

Fluid-Flow-Ansätze zur analytischen Untersuchung eines Fair Queueing Bediensystems

Detlef Saß

Universität Stuttgart
Institut für Nachrichtenvermittlung und Datenverarbeitung
DSass@santiron.de

Zusammenfassung Um unterschiedliche Dienstgüten bei der Übertragung von Daten zu garantieren, muß der ankommende Datenstrom nach der verlangten Dienstgüte klassifiziert und in Warteschlangen isoliert werden. Außerdem muß eine Bediendisziplin vorhanden sein, die den aufgeteilten Datenstrom wieder entsprechend der Dienstgüten zusammenbringt und weitertransportiert. Es wird eine Bediendisziplin, erweitertes Prioritätenmodell, betrachtet, bei der jede Dienstgüteklasse eine Mindestbandbreite zur Verfügung hat. Dabei kann eine Dienstgüteklasse Bandbreite von der höherprioren Klasse leihen, wenn diese Klasse Bandbreite übrig hat. Dadurch kann eine effizientere Ausnutzung der Bandbreiten erreicht werden. Für diese Disziplin wird mit Hilfe der Fluid-Flow-Methode ein mathematisches Modell vorgestellt.

1 Einleitung

Im heutigen Zeitalter der Informationsgesellschaft wird die Welt in immer höherem Grade vernetzt. Dabei besteht die Idee Informationen jeglicher Art zwischen unterschiedlichsten Geräten und Orten auf einer möglichst einheitlichen Infrastruktur zu transportieren. Dies kann durch diensteintegrierenden Netze realisiert werden.

In diensteintegrierenden Netzen werden Informationen unterschiedlichster Dienstgüte gemeinsam transportiert. Um dennoch unterschiedliche Dienste anbieten zu können, warten Informationen in nach (Dienst-) Klassen isolierten Warteschlangen auf freiwerdende Ressourcen. Ein Scheduler kontrolliert die Bedienung der Warteschlangen nach bestimmten Kriterien.

In dieser Arbeit wird ein sogenanntes *erweitertes Prioritätensystem* mathematisch beschrieben bei dem jede Klasse ein reservierte Mindestbandbreite zu Verfügung hat. Um statistisches Multiplexen zu ermöglichen, kann eine niederpriore Klasse Bandbreite von höherprioren Klassen leihen. Unter der Annahme, daß niederpriore Klassen ihre Mindestbandbreite eher ausnutzen werden, können höherpriore von niederprioren Klassen keine Bandbreite leihen. In Abb. 1 wird dies für den Spezialfall mit zwei Klassen dargestellt.

Dieses System erweitert das von Elwalid und Mitra in [3] vorgestellte Prioritätensystem um eine genauere Approximation der Verteilungsfunktionen

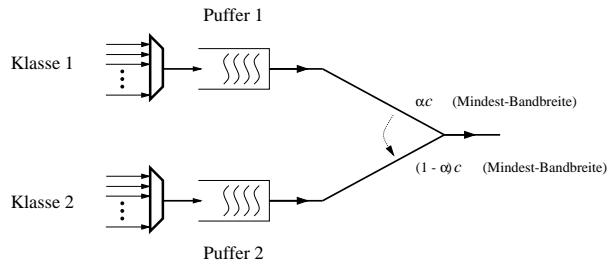


Abbildung 1. Grundmodell des erweiterten Prioritätenmodell

des Pufferinhalts und der Gewährung einer Mindest-Bandbreite der einzelnen Dienstgüteklassen.

Das Papier ist wie folgt strukturiert. In Abschnitt 2 wird die Fluid Flow Methode aufgearbeitet. In Abschnitt 3 wird das erweiterte Prioritätenmodell erläutert und in Abschnitt 4 erfolgt eine Leistungsbewertung.

2 Fluid-Flow-Methode

Durch die Fluid-Flow-Methode kann ein FIFO-Modell (first-in-first-out Modell) analytisch beschrieben werden. Diese Methode wurde zuerst in [1] vorgestellt und danach durch viele weitere Veröffentlichungen weiterentwickelt.

Beim FIFO-Modell wird der ankommende Datenstrom in einen einzigen Puffer geleitet. In diesem werden die Daten nach ihrem zeitlichen Eintreffen eventuell zwischengespeichert und dann in dieser Reihenfolge abgearbeitet. Die Daten werden nur zwischengespeichert, wenn die ankommende Datenstromrate größer ist als die Bedienrate oder wenn sich bereits Daten im Puffer befinden.

Der entscheidende Ansatz bei der Fluid-Flow-Methode ist die Betrachtung des Datenverkehrs, wie der Name schon andeutet, als Flüssigkeitsströme. Dies bedeutet unter anderem, daß letztendlich nicht mehr die Datenpakete gezählt werden, sondern daß der Datenstrom als kontinuierlicher und gleichförmiger Fluß interpretiert wird. Vom mathematischen Standpunkt aus ist dies ein Übergang von einer diskreten Anschauung in eine stetige, kontinuierliche Anschauung. Der wesentliche Unterschied zur paketerorientierten Anschauung ist der, daß sich der Datenstrom beliebig fein granulieren läßt. Damit kann erreicht werden, daß die Bedieneinheit nicht mehr mit einem einzelnen Datenpaket belegt bzw. blockiert ist, da die Bearbeitung jeweils an einer beliebigen Stelle stoppen und starten kann.

Hat der Quellprozeß Zustandsdauern die negativ-exponentiell verteilt sind, dann wird der Quellprozeß als *Markov Modulates Fluid Process* (kurz MMFP) bezeichnet. Entsprechend wird der Bedienprozeß als MMFP bezeichnet. Im folgenden wird der Bedienprozeß als Abflußprozeß bezeichnet, um deutlicher auf die Situation einzugehen, daß Flüssigkeitsprozesse betrachtet werden.

Die Fluid-Flow-Methode beschreibt das FIFO-Modell für eine große Puffergröße sehr gut, da der Effekt, daß kurzfristig phasengleich Datenpakete ankommen können, durch einen genügend großen Puffer kompensiert werden kann.

Sei nun der Ankunftsprozeß ein Markov Modulates Fluid Process (MMFP) der durch die Übergangsratenmatrix \mathbf{M} der Dimension $N+1$ und den Senderratenvektor \mathbf{r} charakterisiert. Sei außerdem c die konstante Abflußrate mit dem der Puffer bedient wird.

Das ij -te Matrixelement $(\mathbf{M})_{ij}$ entspricht der Übergangsrate vom Zustand j in den Zustand i . Die Diagonalelemente der Matrix \mathbf{M} müssen so gewählt werden, daß die Spaltensummen 0 ergeben.¹

Sei $\pi_i(t, x)$ die Verteilungsfunktion des Pufferinhaltes des Systems zum Zeitpunkt t aus, wenn die Markov-Kette sich im Zustand i befindet, d.h.

$$\pi_i(t, x) := \text{P}[\text{''Zustand zum Zeitpunkt } t \text{''} = i \wedge \text{''Pufferinhalt zum Zeitpunkt } t \text{''} \leq x]. \quad (1)$$

Die zukünftige Entwicklung eines Markov-Prozesses hängt wegen der Eigenschaft der Gedächtnislosigkeit nur vom aktuellen Zustand ab, daraus ergibt sich der folgende grundlegende Lösungsansatz. Zur Bestimmung der Verteilungsfunktionen π_i ($i = 0, \dots, N$) wird das System zum Zeitpunkt $t+\Delta t$ betrachtet und die Verteilungsfunktion zu diesem Zeitpunkt durch die Verteilungsfunktionen zum Zeitpunkt t ausgedrückt. Es können während der Zeitspanne $[t, t+\Delta t]$ für π_i zwei wesentliche Ereignisse geschehen. Es kann zum einen ein Zustandsübergang von irgendeinem Zustand in den Zustand i erfolgen oder es kann überhaupt kein Zustandsübergang geschehen. Somit ergibt sich der Ansatz

$$\begin{aligned} \pi_i(t+\Delta t, x) &= \sum_{k=0, k \neq i}^N \pi_k(t, x - (r_k - c)\Delta t) \mathbf{M}_{ik} \Delta t \\ &\quad + \pi_i(t, x - (r_i - c)\Delta t) \left(1 - \sum_{k=0, k \neq i}^N \mathbf{M}_{ki} \Delta t\right) \\ &\quad + \text{O}(\Delta t^2) \end{aligned} \quad (2)$$

Nach geeigneten Umformungen und beim Grenzübergang $\Delta t \rightarrow 0$ und Zusammenfassung der Gleichungen in Matrixschreibweise ergibt sich folgende Gleichung

$$\frac{\partial}{\partial t} \boldsymbol{\pi}(t, x) + \mathbf{D} \frac{\partial}{\partial x} \boldsymbol{\pi}(t, x) = \mathbf{M} \boldsymbol{\pi}(t, x) \quad (3)$$

Wobei Matrix \mathbf{D} als *Driftmatrix* bezeichnet wird und ist durch

$$\mathbf{D}_{kj} := \begin{cases} r_k - c & \text{falls } k = j \\ 0 & \text{sonst} \end{cases}. \quad (4)$$

¹ Die Matrix \mathbf{M} ist eigentlich die transponierte Übergangsratenmatrix. Der Grund dafür liegt in der Links-Multiplikation der Matrix \mathbf{M} in (5). Sie wird aber zur Erleichterung der Notation nur als Übergangsratenmatrix bezeichnet.

definiert. Die gemeinsame Verteilungsfunktion sei $\boldsymbol{\pi}(t, x) := (\pi_0(t, x), \dots, \pi_N(t, x))^T$.

Im stationären Fall, welcher für die weiteren Betrachtungen der wesentliche ist, kann die partielle Ableitung nach t vernachlässigt werden. Bezeichne $\boldsymbol{\pi}(x) := \lim_{t \rightarrow \infty} \boldsymbol{\pi}(t, x)$, dann ergibt sich aus (3)

$$\mathbf{D} \frac{d}{dx} \boldsymbol{\pi}(x) = \mathbf{M} \boldsymbol{\pi}(x) \quad (5)$$

ein System von gewöhnlichen homogenen Differentialgleichungen erster Ordnung.

Dieses System kann durch die Berechnung der Eigenwerte z_i und Eigenvektoren $\boldsymbol{\varphi}_i$ der Matrix $\mathbf{D}^{-1} \mathbf{M}$ gelöst werden. Damit läßt sich die Lösung von (5) in der Spektralform

$$\boldsymbol{\pi}(x) = \sum_{i=0}^N a_i e^{z_i x} \boldsymbol{\varphi}_i. \quad (6)$$

darstellen.

Der stationäre Zustandswahrscheinlichkeitsvektor $\boldsymbol{\omega} := (\omega_0, \dots, \omega_N)^T$, der Eigenvektor zum Eigenwert $z_0 = 0$ ist, läßt sich aus der Normierungsbedingung $\sum_i \omega_i = 1$ und den sogenannten "Gleichungen des statistischen Gleichgewichts für stationäre Markov'sche Prozesse"

$$\sum_{j, j \neq i} \mathbf{M}_{ji} \omega_i \stackrel{!}{=} \sum_{k, k \neq i} \mathbf{M}_{ik} \omega_k \quad (i = 0, 1, \dots, N) \quad (7)$$

berechnen [10].

Die Zustände der Markov-Kette werden anhand des Vorzeichens der Driftwerte unterschieden. Die Zustände in denen die Senderate kleiner ist als die Abflußrate c werden als Unterlast-Zustände (Underload-Zustände, kurz *UL-Zustände*) bezeichnet. Die Zustände, in denen dies nicht der Fall ist, werden als Überlast-Zustände (Overload-Zustände, kurz *OL-Zustände*) bezeichnet.

Die Eigenwerte werden nach folgendem Schema indiziert

$$\underbrace{z_\tau \leq z_{\tau-1} \leq \dots \leq z_1}_{\text{OL-Zustände}} < z_0 = 0 < \underbrace{z_{\tau+1} \leq \dots \leq z_N}_{\text{UL-Zustände}}, \quad (8)$$

wobei τ der Anzahl der OL-Zustände entspricht.

Um die Koeffizienten a_i aus (6) berechnen zu können, müssen die Randbedingungen formuliert werden. Dazu wird angenommen, daß der Puffer die Kapazität S hat.

Damit lauten die Randbedingungen

$$\pi_i(0) = 0 \quad \text{für } i \text{ ist OL-Zustand.} \quad (9)$$

$$\pi_j(S) = \omega_j \quad \text{für } j \text{ ist UL-Zustand.} \quad (10)$$

und bilden ein lineares Gleichungssystem der Dimension $N+1$.

Falls ein Wartesystem betrachtet werden soll, dies bedeutet ein System mit unendlich großem Puffer ($S \rightarrow \infty$), dann vereinfachen sich die Randbedingungen da die Koeffizienten $a_j = 0$ für $j = \tau+1, \dots, N$ sind.

Werden nun mehrere verschiedene Quellen betrachtet, die in einen Puffer senden, dann kann mit Hilfe des Kronecker-Produktes und der Kronecker-Summe eine Ersatzquelle bestimmt werden. Mit dieser Ersatzquelle kann dann die Verteilungsfunktion berechnet werden. Seien die K Quellen durch $(\mathbf{M}_i, \mathbf{R}_i)$ charakterisiert, wobei \mathbf{M}_i die Übergangsratenmatrix und $\mathbf{R}_i := \text{diag}(\mathbf{r}_i)$ Diagonalmatrix und \mathbf{r}_i der Senderatenvektor ist, dann kann die Ersatzquelle durch (\mathbf{M}, \mathbf{R}) , mit

$$\mathbf{M} := \mathbf{M}_1 \oplus \mathbf{M}_2 \oplus \cdots \oplus \mathbf{M}_K \quad (11)$$

$$\mathbf{R} := \mathbf{R}_1 \oplus \mathbf{R}_2 \oplus \cdots \oplus \mathbf{R}_K, \quad (12)$$

charakterisiert werden. Die stationären Zustandswahrscheinlichkeiten werden ebenfalls durch $\boldsymbol{\omega} := \boldsymbol{\omega}_1 \otimes \dots \otimes \boldsymbol{\omega}_K$ bestimmt bzw. durch (7) berechnet.

3 Erweitertes Prioritätenmodell

Beim erweiterten Prioritätenmodell soll jede Klasse eine vorher festgelegte Mindest-Bandbreite zur Verfügung haben. Die Summe über alle Mindest-Bandbreiten soll der Gesamtbandbreite entsprechen. Die Klassen sind nach ihrer Priorität absteigend sortiert. Die Daten einer Klasse werden (eventuell) in einem FIFO-Puffer zwischengespeichert.

Eine Klasse hat Bandbreite übrig, wenn der Puffer dieser Klasse leer ist und die Senderate der Klasse kleiner ist als ihre Mindest-Bandbreite. Dann wird die übrige Bandbreite der nächst-niederen Klasse zur Verfügung gestellt. Somit entsteht eine einseitige Kopplung in Richtung der niederprioreren Klassen.

Das nachfolgende Modell unterscheidet sich von dem in [3] beschriebenen durch die Gewährung der Mindest-Bandbreite, einer unterschiedlichen Anwendung, Einbeziehen bzw. Lösen des modellierten Prozeß bzw. Gesamtsystems.

3.1 Ansatz und mathematisches Modell

Es wird ein System mit zwei Klassen betrachtet, da der wesentliche Ansatz und die wesentlichen Resultate bereits dadurch erklärt und entwickelt werden können.

Die Klasse i , ($i = 1, 2$), soll als Flüssigkeitsstrom (fluid process) interpretiert werden und soll durch das Paar Übergangsratenmatrix und Senderatenvektor

$$(\mathbf{M}_i, \mathbf{r}_i), \quad (13)$$

welche die Dimension $N_i + 1$ haben, charakterisiert werden.

Eine andere anschauliche Darstellung dieses Systems, bei der die einseitige Kopplung besonders deutlich wird, ist in Abb. 2 abgebildet.

In dieser Abbildung wird ein Puffer als ein Gefäß ohne offensichtlichen Abfluß dargestellt. In dieses sendet die Klasse ihre Daten hinein und der eigentliche Abfluß ist ein Prozeß (mit einem Zustand) der die Daten hinausbefördert. Dies bedeutet, daß der Abflußprozeß als Quelle mit einer negative Senderate aufgefaßt wird. Die Unterscheidung Quelle oder Abfluß kann somit auf das Vorzeichen des Senderatenvektors reduziert werden. Dies ermöglicht eine flexiblere Sichtweise

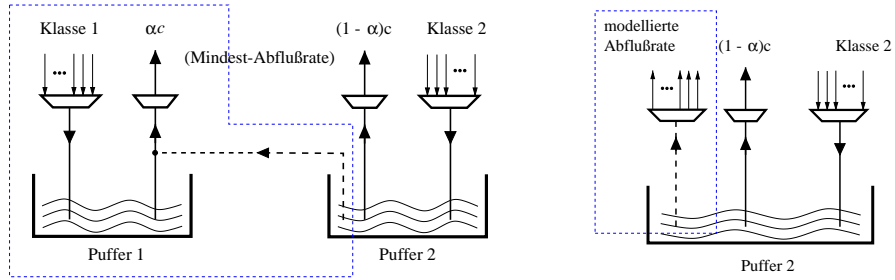


Abbildung 2. Alternative Darstellung einer einseitigen Kopplung (links) und Puffer 2 mit modellierter zusätzlicher Abflußrate (rechts)

bzw. Darstellung und eine vereinfachte Behandlung beim weiteren Vorgehen. Der gestrichelte Arm des Abflußprozesses von Puffer 1 soll die Abgabe der Abflußrate widerspiegeln.

Um die Verteilungsfunktion des Inhalts von Puffer 2 zu bestimmen, muß der gestrichelte Arm des Abflußprozesses von Puffer 1 in einer angemessenen Weise bei der Betrachtung Berücksichtigung finden. Diese muß das dynamische Verhalten der variablen zusätzlichen Abflußrate widerspiegeln, da wie in [13] zu finden ist, ein einfacher, undynamischer Ansatz nicht ausreicht. Dazu wird der gestrichelte Arm bzw. der Teil des Systems, welcher durch eine gestrichelte Linie umgrenzt ist, als ein weiterer Abflußprozeß für Puffer 2 modelliert. Dies ist durch einen weiteren Abflußprozeß in Abb. 2 dargestellt.

Der zusätzliche Abflußprozeß wird als MMFP interpretiert und geeignet modelliert. Damit kann der modellierte Abflußprozeß und der Abflußprozeß mit der konstanten Mindest-Abflußrate zu einem Prozeß überlagert werden. Dieses System kann dann mittels der Fluid-Flow-Methode, wie folgend erläutert, gelöst werden.

3.2 Modellierung des Abflußprozesses

Zur Modellierung des Abflußprozesses wird der Zustandsraum \mathcal{S} der Klasse 1 in die zwei Bereiche Underload und Overload

$$\begin{aligned} \mathcal{S}_{UL} &:= \{s \in \mathcal{S} \mid (r_1)_s < \alpha c\} \\ \mathcal{S}_{OL} &:= \{s \in \mathcal{S} \mid (r_1)_s \geq \alpha c\} \end{aligned} \quad (14)$$

partitioniert.

Die Zustandsübergänge der Klasse 1 innerhalb der Menge \mathcal{S}_{UL} spiegeln das dynamische Verhalten der zusätzlichen Abflußrate wider. Wogegen die Zustandsübergänge der Klasse 1 aus der Menge \mathcal{S}_{UL} in die Menge \mathcal{S}_{OL} oder innerhalb der Menge \mathcal{S}_{OL} keine Änderung der zusätzlichen Abflußrate bedeuten.

Die Modellierung soll anhand einer einfachen Markov-Kette plausibel gemacht werden. Sei Klasse 1 ein Geburts-Sterbe-Prozeß mit N_1+1 Zuständen, der durch

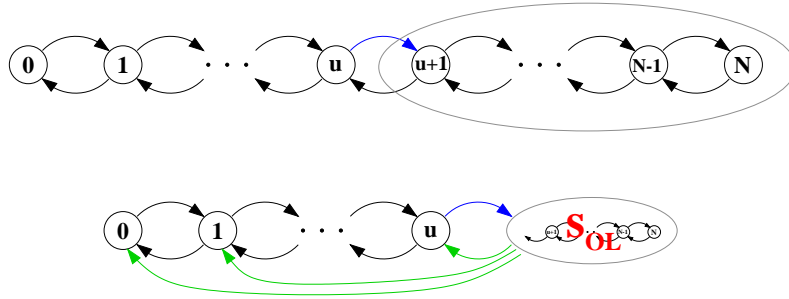


Abbildung 3. Zustandsdiagramm der Klasse 1 (oben) und der modellierten Abflußquelle (unten)

Überlagerung N_1 identischer, unabhängiger On-Off-Quellen entsteht. Sei Zustand u der größte UL-Zustand.

Die zu modellierende Abflußquelle soll die Dynamik der UL-Zustände $\{0, 1, \dots, u\}$ widerspiegeln und werden deshalb in detaillierter Form einfließen. Wogegen die OL-Zustände zu einem Zustand s_{OL} zusammengefaßt werden. Es müssen allerdings zusätzliche Übergänge aus Zustand s_{OL} in $\{0, 1, \dots, u\}$ bestimmt werden, da der Abarbeitung des noch gefüllten Puffer 1 Rechnung zu tragen ist, bevor Bandbreite abgegeben werden kann. Dies ist in Abb. 3 dargestellt.

Um UL-Zustände sehr detailliert einfließen zu lassen wird die Submatrix

$$\mathbf{M}_{UL} := \{(\mathbf{M}_1)_{ss'} \mid s, s' \in \mathcal{S}_{UL}\} \quad (15)$$

gebildet, welche die Übergangsraten innerhalb der Menge \mathcal{S}_{UL} enthält. Die Diagonalelemente müssen allerdings so angepaßt werden, daß die Spaltensummen identisch 0 sind, damit \mathbf{M}_{UL} einer Übergangsratenmatrix entspricht.

Sei $\mathbf{F}(x) := (F_0(x), \dots, F_{N_1}(x))^T$ die stationäre Verteilungsfunktion des Inhaltes von Puffer 1. Bezeichne $\mathbf{F}_{UL}(x) := \{F_s(0) \mid s \in \mathcal{S}_{UL}\}$ den Teilvektor von $\mathbf{F}(x)$, welcher nur die Komponenten der Verteilungsfunktion der UL-Zustände beinhaltet.

Die Situation, daß der Puffer 1 leer ist und leer bleibt, kann nur eintreten, wenn sich Klasse 1 in einem UL-Zustand befindet, der Puffer 1 leer ist und wenn nur Zustandsübergänge in weitere UL-Zustände geschehen. Die Wahrscheinlichkeit dieses obigen Ereignisses für die einzelnen UL-Zustände aus \mathcal{S}_{UL} kann durch

$$\mathbf{q} := \{q_s \mid s \in \mathcal{S}_{UL}\} := \frac{\mathbf{M}_{UL} \mathbf{F}_{UL}(0)}{\langle \mathbf{M}_{UL} \mathbf{F}_{UL}(0), \mathbf{1} \rangle} \quad (16)$$

bestimmt werden. Das Skalarprodukt $\langle \mathbf{M}_{UL} \mathbf{F}_{UL}(0), \mathbf{1} \rangle$ dient der Normierung.

Da die OL-Zustände für das dynamische Verhalten der Abgabe der übrigen Abflußrate kaum eine Rolle spielen, werden diese beim modellierten Abflußprozeß zu dem einen Zustand s_{OL} zusammengefaßt. Dazu muß die mittlere Zeitdauer

bestimmt werden, bei der das System in der Situation verweilt in der keine Abflußrate übrig ist. Dies entspricht der mittleren Vollastdauer (mean busy period) der Klasse 1 und ist der Kehrwert der Übergangsrate, welche durch

$$\lambda_{s_{OL}} := -\frac{\langle \mathbf{M}_{UL} \mathbf{F}_{UL}(0), \mathbf{1} \rangle}{1 - \langle \mathbf{F}(0), \mathbf{1} \rangle} \quad (17)$$

bestimmt werden kann.

Der modellierte Abflußprozeß läßt sich dann durch das Paar $(\mathbf{M}_{out}, \mathbf{r}_{out})$ aus Übergangsratenmatrix und Serverratenvektor charakterisieren. Diese werden wie folgt festgelegt. Sei d_{UL} die Dimension der Matrix \mathbf{M}_{UL} . Dann haben die Matrix \mathbf{M}_{out} und der Serverratenvektor \mathbf{r}_{out} die Dimension $d_{out} := d_{UL} + 1$ und die Gestalt

$$\mathbf{M}_{out} := \left[\begin{array}{c|c} \mathbf{M}_{UL} - \text{diag}(\mathbf{a}) & \mathbf{b} \\ \hline \mathbf{a} & -\lambda_{s_{OL}} \end{array} \right], \quad \mathbf{r}_{out} := [\mathbf{r}_{UL}, 0] \quad (18)$$

Wobei $\mathbf{r}_{UL} := \{(\mathbf{r}_1)_s - \alpha c \mid s \in \mathcal{S}_{UL}\}$,

$$a_i := \sum_{k \in \mathcal{S}_{OL}} (\mathbf{M}_1)_{ki} \quad i = 1, 2, \dots, d_{UL} \quad (19)$$

die Gesamtrate aus dem UL-Zustand i in die OL-Zustände ist und

$$b_i := q_i \lambda_{s_{OL}} \quad i = 1, 2, \dots, d_{UL}, \quad (20)$$

die Übergangsrate vom zusammengefaßten OL-Zustand s_{OL} in den UL-Zustand i ist. Für den stationären Zustandswahrscheinlichkeitsvektor des modellierten Abflußprozesses gilt $\boldsymbol{\omega}_{out} = (\mathbf{F}_{UL}(0), 1 - \langle \mathbf{F}_{UL}(0), \mathbf{1} \rangle)^T$, dieser hat die Dimension d_{out} . Es ist zu beachten, daß $\mathbf{F}_{UL}(0)$ ein Vektor der Länge d_{UL} ist. Damit ist der zusätzliche Abflußprozeß für Puffer 2 vollständig beschrieben und die Verteilungsfunktion des Inhaltes von Puffer 2 kann mittels der Fluid-Flow-Methode bestimmt werden.

3.3 Bestimmung der Lösung

Sei $\boldsymbol{\pi}$ die Verteilungsfunktion des Inhaltes von Puffer 2. Um diese zu bestimmen wird die Fluid-Flow-Methode verwendet. Damit die Fluid-Flow-Methode angewendet werden kann, müssen einige Matrizen bzw. Vektoren definiert werden und Gleichungen aufgestellt werden. Dies wird nur in knapper Form geschehen, um sie an das in diesem Abschnitt betrachtete System anzupassen.

Dazu wird die Übergangsratenmatrix des Gesamtsystems, welches in Abb. 2 dargestellt ist, durch

$$\mathbf{M} := \mathbf{M}_2 \oplus \mathbf{M}_{out} \quad (21)$$

bestimmt, welche die Dimension $(N_2+1)d_{\text{out}} =: N$ hat. Damit ist die Verteilungsfunktion $\boldsymbol{\pi}(y) := (\pi_0(y), \dots, \pi_{N-1}(y))^T$ durch das gewöhnliche Differentialgleichungssystem

$$\mathbf{D} \frac{d}{dy} \boldsymbol{\pi}(y) = \mathbf{M} \boldsymbol{\pi}(y) \quad (22)$$

bestimmbar. Dabei ist

$$\mathbf{D} := \text{diag}(\mathbf{r}_2) \oplus (-\text{diag}(\mathbf{r}_{\text{out}})) \oplus (-(1-\alpha)c) \quad (23)$$

die Driftmatrix. Sie entsteht durch die Kronecker-Summe aus den Diagonalmatrizen der Senderate \mathbf{r}_2 der Klasse 2, der modellierten Serverrate \mathbf{r}_{out} und dem eigentlichen Abflußprozeß $(1-\alpha)c$.

Diese Vorgehensweise liefert für den Fall, daß der mittlere Drift

$$\boldsymbol{\omega}^T \mathbf{D} \mathbf{1} < 0 \quad (24)$$

eine Lösung, wobei der Wahrscheinlichkeitsvektor $\boldsymbol{\omega}$ ein (stehender) Vektor sein soll. Die Lösung kann durch die Spektraldarstellung

$$\boldsymbol{\pi}(y) = \sum_i a_i e^{z_i x} \boldsymbol{\varphi}_i \quad (25)$$

dargestellt werden. Wobei die Eigenwerte mit z_i und die zugehörigen Eigenvektoren mit $\boldsymbol{\varphi}_i$ bezeichnet worden sind. Die Koeffizienten werden durch die Randbedingungen, wie sie in Abschnitt 2 angegeben sind, bestimmt.

3.4 Verwendung der Lösung

Bei Betrachtung einer endlichen Größe von Puffer 2 werden alle N Eigenwerte bzw. Eigenvektoren betrachtet. Dieser Ansatz kann bei großer Puffergröße bzw. bei großen Eigenwerten zu numerischen Problemen führen. Diese Probleme können verringert werden, wenn der Puffer 2 als unendlich groß betrachtet wird und die Verlustwahrscheinlichkeit durch die Überschreitungswahrscheinlichkeit nach oben abgeschätzt wird. Die Überschreitungswahrscheinlichkeit wird durch

$$l(y) = 1 - \langle \boldsymbol{\pi}(y), \mathbf{1} \rangle \quad (26)$$

bestimmt.

Es können unter Umständen aber immer noch numerische Probleme auftreten, trotz der Betrachtung der nicht-positiven Eigenwerte, bei denen keine vernünftige Lösung zu bestimmen ist. Aus diesem Grund kann der Ansatz aus [3], welcher den dominanten (größten negativen) Eigenwert z_1 der Matrix $\mathbf{D}^{-1} \mathbf{M}$ und die Chernoff-Schranke L_0 verwendet, auf das hier betrachtete System übertragen werden. Diese soll im folgenden als ECDE-Approximation (extended Chernoff dominant eigenvalue) bezeichnet werden. Die Bestimmung der Chernoff-Schranke wird in [3], [13] erläutert. Damit kann die Überschreitungswahrscheinlichkeit durch

$$l(y) \leq L_0 e^{z_1 y} \quad (27)$$

abgeschätzt werden.

Der Ansatz für ein System mit zwei Klassen kann auf mehrere angewendet werden. Dazu werden die ersten beiden Klassen, entsprechend der Reihenfolge, wie die Klassen die übrige Abflußrate erhalten sollen, zuerst betrachtet. Dann wird nach obigem Ansatz der zusätzliche Abflußprozeß für die zweite Klasse aus der ersten Klasse modelliert. Dieser wird dann dazu verwendet, die Verteilungsfunktion des Pufferinhaltes der zweiten Klasse zu bestimmen. Der modellierte zusätzliche Abflußprozeß von der ersten Klasse und der konstante Abflußprozeß der zweiten Klasse werden aggregiert. Damit kann dann der zusätzliche Abflußprozeß der übrigen Abflußrate für die dritte Klasse modelliert werden. Dieses Verfahren wird rekursiv auf die noch verbleibenden Klassen angewendet. Dabei entsteht bei jeder Modellierung des zusätzlichen Abflußprozesses ein Fehler, welcher sich dementsprechend verstärkt und fortpflanzt. Dieser Sachverhalt wurde hier nicht untersucht, da in den meisten Fällen tatsächlich zwei verschiedene Klassen ausreichend sind [2].

4 Leistungsbewertung

Die oben beschriebene Approximation wurde mit dem Programm MATLAB bzw. Octave numerisch bestimmt und mit den Simulationsdaten, welche durch die Simulationsumgebung SimLib des Institutes für Nachrichtenvermittlung u. Datenverarbeitung der Universität Stuttgart [9] gewonnen wurden, verglichen.

Die Quellen einer Klasse wurden als identische, unabhängige On-Off-Quellen gewählt mit negativ-exponentiell verteilten Zustandszeiten mit Parameter μ_i, λ_i und Senderate h_i . Die Anzahl der Quellen je Klasse sind identisch.

Mit "Simulation Puffer 1" bzw. "Puffer 2" werden die Graphen der Überschreitungswahrscheinlichkeiten für den entsprechenden Puffer bezeichnet, welche durch die Simulation gewonnen worden sind.

Die beiden Approximationen der Überschreitungswahrscheinlichkeit wurden für (26) mit "Approximation des Puffer 2" und für (27) mit "ECDE Puffer 2", die Fluid-Flow-Lösung für Puffer 1 als "FFM Puffer 1", bezeichnet.

Diese Graphen sollen anhand Abb. 4 exemplarisch für eine Situation mit 20 Quellen je Klasse dargestellt werden.

Die ECDE-Approximation liefert eine konservative Abschätzung allerdings mit einer stärkeren Überschätzung der Überschreitungswahrscheinlichkeit des Puffer 2, da nur ein Eigenwert und die Chernoff-Schranke dort einfließen. Die oben ausführlicher beschriebene Approximation bei der alle (bzw. alle negativen) Eigenwerte einfließen liefert ein deutlich genaueres Ergebnis, allerdings liegt hier eine Unterschätzung der Überschreitungswahrscheinlichkeit vor. Diese ist allerdings eine deutlich geringere Abweichung als die Abweichung der ECDE-Approximation. Somit spiegelt die Approximation aus Abschnitt 3 genauer den Verlauf der Überschreitungswahrscheinlichkeit von Puffer 2 wider. Der bekannte Ansatz der Fluid-Flow-Methode aus [1] entspricht dem Graph "FFM Puffer 1", mit der erwarteten guten Übereinstimmung.

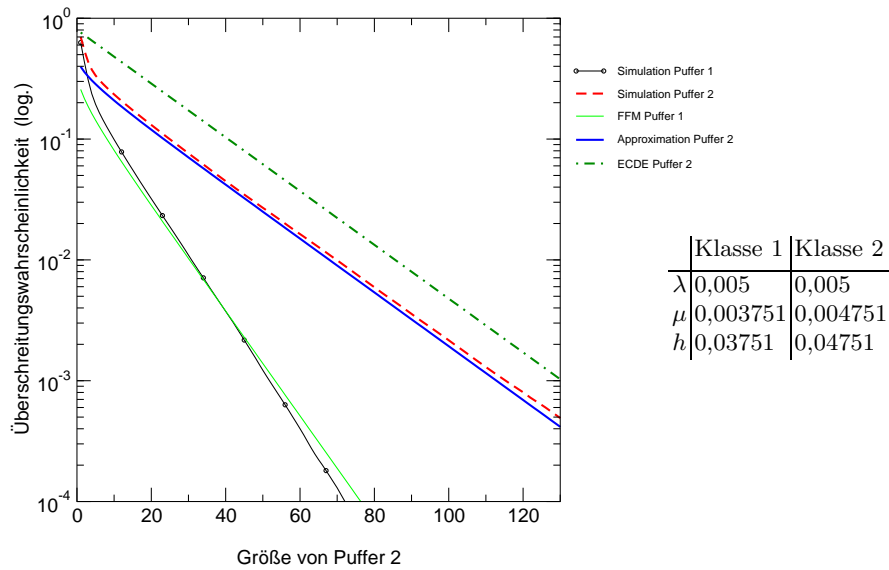


Abbildung 4. Verlustwahrscheinlichkeiten der Puffer ($\alpha = \frac{1}{2}$, 10^7 Pakete)

Der Grund weshalb bei der Approximation von Puffer 2 ein Unterschätzung zur Simulation vorliegt wurde noch nicht ausführlich genug untersucht. Es könnte zum einen sein, daß eventuell die Interpretation des Abflußprozeß von Puffer 1 als MMFP und die damit verbundene Gedächtnislosigkeit des Prozesses nicht vollständig die reale Situation widerspiegelt. Zum anderen könnten es numerische Ungenauigkeiten sein, da Gleichungssysteme gelöst worden sind bzw. Ähnlichkeitstransformationen zur Bestimmung der Eigenwerte durchgeführt worden sind. Vorallem aber, da die numerischen Ergebnisse von Puffer 1 verwendet werden müssen um den Abflußprozeß numerisch zu modellieren. Dieses wird dann wiederum zur Bestimmung der Fluid-Flow-Lösung verwendet.

5 Zusammenfassung und Ausblick

Für die hier betrachtete Bediendisziplin wurde eine Approximation der Verteilungsfunktion für den Inhalt des gekoppelten Puffers, dies entpricht dem Puffer der sich Bandbreite ausleihen kann, bestimmt. Daraus läßt sich die Verlustwahrscheinlichkeit des Puffers und eine Abschätzung dieser bestimmen.

Die Abschätzung (27) ist eine robuste und numerisch schnell zu bestimmen- de Approximation. Dies erweitert das System aus [3] um die Gewährung einer Mindest-Bandbreite der einzelnen Klassen. Dieser Ansatz könnte zur Verbindungsannahmekontrolle (CAC) verwendet werden. Dieses CAC kann aufgrund der Gewährung der Mindest-Bandbreite unabhängig je Klasse durchgeführt werden.

Der Ansatz aus (26) liefert keine konservative, jedoch eine deutlich genauere, als die Approximation, welche nur den dominanten Eigenwert und die Chernoff-Schranke verwendet. Dies ist eine neue Approximation bzw. ein neuer Ansatz und wurde in [3] nicht betrachtet. Dieser Ansatz kann bei Dimensionierungs- und Planungs- bzw. Designfragen Anwendung finden.

Acknowledgement Ich möchte meinen Betreuern Klaus Dolzer, Stefan Bodamer und Martin Lorang für Ihre Unterstützung beim Estellen dieser Arbeit bzw. meiner Diplomarbeit danken.

Literatur

1. D. Anick, D. Mitra and M. Sondhi, *Stochastic theory of a data handling system with multiple sources*, Bell Sys, Tech. J., **61**, pp. 1871-1894, 1982
2. K. Dolzer, W. Payer and M. Eberspächer, *A simulation study on traffic aggregation in multi-services networks*, Proceedings of the IEEE Conference on High Performance Switching and Routing (ATM 2000), Heidelberg, June 2000, pp. 157-165
3. A. Elwalid and D. Mitra, *Analysis, approximation and admission control of a multi-service multiplexing system with priorities*, IEEE INFOCOMM 1995, pp. 463-473
4. A. Elwalid and D. Mitra, *Analysis and design of rate-based congestion control of high speed networks, I: stochastic fluid models, access regulation*, Queueing Systems **9**, pp. 29-64, 1991
5. A. Elwalid and D. Mitra, *Effective bandwidth of general Markovian traffic sources and admission control of high speed networks*, IEEE/ACM Trans. Networking **1**(3), pp. 329-343, 1993
6. M. Fiedler and H. Voss, *How to win the numerical battle against the finite buffer stochastic fluid flow model*, COST 257, 257TD(99)38, 1999
7. M. Fiedler and U. Krieger, *The fluid flow model with variable capacity*, COST 257, 257TD(00)18, 2000
8. L. Kleinrock, *Queueing systems, volume I: theory*, John Wiley and Sons, New York, London, Sydney, Toronto, 1975, ISBN 0-471-49111-X
9. IND, *Simulation Library (INDSimLib)*, www.ind.uni-stuttgart.de/INDSimLib
10. P. J. Kühn, *Teletraffic theory and engineering*, Vorlesungsskript, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, Stuttgart 1999
11. H. Kröner, *Verkehrssteuerung in ATM-Netzen - Verfahren und verkehrstheoretische Analyse zur Zellpriorisierung und Verbindungsannahme*, 62. Bericht über verkehrstheoretische Arbeiten, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, Stuttgart 1995
12. D. Mitra, *Stochastic theory of a fluid model of producer and consumer coupled by a buffer*, Advances in Applied Probability, **20**, pp. 646-676, 1988
13. D. Saß, *Fluid-Flow-Ansätze zur analytischen Untersuchung eines Fair Queueing Bediensystems*, Diplomarbeit Nr. 1676, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 2000
14. M. Schopp, *Analytische Behandlung allgemeiner Fluid-Flow-Modelle für ATM-Netze*, Diplomarbeit Nr. 1194, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1992
15. T. Stern and A. Elwalid, *Analysis of separable markov-modulated rate model for information-handling systems*, Advances in Applied Probability, **23**, pp. 105-139, 1991