

Kapitel 4 M/G/1-System

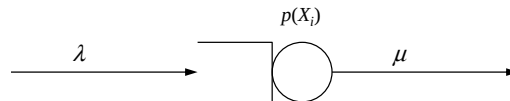
Universität Hannover
Institut für Allgemeine Nachrichtentechnik
Prof. Dr.-Ing. K. Jobmann
Markus Radimirsch

1

2. Das M/G/1-System

- Herleitung der Durchschnittswerte für Warteschlangenlänge und Zeit im System
- Ausflug: Erneuerungstheorie
- Wahrscheinlichkeitsverteilung der Warteschlangenlänge
- Beispiele und Erweiterungen des M/G/1-Systems
 - M/G/1-System mit Leerzeiten
 - Reservierung und Polling
 - Mehrnutzersystem
 - System mit Prioritäten

- System mit einer Bedieneinheit und unendlich großer Warteschlange
- Kunden treffen Poisson-verteilt mit Rate λ ein (M für Markov-Prozess)
- Die Bedienzeiten haben eine beliebige Verteilung (G für General Distribution).
- Kunden werden in der Reihenfolge bedient, in der sie ankommen, und X_i ist die Bedienzeit des i -ten Kunden. Die Bedienzeiten X_i haben die Wahrscheinlichkeitsdichte $p(X_i)$, bzw. die Wahrscheinlichkeitsverteilung $P(X_i)$



Durchschnittliche Bedienzeit:

$$\bar{X} = E\{X\} = \frac{1}{\mu}$$

Zweites Moment der Bedienzeit:

$$\overline{X^2} = E\{X^2\}$$

W_i - Wartezeit des i -ten Kunden in der Queue

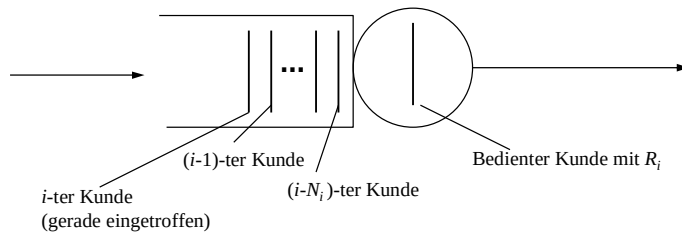
R_i - Restliche Bedienzeit des aktuell bedienten Kunden aus Sicht des i -ten Kunden bei seiner Ankunft

X_i - Bedienzeit des i -ten Kunden

N_i - Anzahl Kunden in der Queue aus Sicht des i -ten Kunden bei seiner Ankunft

Die Wartezeit des i -ten Kunden in der Queue ergibt sich zu:

$$W_i = R_i + \sum_{j=i-N_i}^{i-1} X_j$$



Für den Erwartungswert von W_i gilt, wenn die Zufallsvariablen N_i und X_i unabhängig sind:

$$E\{W_i\} = E\{R_i\} + E\left\{\sum_{j=i-N_i}^{i-1} E\{X_j \mid N_i\}\right\} = E\{R_i\} + \bar{X}E\{N_i\}$$

Die Grenzwertbildung $i \rightarrow \infty$ führt zu:

$$W = R + \frac{1}{\mu} N_Q$$

wobei R die durchschnittliche restliche Bediendauer ist

6

Die Werte W , R und N_Q sind alles Langzeit-Grenzwerte, wenn die Zeit oder der Kundenindex gegen unendlich geht. Wir nehmen an, dass diese Grenzwerte existieren, was für fast alle real auftretenden Systeme der Fall ist, solange $\lambda < \mu$. Beachten Sie, dass N_Q und R aus Sicht eines ankommenden Kunden den Werten entsprechen, die ein außenstehender Betrachter zu einer beliebigen Zeit sieht. Dies gilt aufgrund des Poisson-Charakters des Ankunftsprozesses.

Durchschnittliche Wartezeit in Queue W

Mit Little's Law:

$$N_Q = \lambda W$$

Ersetzen in vorige Formel:

$$W = \lambda \cdot R + \frac{\lambda}{\mu} N_Q = \lambda \cdot R + \rho N_Q \Rightarrow N_Q = \frac{\lambda \cdot R}{1 - \rho} = \lambda \cdot W$$

Und somit:

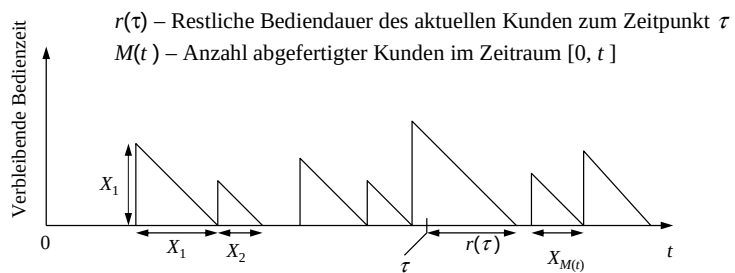
$$W = \frac{R}{1 - \rho}$$

Nun fehlt uns nur noch R .

7

Wie man an den obigen Formeln sehen kann, hängt die durchschnittliche Wartezeit in der Queue nur vom Nutzungsgrad ρ und der durchschnittlichen restlichen Bediendauer R ab. Beachten Sie auch, dass dies Notwendigkeit sich aus der Tatsache ergibt, dass der Bedienprozess nicht gedächtnislos ist!

Wir verwenden dazu eine graphische Methode:



$$\bar{r} = \frac{1}{t} \int_0^t r(\tau) d\tau = \frac{1}{t} \sum_{i=1}^{M(t)} \frac{X_i^2}{2}$$

8

Das Bild zeigt die restliche Bediendauer als eine Funktion der Zeit t . Immer wenn die Abfertigung eines Kunden begonnen wird, beginnt $r(\tau)$ mit X_i und nimmt dann linear ab. Wenn ein Kunde in der Queue ist, wird mit dessen Abfertigung sofort nach Bediende des vorigen Kunden begonnen.

Wir versuchen nun, den Mittelwert von $r(\tau)$ im Intervall von $[0, t]$ zu berechnen. Das Integral entspricht der Summe der Flächeninhalte der jeweiligen dreieckigen Bediendauernverläufe über der Zeit.

Grenzwertbildung (unter der Annahme, dass der Grenzwert existiert) ergibt:

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t r(\tau) d\tau = \frac{1}{2} \lim_{t \rightarrow \infty} \frac{M(t)}{t} \cdot \lim_{t \rightarrow \infty} \frac{\sum_{i=1}^{M(t)} X_i^2}{M(t)}$$

Die rechte Seite ist das Produkt aus der durchschnittlichen Bedienrate (= Ankunftsrate) und dem zweiten Moment der Bediendauer.

Somit:

$$R = \frac{\lambda \overline{X^2}}{2}$$

Die Pollaczek-Khinchin (P-K) Formel ergibt sich für die Wartezeit in der Queue damit zu:

$$W = \frac{R}{1-\rho} = \frac{\lambda \bar{X}^2}{2(1-\rho)}$$

Die Zeit im System ist:

$$T = \bar{X} + W = \frac{1}{\mu} + \frac{\lambda \bar{X}^2}{2(1-\rho)}$$

Die Wartezeit in Queue ist wesentlich bestimmt durch das zweite Moment des Bedienprozesses!

10

Beachten Sie, dass die Herleitung auf zwei wesentlichen Annahmen basiert:

1. Die stationären Durchschnittswerte W , R und N_Q existieren.
2. Die Zeitlichen Durchschnittswerte sind gleich den Ensemble-Durchschnittswerten. Dies setzt die Ergodizität des Prozesses voraus

In den meisten interessierenden Fällen sind diese Voraussetzungen gegeben.

Diskussion der P-K
Formel

$$W = \frac{\lambda \overline{X^2}}{2(1-\rho)}$$

- Was passiert, wenn zwar $\rho < 1$ und $E\{X\}$ endlich, aber $E\{X^2\}$ nicht endlich ist?
 - Dann ist die Bedienzeit eines kleinen Anteils Kunden sehr sehr lang, riesige Anzahlen Kunden stellen sich in der Warteschlange an und werden jeweils signifikant durch die verbleibende Bedienzeit des aktuell bedienten Kunden verzögert.

Diskussion der P-K
Formel (2)

- Die Herleitung der P-K Formel geschah unter der Annahme, dass Kunden in der Reihennfolge ihrer Ankunft bedient werden. Es kann gezeigt werden, dass sie auch gilt, wenn die Bedienreihenfolge nicht der Ankunftsreihenfolge entspricht.
 - Angenommen, Kunde i und j tauschen die Plätze. Dann hat i die restliche Wartezeit von j und j die von i . Der Durchschnitt bleibt gleich.
 - Jede geänderte Reihenfolge kann durch eine endliche Anzahl von Platztauschen repräsentiert werden.

Es gilt:

$$E\{X\} = \bar{X} = 1/\mu$$

$$E\{X^2\} = \bar{X}^2 = 2/\mu^2$$

Damit ist:

$$W = \frac{\lambda \bar{X}^2}{2(1-\rho)} = \frac{2\lambda}{2\mu^2(1-\rho)} = \frac{\rho}{\mu(1-\rho)},$$

$$T = W + \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)} + \frac{1}{\mu} = \frac{\rho}{\lambda(1-\rho)}$$

und:

$$N_Q = \frac{\rho^2}{(1-\rho)}; \quad N = \frac{\rho}{(1-\rho)}$$

Dies entspricht den Ergebnisse für M/M/1-Systeme

D steht für deterministisch, d.h. alle Kunden haben die gleiche Bedienzeit $1/\mu$. Dann gilt:

$$E\{X\} = \bar{X} = 1/\mu$$

$$E\{X^2\} = \bar{X}^2 = 1/\mu^2$$

Damit ist:

$$W = \frac{\lambda \bar{X}^2}{2(1-\rho)} = \frac{\lambda}{2\mu^2(1-\rho)} = \frac{\rho}{2\mu(1-\rho)},$$

$$T = W + \frac{1}{\mu} = \frac{\rho}{\lambda(1-\rho)} + \frac{\rho}{2\lambda}$$

und:

$$N_Q = \frac{\rho^2}{2(1-\rho)}; \quad N = \frac{\rho}{2(1-\rho)} + \frac{\rho}{2}$$

14

Das M/D/1-System ergibt den minimal möglichen Wert für $E\{X^2\}$ bei gegebenem μ . Daher sind die ermittelten Werte für W , T , N_Q und N eine untere Schranke für M/G/1-Systeme mit demselben λ und μ .

Vergleich M/M/1 und
M/D/1

M/M/1:

$$W = \frac{\rho}{\mu(1-\rho)},$$

$$T = \frac{\rho}{\lambda(1-\rho)}$$

$$N_Q = \frac{\rho^2}{(1-\rho)};$$

$$N = \frac{\rho}{(1-\rho)}$$

M/D/1:

$$W = \frac{\rho}{2\mu(1-\rho)},$$

$$T = \frac{\rho}{2\lambda(1-\rho)} + \frac{\rho}{2\lambda}$$

$$N_Q = \frac{\rho^2}{2(1-\rho)};$$

$$N = \frac{\rho}{2(1-\rho)} + \frac{\rho}{2}$$

15

Interessant ist auch, dass W und N_Q für ein M/D/1-System genau die Hälfte der Werte für ein M/M/1-System ergeben. Die Werte für N und T sind bei M/D/1 gleich wie bei M/M/1 für kleines ρ und tendieren gegen die halben Werte für $\rho \rightarrow 1$. Dies liegt daran, dass die erwartete Bediendauer die gleiche ist für beide Fälle, und bei kleinem ρ geschieht das meiste Warten beim Bedienen. Für große ρ ist der Großteil der Wartezeit in der Queue.

Beachten Sie, dass wir bei M/G/1-
Modellen bislang nur
Durchschnittswerte berechnet
haben.

Für M/M/1 kennen wir aber auch Verteilungsdichten für
die Anzahl der Nutzer im System

16

Hier zeigt sich, dass dadurch, dass die Gleichgewichtsgleichungen aufgrund der mangelnden Gedächtnislosigkeit des Bedienprozesses nicht aufgestellt werden können. Stattdessen hängen T und N entscheidend von dem zweiten Moment des Bedienprozesses ab. Es wurde auch keine Markov-Kette zur Ableitung verwendet!

Wie genau macht sich nun die Abhängigkeit vom Bedienprozess bemerkbar? Wir werden die Antwort in der Theorie der Erneuerungen finden (Renewal Theory)

- Herleitung der Durchschnittswerte für Warteschlangenlänge und Zeit im System
- **Ausflug: Erneuerungstheorie**
- Wahrscheinlichkeitsverteilung der Warteschlangenlänge
- Beispiele und Erweiterungen des M/G/1-Systems
 - M/G/1-System mit Leerzeiten
 - Reservierung und Polling
 - Mehrnutzersystem
 - System mit Prioritäten

Situation: Ein Hotelportier tritt seinen Dienst zu einem beliebigen Zeitpunkt an. Hotelgäste kommen Poisson-verteilt mit der Rate λ an. Wie lange muß der Portier warten, bis der nächste Hotelgast ankommt?

1. Durchschnittliche Zeit zwischen Hotelgästen ist $1/\lambda$, also wartet er im Durchschnitt $1/(2\lambda)$
2. Der Poisson-Prozess ist gedächtnislos, also wartet er $1/\lambda$

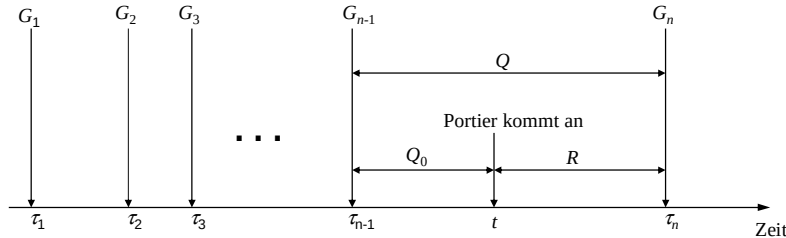
Die zweite Lösung impliziert, dass die Zeit seit dem letzten Hotelgast und der Ankunft des Portiers = $1/\lambda$ plus die Wartezeit von $1/\lambda$ ist, also ist der Abstand zwischen zwei Gästen $2/\lambda$. Das stimmt nicht mit der Ankunftsrate für Gäste von λ überein, wir haben es also mit einem Paradox zu tun

Trotzdem stimmt Lösung 2.

Die Zwischenankunftszeiten dürfen einer beliebigen Statistik gehorchen, die einzelnen Ankünfte müssen statistisch unabhängig sein.

G_k – Ankunft des k-ten Gastes

τ_k – Zeitpunkt der Ankunft des k-ten Gastes



Wahrscheinlichkeitsverteilung: $F(x) \equiv P[\tau_{k+1} - \tau_k \leq x]$

Wahrscheinlichkeitsdichte: $f(x) = \frac{dF(x)}{dx}$

20

Der Portier kommt zum Zeitpunkt t an. G_{n-1} ist der letzte Gast, der vor ihm angekommen ist, der nächste Gast nach t ist G_n . Q ist dabei diese spezielle Zwischenankunftszeit, R ist die Zeit, die der Portier bis zum nächsten Gast warten muss. Die Serie der Ankunftszeiten $\{\tau_k\}$ ist ein Erneuerungsprozess.

Erneuerungsprozesse beschäftigt sich mit der sofortigen Erneuerung von Komponenten. Dann könnte beispielsweise $\{\tau_k\}$ eine Serie von Zeitpunkten darstellen, zu denen die alte Komponente ausfällt und durch eine neue ersetzt wird. Q ist dann die Lebenszeit der Komponente, R ist die verbleibende Lebensdauer und $Q_0 = Q - R$ wird das Alter der Komponente genannt. Diese Terminology soll im folgenden verwendet werden.

Unser Ziel ist es, die Wahrscheinlichkeitsdichte für Q und R zu finden. Da wir nur an den begrenzenden Verteilungen interessiert sind, nehmen wir an, dass der Erneuerungsprozess bereits beliebig lange läuft.

Q ist nicht nach $F(x)$ verteilt, d.h. der
Portier sucht sich kein „typisches“ Intervall
für sein Eintreffen aus

21

Die Lösung für das Paradox liegt darin, dass sich der Portier kein typisches Intervall der Dauer Q für seine Ankunft aussucht. Es ist nämlich wahrscheinlicher, dass der Portier während eines langen Intervalls eintrifft als während eines kurzen.

Sei

$$R(x) \equiv P[R \leq x]$$

die Wahrscheinlichkeitsverteilung für die verbleibende Lebensdauer
und die zugehörige Wahrscheinlichkeitsdichte

$$r(x) = \frac{dR(x)}{dx}$$

Analog für die ausgewählte Lebenszeit Q :

$$Q(x) \equiv P[Q \leq x]$$

$$q(x) = \frac{dQ(x)}{dx}$$

Herleitung Erneuerungsprozesse

Lange Intervalle zwischen Erneuerungszeitpunkten besetzen einen größeren Anteil der Zeitachse. Daher trifft der Portier eher in einem langen als in einem kurzen Intervall ein.

Somit:

$$q(x)dx = Kx \cdot f(x)dx$$

Durch Integrieren der beiden Seiten ergibt sich:

$$K = \frac{1}{m_1} \text{ mit } m_1 = E\{\tau_k - \tau_{k-1}\}$$

und somit haben wir unser erstes gewünschtes Ergebnis:

$$q(x) = \frac{x \cdot f(x)}{m_1}$$

23

Wir nehmen an, dass die Wahrscheinlichkeit, dass der Portier in einem Intervall der Dauer x eintrifft, proportional ist zu dessen Länge x und natürlich zu dessen relativer Auftrittshäufigkeit $f(x) dx$.

Der Faktor K in der Formel ist eine Konstante, die der Normalisierung der Wahrscheinlichkeitsdichte dient.

Nun weiter zur Dichte der verbleibenden Lebensdauer:

$$P[R \leq y | Q = x] = \frac{y}{x}; \quad 0 \leq y \leq x$$

Die Verbundwahrscheinlichkeitsdichte für Q und R ergibt sich dann zu:

$$P[y < R \leq y + dy, x < Q \leq x + dx] = \frac{dy}{x} \cdot \frac{x \cdot f(x) dx}{m_1} = \frac{f(x) dy dx}{m_1}$$

$r(y)$ ergibt sich durch Integration über x :

$$r(y) dy = \int_{x=y}^{\infty} \frac{f(x) dy dx}{m_1} \quad \text{und damit: } r(y) = \frac{1 - F(y)}{m_1}$$

Das ist das zweite gewünschte Ergebnis. Wir suchen nun das erste Moment von $r(y)$.

24

Wenn wir wissen, dass $Q=x$, dann ergibt sich die Wahrscheinlichkeit, dass die verbleibende Lebensdauer den Wert y nicht überschreitet, nach der ersten Formel oben. Dies ergibt sich daraus, dass wir einen beliebigen Punkt im Intervall der Dauer x wählen und er somit gleichverteilt im Intervall liegt.

Dazu verwenden wir die Laplace-Transformierte der Wahrscheinlichkeits-dichten zur Generierung von Momenten:

$$f(x) \Leftrightarrow F_s(s)$$

$$r(x) \Leftrightarrow R_s(s)$$

Es gilt für das n -te Moment:

$$F_s^{(n)}(0) = E\{X^n\} = m_n$$

Somit:

$$R_s(s) = \frac{1 - F_s(s)}{sm_1} \text{ mit } F(x) \Leftrightarrow \frac{F_s(s)}{s}$$

Mit $r_n = E\{R^n\}$:

$$r_1 = R_s^{(1)}(0) = \frac{d}{ds} \frac{1 - F_s(s)}{sm_1} \Big|_{s=0}$$

Mit Hilfe von L'Hospital ergibt sich:

$$r_n = \frac{m_{n+1}}{(n+1)m_1}; \quad r_1 = \frac{m_2}{2m_1} = \frac{m_1}{2} + \frac{\sigma^2}{2m_1}$$
$$(\sigma^2 = m_2 - m_1^2 \text{ Varianz})$$

Das bedeutet, dass die verbleibende Lebensdauer R nicht nur vom Mittelwert der Lebensdauer Q einer Komponente abhängt, sondern auch von der Varianz der Lebensdauer

Poisson-verteilte Lebensdauern: $m_1 = 1/\lambda$, $\sigma^2 = 1/\lambda^2$. Somit:

$$\bar{R} = r_1 = \frac{1}{2\lambda} + \frac{\frac{1}{\lambda^2}}{2 \cdot \frac{1}{\lambda}} = \frac{1}{\lambda}$$

Konstante Lebensdauern: $m_1 = 1/\lambda$, $\sigma^2 = 0$. Somit:

$$\bar{R} = r_1 = \frac{1}{2\lambda}$$

27

Der erste Fall entspricht dem, was aufgrund der Gedächtnislosigkeit des Poisson-Prozesses erwartet werden kann.

Der zweite Fall entspricht dem, was wir intuitiv erwartet hätten. Es stellt sich jedoch heraus, dass diese Lösung nur dann stimmt, wenn die Lebensdauern konstant sind. Bei allen anderen Verteilungen der Lebensdauer ist die verbleibende Lebensdauer größer, so dass dieser Fall als untere Schranke angesehen werden kann.

Zusammenhang Erneuerungsprozess und M/G/1

Erneuerungsprozess:

$$R = \frac{\bar{X}}{2} + \frac{\overline{X^2} - \bar{X}^2}{2m_1}$$

M/G/1:

$$R = \frac{\lambda \overline{X^2}}{2}$$

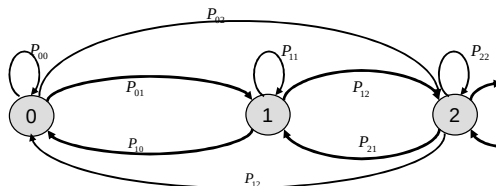
Der Unterschied besteht darin, dass M/G/1 kein wirklicher Erneuerungs-prozess ist. Das liegt daran, dass nach einem abgefertigten Paket in M/G/1 nicht zwangsläufig immer ein Paket zur Bedienung vorhanden ist.

Beiden gemeinsam ist, dass sie von der Varianz des „Bedienprozesses“ abhängen.

2. Das M/G/1-System

- Herleitung der Durchschnittswerte für Warteschlangenlänge und Zeit im System
- Ausflug: Erneuerungstheorie
- **Wahrscheinlichkeitsverteilung der Warteschlangenlänge**
- Beispiele und Erweiterungen des M/G/1-Systems
 - M/G/1-System mit Leerzeiten
 - Reservierung und Polling
 - Mehrnutzersystem
 - System mit Prioritäten

- Zeitdiskrete Markov-Kette geht davon aus, dass der Übergang entlang der Kanten immer zu Zeiten mit konstantem Abstand Δt erfolgt
 - Aufenthaltsdauer folgt geometrischer Verteilung, im Grenzfall $\Delta t \rightarrow 0$ exponentiell verteilt
- Semi-Markov Prozess: Aufenthaltsdauer ist beliebig verteilt. Der Prozess verhält sich aber zu den Zeiten der Zustandsübergänge wie eine Markov-Kette
 - Zu diesen Zeitpunkten sprechen wir von einer eingebetteten Markov-Kette



- Wir betrachten die eingebettete Markov-Kette des M/M/1-Systems zu den Abfertigungszeitpunkten
- Wir definieren eine Matrix der Übergangswahrscheinlichkeiten:

$$\mathbf{K} = [P_{rs}] \equiv P[N_{i+1} = s \mid N_i = r]$$

N_i – Anzahl Kunden im System direkt nach der i -ten Abfertigung

Um von $N_i=r$ zu $N_{i+1}=s$ zu kommen, sind $s - r + 1$ Ankünfte notwendig.

Wir nennen die Übergangswahrscheinlichkeiten:

$$P_{rs} = k_{s-r+1}$$

31

Als Abfertigungszeitpunkt bezeichnen wir die Zeitpunkte unmittelbar nachdem ein Bedienprozess beendet ist und der entsprechende Kunde das System verlassen hat. Wir werden äquivalent auch von „Bedienende“ sprechen.

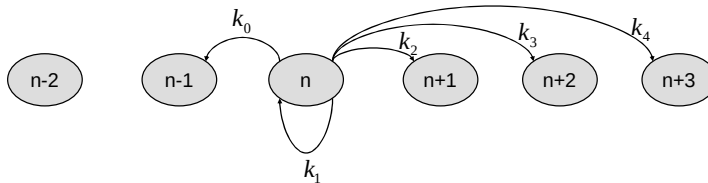
Die „+1“ in $s-r+1$ ergibt sich daraus, dass die Anzahl im System bei der $(i+1)$ -ten Abfertigung um 1 reduziert wird.

Die eingebettete Markov-Kette des M/G/1-Systems

Damit ergibt sich die Matrix der Übergangswahrscheinlichkeiten zu:

$$s : \begin{pmatrix} k_0 & k_1 & k_2 & k_3 & k_4 & k_5 & \dots \\ k_0 & k_1 & k_2 & k_3 & k_4 & k_5 & \dots \\ 0 & k_0 & k_1 & k_2 & k_3 & k_4 & \dots \\ 0 & 0 & k_0 & k_1 & k_2 & k_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \end{pmatrix}$$

Zugehörige eingebettete Markov-Kette:



32

Die Einträge in der Matrix sind Wahrscheinlichkeiten! Beachten Sie, dass die Summe der k 's in den Zeilen immer 1 ergeben muss, da eine Zeile den Übergang aus einem Zustand r in irgendeinen Folgezustand bezeichnet. Ein Zustandsübergang findet in jedem Fall, also mit Wahrscheinlichkeit 1 statt!

Die Einträge ergeben sich folgendermaßen: In der ersten Zeile ist der Ausgangszustand $r=0$. Um von $r=0$ nach der i -ten Abfertigung auf $s=0$ nach der $i+1$ -ten Abfertigung zu kommen, brauchen wir 0 Ankünfte. Dieser Fall kommt nicht wirklich vor, da wir das System eigentlich nur zu den Abfertigungszeitpunkten betrachten und bei 0 Kunden im System auch kein Kunde abgefertigt werden kann. Der Eintrag ist der Vollständigkeit halber mit k_0 eingetragen. Von $r=0$ nach $s=1$ brauchen wir $s-r+1=1$ Ankünfte, repräsentiert durch die Wahrscheinlichkeit k_1 . Die weiteren Einträge ergeben sich analog.

Betrachten wir beispielsweise den Eintrag in der zweiten Reihe ($r=1$) und der vierten Spalte ($s=3$). Dann beginnen wir mit 1 Kunden nach der i -ten Abfertigung. Nach der $i+1$ -ten Abfertigung sind 3 Kunden im System, das bedeutet, dass in der Zwischenzeit 3 Ankünfte waren und 1 Kunde das System zum $i+1$ -ten Abfertigungszeitpunkt verlassen hat.

Sei $\mathbf{p} = [p_0 \ p_1 \ p_2 \ \dots]$ der Vektor der Wahrscheinlichkeiten für die Anzahl Kunden im System. Dann muss im stationären Zustand zu den Abfertigungszeitpunkten gelten:

$$\mathbf{p} = \mathbf{p} \cdot \mathbf{K}$$

Dann ergeben sich die folgenden Gleichungen (siehe Markov-Kette):

$$p_0 = p_0 k_0 + p_1 k_0,$$

$$p_1 = p_0 k_1 + p_1 k_1 + p_2 k_0,$$

$$p_2 = p_0 k_2 + p_1 k_2 + p_2 k_1 + p_3 k_0,$$

$$\vdots \quad \vdots$$

Zusammenfassend:

$$p_r = p_0 k_r + \sum_{s=1}^{r+1} p_s k_{r-s+1}, \quad r = 0, 1, 2, \dots$$

33

Nun wissen wir, wie die Zustandswahrscheinlichkeiten zusammenhängen. Es fehlen uns die Übergangswahrscheinlichkeiten.

Wir können die Wahrscheinlichkeiten für die Anzahl der Ankünfte zwischen dem i -ten und $(i+1)$ -ten Abfertigungszeitpunkt angeben:

$$k_j = \int_0^{\infty} P[\text{Zahl der Ankünfte} = j \mid \tau = t] p_{\mu}(t) dt$$

τ - beobachtete Bediendauer

$p_{\mu}(\tau)$ - Wahrscheinlichkeitsdichte für die Bediendauer τ

Der Ankunftsprozess ist ein Poisson-Prozess, so dass

$$k_j = \int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} p_{\mu}(t) dt$$

Dies ist ein erstes Zwischenergebnis. Nun kann man bei bekannter Bedienstatistik die Übergangswahrscheinlichkeiten berechnen.

Der Weg zur Dichte der Kundenanzahl im System

Der Weg zur Wahrscheinlichkeitsdichte für die Anzahl der Kunden im System zu den Abfertigungszeiten geht über z-Transformierte:

$$Z\{p_i\} = P_z(z) = \sum_{i=0}^{\infty} p_i z^i; \quad K_z(z) = \sum_{i=0}^{\infty} k_i z^i$$

Mit Parameterwechsel ($t=s-1$):

$$p_r = p_0 k_r + \sum_{t=0}^r p_{t+1} k_{r-t}$$

Das sieht aus wie eine Faltung! Es gilt:

$$x_i = \sum_{j=0}^{\infty} f_j g_{i-j} \Leftrightarrow X_z(z) = F_z(z) G_z(z)$$

Einseitige z-Transformation der 2. Gleichung auf der Seite:

$$P_z(z) = p_0 K_z(z) + Z\{p_{t+1}\} K_z(z)$$

35

Bei den z-Transformierten in der unteren Gleichung handelt es sich um einseitige Transformationen, d.h. $i \in [0, \infty]$

Z-Transformation von p_{t+1} :

$$\begin{aligned} Z\{p_{t+1}\} &= \sum_{t=0}^{\infty} p_{t+1} z^t = z^{-1} \sum_{t=0}^{\infty} p_{t+1} z^{t+1} = z^{-1} \sum_{u=0}^{\infty} p_u z^u - z^{-1} p_0 = \\ &= z^{-1} P_z(z) - z^{-1} p_0 \end{aligned}$$

Somit:

$$P_z(z) = p_0 K_z(z) + [z^{-1} P_z(z) - z^{-1} p_0] K_z(z)$$

Nach weiterer Umformung:

$$P_z(z) [1 - z^{-1} K_z(z)] = p_0 [1 - z^{-1}] K_z(z)$$

Auflösen nach $P_z(z)$:

$$P_z(z) = \frac{p_0 [1 - z^{-1}] K_z(z)}{1 - z^{-1} K_z(z)} = \frac{p_0 [1 - z] K_z(z)}{K_z(z) - z}$$

Mit

$$p_0 = 1 - \frac{\lambda}{\mu} = 1 - \rho$$

Ergibt sich:

$$P_z(z) = \frac{(1 - \rho)(1 - z)K_z(z)}{K_z(z) - z}$$

Wir haben nun also eine Gleichung für die z-Transformierte der Wahrscheinlichkeitsdichte der Wartezeit in der Queue zu den Abfertigungszeitpunkten. Sie hängt von λ , μ und der Statistik des Bedienprozesses ab. Der Bedienprozess wird repräsentiert durch k_j mit:

$$k_j = \int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} p(t) dt$$

37

Das bedeutet, dass wir keine allgemeine Lösung im Zeitbereich angeben können, wie das bei M/M/1 der Fall war. Wir müssen also immer (außer bei M/M/1) den Umweg über die z-Transformierte der Ankunftsstatistik zwischen zwei Abfertigungen gehen!

Hier ergänzen:

Gilt die Dichtefunktion für jeden Zeitpunkt?

- Es lässt sich leicht nachweisen, dass für einen Prozess, der potentiell unendlich dauert, die Durchschnittswerte und die Verteilungsdichte zu den Abfertigungszeiten gleich der Werte für die Ankunftszeiten sind:

$$\lim_{T \rightarrow \infty} |a_n(T) - d_n(T)| \leq 1$$

$$d(T) = a(T) + n(0) - n(T)$$

$$\lim_{T \rightarrow \infty} \frac{d_n(T)}{d(T)} = \lim_{T \rightarrow \infty} \frac{a_n(T) + d_n(T) - a_n(T)}{a(T) + n(0) - n(T)} = \lim_{T \rightarrow \infty} \frac{a_n(T)}{a(T)}$$

- Da die Ankunftszeiten Poisson-verteilt und damit unabhängig vom Zustand des Systems sind, gelten die Werte für jeden Zeitpunkt.

38

$a_n(t)$, $d_n(t)$ – Anzahl der Ankünfte, bzw. Abfertigungen im Zeitraum $[0, t]$, wenn n Kunden im System sind

$a(t)$, $d(t)$ – Anzahl der Ankünfte bzw. Abfertigungen im Zeitraum $[0, t]$

- Herleitung der Durchschnittswerte für Warteschlangenlänge und Zeit im System
- Ausflug: Erneuerungstheorie
- Wahrscheinlichkeitsverteilung der Warteschlangenlänge
- **Beispiele und Erweiterungen des M/G/1-Systems**
 - M/G/1-System mit Leerzeiten
 - Reservierung und Polling
 - Mehrnutzersystem
 - System mit Prioritäten

- Angenommen, die Bedieneinheit hat nach jeder Abfertigung eine Leerzeit
 - Ein neuer Kunde wird erst nach Ablauf der Leerzeit bedient
 - Wenn bis Ablauf der Leerzeit kein Kunde da ist, beginnt eine neue Leerzeit
 - Ein ankommender Kunde muss auf die vor ihm wartenden Kunden und die zusätzlichen Leerzeiten warten, bis er bedient wird
- Anwendungen:
 - Datennetze, in denen Steuerdaten zwischen Nutzdaten verschickt werden
 - TDM-Netze, bei denen periodisch wiederkehrende Zeitschlitzte zur Verfügung stehen.

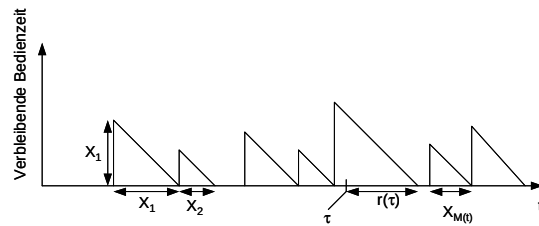
Seien $V_1, V_2, \dots$ die Dauern der Leerzeiten 1, 2, ... (unabhängig voneinander und identisch verteilt).

Es gilt weiterhin:

$$W = \frac{R}{1 - \rho}$$

wobei R nun neu berechnet werden muss. R ist bestimmt entweder durch die restliche momentane Bediendauer oder die restliche Leerzeit

Wir gehen wieder grafisch vor.

M/G/1-System mit
Leerzeiten

$$R = \frac{1}{t} \int_0^t r(\tau) d\tau = \frac{1}{2t} \left[\sum_{i=1}^{M(t)} X_i^2 + \sum_{i=1}^{L(t)} V_i^2 \right] = \frac{M(t)}{2t} \frac{\sum_{i=1}^{M(t)} X_i^2}{M(t)} + \frac{L(t)}{2t} \frac{\sum_{i=1}^{L(t)} V_i^2}{L(t)}$$

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t r(\tau) d\tau = \frac{1}{2} \lim_{t \rightarrow \infty} \frac{M(t)}{t} \cdot \lim_{t \rightarrow \infty} \frac{\sum_{i=1}^{M(t)} X_i^2}{M(t)} + \frac{1}{2} \lim_{t \rightarrow \infty} \frac{L(t)}{t} \cdot \lim_{t \rightarrow \infty} \frac{\sum_{i=1}^{L(t)} V_i^2}{L(t)}$$

42

$M(t)/t$ geht letztendlich gegen λ , so dass der erste Term gegen den bereits in der Herleitung von M/G/1 angegebenen Wert strebt. Wenn $t \rightarrow \infty$, nähert sich der zeitliche Anteil, während dessen Kunden bedient werden, gegen ρ und daher geht der Anteil an Leerzeiten gegen $(1-\rho)$.

M/G/1-System mit
Leerzeiten (2)

Es gilt nun:

$$\frac{1}{2} \lim_{t \rightarrow \infty} \frac{M(t)}{t} \cdot \lim_{t \rightarrow \infty} \frac{\sum_{i=1}^{M(t)} X_i^2}{M(t)} = \frac{\lambda \overline{X^2}}{2}$$

$$\frac{1}{2} \lim_{t \rightarrow \infty} \frac{L(t)}{t} \cdot \lim_{t \rightarrow \infty} \frac{\sum_{i=1}^{L(t)} V_i^2}{L(t)} = \frac{(1-\rho) \overline{V^2}}{2\overline{V}}$$

$$\text{mit } \lim_{t \rightarrow \infty} \frac{t(1-\rho)}{L(t)} = \overline{V} \Rightarrow \lim_{t \rightarrow \infty} \frac{L(t)}{t} = \frac{1-\rho}{\overline{V}}$$

Somit:

$$W = \frac{R}{1-\rho} = \frac{\lambda \overline{X^2}}{2(1-\rho)} + \frac{\overline{V^2}}{2\overline{V}}$$

2. Das M/G/1-System

- Herleitung der Durchschnittswerte für Warteschlangenlänge und Zeit im System
- Ausflug: Erneuerungstheorie
- Wahrscheinlichkeitsverteilung der Warteschlangenlänge
- **Beispiele und Erweiterungen des M/G/1-Systems**
 - M/G/1-System mit Leerzeiten
 - **Reservierung und Polling**
 - **System mit einem Nutzer**
 - Mehrenutzersystem
 - System mit Prioritäten

44

Weiter mit Textdokument

- Datenintervalle, Reservierungsintervalle:
 - Ein Kommunikationskanal wird beispielsweise von mehreren räumlich getrennten Nutzern verwendet.
 - Es kann jeweils nur ein Nutzer zu einer Zeit übertragen
 - Der Kommunikationskanal wird in Zeitrichtung unterteilt. Hierbei wird ein Teil zur Reservierung der Ressourcen verwendet, während ein anderer zur Datenübertragung dient.
- Annahme:
 - m Nutzer, m Datenströme
 - in jedem Datenintervall werden Pakete eines einzelnen Nutzers bedient
 - alle Nutzer werden zyklisch bedient (Polling)
 - jedem Datenintervall geht ein Reservierungsintervall voraus

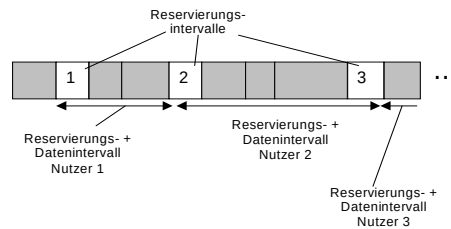
45

Reservierung und Polling gehören in die Welt des scheduling. Damit sind Bedienstrategien gemeint, die dafür sorgen, dass Daten mit unterschiedlichen Eigenschaften richtig gehandhabt werden, so dass eine gewisse Dienstgüt (Quality of Service, QoS) gewährleistet werden kann.

Für ein typisches Beispiel nehmen wir an, dass der Kommunikationskanal von mehreren räumlich getrennten Nutzern verwendet werden kann. Allerdings kann nur ein Nutzer zu einer Zeit übertragen. Der Kommunikationskanal kann somit in Zeitrichtung eingeteilt werden in einen Anteil, der für Datenübertragung, und einen Teil, der für die Reservierung von Ressourcen verwendet wird. Wir nennen diese Intervalle im folgenden Datenintervalle, bzw. Reservierungsintervalle.

Wir werden uns letztendlich mit m Datenströmen von m Nutzern befassen und annehmen, dass in jedem Datenintervall Pakete eines einzigen Nutzers bedient werden. All Nutzer werden zyklisch bedient. Jedem Datenintervall geht ein zugehöriges Reservierungsintervall voraus.

- Zeitverzögert (gated). Es werden nur Pakete bedient, die vor dem Reservierungsintervall eingetroffen sind.
- Unverzögert (exhaustive): Es werden alle Pakete bedient, die bis zum Ende des Datenintervalls eingetroffen sind.
- Teilverzögert (partially gated): Pakete die bis zum Ende des Reservierungsintervalls eingetroffen sind werden bedient.

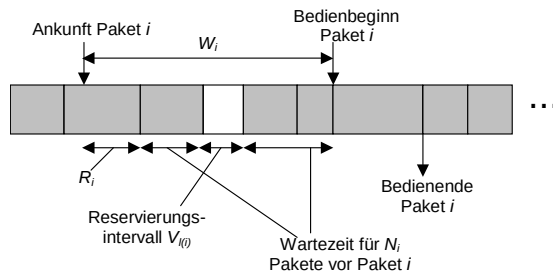


System mit einem Nutzer (1)

- Annahmen:
 - V_i die Dauer des i -ten Reservierungsintervall mit dem 1. Moment $\bar{V}$ und 2. Moment $\bar{V}^2$
 - Zeitverzögertes System, Reservierungsintervalle statistisch unabhängig von Ankunfts- und Bedienprozeß
 - FIFO

47

Wir nähern uns der Lösung zunächst durch Betrachtung eines Systems mit einem Nutzer. Sei V_i die Dauer des i -ten Reservierungsintervall mit dem ersten Moment und zweiten Moment. Wir betrachten zunächst ein zeitverzögertes System und nehmen an, dass die Reservierungsintervalle statistisch unabhängig von Ankunfts- und Bedienprozeß sind. Schließlich nehmen wir zur einfacheren Handhabung an, dass die Pakete in der Reihenfolge ihrer Ankunft bedient werden. Die hergeleiteten Formeln gelten allerdings auch für Fälle, in denen die Bedienreihenfolge nicht zwangsweise der Ankunftsreihenfolge entspricht.

System mit einem Nutzer:
Wartezeit (1)

Wartezeit bis zur Bedienung („Queueing Delay“):

$$E\{W_i\} = E\{R_i\} + \frac{E\{N_i\}}{\mu} + E\{V_{l(i)}\}$$

Wir betrachten das i -te im System eintreffende Paket, wie im Bild dargestellt. Das Paket muss in der Queue die verbleibende Bedien- oder Reservierungszeit des aktuellen Intervalls R_i , sowie die Abfertigung der Pakete in der Queue vor ihm abwarten. Schließlich muss das Paket das eigene Reservierungsintervall mit Dauer $V_{l(i)}$ abwarten (zeitverzögertes System!). Die Wartezeit sieht einem M/G/1-System mit Leerzeiten recht ähnlich! Der Unterschied ist, dass im Reservierungssystem jedes Pakte genau ein Reservierungsintervall abwarten muss, während beim System mit Leerzeiten nur die Pakete, die ein leeres System vorfinden, einen Teil einer Leerzeit abwarten müssen. Beachten Sie, dass das unverzögerte System dem System mit Leerzeiten entspricht!

**System mit einem Nutzer:
Wartezeit (2)**

Die verbleibende Wartezeit R_i , entspricht R für ein System mit Leerzeiten:

$$R = \frac{\lambda \bar{X}^2}{2} + \frac{(1-\rho)\bar{V}^2}{2\bar{V}}$$

Der Grenzwert $\lim_{i \rightarrow \infty} E\{N_i\}/\mu$ entspricht ρW und $\lim_{i \rightarrow \infty} E\{V_{l(i)}\} = \bar{V}$

Damit ist:

$$W = \frac{\lambda \bar{X}^2}{2(1-\rho)} + \frac{\bar{V}^2}{2\bar{V}} + \frac{\bar{V}}{1-\rho} \quad (\text{Ein Nutzer, zeitverzögert})$$

Häufig ist die Dauer des Reservierungsintervalls eine Konstante A .

Dann gilt:

$$W = \frac{\lambda \bar{X}^2}{2(1-\rho)} + \frac{A}{2} \left(\frac{3-\rho}{1-\rho} \right)$$

Beachten Sie auch hier wieder den Einfluss des Erneuerungsprozesses!

2. Das M/G/1-System

- Herleitung der Durchschnittswerte für Warteschlangenlänge und Zeit im System
- Ausflug: Erneuerungstheorie
- Wahrscheinlichkeitsverteilung der Warteschlangenlänge
- **Beispiele und Erweiterungen des M/G/1-Systems**
 - M/G/1-System mit Leerzeiten
 - **Reservierung und Polling**
 - System mit einem Nutzer
 - **Mehrnutzersystem**
 - System mit Prioritäten

50

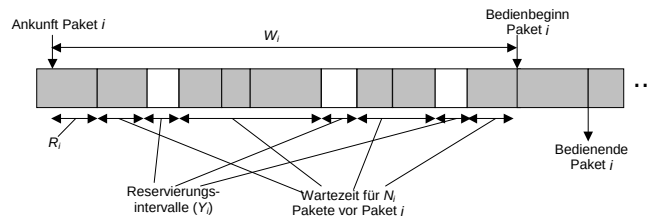
Weiter mit Textdokument

- Annahmen:
 - System mit m Nutzern $0, 1, \dots, m-1$
 - m unabhängige poissonverteilte Ankunftsprozesse der Rate λ/m
 - Nutzungsfaktor: $\rho = \lambda / \mu$
 - $\bar{X}, \bar{X}^2$: 1. und 2. Moment des Bedienprozesse von Nutzer i
 - $\bar{V}, \bar{V}^2$: 1. und 2. Moment des Bedienprozesse von Nutzer i
 - Das l -te Reservierungsintervall u. das l -te Datenintervall gilt für Nutzer $l \bmod m$

51

Wir nehmen nun an, dass das System m Nutzer hat, jeder mit einem unabhängigen Poisson-Ankunfts-prozess mit Rate λ/m . Die Nutzer werden mit $0, 1, \dots, m-1$ durchnummeriert. Mit $\bar{X}$ und $\bar{X}^2$ bezeichnen wir die ersten beiden Momente des Bedienprozesses, und sind die ersten beiden Momente der Dauer des Reservierungsintervalls von Nutzer i . Das l -te Reservierungsintervall ebenso wie das unmittelbar folgende l -te Datenintervall gilt für den Nutzer $l \bmod m$.

Mehrnutzersystem : Wartezeit (1)



Wartezeit bis zur Bedienung („Queueing Delay“):

$$E\{W_i\} = E\{R_i\} + \frac{E\{N_i\}}{\mu} + E\{Y_i\}$$

52

Wir betrachten die i -te Paketankunft. Wie vorher besteht, wie in Bild 3 dargestellt, die Verzögerung des Pakets aus drei Termen, nämlich

- Die Restzeit für die aktuell bearbeitete Paketreservierung (Reservierungs- + Datenintervall);
- Der Mittelwert der Bedienzeit der N_i Pakete, die vor Paket i bedient werden (Beachten Sie, dass dies Pakete von anderen wie auch dem betroffenen Nutzer sind!);
- Der Erwartungswert für die Dauer der Reservierungsintervalle.

Dabei ist Y_i die Dauer aller (ungeteilten) Reservierungsintervalle, die das Paket i abwarten muss, bevor es bedient wird.

$E\{R_i\}$ wird wie gehabt berechnet und ergibt sich zu:

$$R = \frac{\lambda \overline{X^2}}{2} + \frac{(1-\rho) \sum_{l=0}^{m-1} \overline{V_l^2}}{2 \sum_{l=0}^{m-1} \overline{V_l}}$$

Wenn wir $Y = \lim_{i \rightarrow \infty} E\{Y_i\}$ definieren, dann ist

$$W = R + \rho W + Y, \text{ bzw. } W = \frac{R + Y}{1 - \rho}$$

Nun fehlt uns nur noch Y , das wir zunächst für ein unverzögertes System berechnen.

Berechnung von Y :

Es sei:

$$\alpha_{ij} = E\{Y_i \mid \text{Paket } i \text{ trifft in Nutzer } l\text{'s Reservierungs- oder Datenintervall ein und gehört zu Nutzer } (l+j) \bmod m\}$$

mit:

$$\alpha_{ij} = \begin{cases} 0; & j = 0 \\ \bar{V}_{(l+1) \bmod m} + \dots + \bar{V}_{(l+j) \bmod m}; & j > 0 \end{cases}$$

(j ist quasi die Distanz zwischen Nutzer l und dem Nutzer, dem das Paket gehört. Pakete des gerade bedienten Nutzers werden sofort mitbedient, da wir ein unverzügliches System betrachten.)

Da Paket i zu jedem Nutzer mit Wahrscheinlichkeit $1/m$ gehört, gilt (summiert über alle Nutzer $(l+j) \bmod m$):

$E\{Y_i \mid \text{Paket } i \text{ trifft in Nutzer } l\text{'s Reservierungs- oder Datenintervall ein}\} =$

$$\begin{aligned} &= \frac{1}{m} \sum_{j=1}^{m-1} \alpha_{lj} = \frac{1}{m} [(\bar{V}_{(l+1) \bmod m}) + (\bar{V}_{(l+1) \bmod m} + \bar{V}_{(l+2) \bmod m}) + \dots + (\bar{V}_{(l+1) \bmod m} + \dots + \bar{V}_{(l+m-1) \bmod m})] \\ &= \sum_{j=1}^{m-1} \frac{m-j}{m} \bar{V}_{(l+j) \bmod m} \end{aligned}$$

Für die Wahrscheinlichkeit des Eintreffens eines Pakets während das Datenintervall des l -ten Nutzers gilt: ρ/m

Und während des Reservierungsintervalls: $(1-\rho)\bar{V}_l / \sum_{k=0}^{m-1} \bar{V}_k$

Damit folgt:

$$\begin{aligned}
 Y &= \sum_{l=0}^{m-1} \left(\frac{\rho}{m} + \frac{(1-\rho)\bar{V}_l}{\sum_{k=0}^{m-1} \bar{V}_k} \right) \sum_{j=1}^{m-1} \frac{m-j}{m} \bar{V}_{(l+j) \bmod m} = \\
 &= \frac{\rho}{m} \sum_{j=1}^{m-1} \frac{m-j}{m} \left(\sum_{l=0}^{m-1} \bar{V}_l \right) + \frac{(1-\rho)}{\sum_{k=0}^{m-1} \bar{V}_k} \sum_{l=0}^{m-1} \sum_{j=1}^{m-1} \frac{m-j}{m} \bar{V}_l \bar{V}_{(l+j) \bmod m}
 \end{aligned}$$

56

Da alle Nutzer die gleiche Datenrate haben, sind im stationären Zustand die Datenintervalle aller Nutzer im Mittel gleich.

Die zweite Summe kann umgeformt werden zu:

$$\sum_{l=0}^{m-1} \sum_{j=1}^{m-1} \frac{m-j}{m} \bar{V}_l \bar{V}_{(l+j) \bmod m} = \frac{1}{2} \left[\left(\sum_{l=0}^{m-1} \bar{V}_l \right)^2 - \sum_{l=0}^{m-1} \bar{V}_l^2 \right]$$

(Die rechte Seite ist die Summe aus allen möglichen Produkten $\bar{V}_l \bar{V}_{l'}$ mit $l \neq l'$ und die linke Seite ist die Summe aller möglichen Terme $(j/m) \bar{V}_l \bar{V}_{l'}$ und $((m-j)/m) \bar{V}_l \bar{V}_{l'}$ mit $j=|l-l'|$ und $l \neq l'$.)

$$\begin{aligned} \text{Mit } \bar{V} = \frac{1}{m} \sum_{l=0}^{m-1} \bar{V}_l \text{ folgt: } Y &= \frac{\rho \bar{V} (m-1)}{2} + \frac{(1-\rho) m \bar{V}}{2} - \frac{(1-\rho) \sum_{l=0}^{m-1} \bar{V}_l^2}{2m\bar{V}} = \\ &= \frac{(m-\rho) \bar{V}}{2} - \frac{(1-\rho) \sum_{l=0}^{m-1} \bar{V}_l^2}{2m\bar{V}} \end{aligned}$$

Die Wartezeit W für das unverzögerte System lässt sich berechnen durch:

$$W = \frac{\lambda \overline{X}^2}{2(1-\rho)} + \frac{(m-\rho)\overline{V}}{2(1-\rho)} + \frac{\sum_{l=0}^{m-1} (\overline{V}_l^2 - \overline{V}_l^2)}{2\overline{V}} = \frac{\lambda \overline{X}^2}{2(1-\rho)} + \frac{(m-\rho)\overline{V}}{2(1-\rho)} + \frac{\sigma_v^2}{2\overline{V}}$$

mit der Varianz der Reservierungsdauern gemittelt über alle Nutzer, . Dies ist unser erstes Ergebnis für das unverzögerte System.

Teilverzögertes System:

Ein Paket, das während des eigenen Datenintervalls eines Nutzers eintrifft, wird lediglich um durchschnittlich $m\bar{V}$ verzögert. Das Ereignis tritt mit Wahrscheinlichkeit ρ/m ein, so dass Y um $\rho\bar{V}$ zusätzlich verzögert wird.

Es gilt:

$$W = \frac{\lambda \bar{X}^2}{2(1-\rho)} + \frac{(m+\rho)\bar{V}}{2(1-\rho)} + \frac{\sigma_v^2}{2\bar{V}}$$

Verzögertes System:

Ein Paket, das während des eigenen Reservierungsintervalls eines Nutzers eintrifft, wird um durchschnittlich $m\bar{V}$ verzögert. Das Ereignis tritt mit Wahrscheinlichkeit $(1-\rho)/m$ ein, so dass Y um $(1-\rho)\bar{V}$ zusätzlich verzögert wird. Es gilt:

$$W = \frac{\lambda \bar{X}^2}{2(1-\rho)} + \frac{(m+2-\rho)\bar{V}}{2(1-\rho)} + \frac{\sigma_v^2}{2\bar{V}}$$

Vergleich: Reservierungssysteme

Angenommen, alle Reservierungsintervalle haben eine konstante Dauer von A . Dann ist $\bar{V} = A/m$ und $\sigma_v^2 = 0$:

$$W = \frac{\lambda \bar{X}^2}{2(1-\rho)} + \frac{A}{2} \cdot \frac{1-\rho/m}{(1-\rho)} \quad \text{unverzögert}$$

$$W = \frac{\lambda \bar{X}^2}{2(1-\rho)} + \frac{A}{2} \cdot \frac{1+\rho/m}{(1-\rho)} \quad \text{teilverzögert}$$

$$W = \frac{\lambda \bar{X}^2}{2(1-\rho)} + \frac{A}{2} \cdot \frac{1+(2-\rho)/m}{(1-\rho)} \quad \text{verzögert}$$

Im Vergleich zum System mit einem Nutzer zeigt sich, dass im Fall mehrerer Nutzer die Verzögerung etwas verringert wird.

System mit Einzelpaket- Bedienung

System, in dem in jedem Datenintervall eines Nutzers nur 1 Paket bedient wird, nämlich das erste Paket in der Queue eines Nutzers. Hier treten nur die verzögerte und teilverzögerte Version auf.

Wie vorher gilt für $t \rightarrow \infty$:

$$W = R + \rho W + Y, \text{ bzw. } W = \frac{R+Y}{1-\rho}$$

61

R errechnet sich wie zuvor. Um Y für dieses System zu berechnen, gehen wir davon aus, dass ein Paket, das während des Reservierungs- oder Datenintervalls von Nutzer i eintrifft, zu irgendeinem der Nutzer mit gleicher Wahrscheinlichkeit $1/m$ gehört.

System mit Einzelpaket- Bedienung: teilverzögert

Annahme:

Wahrscheinlichkeit für das Eintreffen eines Datenpakets während des Reservierungs- oder Datenintervalls des l -ten Nutzers ist gleichverteilt und damit $1/m$.

Für den stationären Zustand gilt dann:

Erwartungswert {Anzahl Pakete in der Queue des Nutzers dessen Paket im l -ten Intervall eintrifft }:

$$\lim_{i \rightarrow \infty} E\{N_i\} = \lambda W / m$$

Jedes dieser Pakete verursacht einen extra Reservierungszyklus $m\bar{V}$ so dass Y um den Betrag $\lambda W \bar{V}$ vergrößert wird.

Es ergibt sich:

$$W = \frac{R + \tilde{Y}}{1 - \rho - \lambda \bar{V}}$$

62

R errechnet sich wie zuvor. Um Y für dieses System zu berechnen, gehen wir davon aus, dass ein Paket, das während des Reservierungs- oder Datenintervalls von Nutzer l eintrifft, zu irgendeinem der Nutzer mit gleicher Wahrscheinlichkeit $1/m$ gehört.

Erklärung jedes paket brauch nun eigenes Reservierungsintervall,übertragung des paketes gehört nicht mehr zur wateszeit es muß natürlich die übrigen Reservierungszeiten wie oben abwarten aber keine zusätzlichen übertragungen

Generell hat sich die Verzögerung um den Faktor

$$\frac{1-\rho}{1-\rho-\lambda\bar{V}}$$

vergrößert.

Eingesetzt in die obigen Formel ergibt dies:

$$W = \frac{\lambda \bar{X}^2}{2(1-\rho-\lambda\bar{V})} + \frac{(m+\rho)\bar{V}}{2(1-\rho-\lambda\bar{V})} + \frac{\sigma_v^2(1-\rho)}{2\bar{V}(1-\rho-\lambda\bar{V})} \quad (\text{teilverzögert})$$

63

Betrachten wir nun die verzögerte Version. Y_i ist gleich wie bei der teilverzögerten Lösung plus ein zusätzlicher Zyklus von Reservierungsintervalls der mittleren Länge, der auftritt, wenn Paket i während des Reservierungsintervalls seines Besitzers eintrifft und das darauffolgende Datenintervall leer bleibt. Dieser Fall tritt mit Wahrscheinlichkeit auf, wie leicht gezeigt werden kann.

Verzögertes System:

Y_i entspricht teilverzögerter Lösung plus zusätzlichem Reservierungsintervall der mittleren Länge $m\bar{V}$, wenn Paket i während des Reservierungsintervalls seines Besitzers eintrifft und das darauffolgende Datenintervall leer bleibt.

Die Wahrscheinlichkeit hierfür ist: $(1 - \rho - \lambda\bar{V}) / m$

Es folgt:

$$W = \frac{\lambda \bar{X}^2}{2(1 - \rho - \lambda\bar{V})} + \frac{(m + 2 - \rho - 2\lambda\bar{V})\bar{V}}{2(1 - \rho - \lambda\bar{V})} + \frac{\sigma_v^2(1 - \rho)}{2\bar{V}(1 - \rho - \lambda\bar{V})} \quad (\text{verzögert})$$

64

Betrachten wir nun die verzögerte Version. Y_i ist gleich wie bei der teilverzögerten Lösung plus ein zusätzlicher Zyklus von Reservierungsintervalls der mittleren Länge, der auftritt, wenn Paket i während des Reservierungsintervalls seines Besitzers eintrifft und das darauffolgende Datenintervall leer bleibt.

Beachten Sie, dass es hier nicht mehr ausreicht, wenn $\rho = \lambda/\mu < 1$, sondern es muss gelten

$$\rho + \lambda \bar{V} < 1, \text{ bzw. } \lambda \left(\frac{1}{\mu} + \bar{V} \right) < 1$$

Dies liegt daran, dass jedes einzelne Paket ein eigenes Reservierungsintervall der mittleren Dauer braucht, was die mittlere effektiv benötigte Bediendauer auf $1/\mu + \bar{V}$ erhöht.

2. Das M/G/1-System

- Herleitung der Durchschnittswerte für Warteschlangenlänge und Zeit im System
- Ausflug: Erneuerungstheorie
- Wahrscheinlichkeitsverteilung der Warteschlangenlänge
- **Beispiele und Erweiterungen des M/G/1-Systems**
 - M/G/1-System mit Leerzeiten
 - **Reservierung und Polling**
 - System mit einem Nutzer
 - Mehrnutzersystem
 - **System mit Prioritäten**

66

Weiter mit Textdokument

System mit Prioritäten

Erweiterung des M/G/1 Systems:

- eintreffende Kunden werden in n Prioritätsklassen eingeteilt, wobei Klasse 1 die höchste und Klasse n die niedrigste Priorität hat
- Die Ankunftsrate und die ersten beiden Momente der Bediendauern für jede Klasse k werden mit λ_k , $\bar{X}_k$ und $\bar{X}_k^2$ bezeichnet.
- Die Ankunftsprozess für alle Klassen werden als Poisson-verteilt und untereinander und von den Bedienprozessen unabhängig angenommen.

Unterbrechungsfreier Fall :

- Es wird jeder Kunde zunächst abgefertigt, auch wenn ein Kunde mit höherer Priorität eintrifft.
- Eine separate Warteschlange wird für jede Klasse angenommen.
- Wenn die Bedieneinheit frei wird, wird der erste Kunde der höchsten Klasse bedient, und nur wenn die Warteschlange der nächsthöheren Klasse leer ist, kann ein Kunde einer bestimmten Klasse bedient werden

68

Im unterbrechungsfreien Fall wird jeder Kunde zunächst abgefertigt, auch wenn ein Kunde mit höherer Priorität eintrifft. Eine separate Warteschlange wird für jede Klasse angenommen. Wenn die Bedieneinheit frei wird, wird der erste Kunde der höchsten Klasse bedient, und nur wenn die Warteschlange der nächsthöheren Klasse leer ist, kann ein Kunde einer bestimmten Klasse bedient werden. Solche Verhältnisse finden sich häufig in Datenübertragungssystemen. Wir werden eine Formel für die mittlere Verzögerung für jede Prioritätsklasse entwickeln, die ähnlich der P-K Formel ist und auf ähnliche Weise abgeleitet wird.

Es sei:

N_Q^k - Mittlere Anzahl Kunden in der Warteschlange für Klasse k

W_k - Mittlere Zeit in der Warteschlange für Klasse k

$\rho_k = \lambda_k / \mu_k$ - Nutzungskoeffizient für Klasse k

R - Mittlere verbleibende Bediendauer

Wir nehmen an, dass die mittlere Gesamtnutzung
des Systems kleiner 1 ist, also:

$$\sum_{i=1}^n \rho_k < 1$$

69

⇒ Wenn diese Annahme nicht eingehalten wird, ist die Konsequenz, dass es eine Klasse k gibt, für die gilt, dass die mittlere Verzögerung dieser und aller niedrigeren Prioritätsklassen unendlich wird, während nur Kunden mit Prioritäten höher als k endliche Verzögerungen haben.

Wie bei der Herleitung der P-K Formel ergibt sich für die höchste Klasse $k=1$:

$$W_1 = R + \frac{1}{\mu_1} N_Q^1$$

Mit Little's Law gilt $N_Q^1 = \lambda_1 W_1$ und damit ist:

$$W_1 = R + \rho_1 W_1 \Rightarrow W_1 = \frac{R}{1 - \rho_1}$$

Für die Verzögerung der Klasse 2 gilt:

$$W_2 = R + \frac{1}{\mu_1} N_Q^1 + \frac{1}{\mu_2} N_Q^2 + \frac{1}{\mu_1} \lambda_1 W_2$$

70

Für die zweite Klasse erhalten wir einen ähnlichen Ausdruck, außer dass die zusätzlich Verzögerung durch Kunden der Klasse 1 berücksichtigt werden müssen, die während der Wartezeit in der Warteschlange eintreffen

Nach Umformung und unter Verwendung von Little's Law erhalten wir:

$$W_2 = \frac{R + \rho_1 W_1}{1 - \rho_1 - \rho_2}$$

Unter Verwendung des Ausdrucks für W_1 folgt:

$$W_2 = \frac{R}{(1 - \rho_1)(1 - \rho_1 - \rho_2)}$$

Die Herleitung ist nun ähnlich für alle Prioritätsklassen mit $k > 1$. Die Formel für die Verzögerung in der Warteschlange ist somit:

$$W_k = \frac{R}{(1 - \rho_1 - \dots - \rho_{k-1})(1 - \rho_1 - \dots - \rho_k)}$$

Nun fehlt uns noch ein Ausdruck für R . Wir gehen vor wie bei der Herleitung der P-K Formel:

$$R = \frac{1}{2} \sum_{i=1}^n \lambda_i \overline{X_i^2}$$

Es folgt:

$$W_k = \frac{\sum_{i=1}^n \lambda_i \overline{X_i^2}}{2(1 - \rho_1 - \dots - \rho_{k-1})(1 - \rho_1 - \dots - \rho_k)}$$

Zusammenfassend kann gesagt werden, dass allgemein Kunden mit höherer Priorität eine geringere mittlere Verzögerung haben als solche mit niedrigerer Priorität.

72

R errechnet sich wie zuvor. Um Y für dieses System zu berechnen, gehen wir davon aus, dass ein Paket, das während des Reservierungs- oder Datenintervalls von Nutzer l eintrifft, zu irgendeinem der Nutzer mit gleicher Wahrscheinlichkeit $1/m$ gehört.

Beispiel: ATM.

Zerlegung großer Pakete in kleine. Übertragung der Fragmente wird unterbrochen, wenn ein Paket höherer Priorität eintrifft.

73

Eine der Eigenschaften der unterbrechungsfreien Bedienstrategie ist, dass die mittlere Verzögerung einer Prioritätsklasse von den Ankünften niedrigerer Klassen abhängt (auch bei Prioritätsklasse 1, wo R durch die Bedienprozesse aller Klassen beeinflusst wird). Dies liegt daran, dass ein laufender Bedienprozess durch die Ankunft eines Kunden mit höherer Priorität nicht unterbrochen wird. Wie ist das aber, wenn eine solche Unterbrechung erlaubt ist?

Solch ein Verfahren ist annähernd im Fall von ATM gegeben, wo große Pakete in kleine Pakete (Zellen) zerteilt werden. Wird gerade ein Paket in Form einer Gruppe von Zellen bedient und trifft ein Paket höherer Priorität ein, wird die Übertragung des ersten Pakets unterbrochen, d.h. die übrigen Zellen warten so lange, bis alle Zellen des Pakets höherer Priorität bedient worden ist.

Mittleren Zeit im System T_k für Klasse k :

- Kunden der Prioritäten $k+1$ bis n beeinflussen die Berechnung nicht.
- Jede dieser Klassen kann so behandelt werden, als ob sie die niedrigste im System sei.

T_k wird durch drei Terme ausgedrückt:

1. Die durchschnittliche Bediendauer für Klasse k , $1/\mu_k$.
2. Die mittlere Zeit, die benötigt wird, um bei Ankunft eines Kunden der Klasse k die Kunden der Klassen $1, \dots, k-1$ zu bedienen, die sich bereits im System befinden:

$$\frac{R_k}{1 - (\rho_1 + \rho_2 + \dots + \rho_k)}$$

mit der mittleren verbleibenden Bediendauer $R_k = \frac{1}{2} \sum_{i=1}^k \lambda_i \overline{X_i^2}$

74

Wir betrachten nun nicht die Wartezeit in der Queue, sondern die Zeit im System (inklusive Bedienzeit). Bei der Berechnung der mittleren Zeit im System für Klasse k , T_k , beeinflussen Kunden der Prioritäten $k+1$ bis n die Berechnung nicht. Daher können wir jede Klasse so behandeln, als ob sie alleine im System wäre. T_k wird durch drei Terme ausgedrückt:

1. Die durchschnittliche Bediendauer für Klasse k , $1/\mu_k$.
2. Die mittlere Zeit, die benötigt wird, um bei Ankunft eines Kunden der Klasse k die Kunden der Klassen $1, \dots, k-1$ zu bedienen, die sich bereits im System befinden. Diese entspricht, wie man leicht einsieht, der mittleren Wartezeit in der Queue des entsprechenden normalen M/G/1-System ohne Prioritäten, wobei die Kunden der Prioritäten $k+1 \dots n$ vernachlässigt werden.

(Beachten Sie hierzu, dass die unbeendete Arbeit zu allen Zeiten in einem M/G/1-System unabhängig vom Prioritätsschema des Systems ist. Dies gilt für jedes System, in dem die Bedieneinheit immer beschäftigt ist, solange das System nicht leer ist, und Kunden das System erst verlassen, wenn sie zu Ende „bearbeitet“ worden sind.

3. Die mittlere Wartezeit für Kunden der Klassen 1 bis $k-1$, die eintreffen, während der Kunde der Klasse k im System ist. Dieser Term ist:

$$\sum_{i=1}^{k-1} \frac{1}{\mu_i} \lambda_i T_k = \sum_{i=1}^{k-1} \rho_i T_k$$

für $k > 1$ und 0 für $k=1$

Zusammengefasst erhalten wir die Gleichung:

$$T_k = \frac{1}{\mu_k} + \frac{R_k}{1 - (\rho_1 + \rho_2 + \dots + \rho_k)} + T_k \sum_{i=1}^{k-1} \rho_i$$

Das letztendliche Ergebnis für $k=1$ ist:

$$T_k = \frac{(1/\mu_1)(1-\rho_1) + R_1}{1-\rho_1}$$

und für $k>1$:

$$T_k = \frac{(1/\mu_k)(1-\rho_1-\rho_2-\dots-\rho_k) + R_k}{(1-\rho_1-\rho_2-\dots-\rho_{k-1})(1-\rho_1-\rho_2-\dots-\rho_k)}$$

mit R_k wie oben angegeben.

G/G/1-System

- Ankunfts- und Bedienprozess haben eine beliebige Statistik
- Annahme: Zwischenankunfts- und Bedienzeiten sind statistisch unabhängig

77

Zu guter Letzt wollen wir einen kurzen Blick auf das G/G/1-System werfen. In diesem System haben sowohl der Ankunfts- als auch der Bedienprozess eine beliebige Statistik. Wie immer nehmen wir an, dass die Zwischenankunftszeiten und die Bedienzeiten unabhängig sind. Wir werden sehen, dass wir für diese Art System nicht einmal exakte Durchschnittswerte angeben können, geschweige denn Wahrscheinlichkeitsdichten der Wartezeit oder ähnliches. Stattdessen gibt es verschiedene Ausdrücke für obere Schranken, die vom Mittelwert und der Varianz des Ankunfts- und Bedienprozesses abhängen.

Definitionen

σ_a^2 - Varianz der Zwischenankunftszeiten

σ_b^2 - Varianz der Bedienzeiten

λ - Mittlere Ankunftsrate

ρ - Nutzungskoeffizient = λ / μ , wobei $1/\mu$ die mittlere Bediendauer ist

W_k - Wartezeit des k - ten Kunden in Queue

X_k - Bediendauer des k - ten Kunden

τ_k - Zwischenankunftszeit zwischen dem k - ten und $(k + 1)$ - ten Kunden

Definitionen

Zur Vereinfachung der Analyse definieren wir für beliebige Zufallsvariablen Y :

$$Y^+ = \max\{0, Y\}; \quad Y^- = -\min\{0, Y\}$$

$$\bar{Y} = E\{Y\}; \quad \sigma_Y^2 = E\{Y^2 - \bar{Y}^2\}$$

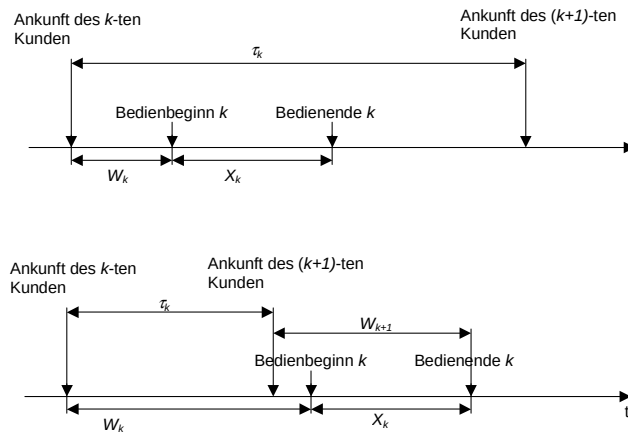
Daraus folgt, dass

$$Y = Y^+ - Y^-; \quad Y^+ \cdot Y^- = 0$$

und deswegen auch:

$$\bar{Y} = \bar{Y}^+ - \bar{Y}^-; \quad \sigma_Y^2 = \sigma_{Y^+}^2 + \sigma_{Y^-}^2 + 2\bar{Y}^+ \cdot \bar{Y}^-$$

Zwei Fälle für den Zusammenhang der Wartezeiten zweier aufeinanderfolgender Kunden



Im Bild ist ersichtlich, dass

$$W_{k+1} = \max\{0, W_k + X_k - \tau_k\} = (W_k + V_k)^+$$

mit $V_k = X_k - \tau_k$. Wir definieren außerdem:

$$I_k = (W_k + V_k)^-$$

als die Zufallsvariable der Dauer der ungenutzten
Periode zwischen der Ankunft des k -ten und des $(k+1)$ -
ten Kunden

Mit den obigen Definitionen können wir schreiben:

$$\sigma_{(W_k+V_k)}^2 = \sigma_{(W_k+V_k)^+}^2 + \sigma_{(W_k+V_k)^-}^2 + 2\overline{(W_k+V_k)^+} \cdot \overline{(W_k+V_k)^-}$$

Durch die Unabhängigkeit von W_k und V_k können wir schreiben:

$$\sigma_{(W_k+V_k)}^2 = \sigma_{W_k}^2 + \sigma_{V_k}^2 = \sigma_{W_k}^2 + \sigma_a^2 + \sigma_b^2$$

Wenn wir alle obigen Gleichungen kombinieren, ergibt sich:

$$\sigma_{W_k}^2 + \sigma_a^2 + \sigma_b^2 = \sigma_{W_{k+1}}^2 + \sigma_{I_k}^2 + 2\overline{W_{k+1}} \cdot \overline{I_k}$$

$$\text{mit } I_k = (W_k + V_k)^- \text{ und } W_{k+1} = (W_k + V_k)^+$$

Wenn wir nun den Grenzwert für $k \rightarrow \infty$ bilden und annehmen, dass der Grenzwert existiert, d.h.

$$\bar{W}_k \rightarrow W; \quad \sigma_{W_k}^2 \rightarrow \sigma_W^2; \quad \bar{I}_k \rightarrow I; \quad \sigma_{I_k}^2 \rightarrow \sigma_I^2$$

dann erhalten wir

$$W = \frac{\sigma_a^2 + \sigma_b^2}{2I} - \frac{\sigma_I^2}{2I}$$

Die mittlere ungenutzte Zeit I zwischen zwei aufeinanderfolgenden Ankünften ist $I = (1-\rho)/\lambda$, so dass wir schreiben können:

$$W = \frac{\lambda(\sigma_a^2 + \sigma_b^2)}{2(1-\rho)} - \frac{\lambda\sigma_I^2}{2(1-\rho)}$$

Mit $\sigma_I^2 \geq 0$ gilt:

$$W \leq \frac{\lambda(\sigma_a^2 + \sigma_b^2)}{2(1-\rho)}$$

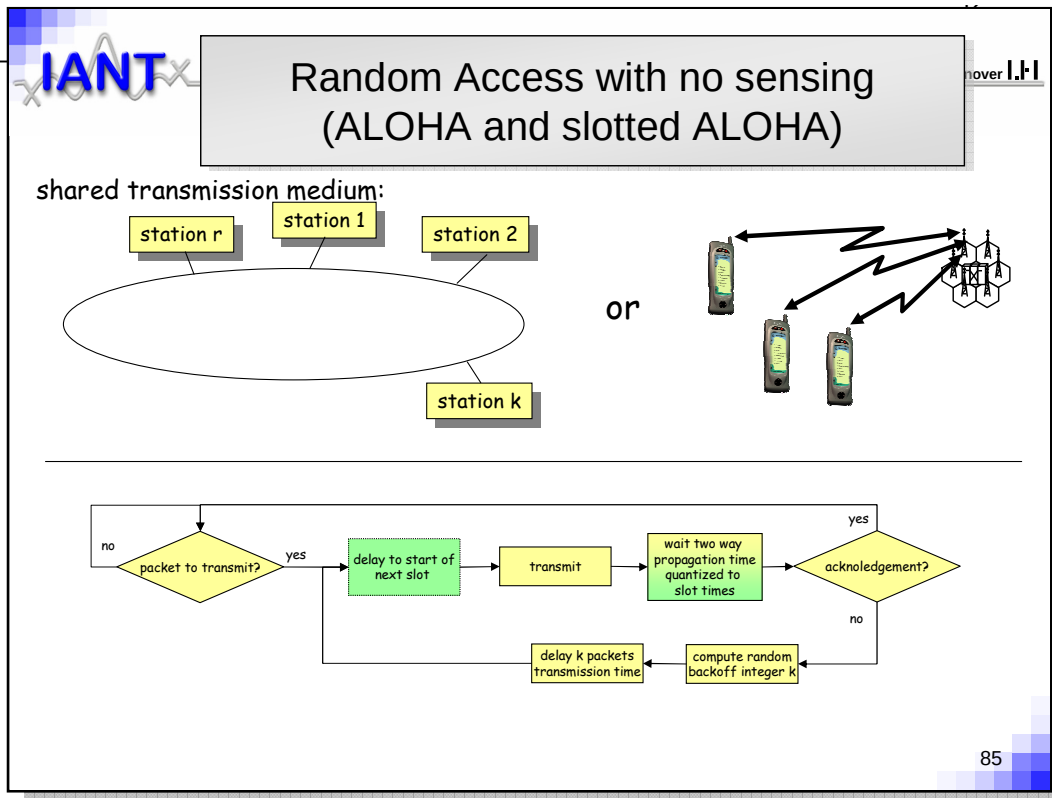
Wir geben nun ohne Herleitung eine weitere, etwas exaktere obere Grenze für G/G/1-Systeme an:

$$W \leq \frac{\lambda(\sigma_a^2 + \sigma_b^2)}{2(1-\rho)} - \frac{\lambda(1-\rho)\sigma_a^2}{2}$$

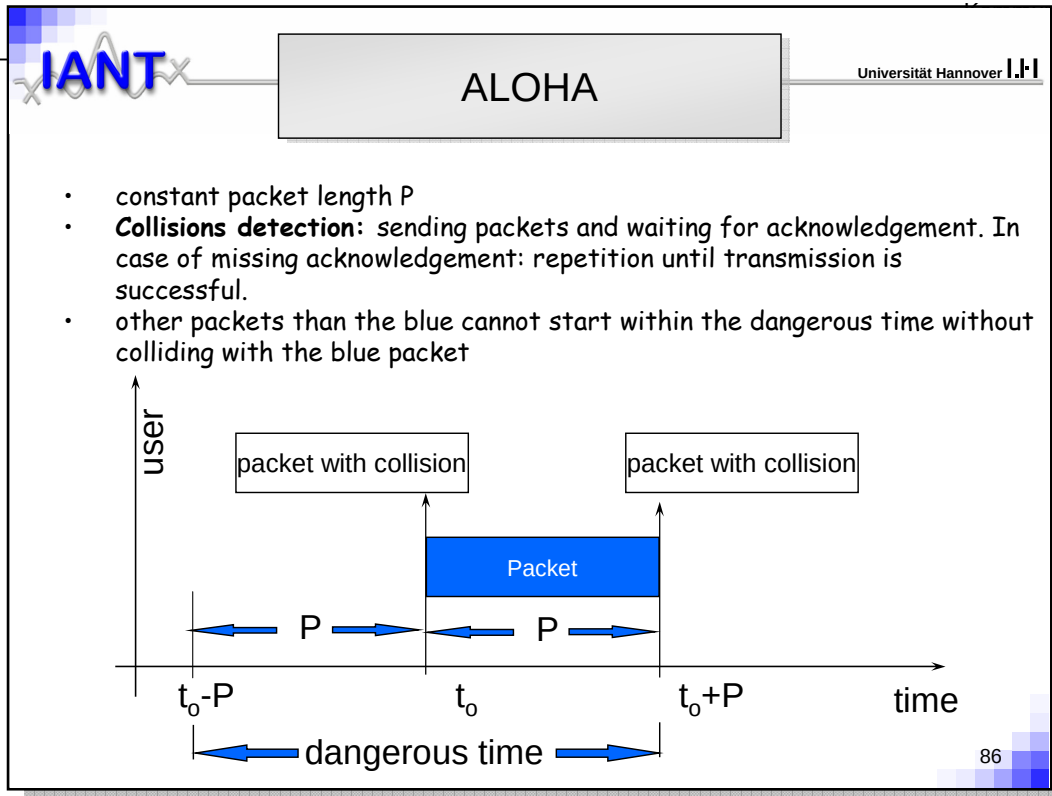
Es sein noch angemerkt, dass es in der Literatur noch weitere und exaktere Annäherungsformeln für G/G/1-Systeme gibt.

84

Leider ist es nicht auf einfache Art und Weise möglich, die Statistik der ungenutzten Zeiten allgemein anzugeben. Daher benutzt man die obige Gleichung, um eine obere Schranke für W anzugeben. Beachten Sie, dass mit zunehmender Last des Systems die ungenutzten Zeiten und damit auch ihre Varianz kleiner werden, so dass dann die oberer Ungleichung zunehmend genau wird.



The slotted ALOHA- mechanism is implemented in the registration phase of the GSM system.



Die gefährliche Zeit ist gleich der doppelten Nachrichtenlänge:

gerade etwas weniger als eine Nachrichtenlänge vor der Übertragung darf nichts von einer anderen Station gesendet werden, und natürlich nicht während der Paketübertragung selber, damit keine Kollision auftritt

kein vorheriges Abhören des Kanals!

Laufzeit zum und Bearbeitungszeit im Empfänger werden vernachlässigt

Effizienz ALOHA (1)

- während der konstanten Rahmenzeit t werden von unendlich vielen Teilnehmern genau S Rahmen nach einer Poissonverteilung erzeugt:
 - für $S > 1$ ist kein geordneter Verkehr möglich
 - Bedingung ist deshalb: $0 < S \leq 1$
- k poissonverteilte (Annahme!) Übertragungsversuche (neue und wiederholte) führen zu einer mittleren Rahmenzahl G während einer Rahmenlänge t
 - bei wenig neuen Paketen $S \sim 0$ $G \sim S$
 - Allgemein:

p_0 : Wahrscheinlichkeit, daß keine Kollision stattfindet

$$S = G \cdot p_0$$

87

Rahmenzeit: die Zeit, die benötigt wird, um einen Standardrahmen zu übertragen (Länge des Rahmens/Bitrate)

$S > 1$: es finden nur noch Kollisionen statt

Effizienz ALOHA (2)

- die Wahrscheinlichkeit, daß während eines Rahmens k Rahmen produziert werden ist poissonverteilt:

$$p_k = \frac{G'^k}{k!} \cdot e^{-G'}$$

- daraus folgt mit $G'=2G$, da "gefährliche Zeit" = $2t$

$$p_0 = \frac{(2G)^0}{0!} \cdot e^{-2G}$$

- und damit

$$p_0 = e^{-2G}$$

vgl. Verkehrstheorie NVT2:

Poissonverteilung:

Effizienz ALOHA (3)

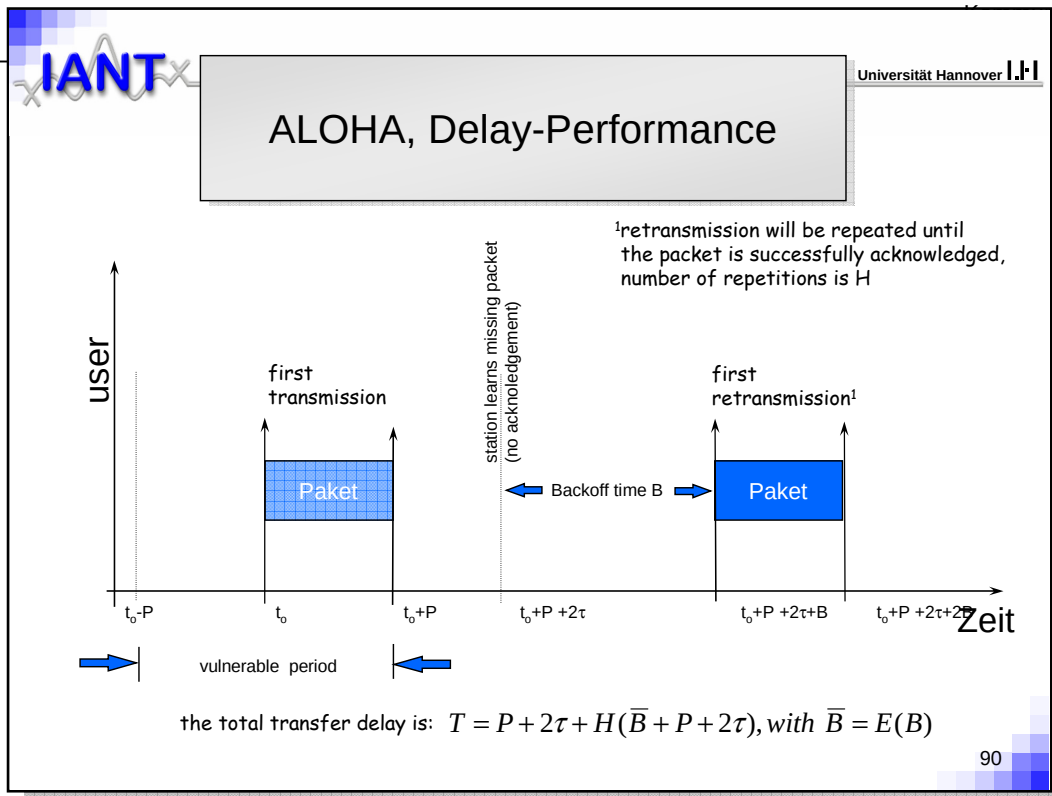
Wahrscheinlichkeit, daß keine andere Sendeanforderung
während unseres Datenpaketes vorliegt (keine Kollision)
ist

$$p_0 = e^{-2G}$$

damit folgt für S: $S = G \cdot e^{-2G}$

mit dem Maximum: $\frac{dS}{dG} = 0$ $\Rightarrow$ $G = 0,5$
 $\Rightarrow$ $S = 0,5e^{-0,5 \cdot 2} = 0,1$

die beste Performance des ALOHA liegt also bei einer
Wahrscheinlichkeit für die Sendeanforderung von 0,18.



T: Total transfer delay

P: packet length in time

τ : propagation time on transmission media

H: number of transmission retries

B: Backoff time

B_{quer} : average Backoff time

ALOHA, Delay-Performance

The total transfer delay (see last slide):

$$T = P + 2\tau + H(\bar{B} + P + 2\tau), \text{ with } \bar{B} = E(B)$$

The average number of retransmissions is the ratio of offered load G^1 and the throughput S reduced by one for the first transmission:

$$H = \frac{\bar{G}}{S} - 1 \quad 6.3.29$$

Using equation :

$$H = \frac{G}{S} - 1 = e^{2G} - 1$$

Substitution :

$$T = P + 2\tau + (e^{2G} - 1)(\bar{B} + P + 2\tau),$$

Normalizing:

$$T = 1 + 2\frac{\tau}{p} + (e^{2G} - 1)(\frac{\bar{B}}{p} + 1 + 2\frac{\tau}{p}),$$

with $\alpha = \frac{\tau}{P}$

$$T = 1 + 2\alpha + (e^{2G} - 1)\left(\frac{\bar{B}}{P} + 1 + 2\alpha\right)$$

¹ The offered load G includes the attempts to repeat

ALOHA, Delay-Performance

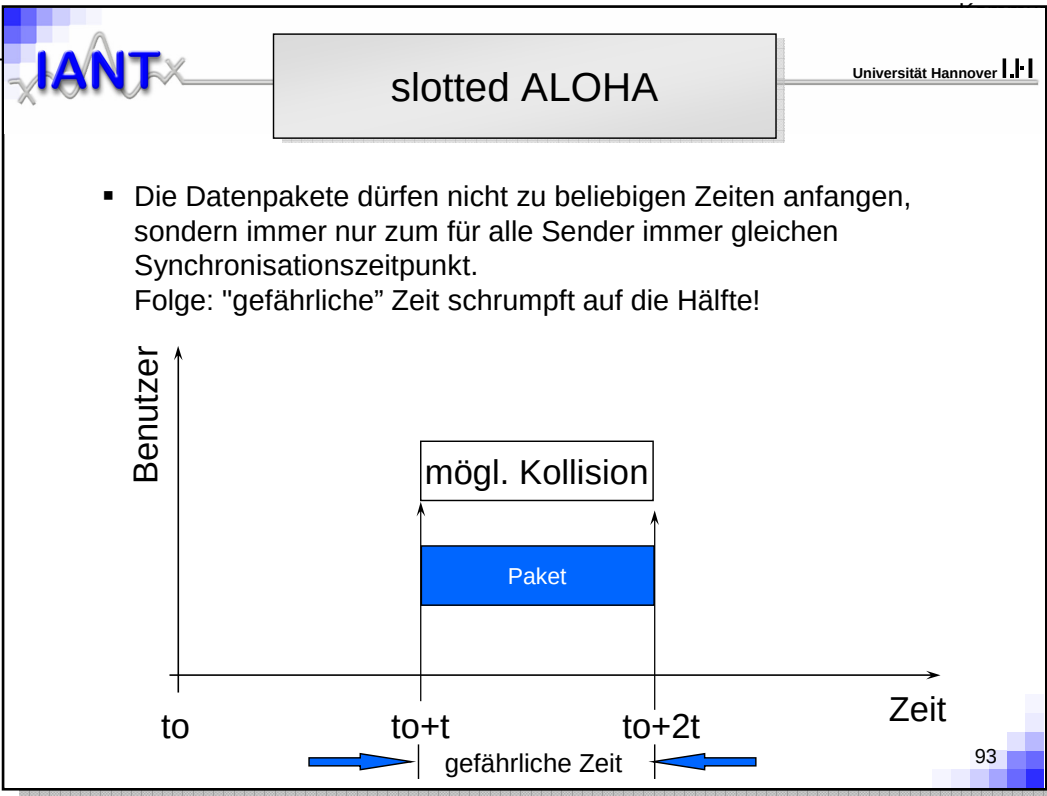
The average backoff delay still needs to be solved. The backoff time is determined with a random integer figure k , which can take the value between 0 and $K-1$. The value of K determines the amount of collisions. A bigger K produces less collisions.

The average backoff time:

$$\bar{B} = \frac{\sum_{k=0}^{K-1} k}{K} P = \frac{K-1}{2} P$$



$$\hat{T} = (1 + 2\alpha)e^{2G} + \left(\frac{k-1}{2}\right)(e^{2G} - 1)$$



Wahrscheinlichkeit, daß keine andere Sende-anforderung während unseres Datenpaketes vorliegt (keine Kollision) ist $P_0 = e^{-G}$

$$\sigma_0 = e^{-G}.$$

Damit wird die Zahl der während eines Rahmens vorliegenden Sendeansforderungen:

$$S = G \cdot e^{-G}$$

$$S = G \cdot e^{-G}$$

mit dem Maximum: $\frac{dS}{dG} = 0$

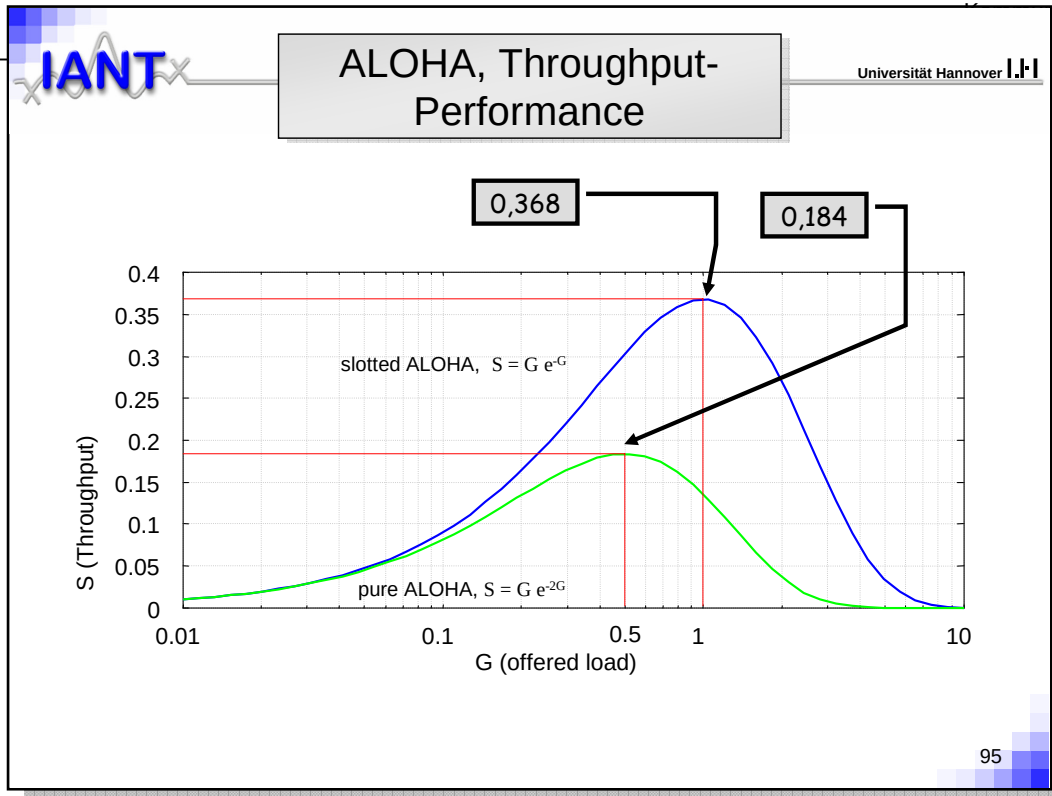


G=1



$$S = e^{-1} = 0,37$$

Die beste Performance des slotted ALOHA liegt also bei einer Wahrscheinlichkeit für die Sende-anforderung von 0,37. Gegenüber dem reinen ALOHA ist eine Verbesserung von 0,18 0,37 zu verzeichnen



#Bildgroesse:

post landscape: 10inch*7inch=25.4cm*17.78cm

eps: Alles halb so gross

Breite 12.7cm, Hoehe mitskaliert:

set autoscale xy

Anzahl der Markierungen auf den Achsen (mit Beschriftung) festlegen

set ytics 0,0.05,0.4

Abtastwertanzahl

set sample 50

Positionierung der Legende

set key 9,2.5

Logarithmische Skalierung

set logscale x

Achsenbeschriftung

set xlabel 'G (Versuche pro Paket)'

set ylabel 'S (Durchsatz pro Rahmen)' 0

set grid

set term windows color "Arial" 16

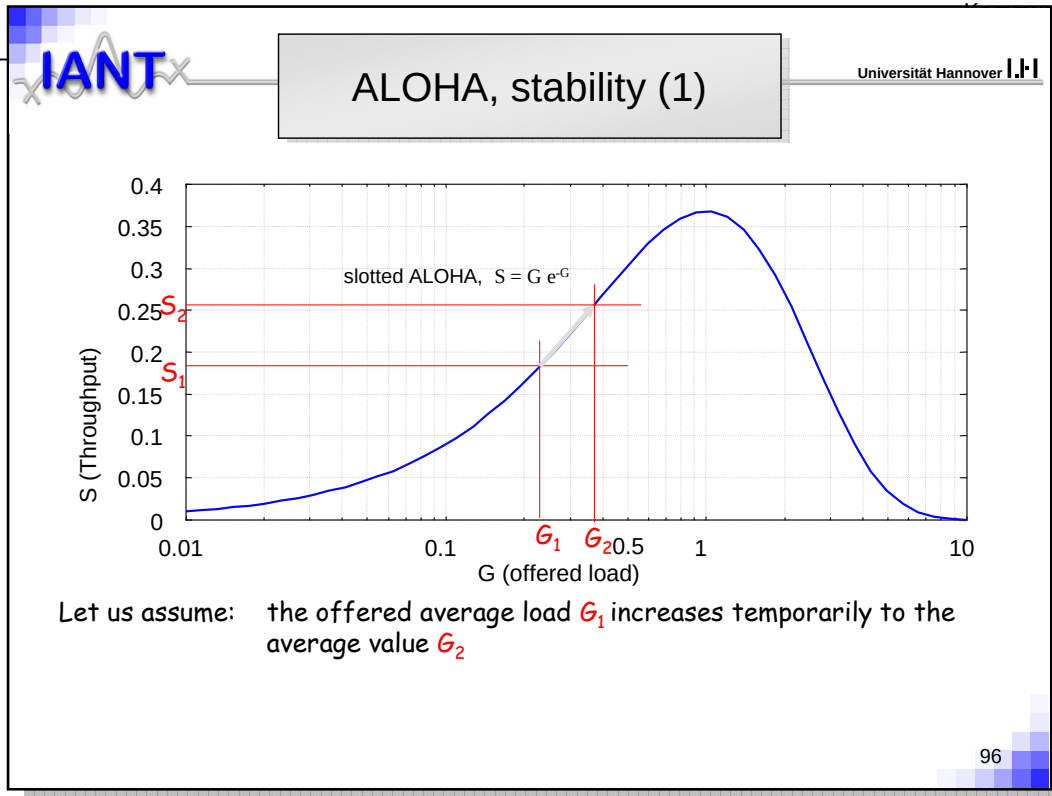
hier bitte die zuplotenden Funktionen angeben

mit Titel und Beschriftung der Achsen

und Liniensytle

plot [0.01:10] [0:0.4] x*exp(-x) title 'Unterteiltes ALOHA' with lines, \

x*exp(-2*x) title 'Reines ALOHA' with lines



#Bildgroesse:

post landscape: 10inch*7inch=25.4cm*17.78cm

eps: Alles halb so gross

Breite 12.7cm, Hoehe mitskaliert:

set autoscale xy

Anzahl der Markierungen auf den Achsen (mit Beschriftung) festlegen

set ytics 0,0.05,0.4

Abtastwertanzahl

set sample 50

Positionierung der Legende

set key 9,2.5

Logarithmische Skalierung

set logscale x

Achsenbeschriftung

set xlabel 'G (Versuche pro Paket)'

set ylabel 'S (Durchsatz pro Rahmen)' 0

set grid

set term windows color "Arial" 16

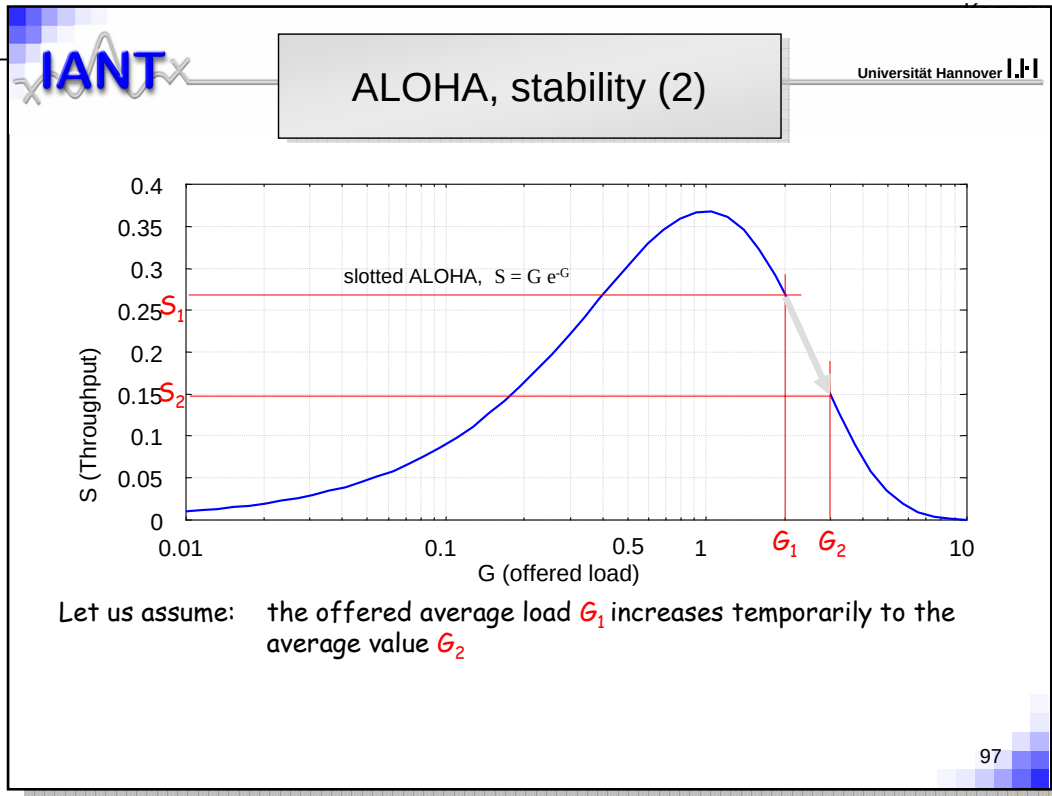
hier bitte die zuplotenden Funktionen angeben

mit Titel und Beschriftung der Achsen

und Liniensytle

plot [0.01:10] [0:0.4] x*exp(-x) title 'Unterteiltes ALOHA' with lines, \

x*exp(-2*x) title 'Reines ALOHA' with lines



#Bildgroesse:

post landscape: 10inch*7inch=25.4cm*17.78cm

eps: Alles halb so gross

Breite 12.7cm, Hoehe mitskaliert:

set autoscale xy

Anzahl der Markierungen auf den Achsen (mit Beschriftung) festlegen

set ytics 0,0.05,0.4

Abtastwertanzahl

set sample 50

Positionierung der Legende

set key 9,2.5

Logarithmische Skalierung

set logscale x

Achsenbeschriftung

set xlabel 'G (Versuche pro Paket)'

set ylabel 'S (Durchsatz pro Rahmen)' 0

set grid

set term windows color "Arial" 16

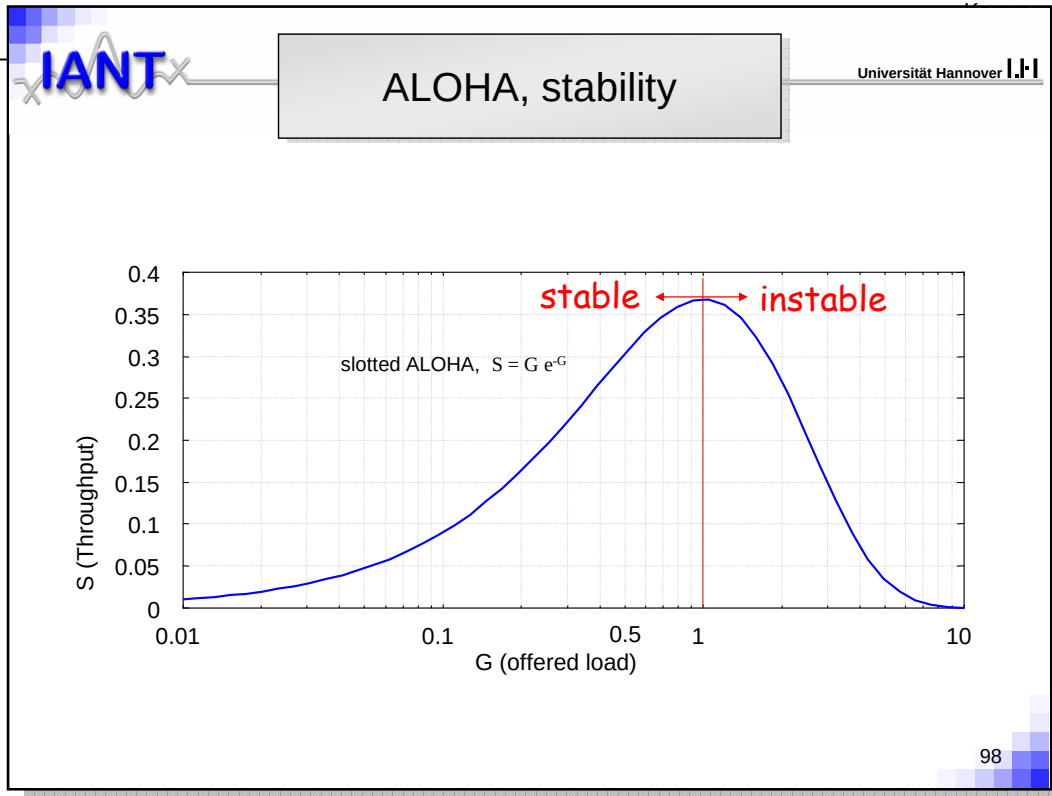
hier bitte die zuplotenden Funktionen angeben

mit Titel und Beschriftung der Achsen

und Liniensyle

plot [0.01:10] [0:0.4] x*exp(-x) title 'Unterteiltes ALOHA' with lines, \

x*exp(-2*x) title 'Reines ALOHA' with lines



#Bildgroesse:

post landscape: 10inch*7inch=25.4cm*17.78cm

eps: Alles halb so gross

Breite 12.7cm, Hoehe mitskaliert:

set autoscale xy

Anzahl der Markierungen auf den Achsen (mit Beschriftung) festlegen

set ytics 0,0.05,0.4

Abtastwertanzahl

set sample 50

Positionierung der Legende

set key 9,2.5

Logarithmische Skalierung

set logscale x

Achsenbeschriftung

set xlabel 'G (Versuche pro Paket)'

set ylabel 'S (Durchsatz pro Rahmen)' 0

set grid

set term windows color "Arial" 16

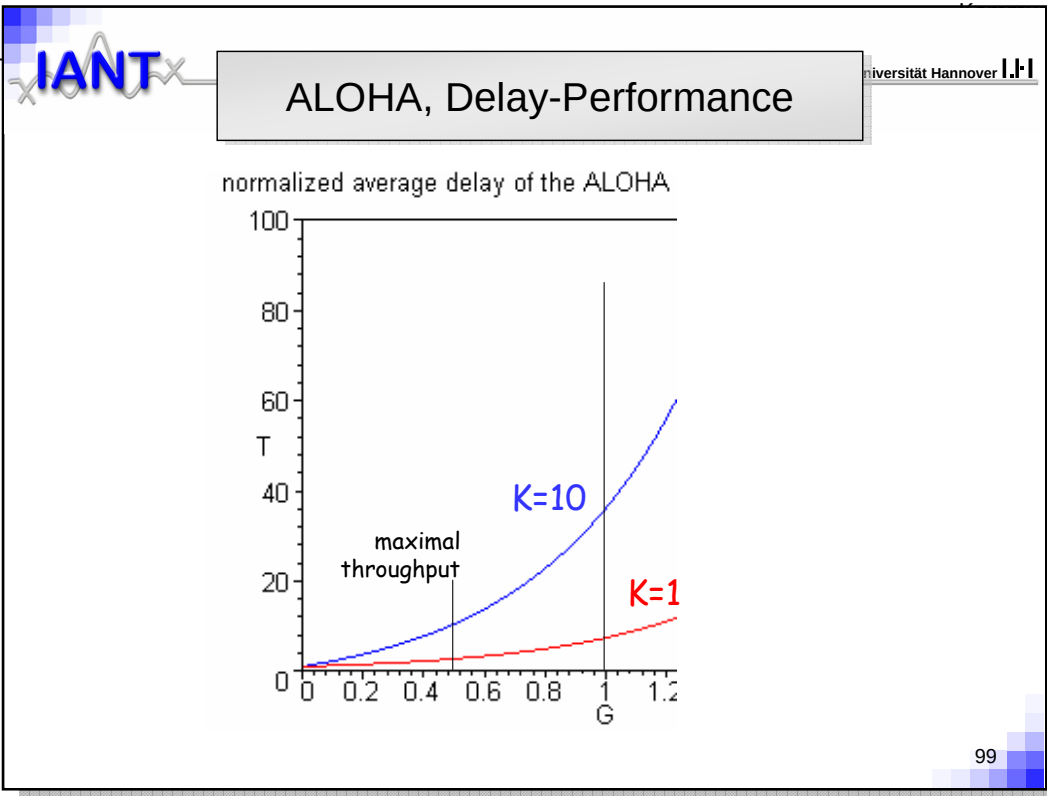
hier bitte die zuplotenden Funktionen angeben

mit Titel und Beschriftung der Achsen

und Liniensytle

plot [0.01:10] [0:0.4] x*exp(-x) title 'Unterteiltes ALOHA' with lines, \

x*exp(-2*x) title 'Reines ALOHA' with lines



Maple Skript:

with(plots):

```
> setoptions(title=` normalized average delay of the ALOHA MAC over offered traffic `, style=line,
axes=BOXED);
```

```
> alpha:=0;
```

```
>
```

```
> K:=1;
```

```
delay1:=plot((1+2*alpha)*exp(2*G)+((K-1)/2)*(exp(2*G)-1),G=0..2,T=0..100,colour=red);
```

```
> K:=10;
```

```
delay2:=plot((1+2*alpha)*exp(2*G)+((K-1)/2)*(exp(2*G)-1),G=0..2,T=0..100,colour=blue);
```

```
> display(delay1,delay2);
```

Slotted Aloha Durchsatz: Definitionen

M - Anzahl der Stationen

σ - Wahrscheinlichkeit für die Übertragung
einer Station ohne vorausgegangene Kollision

p - Wahrscheinlichkeit für die Übertragung
einer Station mit vorausgegangene Kollision

λ - Mittlere Ankunftsrate

ρ - Nutzungskoeffizient = λ / μ , wobei $1 / \mu$ die mittlere Bediendauer ist

N^t - Zustandvariable für die Anzahl kollidierter Pakete

Definitionen

Für die Wahrscheinlichkeit eines Senderversuchs für ein backlogged Packets im k -ten Timeslot ergibt sich eine geometrische Verteilung bei festem p :

$$P(K = k) = p \cdot (1-p)^{k-1}$$

Bei einem poissonverteilten Ankunftsprozess ist die Wahrscheinlichkeit dafür, dass eine Station keine Daten zu senden hat $e^{-\frac{\lambda}{M}}$ und es gilt:

$$\sigma = 1 - e^{-\frac{\lambda}{M}}$$

Unter der Voraussetzung, dass σ und M zeit-invariant sind,
ist N^t ein Markov-Prozess.

Übergangswahrscheinlichkeiten für N^t mit $i = 0, 1, 2, \dots, M$:

$$p_{ij} = \text{Prob}[N^{t+1} = j | N^t = i]$$

$$= \begin{cases} 0 & j \leq i-2 \\ ip(1-p)^{i-1}(1-\sigma)^{M-i} & j = i-1 \\ (1-p)^i(M-i)\sigma(1-\sigma)^{M-i-1} + [1-ip(1-p)^{i-1}](1-\sigma)^{M-i} & j = i \\ (M-i)\sigma(1-\sigma)^{M-i-1}[1-(1-p)^i] & j = i+1 \\ \binom{M-i}{j-i}\sigma^{j-i}(1-\sigma)^{M-j} & j \geq i+2 \end{cases}$$

102

Markov-Prozess: Grenzwert

Mit $M \rightarrow \infty$ und $\sigma \rightarrow 0$, gilt $M\sigma = S = \text{const. (channel input rate)}$.

Da $e^{-\tilde{S}} = \lim_{M \rightarrow \infty, \sigma \rightarrow 0} (1-\sigma)^{\frac{\sigma}{M}}$ folgt:

$$p_{ij} = \text{Prob}[N^{t+1} = j | N^t = i]$$

$$= \begin{cases} 0 & j \leq i-2 \\ ip(1-p)^{i-1}e^{-S} & j = i-1 \\ (1-p)^i Se^{-S} + [1-ip(1-p)^{i-1}]e^{-S} & j = i \\ Se^{-S}[1-(1-p)^i] & j = i+1 \\ \frac{S^{j-i}}{(j-i)!}e^{-S} & j \geq i+2 \end{cases}$$

Durchsatz

Sei D_n die erwartete Abweichung der Anzahl kollidierter Stationen im nächsten Slot im Zustand $N^t = n$, dann gilt :

$$D_n = (M - n)\sigma - S_{out}$$

mit S_{out} für die Wahrscheinlichkeit eines erfolgreichen Sendeversuchs.

Durchsatz

Geht man von einer Anzahl von $N^t = n$ Paketen aus, so lässt sich der erwartete Durchsatz $S_{out}(n, \sigma)$ durch die Wahrscheinlichkeit für die Übertragung eines einzelnen Pakets ausdrücken :

$$S_{out}(n, \sigma) = (1-p)^n (M-n)\sigma(1-\sigma)^{M-n-1} + np(1-p)^{n-1}(1-\sigma)^{M-n}$$

Für den Grenzwertfall $M \rightarrow \infty$ und $\sigma \rightarrow 0$ gilt dann :

$$S_{out}(n, \sigma) = (1-p)^n S e^{-S} + np(1-p)^{n-1} e^{-S}$$

Durchsatz

Mit $G(n)$ als Anzahl aller Sendeveruche
in einem Slot im Zustand n gilt dann :

$$G(n) = (M - n)\sigma + np$$

Unter Annahme dass $p \rightarrow 0$ und $\sigma \rightarrow 0$ mit $S \approx G$
lässt sich der Ausdruck

$$S_{out}(n, \sigma) = (1 - p)^n S e^{-S} + np(1 - p)^{n-1} e^{-S}$$

zu

$$S_{out} \approx G e^{-G}$$

umformulieren.

Slotted Aloha Delay-Performance: Definitionen

D mittlere Verzögerungszeit in Slots

$\bar{B}$ mittlere Anzahl Slots für Backhoff

$$\bar{B} = \frac{\sum_{k=0}^{K-1} k}{K} = \frac{K-1}{2}$$

WT Quittierungszeit bei erfolgreicher Übertragung

H Wiederholungen bei Kollision

Slotted Aloha Delay-Performance

Für die Anzahl der Slots bis zur erfolgreichen Übertragung eines Pakets gilt :

$$D = WT + 1 + R + H \cdot (WT + 1 + \frac{K-1}{2})$$

$$\text{mit } R = \frac{1}{2}$$

Mit der Wahrscheinlichkeit p_i , dass genau i Wiederholungen zu einer erfolgreichen Übertragung führen :

$$p_i = (1-\sigma)p(1-p)^{i-1}$$

berechnet sich H zu

$$H = \sum_{i=1}^{\infty} i \cdot p_i = \frac{1-\sigma}{p}$$

Slotted Aloha Delay-Performance

Es resultiert :

$$D = WT + 1 + \frac{1}{2} + \frac{1-\sigma}{p} \cdot \left(WT + 1 + \frac{K-1}{2} \right)$$

