

Automatisierte Ressourcen- planung in dienstorientierten Netzinfrastrukturen

In Kooperation mit dem BMBF Projekt MAMS/MAMSPplus

ITG FG 5.2.1 – Workshop „Network Resource Management“ 2009

Andreas Reifert

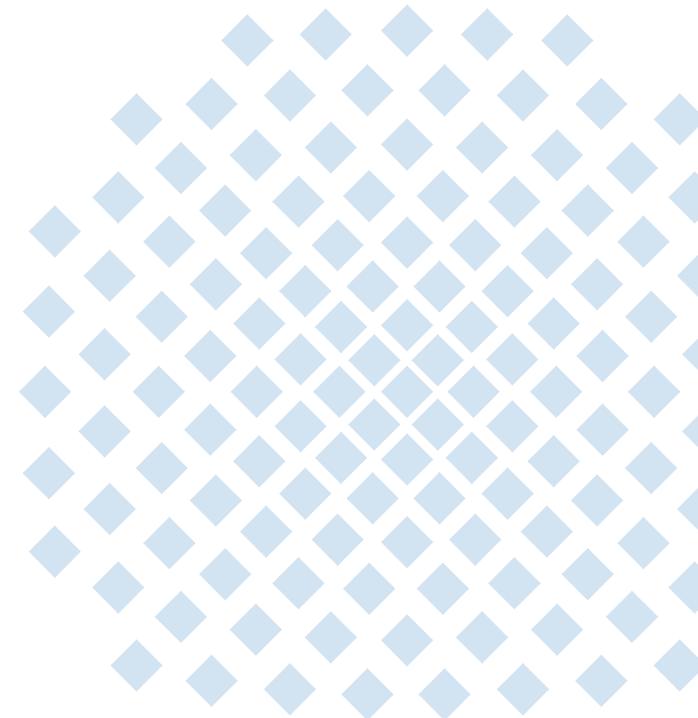
andreas.reifert@ikr.uni-stuttgart.de

19. März 2009

Universität Stuttgart

Institut für Kommunikationsnetze
und Rechnersysteme (IKR)

Prof. Dr.-Ing. Dr. h.c. mult. P. J. Kühn



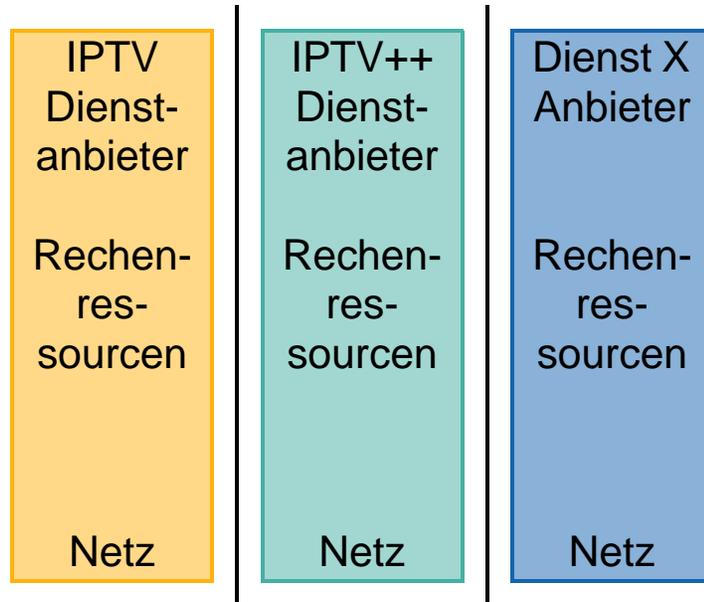
Agenda

- Motivation – Bereitstellung von Diensten
 - Silo-Ansatz
 - Trennung Dienstanbieter (Dienst und Dienstkomponenten) und Dienstbetreiber (dienstorientierte Netzinfrastruktur)
 - Platzierung von Dienstkomponenten auf Ressourcen der Infrastruktur
- Modelle für automatisierte Platzierung
 - Dienstmodell
 - Infrastrukturmodell
- Platzierungsstrategien
 - Untersuchungsraum
 - Zwei Strategien als Grenzfälle
- Untersuchungen und vorläufige Ergebnisse
- Zusammenfassung

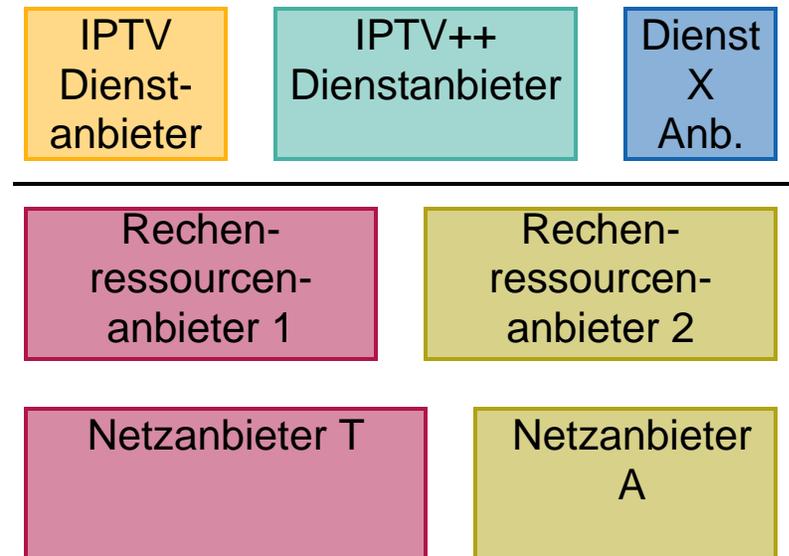
Motivation

Bereitstellung von Diensten

Silo Ansatz



Vertikale Trennung



Bewertung

- Dedizierte Ressourcen für Komponenten eines Dienstes
- Fachwissen für Dienstimplementierung notwendig
- Individuell schnelle Dienstentwicklung
- Klare Trennung Dienstanbieter / Dienstbetreiber („Infrastrukturanbieter“)
- Framework an Schnittstellen notwendig
- In Summe schnelle Dienstentwicklung
- Gute Ressourcenauslastung

Motivation

Beispiel IPTV

Teildienste

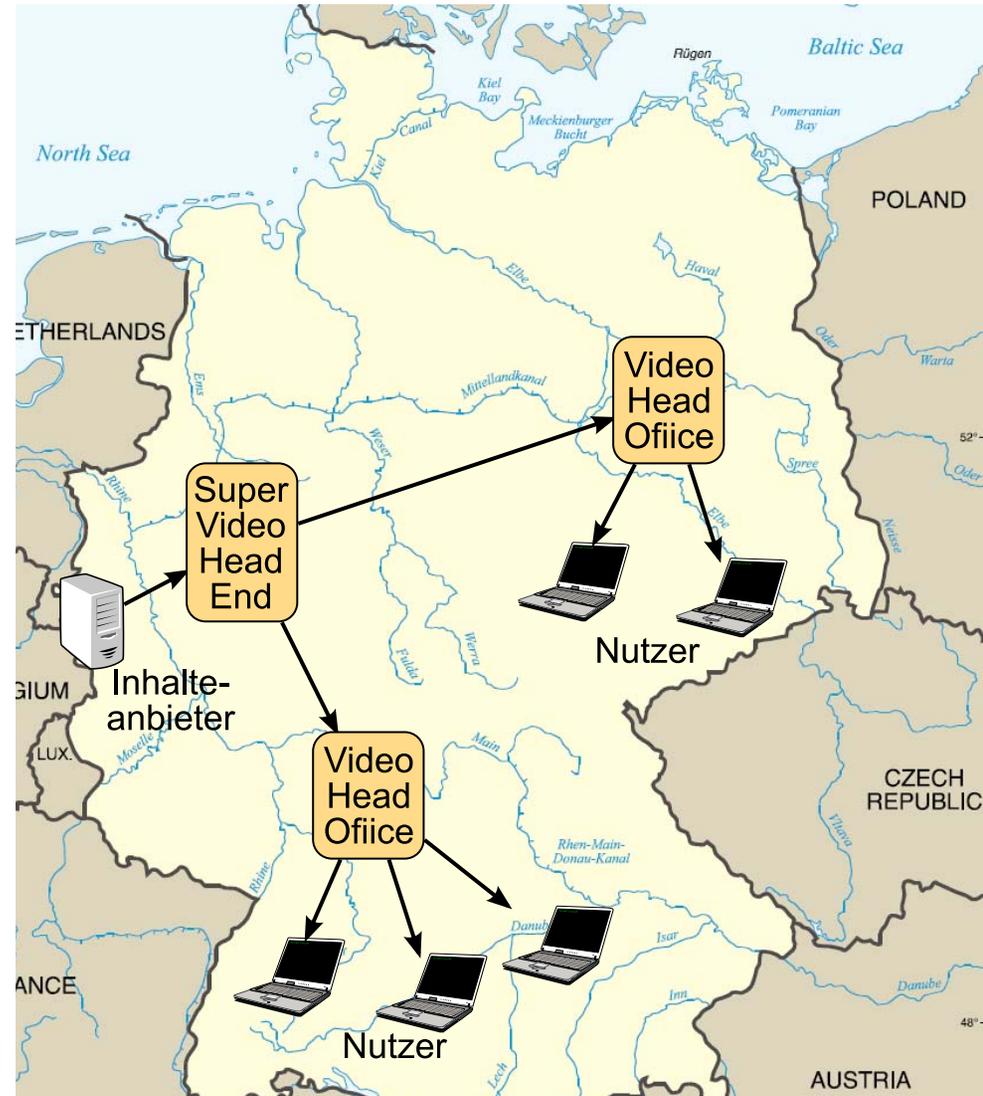
- Übermittlung von Fernsehprogrammen (Aufzeichnung oder Live)
- Video-on-Demand
- Time-shift Video

Anbindung von

- Inhaltenanbieter
- Nutzer

Architektur

- Super Video Head End
- Video Head Office



Motivation

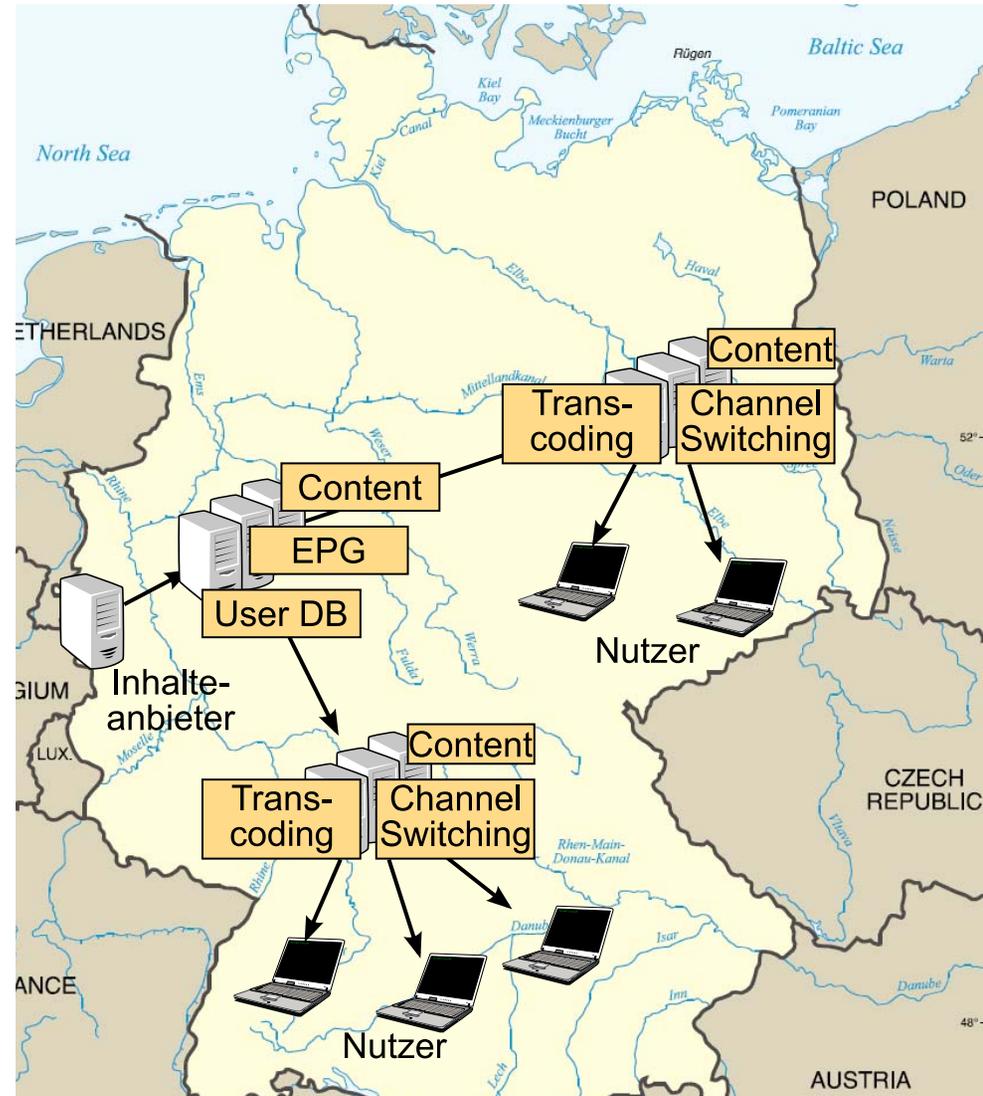
Beispiel IPTV

Funktionen / Dienstkomponenten

- Datenbank und lokale Replikationen für Inhalte
- Benutzerdatenbank
- Electronic Programming Guide (EPG)
- Transcoding
- Kanalwechsel

Realisierung

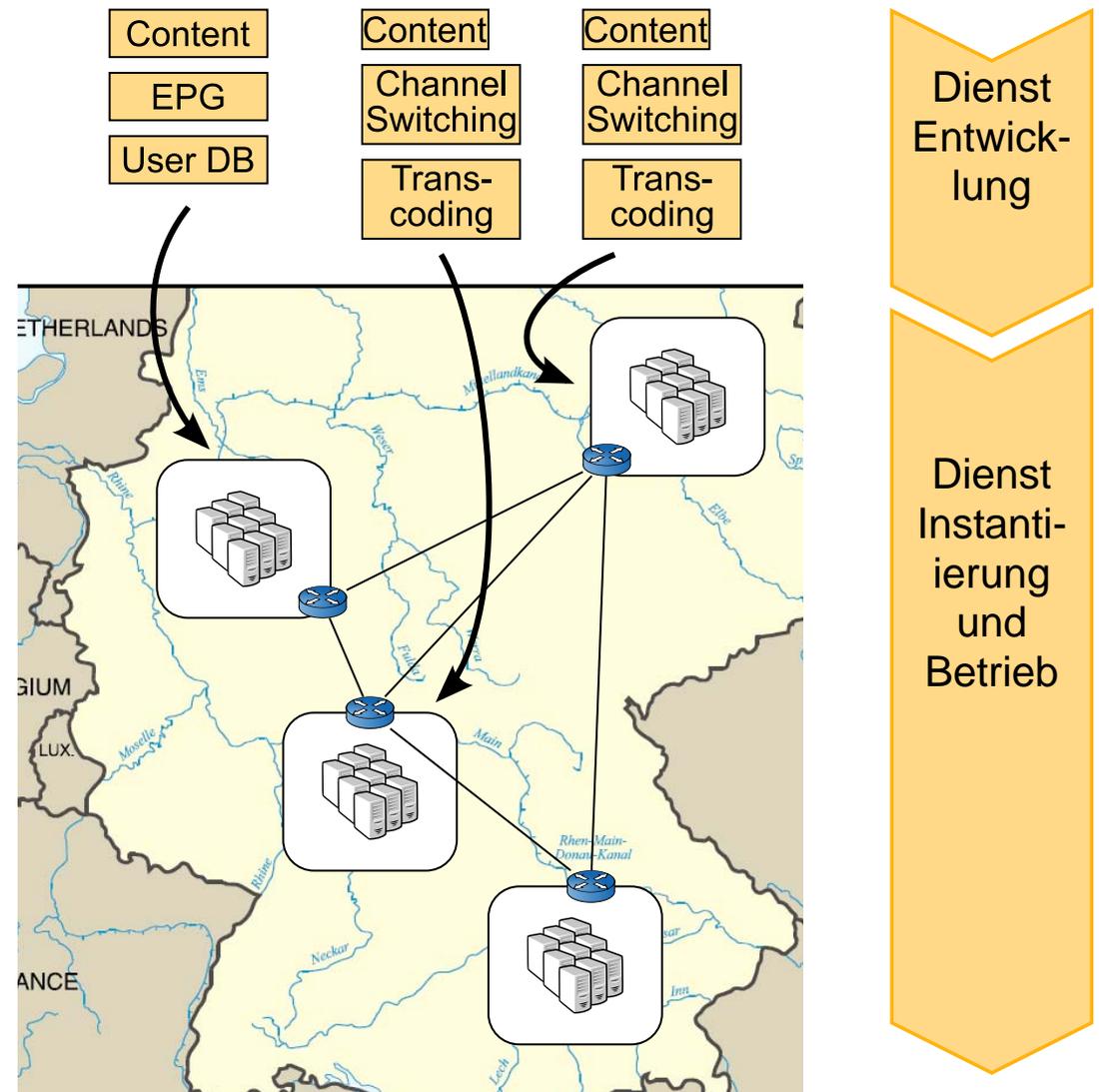
- Eigene Ressourceninfrastruktur für Komponenten
 - Feste Zuordnung der Dienstkomponenten zu Ressourcen
- Dimensionierungsfrage



Motivation

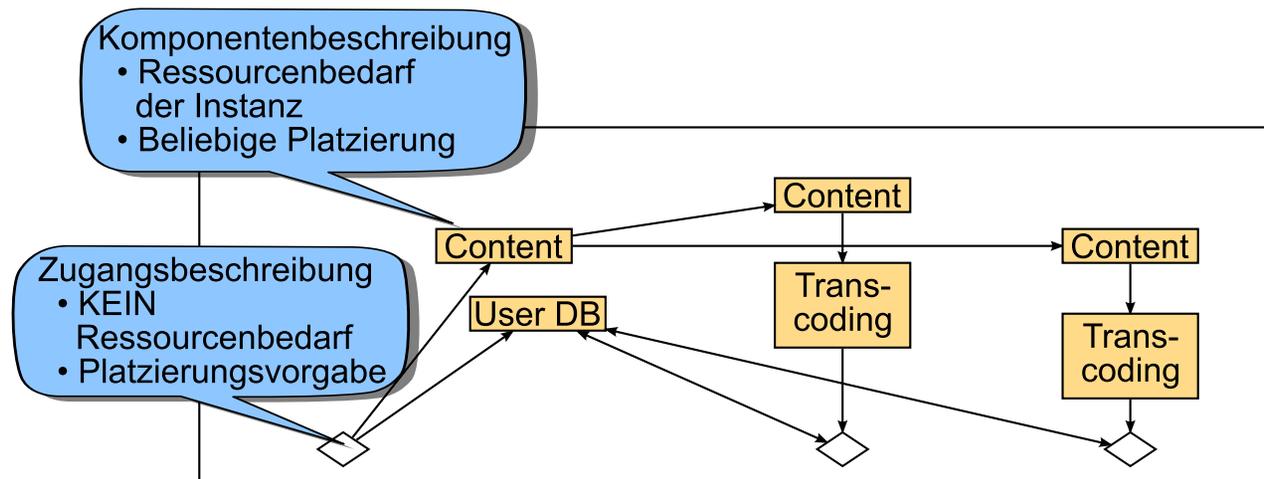
Dienstorientierte Netzinfrastruktur

- Trennung der Kopplung zwischen Komponenten und Ressourcen
- Schmale und flexible Schnittstelle durch Virtualisierung der Infrastruktur
- Entwicklung des Dienstes durch Dienstanbieter
- Bereitstellung der Ressourcen durch Dienstbetreiber (= „Infrastrukturanbieter“)
- Automatisierung der Platzierung der Dienstkomponenten
→ Modell für Dienst und Infrastruktur notwendig



Modellierung

... eines Dienstes (Dienstbeschreibung)

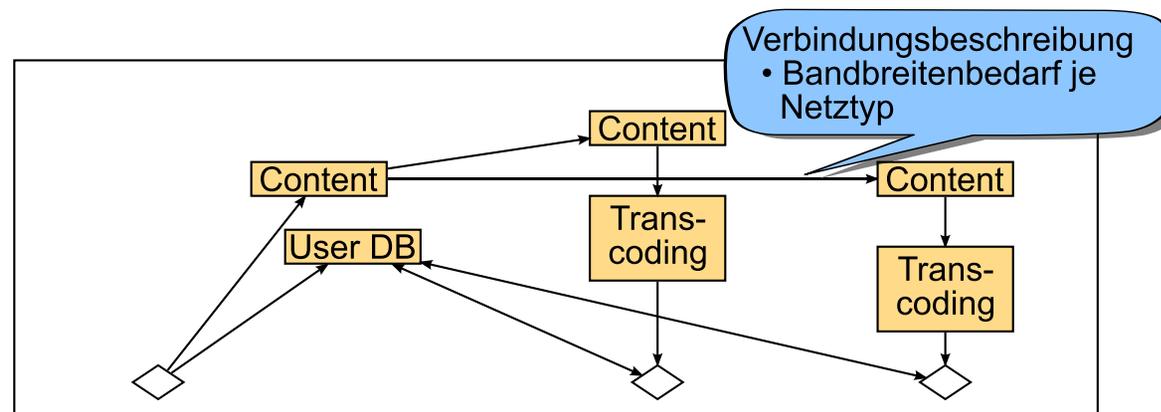


Diensttopologie bestehend aus

- Komponentenbeschreibungen
 - Ressourcenbedarf der zugehörigen Komponente
- Platzierungsanforderungen
- Verbindungsbeschreibungen
 - Bandbreitenbedarf
 - Je nach Platzierung der Komponenten mehrere Angaben möglich
- Verzögerungsanforderungen
 - Von Verbindungen
 - Auf Verbindungspfaden

Modellierung

... eines Dienstes (Dienstbeschreibung)

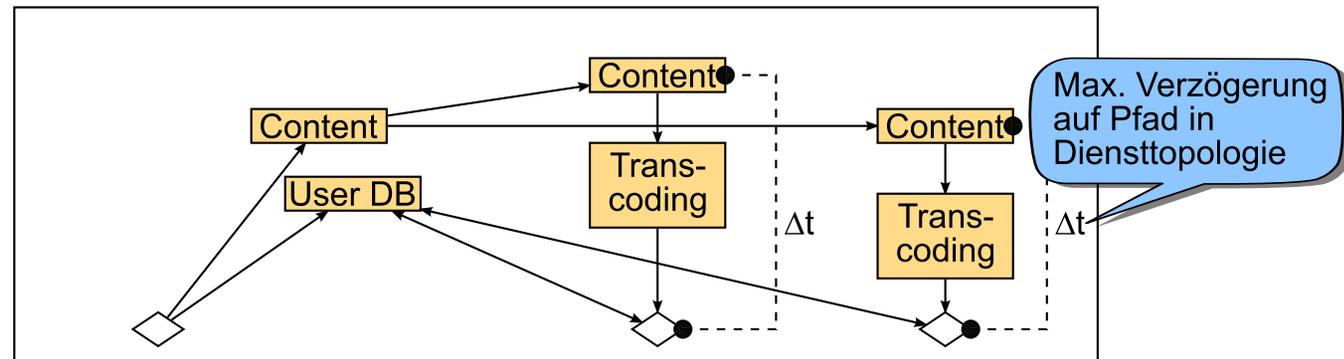


Diensttopologie bestehend aus

- Komponentenbeschreibungen
 - Ressourcenbedarf der zugehörigen Komponente
- Platzierungsanforderungen
- Verbindungsbeschreibungen
 - Bandbreitenbedarf
 - Je nach Platzierung der Komponenten mehrere Angaben möglich
- Verzögerungsanforderungen
 - Von Verbindungen
 - Auf Verbindungspfaden

Modellierung

... eines Dienstes (Dienstbeschreibung)



Diensttopologie bestehend aus

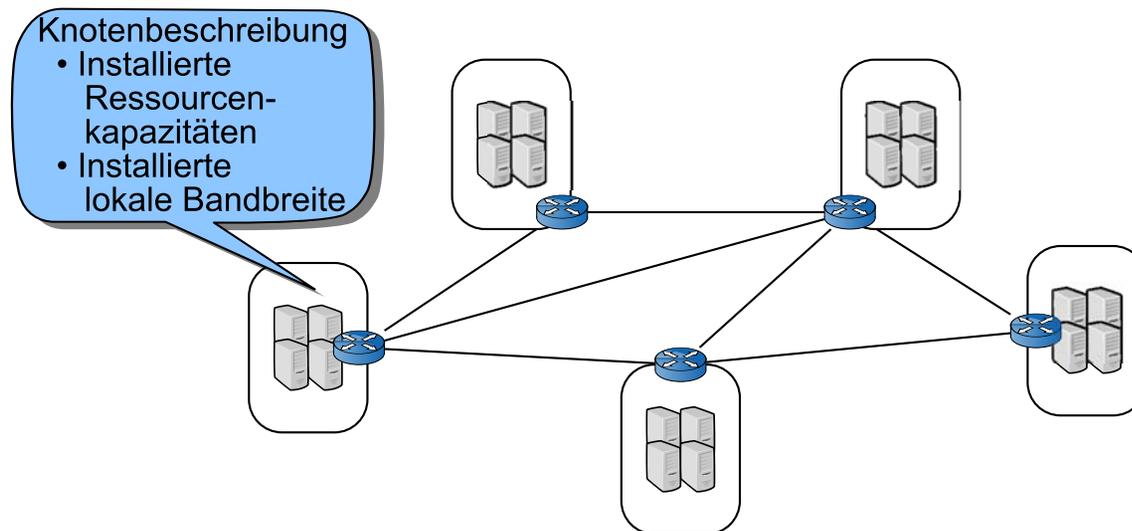
- Komponentenbeschreibungen
 - Ressourcenbedarf der zugehörigen Komponente
- Platzierungsanforderungen
- Verbindungsbeschreibungen
 - Bandbreitenbedarf
 - Je nach Platzierung der Komponenten mehrere Angaben möglich
- Verzögerungsanforderungen
 - Von Verbindungen
 - Auf Verbindungspfaden

Modellierung

... der Ressourcen (Ressourcenbeschreibung)

Ressourcentopologie bestehend aus

- Knoten
 - Ausführungsumgebung für Komponenten
 - Ressourcenkapazitäten
 - Links
 - Installierte Bandbreite
 - Verzögerung
- Zur Instantiierung von Komponenten

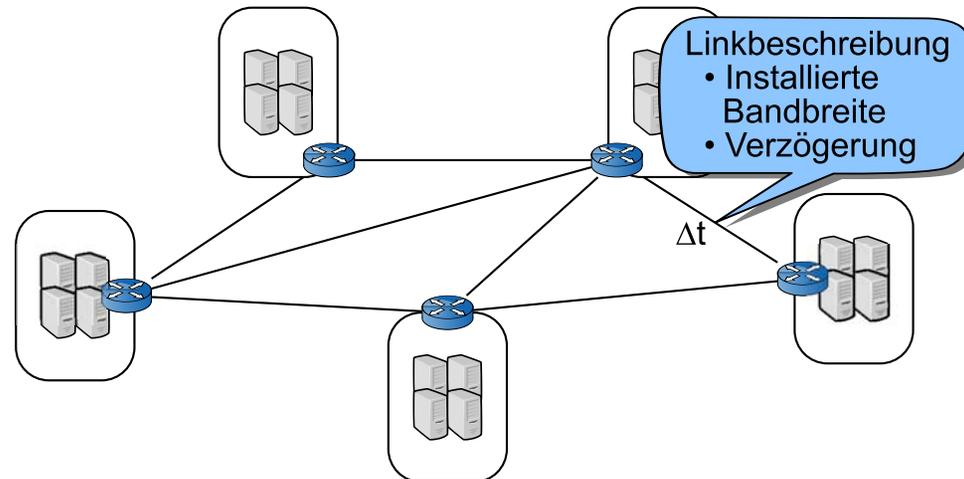


Modellierung

... der Ressourcen (Ressourcenbeschreibung)

Ressourcentopologie bestehend aus

- Knoten
 - Ausführungsumgebung für Komponenteninstanzen
 - Ressourcenkapazitäten
 - Links
 - Installierte Bandbreite
 - Verzögerung
- Zur Instantiierung von Komponenten



Rollen

Dienstanbieter

- Bereitstellung der Dienstkomponenten
 - Komponentenbeschreibungen
 - Implementierungen
- Beschreibung der Dienstopologie

Umsetzung

WS-ähnliches Komponentenmodell

Dienstbetreiber/„Infrastrukturanbieter“

- Auswahl der Komponenten
- Reservierung von Ressourcenkapazitäten
- Platzierung von Komponenten
 - Auf bereitgestellten Knoten und Links

Umsetzung

Virtualisierung (Netz, Rechenressourcen)



Rollen

Dienstanbieter

- Bereitstellung der Dienstkomponenten
 - Komponentenbeschreibungen
 - Implementierungen
- Beschreibung der Dienstopologie

Umsetzung

WS-ähnliches Komponentenmodell

Dienstbetreiber/„Infrastrukturanbieter“

- Auswahl der Komponenten
- Reservierung von Ressourcenkapazitäten
- Platzierung von Komponenten
 - Auf bereitgestellten Knoten und Links

Umsetzung

Virtualisierung (Netz, Rechenressourcen)



Platzierung der Komponenten auf Ressourcen

Überblick

Untersuchungsraum

- Platzierungsstrategie
- Dynamisches Verhalten
- Reorganisation

Optimierungskriterien

- Ressourcenauslastung
- Fairness
- Anzahl aktiver Knoten
- Betriebskosten

Mögliche Ziele

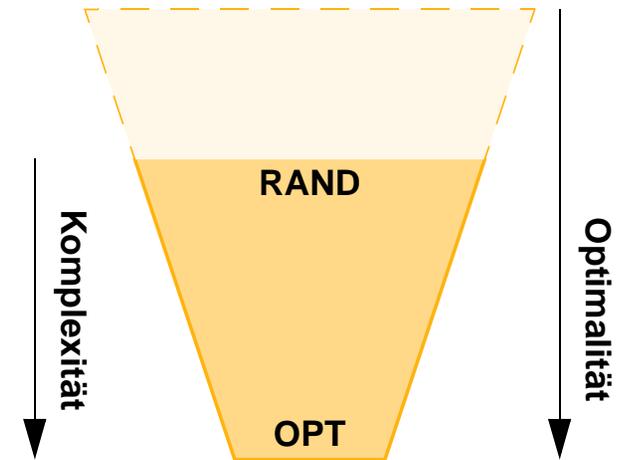
- Minimierung (bzw. Maximierung) der Optimierungskriterien
- Wirtschaftlichkeit (CAPEX und OPEX)

Platzierung der Komponenten auf Ressourcen

Platzierungsstrategie – Algorithmen

Algorithmische „Schranken“

- Zufällige Platzierung als intuitive **obere Schranke** (RAND)
 - „Schlecht, aber schnell“
- Optimale Platzierung als **untere Schranke** (OPT)
 - Gemischt-ganzzahliges lineares Programm (Mixed Integer Linear Program)
 - „Langsam, aber gut“



Metriken

- Wahrscheinlichkeit der Ablehnung
- (Link-) Ressourcenbedarf einer Platzierung
- Algorithmische Komplexität

Vergleich der Ablehnewahrscheinlichkeiten

Vorgehen

Methode

Monte-Carlo Simulation

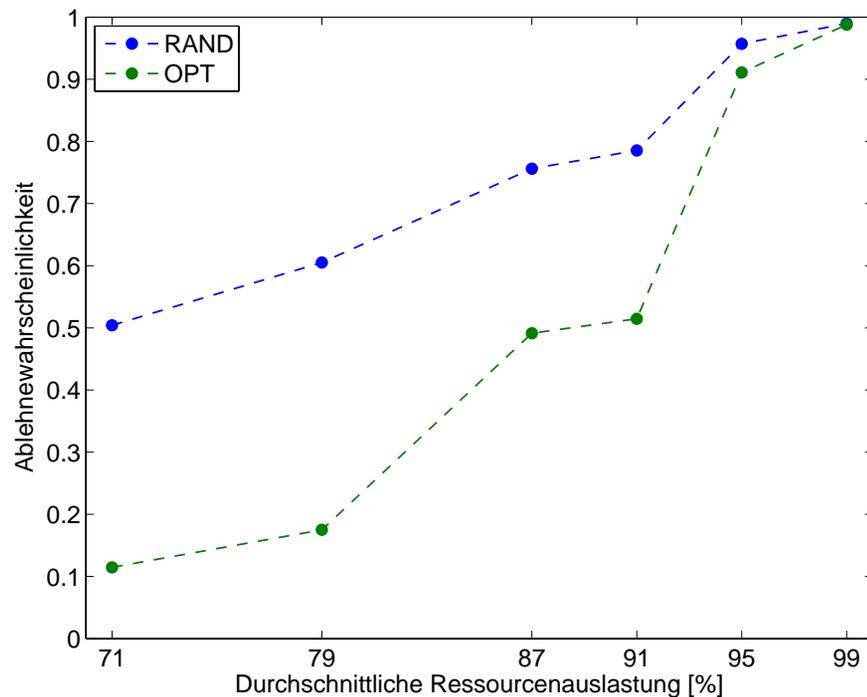
Generierung von Dienstopologie und Ressourcentopologie

- Zufällig
- Identische Knotenressourcen und installierte Linkbandbreiten
- Klassifiziert nach
 - Durchschnittliche Auslastung der Ressourcen
 - Anzahl Komponentenbeschreibungen in Dienstopologie

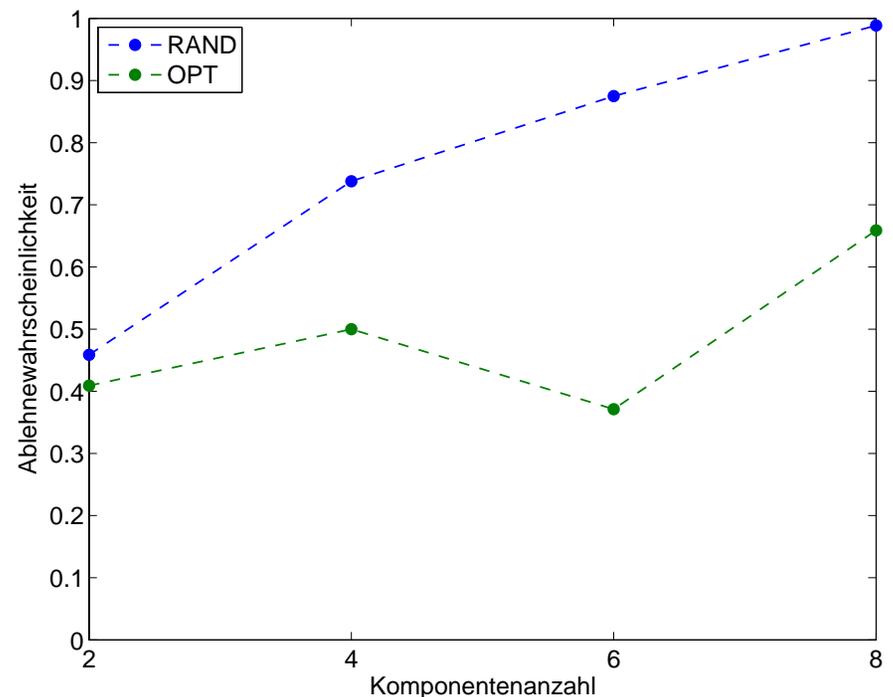
Vergleich der Ablehnewahrscheinlichkeiten

Vorläufige Ergebnisse

Auslastung



Dienstgröße



- RAND selten unter 50%
- Größter Gewinn durch OPT bei mittlerer Auslastung und den größeren Dienstopologien
→ Relevante Szenarien

Vergleich des Ressourcenbedarfs OPT : RAND

Vorgehen

Methode

Monte-Carlo Simulation

Generierung von Dienstopologie und Ressourcentopologie

- Zufällig
- Identische Knotenressourcen und installierte Linkbandbreiten
- Klassifiziert nach
 - Durchschnittliche Auslastung der Ressourcen
 - Anzahl Komponentenbeschreibungen in Dienstopologie (Scharparameter)

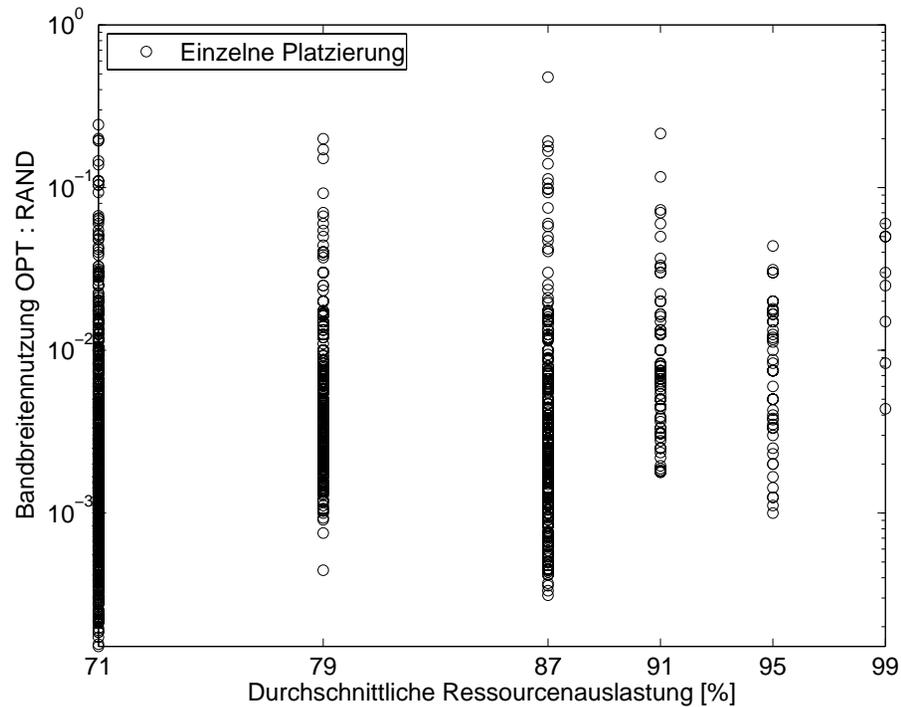
Metrik

- Nur Linkressourcen (Bandbreite BW)
(Keine Verbesserung der Nutzung der Knotenressourcen durch OPT in diesem Modell)
- Bandbreitennutzung = $\frac{\sum BW(OPT)}{\sum BW(RAND)}$
- Bandbreiteneinsparung = $1 - \text{Bandbreitennutzung}$

Vergleich des Bandbreitenbedarfs OPT : RAND

Vorläufige Ergebnisse

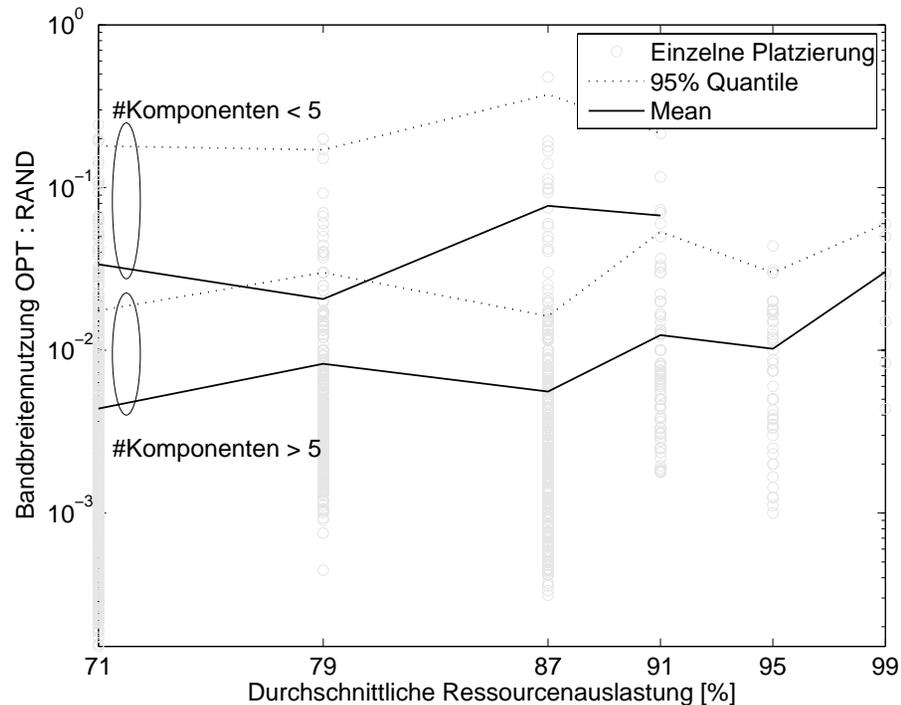
Bandbreitennutzung



Vergleich des Bandbreitenbedarfs OPT : RAND

Vorläufige Ergebnisse

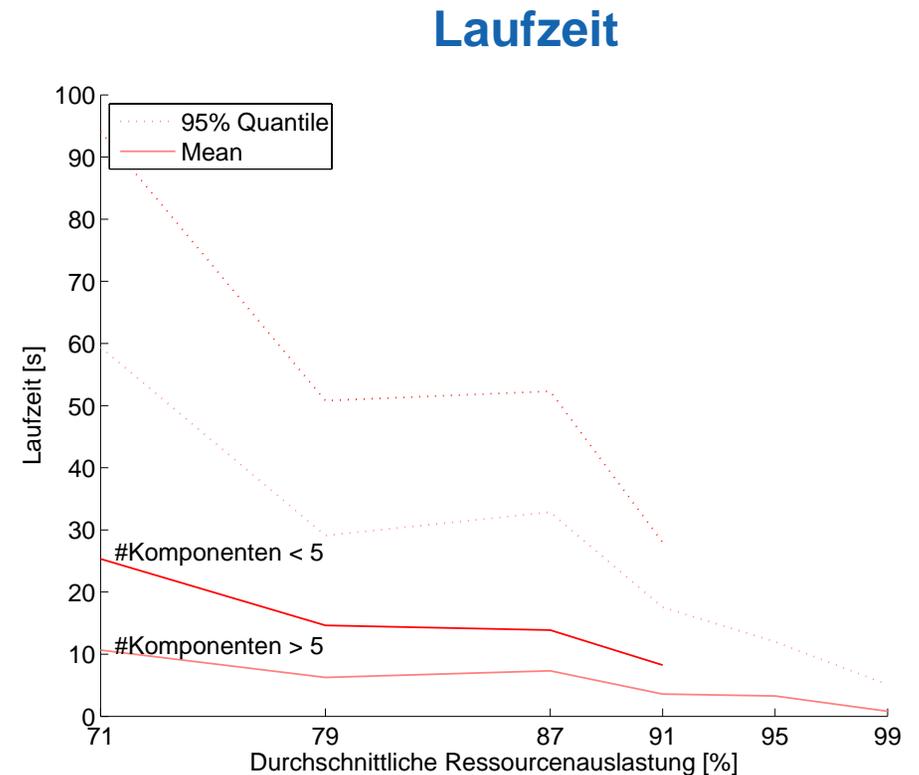
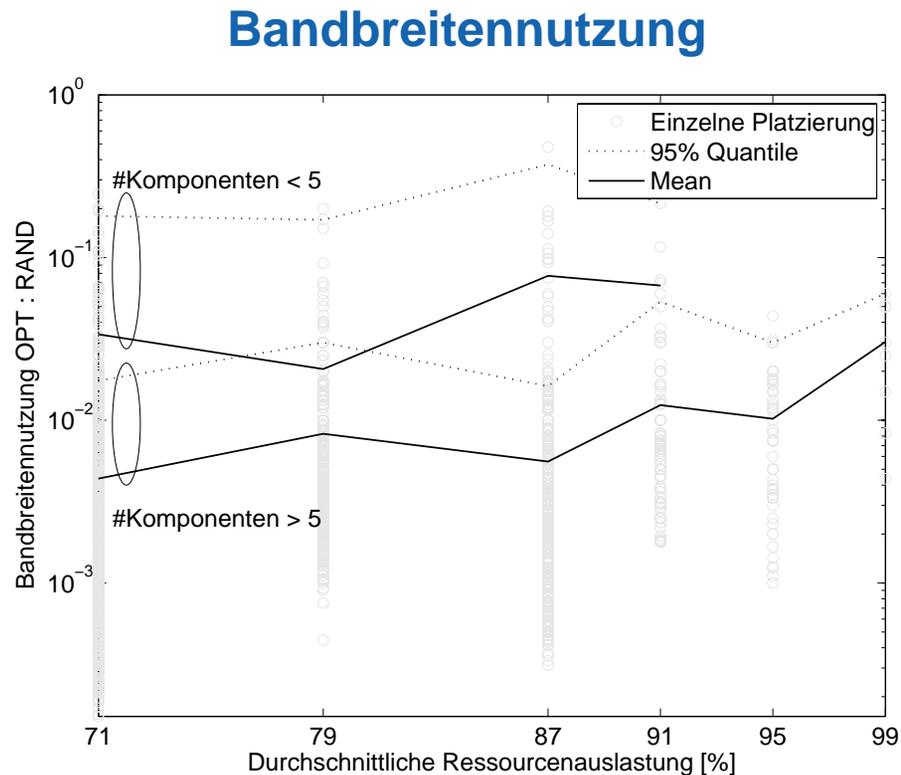
Bandbreitennutzung



- Über 80% Bandbreiteneinsparung
- Größte Einsparung bei mittlerer Auslastung und den größeren Dienstopologien

Vergleich des Bandbreitenbedarfs OPT : RAND

Vorläufige Ergebnisse



- Über 80% Bandbreiteneinsparung
 - Größte Einsparung bei mittlerer Auslastung und den größeren Dienstopologien
 - Aber: Höhere Rechenzeiten
- Heuristiken notwendig

Zusammenfassung

- Dienstorientierte Netzinfrastruktur besteht aus Netz und Rechenressourcen
 - Trennung von Dienstanbieter und Dienstbetreiber
 - Schlanke aber flexible Schnittstelle durch
 - Virtualisierung
 - WS-Komponentenmodell
- Methode zur wissenschaftlichen Untersuchung des Platzierungsproblems
 - Modellierung durch Dienst- und Ressourcenbeschreibungen
 - Monte-Carlo Simulation zur Bewertung neuer Platzierungsstrategien
 - Vergleich mit „unteren“ und „oberen“ algorithmischen Schranken
- Vorläufige Ergebnisse zeigen folgende Tendenzen auf
 - Optimierungspotenzial zwischen den beiden Schranken vorhanden
 - Hohe Einsparung von Linkressourcen von OPT bzgl. RAND (mehr als 80%)
 - Hohe Laufzeiten von OPT in relevanten Szenarien erwartet

Fazit

- Zunehmende Bedeutung der Fragestellung (vgl. Cloud Computing)
- Unzureichend erforscht
- Weitere Entwicklung der Untersuchungsmethodik