



Universität Augsburg  
Prof. Dr. Hans Ulrich Buhl  
Kernkompetenzzentrum  
Finanz- & Informationsmanagement  
Lehrstuhl für BWL, Wirtschaftsinformatik,  
Informations- & Finanzmanagement

**UNIA**  
Universität  
Augsburg  
University

Diskussionspapier WI-78

## **Elektronischer Vertrieb von Dienstleistungen: Optimaler Serverstandort als Warehouse Location Problem**

von

Melanie Frank, Dennis Kundisch, Axel Schell

Juni 2000

in: Bodendorf, F., Grauer, G., Hrsg., *Verbundtagung Wirtschaftsinformatik (WINFO) 2000*, Siegen, Oktober 2000, Shaker, Aachen, 2000, S.208-222



# **Elektronischer Vertrieb von Dienstleistungen: Optimaler Serverstandort als Warehouse Location Problem**

Melanie Frank, Dennis Kundisch, Axel J. Schell

Universität Augsburg, Lehrstuhl für Betriebswirtschaftslehre mit dem Schwerpunkt Wirtschaftsinformatik

Universitätsstraße 16, 86135 Augsburg

Telefon: ++49/821/598-4140, Telefax: ++49/821/598-4225

`melanie.frank@student.uni-augsburg.de,`  
`{dennis.kundisch | axel.schell}@wiso.uni-augsburg.de`

## Elektronischer Vertrieb von Dienstleistungen: Optimaler Serverstandort als Warehouse Location Problem

**Abstract:** In diesem Beitrag wird die These aufgestellt, daß neben einer Vielzahl von qualitativen und quantitativen Standortfaktoren auch die dem Kunden entstehende Verzögerungszeit beim Kauf oder der Nutzung von elektronischen Dienstleistungen in die Standortentscheidung für die Einrichtung eines Servers mit einkalkuliert werden sollte. Die Verzögerungszeit verursacht dem Endkunden bei der Nutzung elektronischer Dienstleistungen (Opportunitäts-)Kosten und wird daher vom Kunden in seine Kauf-/Downloadentscheidung einbezogen. Zur Stützung der These wird ein unkapazitiertes und ein kapazitiertes Modell bzgl. der Bandbreiten der Übertragungsleitungen aus dem Operations Research (hier: Warehouse Location Problem) auf das Internet übertragen. Die Ergebnisse bestätigen tendenziell einen substantiellen Einfluß der Serverstandortwahl auf den Erfolg des Vertriebs von elektronischen Dienstleistungen.

### 1 Einleitung

Der Internet Boom hält auch nach Jahren des rasanten Wachstums weiterhin an. Nach Prognosen von EITO wird im Jahr 2002 eine Internet-Gemeinde von insgesamt 328 Millionen Nutzern erreicht - in Deutschland von fast 20 Millionen [ECIN00]. Nicht nur die Nutzerzahl, sondern auch die Anzahl der Websites nimmt ständig zu. So hat sie sich im Zeitraum 12/98 bis 12/99 weit mehr als verdoppelt [ZAKON00]. Insbesondere für Unternehmen stellt sich heute die Frage, *ob* man ins Internet gehen soll, nicht mehr, sondern nur noch *wie* die Web-Präsenz gestaltet und welche Informationen, Produkte und Dienstleistungen angeboten werden sollen. Auch die Zahl der ans Netz angeschlossenen Rechner (Hosts) steigt deswegen seit Jahren stark an. Täglich werden mehr als ca. 70.000 neue Hosts ans Netz angeschlossen [ZAKON00]. Insbesondere bei Servern müssen dabei jeweils Standortentscheidungen getroffen werden, an welchen Internet Service Provider (ISP) bzw. welchen Backbone der Server angeschlossen werden soll. Des Weiteren ist die Entscheidung zu treffen, an welchen Knoten innerhalb dieses ISP-Netzwerks bzw. Backbones der Server angeschlossen werden soll.

Für eine ökonomisch fundierte Standortentscheidung sind eine Vielzahl von qualitativen (z.B. Sicherheitsüberlegungen, Convenience, rechtliche Aspekte, vorhandenes Know-how, Nähe zu zentralen Datenaustauschpunkten CIX) und quantitativen (z.B. Kosten, Zeit, steuerliche Aspekte) Einflußfaktoren einzubeziehen. Faßt man die Entscheidung für einen Webauftritt bzw. die

Anschaffung eines (weiteren) Web Servers als finanzwirtschaftliche Investition auf, ergibt sich das Problem, daß die qualitativen Einflußfaktoren nur schwer im Entscheidungskalkül angemessen und objektiv berücksichtigt werden können. Zum Beispiel ist der mittelfristige Umsatzeffekt bei einer Verlagerung des Serverstandorts von Deutschland auf die Cayman Islands aufgrund der unklaren Rechtslage schwer erfaßbar. Wesentlich einfacher ist die Bestimmung der steuerlichen Effekte durch die Verlegung. Im Bereich solcher steuerlichen Wirkungen ist bereits einiges an Forschungsarbeit geleistet worden [SATZGER99a], [SATZGER99b]. Auch bezüglich der Kosten sind normalerweise umfassende Informationen – insbesondere in Fachzeitschriften mit ausführlichen Vergleichen – erhältlich [HIMMELEIN99].

Der quantitative Einflußfaktor „Zeit“ als wichtiger Entscheidungsaspekt für die Serverstandortwahl hat in der Wissenschaft bisher – unseres Erachtens zu Unrecht – relativ wenig Beachtung gefunden. Dies ist um so verwunderlicher, wenn man die Meldungen der letzten Monate verfolgt hat. So meldete z.B. Zona Research, daß lange Web Page Ladezeiten schon heute die Unternehmen Milliarden US Dollar kosten. Zona Research fand heraus, daß Webpage Ladezeiten, die 8 Sekunden überschreiten, bei mehr als 33% der Internet User dazu führt, daß diese den Download abbrechen und somit keine Geschäfte dort tätigen [ZONA99].

Alarmierend ist zudem, daß eine aktuelle Studie von Northeast Consulting Resources feststellt, daß das momentan schon langsame Internet die erfolgreiche Zukunft des Electronic Commerce gefährdet und außerdem in den nächsten Jahren noch langsamer zu werden droht. Ursächlich für diese Verzögerungen ist die im Internet verwendete Versendetechnik. Die zu übertragenden Daten werden dabei in Datenpakete aufgeteilt, jeweils mit einem Header ausgestattet, der eindeutige Absender- und Adressateninformation enthält und unabhängig voneinander an den Zielort geroutet. Jeder Router schickt ankommende Pakete anhand seiner Routingtabelle weiter, die periodisch oder dynamisch aktualisiert wird, bis das Paket seinen Bestimmungsort erreicht hat. Diese „packet switching“ Technologie [TANENEBAUM96, S. 133-134] führt dazu, daß es zu Verzögerungen aus folgenden Gründen kommen kann.

- An den Routern kommen gleichzeitig zu viele Datenpakete an und eine Zwischenspeicherung in Puffern wird erforderlich.
- Die verfügbare Bandbreitenkapazität<sup>1</sup> ist zu gering.

---

<sup>1</sup> Ein einzelnes Datenpaket, das sich bereits auf einem Datenübertragungsmedium (z.B. Lichtwellenleiter) befindet, erfährt bei diesem Transport zwischen zwei Routern praktisch keine Verzögerung, da es sich (nahezu) mit Lichtgeschwindigkeit fortbewegt. Durch die begrenzte Bandbreite können jedoch in der Regel nicht alle Pakete eines zusammengehörigen Datenflusses parallel versendet werden.

- Die Verarbeitungskapazität der Hard- und Software eines Servers, der einen Service anbietet, ist unterdimensioniert.

Da der letzte Punkt relativ einfach in den Griff zu bekommen ist, wird für die Betrachtung im folgenden für die erwartete Übertragungszeit nur wichtig sein, wie viele Router auf dem Weg vom Sender zum Empfänger zu überwinden sind und welche Bandbreitenkapazitäten auf diesem Weg zur Verfügung stehen. Wir stellen in der vorliegenden Arbeit folgende These auf:

*Die physische Standortentscheidung für einen Server hat aufgrund der unterschiedlichen Übertragungszeiten substantielle betriebswirtschaftliche Konsequenzen und ist somit in die Entscheidung einzubeziehen.*

Für die Untersuchung der These auf Validität werden wir versuchen, Methoden des OR auf die Standortentscheidung anzuwenden. Faßt man einen Server als ein Warehouse auf, können Überlegungen aus der normativen Standorttheorie, genauer den Warehouse Location Problemen, verwendet werden, um eine zu definierende Zielfunktion zu optimieren. Der Beitrag ist wie folgt organisiert: Nach dieser Einleitung wird in Abschnitt 2 die Modellanalyse vorgestellt. Anschließend werden Limitierungen der Analyse diskutiert und mögliche Erweiterungen für zukünftige Forschungsaktivitäten aufgezeigt (Abschnitt 3), um im Abschnitt 4 mit einer kurzen Zusammenfassung der Ergebnisse zu konkludieren.

## 2 Modellanalyse

Zunächst werden die Grundannahmen getroffen, anschließend wird ein unkapazitiertes Modell vorgestellt. Darauf aufbauend wird ein kapazitiertes Modell entwickelt.

### 2.1 Annahmen

Wir gehen im einzelnen von folgenden Annahmen aus:

#### (AT) Netztopologie

Es sei eine Netztopologie gegeben, die formal durch einen ungerichteten<sup>2</sup> knotenbewerteten Graphen  $G = [V, E, c; b]$  mit  $V = \{1, \dots, n\}$  als Knotenmenge,  $E = \{e_1, \dots, e_m\}$  als Kantenmenge,  $c_i \geq 0$  mit  $i = 1, \dots, n$  als erwartete Verzögerungszeit des Knoten  $i$ <sup>3</sup>,  $b_i \geq 0$  mit  $i = 1, \dots, n$  als Nachfragemenge am Knoten  $i$  dargestellt ist. Es gilt  $e \in E$  wobei  $e = [i, j] = [j, i]$  und  $i, j \in V$ .

---

<sup>2</sup> Dies geschieht zum einen aus Vereinfachungsgründen, zum anderen ist das Internet in der Regel auf diese symmetrische Weise aufgebaut [vgl. ODLYTZKO98].

<sup>3</sup> Abweichend von der sonst im OR üblichen Kostenbewertung der Kanten, werden hier die Knoten als Kostentreiber verwendet und mit der am Knoten  $i$  erwarteten Verzögerungszeit „bewertet“.

Eine Kante kann mit einer Kapazitätsrestriktion  $(l_{ij}, k_{ij})$  belegt werden, wobei  $l_{ij} \geq 0$  die Unter- und  $k_{ij} \geq l_{ij}$  die Obergrenze bilden.

#### **(AR) Routing**

Die Knoten im Netzwerk kommunizieren auf Basis des TCP/IP Protokolls miteinander, wobei wir vereinfachend annehmen, daß die Pakete im Sinne eines kürzesten Wege Routings (shortest path routing) [TANENBAUM96, S. 348-352] versendet werden, d.h. der Aufbau der Routingtabellen ist dergestalt, daß von jedem Startknoten  $i$  die Anzahl der Hops  $d_{ij}$  zum Zielknoten  $j$  minimiert wird. Knoten- und Kabelausfälle sowie Pufferüberläufe sind definatorisch ausgeschlossen. Die Entfernungsmatrix, hier kürzeste Verzögerungsmatrix, wird durch  $D(G) = [t_{ij}]$  beschrieben.

#### **(AP) Produkt**

Die angebotene elektronische Dienstleistung<sup>4</sup> hat den (vom Unternehmen exogen festgelegten) Preis  $p \geq 0$  in Geldeinheiten und kann zu Grenzkosten null hergestellt werden ( $MC = 0$ ).

#### **(AK) Kunden**

Alle Kunden beziehen in ihre Kauf- bzw. Downloadentscheidung neben dem Preis  $p$  auch die erwarteten Kosten der Verzögerung  $v$  gleichermaßen mit ein.  $v(t)$  ist eine mit Geldeinheiten bewertete Funktion der erwarteten Verzögerungszeit  $t$ . Der Kunde sieht sich also einem Preis  $p' = p + v(t)$  gegenüber.

#### **(AN) Nachfrage**

Die jeweils am Knoten  $i$  aggregierte Nachfrage  $N_{ij}$  mit  $i, j = 1, \dots, n$  sei bekannt und eine Funktion von  $p'$ , d.h.  $N_{ij} = N_{ij}(p, v(t_{ij})) = N_{ij}(p'_{ij}) = a_i - r \cdot p'_{ij}$  falls  $j$  als Serverstandort gewählt wird und mit  $a_i$  als der autonomen Nachfrage am Knoten  $i$  und  $r$  dem Sensitivitätsparameter bzgl. des Produktpreises  $p'$ .  $N(G) = [N_{ij}]$  stellt die Nachfragematrix dar.

#### **(AZ) Zielfunktion**

Das Unternehmen versucht seinen Absatz  $Q$  der elektronischen Dienstleistung durch das intelligente Aufstellen genau eines Servers in  $G$  zu maximieren.<sup>5,6</sup>

Andere Einflußfaktoren als der Zeitfaktor seien nicht relevant.

---

<sup>4</sup> Unter elektronischen Dienstleistungen wird im folgenden jegliche entgeltliche und unentgeltliche Bereitstellung von Informationen (z.B. Webseiten) und Informationsprodukten (z.B. Musik, Finanzdienstleistungen, Software) verstanden.

<sup>5</sup> Die Definition der Zielfunktion hat einen entscheidenden Einfluß auf das spätere Ergebnis. Es sind hier durchaus auch andere Zielfunktionen vorstellbar, z.B. könnte man versuchen, die erwartete Verzögerungszeit für jeden Internet-User auf maximal 3 Sekunden zu begrenzen.

<sup>6</sup> Da hier  $MC=0$  gilt, entspricht die Nachfragemaximierung gleichzeitig der Gewinnmaximierung.

## 2.2 Unkapazitiertes Modell

Im einfachsten Fall unseres Modells wollen wir von einer unbeschränkten Bandbreitenkapazität ausgehen, d.h. Verzögerungen können nur an den Routern entstehen. Zum einen ist dieser Fall insbesondere dann interessant, wenn es sich nur um ein zu übertragendes Datenpaket bzw. sehr wenige handelt. Beispiele hierfür können Fließtext-E-mails oder Telnet Anwendungen, aber auch die Übertragung von einfachen HTML-Seiten ohne eingebettete Bild-Objekte sein. Zum anderen kann es sich auch um größere zu übertragende Datenmengen handeln, wenn das Netzwerk in der Regel wenig ausgelastet ist.

Der erste Schritt in der Optimierung ist grundsätzlich die Bestimmung der kürzesten Verzögerungsmatrix. Hierzu läßt sich beispielweise das Verfahren von Dijkstra anwenden [DOMSCHKE95, S 101 - 110]. Dieses Vorgehen soll an einem Beispiel verdeutlicht werden.

**Beispiel 1:** Gegeben sei das Backbone Netzwerk des Deutschen Forschungs-Netzwerks (DFN) (siehe Abbildung 1). Das DFN plane die kostenlose Bereitstellung eines Forschungsergebnisses auf einem eigens dafür bereit gestellten Server. Alle Knoten bis auf Frankfurt seien (fiktiv) mit  $c_i = 1$  bewertet; Frankfurt als Gateway zu anderen Backbone Netzwerken erhält die (fiktive) Bewertung  $c_6 = 5$ . Die Leitungen unterliegen keiner Kapazitätsbeschränkung, also  $k_{ij} = \infty$ .

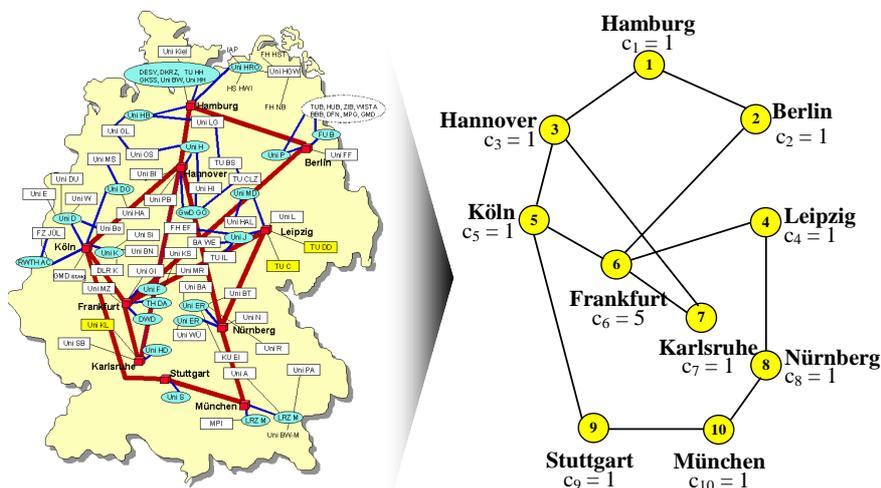


Abbildung 1: Schematischer Aufbau des DFN Backbone

Die kürzeste Wegematrix läßt sich nun mit dem Verfahren von Dijkstra berechnen.



**Beispiel 2:** Wir verwenden die Daten und Ergebnisse aus Beispiel 1. Weiterhin sei  $p = 0$  (kostenloser Forschungsbericht) und die Nachfragen bzw. Nachfragefunktionen an den Knoten wie folgt gegeben:

Knotenname	Knotennummer	Nachfragefunktion
Hamburg	1	$100 - 3p'$
Leipzig	4	$30 - 3p'$
Frankfurt	6	$120 - 3p'$
Karlsruhe	7	$150 - 3p'$
München	10	$100 - 3p'$
alle anderen	2, 3, 5, 8, 9	0

**Tabelle 1: Nachfragedaten an den Knoten im Beispiel 2**

Vereinfachend sei hier angenommen, daß die errechnete erwartete Verzögerungszeit genau der mit Geldeinheiten bewerteten Verzögerung entspricht, d.h.  $D(G) = V(G)$ . Da  $p = 0$ , entspricht  $V(G)$  auch gleichzeitig der Matrix der erwarteten Produkt-/Downloadgesamtkosten  $[p'_{ij}]$ . Die Nachfragematrix sieht demnach wie folgt aus:

$$N(G) = \begin{matrix} & \begin{matrix} \text{Zielknoten } j \\ \hline 97 & 0 & 0 & 9 & 0 & 99 & 141 & 0 & 0 & 85 \\ 94 & 0 & 0 & 9 & 0 & 102 & 138 & 0 & 0 & 82 \\ 94 & 0 & 0 & 12 & 0 & 99 & 144 & 0 & 0 & 88 \\ 79 & 0 & 0 & 27 & 0 & 102 & 129 & 0 & 0 & 91 \\ 91 & 0 & 0 & 15 & 0 & 102 & 141 & 0 & 0 & 91 \\ 79 & 0 & 0 & 12 & 0 & 105 & 132 & 0 & 0 & 76 \\ 91 & 0 & 0 & 9 & 0 & 102 & 147 & 0 & 0 & 85 \\ 82 & 0 & 0 & 24 & 0 & 99 & 132 & 0 & 0 & 94 \\ 88 & 0 & 0 & 18 & 0 & 99 & 138 & 0 & 0 & 94 \\ 85 & 0 & 0 & 21 & 0 & 96 & 135 & 0 & 0 & 97 \end{matrix} \\ \left. \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix} \right\} \text{Startknoten } i \end{matrix}$$

Dies ergibt folgende (transponierte) Gesamtnachfragematrix:

$$GN(G)^T = (431 \ 425 \ 437 \ 428 \ 440 \ 404 \ 434 \ 431 \ 437 \ 434)$$

$$\Rightarrow Q^* = \max GN(G)^T = 440$$

und  $\max GN(G)$  liefert Knoten 5 (Köln) mit einer Nachfrage von  $Q^* = 440$  als optimalen Serverstandort. Auf Basis dieses unkapazitierten Modells kann das Zwischenfazit gezogen werden, daß der physische Standort eines Servers einen Einfluß auf den Unternehmenserfolg beim Vertrieb von elektronischen Dienstleistungen haben kann.

Aufbauend auf dieses Modell wenden wir uns nun dem kapazitierten Fall in Abschnitt 2.3 zu.

## 2.3 Kapazitiertes Modell

Im Gegensatz zu oben dargestelltem Modell beschreiben wir in diesem Absatz den wesentlich realitätsnäheren Fall, in dem wir sowohl unterschiedliche Bandbreiten [KAUFFELS97, S. 42], i.d.R. von 56 Kbit/s bei analogen Leitungen bis 622 Mbit/s bei Backbone Verbindungen [KAUFFELS97, S. 43], unterschiedliche Übertragungsraten als auch die Verarbeitungsraten der Router miteinbeziehen. Wie ebenfalls schon oben erwähnt, entstehen Verzögerungen hauptsächlich an Routern, da die Übertragungsraten eines Netzes in der Regel nicht an die Verarbeitungsraten der Router angepaßt sind und die Netzauslastungen stark variieren. Wenn aber die Übertragungsraten nicht den Verarbeitungsraten entsprechen, entstehen Warteschlangen, wie sie auch aus dem Alltag von Bankschaltern, Straßenampeln oder in der Produktion bekannt sind. Daher wird im folgenden das Konzept der Warteschlangentheorie auf dieses Problem übertragen.

### 2.3.1 Warteschlangentheorie

Ein Warteschlangensystem läßt sich auch als Input-Output-System darstellen, mit Warteraum und Abfertigung [DOMSCHKE91, S. 185]. Übertragen auf den vorliegenden Fall entspricht der Warteraum des Warteschlangensystems den Puffern des Routers und die Abfertigung kann als das Weiterversenden von Datenpaketen auf Basis der Routingtabelle betrachtet werden.

Die Warteschlangentheorie führt das Entstehen von Warteschlangen darauf zurück, daß zufällige Schwankungen in der Bediendauer und den zeitlichen Abständen zwischen zwei Ankünften gegeben sind. Übertragen auf unseren Fall bedeutet dies, daß ein nicht unerheblicher Unterschied zwischen den Kapazitäten der einzelnen Leitungen und den Verarbeitungsraten der Router besteht. Vor allem sind auch die sogenannte Peaks von großem Interesse, d. h. wenn auf den Datenautobahnen Staus auf besonders frequentierten Leitungen entstehen. Lassen sich die Staus nicht vermeiden und die Warerräume (Puffer) der einzelnen Router laufen über, gehen auch Datenpakete verloren [TANENBAUM96, S. 374]. In unserer Arbeit vernachlässigen wir aus Komplexitätsgründen den gerade erwähnten Fall.

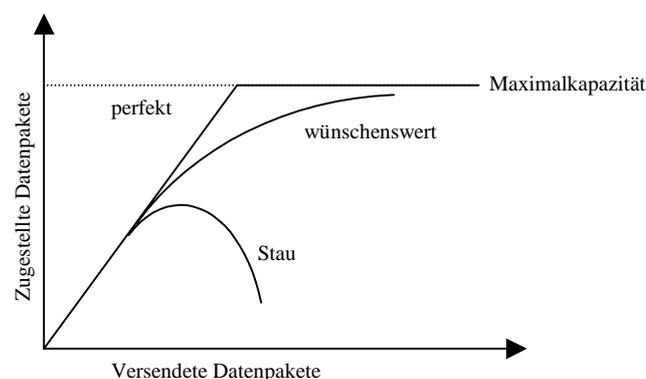


Abbildung 2: Netzwerkperformance [TANENEBAUM96, S.374]

Oben abgebildete Darstellung verdeutlicht den Zusammenhang zwischen möglicherweise auftretenden Staus auf den Datenautobahnen und den damit einhergehenden Performanceeinbußen. Je besser das Verhältnis zwischen versendeten und zugestellten Datenpaketen ausfällt, desto besser fällt auch die allgemeine Netzperformance aus. Der ideale Fall ist gegeben, falls alle Datenpakete, die versendet werden, auch bei den entsprechenden Empfängern ankommen. Somit ist eine Auslastung der maximal gegebenen Kapazitäten garantiert. Der worst case tritt dann auf, wenn sich das Verhältnis zwischen gesendeten und zugestellten Paketen derart verschlechtert, daß sich Staus bilden, was für den Nutzer lange Wartezeiten bedeutet.

Für die Betrachtung ist es unerheblich, ob von Außen, d.h. ein User entscheidet sich für die Nutzung eines Dienstes des Internets, oder von Innen, d.h. von einem vorangehenden Router, Datenpakete an einer Station/Router ankommen. Datenpakete treten an einem Knoten in das System ein, passieren einen oder auch mehrere Router und verlassen das System wieder, wobei Eintritts- und Austrittsstelle nicht identisch sein müssen.

### 2.3.2 Warteschlangenmodell

Zusätzlich zu den bereits getroffenen sind zwei weitere Annahmen nötig:

#### **(AW) Warteschlange**

Ankünfte von „Außen“ an einem Knoten  $i$  folgen einer Poisson-Verteilung mit einer *mittleren Ankunftsrate* von  $\lambda$  pro Zeiteinheit. Die *Abfertigungsraten* an jedem Knoten  $i$  sind unabhängig voneinander und mit einem Parameter  $\mu$  (mittlere Anzahl der Einheiten, die pro Zeiteinheit abgefertigt werden können) exponential verteilt. Dabei muß die Möglichkeit des Wartens bestehen und das Anstellen erfolgt in geordneten Reihen. Es gilt das strenge First-In First-Out Prinzip, d.h. zuerst ankommende Pakete werden auch zuerst weiter versendet.

#### **(AS) Warteraum**

Es handelt sich um ein System mit begrenztem Warteraum, in dem nur eine bestimmte endliche Zahl von Einheiten zugelassen wird. Die Zahl der Ankünfte ist abhängig von der Zahl der bereits im Warteraum befindlichen Einheiten und die maximal mögliche Schlangenlänge ist gleich der Warteraumkapazität  $m$ .

Aus obigen Annahmen läßt sich folgern, daß sich ein Netzwerk so verhält, als ob jeder Knoten als ein unabhängiges  $M/M/1$ -System mit den Parametern  $\lambda$  und  $\mu$  angesehen werden kann. Somit kann das gesamte Netzwerk als eine Ansammlung einzelner  $M/M/1$ -Systeme betrachtet wer-

den. Es werden daher die allgemeingültigen formalen Zusammenhänge eines  $M/M/1$ -Systems<sup>7</sup> zugrunde gelegt [ZIMMERMANN86, S. 366].

Für die Wahrscheinlichkeit, daß das System leer ist, gilt demzufolge

$$P_0 = \frac{1}{\sum_{n=0}^m \left[ \frac{m!}{(m-n)!} \left( \frac{\lambda}{\mu} \right)^n \right]} \tag{5}$$

und die mittlere Schlangenlänge bezogen auf alle Einheiten ist

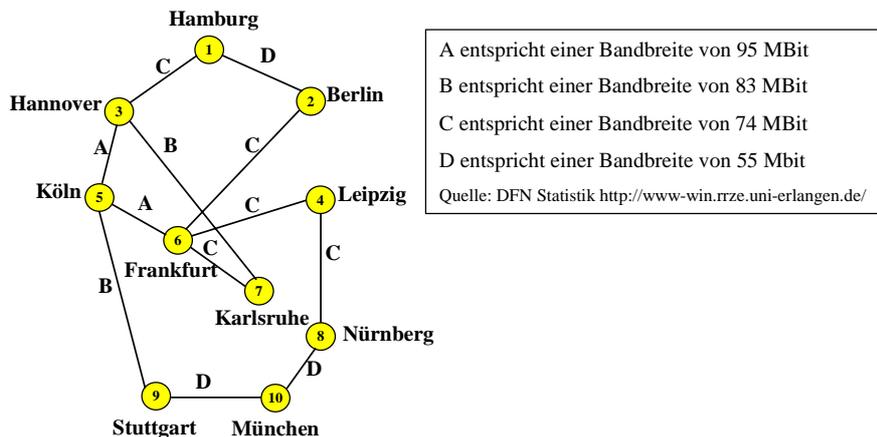
$$L(q) = m - \frac{\lambda + \mu}{\lambda} (1 - P_0) \tag{6}$$

Über diese Zusammenhänge hinaus wird die durchschnittliche Wartezeit  $t(q)$  eines Datenpaketes im Warteraum als Quotient der mittleren Schlangenlänge  $L(q)$  bezogen auf alle Einheiten  $q$  und der Ankunftsrate  $\lambda$  bestimmt.

$$t(q) = \frac{L(q)}{\lambda} \tag{7}$$

Damit ist die durchschnittliche Wartezeit  $t$  für ein  $M/M/1$ -System ermittelt und die Formeln (1) – (4) können analog angewendet werden. Dies wird in einem weiteren Beispiel verdeutlicht.

**Beispiel 3:** Aufbauend auf Beispiel 1 (DFN-Netzwerk) wird unterstellt, daß keine Technologievorteile einzelner Router bestehen ( $m$  wird fix auf 3 gesetzt). In Abhängigkeit der gegebenen Bandbreiten werden unten stehende Werte für  $\mu$  und  $\lambda$  angenommen (Tabelle 2).



**Abbildung 3: DFN Backbone mit 4 Bandbreitenkategorien**

Mit den getroffenen Annahmen und unter Anwendung von (5), (6) und (7) lassen sich folgende Werte ermitteln.

<sup>7</sup> Genau genommen liegt ein  $M/M/1:(m/FIFO)$  System vor, d.h. ein Ein-Kanal-System mit begrenztem Warteraum, bei dem die Abfertigungsreihenfolge gemäß dem First-In First-Out Prinzip folgt.

	$m$	$\mu$	$\lambda$	$P_0$	$L(q)$	$t(q)$
A	3	0,95	0,5	0,19	0,651	1,3
B	3	0,83	0,5	0,16	0,766	1,5
C	3	0,74	0,5	0,13	0,842	1,7
D	3	0,55	0,5	0,08	1,068	2,1

**Tabelle 2: Datentableau zu Beispiel 3**

Am Beispiel der Bandbreitenkategorie D werden die berechneten Parameter wie folgt interpretiert. Die Wahrscheinlichkeit, daß sich null Elemente im System befinden und somit kein Stau auftritt beträgt 8%, die mittlere Schlangelänge bezogen auf alle Einheiten beläuft sich auf 1,068 Elemente und die durchschnittliche Wartezeit eines Elementes, d.h. die Zeit, bis das Paket den Standort verläßt, beträgt bei einer Bandbreite der Kategorie D 2,1 ZE (Zeiteinheiten). Eine Interpretation der Ergebnisse für die Kategorien A bis C erfolgt analog zu D.

Zur Ermittlung der kürzesten Verzögerungsmatrix werden die für  $t(q)$  ermittelten Werte verwendet. Die berechneten Werte  $t(q)$  unterstützen die Einteilung der Anbindungen in die vier verschiedenen Kategorien, da  $t(q)$  wie folgt interpretiert wird: Die durchschnittliche Wartezeit eines Datenpaketes bei einer Anbindung der Klasse A ist um etwa 62% geringer als bei einer Anbindung der Klasse D. Damit errechnet sich die kürzeste Verzögerungsmatrix mittels des Verfahrens von Dijkstra zu:

$$D(G) = \begin{pmatrix} 0 & 2,1 & 1,7 & 5,5 & 3 & 3,8 & 3,2 & 7,2 & 4,5 & 6,6 \\ 2,1 & 0 & 3,8 & 3,4 & 3,0 & 1,7 & 3,4 & 5,1 & 4,5 & 6,6 \\ 1,7 & 3,8 & 0 & 4,3 & 1,3 & 2,6 & 1,5 & 6,0 & 2,8 & 4,9 \\ 5,5 & 3,4 & 4,3 & 0 & 3,0 & 1,7 & 3,4 & 1,7 & 4,5 & 3,8 \\ 3 & 3,0 & 1,3 & 3,0 & 0 & 1,3 & 3,0 & 4,7 & 1,5 & 3,6 \\ 3,8 & 1,7 & 2,6 & 1,7 & 1,3 & 0 & 1,7 & 3,4 & 3,8 & 4,9 \\ 3,2 & 3,4 & 1,5 & 3,4 & 3,0 & 1,7 & 0 & 5,1 & 4,5 & 6,6 \\ 7,2 & 5,1 & 6,0 & 1,7 & 4,7 & 3,4 & 5,1 & 0 & 4,2 & 2,1 \\ 4,5 & 4,5 & 2,8 & 4,5 & 1,5 & 2,8 & 4,5 & 4,2 & 0 & 2,1 \\ 6,6 & 6,6 & 4,9 & 3,8 & 3,6 & 4,9 & 6,6 & 2,1 & 2,1 & 0 \end{pmatrix}$$

Nun kann in einem zweiten Schritt auf Basis der neuen Verzögerungsmatrix und der in Tabelle 1 definierten Nachfragefunktionen folgende Nachfragematrix aufgestellt werden.

$$N(G) = \begin{pmatrix} 100 & 0 & 0 & 13,5 & 0 & 108,6 & 140,4 & 0 & 0 & 80,2 \\ 93,7 & 0 & 0 & 19,8 & 0 & 114,9 & 139,8 & 0 & 0 & 80,2 \\ 94,9 & 0 & 0 & 17,1 & 0 & 112,2 & 145,5 & 0 & 0 & 85,3 \\ 83,5 & 0 & 0 & 30 & 0 & 114,9 & 139,8 & 0 & 0 & 88,6 \\ 91 & 0 & 0 & 21 & 0 & 116,1 & 141 & 0 & 0 & 89,2 \\ 88,6 & 0 & 0 & 24,9 & 0 & 120 & 144,9 & 0 & 0 & 85,3 \\ 90,4 & 0 & 0 & 19,8 & 0 & 114,9 & 150 & 0 & 0 & 80,2 \\ 78,4 & 0 & 0 & 24,9 & 0 & 109,8 & 134,7 & 0 & 0 & 93,7 \\ 86,5 & 0 & 0 & 16,5 & 0 & 111,6 & 136,5 & 0 & 0 & 93,7 \\ 80,2 & 0 & 0 & 18,6 & 0 & 105,3 & 130,2 & 0 & 0 & 100 \end{pmatrix}$$

Daraus ergibt sich dann folgende (transponierte) Gesamtnachfragematrix:

$$GN(G)^T = (442,7 \quad 448,4 \quad 455 \quad 456,8 \quad 458,3 \quad 463,7 \quad 455,3 \quad 441,5 \quad 444,8 \quad 434,3)$$

$$\Rightarrow Q^* = \max GN(G)^T = 463,7$$

Und  $\max GN(G)$  liefert Knoten 6 (Frankfurt) mit einer Nachfrage von  $Q^* = 463,7$  als optimalen Serverstandort.

Mit dem Ergebnis, daß auch im kapazitierten Fall in der Regel *ein* optimaler Serverstandort bzgl. der spezifizierten Zielfunktion formal bestimmt werden kann, wenden wir uns nun den Limitierungen der Analyse und möglichen Erweiterungen für zukünftige Arbeiten zu.

### 3 Limitierungen und Erweiterungen

- **Nachfrage:** Durch die Annahme der statischen, homogenen und linearen Nachfrage wird das Modell zwar erst handhabbar, gleichzeitig schränkt diese die Aussagekraft der Ergebnisse stark ein. Ziel war es jedoch die These zu stützen, daß die unterschiedlichen Übertragungszeiten substantielle betriebswirtschaftliche Konsequenzen haben können. Auf Basis des Modells ist eine Tendenzaussage möglich, eine Quantifizierung könnte in einem weiteren Schritt untersucht werden.
- **Netzwerk und Routing:** Des weiteren ist ein statisches Netzwerk und statisches Routing unterstellt, was so in der Realität nicht zutrifft. Ständig werden Subnetze erweitert und umgebaut. Routingtabellen werden dynamisch an die momentane Netzauslastung angepaßt. Aus Komplexitätsgründen wurde hiervon in der Analyse abstrahiert.
- **Verteilungsannahme:** Die Annahme der Exponentialverteilung im Rahmen des kapazitierten Modells konnte empirisch nicht bestätigt werden [WILLINGER95]. Vielmehr wurde eine „long range dependence“ und eine unendliche Varianz der Verteilung der Ankunftszeiten gemessen.

- **Datenpaketbetrachtung:** In beiden Modellen wurde nur ein einzelnes Datenpaket betrachtet, insbesondere bei Kapazitätsbeschränkungen hinsichtlich der Bandbreite erscheint jedoch eine Datenflußbetrachtung geeigneter.
- **Eindimensionale Zielfunktion:** Von Einflußfaktoren außer der Zeit wurde in der Analyse abstrahiert. Für eine fundierte finanzwirtschaftliche Betrachtung müßten jedoch alle Standortfaktoren, welche die Standortentscheidung beeinflussen, berücksichtigt werden.

Mit diesen Limitierungen wenden wir uns möglichen Erweiterungen zu. Erstes Ziel sollte die Einbeziehung der (neuen) Erkenntnisse über die Verteilung der Ankunftsrate und die Anpassung des kapazitierten Modells auf eine Datenflußbetrachtung sein. Des weiteren sollte ein Modell mit mehrdimensionaler Zielfunktion entwickelt werden, um die verschiedenen Einflußfaktoren zu berücksichtigen. Denkbar wäre auch eine Erweiterung des Verfahrens auf mehrere Serverstandorte unter Ausnutzung von State-of-the-Art Caching Technologien [z.B. CACHEFLOW00]. Eine andere Erweiterungsmöglichkeit stellt die Einbeziehung unterschiedlicher Kosten in die Betrachtung dar. Hierbei sind zwei Kostenbereiche sowie verschiedene Sichtweisen zu unterscheiden. Zum einen sind standortabhängige Kosten zu nennen, die je nach Geschwindigkeit der Anbindung und maximaler Bandbreitennutzungsmöglichkeit entstehen, d.h. je nach Wahl des Providers und dessen Anbindung entstehen unterschiedlich hohe Kosten. Beispielsweise belaufen sich Server Hosting Kosten einer 2 MBit Anbindung auf einen wesentlich geringeren Betrag als die Kosten bei einem Anschluß an einen 34 MBit Backbone. Zum anderen fallen auch sogenannte "Durchleitungskosten" an, d.h. die Backbone Netzbetreiber berechnen ein volumenabhängiges Entgelt für die Nutzung ihrer Leitungen. Je nach Anzahl der übertragenen Daten für die Nutzung ihres Netzes wird ein bestimmter Betrag angesetzt. Die so entstehenden Durchleitungskosten werden durch den Provider an den Serverbetreiber umgelegt.

#### 4 Zusammenfassung

Ausgangspunkt der Überlegungen dieses Beitrags war die These, daß die physische Standortentscheidung für einen Server aufgrund unterschiedlicher Übertragungszeiten substantielle betriebswirtschaftliche Konsequenzen hat und daher in die Standortentscheidung mit einzubeziehen ist. Hierzu wurden Modelle des Operations Research auf das Internet übertragen (hier: Warehouse Location Problem). Sowohl im unkapazitierten Modell als auch im kapazitierten Modell unter Anwendung der Warteschlangentheorie bestätigt sich die These. Es kann also davon ausgegangen werden, daß die Standortentscheidung für einen Server keinesfalls nur alleine von den Kosten oder den steuerlichen Rahmenbedingungen, sondern auch die erwartete Verzögerungszeit, die

für den Konsumenten beim Kauf bzw. Download von elektronischen Dienstleistungen entsteht, in die Standortentscheidung einzubeziehen ist. Auf Basis des vorgestellten Modells kann in der Regel *ein* optimaler Serverstandort bestimmt werden.

## Bibliographie

- [CACHEFLOW00] CacheFlow Inc., White Papers. Erhältlich bei <http://www.cacheflow.com/technology/whitepapers/web.cfm>, Download am 28.08.2000
- [DOMSCHKE95] Domschke, Wolfgang: Logistik: Transport, 4. Auflage, Oldenbourg Verlag, 1995.
- [DOMSCHKE91] Domschke, Wolfgang, Drexl, Andreas: Einführung in Operations Research, 2. Auflage, Springer Verlag, 1991.
- [ECIN00] Electronic Commerce InfoNet (ECIN): Nutzerzahlen - wie hoch ist die Reichweite des Internets? Erhältlich bei: <http://www.ecin.de/marktbarometer/daten/nutzer.html>, Download am 04.02.2000.
- [HIMMELEIN99] Gerald Himmelein, Johannes Endres Kost und Logis frei 18 kostenlose Webhosting-Angebote ab 10 Megabyte, in c't 18/99, S. 126
- [ZAKON00] Robert H'obbes' Zakon: Hobbes' Internet Timeline v5.0. Erhältlich bei: <http://www.isoc.org/zakon/Internet/History/HIT.html>
- [KAUFFELS97] Kauffels, Franz-Joachim: Moderne Datenkommunikation, Eine strukturierte Einführung, 2. Auflage, International Thomson Publishing, Bonn, 1997.
- [ODLYZKO98] Odlyzko, Andrew, The low utilization and high cost of data networks, 1998, erhältlich bei <http://www.research.att.com/~amo/doc/recent.html>.
- [PAXSON97] Paxson, Vern: Measurement and Analysis of End-to-End Internet Dynamics., Ph.D. Thesis, Computer Science Division, University of California, Berkeley, April 1997.
- [SATZGER99a] Satzger, Gerhard: Electronic Commerce: Ein Instrument zur Steuergestaltung? in: Wirtschaftsinformatik, 41, 1, 1999, S. 40-47.
- [SATZGER99b] Satzger, Gerhard: Die Besteuerung des internationalen elektronischen Handels und deren Auswirkungen auf Leistungsangebote und Standortentscheidungen, in: Egger, A., Grün, O., Moser, R., Hrsg., Managementinstrumente und -konzepte, Proceedings der 60. Wissenschaftlichen Jahrestagung des Verbandes der Hochschullehrer für Betriebswirtschaft e.V., Wien, 1999, S. 225-240.
- [TANENBAUM96] Tanenbaum, Andrew: Computer Networks, 3. Auflage, Prentice Hall, 1996.
- [WILLINGER95] Willinger, Walter, Taqqu, Murad, Leland, Will, Wilson, Daniel, Self-Similarity in High-Speed Packet Traffic: Analysis and Modelling of Ethernet Traffic Measurements, Statistical Science, Vol. 10, No. 1, 1995, S. 67 – 85.
- [ZONA99] Zona Research: The need for speed. Erhältlich bei <http://www.zonaresearch.com/promotions/needforspeed>, Juni 1999.
- [ZIMMERMANN86] Zimmermann, Werner: Operations Research, Quantitative Methoden zur Entscheidungsvorbereitung, 3. Aufl., Oldenbourg Verlag, München, 1986.