Prozeßrechner: statische und terminorientierte Prioritäten im Vergleich*

Process-control computers: static and due-date oriented priorities compared

Elektron. Rechenanl. 20 (1978), H. 6, S. 283–289
Manuskripteingang: 16. Juni 1978

von B. WALKE
AEG-TELEFUNKEN
Forschungsinstitut Ulm

Für den ereignisgesteuerten Betrieb eines Prozeßrechners wird unterstellt, daß unterschiedlich wichtige Anforderungen auftreten. Das wird einerseits durch externe statische, andererseits durch wartezeitabhängige (terminorientierte) Prioritäten berücksichtigt. Ergebnisse für die mittleren Wartezeiten und Wartezeitverteilungen unterschiedlich wichtiger Anforderungen werden für beide Abfertigungsstrategien angegeben und verglichen. Es wird demonstriert, daß statische Prioritäten ein Sonderfall terminorientierter Prioritäten sind.

We consider the event driven operation of a process-control computer system and assume service requests having different urgencies. This is taken into account by assigning either static or due-date oriented priorities to such requests. Results are presented and compared for the mean wait and the waiting-time distribution function of different urgent requests for both types of priority assignments. Static priorities are demonstrated to be included as a limiting case of due-date oriented priorities.

1. Einleitung

Bei Prozeßrechnern (Realzeitrechnern) wird besonders großer Wert auf das „Echtzeitverhalten" gelegt. Kenngrößen hierzu sind – vgl. DIN 44300, S. 15 – 1. die *Reaktionszeit* (Zeitspanne zwischen dem Ende des Eintreffens einer Anforderung und dem Beginn der Bearbeitung), 2. die Bediendauer (Zeitspanne zwischen Beginn und Ende der als nicht-unterbrechbar unterstellten Bearbeitung einer Anforderung), sowie bisher nicht genormte Größen, z. B. die *Termineinhaltung* (Wahrscheinlichkeit für die Einhaltung von Terminen). Während die Bediendauer allein von der Verarbeitungsgeschwindigkeit der Zentraleinheit des Prozeßrechners abhängt, können die Reaktionszeit (im folgenden *Wartezeit* genannt) und die Termineinhaltung durch strategische Maßnahmen bei der Festlegung der Abfertigung durch die Zentraleinheit stark beeinflußt werden. Das wird im folgenden anhand von Beispielen erläutert. Dabei wird nicht nur der Mittelwert der Wartezeit, sondern auch ihre Verteilungsfunktion betrachtet werden.

Aus der praktischen Anwendung von Prozeßrechnern ist geläufig, daß das angestrebte Echtzeitverhalten meist durch Anwendung externer statischer Prioritäten erreichbar ist. Dabei teilt man unterschiedlich wichtigen Anforderungen verschiedene Prioritäten zu und wählt zu Entscheidungszeitpunkten jeweils die längst wartende Anforderung mit höchster Priorität (kleinster Prioritätszahl) zur Bearbeitung aus. Von nicht-unterbrechender Priorität spricht man, wenn ein Entscheidungszeitpunkt nur durch Fertigstellung einer Anforderung durch die Zentraleinheit erreicht werden kann. Unterbrechende Priorität liegt vor, wenn jede Ankunft einer Anforderung bei der Zentraleinheit einen Entscheidungszeitpunkt zur Folge hat.

Aus der Tatsache, daß man mit Hilfe statischer Prioritäten ein angestrebtes Echtzeitverhalten erreichen kann, darf nicht unmittelbar gefolgert werden, daß alle Probleme damit gelöst sind. Unter dem Gesichtspunkt des wirtschaftlichen Einsatzes von Prozeßrechnern sind solche Lösungen u. U. sehr unbefriedigend. Statische Prioritäten sind ein zu wenig flexibles Werkzeug für eine zeitgerechte Ablaufsteuerung. In manchen denkbaren Belastungsfällen eines Prozeßrechners darf die Zentraleinheit nur zu einem geringen Prozentsatz (z. B. 20%) ausgelastet werden, um das geforderte Zeitverhalten mit Hilfe statischer Prioritäten zu erreichen [1].

Besonders unter dem Gesichtspunkt der optimalen Termineinhaltung „ist es bei der Erstellung solcher Systeme notwendig, starken Einfluß auf die Warteschlangenstrategien nehmen zu können. In den meisten Fällen läuft dies auf eine dynamische und warteschlangen-spezifische Prioritäten-Vergabe hinaus." (VDI/VDE Richtlinie 3554, Entwurf 1976.) Eine solche terminorientierte Strategie mit dynamischen Prioritäten, die im Grenzfall in die bekannte Strategie mit statischen Prioritäten übergeht, wird hier anhand von Ergebnissen besprochen. Die gezeigten Resultate sind in guter Näherung berechenbar (vgl. [1], Kap. 12) und durch Simulation bestätigt worden. Der Schwerpunkt liegt in diesem Beitrag auf dem Vergleich von statischen und terminorientierten Prioritäten bezüglich der erreichten Wartezeit. Die Herleitung der Näherungsrechnungen findet man in [2] und [3].

2. Modellbeschreibung

Wir betrachten das einfache Modell entsprechend Bild 1. Dort unterscheiden wir N Typen von Anforderungen, die je gemäß einem Poisson-Prozeß (mit negativ exponentiell verteilten Ankunftsabständen) mit der Rate λ_i Anforderungen pro Zeiteinheit ($1 \leq i \leq N$) ankommen und sich an das Ende der Warteschlange i anstellen. Diese Annahme eines Poisson-Prozes-

* Die Arbeit ist zu 50% aus Mitteln des Bundesministeriums für Forschung und Technologie (3. DV-Programm, Teil B. I. 1.3) gefördert worden.

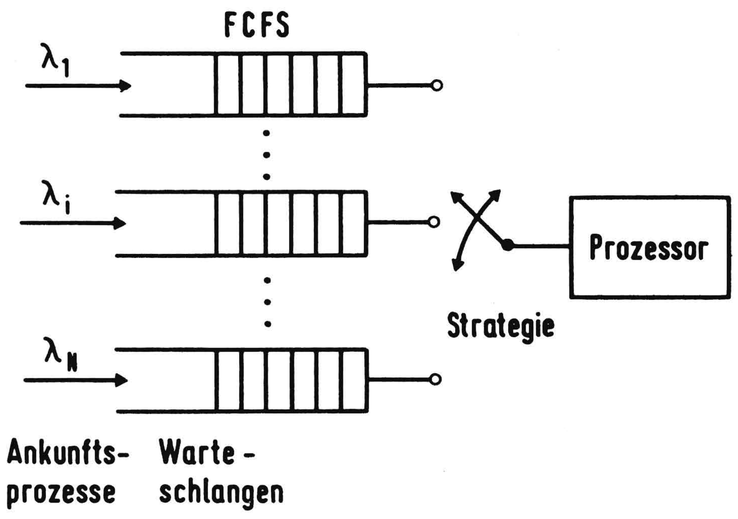

Bild 1. Modell eines Prozeßrechners mit $N=4$ Typen von Anforderungen und je einer unbegrenzten Warteschlange. Die Abfertigungsreihenfolge innerhalb jeder Warteschlange entspricht der Reihenfolge des Eintreffens (FCFS).

ses ist immer dann gerechtfertigt, wenn viele unabhängige Quellen spontan Anforderungen erzeugen. Die Abfertigung innerhalb jeder Warteschlange entspricht der Reihenfolge des Eintreffens. Die Bediendauer b_i einer Anforderung vom Typ i sei nicht bekannt, sondern durch eine beliebige Verteilungsfunktion $F_i(t)=P(b_i \leq t)$ beschrieben. Der Prozessor (Zentraleinheit des Prozeßrechners) wählt nach Fertigstellung einer Anforderung jeweils diejenige zur Bearbeitung aus, welche dann die höchste Priorität hat. Wir betrachten also den Fall nicht-unterbrechender Prioritäten. Der Fall unterbrechender Prioritäten könnte ähnlich behandelt werden, vgl. [1], Kap. 12.2. Drei Abfertigungsstrategien werden betrachtet:

- NONPRE: (non-preemptive priority):

  Eine Anforderung vom Typ i hat die externe statische Priorität i, wobei die Priorität um so höher ist, je kleiner die Zahl i ist. Diese Strategie wurde erstmals in [4] behandelt und wird auch „head of the line discipline" genannt.

- RU (relative urgency):

  Die Priorität $q_i(t)$ einer Anforderung vom Typ i zum Zeitpunkt t gehorcht dem Gesetz

$$q_i(t)=\omega_i-w_i(t), \tag{1}$$

wobei ω_i die – für alle Anforderungen vom Typ i gleichgroße – Wichtigkeit und $w_i(t)$ die zum Zeitpunkt t aufgelaufene Wartezeit ist. Ohne Einschränkung der Allgemeinheit kann man festlegen

$$0 \leq \omega_1 \leq \omega_2 \leq \ldots \omega_i \ldots \leq \omega_N\,. \tag{2}$$

Wie bei Prioritäten bedeuten kleine Werte von ω_i hohe Wichtigkeiten.

Unterschiedliche Wichtigkeiten haben zur Folge, daß gleichzeitig wartende Anforderungen verschiedenen Typs unterschiedlich lange bis zum Bearbeitungsbeginn warten müssen. Zwei Anforderungen vom Typ k bzw. j $(k<j)$ haben nur dann gleiche Priorität, wenn die Anforderungen die Zeiten w_k bzw. w_j gewartet haben und $w_j-w_k=\omega_j-\omega_k$ ist. Ist $w_j-w_k>\omega_j-\omega_k$, dann ist und bleibt die Priorität der Anforderung vom Typ j am höchsten. Die Differenz der Wichtigkeiten $\omega_j-\omega_k$ zweier Typen von Anforderungen gibt die Wartezeit des unwichtigeren Typs j an, die zugunsten des Typs k in Kauf genommen wird. Man kann die Wichtigkeit ω_i als freiwillige Wartezeit einer Anforderung ansehen, so daß sich ihr Termin T_i für den Bearbeitungsbeginn aus Ankunftszeitpunkt t_i und Wichtigkeit ω_i zu $T_i=t_i+\omega_i$ ergibt. Eine unwichtigere Anforderung vom Typ j muß gegenüber einer vom Typ k genau $\omega_j-\omega_k$ Zeiteinheiten länger warten, um den gleichen Termin (d. h. gleiche dynamische Priorität) zu haben. Im Rahmen der folgenden Betrachtungen ist es gleichgültig, ob man den Bearbeitungsbeginn T_i oder den Fertigstellungszeitpunkt $\tau_i = t_i + \omega_i + b_i$ als Termin ansieht. Von dieser Interpretation rührt der Name terminorientierte Abfertigung her. Diese Strategie wurde erstmals in [5] behandelt.

Die dynamische Priorität wächst für jede Anforderung linear mit der Zeit t. Die Steigung ist, unabhängig von der Wichtigkeit ω_i, für alle Anforderungen gleich groß. Bild 2 zeigt ein Beispiel mit zwei Wichtigkeiten $\omega_1=2$ und $\omega_2=4$ und gleichen Bediendauern (je 3 Zeiteinheiten). An der Ordinate ist die dynamische Priorität und an der Abszisse die Zeit dargestellt. Beim Eintreffen einer Anforderung hat sie die Priorität ω_i, weil $w_i(t)=0$ ist. Muß eine Anforderung warten, dann nimmt dabei ihre Priorität zu. Zu den

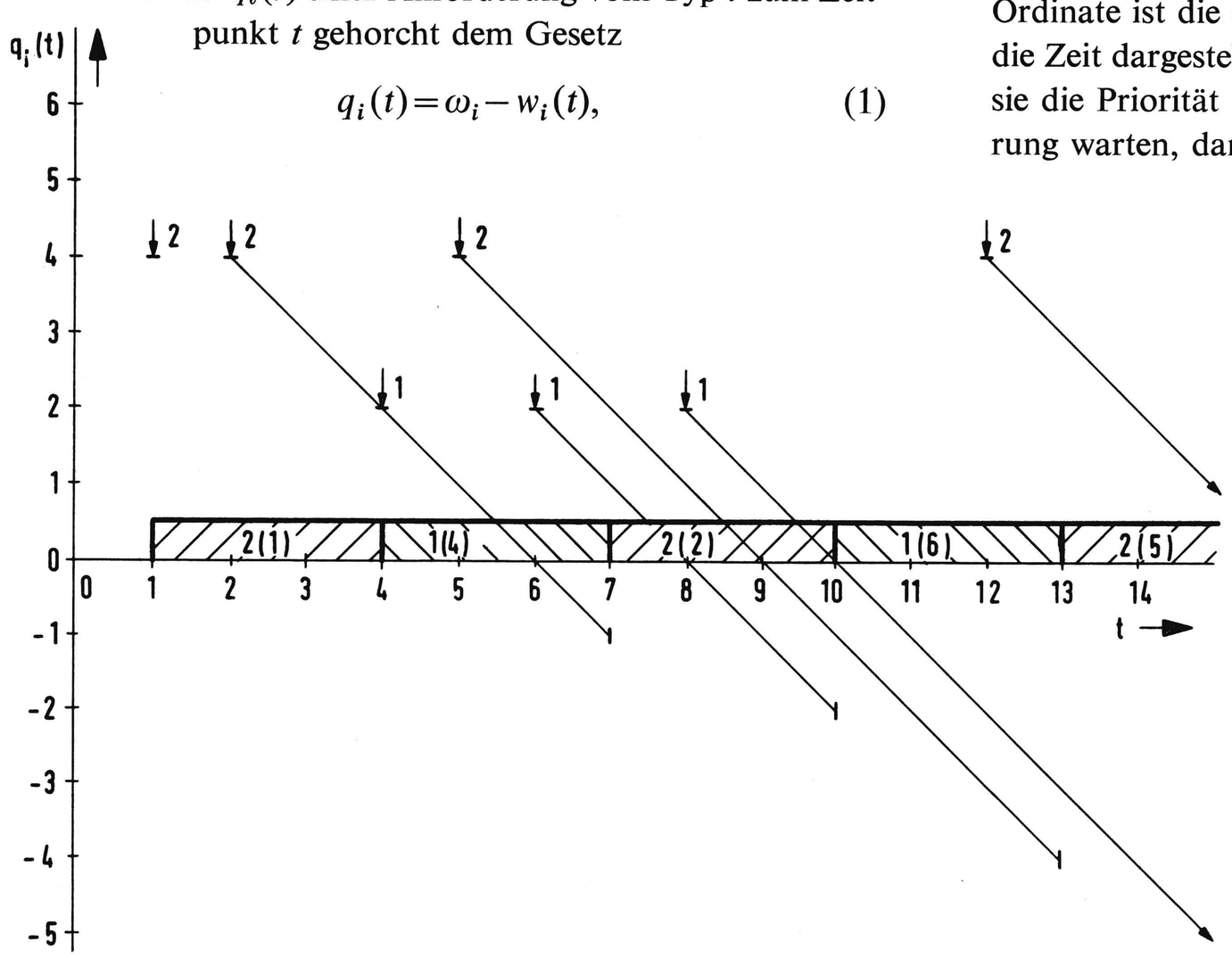

Bild 2. Beispiel für den Verlauf der dynamischen Prioritäten $q_i(t)$ in Abhängigkeit von der Zeit t für ein Modell mit $i=2$ Typen von Anforderungen und den Wichtigkeiten $\omega_1=2$, $\omega_2=4$. Die Bediendauern sind gleich und je 3 Zeiteinheiten lang. Eintreffzeitpunkte sind markiert $\downarrow i$. Die Prozessorarbeit an einer zum Zeitpunkt t eingetroffenen Anforderung vom Typ i ist schraffiert angegeben mit dem Kennzeichen $i(t)$.

Zeitpunkten $t=1, 2, 4, 5, 6, 8$ und 12 treffen Anforderungen ein. Der individuelle Prioritätsverlauf ist, falls sie warten müssen, dargestellt. Nach Fertigbearbeitung durch den Prozessor wird jeweils als nächste die Anforderung mit höchster dynamischer Priorität ausgewählt. Die Prozessorbelegung ist durch einen schraffierten Balken dargestellt. Eingetragen ist, an welcher Anforderung jeweils gearbeitet wird: 1(4) bedeutet Arbeit an einer Anforderung $i=1$, die zum Zeitpunkt $t=4$ eingetroffen ist. Im Beispiel sind zum Zeitpunkt $t=4$ die dynamischen Prioritäten der Anforderungen 2(2) und 1(4) gleich groß. Dann wird bei der Strategie RU immer die wichtigere Anforderung 1(4) bevorzugt. Beim Nulldurchgang der dynamischen Priorität ist der jeweilige Termin abgelaufen.

- FCFS: Die Abfertigung erfolgt in der Reihenfolge des Eintreffens.

Bezeichnet man die Varianz der Wichtigkeitszahlen ω_i mit Var (ω), dann werden als Grenzfälle der Strategie RU für Var $(\omega) \to \infty$ die Strategie NONPRE bzw. für Var $(\omega) \to 0$ die Strategie FCFS approximiert. Es reicht also eigentlich aus, nur die Strategie RU einschließlich ihrer Grenzfälle zu betrachten, um Aussagen über alle drei Strategien zu erhalten.

Als charakteristische Größen unseres Modells nach Bild 1 betrachten wir die mittlere Wartezeit W_i von Anforderungen des Typs i, die Verteilungsfunktion (abgekürzt VF) $W_i(t)=P(w_i \leq t)$ ihrer Wartezeiten w_i und die Terminüberschreitungs-Wahrscheinlichkeit $P(w_i > \omega_i)$ bei allen drei Strategien.

3. Ergebnisse für charakteristische Größen

Definiert man

$$W_i(t), F_i(t)$$

als Wartezeit- bzw. Bediendauer-VFn von Anforderungen des Typs i,

$$W_i^{(r)}, \beta_i^{(r)}$$

als r-tes Moment der Wartezeit- bzw. Bediendauer-VF, mit $W_i = W_i^{(1)}$ und $\beta_i = \beta_i^{(1)}$,

$$\varrho_{\leq k} = \sum_{i=1}^{k} \varrho_i$$

als Auslastung des Prozessors durch Anforderungen vom Typ 1 bis k (zusammengefaßtes Angebot),

$$\varrho_i = \lambda_i \beta_i$$

als Auslastung des Prozessors durch Anforderungen vom Typ i,

$$\bar{\omega} = \varrho_{\leq N}^{-1} \sum_{i=1}^{N} \varrho_i \omega_i$$

als Erwartungswert der gewichteten*) Wichtigkeiten,

$$\mathrm{Var}(\omega) = \varrho_{\leq N}^{-1} \sum_{i=1}^{N} \varrho_i \omega_i^2 - \bar{\omega}^2$$

als Varianz der gewichteten*) Wichtigkeiten,

$$\lambda_{\leq N} = \sum_{i=1}^{N} \lambda_i$$

als gesamte Ankunftsrate,

$$\beta_{\leq N}^{(r)} = \lambda_{\leq N}^{-1} \sum_{i=1}^{N} \lambda_i \beta_i^{(r)}$$

als r-tes Moment der gemeinsamen Bediendauer-VF $F_{\leq N}(t)$,

$$F_{\leq N}(t) = \lambda_{\leq N}^{-1} \sum_{i=1}^{N} \lambda_i F_i(t)$$

als durch Gewichtung der Bediendauer-VFn $F_i(t)$ gebildete gemeinsame Bediendauer-VF,

dann kann man bekannte Ergebnisse (vgl. [1]) für die ersten beiden Momente der Wartezeit-VF schreiben bei der Strategie

FCFS:

$$W_{\mathrm{FCFS}} = \frac{\lambda_{\leq N} \beta_{\leq N}^{(2)}}{2(1-\varrho_{\leq N})} \tag{3}$$

$$W_{\mathrm{FCFS}}^{(2)} = 2\, W_{\mathrm{FCFS}}^2 + \frac{\lambda_{\leq N} \beta_{\leq N}^{(3)}}{3(1-\varrho_{\leq N})} \tag{4}$$

NONPRE:

$$W_{i\mathrm{NONPRE}} = \frac{\lambda_{\leq N} \beta_{\leq N}^{(2)}}{2(1-\varrho_{\leq i})(1-\varrho_{\leq i-1})} \tag{5}$$

$$W_{i\mathrm{NONPRE}}^{(2)} = \frac{\lambda_{\leq N} \beta_{\leq N}^{(3)}}{3(1-\varrho_{\leq i})(1-\varrho_{\leq i-1})^2} + \frac{\lambda_{\leq N} \beta_{\leq N}^{(2)} \lambda_{\leq i} \beta_{\leq i}^{(2)}}{2(1-\varrho_{\leq i})^2 (1-\varrho_{\leq i-1})^2} + \frac{\lambda_{\leq N} \beta_{\leq N}^{(2)} \lambda_{\leq i-1} \beta_{\leq i-1}^{(2)}}{2(1-\varrho_{\leq i})(1-\varrho_{\leq i-1})^3}. \tag{6}$$

Für die Strategie RU sind nur obere und untere Grenzen bzw. eine gute Näherung für die mittlere Wartezeit bekannt. Diese Näherung verwenden wir hier

RU:

$$W_i = \frac{\lambda_{\leq N} \beta_{\leq N}^{(2)}/2 + \sum_{j=1}^{i} \varrho_j W_j + \sum_{j=i+1}^{N} \varrho_j W_j \exp\left[\varrho_{\leq N}(\omega_i - \omega_j)/W_j\right]}{1 - \sum_{j=1}^{i-1} \varrho_j \left(1 - \exp\left[\varrho_{\leq N}(\omega_j - \omega_i)/W_i\right]\right)}$$

$$\text{für} \quad 0 \leq \varrho_{\leq N} < 1. \tag{7}$$

In diesem Ausdruck treten nur Differenzen von Wichtigkeiten auf. Das bedeutet, bezüglich der mittleren Wartezeiten sind folgende Sätze von Wichtigkeiten äquivalent ({...} = ein Satz)

$$\{\omega_1, \omega_2, \ldots \omega_i, \ldots \omega_N\} \approx \{\omega_1 + r, \omega_2 + r, \ldots \omega_i + r, \ldots \omega_N + r\} \tag{8}$$

mit r beliebig reell unter Beachtung von Gl. (2).

Außerdem gilt näherungsweise für große Wartezeiten [3] der Strategie RU die Wartezeit-VF

$$W_i(t) = 1 - (1-p)\, \mathrm{e}^{-\mu(\bar{\omega} - \omega_i)}\, \mathrm{e}^{-\mu t}, \tag{9}$$

mit

$$\mu = 2/[W_{\mathrm{FCFS}}(1 + C_w^2)],$$

*) Man kommt zu diesen Ergebnissen, wenn man von einer Verteilung der Wichtigkeitszahlen mit Sprungstellen bei ω_i und jeweiliger Sprunghöhe $\varrho_i/\varrho_{\leq N}$ ausgeht.

so daß sich mit

$$C_W^2 = W_{\mathrm{FCFS}}^{(2)} / W_{\mathrm{FCFS}}^2 - 1$$

ergibt

$$\mu = 2\, W_{\mathrm{FCFS}} / W_{\mathrm{FCFS}}^{(2)} .$$

Für kleine Wartezeiten gilt näherungsweise [2] die Wartezeit-VF

$$W_i(t) = 1 - \varrho_{\leq N}\, \mathrm{e}^{-\varrho_{\leq N} t / W_i} . \tag{10}$$

4. Vergleich und Diskussion charakteristischer Größen

Die in Abschnitt 3 angegebenen Ausdrücke für die charakteristischen Größen sind wenig anschaulich. Außerdem fehlen Resultate zur exakten Beschreibung der Wartezeit-VFn. Wir betrachten ein Belastungsbeispiel und verwenden neben exakten auch Simulations-Ergebnisse für dieses

Beispiel: $N = 4$ Typen von Anforderungen,

konstante aber verschiedene Bediendauern der Länge β_i, gleiche Prozessorbelastung ϱ_i durch jeden Typ von Anforderungen, unterschiedliche Wichtigkeiten ω_i, die jeweils extra angegeben werden, vgl. z. B. Bild 3.

i	1	2	3	4
β_i	1	3	5	7

$\varrho_i = \varrho_{\leq N} / N$.

Dafür berechnet man mit

$$\beta_i^{(r)} = \beta_i^r \quad \text{und mit} \quad \sum_{i=1}^{N} = \sum ,$$

$$\lambda_{\leq N} = \sum \lambda_i = \sum \varrho_i / \beta_i ,$$

bzw. mit

$$\varrho_i = \varrho_{\leq N} / N , \quad \lambda_{\leq N} = \varrho_{\leq N} / N \sum 1/\beta_i .$$

Außerdem ist

$$\sum \lambda_i \beta_i^{(r)} = \sum \lambda_i \beta_i^r = \sum \varrho_i \beta_i^{r-1} = \varrho_{\leq N} / N \sum \beta_i^{r-1} ,$$

so daß

$$\beta_{\leq N}^{(r)} = \sum \beta_i^{r-1} / \sum 1/\beta_i$$

wird. Man erhält dann

$$\beta_{\leq N} = 2{,}39 , \quad \beta_{\leq N}^{(2)} = 9{,}43 .$$

Bei der Strategie NONPRE ist i die statische Priorität. Die mittlere Wartezeit dieser Strategie berechnet man nach Gl. (5) für dieses Beispiel – mit den eben bestimmten Ausdrücken – und

$$\varrho_{\leq k} = k \varrho_{\leq N} / N$$

zu

$$W_{i\mathrm{NONPRE}} = \frac{\varrho_{\leq N} / N \sum \beta_i}{2(1 - i\varrho_{\leq N}/N)(1 - (i-1)\,\varrho_{\leq N}/N)} .$$

Die zugehörigen Kurven ($i = 1$ bis 4) sind in Bild 3 dargestellt. Bezüglich dieser Strategie sind die Bilder 3a) bis 3c) identisch. Das gilt auch für die gestrichelt gezeichnete Kurve der Strategie FCFS, vgl. Gl. (3). Die strichpunktierten Kurven gelten für die Strategie RU und unterscheiden sich in den Bildern 3a), b) und c) in der konstant angenommenen Differenz $\Delta\omega = \omega_{j+1} - \omega_j$ (mit $1 \leq j \leq 3$) der Wichtigkeiten, die dort 5, 10 und 50 beträgt.

Eine wesentliche Eigenschaft der Strategie NONPRE ist, daß auch bei sehr starker Prozessorauslastung $\varrho_{\leq N} \to 1$ die mittlere Wartezeit von Anforderungen mit hoher Priorität klein bleibt. Dieses Verhalten kann – durch Vergrößerung der Streuung Var (ω) der Wichtigkeiten – durch die Strategie RU in jedem gewünschten Maße erreicht werden. Gleichzeitig werden dabei auch die mittleren Wartezeiten für Anforderungen mit niedrigeren Prioritäten gut approximiert.

In Bild 3c) ist das – bei Var (ω) = 3125 – für Gesamtangebote etwa im Bereich $0 \leq \varrho_{\leq N} \leq 0{,}8$ erfüllt. In Bild 3b) ist die Streuung Var (ω) = 125 deutlich kleiner und deshalb sind die mittleren Wartezeiten der Strategien NONPRE und RU nur im Bereich $0 \leq \varrho_{\leq N} \leq 0{,}6$ etwa gleichgroß. Bei noch kleinerer Streuung der Wichtigkeiten Var (ω) = 31,25, vgl. Bild 3a), gilt das nur noch für etwa den Bereich $0 \leq \varrho_{\leq N}$

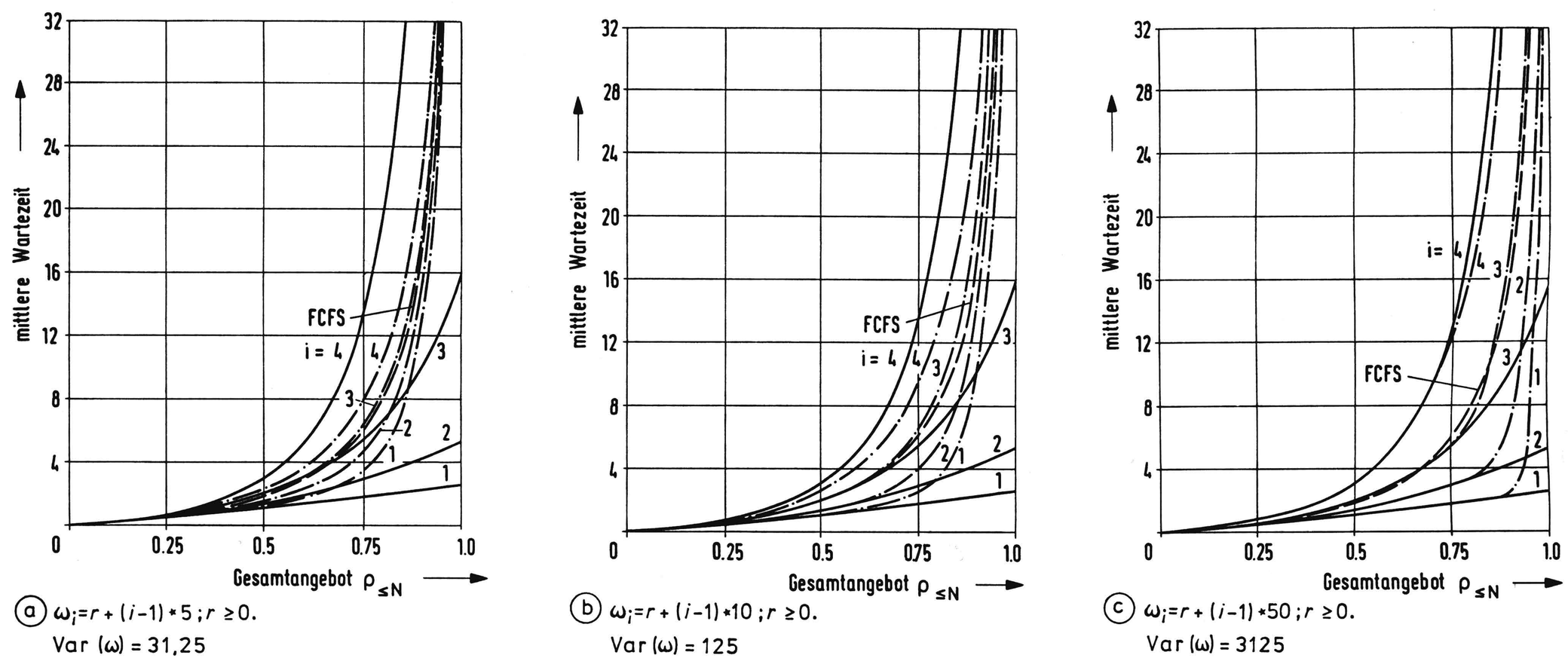

Bild 3. Mittlere Wartezeit der Strategien FCFS (gestrichelt), RU (strichpunktiert) und NONPRE (ausgezogen). Die Wichtigkeiten ω_i in den Bildern a), b) und c) haben nur Bedeutung für die Strategie RU. Die Differenz $\Delta\omega$ der Wichtigkeiten $\omega_{j+1} - \omega_j$ mit ($1 \leq j \leq 3$) beträgt in Bild a) $\Delta\omega = 5$, b) $\Delta\omega = 10$ und c) $\Delta\omega = 50$.

$\leq 0{,}4$. Läßt man die Streuung gegen Null gehen, so ergibt sich für alle Typen von Anforderungen als Grenzfall die mittlere Wartezeit der Strategie FCFS.
Aus letztgenannter Tatsache kann nicht gefolgert werden, daß die Kurven der Strategie RU bei Var $(\omega) > 0$ symmetrisch zur Kurve der Strategie FCFS verlaufen, sie beispielsweise nicht schneiden dürfen.
Das hier an einem Beispiel demonstrierte Verhalten der Strategie RU in Abhängigkeit von der Streuung der Wichtigkeiten gilt für beliebige andere Beispiele ebenso. Mit der Strategie RU läßt es sich z. B. erreichen, daß unterschiedlich wichtige Anforderungen bei normaler Last eines Prozeßrechners bedient werden, wie durch die Strategie NONPRE, daß aber bei vorübergehend ansteigendem Gesamtangebot die unwichtigsten Typen von Anforderungen auf Kosten der wichtigsten noch mit akzeptabler mittlerer Wartezeit bedient werden.

Simulationsuntersuchungen an Wartezeit-VF*n* der Strategien NONPRE, RU und FCFS ergaben folgendes:

- Die Wartezeit-VF*n* der Strategien NONPRE und RU verlaufen sehr ähnlich, solange die mittleren Wartezeiten W_i ähnlich groß sind. Das ist immer bei „kleinen" Gesamtangeboten der Fall. Erst bei größerem Gesamtangebot ergaben sich unterschiedliche Wartezeit-VF*n*. Mit zunehmendem Gesamtangebot weichen die Wartezeit-VF*n* der Strategie RU von den entsprechenden VF*n* der Strategie NONPRE zunehmend ab. Dabei schwenken sie für große Wartezeiten in einen Verlauf ein, der parallel zu dem der Wartezeit-VF der Strategie FCFS ist, vgl. ausgezogene Kurven in Bild 4 für Wichtigkeiten ω_i entspr. Bild 3b. Das Hinbiegen der Wartezeit-VF*n* der Strategie RU zu der Wartezeit-VF der Strategie FCFS tritt dadurch auf, daß unwichtigere Typen von Anforderungen so lange warten müssen, daß ihre dynamische Priorität höher als die wichtigerer Anforderungen geworden ist. Aus Bild 5 sieht man, daß dieses Umbiegen unabhängig vom Gesamtangebot für jeden Typ i bei einer bestimmten Wartezeit $t = w_i$ erfolgt. Ist die Streuung der Wichtigkeiten klein, dann beobachtet man diesen Effekt schon bei kleineren Wartezeiten. Je größer diese Streuung ist, bei um so größeren Wartezeiten erst wird der Parallel-Lauf mit der Wartezeit-VF der Strategie FCFS eingeleitet, vgl. Bild 6. Für den Satz von Wichtigkeiten entsprechend Bild 3c), wo Var $(\omega) = 3125$ ist, verlaufen die Wartezeit-VF*n* der Strategien RU und NONPRE bei $\varrho_{\leq N} = 0{,}75$ noch deckungsgleich, vgl. gestrichelte Kurven in Bild 4. Das entspricht der Beobachtung in Bild 3c), daß die mittleren Wartezeiten beider Strategien bei diesem Gesamtangebot noch identisch sind.
- Aus dem beobachteten Parallelverlauf der Wartezeit-VF*n* der Strategien RU und FCFS kann eine Näherung für den Verlauf bei großen Wartezeiten abgeleitet werden, vgl. Gl. (9).

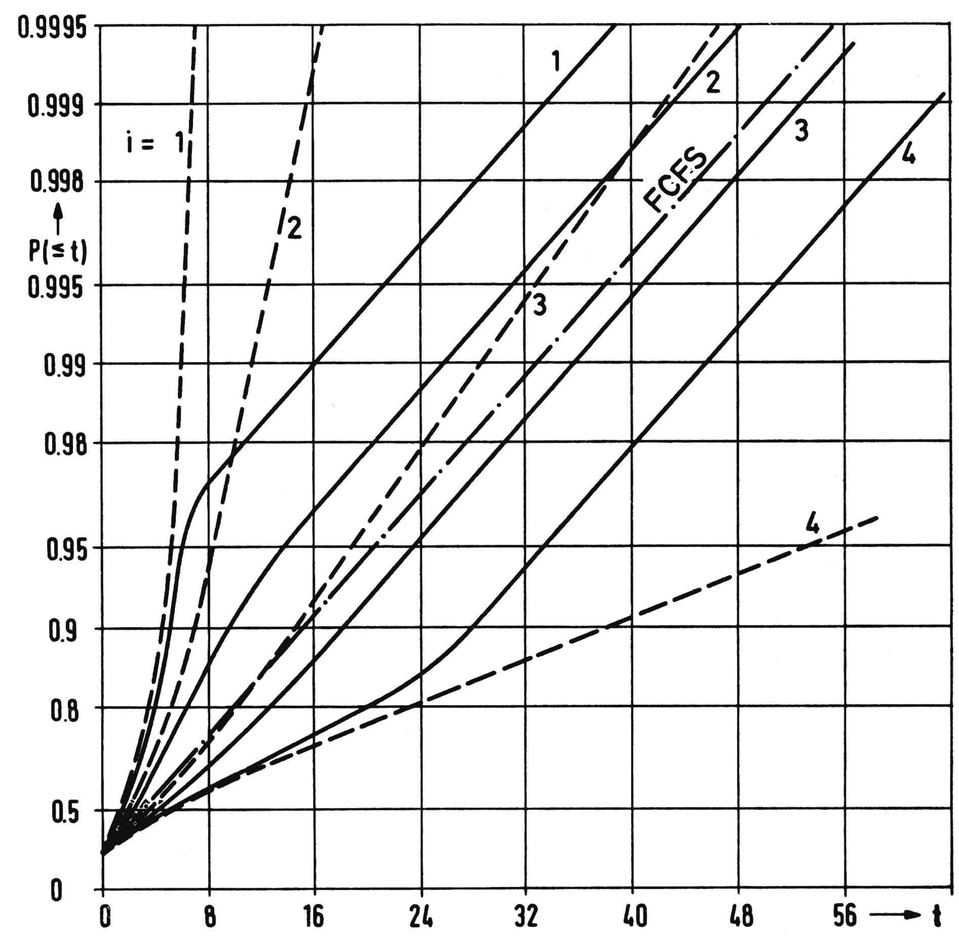

Bild 4. Simulierte Wartezeit-VF*n* bei nichtunterbrechenden statischen (gestrichelt) und terminorientierten (ausgezogen) Prioritäten. Die strichpunktierte Kurve gilt für die Strategie FCFS. $\varrho_{\leq N} = 0{,}75$. Die Wichtigkeiten sind wie im Bild 3b) angegeben.

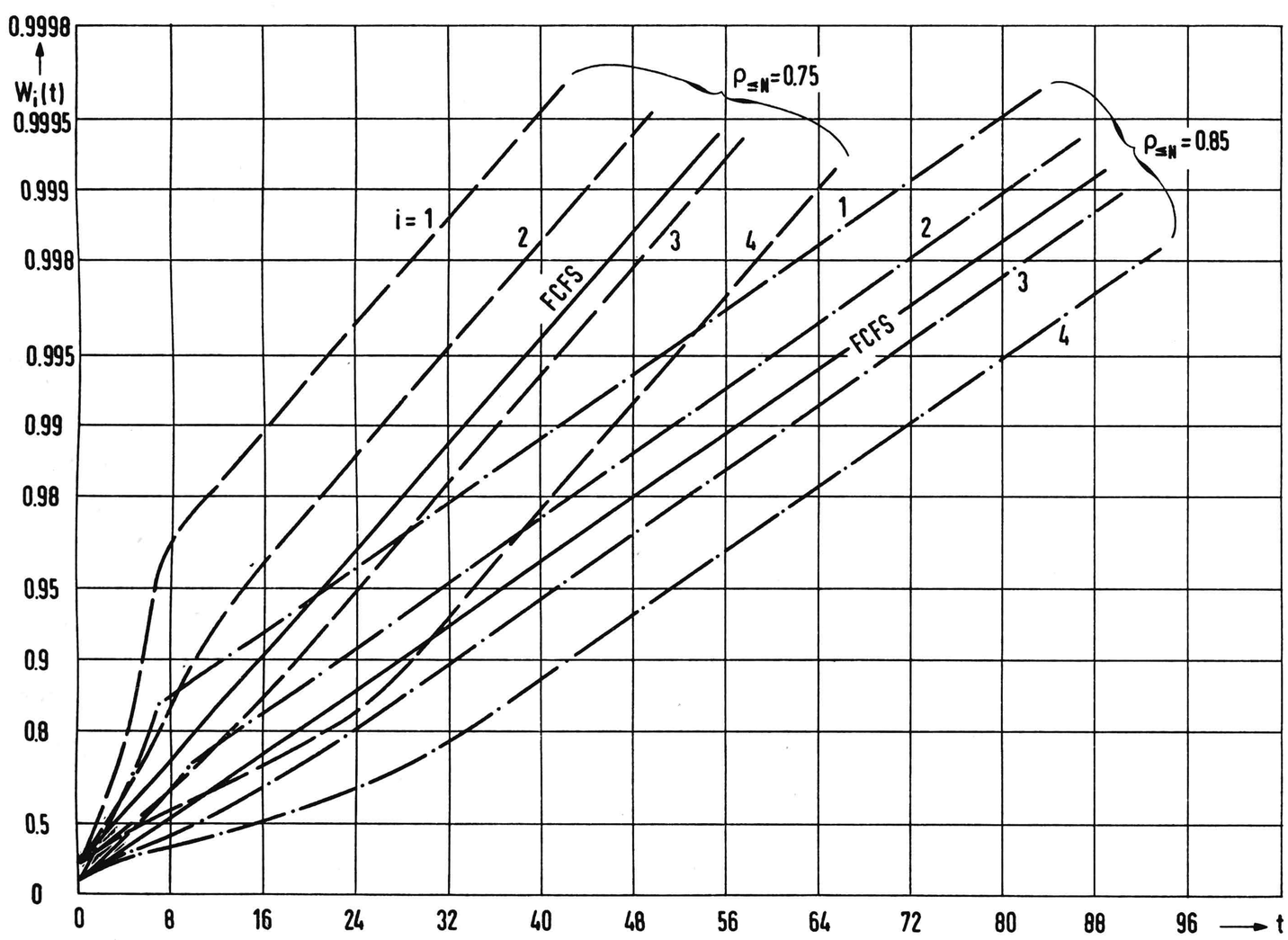

Bild 5. Wartezeit-VF*n* für 2 Gesamtangebote $\varrho_{\leq N} = 0{,}75$ und 0,85. Ausgezogene Kurve: Strategie FCFS, unterbrochene Kurven: Strategie RU. Die Wichtigkeiten sind wie in Bild 3b).

- Die zu einer Wahrscheinlichkeit $W_i(t) = P(w_i \leq t)$ gehörigen (z. B. aus Bild 5 ablesbaren) Wartezeiten w_i der verschiedenen Typen von Anforderungen unterscheiden sich für große t näherungsweise um die Differenz der entsprechenden Wichtigkeiten. Das bedeutet, für große t werden alle Typen von Anforderungen gleich gerecht behandelt – unter Berücksichtigung der unterschiedlichen freiwilligen Wartezeiten ω_i.

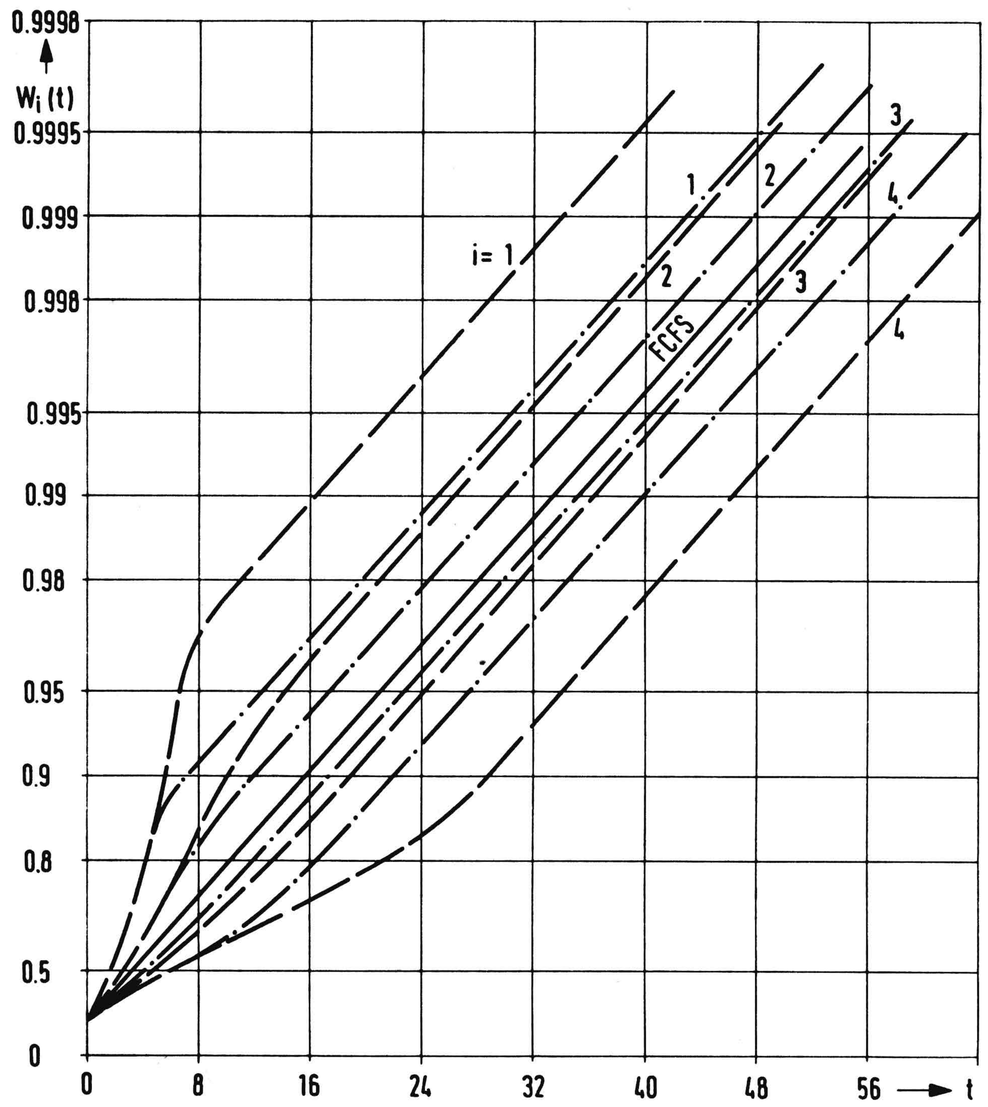

Bild 6. Wartezeit-VFn bei $\varrho_{\leq N} = 0{,}75$ und 2 Sätzen von Wichtigkeiten, nämlich 1. $\Delta\omega = \omega_{j+1} - \omega_j = 5$ (strichpunktiert), $\mathrm{Var}(\omega) = 31{,}25$, bzw. 2. $\Delta\omega = 10$ (gestrichelt), $\mathrm{Var}(\omega) = 125$, mit $\omega_1 = 0$ in beiden Fällen.

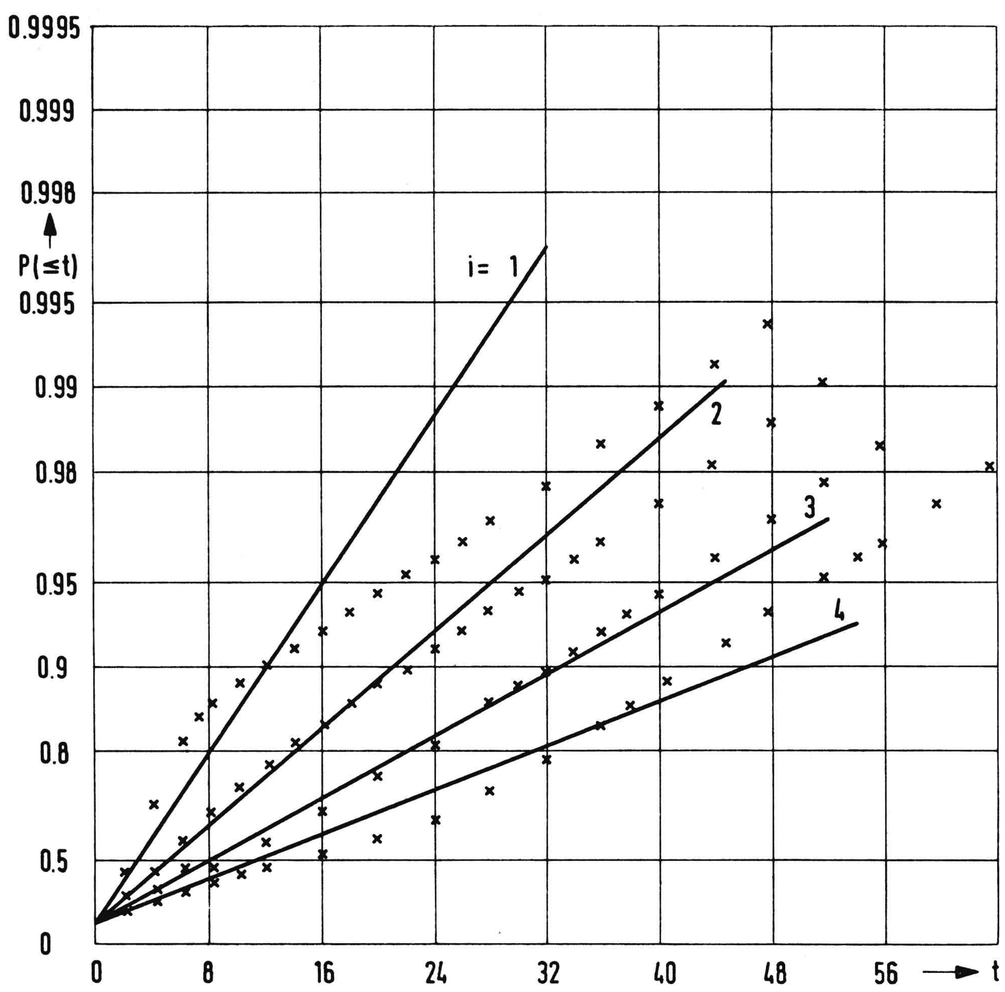

Bild 7. Wartezeit-VFn der Strategie RU (Kreuze), angenähert nach Gl. (10) (Geraden), $\varrho_{\leq N} = 0{,}85$. Die Wichtigkeiten sind wie in Bild 3b).

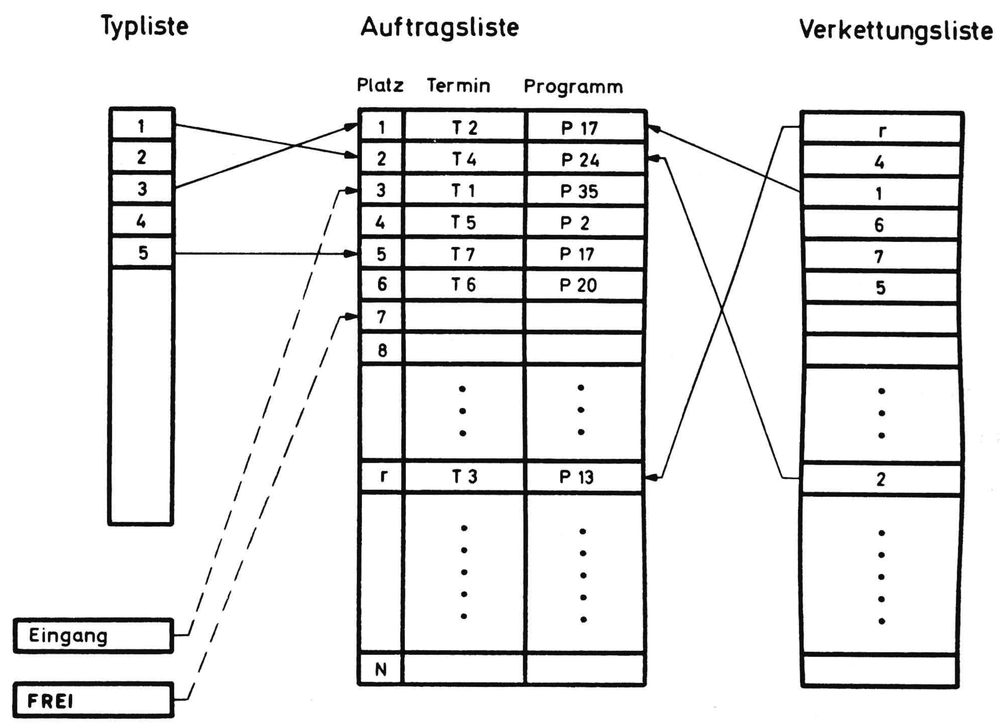

Bild 8. Auftragsliste (AL) mit Starttermin Tk und zugehöriger Programmnummer (z. B. P 17), Verkettungsliste (VL) mit gleicher Zahl von Plätzen wie in der AL, Typliste (TL) mit Zeigern auf die prozessornächste Anforderung jedes Typs und Eingangs- und Freizeiger. Für jeden belegten Platz in der AL gibt der entsprechende Platz der VL einen Platz der AL an, in dem Termin und Nummer des nächstwichtigsten Nachfolgeprogramms steht.

Eine ankommende Anforderung des Typs i belegt den im Zeiger FREI angegebenen Platz und wird entweder ① (logisch) nach dem Zeiger i der TL, sonst (falls i unbesetzt in TL) ② nach dem Zeiger für nächstwichtigeren Typ $j < i$, sonst (falls alle Elemente $j < i$ in TL unbesetzt) ③ vom Listenanfang von AL an eingeordnet. Bei ② und ③ ist dazu paarweiser Vergleich von Ti mit dem Termin der den jeweiligen AL-Platz belegenden Anforderung nötig. Die VL muß dabei aktualisiert werden. Nach Fertigstellung einer Anforderung wird der Zeiger EINGANG auf das Nachfolgeprogramm umgesetzt.

- Anforderungen unterschiedlicher Priorität erfahren bei der Strategie NONPRE sehr unterschiedliche Wartezeiten: Im Gegensatz zur Strategie RU streben hier die Wartezeit-VFn auseinander, vgl. gestrichelte Kurven in Bild 4.
- Bei kleinen Wartezeiten verlaufen die Wartezeit-VFn der Strategie RU *nicht* parallel und sind durch Gl. (10) annäherbar, vgl. Bild 7.

5. Verwaltungsaufwand der Strategie RU

Für die praktische Anwendung hat die Strategie RU gegenüber der Strategie NONPRE den Vorteil, daß sie durch Wahl der Parameter ω_i die Möglichkeit bietet, das Echtzeitverhalten genau den Bedürfnissen anzupassen. Als Grenzfall kann auch die Strategie NONPRE erzeugt werden. Es stellt sich die Frage nach dem dafür notwendigen Verwaltungsaufwand.

Man kann sich alle wartenden Anforderungen, abweichend von der Darstellung nach Bild 1, auch als in einer Warteschlange wartend vorstellen:

Bei der Strategie NONPRE muß eine ankommende Anforderung der Priorität i so in diese Warteschlange eingeordnet werden, daß sie vor allen Anforderungen niedrigerer Priorität und nach allen Anforderungen gleicher oder höherer Priorität liegt. Der Prozessor muß dann lediglich eine nach der anderen Anforderung vom Kopf der Warteschlange nehmen und abarbeiten. (Deshalb wird diese Strategie in der Literatur auch *head of the line discipline* genannt.)

Bei der Strategie RU muß eine ankommende Anforderung, welche wegen $w_i(t)=0$ die Priorität $q_i(t)=\omega_i$ hat, *nach* allen wartenden Anforderungen plaziert werden, die ebenfalls diese oder eine höhere Priorität haben und *vor* allen unwichtigeren. Dieses Einsortieren muß spätestens nach Fertigstellung der gerade in Arbeit befindlichen Anforderung erfolgen. Vor dem Einsortieren werden die dynamischen Prioritäten aller wartenden Anforderungen durch Subtraktion des zwischen diesem und dem letzten Aktualisierungszeitpunkt liegenden Zeitbetrages von den inzwischen veralteten Prioritäten aktualisiert. Danach kann die Einsortierung neu angekommener, unter Berücksichtigung ihrer eventuellen Wartezeit, erfolgen.

Ein eleganteres Verfahren verzichtet auf diese Aktualisierung und betrachtet nur Anforderungen der Typen $j > i$, denn nur diese können evtl. durch Warten eine höhere Priorität als die einzuordnende Anforderung des Typs i haben. Anstelle der aktuellen Priorität wird für jede Anforderung (neben dem Typ) der Termin $T_k (1 \leq k \leq N)$ gespeichert. Stellt man sich alle wartenden Anforderungen in einer Warteschlange so angeordnet vor, daß die Strategie FCFS zur angestrebten Abarbeitung führt, dann braucht man nur die Typen j zu betrachten, *nach* denen aus der Sicht des Prozessors keine Anforderung vom Typ i mehr wartet. Dabei ist berücksichtigt, daß nach genügend langem Warten Anforderungen vom Typ j wichtiger geworden sein können als später angekommene vom Typ i und deshalb vor diesen bedient werden müssen. Zur Vereinfachung dieser Suche sind Zeiger zweckmäßig, die für jeden Typ die jeweils prozessorentfernteste Position in der Warteschlange anzeigen, vgl. Bild 8. Diese Zeiger entsprechen den Marken, die man bei der Strategie NONPRE als Hilfsmittel zur Verwaltung von Anforderungen unterschiedlicher Priorität in einer gemeinsamen Warteschlange benötigt. Bei kleiner augenblicklicher Warteschlangenlänge kann das Einordnen evtl. noch schneller durch Vergleich der angekommenen Anforderung mit den an Kopf und Schwanz wartenden Anforderungen erfolgen.

Offensichtlich ist der Verwaltungsaufwand der Strategie RU etwas größer als bei NONPRE, scheint aber noch erträglich zu sein.

Anmerkung: Die Simulationsergebnisse verdanke ich den Arbeiten meines Kollegen *W. Rosenbohm*, vgl. [3].

Literatur

[1] *Walke, B.:* Realzeitrechner-Modelle, Theorie und Anwendung. Reihe Datenverarbeitung, R. Oldenbourg Verlag, München/Wien 1978.

[2] *Walke, B.:* Improved bounds and an approximation for a dynamic priority queue. Proceed. Measur., Modell. and Perform. Evaluat. Comp. Syst., Bonn, Oct. 1977, North-Holland Publishing Company (1977), pp. 321–346.

[3] *Walke, B.; Rosenbohm, W.:* Waiting-time distributions for deadline oriented servicing. 4th Int. Sympos. Modell. and Perform. Evaluat. Comp. Systems, 6.–8. Febr. 1979, Wien-Laxenburg, Proceed. by North-Holland Publishing Company, 1979.

[4] *Cobham, A.:* Priority assignment in waiting line problems. Oper. Res. (USA), Vol. 2 (Febr. 1954), pp. 70–76.

[5] *Jackson, J. R.:* Some problems in queueing with dynamic priorities. Naval Research Logistics Quarterly. Sept. 1960, pp. 235–249.