

Netzwerke von Warteschlangen über Graphen mit ausfallenden Knoten

Cornelia Sauer

Fachbereich Mathematik, Universität Hamburg

Bundesstraße 55, 20146 Hamburg

In dieser Arbeit werden das Leistungsverhalten von Warteschlangennetzen und ihre Zuverlässigkeit in einem gemeinsamen Modell beschrieben. Sie kann daher dem neueren Forschungsgebiet **Performability** zugeordnet werden, siehe [4]. Bisherige Ergebnisse in der Literatur führten bei einer exakten Modellierung von Netzwerken mit unzuverlässigen Knoten nicht zu stationären Verteilungen in Produktform, die einfache Leistungsanalysen ermöglichen. Die Arbeit [2] gibt einen Überblick über Approximationsmethoden, die deshalb entwickelt wurden. Im Gegensatz dazu wird hier ausgehend vom Standard Jackson Netzwerk der Ausfall und die anschließende Reparatur der Knoten gleich in die Markovsche Systembeschreibung so eingebaut, dass das Gleichgewichtsverhalten des Netzwerks in einer neuartigen Produktformstruktur angebar ist. Zunächst sei kurz das zugrundegelegte Netzwerk skizziert.

Definition 1 *Ein Jackson Netzwerk besteht aus J Knoten, gegeben in der Menge $\bar{J} := \{1, 2, \dots, J\}$, jeder Knoten j repräsentiert ein $|M|1|\infty$ – FCFS-System, an dem die Servicekapazität $\mu_j(n_j) > 0$ bereitgestellt wird, wenn sich dort genau $n_j \in \mathbb{N}$ Kunden befinden, $j \in \bar{J}$. Die ununterscheidbaren Kunden treffen am Knoten j in einem Poisson- λ_j -Strom aus der Außenwelt ein, $\lambda_j \geq 0$, und fordern an jedem Knoten i , den sie während ihrer Zeit im Netzwerk aufsuchen, eine Bedienung, die exponential verteilt ist mit Erwartungswert 1. Die Kundenbewegungen im Netzwerk werden bestimmt durch einen Markovschen Routingmechanismus, gegeben durch die stochastische, irreduzible Matrix $\mathcal{R} := (r(i, j) : i, j \in \bar{J}_0)$ mit $r(0, 0) := 0$, $\lambda := \sum_{j=1}^J \lambda_j$, $r(0, j) := \frac{\lambda_j}{\lambda}$ und $\bar{J}_0 := \bar{J} \cup \{0\}$. Die Folgen der Zwischenankunftszeiten und Bedienzeitforderungen bilden eine unabhängige Familie von Zufallsvariablen und die Routingentscheidungen sind unabhängig von der Vergangenheit, gegeben den gegenwärtigen Zustand. Mit $X_j(t)$ als der Anzahl der an Knoten j anwesenden Kunden zur Zeit $t \geq 0$, sei $X(t) := (X_j(t) : j \in \bar{J})$ der gemeinsame Schlangenlängenvektor zur Zeit t .*

Satz 2 ([3]) *Der gemeinsame Schlangenlängenprozess $X = (X(t) : t \geq 0)$ des Jackson Netzwerks ist ein Markov Prozess auf $E = \mathbb{N}^J$, der als ergodisch vorausgesetzt sei. Dann hat die Verkehrsgleichung des Netzwerks eine eindeutige Lösung $\eta = (\eta_1, \dots, \eta_J)$, gegeben durch*

$$\eta_j = \lambda_j + \sum_{i=1}^J \eta_i r(i, j), \quad j \in \bar{J}, \quad (1)$$

und die eindeutige stationäre und Grenzverteilung von X ist (mit Normierungskonstante $K(J)$):

$$\pi(n_1, \dots, n_J) = K(J)^{-1} \prod_{j=1}^J \prod_{k=1}^{n_j} \left(\frac{\eta_j}{\mu_j(k)} \right), \quad \text{für } (n_1, \dots, n_J) \in E.$$

In Erweiterung des Jackson Netzwerks seien nun die Knoten als unzuverlässig vorausgesetzt, d.h. die Bediener fallen in zufälligen Zeitpunkten aus und werden dann repariert. Genauer: die Knoten können einzeln oder in Gruppen gemeinsam ausfallen und auch die anschließende Reparatur kann für einzelne Knoten oder für eine Knotengruppe zeitgleich enden. Es ist dabei nicht notwendig, dass Knoten, die zusammen ausgefallen sind, auch gemeinsam ihre Reparatur beenden.

Ausgefallene Knoten nehmen während ihrer Reparatur weder weitere Kunden auf noch wird die Bedienung von evtl. anwesenden Kunden fortgesetzt. Letztere verbleiben am Knoten, bis der Server repariert ist und die Bedienung wieder aufgenommen wird. Weiter unten werden drei Verfahren beschrieben, mit denen neue

Routingmatrizen hergeleitet werden, die das Kundenverhalten im Netzwerk in Bezug auf ausgefallene Knoten regeln.

Um die zeitliche Entwicklung des Netzwerks darstellen zu können, muss der Zustandsraum für den beschreibenden Markovprozess erweitert werden zu $\tilde{E} = E \cup (\mathcal{P}(\bar{J}) \setminus \{\emptyset\} \times \mathbb{N}^J)$. Ein neuer Zustand der Form

$$(\bar{I}, n_1, n_2, \dots, n_J) \in \tilde{E} \setminus E$$

bedeutet dann, dass die Knotengruppe $\bar{I} \in \mathcal{P}(\bar{J}) \setminus \{\emptyset\}$ ausgefallen ist und repariert wird und dass an diesen Knoten $i \in \bar{I}$ jeweils n_i Kunden auf die Reparatur des Bedieners warten, während sich an den normal arbeitenden Knoten i je n_i Kunden befinden, $i \in \bar{J} \setminus \bar{I}$.

Sei $\bar{K} \subset \bar{J}$ die Menge der momentan ausgefallenen Knoten. Dann fällt die Menge $\bar{I} \subset \bar{J} \setminus \bar{K}$, $\bar{I} \neq \emptyset$, normal arbeitender Knoten zusammen mit Intensität $\alpha_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{K} \cup \bar{I})$ aus, wenn sich n_i Kunden an Knoten i , $i \in \bar{K} \cup \bar{I}$, befinden und im Fall $\bar{K} \neq \emptyset$ endet für die Knotenmenge $\bar{I} \subset \bar{K}$, $\bar{I} \neq \emptyset$, die Reparatur mit Intensität $\beta_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{K})$, wenn im Moment des jeweiligen Ausfalls n_i Kunden am Knoten i , $i \in \bar{K}$, waren, die sich immer noch wartend dort befinden. Die Ausfall- und Reparaturraten können also insbesondere von der lokalen Kundensituation im Netzwerk abhängen.

Jedoch können die Intensitäten $\alpha_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{J})$ und $\beta_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{J})$ in dieser komplexen Situation nicht mehr beliebig gewählt werden. Noch sehr allgemeine Formen erhält man, wenn man die Ausfälle als Geburten und die Reparaturen als Tode auffasst und bekannte Intensitäten aus der Theorie mehrdimensionaler Geburts- und Todesprozesse (siehe [6], Tab. 1), auf unsere Situation überträgt. Es sei eine Möglichkeit angegeben:

Definition 3 *Die Intensitäten für Ausfall bzw. Reparatur der Knotengruppe $\bar{I} \neq \emptyset$ haben, wenn die Knotengruppe $\bar{K} \subset \bar{J}$ als momentan ausgefallen vorausgesetzt sei, die folgende Form:*

$$\alpha_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{J}) := \frac{A(n_i : i \in \bar{K} \cup \bar{I})}{A(n_i : i \in \bar{K})} \quad \text{bzw.} \quad \beta_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{J}) := \frac{B(n_i : i \in \bar{K})}{B(n_i : i \in \bar{K} \setminus \bar{I})}. \quad (2)$$

Hierbei sind A and B beliebige nichtnegative reelle Funktionen: $A, B : \bigcup_{l \in \bar{J}_0} \mathbb{N}^l \rightarrow \mathbb{R}_+ = [0, \infty)$.

$\alpha_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{J})$ und $\beta_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{J})$ dienen als Übergangintensitäten für den erweiterten Markovprozess auf \tilde{E} und werden als endlich vorausgesetzt. Zur Normierung setzen wir: $A(n_i : i \in \emptyset) := 1 =: B(n_i : i \in \emptyset)$.

Es ist nun das Routingverhalten der Kunden im Netzwerk für Zeiten, in denen mindestens ein Knoten ausgefallen ist, festzulegen. Dazu werden drei mögliche Lösungsprinzipien vorgestellt. Die ersten beiden stammen ursprünglich aus der Theorie des Information-Blockings, [1], wo sie für die Behandlung von Netzwerken mit Kapazitätsbeschränkungen entwickelt wurden. Es ist überraschend, dass sie auch hier im Netzwerk, das dem dynamischem Wechselspiel von Ausfall und Reparatur der Knoten unterworfen ist, geeignete Umleitungsregeln für die Kunden definieren.

Definition 4 (Stalling[5]) *Sobald ein einzelner Knoten oder eine Knotengruppe ausgefallen ist, werden alle externen Ankunftsströme und alle Bedienprozesse an arbeitenden Knoten augenblicklich unterbrochen und erst wieder aufgenommen, wenn alle ausgefallenen Knoten wieder repariert sind.*

Dieses Prinzip des erzwungenen Stillstandes findet in der Praxis durchaus seine Anwendung. So wird in großen Produktionsanlagen, z.B. bei der Fließbandfertigung in der Automobilindustrie, bei Ausfall oder Störung einer Fertigungseinheit die gesamte Produktion gestoppt und erst nach erfolgter Reparatur wieder aufgenommen.

Das zweite Lösungsverfahren hat sich vor allem in der Nachrichtenverkehrstheorie etabliert, wo es bei der Modellierung von Systemen mit begrenzter Pufferkapazität verwendet wird.

Definition 5 (Repetitive Service – Random Destination (RS-RD)) *Ein Kunde, der nach Ende seiner Bedienung an Knoten $i \in \bar{J} \setminus \bar{I}$ gemäß seiner Routingvorschrift einen gerade in Reparatur befindlichen Knoten $j \in \bar{I}$ als nächsten Aufenthaltsknoten gewählt hat, verbleibt an Knoten i , um eine weitere Bedienung zu erhalten, deren Dauer neu bestimmt wird. Wenn diese zusätzliche Bedienung endet, wählt er aufs Neue seinen nächsten anzusteuern Knoten gemäß seiner Routingmatrix.*

Die formale Anwendung dieses Prinzips ergibt als neue Routingmatrix $\tilde{R}^{\bar{I}} = (\tilde{r}^{\bar{I}}(i, j) : i, j \in \bar{J}_0 \setminus \bar{I})$, mit

$$\tilde{r}^{\bar{I}}(i, j) = \begin{cases} r(i, j), & i, j \in \bar{J}_0 \setminus \bar{I}, i \neq j, \\ r(i, i) + \sum_{k \in \bar{I}} r(i, k), & i \in \bar{J}_0 \setminus \bar{I}, i = j. \end{cases} \quad (3)$$

Die Routingmatrix des ursprünglichen Prozesses muss hier als reversibel vorausgesetzt werden.

Das dritte Prinzip beruht auf der Theorie von Tabuwahrscheinlichkeiten in Zusammenhang mit Markov-Ketten. Als allgemeine Methode wird es in [5] eingeführt. Dabei wird der Routingprozess des Jackson Netzwerks als eine Markov Kette auf dem Zustandsraum \bar{J}_0 mit Übergangsmatrix $\mathcal{R} = (r(i, j) : i, j \in \bar{J}_0)$ betrachtet. Verboten man nun den Kunden die Knotenmenge $\bar{I} \subset \bar{J} (\bar{I} \neq \emptyset, \bar{I} \neq \bar{J})$ zu betreten, d.h., \bar{I} ist der Tabubereich, dann steuert die Skipping-Regel ein spezielles Überspringen des verbotenen Bereiches wie folgt:

Definition 6 (Skipping) Wenn ein Kunde an Knoten $i \in \bar{J}_0 \setminus \bar{I}$ als nächsten anzusteuern den Knoten $j \in \bar{J}_0 \setminus \bar{I}$ wählt, (mit Wahrscheinlichkeit $r(i, j)$), dann springt er sofort zu diesem, wählt er aber den ausgefallenen Knoten $k \in \bar{I}$ (mit Wahrscheinlichkeit $r(i, k)$), kann er sich nicht zu diesem begeben, sondern muss sofort seinen nächsten Aufenthaltsknoten l wählen, mit Wahrscheinlichkeit $r(k, l)$, etc. Dieses Skipping-Prinzip führt zu einer neuen Markovschen Routingmatrix $\hat{\mathcal{R}}^{\bar{I}} = (\hat{r}^{\bar{I}}(i, j) : i, j \in \bar{J}_0 \setminus \bar{I})$, gegeben durch

$$\hat{r}^{\bar{I}}(i, j) = r(i, j) + \sum_{k \in \bar{I}} r(i, k) \hat{r}^{\bar{I}}(k, j), \quad i \in \bar{J}_0, j \in \bar{J}_0 \setminus \bar{I}. \quad (4)$$

Das nachstehende Hauptergebnis zeigt, dass die vorgestellte Modellierung zu einer stationären Verteilung in Produktform für das Netzwerk mit unzuverlässigen Knoten führt, die die gemeinsame oder getrennte Behandlung der Aspekte Zuverlässigkeit und Leistungsverhalten des Netzwerks ermöglicht.

Satz 7 Gegeben sei ein Jackson Netzwerk mit Ausfall- und Reparaturverhalten gemäß Definition 3, für das im Falle des Ausfalls von Knoten eine neue Routingmatrix gemäß dem **Stalling-Prinzip**, dem **RS-RD-Prinzip** oder dem **Skipping-Prinzip** erzeugt wird. Dann besitzt der beschreibende, erweiterte Markovprozess auf \tilde{E} die folgende stationäre Verteilung in Produktform.

Definition 8 (Stationäre Verteilung für unzuverlässige Jackson Netzwerke)

Auf dem erweiterten Zustandsraum \tilde{E} sei die folgende Verteilung in Produktform definiert:

$$\pi(n_1, \dots, n_J) = C^{-1}(J) \prod_{j=1}^J \prod_{i=1}^{n_j} \left(\frac{\eta_j}{\mu_j(i)} \right) \quad \text{für } (n_1, n_2, \dots, n_J) \in E,$$

$$\pi(\bar{I}, n_1, n_2, \dots, n_J) = C^{-1}(J) \frac{A(n_i : i \in \bar{I})}{B(n_i : i \in \bar{I})} \prod_{j=1}^J \prod_{i=1}^{n_j} \left(\frac{\eta_j}{\mu_j(i)} \right) \quad \text{für } (\bar{I}, n_1, n_2, \dots, n_J) \in \tilde{E} \setminus E.$$

Dabei ist $\eta = (\eta_1, \eta_2, \dots, \eta_J)$ die Lösung der Verkehrsgleichung des Standard-Netzwerks, (1), und $C(J)$ ist die Normierungskonstante, die als endlich vorausgesetzt sei.

Literatur

- [1] S. Balsamo, V. De Nitto Persone, R. Onvural. *Analysis of Queueing Networks with Blocking*. Kluwer academic publishers, 2001.
- [2] R. Chakka and I. Mitrani. Approximate solutions for open networks with breakdowns and repairs. In F. P. Kelly, S. Zachary and I. Ziedins, editors, *Stochastic Networks - Theory and applications*, chapter 16, pp. 267-280, Clarendon Press, Oxford, 1996.
- [3] J. R. Jackson. Networks of waiting lines. *Operations Research*, 5, pp. 518-521, 1957.
- [4] J.F. Meyer. Performability modeling of distributed real-time systems. In G. Iazeolla, P.J. Courtois and A. Hordijk, editors, *Mathematical Computer Performance and Reliability*, pp. 361-372, North-Holland, Amsterdam, 1984.
- [5] R. Schassberger. Decomposable stochastic networks: Some observations. In F. Baccelli and G. Fayolle, editors, *Modelling and Performance Evaluation Methodology*, volume 60 of *Lecture Notes in Control and Information Sciences*, chapter IV, pp. 137-150, Springer, Berlin, 1984.
- [6] R. F. Serfozo. Queueing networks with dependent nodes and concurrent movements. *Queueing Systems*, 13, pp. 143-182, 1993.