modelle bzw. Leistungsbewertungsmethoden in Kommunikationsnetzen wie dem Internet in Frage gestellt. Diese Arbeit liefert obere Schranken für Leistungsmetriken in Netzwerken mit langzeitkorreliertem Datenverkehr basierend auf dem stochastischen Netzwerkkalkül. Ein grundlegendes Resultat dieser Arbeit beziffert den Zuwachs von Ende-zu-Ende Latenzen für Netzwerkpfade mit langzeitkorreliertem Verkehr. Außerdem werden Methoden zur abtafbasierten Schätzung der statistischen Eigenschaften des Datenverkehrs vorgestellt. Praktische Messergebnisse untermauern die analytischen Resultate zur Anwendbarkeit und Genauigkeit der vorgeschlagenen Schätzer.

1 Einleitung und verwandte Arbeiten

Die Eigenschaften der Selbstähnlichkeit und Langzeitkorrelation im Internetdatenverkehr können schwer auf traditionelle Verkehrsmodelle bzw. Leistungsbewertungsmethoden für Kommunikationsnetzwerke abgebildet werden [LTWW94, FGHW99]. Approximationen und asymptotische Resultate aus der Literatur weisen auf einen starken Einfluss der Langzeitkorrelation auf die Dienstgüte hin [Nor94, FMN00, DO95]. Langzeitkorrelationen “Long Range Dependence (LRD)” werden durch den sogenannten Hurstparameter $H \in (0.5, 1)$ charakterisiert. Ein etabliertes, stochastisches Modell für den selbstähnlichen LRD Internetdatenverkehr ist die Superposition aus einem linearen Prozess mit der Rate λ und einer fraktalen Brownschen Bewegung “fractional Brownian motion (fBm)” mit einem Varianzparameter σ^2 und dem Hurstparameter $H \in (0.5, 1)$ [Nor94]. Der entsprechende, stationäre Inkrementprozess $Y(t)$ besitzt eine abklingende Autokovarianzfunktion¹, so dass $c_Y(\tau) \simeq H(2H - 1)\sigma_Y^2\tau^{2H-2}$ für $\tau \rightarrow \infty$ gilt [Ber94]. Dies steht im Kontrast zur traditionellen Annahme der exponentiell abklingenden Autokovarianz für z.B. Markov oder ARMA Verkehrsmodelle. Das Entstehen von Selbstähnlichkeit und LRD im

*Englischer Titel der Dissertation: “Non-asymptotic Performance Evaluation and Sampling-based Parameter Estimation for Communication Networks with Long Memory Traffic”

¹Definiert für Prozess $Y(t)$ und Abstand τ als $c_Y(\tau) := \mathbb{E}[Y(t)Y(t + \tau)] - \mathbb{E}[Y(t)]\mathbb{E}[Y(t + \tau)]$.

Datenverkehr wurde sowohl analytisch untersucht als auch immer wieder in Internetverkehrsmitschnitten festgestellt und nachemuliert [GMR09, LGD⁺10].

Praktische Studien, Approximationen und asymptotische Resultate deuten darauf hin, dass sich Warteschlangensysteme im Zusammenhang mit LRD Verkehr fundamental anders verhalten als in Verbindung mit gedächtnislosem oder Markov Verkehr. Für fBm Verkehr mit den Parameter $\lambda, \sigma^2, H \in (0.5, 1)$ beziffert ein prominentes Resultat die Wahrscheinlichkeit, dass der Pufferfüllstand B an einem Bediensystem mit konstanter Kapazität C eine Grenze b überschreitet mit

$$P[B > b] \approx \exp \left(-\frac{1}{2\sigma^2} \left(\frac{C - \lambda}{H} \right)^{2H} \left(\frac{b}{1 - H} \right)^{2-2H} \right) := \varepsilon_a. \quad (1)$$

Im Vergleich ist gedächtnisloser bzw. Markov Verkehr für den exponentiellen Tail-Abfall der Pufferfüllstandverteilung bekannt. Das Resultat (1) wurde als Approximation in [Nor94] und als asymptotische Grenze für $b \rightarrow \infty$ in [DO95] hergeleitet.

In den letzten Jahren hat sich das stochastische Netzwerkalkül als mathematisches Rahmenwerk für die Leistungsbewertung von Kommunikationsnetzen entwickelt [Cha00, BLP06, CBL06]. Dieses basiert auf der separaten Beschreibung von dem Datenverkehr einerseits und dem von Netzwerkkomponenten bereitgestellten Dienst andererseits. Diese Betrachtung führt zu einer hohen Anwendbarkeit des Netzwerkalküls, da somit unterschiedliche Verkehrs- und Dienstmodelle kombiniert werden können. Weiterhin ermöglichen Kompositionsresultate die Abstraktion von Netzwerkpfeilen als ein zusammengefasstes Äquivalentsystem und somit die Ausweitung der Analyse auf Ende-zu-Ende Pfade. In dieser Arbeit wird das stochastische Netzwerkalkül benutzt, um nicht-asymptotische, obere Schranken für Dienstgütemetriken für den LRD Verkehr zu bestimmen. Es werden Szenarien der Ressourcenverteilung (Scheduling) und der Ende-zu-Ende Performanz analysiert. Die hier zusammengefassten Ergebnisse sind in [RF12, RBF13] erschienen.

Das Netzwerkalkül beschreibt die Beziehung zwischen den kumulativen Datenankünften A und Abgängen D an einem Warteschlangensystem wie in Abb. 1. Die Ankünfte und Abgänge in einem Zeitintervall $[s, t]$ lauten $A(s, t)$ bzw. $D(s, t)$ für $0 \leq s \leq t$. $A(0, t)$ wird als $A(t)$ abgekürzt. Angenommen werden hier verlustfreie Systeme und ein fluid Datenmodell. Zeitfunktionen sind nicht-negativ, nicht-fallend, linksstetig angenommen und gehen durch den Ursprung, z.B. $0 \leq A(s) \leq A(t), \forall s \leq t$ und $A(0) = 0$. Bediensysteme sind “work-conserving” und haben eine Kapazität C . Angenommen werden Topologien ohne zyklische Abhängigkeiten wie z.B. in Abb. 1. Statistische, punktweise Einhüllende sind Schranken für die stationären Ankünfte im Sinne von $P[A(s, t) > E(t - s) + b] \leq \varepsilon_p(b)$ [YS93]. Ankünfte der Verkehrsklasse “Exponentially Bounded Burstiness (EBB)” besitzen eine lineare Einhüllende $E(t) = \rho t$ mit $\varepsilon_p(b) = \phi e^{-\theta b}$ und $\phi, \theta > 0$ wie z.B. gedächtnisloser bzw. Markov Verkehr. Das Ziel der Analyse sind Leistungsschranken, die als Dienstgütekriterien gelten wie z.B. $P[B > b] \leq \varepsilon$. Diese kann mithilfe der Lindley Gleichung umformuliert werden als $P[B > b] = P \left[\sup_{s \in [0, t]} \{A(s, t) - C(t - s)\} > b \right]$. Die Approximation in (1) zieht das Supremum aus dem Wahrscheinlichkeitsausdruck heraus und verwendet die punktweise Einhüllende der Ankünfte. Dies führt jedoch zu einer unerwünschten unteren Schranke für $P[B > b]$. Ein Ansatz zur Bestimmung einer oberen Schranke für diesen Ausdruck basiert auf der *pfadweise* Einhüllenden $E(t) \leq b + Ct$, so

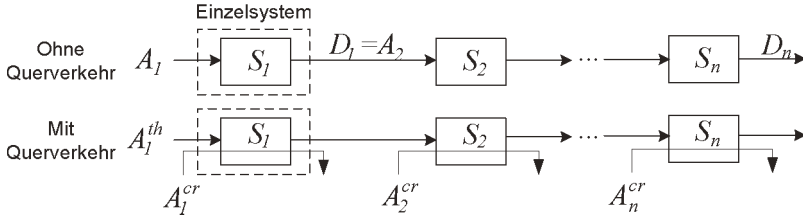


Abbildung 1: Kumulative Ankünfte A bzw. Abgänge D in einem Network von Warteschlangensystemen. Ankünfte und Abgänge sind über die Dienstkurve S verknüpft. Durchgehender Verkehr A^{th} teilt sich Ressourcen an jedem System mit Querverkehr A^{cr} .

dass $P \left[\sup_{s \in [0, t]} \{A(s, t) - C(t - s)\} > b \right] \leq P \left[\sup_{s \in [0, t]} \{A(s, t) - E(t - s)\} > 0 \right]$ gilt. Das Pendant zur Einhüllenden ist die statistische Dienstkurve $P[D(t) < \inf_{s \in [0, t]} \{A(s) + [S(t - s) - b]_+\}] \leq \varepsilon_d(b)$ mit $[x]_+ = \max\{x, 0\}$ [CBL06]. Zur Herleitung von Leistungsschranken werden Einhüllende und Dienstkurven wie folgt kombiniert: Für Ankünfte mit einer Einhüllenden $P \left[\sup_{s \in [0, t]} \{A(s, t) - E(t - s)\} > b \right] \leq \varepsilon_s(b)$ zu einem System mit der obigen Dienstkurve gilt folgende Schranke für den Pufferfüllstand $P[B > \sup_{s \geq 0} \{E(s) - S(s)\} + b] \leq \varepsilon(b)$, mit $\varepsilon(b) = \inf_{x \in [0, b]} \{\varepsilon_s(x) + \varepsilon_d(b - x)\} := \varepsilon_s \otimes \varepsilon_d(b)$. Die Operation $\otimes$ ist bekannt als Min-Plus Faltung [Cha00]. Abbildung 1 zeigt den Fall von durchgehendem Verkehr $A^{th}(t)$ an Systemen mit Querverkehr $A^{cr}(t)$ mit pfadweise Einhüllenden $E^{cr}(t)$ und Verletzungswahrscheinlichkeiten $\varepsilon_s(b)$. Ohne Scheduling Annahmen liefert $S^{lo}(t) = [Ct - E^{cr}(t)]_+$ mit $\varepsilon_s(b)$ eine gültige Dienstkurve für den verfügbaren Dienst für den durchgehenden Verkehr $A^{th}(t)$.

Eine stochastische Netzwerkdienstkurve für eine Reihenschaltung (wie in Abb. 1) von n Systemen jeweils mit obiger Dienstkurve ist in [CBL06] hergeleitet als $S^{net}(t) = S_1 \otimes S_2^{-\varrho} \otimes \dots \otimes S_n^{-(n-1)\varrho}(t)$ mit $S_i^{-\varrho} = S(t) - \varrho t$, den freien Ratenparameter $\varrho > 0$ und der Verletzungswahrscheinlichkeit $\varepsilon^{net}(b) = \varepsilon_1^\varrho \otimes \varepsilon_2^\varrho \otimes \dots \otimes \varepsilon_{n-1}^\varrho \otimes \varepsilon_n(b)$. Die Netzwerkdienstkurve basiert auf einer sogenannten pfadweisen Dienstkurve, die wie folgt definiert ist: $P \left[\sup_{t \in [0, u]} \{ \inf_{s \in [0, t]} \{A(s) + [S(t - s) - \varrho(u - t) - b]_+\} - D(t) \} > 0 \right] \leq \varepsilon^\varrho(b)$ mit $\varepsilon^\varrho(b) = \frac{1}{\varrho} \int_b^\infty \varepsilon_d(x) dx$ und $0 \leq s \leq t \leq u$. Die Netzwerkdienstkurve bildet ein Äquivalentsystem und erlaubt somit die Herleitung von Ende-zu-Ende Leistungsschranken für Netzwerkpfade.

2 Problemstellung

Der erhebliche Einfluss der statistischen Eigenschaften von langzeitkorreliertem Datenverkehr auf die Dienstgüte begründet die Erforschung von Methoden zur Schätzung dieser Eigenschaften. Traditionelle Methoden basieren auf einer vollständigen Aufnahme des Datenverkehrs. Abschnitt 3 zielt auf die Konstruktion von statistischen Schätzmethoden für den netzweiten Einsatz ohne administrativen Zugriff auf den Datenverkehr.

Die Approximation (1) deutet darauf hin, dass LRD einen wesentlichen Einfluss auf die Netzwerkperformanz besitzt. Eine punktweise Einhüllende für fBm Verkehr mit $H \in$

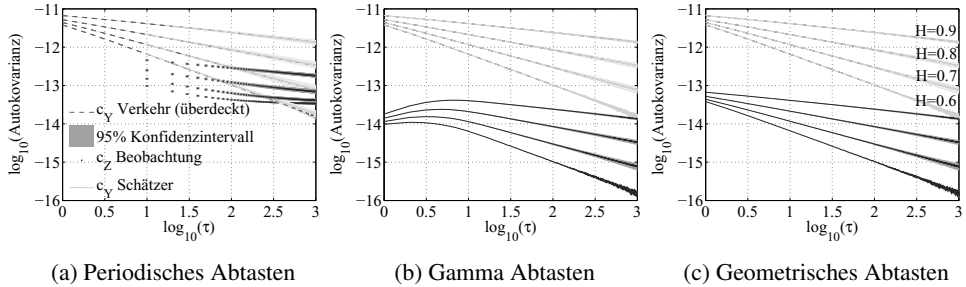


Abbildung 2: Autokovarianzinferenz aus abgetasteten LRD Prozessen für verschiedene Hurstparameter H . Die Zwischenabtastzeiten sind (a) konstant, (b) Gamma bzw. (c) geometrisch verteilt mit $\mu_X = 0.1$. Die Autokovarianz " c_Z Beobachtung" ist in (a) und (b) verzerrt und in (c) nur verschoben. Die geschätzte Autokovarianz " c_Y Schätzer" stimmt mit der Autokovarianz des Verkehrs " c_Y traffic" überein.

(0.5, 1) existiert in der Literatur [FMN00]. Standardmethoden zur Konstruktion von pfadweise Einhüllenden verletzen die Nebenbedingungen der Herleitung von Leistungsschranken für fBm Verkehr [Riz13]. Die Herleitung von nicht-asymptotischen, oberen Leistungsschranken für LRD Verkehr für Einzelsysteme und Netzwerkpfade war für eine gewisse Zeit ein offenes Problem, welches durch einen neuen Ansatz in den Abschnitten 4 und 5 bewältigt wird.

3 Abtastbasierte Schätzung statistischer Eigenschaften des langzeitkorrelierten Verkehrs

Im Folgenden wird ein Modell zur Schätzung von LRD Verkehrseigenschaften mittels Abtastung präsentiert. Ein fundamentales Abtastresultat aus der Warteschlangentheorie beweist, dass der Anteil der Poisson Ankünfte, die ein Warteschlangensystem in einem bestimmten Zustand erfahren, im Durchschnitt gleich dem Anteil der Zeit ist, die das System in diesem Zustand verbringt. Dieses Resultat ist bekannt als PASTA, "Poisson Arrivals see Time Averages" [Wol82]. Eine Verallgemeinerung dieses Resultats ist bekannt als NI-MASTA "Non-intrusive Mixing Arrivals See Time Averages" [BMVB09] und bildet die Grundlage für erwartungstreue, abtastbasierte Schätzer von beliebigen Momenten von Zufallsvariablen.

Das Ziel dieser Arbeit ist eine stochastische Abtastung zur erwartungstreuen Inferenz der Kovarianz des LRD Verkehrs. Das Abtastmodell ist definiert auf Basis von drei stationären, zeitdiskreten Zufallsprozessen, nämlich $Y(t)$, $X(t)$ und $Z(t)$ für $t \in \mathbb{N}_0$. Der LRD Verkehrsinkrementprozess $Y(t)$ ist charakterisiert durch die Autokovarianzfunktion aus Abschnitt 1. Der Abtastprozess $X(t)$ ist ein Kronecker Delta Prozess, wobei jedes Kronecker Delta eine erfolgte Abtastung beschreibt. Der beobachtete Prozess $Z(t)$ ist gegeben als $Z(t) = X(t)Y(t)$, wobei $X(t)$ und $Y(t)$ unabhängig sind. Die Zeit zwischen zwei Abtastwerten ist unabhängig und identisch verteilt und charakterisiert den Abtastprozess. Die Intensität $\mu_{(\cdot)}$ von Prozess $(\cdot)$ ist gegeben durch den Erwartungswert $E[(\cdot)]$. Der Zusam-

menhang zwischen der Autokovarianz $c_Z(\tau)$ des beobachteten Prozesses $Z(t)$ und der unbekannten Autokovarianz $c_Y(\tau)$ des Verkehrsprozesses $Y(t)$ ist gegeben als [Riz13]

$$c_Z(\tau) = (c_X(\tau) + \mu_X^2) c_Y(\tau) + c_X(\tau) \mu_Y^2. \quad (2)$$

Aus (2) ist zu erkennen, dass die Wahl des Abtastprozesses einen Einfluss auf die beobachtete Autokovarianz hat. Dieser Einfluss ist unter bestimmten Bedingung reversibel, so dass die Inferenz von $c_Y(\tau)$ möglich ist.

Beispiele verschiedener Abtastprozesse mit entsprechenden Rekonstruktionsformeln für die Autokovarianz des Verkehrsprozesses $c_Y(\tau)$ sind in [RBF13] gegeben. Abbildung 2 zeigt beispielhaft die erfolgreiche Rekonstruktion der Autokovarianz aus den Beobachtungen verschiedener Abtastprozesse. Das geometrische Abtasten führt zu einer Skalierung der Autokovarianzfunktion des Verkehrs, so dass z.B. der Hurst Parameter H direkt aus der Steigung von $c_Z(\tau)$ in Abb. 2c geschätzt werden kann.

Die Untersuchung des Schätzers der Autokovarianzfunktion für eine endliche Anzahl an Beobachtungen T ergibt, dass der Schätzer asymptotisch erwartungstreu ist. Der Bias des Schätzers nimmt für steigendes T mit $(T - \tau)^{2H-2}$ ab. Abtastprozesse, welche die schwach-mixing Eigenschaft besitzen, liefern asymptotisch erwartungstreue Schätzer für die Autokovarianz des LRD Datenverkehrs.

Eine Implementierung des geometrischen Abtastens wird in kontrollierter Umgebung getestet und später in Internetmessungen eingesetzt [Riz13]. Das Testverfahren basiert auf Ende-zu-Ende Latenzen injizierter Testpakete, um aus der Interaktion mit dem Datenverkehr auf die dominante Verkehrsautokovarianz, d.h. das größte H , entlang des Netzwerkpfades zu schließen. Dabei besteht kein direkter Zugang zum Verkehr. Ein Vergleich der Hurstparameterschätzung sowohl bei direktem Zugang zum Datenverkehr als auch beim Testverfahren ist auf Abb. 3a zu sehen. Hierfür wurde Datenverkehr mit verschiedenem Hurstparameter synthetisiert und in einer kontrollierten Testumgebung FiLab [IL] getestet. Abbildung 3a zeigt die erfolgreiche Schätzung des Hurstparameters sowohl bei direktem Abtasten als auch unter Verwendung des Testverfahrens. Abbildung 3b zeigt den typischen LRD Verlauf der Datenverkehrsautokovarianz bei einer Momentaufnahme auf einem Internetpfad zu einem PlanetLab Rechner an einer US Universität.

4 Warteschlangensystemanalyse mit langzeitkorreliertem Verkehr

In diesem Abschnitt werden obere, nicht-asymptotische Dienstgüteschranken für LRD Ankunftsverkehr an einem Warteschlangensystem hergeleitet. Der LRD Verkehr ist modelliert durch einen fBm Prozess mit einem Hurstparameter $H \in (0.5, 1)$, eine mittlere Rate $\lambda > 0$ und einem Varianzparameter $\sigma^2 > 0$. Es wird ein zeitdiskretes Modell angenommen. Im Folgenden wird eine pfadweise Einhüllende für den fBm Verkehr hergeleitet. Man betrachte die punktweise Einhüllende $E(t) = \lambda t + \sqrt{-2 \log \varepsilon_p(t)} \sigma t^H$ mit zeitabhängiger Verletzungswahrscheinlichkeit, die durch die Einführung der freien Parameter $\beta \in (0, 1 - H)$ and $\eta \in (0, 1)$ als $\varepsilon_p(t) = \eta^{t^{2\beta}}$ ausgedrückt werden kann.

Theorem 4.1 (FBM Pfadweise Einhüllende). *Gegeben sei fBm Verkehr charakterisiert durch das Tripel λ, σ^2 , und $H \in (0.5, 1)$. Eine pfadweise Einhüllende ist gegeben als*

$$E(t) = \lambda t + \sqrt{-2 \log \eta} \sigma t^{H+\beta}$$

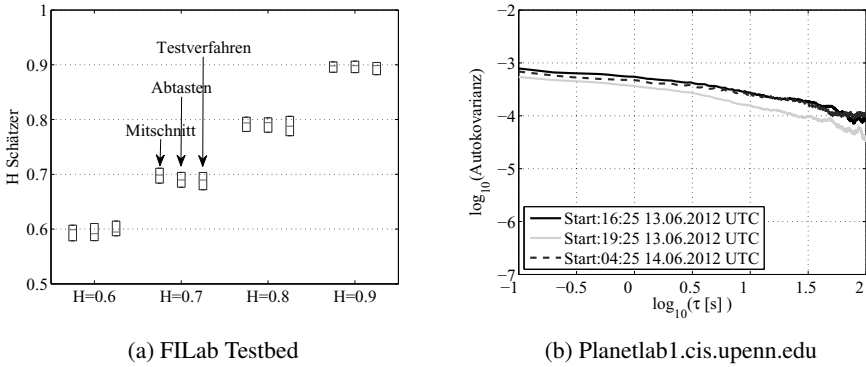


Abbildung 3: (a) Hurstparameterschätzungen bei direktem Zugang zum Datenverkehr als auch über das Testverfahren. (b) Beispiel einer Autokovarianzschätzung aus einer Ende-zu-Ende Internetmessung.

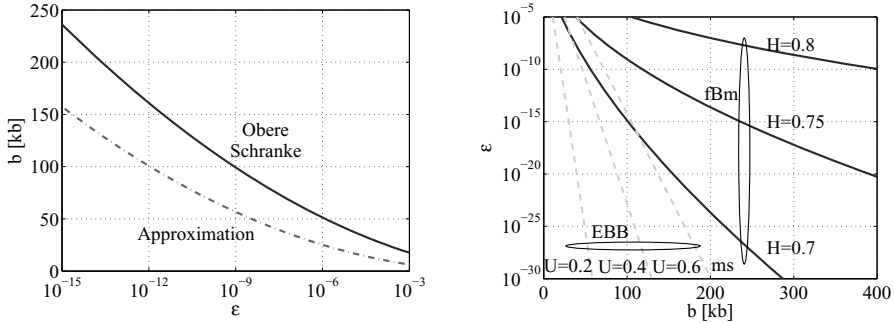
mit der Verletzungswahrscheinlichkeit $\varepsilon_s = \frac{\Gamma(\frac{1}{2\beta})}{2\beta(-\log \eta)^{\frac{1}{2\beta}}}$ und den freien Parameter $\beta \in (0, 1 - H)$ und $\eta \in (0, 1)$.

Der Beweis von Theorem 4.1 ist gegeben in [Riz13]. Eine intuitive Skizze des Beweises ist wie folgt: Es werden Chernoffs Theorem und Booles Ungleichung benutzt. Die punktweise Einhüllende wird durch den Parameter β relaxiert, so dass die punktweise Verletzungswahrscheinlichkeit $\varepsilon_p(t)$ mit t abklingt. Ein Lemma in [Riz13] liefert die obere Schranke für die Verletzungswahrscheinlichkeit. Die Grenzen der freien Parameter erlauben die Herleitung von Leistungsschranken für Pufferfüllstand oder Latenz. Eine Parametrisierung in geschlossener Form ist gegeben in [Riz13].

Eine obere Schranke für den Pufferfüllstand an einem Bediensystem mit konstanter Kapazität C kann gefunden werden als $P[B > b] \leq \varepsilon_s$ mit ε_s und β aus Theorem 4.1. Die entsprechende Parametrisierung lautet $\eta = \exp\left(-\frac{1}{2\sigma^2} \left(\frac{C-\lambda}{H+\beta}\right)^{2(H+\beta)} \left(\frac{b}{1-(H+\beta)}\right)^{2-2(H+\beta)}\right)$.

Entsprechend ist die Latenz beschränkt durch $P[W > b/C] \leq \varepsilon_s$. Abbildung 4a vergleicht die Pufferfüllstandsschranke zur Approximation (1) für ein Beispielszenario mit den Parameter $C = 1$ Gbps, $\lambda = 0.5$ Gbps, $\sigma = \lambda/2$, und $H = 0.75$ und eine Zeitschlitzlänge von $10 \mu\text{s}$. Beide Verletzungswahrscheinlichkeiten besitzen einen Weibull-Tail. In [Riz13] wird bewiesen, dass der Tail-Abfall beider Formulierung log-asymptotisch gleich ist.

Die Formulierung der Pufferfüllstandsschranke nach der Optimierung von β liefert den Ausdruck $\tilde{\varepsilon}_s = \frac{b}{C-\lambda} \varepsilon_a^{1+\log \chi} \sqrt{2\pi(-\log \varepsilon_a)}$ mit $\chi \in (\frac{1}{4}, 1)$. Diese rigorose, obere Schranke lässt ähnlich wie Approximation (1) aus [Nor94] folgendes ableiten: Die mittlere verfügbare Bandbreite $C - \lambda$ hat einen stärkeren Einfluss auf die Verletzungswahrscheinlichkeit als die Puffergröße b bei LRD Verkehr. Für $H = 0.5$, d.h. EBB Verkehr, ist der Einfluss beider Faktoren gleich. Dieses Resultat untermauert die Bedeutung von verfügbarer Bandbreite und unterstützt moderne Aufforderungen zu kleinen Routerpuffer. Abbildung 4b vergleicht den Tail-Abfall für EBB und LRD Verkehr.



(a) Pufferfüllstandsschranke vs. Approximation. (b) Tail-Verhalten der Pufferfüllstandsschranken.

Abbildung 4: (a) Numerisches Beispiel für die Pufferfüllstandsschranke. (b) Ineffizienz von großen Puffer bei LRD Datenverkehr: Der Weibull-Tail entspricht $\log \varepsilon \simeq -K_1 b^{2-2H}$ im Vergleich zu $\log \varepsilon \simeq -K_2 b$ für den EBB Verkehr mit den Konstanten K_1, K_2 . Die Änderung der Bursthaftigkeit U des Markov EBB Verkehrs ändert die Steigung.

Für Einzelsysteme mit LRD Querverkehr, wie in Abb. 1, kann eine verfügbare Dienstkurve für den durchgehenden Verkehr als $S(t) = (C - r)t$ gefunden werden, wobei der LRD Querverkehr eine affine Einhüllende $E(t) = rt$ besitzt. Die affine Einhüllende kann geometrisch bestimmt werden als obere Schranke für die Einhüllende in Theorem 4.1 [Riz13]. Die Verletzungswahrscheinlichkeit der Dienstkurve entspricht $\varepsilon_s(b)$ der affinen Einhüllenden. Mit Hilfe dieser Dienstkurve kann im nächsten Abschnitt die Analyse auf gesamte Pfade mit LRD Querverkehr ausgeweitet werden.

5 Netzwerkpfadanalyse mit langzeitkorreliertem Datenverkehr

Im Folgenden werden obere Dienstgüteschranken für Netzwerkpfade mit LRD Verkehr wie in Abb. 1 vorgestellt. Hierfür wird eine Netzwerkdienstkurve durch die Faltung der Dienstkurven der einzelnen Systeme hergeleitet. Die Netzwerkdienstkurve erlaubt die Herleitung von oberen Ende-zu-Ende Leistungsschranken z.B. für die Latenz.

Zur Bestimmung der Netzwerkdienstkurve wird eine pfadweise verfügbare Dienstkurve aufbauend auf Abschnitt 4 hergeleitet. Für ein System mit Kapazität C und LRD Querverkehr mit der affinen Einhüllenden $E(t) = rt$ ist der Ausdruck $S(t) = (C - r)t$ eine gültige, pfadweise verfügbare Dienstkurve mit Verletzungswahrscheinlichkeit $\varepsilon^\ell(b) = \frac{\Gamma(\frac{1}{2\beta})}{2\varrho^\vartheta \frac{1}{2\beta} (1-(H+2\beta))} b^{-\frac{1-(H+2\beta)}{\beta}}$ mit den freien Parameter $\varrho \in (0, C - r)$ und $\beta \in (0, \frac{1-H}{2})$ und mit $\vartheta = \frac{1}{2\sigma^2} ((r - \lambda)/(H + \beta))^{2(H+\beta)} (1/(1 - (H + \beta)))^{2-2(H+\beta)}$.

Im Folgenden werden n homogene Systeme angenommen. Weitere Szenarien sind in [Riz13] beschrieben. Der durchgehende Verkehr besitzt eine affine Einhüllende mit der Rate r^{th} . Jedes System besitzt eine Kapazität C und der LRD Querverkehr wird durch das Tripel λ, σ^2 und $H^{cr} \in (0.5, 1)$ beschrieben. Die Abkürzungen th und cr werden für den durchgehenden bzw. für den Querverkehr benutzt. An jedem System $i \in [1, n]$ ist eine verfügbare Dienstkurve $S_i(t) = (C - r^{cr})t$ zu finden, wobei r^{cr} die Rate der affi-

nen Einhüllenden des Querverkehrs wiedergibt. Eine Netzwerkdienstkurve im Sinne von Abschnitt 1 ist gegeben als $S^{net}(t) = (C - r^{cr} - \Delta)t$ mit konstantem $\Delta := (n - 1)\varrho$ und mit $\Delta \in (0, C - r^{cr})$ für die Stabilität. Wie in Abschnitt 1 beschrieben, ergibt sich die Verletzungswahrscheinlichkeit für $S^{net}(t)$ aus der Min-Plus Faltung als $\varepsilon^{net}(b) = \frac{\Gamma(\frac{1}{2\beta})}{2\vartheta^{\frac{1}{2\beta}}} \inf_x \left\{ \frac{(n-1)^2 \left(\frac{b-x}{n-1}\right)^{-\frac{1-(H+2\beta)}{\beta}}}{\Delta(1-(H+2\beta))} + \frac{x^{-\frac{1-(H+2\beta)}{\beta}}}{\beta} \right\}$, mit $x \in (0, b)$ und $\beta \in (0, \frac{1-H}{2})$.

Die Netzwerkdienstkurve erlaubt die Herleitung von Ende-zu-Ende Leistungsschranken wie in Abschnitt 1 dargestellt. Zum Beispiel ist eine obere Latenzschranke für durchgehenden Verkehr mit konstanter Rate r^{th} gegeben als $P[W > b/(C - r^{cr} - \Delta)] \leq \varepsilon^{net}(b)$ mit der Stabilitätsbedingung $r^{th} + r^{cr} + \Delta \leq C$. Eine direkte Erweiterung für durchgehenden Verkehr mit der affinen Einhüllenden mit der Rate r^{th} und der Verletzungswahrscheinlichkeit ε^{th} ist gegeben als $P[W > b/(C - r^{cr} - \Delta)] \leq \varepsilon^{net} \otimes \varepsilon^{th}(b)$.

Abbildung 5a zeigt ein numerisches Beispiel für Ende-zu-Ende Latenzschranken für durchgehenden Verkehr mit konstanter Rate von 250 Mbps und LRD Querverkehr mit $\lambda = 250$ Mbps, $\sigma = \lambda/2$ und unterschiedlichen Hurstparameter. Die Kapazität C beträgt 1 Gbps und die Verletzungswahrscheinlichkeit ist festgelegt auf $\varepsilon = 10^{-12}$. Die freien Parameter der Einhüllenden werden numerisch optimiert. Man erkennt, dass die End-zu-Ende Latenzschranken superlinear in Abhängigkeit von der Pfadlänge n wachsen.

Das nachfolgende Theorem beschreibt den Zuwachs von Ende-zu-Ende Latenzen für Netzwerkpfade mit n Systemen in Folge und LRD Querverkehr.

Theorem 5.1 (Zuwachs von Ende-zu-Ende Dienstgüteschranken). *Gegeben seien n homogene Systeme jeweils mit LRD Querverkehr mit Hurstparameter $H \in (0.5, 1)$. Ende-zu-Ende Dienstgüteschranken für den durchgehenden Verkehr skalieren für gleichbleibende Verletzungswahrscheinlichkeit mit $\mathcal{O}\left(n(\log n)^{\frac{1}{2-2H}}\right)$.*

Der Beweis zu Theorem 5.1 ist gegeben in [Riz13]. Folgend ist eine Skizze des Beweises: Der Zuwachs ist durch den Tail-Abfall der Verletzungswahrscheinlichkeiten der einzelnen, verfügbaren Dienstkurven des Pfades charakterisiert. Diese werden für die Netzwerkdienstkurve mittels der Min-Plus Faltung verknüpft. Der Einfluss des Hurstparameters H des Querverkehrs ist moderat aufgrund der polylogarithmischen Komponente in Theorem 5.1. Für $H = 0.5$ gibt Theorem 5.1 das aus der Literatur bekannte Zuwachsresultat für EBB Verkehr wieder [CBL06]. Abbildung 5b zeigt die Bedeutung der mittleren verfügbaren Bandbreite für Ende-zu-Ende Latenzschranken im Zusammenhang mit LRD Querverkehr. Für eine festgelegte Verletzungswahrscheinlichkeit führt das Halbieren der verfügbaren Bandbreite zur Verdoppelung der Latenzschranke w für $H = 0.5$. Dieser Zusammenhang verschlechtert sich wesentlich für $H \in (0.5, 1)$. Dieses Beispiel untermauert die Bedeutung von Überkapazitäten in Rechnernetzen.

6 Schlüsselbeiträge

Im Rahmen dieser Arbeit wurde eine abtastbasierte Schätzmethode für die Langzeitkorrelationen des Datenverkehrs vorgestellt. Die Rückgewinnung der statistischen Eigenschaften des Verkehrs aus den zufälligen Beobachtungen wurde analytisch bewiesen. Der Einfluss einer begrenzten Anzahl an Beobachtungen auf die Schätzgenauigkeit wurde quan-

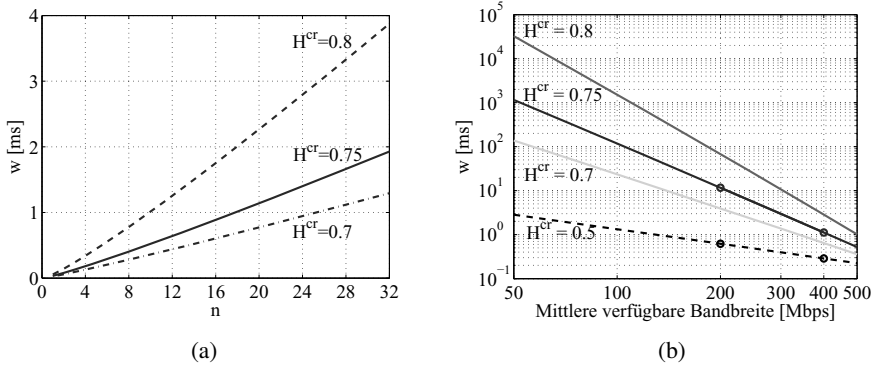


Abbildung 5: (a) Ende-zu-Ende Latenz für LRD Querverkehr wächst superlinear mit der Pfadlänge n an. (b) Der starke Einfluss der mittleren, verfügbaren Bandbreite auf die Dienstgüte im Zusammenhang mit LRD Querverkehr.

tifiziert. Weiterhin wurden asymptotisch erwartungstreue Schätzer für die Langzeitkorrelationen des Verkehrs vorgestellt. Diese Arbeit ergänzt traditionelle Schätzmethode aus der Literatur durch ein Verfahren, welches Testpakete in das Netzwerk injiziert, auf zufälligem Abtasten aufbaut und keine administrative Berechtigung an Router erfordert. Im Gegensatz zu traditionellen Methoden, welche die statistischen Eigenschaften an einem einzelnen Router wiedergeben, ermöglicht das vorgestellte Verfahren die dominanten Eigenschaften entlang eines untersuchten Pfades zu schätzen. Das präsentierte Verfahren wird in einer kontrollierten Umgebung verifiziert und es werden beispielhafte Ergebnisse einer Langzeit-Internetstudie vorgestellt.

Im Rahmen des stochastischen Netzwerkalküls wurden Leistungsschranken für Ende-zu-Ende Latenzen in Netzwerken mit langzeitkorreliertem Datenverkehr hergeleitet. Ein grundlegendes Resultat dieser Arbeit bezieht sich auf den Zuwachs von Ende-zu-Ende Latenzen für Netzwerkpfade mit n Systemen in Folge auf $\mathcal{O}\left(n(\log n)^{\frac{1}{2-2H}}\right)$. Vergleichbare Resultate aus der Literatur sind z.B. $\Theta(n)$ unter statistischer Unabhängigkeit und $\Theta(n \log n)$ für EBB Verkehr. Die Ergebnisse dieser Arbeit quantifizieren den Einfluss der Langzeitkorrelationen auf die Netzwerkperformanz. Die in geschlossener Form hergeleiteten Resultate haben eine grundlegende Auswirkung auf die Dimensionierung und den Betrieb von Kommunikationsnetzwerken.

Literatur

- [Ber94] J. Beran. *Statistics for Long-Memory Processes*. Chapman & Hall/CRC, Oktober 1994.
- [BLP06] A. Burchard, J. Liebeherr und S. D. Patek. A Min-Plus Calculus for End-to-End Statistical Service Guarantees. *IEEE Trans. Inform. Theory*, 52(9):4105–4114, September 2006.
- [BMVB09] F. Baccelli, S. Machiraju, D. Veitch und J. Bolot. The Role of PASTA in Network Measurement. *IEEE/ACM Trans. Networking*, 17(4):1340–1353, August 2009.

- [CBL06] F. Ciucu, A. Burchard und J. Liebeherr. Scaling Properties of Statistical End-to-end Bounds in the Network Calculus. *IEEE/ACM Trans. Networking*, 14(6):2300–2312, Juni 2006.
- [Cha00] C.-S. Chang. *Performance Guarantees in Communication Networks*. Springer, 2000.
- [DO95] N. G. Duffield und N. O’Connell. Large Deviations and Overflow Probabilities for the General Single-Server Queue, with Applications. *Math. Proc. Camb. Phil. Soc.*, 118(2):363–375, September 1995.
- [FGHW99] A. Feldmann, A. C. Gilbert, P. Huang und W. Willinger. Dynamics of IP traffic: A Study of the Role of Variability and the Impact of Control. *SIGCOMM Comput. Commun. Rev.*, 29(4):301–313, August 1999.
- [FMN00] N. Fonseca, G. Mayor und C. Neto. On the Equivalent Bandwidth of Self-similar Sources. *ACM Trans. Model. Comput. Simul.*, 10(2):104–124, April 2000.
- [GMR09] H. Gupta, A. Mahanti und V. Ribeiro. Revisiting Coexistence of Poissonity and Self-similarity in Internet Traffic. In *Proc. of IEEE/ACM MASCOTS*, September 2009.
- [IL] Future Internet Lab. <http://www.filab.uni-hannover.de/>.
- [LGD⁺10] P. Loiseau, P. Goncalves, G. Dewaele, P. Borgnat, P. Abry und P. Primet. Investigating Self-Similarity and Heavy-Tailed Distributions on a Large-Scale Experimental Facility. *IEEE/ACM Trans. Networking*, 18(4):1261–1274, August 2010.
- [LTWW94] W. Leland, M. Taqqu, W. Willinger und D. Wilson. On the Self-similar Nature of Ethernet Traffic (Extended Version). *IEEE/ACM Trans. Networking*, 2(1):1–15, Februar 1994.
- [Nor94] I. Norros. A Storage Model with Self-similar Input. *Queueing Systems*, 16(3):387–396, September 1994.
- [RBF13] A. Rizk, Z. Bozakov und M. Fidler. Estimating Traffic Correlations from Sampling and Active Network Probing. In *Proc. of IFIP Networking*, Mai 2013.
- [RF12] A. Rizk und M. Fidler. Non-asymptotic End-to-end Performance Bounds for Networks with Long Range Dependent FBM Cross Traffic. *Computer Networks*, 56(1):127–141, 2012.
- [Riz13] Amr Rizk. *Non-asymptotic Performance Evaluation and Sampling-based Parameter Estimation for Communication Networks with Long Memory Traffic*. Dissertation, Leibniz Universität Hannover, 2013.
- [Wol82] R. Wolff. Poisson Arrivals See Time Averages. *Operations Research*, 30(2):223–231, März 1982.
- [YS93] O. Yaron und M. Sidi. Performance and Stability of Communication Networks via Robust Exponential Bounds. *IEEE/ACM Trans. Networking*, 1(3):372–385, Juni 1993.



Amr Rizk schloss im Jahr 2008 das Wirtschaftsingenieur-Studium an der Technischen Universität Darmstadt ab. Im Jahr 2013 wurde er an der Fakultät für Elektrotechnik und Informatik der Leibniz Universität Hannover zum Dr.-Ing. (mit Auszeichnung) promoviert. Seine Forschungsschwerpunkte liegen in den Bereichen der Leistungsbewertung von Rechnernetzen, der stochastischen Modellierung von Kommunikationssystemen und der Nachrichtenverkehrstheorie.