Martin Zaddach
Fachbereich Informatik - TKRN, UniHH

Lastcharakterisierung und Lastmodellierung von verteilten Videoanwendungen als Basis für ein adaptives Management in IP-basierten Rechnernetzen

Martin Zaddach

INHALT

Warum Lastmodellierung von Videoströmen ?

- Videoapplikationen (VA) stellen hohe Forderungen an Übertragungsresourcen
- VA weisen starke Schwankungen in ihren Forderungen auf
- In IP-basierten Netzen besteht Abhängigkeit zwischen Verzögerungs-/Verlustverhalten und Struktur des Paketstromes
- Erzeugung von TCP-*friendly* Verkehrslasten
- Spezifikation von Filtern (z.B. Token-Bucket) benötigen Kenntnisse der Verkehrslasten
- Nutzung der Parametrisierungsmöglichkeiten der Videoapplikation zur Kontrolle des Paketstromes

Spezifikation der Last

- Definition eindeutiger Schnittstellen, z.B.
  - Anwendungsnahe Schnittstelle (Primärlastschnittstelle, IF_P) durch Dienstübergabepunkt des Transportdienstes (RTP, UDP,...)
  - Netznahe Schnittstelle (Sekundärlastschnittstelle, IF_S) unterhalb von IP
- Charakterisierung des Ankunftsprozesses der Aufträge an IF
- Modellierung der Betriebsmittelforderungen und anderer relevanter Auftragsattribute

In IP-Netzen eignet sich zur Spezifikation der Aufträge der Ankunftszeitpunkt sowie ihre Größe, daher...

- Ankunftsprozeß: deterministisches, isochrones Zeitraster
- Eingebettete *Transform Expand Sample*-Prozesse oder Autoregressive Prozesse zur Modellierung der Auftrags- bzw. Framegrößen

Lastmessungen Längenapproximationen

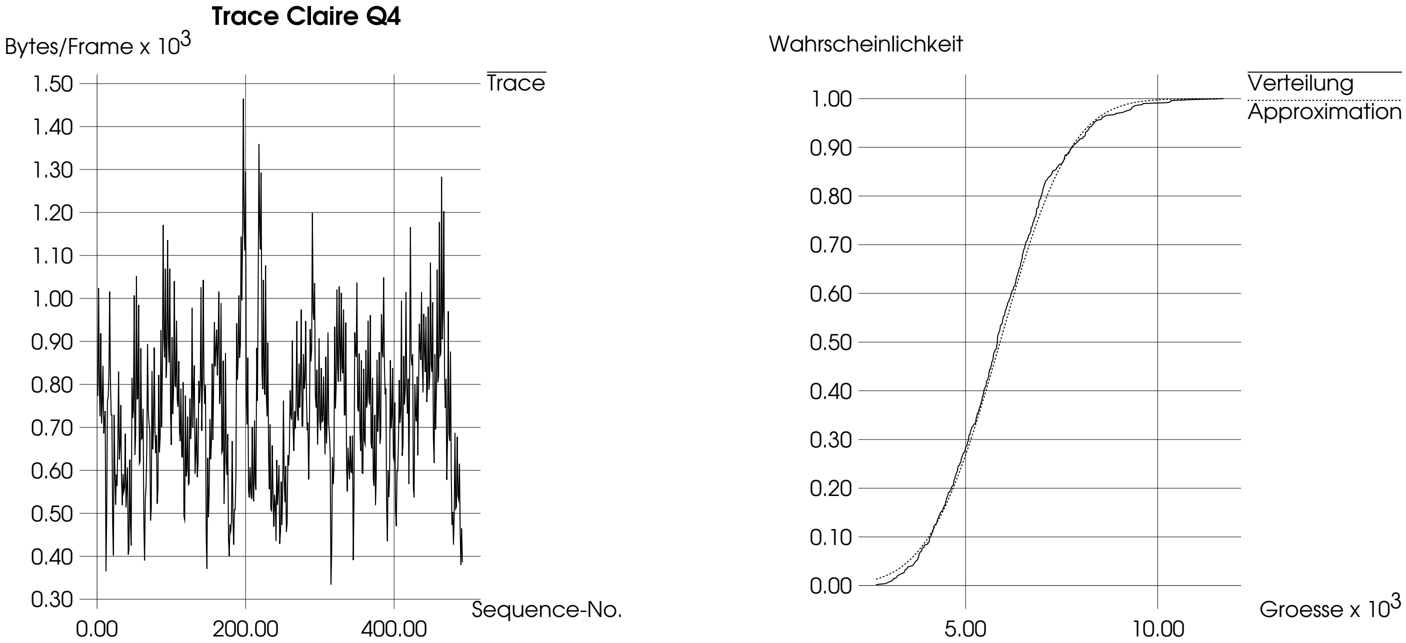


Abbildung 1: H.261-Trace Claire Q4 (links) und Framelängenverteilung (rechts)

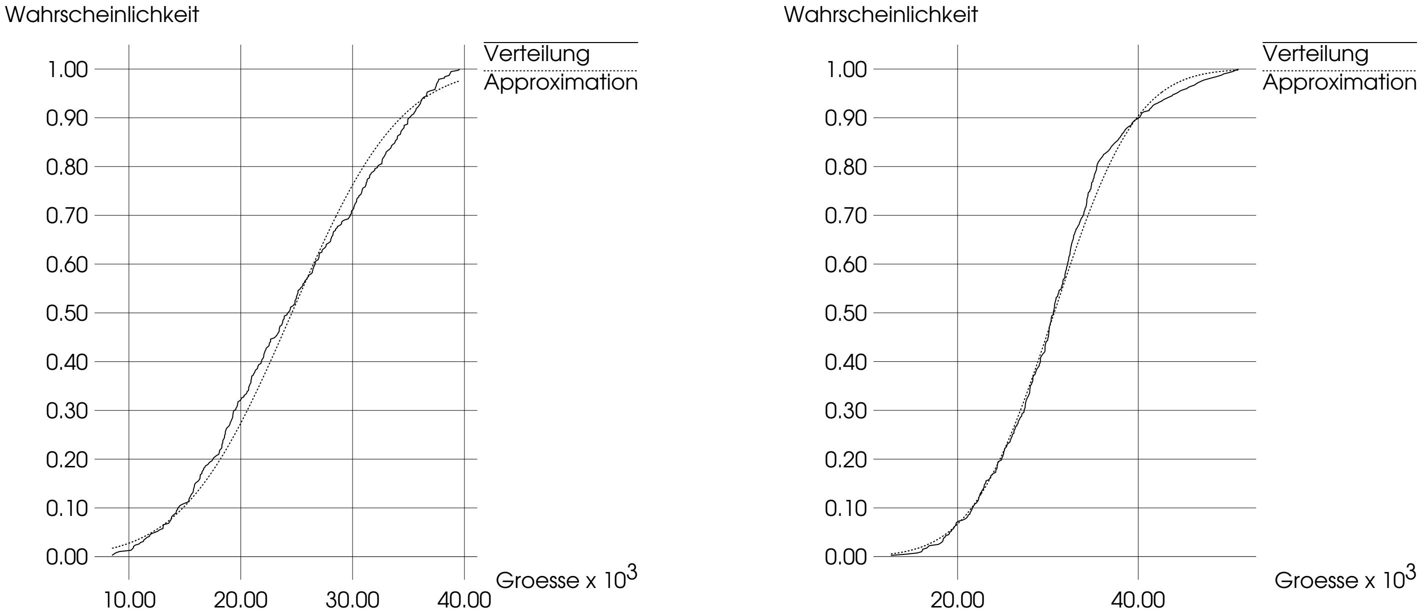


Abbildung 2: Framelängenverteilung Carphone (links) und Foreman (rechts)

Güte der Approximationen

Test auf Normalverteilung - χ^2-Test

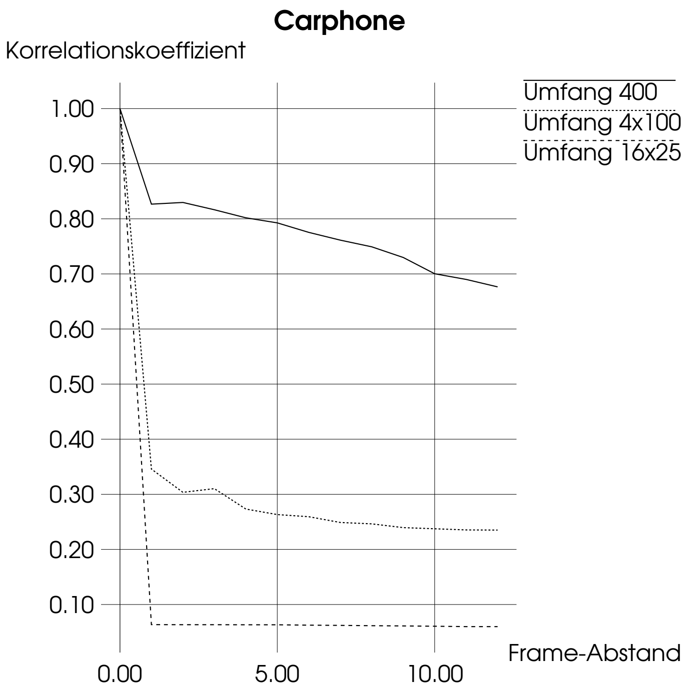
- 10 Freiheitsgrade bei zwei geschätzten Parametern (μ, σ)
- gewähltes Signifikanzniveau $\alpha = 1\%$
- Signifikanzgröße $\chi^2_{0.01,10} = 23.209$

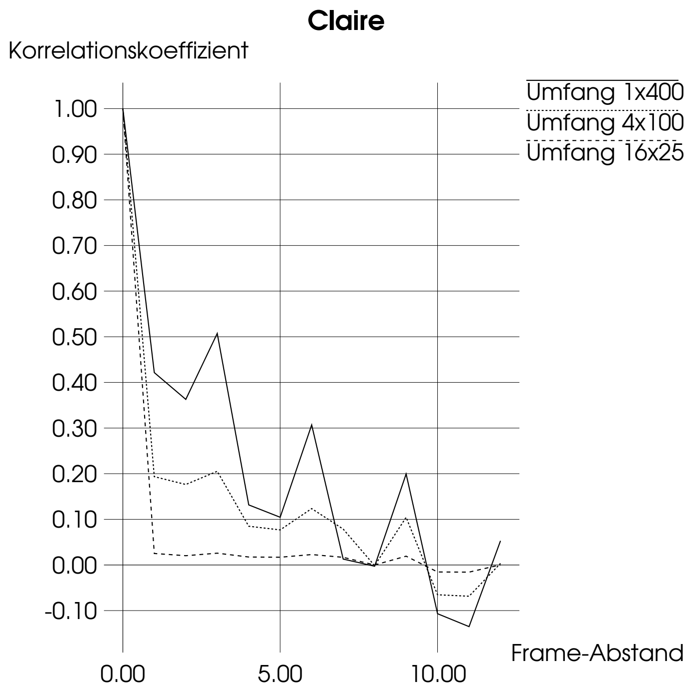
Resultate des χ^2-Tests

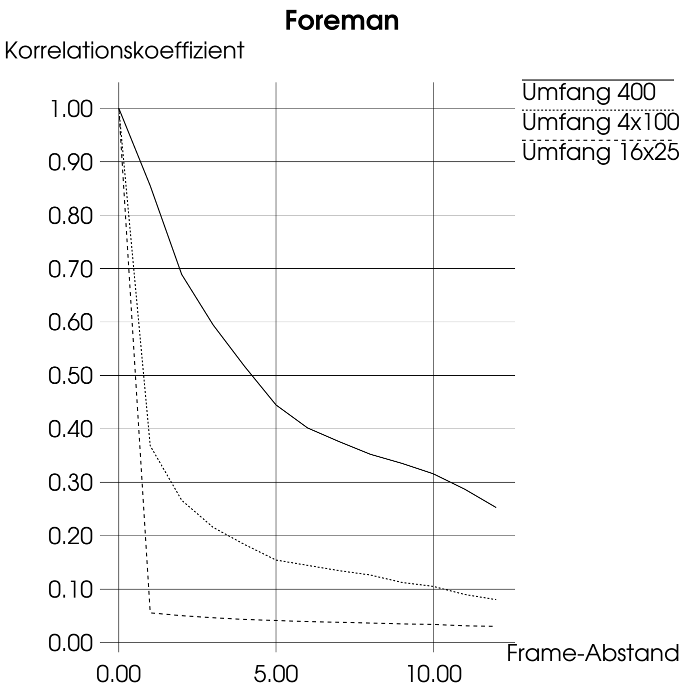
Standard	Quantisierung	Videosequenz		
		Claire	Foreman	Carphone
H.261	Q1	4.2502	13.876	12.9953
	Q4	6.5248	14.0763	13.4218
	Q10	9.7671	13.1631	14.0233
H.263	Q1	4.2887	16.876	14.1479
	Q4	6.5352	18.1099	14.2638
	Q10	9.9816	14.4627	14.5210

Tabelle 1: Ergebnisse der χ^2-Tests auf Normalverteilung

Korrelationsstruktur und Stichprobenumfang







Anmerkungen zur Korrelationsstruktur

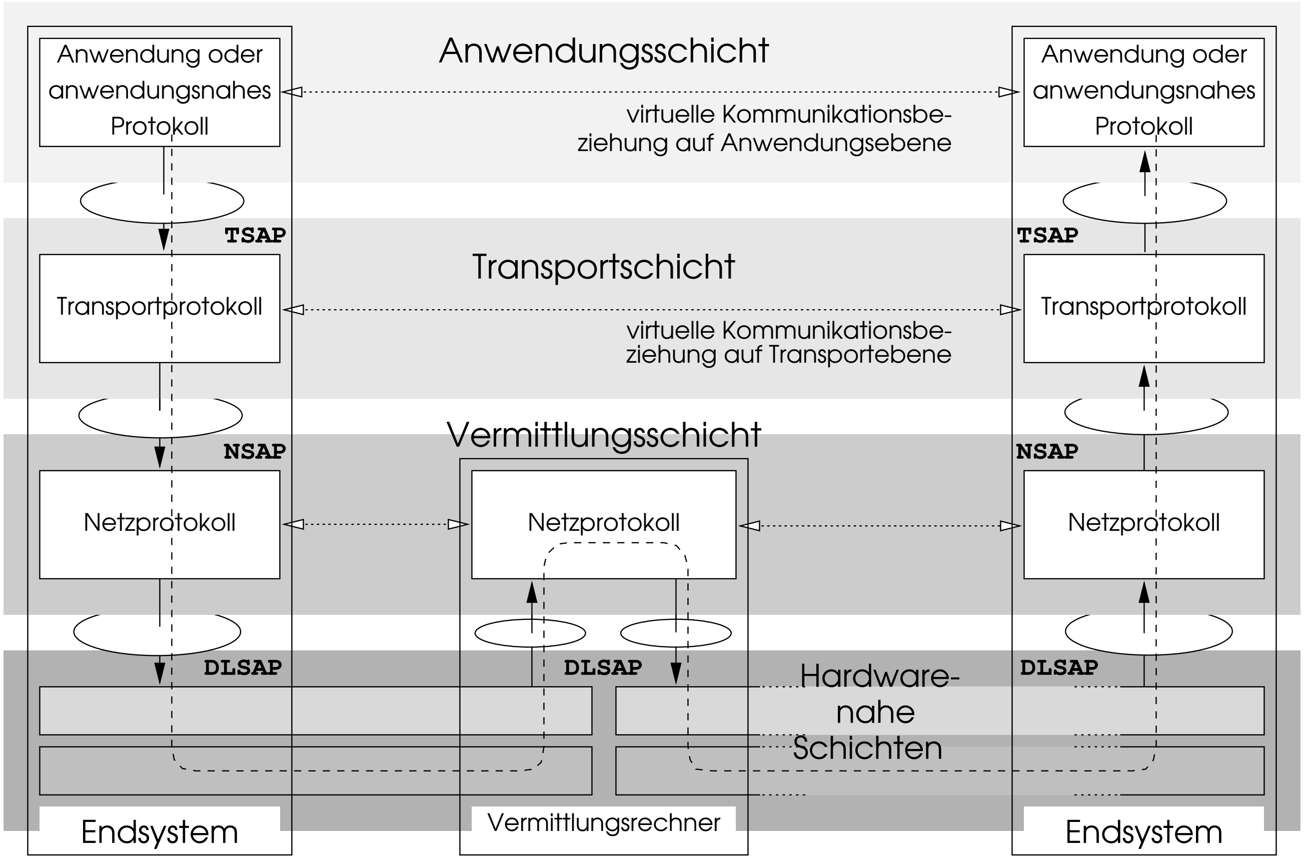
Ergebnisse der Untersuchungen:

- Korrelationsstruktur wird durch Langzeitkorrelationen aufgebaut
- Messungen geringen Stichprobenumfangs weisen weitgehende stochastische Unabhängigkeit auf
- Kurzzeitige Schwankungen sind rein zufällig

Anwendungsmöglichkeiten:

- *Allgemeine* Nutzbarkeit:
  - einfache Parametrisierbarkeit der Modelle zur Laufzeit (Ringpuffer)
- *Spezielle* Nutzbarkeit:
  - Ratenkontrollalgorithmen
  - Anpassung an Zielrate, Filter, Netzverhalten,...

Protokollarchitektur

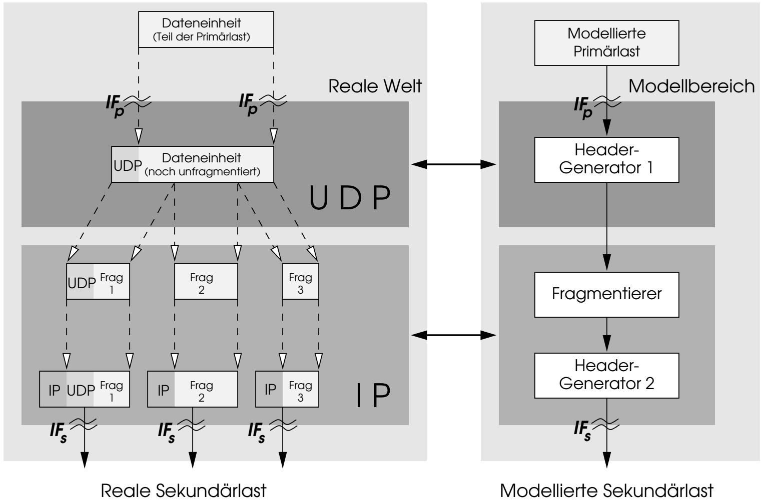


Stochastische Lasttransformationen

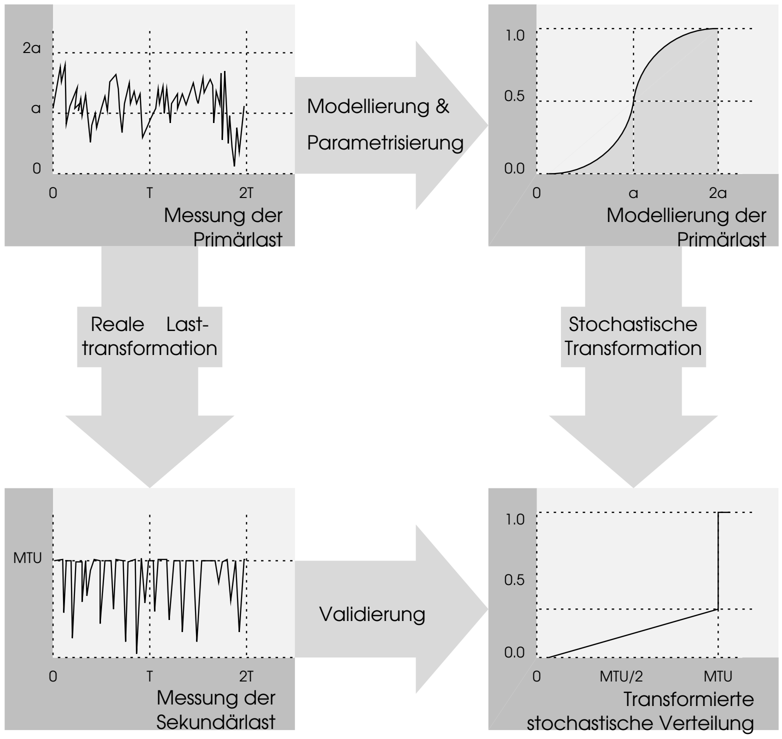
- Netznahe Sekundärlasten sind für Verzögerungs- und Verlustverhalten, sowie für Filterspezifikationen relevant
- Eine mögliche Lösung: Stochastische Transformation der anwendungsnahen Lasten in netznahe Lasten
- Nachbildung von realen Transformationsmechanismen durch mathematische Transformationen auf Attributmengen
- Gewinnung von Kenngrößen und Randverteilungen der netznahen Sekundärlasten

Martin Zaddach
Fachbereich Informatik - TKRN, UniHH

Ein Ansatz zur Transformation von Primär- in Sekundärlasten



Analytische Modellierung von Lasttransformationen



Analytische Lasttransformation

Gegeben:
$\mathcal{L}(x)$ = (Wahrscheinlichkeits-)Verteilung der Dateneinheitenlänge, z.B. zu übertragende Videoframes

- Transformation 1:

  Generierung des UDP-Headers der Länge $\upsilon = 8$ Byte
  neue Verteilung $\mathcal{U}(x) = \int_{-\infty}^{x} d\mathcal{L}(s-\upsilon) = \mathcal{L}(x-\upsilon)$

- Transformation 2: IP-Fragmentierung gemäß gegebener MTU-Länge

  Es bezeichne

  - Θ die MTU-Länge (in Byte)
  - ψ die IP-Headerlänge
  - ϑ die max. Länge der Nutzdaten im IP-Paket, d.h. $\vartheta = \Theta - \psi$
  - $\beta(n)$ die Wahrscheinlichkeit, daß die UDP-Dateneinheit in n Segmente fragmentiert wird, mit dem Erwartungswert β

  $$\rightarrow \quad \beta(n) = \int_{(n-1)\,\vartheta}^{n\,\vartheta} d\mathcal{U}(x) = \mathcal{U}(n\,\vartheta) - \mathcal{U}((n-1)\,\vartheta)$$

  und somit

  $$\beta = \sum_{i=1}^{\infty} i\,\beta(i) = \sum_{i=1}^{\infty} i\,\big(\mathcal{U}(i\,\vartheta) - \mathcal{U}((i-1)\,\vartheta)\big) \approx$$

  für ein hinreichend großes l

  $$\approx \sum_{i=1}^{l} i\,\big(\mathcal{U}(i\vartheta) - \mathcal{U}((i-1)\vartheta)\big) = \sum_{i=1}^{l} i\,\mathcal{U}(i\vartheta) - \sum_{i=0}^{l-1}(i+1)\,\mathcal{U}(i\vartheta) =$$
  $$= l\,\mathcal{U}(l\vartheta) - \sum_{i=0}^{l-1}\mathcal{U}(i\vartheta) \approx \sum_{i=0}^{l-1} 1 - \mathcal{U}(i\vartheta)$$

$\rightarrow$ neue Verteilung für die Länge der Fragmente

$$\mathcal{F}(x) = \begin{cases} 0, & x \leq 0, \\ \frac{1}{\beta} \sum_{i=0}^{l-1} \mathcal{L}(x + i\,\vartheta - \upsilon) - \mathcal{L}(i\,\vartheta - \upsilon), & 0 < x < \vartheta, \\ 1, & x \geq \vartheta. \end{cases}$$

- Transformation 3:

  Generierung des IP-Headers der Länge $\psi = 20$ Byte bei IPv4 bzw. $\psi = 40$ Byte bei IPv6
  $\rightarrow$ neue Verteilung für Länge der IP-Pakete $\mathcal{I}(x) = \mathcal{F}(x - \psi)$, ergo

$$\mathcal{I}(x) = \begin{cases} 0, & x \leq \psi, \\ \frac{1}{\beta} \sum_{i=0}^{l-1} \mathcal{L}(x + i\,\vartheta - \upsilon - \psi) - \mathcal{L}(i\,\vartheta - \upsilon - \psi), & \psi < x < \vartheta + \psi, \\ 1, & x \geq \vartheta + \psi. \end{cases}$$

Anwendung der analytischen Lasttransformation: Elementare Primärlast

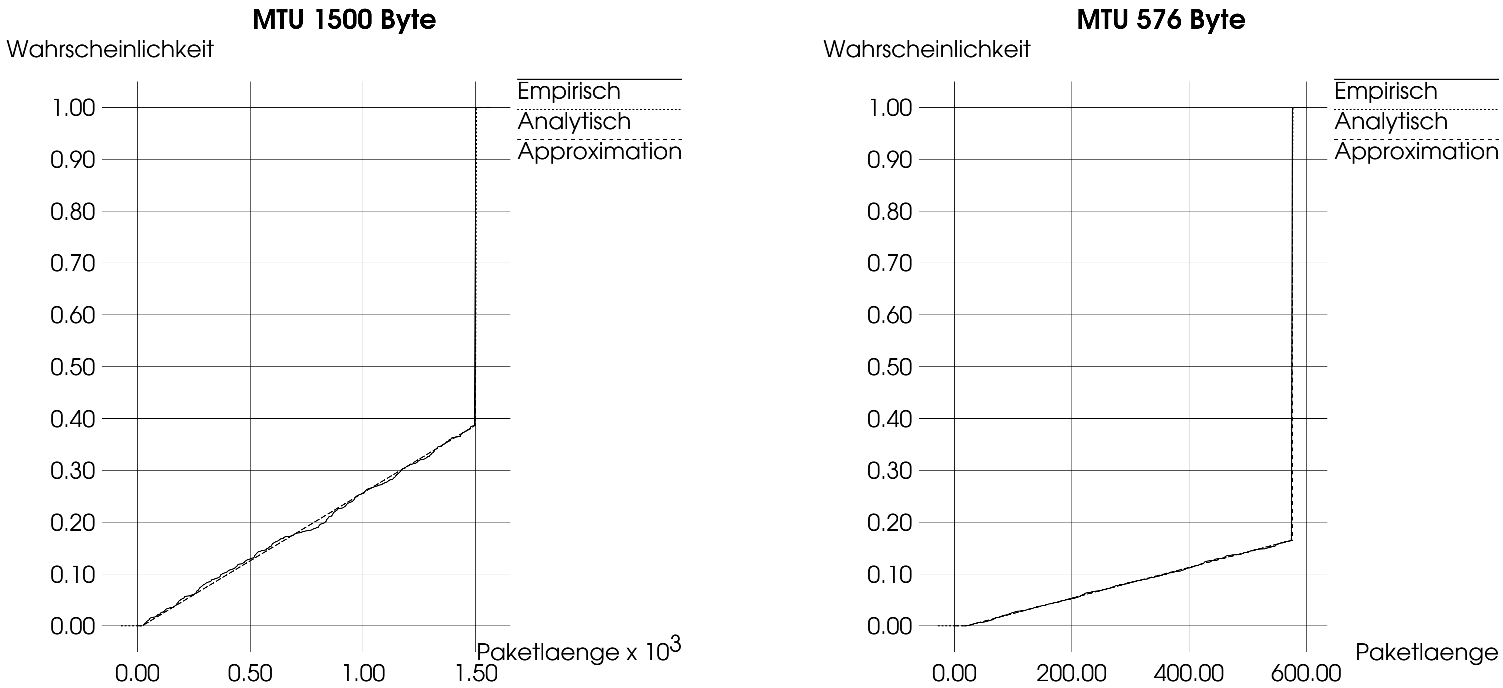
Voraussetzungen:

- nur eine Einzelquelle ($n = 1$), H.261-codiert, Quantisierungsniveau 4 (Q4)
- beobachtete Sequenzen Carphone und Foreman
- MTU-Größen 576 Byte und 1500 Byte

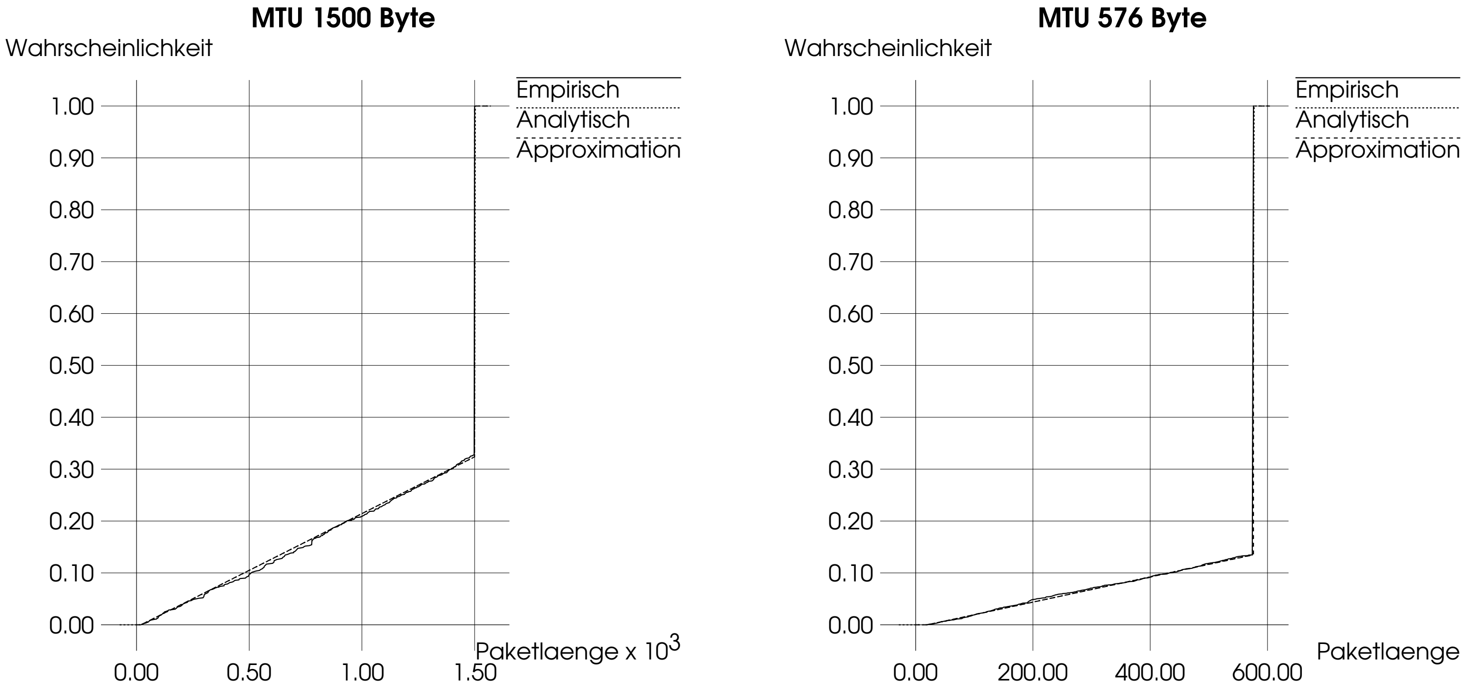
Validationsresultate

- empirisch, d.h. gemessene Sekundärlast (SL)
- analytisch, d.h. mittels analytischer Lasttransformation prognostizierte SL
- Approximation 1. Ordnung, d.h. lineare Approximation der prognostizierten SL mittels der Stützpunkte $(\psi, 0.0)$, $(\Theta, \frac{1}{\beta})$, $(\Theta, 1.0)$.

a) Carphone



b) Foreman

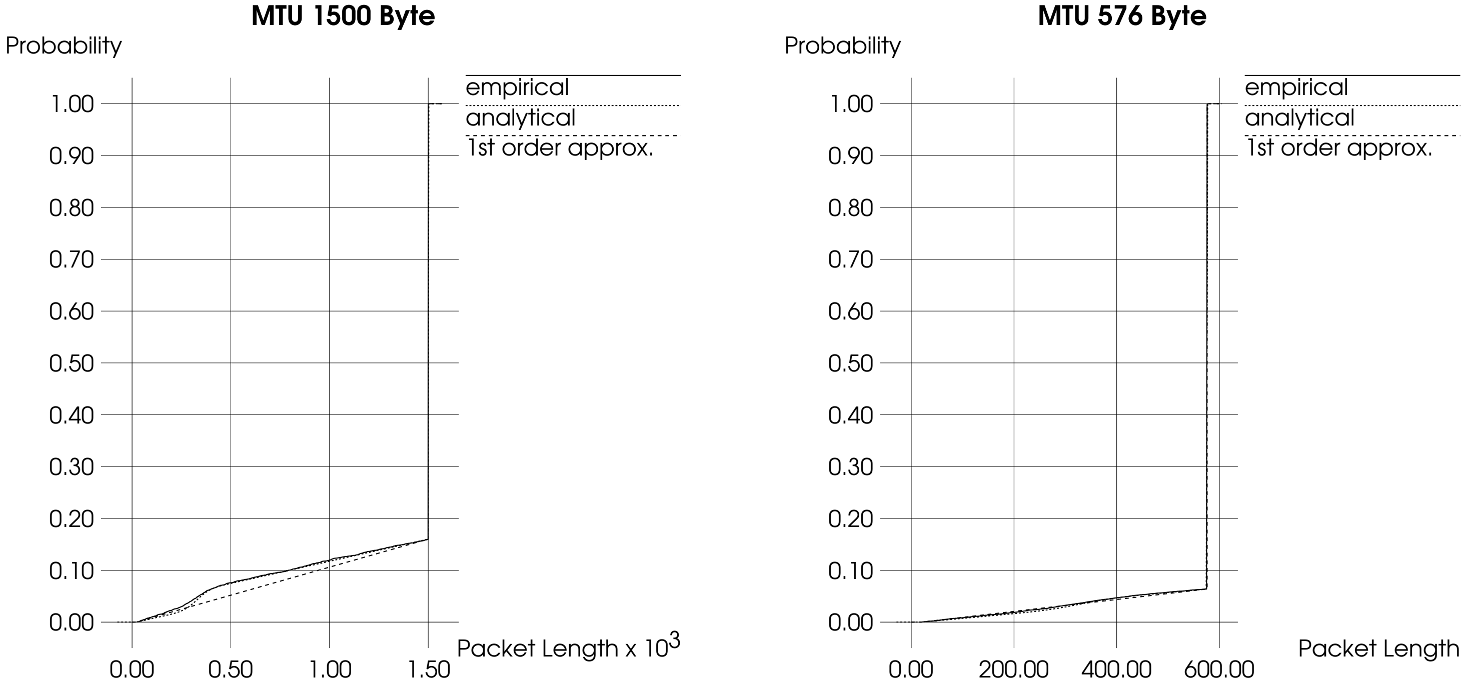


Anwendung der analytischen Lasttransformation: Komplexe Primärlast

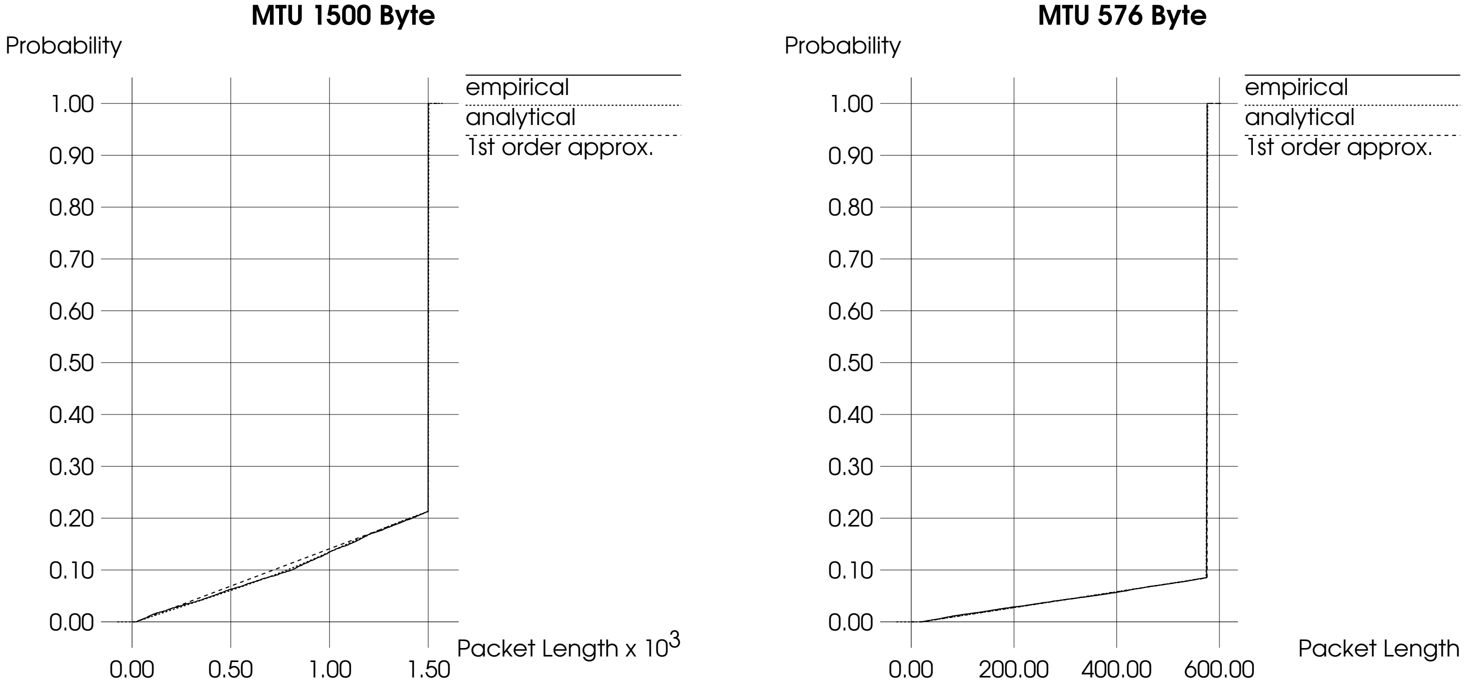
Voraussetzungen, u.a.:

- Überlagerung von n Einzelquellen, $n \in \{3, 30\}$
- H.261-Codierung bei $n = 30$
- H.261-Codierung (20 EQ) gemischt mit H.263-Codierung (10 EQ)

a) $n = 3$


MTU 1500 Byte
Probability
empirical
analytical
1st order approx.
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Packet Length x 10^3
0.00
0.50
1.00
1.50
MTU 576 Byte
Probability
empirical
analytical
1st order approx.
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Packet Length
0.00
200.00
400.00
600.00

b) $n = 30$


MTU 1500 Byte
Probability
empirical
analytical
1st order approx.
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Packet Length x 10^3
0.00
0.50
1.00
1.50
MTU 576 Byte
Probability
empirical
analytical
1st order approx.
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Packet Length
0.00
200.00
400.00
600.00

Ausblick: Managementmöglichkeiten

Realitätsnahe Modelle ermöglichen Einflußnahme auf Primärlast unter Kenntnis der durch sie induzierten Sekundärlast.

Beispiel: Anpassung der Datenrate an einen durch (r, b) (Rate, *Bucket*tiefe) spezifizierten *Token Bucket*

- Framelängen gemäß der Randverteilung $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$ auf Primärlastebene induziert Gruppe von `IP`-Pakete, deren Summe der Längen approximativ $\mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$ verteilt ist, mit $\tilde{\mu} = (1 + \frac{\psi}{\Theta})\hat{\mu}$ und $\tilde{\sigma} = (1 + \frac{\psi}{\Theta})\hat{\sigma}$
- Aus $\tilde{\mu}, \tilde{\sigma}, r, b$ lassen unter Anwendung von Lindley-Prozessen die Verteilung $F(\cdot)$ der *Bucket*füllung herleiten
- Management:
  - Ziel: $\tilde{\mu} \to r$
  - Randbedingung: $(F * \mathcal{N})(b) \geq 1 - \alpha$
- Alternative: Sei mit $\hat{f}$ der aktuelle Füllungsstand gegeben, $\mathcal{N}(\tilde{\mu} - \hat{f}, \tilde{\sigma}^2)(b) \geq 1 - \alpha$

Weitere Möglichkeiten:

- Abschätzen der Zusatzlast bei Verwendung vorwärtsfehlerkorrigierender Codes
- Entwicklung TCP-freundlicher Applikationen
- ...