



HOCHSCHULE FÜR TECHNIK UND WIRTSCHAFT

BERLIN

**Wirtschaftswissenschaften II - Angewandte Informatik
Fachbereich 4**

Epidemische Informationsausbreitung in P2P-Systemen

Forschungsarbeit im Masterstudium betreut von Prof. Dr. Thomas Schwotzer

Verfasser Jonas M. Kress (Jonas-Kress.de)

Inhaltsverzeichnis

1	Einleitung	1
2	Kontext	2
2.1	Netztypen unterhalb von P2P-Systemen	2
2.1.1	Statische Netze	2
2.1.2	Mobile Ad-hoc-Netze (MANET)	3
2.1.3	Opportunistische Netze (OPPNETs) bzw. partitionierte MANETs	5
2.2	Epidemische Systeme	6
2.2.1	Einleitung	6
2.2.2	Grundlagen	7
2.2.3	Zustands-Modelle aus der Epidemiologie	9
2.2.4	Anruf Modell (Random Call)	11
2.2.5	Terminierungs-Strategien	17
3	Epidemische Algorithmen in P2P-Systemen	20
3.1	Einsatz in statischen Netzen	20
3.1.1	Datenbank-Replikation	20
3.1.2	File-Sharing (Gnutella)	20
3.2	Einsatz in MANET-Netzen	24
3.2.1	Routing Protokolle	24
3.2.2	Hardware-Simulation	26
3.3	Opportunistische Netze (OPPNETs) bzw. partitionierte MANETs	30
3.3.1	Hypergossiping	30
3.3.2	Epidemie Middleware	31
3.3.3	Modell zur Informationsausbreitung	32
4	Zusammenfassung und Ausblick	34
	Literaturverzeichnis	36

1 Einleitung

Die Verbreitung von mobilen Geräten, die modernste Technologien zur drahtlosen Kommunikation bereitstellen, nimmt stetig zu. Dies ermöglicht neue Formen spontaner Vernetzung und Kollaboration, bei der sich Benutzer aktiv an der Verbreitung von Informationen beteiligen oder deren technische Begleiter autonom die Basis für eine Datenverbreitung schaffen können, indem diese mit Pendants in ihrer Kommunikationsreichweite Verbindungen aufbauen.

Die sich daraus ergebenden mobile ad hoc network (MANET) Systeme oder opportunistischen Netzwerke (engl. opportunistic networks) haben das Potential gewöhnliche statische Netze zu erweitern und erlauben das zur Verfügung stellen neuer innovativer Dienste.

Die neuen Herausforderungen dieses Kommunikations-Paradigmas liegen vor allem in der enormen Flexibilität und Mobilität des Systems, sowie in der hoch komplexen Form der Informations-Dissemination.

Bei dem Aufbau von P2P-Systemen in diesen Infrastrukturen gilt es neuartige Algorithmen und Herangehensweisen zu entwickeln, die die Kommunikation trotz großer Hürden sicherstellen können. Ein Hoffnungsträger in diesem Gebiet sind epidemische Algorithmen, weil sich diese durch ihre Robustheit, gerade in unzuverlässigen Medien, auszeichnen.

Diese Forschungsarbeit untersucht die Möglichkeiten der epidemischen Informationsausbreitung in Peer-2-Peer Systemen unter Berücksichtigung der zugrunde liegenden Netzstruktur. Hierzu werden Erkenntnisse und Methoden aus der Epidemiologie aufgezeigt, diese auf moderne P2P-Systeme übertragen, um anschließend Forschungsergebnisse und Implementierungen aus diesem Bereich aufzuführen.

Es soll aufgezeigt werden welche Strukturen die vorgefundenen Datenströme und Informationsflüsse annehmen und wie die Technologie der epidemischen Informationsausbreitung sinnvoll genutzt werden kann.

2 Kontext

In diesem Abschnitt wird der Kontext definiert, in dem die Forschungsarbeit stattfindet. Im ersten Teil wird auf verschiedene Infrastrukturen eingegangen, in denen P2P-Systeme Verwendung finden, und dargestellt, welche unterschiedlichen Strukturen einen Einfluss auf die Informationsverteilung nehmen.

Anschließend wird aufgezeigt, was sich hinter epidemischen Systemen verbirgt und wie diese Technologie den Einsatz im Computersystem finden kann.

2.1 Netztypen unterhalb von P2P-Systemen

Da die dem P2P-System zugrunde liegende Struktur eine enorme Einwirkung auf die Form und den Fluss der Information hat, werden in den folgenden Absätzen unterschiedliche Netztypen betrachtet. Dies schafft ein Bewusstsein für die verschiedenen Aspekte des Einsatzes von epidemischen Algorithmen und bildet die Grundlage für die weiteren Kapitel.

2.1.1 Statische Netze

In statischen Netzen sind die Teilnehmer direkt miteinander verbunden, d.h. die Kommunikation von zwei Peers kann direkt erfolgen, ohne vorher eine Route über andere Peer-Nodes etabliert zu haben.

Zusätzlich ist der Aufenthaltsort der Knoten meist räumlich fixiert, dies hat zur Konsequenz, dass fast immer eine externe Stromversorgung vorhanden ist und kabelgebundene Netze mit hoher Übertragungsrate und -stabilität genutzt werden können.

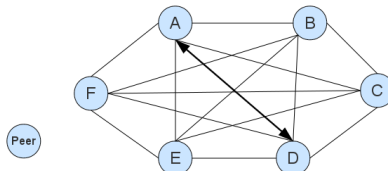


Abb. 2.1: Peer-Kommunikation im statischen Netz A kann direkt mit B kommunizieren

2.1.2 Mobile Ad-hoc-Netze (MANET)

Bei P2P-System in sogenannten mobilen Ad-hoc oder auch spontanen Netzen, findet eine Mobilisierung und drahtlose Kommunikation der Knoten statt. Dies hat zur Folge, dass einzelne Peers ausschließlich durch Routen erreicht werden können, die durch Dritte zur Verfügung gestellt werden, weil jeder Node auf die Kommunikation mit Peers in Reichweite, beschränkt ist.

Neben den gängigen Unicast, Multicast und Broadcast Methoden bieten sich in MANETs Geocasts an, die Knoten in einem bestimmten geografischen Bereich zu adressieren.

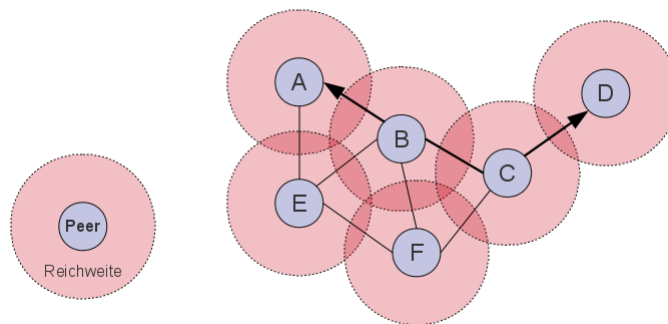


Abb. 2.2: Peer-Kommunikation im mobilen Ad-hoc-Netz
A kommuniziert mit D über die Knoten B und C

Zum Routing in mobilen Ad-hoc-Netzen kommen folgende Routingverfahren zum Einsatz:

- Positionsbasiert

Wertet geodätische Informationen, wie z.B. Daten des GPS -Empfängers der Knoten aus, um Routen aufzubauen.

- Topologiebasiert

Analysiert die Topologie des Netzwerks, indem es Hops und Nachbarschaftsbeziehungen der Knoten aufzeichnet.

- Energieverbrauch berücksichtigende Verfahren

Misst die aufgewendete Energie der Verbindung und berücksichtigt diese Erkenntnisse beim Routen, um ein Netz mit geringem Energiebedarf zu etablieren.

Die aufgezählten Routingverfahren können zusätzlich in folgende Kategorien eingestuft werden:

- Proaktive Verfahren

Die entsprechenden Routen werden bereits vor dem aktiven Datenaustausch erstellt. Hierzu tauschen die einzelnen Knoten Traffic Control Nachrichten aus, um die jeweils erkundeten Routen zu publizieren.

- Reaktive Verfahren

Eine Route wird stets unmittelbar vor der eigentlichen Übertragung der Information ermittelt. Fluss-Orientiertes Verfahren Durch Auswertung der von anderen Knoten verschickten Meta-Daten wird eine Routingtabelle erstellt.

- Adaptive Verfahren

Adaptiert sich an den Status des Netzwerkes. Meist wird beim Eintreten ins Netz erst proaktiv agiert und im Laufe der Kommunikation auf ein reaktives Verfahren gewechselt.

- Hybrid Verfahren

Nutzt sowohl aktive, reaktive und fluss-orientierte Routing-Mechanismen.

MANETs gibt es in weiteren Ausprägungen:

- Internet Based Mobile Ad-hoc Networks (iMANET)

Netze in denen neben Verbindungen zu mobilen Knoten, auch eine Verbindung zu einem Internet-Gateway besteht.

- Vehicular Ad Hoc Networks (VANETs)

System zur Kommunikation von Fahrzeugen untereinander, um die Sicherheit und den Komfort von Passagieren zu erhöhen.

- Intelligent vehicular ad hoc networks (InVANETs)

Baut auf VANETS auf und stellt Dienste zur Verfügung, die es Verkehrsmitteln erlauben sollen, auf Kollisionen und Unfälle autonom zu reagieren.

Da gewöhnliche Laptops für die Verwendung in mobilen Ad-hoc-Netzen meist nicht zu gebrauchen sind, weil ihnen die entsprechenden Chipsätze und zusätzlichen Antennen fehlen, entwickelte die One Laptop per Child (OLPC) Initiative [OLP] ein Gerät, das dies ermöglicht, um den Einsatz in dritte Welt Ländern zu fördern und eine Vernetzung der Geräte untereinander zu ermöglichen.

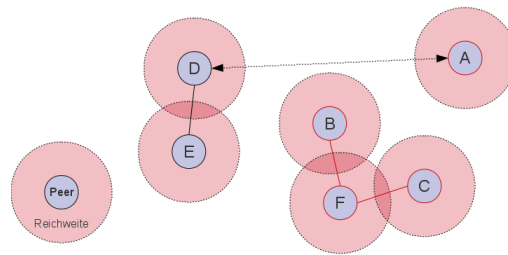


Abb. 2.4: Peer-Kommunikation in OPPNETs bzw. partitionierten MANETs
Information wandert von A nach C und nach dem Partitionswechsel von C weiter zu F und B

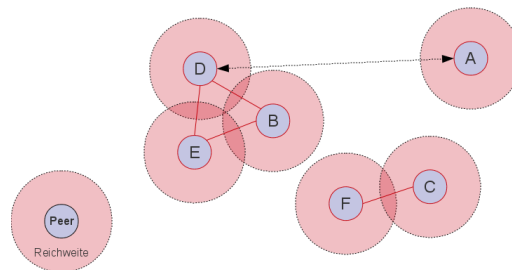


Abb. 2.5: Peer-Kommunikation in OPPNETs bzw. partitionierten MANETs
B wechselt die Partition und die Information erreicht schließlich D

2.2 Epidemische Systeme

2.2.1 Einleitung

Die Natur als Quelle der Inspiration gilt als ein probates Mittel, auch in Zeiten der Hochtechnologie. So gesehen ist es wenig verwunderlich, dass ein über mehrere Milliarden von Jahren „entwickeltes“ System zur Informationsausbreitung der Natur hohen Anklang in der Forschung unseres Informationszeitalters findet.

Auch wenn die negative Konnotation des Begriffes Epidemie, die aus epidemiologischer Sicht eine Infektionskrankheit beschreibt, deren Anzahl an Erkrankungsfällen in einem bestimmten Zeitraum zunimmt, die Brillanz dieser Technologie zu mindern scheint, gibt es wohl wenige Mechanismen der Informationsausbreitung, die sich in ihrer Effizienz mit dieser herausragenden Entwicklung der Natur messen können.

Deswegen ist es wenig verwunderlich, dass dieses Prinzip in P2P-Systemen adaptiert wird, um Information zu verbreiten.

2.2.2 Grundlagen

Das Konzept einer Epidemie beschreibt eine Basisreproduktionsrate, sprich die Anzahl der Sekundärinfektionen eines Infizierten, die im statistischen Mittel größer als eins ist. Dies bedeutet, dass nach jeder Infektion der Betroffene mehr als einen Kontakt infiziert.

So erreicht die Verbreitung in der Startphase eine hohe Reproduktionsrate bis schließlich eine Sättigung des Systems stattfindet, weil sich die Anzahl der empfänglichen Teilnehmer stark verringert hat.

Tab. 2.1: Definition einer Epidemie

	Epidemiologie	Informationsverbreitung
$S(t)$	Anzahl der empfänglichen (susceptible) Individuen zu einer bestimmten Zeit t	Anzahl der Peer-Knoten zum Zeitpunkt t , die an einer Information interessiert sind
$I(t)$	Anzahl der infizierten Individuen zum Zeitpunkt t	Anzahl der Peer-Knoten zum Zeitpunkt t , denen die Information bekannt ist
$s(t) = S(t) / n$	Relative Anzahl der empfänglichen (susceptible) Individuen zu einer bestimmten Zeit t	Relative Anzahl der Peer-Knoten zum Zeitpunkt t , die an einer Information interessiert sind
$i(t) = I(t) / n$	Relative Anzahl der infizierten Individuen zum Zeitpunkt t	Relative Anzahl der Peer-Knoten zum Zeitpunkt t , denen die Information bekannt ist
χ	Kontakte eines Individuums pro Zeiteinheit	Kontakte eines Peers pro Zeiteinheit
β	Wahrscheinlichkeit einer Infektion zwischen einem infiziertem und einem empfänglichem Individuum	Wahrscheinlichkeit der Informationsübertragung zwischen einem informierten und einem uninformierten Peer-Knoten
$a = \beta \chi / n$	Infektionsrate	Informationsausbreitungsrate
$\alpha = a * I(t)$	Mächtigkeit der Infektion: Die Wahrscheinlichkeit eines Individuums pro Zeiteinheit infiziert zu werden	Mächtigkeit der Informationsausbreitung: Die Wahrscheinlichkeit eines Peers pro Zeiteinheit informiert zu werden

Initialisierung:

$$I(0) = 1, S(0) = n - 1, i(0) = 1/n, s(0) = 1 - 1/n$$

Terminierung:

$$I(t\text{-max}) = n, S(t\text{-max}) = 0, i(t\text{-max}) = 1, s(t\text{-max}) = 0$$

2.2.3 Zustands-Modelle aus der Epidemiologie

Ein klassischer Ansatz zur Beschreibung von Epidemien in der mathematischen Epidemiologie sind Zustands-Modelle, die es ermöglichen die Ausbreitungsarten von ansteckenden Krankheiten zu beschreiben. In diesem Abschnitt werden verbreitete Zustands-Modelle illustriert, dies ermöglicht anschaulich den Verlauf von Infektionen darzustellen und die verschiedenen Phasen einer Epidemie zu beschreiben.

Zur Modellierung eines Krankheitsverlaufes ist im Übrigen auch eine Open Source Software vorhanden, die auf dem Eclipse Modeling Framework aufsetzt. [Fou]

Quellen und weitere Informationen unter:

[Ste09] [BAP10]

2.2.3.1 SIR-Modell

Beim SIR-Modell wird von drei Zuständen ausgegangen: empfänglich (susceptible), infektiös (infectious) und genesen (recovered). Dieses robuste und einfache Modell findet z.B. Anwendung bei der Beschreibung der Ausbreitung von Krankheit mit Immunitätsbildung wie Mumps, Masern, Keuchhusten, Röteln, Windpocken und Poliomyelitis.

Neben den bekannten Funktionen $S(t)$ und $I(t)$ wird eine neue Funktion $R(t)$ eingeführt, die die Zahl der genesenen Individuen zum Zeitpunkt t beschreibt. Zu der vorgestellten Ansteckungsrate wird ein neuer Wert belegt, der die Genesungsrate beschreibt und den Übergang von I zu R darstellt.

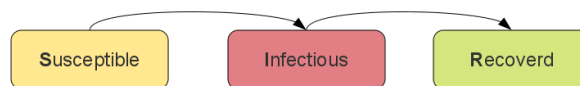


Abb. 2.6: Epidemiologie SIR-Zustands-Modell

2.2.3.2 SEIR-Modell

Das SEIR-Modell hat neben den Zuständen des SIR-Modells den Status ausgesetzt (exposed), in dem das Individuum zwar infiziert, jedoch nicht infektiös ist, d.h. er kann in diesem Zustand die Infektion nicht weitergeben.

Beispiele für diese Art von Krankheiten sind Hepatitis B, Chagas-Krankheit, HIV/Aids und Tuberkulose. Bei den letzten Beiden kann dieser Zustand über Jahre andauern.

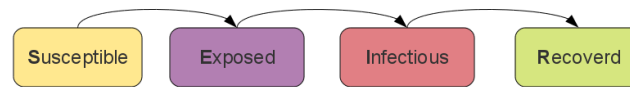


Abb. 2.7: Epidemiologie SEIR-Zustands-Modell

2.2.3.3 MSIR-Modell

In diesem Modell wird ein Zustand (maternally derived immunity) beschrieben, der dem Individuum Immunität für einen begrenzten Zeitraum gewährt, weil durch die Aufnahme von Antikörpern über die Plazenta oder das Kolostrum der Schutz der Mutter auf das Kind übertragen wurde bzw. wird. Diese temporäre Immunität tritt zum Beispiel bei den Windpocken auf.

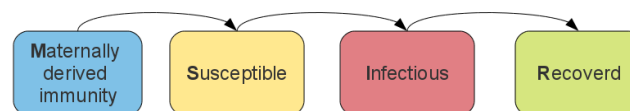


Abb. 2.8: Epidemiologie MSIR-Zustands-Modell

2.2.3.4 Träger-Modell

Bei manchen Individuen kann es auftreten, dass nach einer Infektion die Genesung ausbleibt und der Betroffene in einem Träger-Zustand (carrier) verharrt, bei dem er nicht an der Krankheit leidet aber eventuell Symptome zurückkehren oder sogar ein Übergang in den infektiösen Zustand eintritt. Beispiele für dieses Modell sind Tuberkulose und Typhus.

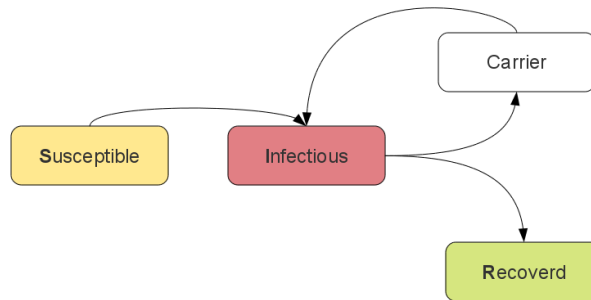


Abb. 2.9: Epidemiologie Träger-Zustands-Modell

2.2.3.5 SIS-Modell

Es kann vorkommen, dass bei manchen Krankheiten wie z.B. Erkältung, die Immunität ausbleibt und das Individuum zwischen dem Zustand der Anfälligkeit und der Infektiösität wechselt. In diesen Fällen wird auf das SIS-Modell zur Beschreibung zurückgegriffen.

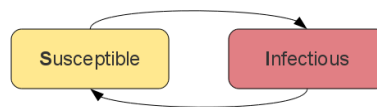


Abb. 2.10: Epidemiologie SIS-Zustands-Modell

2.2.4 Anruf Modell (Random Call)

Die nachfolgenden Betrachtung der Modelle findet auf der Basis von Runden (synchron) statt, um die Ausbreitung der Informationen strukturiert veranschaulichen zu können und weil die mathematische Berechnung dies zur Basis nimmt.

Da die Betrachtungen von einem synchronen Netzmodell ausgehen, können die gewonnenen Erkenntnisse nur bedingt auf die Praxis übertragen werden, da dort zumeist asynchrone Netzwerke vorherrschen.

Beim Random Call findet die Kommunikation zwischen zufällig gewählten Knoten statt. In jeder Runde kontaktiert jeder Peer einen Knoten aus der Liste aller Knoten. Dies impliziert auch die Möglichkeit, dass eine Verbindung zwischen dem gleichen Anfangs- und Endpunkt (Loop) aufgebaut wird.

Daraus ergeben sich beim Random Call für einen Knoten folgende Möglichkeiten:

- Er ist der Einzige, der einen Knoten kontaktiert.
- Er ist Teil einer Gruppe von Knoten, die einen Knoten kontaktieren.
- Er wird von keinem Knoten kontaktiert.
- Er wird von genau einem Knoten kontaktiert
- Er wird von einer Gruppe von Knoten kontaktiert.

Der anschließende Informationsaustausch erfolgt in folgenden Formen:

- Push-Modell Der Anrufer sendet die Informationen an den Angerufenen.

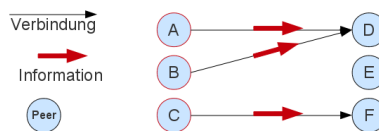


Abb. 2.11: Random Call - Push-Modell

- Pull-Modell Der Anrufer empfängt die Informationen vom Angerufenen.

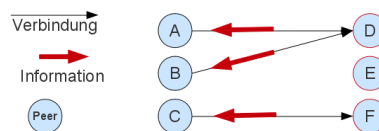


Abb. 2.12: Random Call - Pull-Modell

- Push & Pull-Modell Anrufer sendet oder empfängt Informationen vom Angerufenen.

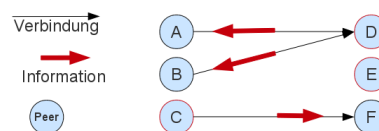


Abb. 2.13: Random Call - Push & Pull-Modell

In der ersten Runde wird mit genau einem infizierten Knoten gestartet, der die Quelle der Epidemie ausmacht.

Die Ausbreitung der Infektion findet pro Runde jeweils ausschließlich über eine Ebene statt. Dies bedeutet, dass Peer-C nicht von Peer-B infiziert wird, wenn Peer-B in der selben Runde von Peer-A angesteckt wurde.

[CS04] [BV00]

2.2.4.1 Push-Modell

Beim Push-Modell fließen die Informationen, die ausgetauscht werden sollen, in Richtung der Kontaktaufnahme. Folglich kann nach Beendigung der zweiten Runde die Anzahl der informierten Knoten maximal zwei betragen.

Es können folgende Fälle eintreten:

- Ein uninformierter Knoten informiert uninformierten Knoten → keine Infektion
- Ein uninformierter Knoten informiert einen informierten Knoten → keine Infektion
- Ein informierter Knoten informiert einen uninformierten Knoten → Infektion
- Ein informierter Knoten informiert einen informierten Knoten → keine Infektion

Nach [Pit87] benötigt eine Epidemie im Push-Modell $\log_2 n + \ln n + O(1)$ Runden zur vollständigen Infektion des Systems.

[TDH04]

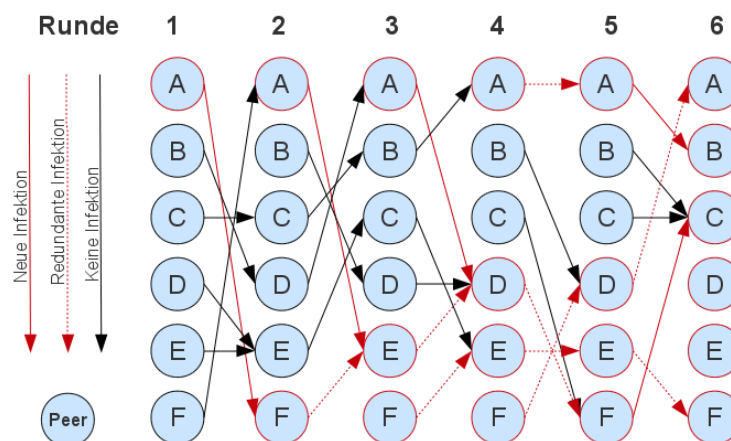


Abb. 2.14: Informationsausbreitung im Push-Modell

Simulator: <http://www.jonas-kress.de/epidemic/>

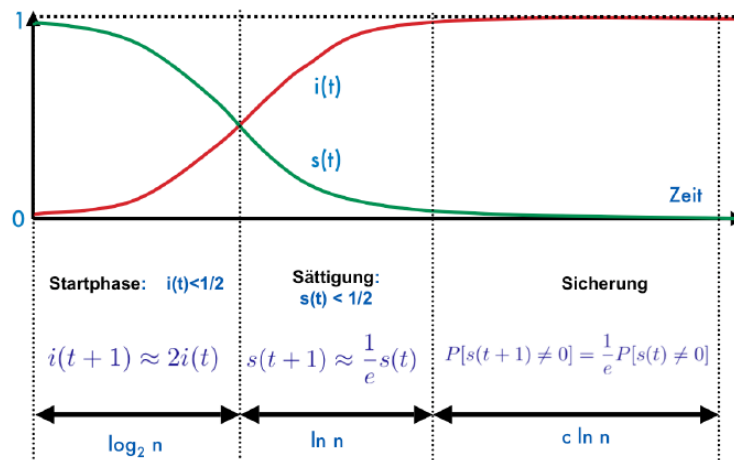


Abb. 2.15: Wachstumsphasen im Push-Modell Quelle: [CS04]

2.2.4.2 Pull-Modell

Äquivalent zum Push-Modell werden in jeder Runde Informationen zwischen zwei Knoten ausgetauscht, jedoch findet die Übertragung genau in entgegengesetzter Richtung der Kontaktaufnahme statt.

Im ersten Moment könnte man annehmen, dass sich beide Verfahren in ihrer Effektivität gleichen, dies ist allerdings nicht der Fall, weil beim Pull-Modell, im Gegensatz zur Push-Variante, theoretisch nach Abschluss der zweiten Runde alle Knoten informiert sein könnten, wenn diese allesamt die Quelle der Epidemie kontaktiert haben.

Es können folgende Fälle eintreten:

- Ein uninformierter Knoten fragt einen uninformierten Knoten → keine Infektion
- Ein uninformierter Knoten fragt einen informierten Knoten → Infektion
- Ein informierter Knoten fragt einen uninformierten Knoten → keine Infektion
- Ein informierter Knoten fragt einen informierten Knoten → keine Infektion

Es wird davon ausgegangen, dass sich ein Gerücht innerhalb von $\log n + O(\log \log n)$ Runden im System ausgebreitet hat.

[TDH04]

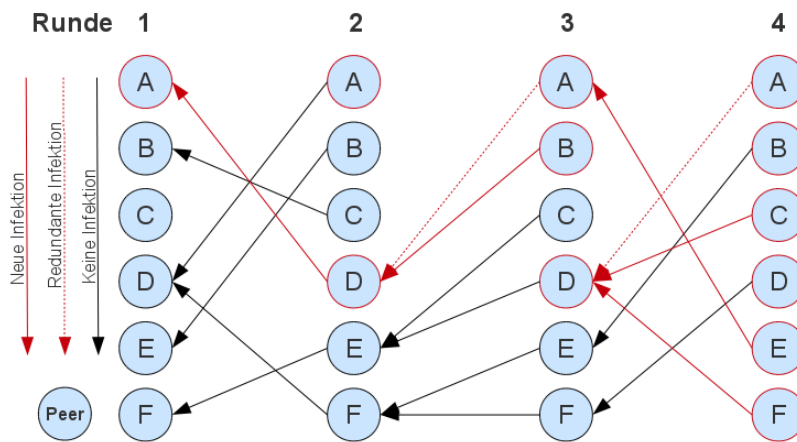


Abb. 2.16: Informationsausbreitung im Pull-Modell
 Simulator: <http://www.jonas-kress.de/epidemic/>

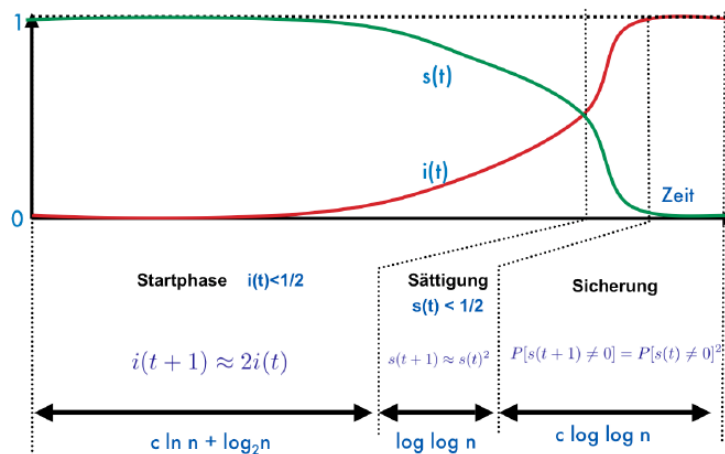


Abb. 2.17: Wachstumsphasen im Pull-Modell Quelle: [CS04]

2.2.4.3 Push & Pull -Modell

In diesem Modell findet eine Kombination der bisher vorgestellten Mechanismen statt. Die Wahl der Kommunikationspartner bleibt zufällig, die Ausbreitung der Informationen findet sowohl in, als auch entgegen der Richtung der Kontaktaufnahme statt.

Laut [BV00] werden beim Push & Pull-Modell alle Knoten innerhalb von $\log_3 n + O(\ln \ln n)$ Runden informiert und dabei $O(n \ln \ln n)$ Nachrichten versendet.

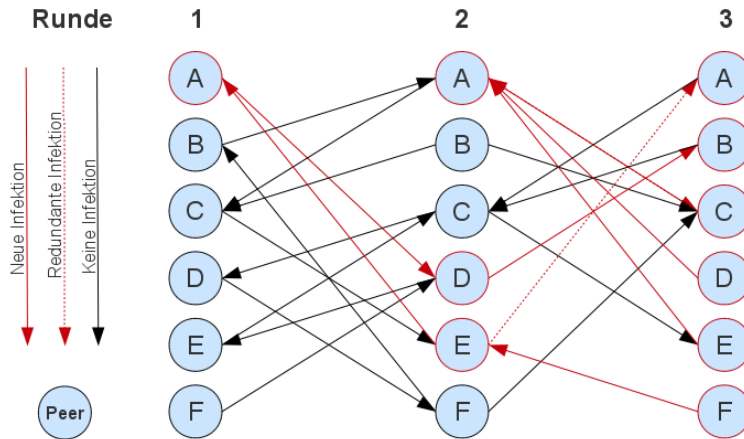


Abb. 2.18: Informationsausbreitung im Push & Pull-Modell
 Simulator: <http://www.jonas-kress.de/epidemic/>

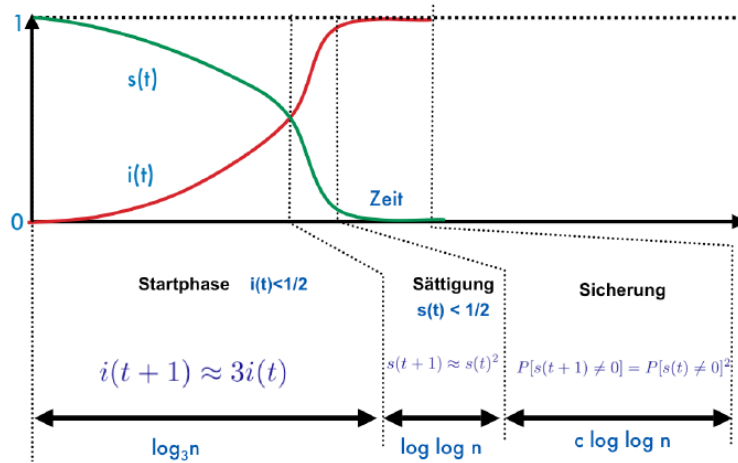


Abb. 2.19: Wachstumsphasen im Push & Pull-Modell Quelle: [CS04]

2.2.5 Terminierungs-Strategien

Es ist offensichtlich, dass bei der epidemischen Ausbreitung von Informationen innerhalb von Peer-2-Peer Systemen Technologien eingesetzt werden müssen, die verhindern, dass das Netz durch die exorbitante Flut an Nachrichten zum Erliegen kommt.

Da jeder Knoten zur Vervielfältigung der ursprünglichen Botschaft beiträgt, muss eine Methodik implementiert werden, die diese Replikation stoppt, wenn davon ausgegangen werden kann, dass genug Kopien im System vorhanden sind, sogenannte Terminierungs-Strategien.

Hierbei gelten zwei Herangehensweisen zu unterscheiden

- Blind

Die Knoten erfahren den Infektionszustand ihrer Nachbarn nicht, das bedeutet die Entscheidung ob ein Paket weitergegeben wird, erschließt sich ausschließlich durch die Auswertung der internen Zustände eines Peers.

- Feedback

Nachbarn kommunizieren ob bei ihnen eine Infektion vorliegt. Aus diesen gewonnenen Informationen kann entschieden werden, ob eine ausreichende Sättigung des System vorliegt, die den weiteren Versand von Nachrichten nicht rechtfertigt.

2.2.5.1 Probabilistisch

Es wird mit einer gewissen Wahrscheinlichkeit entschieden, ob die Infektion weiter gegeben werden soll oder nicht. Diese Strategie sorgt zwar dafür, dass ein Gerücht nach einer gewissen Zeit ausstirbt, es kann aber nicht exakt bestimmt werden, wann dies der Fall sein wird, geschweige denn sichergestellt werden, dass alle Knoten infiziert worden sind.

Durch einen Feedback-Mechanismus, der die Chance für das Aussterben bzw. Fortleben nach den Erkenntnissen der Nachbarn beeinflusst, kann zwar die Leistungen dieser Methode verbessern, um sicherzustellen, dass ein Gerücht das gesamte System durchströmt, ist diese Herangehensweisen offensichtlich ungeeignet.

2.2.5.2 Zeit

Beim Erstellen eines neuen Gerüchts wird dieses einmalig mit einem Zeitstempel versehen. Beim Eintreffen neuer Nachrichten an einem Knoten wird die Zeitdifferenz der lokalen Zeit mit dem Zeitstempel berechnet.

Liegt die Differenz oberhalb eines bestimmten Schwellwertes, wird das Gerücht nicht mehr weiter verbreitet. Es muss sichergestellt sein, dass die Zeiten der verschiedenen Peers sich gleichen, um eine exakte Aussage über die Zeitdifferenz treffen zu können.

2.2.5.3 Time To Live (TTL)

Jede Nachricht bekommt einen Initiierungswert, der bei jeder Weitergabe dekrementiert wird. Ist der Wert null, wird die Nachricht nicht weitergegeben. Der Sender kann somit steuern, wie viele Kopien in einem System erstellt werden sollen. In einem Push-Modell wäre eine Initialisierung des TTL mit $O(\log n)$ denkbar.

2.2.5.4 Time Lived (TL)

Jede Infektion hat einen Wert, der dessen Generation, sprich wie oft eine Weitergabe statt gefunden hat, angibt. Bei jeder Kopie wird dieser Wert inkrementiert. Ab einer bestimmten Schwelle wird die Ausbreitung unterbunden.

Jeder Teilnehmer kann eventuell selbst festlegen, wann dies der Fall ist. Wird der Schwellwert zu niedrig gesetzt, kann die Nachricht eventuell nicht alle Knoten im Netzwerk erreichen, ist er zu hoch, so werden unnötig viele Botschaften versandt.

Wird eine globale Schwelle \max gesetzt, so werden im System $n \cdot \max$ Nachrichten versandt. Bei der Wahl des Schwellwertes muss das Ausbreitungs-Modell berücksichtigt werden.

2.2.5.5 Sheners Min-Counter-Algorithmus

Ist eine Adaptierung des Time Lived Prinzips (TL), bei der der Zähler jeder Nachricht nur dann inkrementiert wird, falls in der vergangenen Austauschrunde alle kontaktierten Teilnehmer mindestens die selben Zählerstände besaßen.

Der Vorteil hierbei ist, dass der Algorithmus auch dann funktioniert, wenn die Partnerwahl der kontaktierten Teilnehmer nicht mit einer gleichmäßigen Wahrscheinlichkeitsverteilung durchgeführt wird.

Analysen [BV00] haben gezeigt, dass das gesamte System innerhalb von $\log^3 n + O(\log \log n)$ Runden mit einer Wahrscheinlichkeit von $1 - n^{-O(1)}$ im Push & Pull Modus mit nur $O(n \log \log n)$ versendeten Nachrichten informiert werden kann.

2.2.5.6 Median-Counter

Algorithmus Dieser Algorithmus entstand bei die Modifikation des Min-Counter Algorithmus von Shenker durch Berthold Vöcking. Beim Vergleich des Zählerstände wird nicht das Minimum berücksichtigt, sondern der Durchschnitt (Median) gebildet.

Zusätzlich verändern Knoten im Laufe einer Epidemie ihr Kommunikations-Modell, so wird Anfangs das Gerücht mittels Push & Pull verbreitet, im weiteren Verlauf auf die Pull-Kommunikation umgestellt, bis schließlich die Verbreitung eingestellt wird.

Zur vollständigen Verteilung eines Gerüchts wird von $O(n \log \log n)$ Nachrichten in $O(\log n)$ Runden ausgegangen. [BV00] [Kar00]

3 Epidemische Algorithmen in P2P-Systemen

Nachdem die letzten Kapitel vorwiegend theorielastig waren, wird nun auf den praktischen Einsatz von epidemischen Algorithmen in P2P-Systemen eingegangen.

Die Betrachtung findet strukturiert auf Basis der Netztypen statt, da diese großen Einfluss auf die Form und Struktur der Datenströme nehmen.

3.1 Einsatz in statischen Netzen

3.1.1 Datenbank-Replikation

Den ersten Einsatz in Computer-Systemen fanden epidemischen Algorithmen bei der Datenbank-Replikation. In dem 1987 veröffentlichten Paper Epidemic Algorithms in Replicated Databases [Dem87] wurde nach Möglichkeiten geforscht, ein robustes und dezentrales System zu etablieren, das es ermöglicht den Datenbestand in verteilten Datenbanken zu replizieren.

Ziel war es, in lokalen Datenbanken erstellte oder geänderte Einträge einem Netz von Datenbanken bekannt zu machen, um im Idealfall eine exakte Kopie einer Datenbasis auf allen Knoten vorzufinden. Dies sollte auch durch den Ausfall von Netz- und Hardware-Ressourcen nicht negativ beeinflusst werden.

3.1.2 File-Sharing (Gnutella)

Im folgenden Abschnitt wird der Einsatz von epidemischen Algorithmen in P2P File-Sharing Netzen anhand von Gnutella erläutert.

Es wird auf Protokoll-Spezifikationen eingegangen und verschiedene Implementierungs-Ausprägungen vorgestellt.

3.1.2.1 Gnutella-Protokoll (Specification 0.4)

Beim P2P File-Sharing System Gnutella wird die einfachste Art eines epidemischen Algorithmus verwendet, das Flooding. Um eine Überflutung des Netzes zu verhindern, wird als Terminierungs-Strategie ein TTL-Feld im Descriptor-Header verwendet.

Folgende Deskriptoren-Pakete durchfluten das P2P-Netz:

- Ping

Wird benutzt, um aktive Hosts im Netz zu entdecken. Von einem Knoten, der ein Ping-Paket erhält wird erwartet, mit einem oder mehreren Pong-Deskriptoren zu antworten.

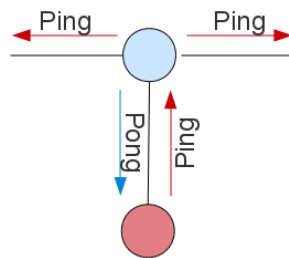


Abb. 3.1: Schematik der Ausbreitung von Ping/Pong Deskriptoren im Gnutella-Protokoll

- Query

Dient zur Suche im verteilten Netz und löst bei Knoten, die beim Vergleich mit ihrem lokalen Datenbestand einen Treffer bemerken, den Versand eines Query-Hit Pakets aus.

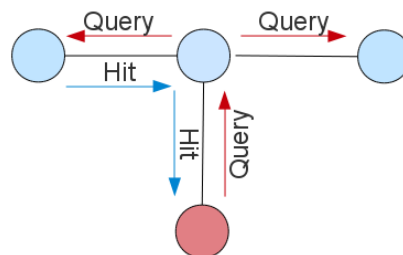


Abb. 3.2: Schematik der Ausbreitung von Query/Hit Deskriptoren im Gnutella-Protokoll

Die jeweiligen Antworten wandern auf den vorher erkundeten Routen oder werden laut Spezifikationen verworfen, um Spoofing und Flooding zu vermeiden.

Da die Strategie beim Durchfluten des Netzes sehr primitiv ist, kann es vor allem bei seltenen Dateien vorkommen, dass Query-Anfragen nicht die Knoten erreichen, die im Besitz der gewünschten Informationen sind. Dies ist auch der Grund, warum das P2P-Filesharing mit Gnutella Protokoll Specification 0.4 nicht skaliert.

Quellen und weitere Informationen unter:

[AD03][Gnua]

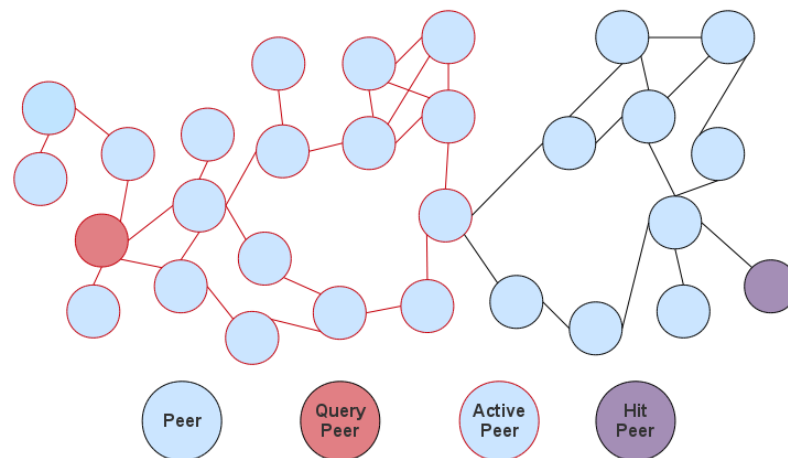


Abb. 3.3: Ausbreitung der Query Deskriptoren im Gnutella-Protokoll mit einer TTL von 5

3.1.2.2 Gnutella2-Protokoll (Specification 0.6)

Die wesentliche Erweiterung, die in dieser Spezifikation erfolgte, ist die Einführung eines Hybrid P2P-Systems, bei dem Knoten die Rolle sogenannter Ultrapeers einnehmen. Diese spezielle Art von Knoten übernimmt zum Teil Aufgaben, die normalerweise auf mehrere normale Knoten (leaf) verteilt sind.

Dies hat den Effekt, dass das Netz skaliert, weil die Anzahl der Knoten, die in das Bearbeiten von Nachrichten und das Bereitstellen von Routen involviert sind, stark reduziert wird. So leitet ein Ultrapeer Such-Anfragen an andere Ultrapeers weiter aber kontaktiert nur die Leaf-Knoten, von denen es annimmt, dass diese im Besitz von relevanten Informationen sind.

Durch diese Technologie, das sogenannte Query Routing Protokoll (QRP), werden Suchanfragen vorwiegend an Ultrapeers, die sich untereinander vernetzen, weitergeleitet

und verarbeitet. Dadurch ist es auch in sehr großen Netzen möglich seltene Informationen aufzuspüren.

Kriterien die ein Knoten erfüllen muss, um als Ultrapeer ausgewählt zu werden:

- Offene Firewall
- Hohe Anzahl möglicher Sockets
- Große Bandbreite
- Hohe Verfügbarkeit
- Ausreichend RAM und CPU-Geschwindigkeit

[Gnub][Ber03a]

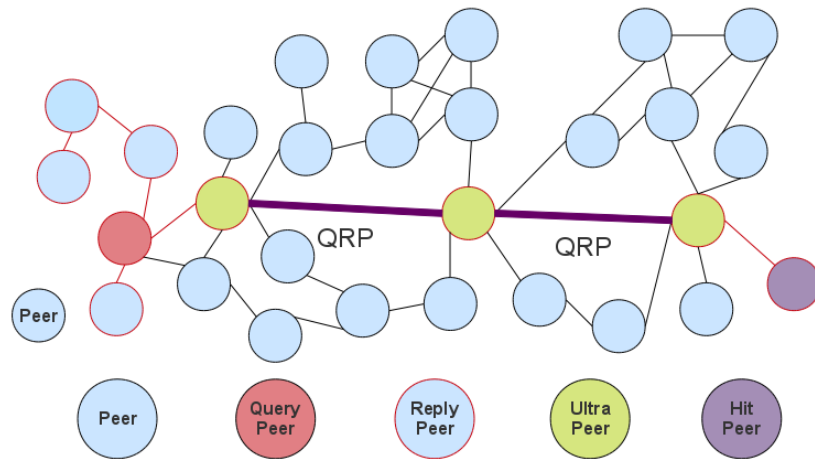


Abb. 3.4: Ausbreitung der Query Deskriptoren im Gnutella-Protokoll mit einer TTL von 5 und QRP

3.1.2.3 Epitella

Epitella ist eine Protokollerweiterung, die bei der Suche verbesserte Ergebnisse liefert, indem es, anstatt das Netzwerk blind zu fluten, die zugrunde liegende Struktur (Anzahl der Knoten und deren Konnektivität) berücksichtigt wird, um sowohl populäre, als auch wenig verbreitete Dateien zu finden. So wird, die durch eine Suchanfrage ausgelöste Epidemie speziell an die Erkenntnisse aus menschlichen Epidemiologie Forschung [AM92] angepasst und das SIR-Modell, bei dem immunisierte Knoten nicht mehr anfällig für Infektionen sind, als Grundlage verwendet.

Die eigentliche Suche verläuft in zwei Schritten. In der ersten Phase wird eine Prüf-Suchanfrage ausgelöst, um die Verbreitung der angefragten Datei zu bestimmen. Anschließend wird die eigentliche Suche aus den gewonnenen Informationen entworfen, deren Ziel es ist, die Benutzeranfrage zu befriedigen, die typischerweise 50-150 Treffer erwartet. Falls die Prüf-Anfrage diese Anzahl von Ergebnissen bereits liefert, wird der zweite Schritt nicht ausgeführt.

Die wesentliche Innovation liegt in der Terminierungs-Strategie, diese wird nicht durch die einfache TTL realisiert, sondern wird durch eine dynamische Infektionsrate und einen bestimmten Abbruchzeitpunkt nach einer festen Anzahl von Runden spezifiziert .

Bei der Auswertung der Ergebnisse wurde die Genauigkeit (umso kleiner, umso genauer) der Suche, die versendeten Nachrichten und das Verhältnis der gesendeten und empfangenen Nachrichten (umso kleiner, umso besser) verglichen. Im Versuch wurden 10000 Knoten mit einem Verbindungsgrad von 30 innerhalb von 100 Runden gemessen.

[HK07]

3.1.2.4 Soziale Netzwerke

Auch in sozialen Netzwerken werden Epidemien genutzt, um Informationen effektiv zu verbreiten. So gibt es beispielsweise auf der Plattform Facebook die Möglichkeit Inhalte durch den „like“ bzw. „gefällt mir“ Button zu kennzeichnen, was zur Folge hat, dass das markierte Objekt im News-Feed aller Freunde sichtbar wird. Falls mehr als ein Kontakt anschließend die Nachricht durch weiteres publizieren verbreitet, wird eine Epidemie ausgelöst bei der die Benutzer die Knoten eines Peer-2-Peer Systems darstellen.

Laut einer Veröffentlichung [Cen10] von Forschern des MIT-Instituts spielt dabei die Verknüpfung der einzelnen Cluster eine starke Rolle. Es wurde herausgefunden, dass die Knoten an der Verbreitung stärker teilnehmen, wenn mehrere ihrer Nachbarn des sozialen Netzwerks bereits ihre Zustimmung preisgegeben hatten. Außerdem wurde festgestellt, dass die Epidemie besonders stark und weitläufig war, wenn das Cluster eine Gitternetzstruktur hatte.

Weitere Informationen zu Clustering in P2P-Systemen:

[WK09]

3.2 Einsatz in MANET-Netzen

3.2.1 Routing Protokolle

Die Liste der Routing Protokolle für Ad-Hoc-Netze [WIK] ist enorm lang. Daraus ist ableitbar, dass es offensichtlich kein Omni-Routing-Protokoll gibt, das alle Bedürfnisse

der MANET-Teilnehmer befriedigt.

Das wird noch deutlicher, wenn man die Kategorien der Routing-Protokolle aus Kapitel 2 betrachtet. So sind eben Merkmale wie Energieberücksichtigung und Geografieabhängigkeit schwer miteinander vereinbar, da sie stark vom Anwendungsfall abhängen. Auch kann das reaktive oder proaktive Erstellen von Routing-Tabellen bedingt durch das Anwendungs-Szenario beeinflusst werden.

3.2.1.1 Optimized Link State Routing

Ein weit verbreitetes Verfahren ist das Optimized Link State Routing (OLSR) [RFC], eine Abwandlung des Link State Routing Algorithmus, speziell für mobile drahtlose Netzwerke. Bei diesem proaktiven Verfahren werden im ersten Schritt Nachbarn mit Hello-Nachrichten aufgespürt und anschließend deren Existenz mit Traffic-Control Nachrichten im Netz verbreitet.

Die TC-Informationen werden per Epidemie (flooding) verbreitet, allerdings nur dann weitergeleitet, wenn der Ziel-Knoten in der Liste der ausgewählten Multiple Relays (MR) vom Quell-Knoten ist. MR sind ein Teil der Knoten in unmittelbarer Nachbarschaft eines Nodes. Der Nachrichtenaustausch wird periodisch wiederholt, um Veränderungen in der Netzstruktur zu publizieren.

Die eigentliche Route der Pakete wird durch den Dijkstra-Algorithmus anhand der Hops berechnet und mit Source-Routing im Paket hinterlegt.

Das OLSR berücksichtigt die Verbindungsqualität nicht und durch seine Proaktivität werden mehr Netz- und Energieressourcen verbraucht, als eventuell benötigt werden. Da nur ein selektives Fluten stattfindet, wird zwar das Netzwerk nicht überflutet, allerdings kann es vorkommen, dass durch die fehlende Redundanz bei hohem Paket-Verlusten im Netzwerk, nicht alle Knoten optimal erreichbar sind.

[MA09]

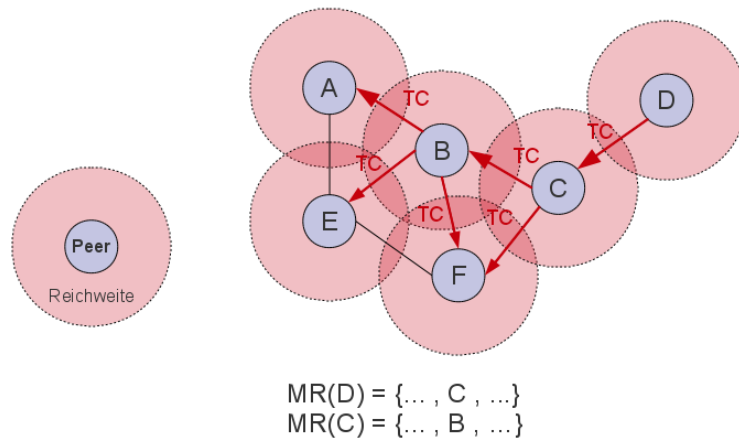


Abb. 3.5: Ausbreitung Traffic-Control Nachrichten (TC) beim OLSR

3.2.2 Hardware-Simulation

Zur Auswertung der Effektivität von epidemischen Algorithmen werden oft Simulation eingesetzt, doch können diese nur einen Teil der Realität abbilden, was die Genauigkeit der gewonnenen Erkenntnisse einschränkt. In diesem Abschnitt wird eine Hardware-Simulation (916 Mhz Einkanal Sender von RF Monolithics www.rfm.com/products/data/tr1000.pdf) vorgestellt, die aus mehr als hundert Transmitter-Knoten besteht, die mit verschiedenen Sendeleistung funken können.



Abb. 3.6: Transmitter-Knoten mit Stromversorgung (916 Mhz Einkanal) Quelle: [DG02]

Das System wurde auf den unterschiedlichen ISO/OSI Schichten analysiert und daraus Statistiken erstellt. Auf dem Layer 1 und Layer 2 (physical / data link) wurden die Paket-Verluste, die effektive Reichweite und Verbindungs-Asymmetrie gemessen. Auf dem MAC-Layer (Sublayer von Layer 2) wurde die Kollisionen und Verzögerungen ausgewertet. Auf Layer 3 und Layer 7 (network / application) wurden die Verbindungsstrukturen der Epidemie aufgezeigt.

Auch wenn der eingesetzte Algorithmus (Fluten) primitiv erscheint zeigen die gewonnenen Ergebnisse, dass sich das resultierende Verhalten auf echter Hardware überraschend komplex gestaltet.

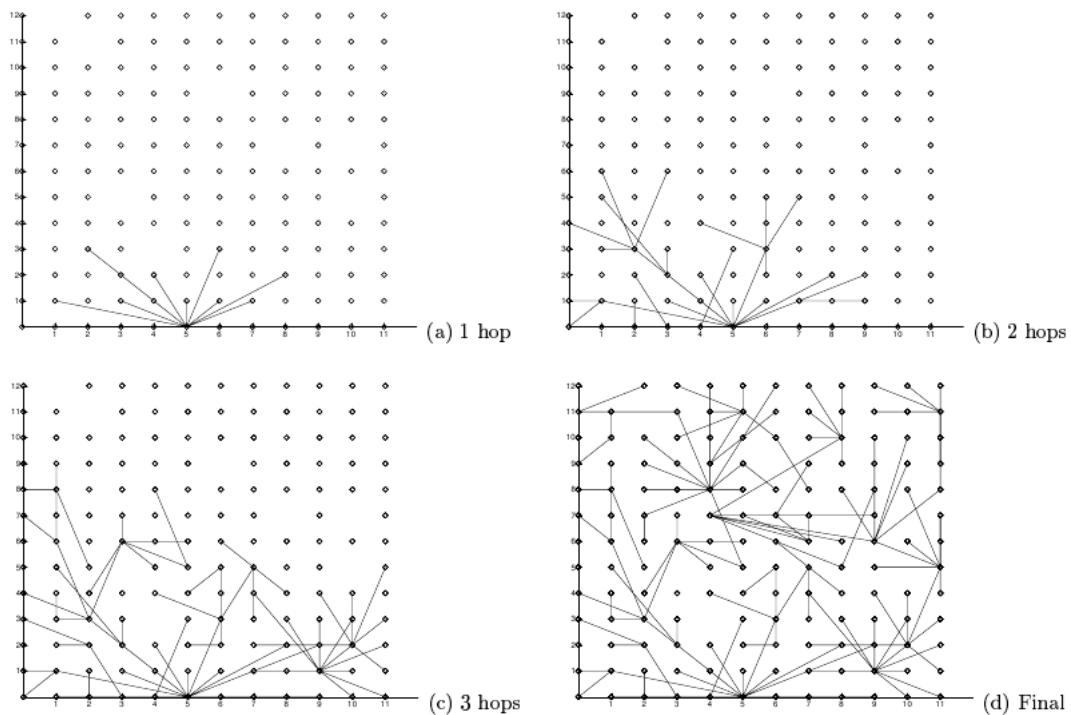


Abb. 3.7: Ausbreitung der Epidemie mit 160 Knoten in einem 12x13 quadratischen Netz mit 2 Fuß Abstand Quelle: [DG02]

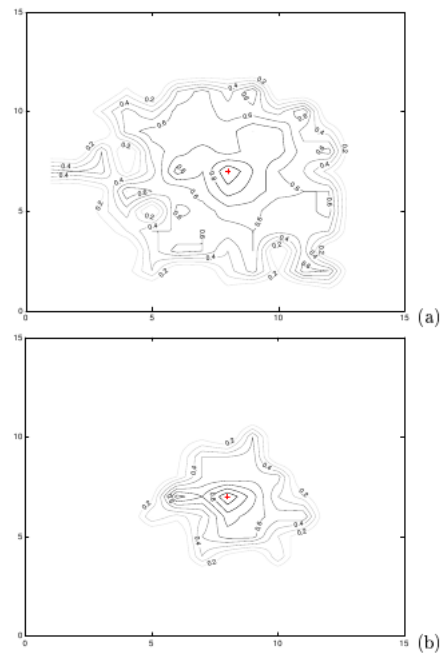


Abb. 3.8: Paket-Verluste (CRC-Checksumme incorrect) mit zwei unterschiedlichen Sendeleistungen Quelle: [DG02]

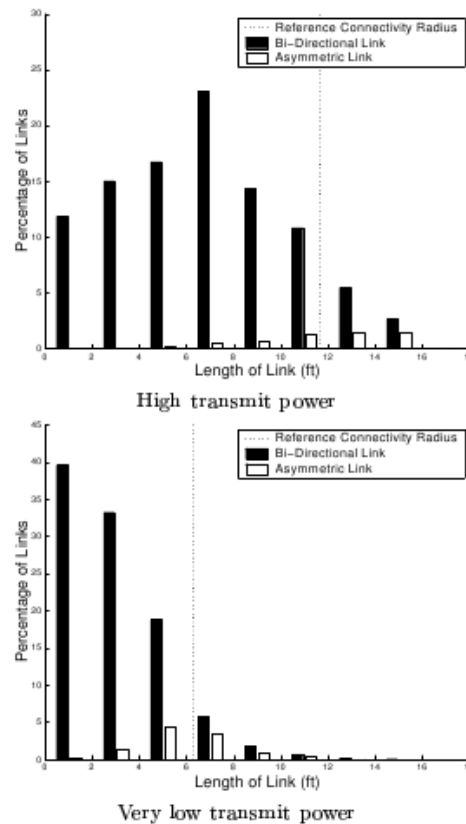


Abb. 3.9: Verteilung der bidirektionalen und asymmetrischen Verbindungen relativ zum Abstand Quelle: [DG02]

Unter anderem wurde herausgefunden, dass bei der epidemischen Ausbreitung Verbindungen über weite Distanzen vermieden werden sollten, da diese meist asymmetrisch sind. Dies führt bei Routing-Protokollen, die sich stark an Hops orientieren, oft zu falschen Annahmen für die Übertragungsgeschwindigkeit.

Auch wurde gezeigt, dass der Energieverbrauch einer Epidemie gesenkt werden kann, wenn eine niedrige Sendeleistung gewählt wird und unnötige Broadcasts vermieden werden.

Außerdem wurde deutlich gezeigt, dass selbst einfachste Protokolle mit wenigen Zuständen eine unerwartet hohe Komplexität in realen Welten erzeugen.

Verbindungs-Asymmetrie ist zu erwarten und bei dem Entwurf neuer Algorithmen sollte berücksichtigt werden, dass diese mit diesem Effekt zurecht kommen.

[DG02]

3.3 Opportunistische Netze (OPPNETs) bzw. partitionierte MANETs

3.3.1 Hypergossiping

Da verbreitete Epidemie basierte Mechanismen in MANETs auf spezielle Anwendungs-Szenarien beschränkt sind, gilt der Erforschung von allgemeingültigen Systemen großes Interesse. Hypergossiping Routing beschreibt eine verallgemeinerte Broadcast-Methode, deren Einsatz es ermöglicht, adaptiv mit einer breiten Reihe von Cluster-Formen, Knoten-Dichten, Mobilitäts-Arten, Übertragungsraten und Netzwerklasten zurecht zu kommen.

Auch wird das Verbreiten von Informationen in partitionierten mobilen Ad-hoc-Netzen unterstützt, indem benötigte Nachrichten wiederholt verbreitet werden, was unter dem Begriff Hyperflooding [KOV01] bekannt wurde, wenn erkannt wird, dass einzelne Knoten nicht erreichbar sind. Dabei wird die Zeitgültigkeit der Daten berücksichtigt, da diese je nach Einsatzgebiet stark variieren kann, so haben z.B. Sensordaten meist nur eine geringe Gültigkeitsdauer.

Durch die Berechnung der Relevanz der vorgehaltenen Informationen kann zusätzlich entschieden werden, ob die empfangene Nachricht gepuffert werden soll. Hierbei beschränkt sich die Bewertung auf das Auswerten von Mobilitätsmustern, so sollten sich Knoten, die die Partition häufig wechseln, durch eine hohe Pufferzeit auszeichnen. Außerdem macht es wenig Sinn innerhalb von mobilen Gruppen allen Knoten alle Nachrichten vorhalten zu lassen.

Folgende Erkenntnisse könnten gewonnen werden:

- Die Adaption von mathematische Modellen aus der Epidemiologie für die Informationsverbreitung in MANETs erwies sich als praktikabel, daraus lässt sich schließen, dass eine gewisse Ähnlichkeit zwischen der Verbreitung von Infektionskrankheiten und der Verteilung von Daten in mobilen Ad-Hoc-Netzen besteht.
- Die adaptierten Modelle halfen bei der Konfiguration der Protokolle an den entsprechenden Schlüsselwerten der MANET-Eigenschaften, so konnte die Infektionsrate an die Knoten-Dichte angepasst werden.
- Hypergossiping erwies sich als Anwendbar, da es sowohl mit unterschiedlichen Knoten-Dichten, als auch Mobilitätsgraden zurecht kam und dabei Überflutung vermied und Partitionierungen überwinden konnte.
- Neuartige Metriken konnten definiert werden, die das paarweise Kommunikationsverhalten von Knoten charakterisieren. Dadurch wird die Entwicklung und Anwendung von neuartigen Algorithmen erleichtert.
- Es konnten Puffer-Strategien entwickelt werden, die sich an der Mobilität von Knoten orientieren.

[Khe07]

3.3.2 Epidemie Middleware

Für den Fall, dass man auf epidemische Algorithmen zurückgreifen will, um Nachrichten in mobilen Ad-hoc-Netzen zu verschicken, sich aber mit den zugrunde liegenden Technologien nicht auseinandersetzen möchte, gibt es eine Middleware. Durch die zur Verfügung gestellte API kann der Versand von Botschaften sehr leicht implementiert werden und mit Parametern, die den Verlauf der Epidemie festlegen, bequem gesteuert werden.

Die Schnittstelle bietet die Möglichkeit zum Versand einer epidemischen Nachricht `epsend(message, recipient, reliability, time)` und eines epidemischen Broadcasts `epcast(message, percentageOfHosts, time)`. In der Variable `message` wird der Text der Botschaft hinterlegt, `recipient` ist die Adresse des Empfängers, `reliability` ist ein Wert $([0,1])$ der bestimmt, wie hoch die Wahrscheinlichkeit ist, dass die Nachricht bis zum Zeitpunkt `time` zugestellt wird, `time` gibt die Zeit an, bis zu der die Epidemie fortgeführt wird und `percentageOfHost` gibt an, wie viele Knoten bei der Broadcast-Methode innerhalb der Zeit `time` informiert werden sollen.

Das zugrunde liegende System greift hierzu auf empirische Erkenntnisse aus der Epidemiologie [Fra80] zurück, verwendet das SIS-Modell um den Infektions-Status in den einzelnen Knoten zu repräsentieren und wertet den Verbindungsgrad der Nodes im Netzwerk aus, um die Replikation und Persistenz der Nachrichten dynamisch an die gegebene Netzstruktur anzupassen. Außerdem wird der begrenzte Speicher der mobilen Geräte beim Puffern der Information berücksichtigt.

Im Modell wird von folgenden Voraussetzungen ausgegangen:

- Das Infektionsrisiko aller Knoten im Bezug zur Quelle der Epidemie ist gleich.
- Die Infektiosität eines Nodes zu einer bestimmten Nachricht ist gleich, jedoch werden Messages unterschiedlicher Infektiosität im Puffer beherbergt.
- Alle Teilnehmer des System tragen zur Verbreitung der Information bei, es gibt keine Saboteure.
- Die Buffergröße aller Endgeräte ist gleich. Die Anzahl der Netzknoten und deren Ausfällen ist unbekannt, konstant und gleich verteilt.

Mit einer Simulation in Omnet++ [OMN] konnte bestimmt werden, dass das entwickelte Protokoll, basierend auf statistischen Werten, die gewünschte Zuverlässigkeit in homogenen Netzwerken erreichen konnte. Dies geschah sowohl bei der anycast, als auch multicast Kommunikation. [MM06]

3.3.3 Modell zur Informationsausbreitung

In 3.2.2 wurde ersichtlich, dass selbst bei statischen Knoten und einem primitiven epidemischen Algorithmus die Struktur der Datenströme eine außerordentlich Komplexität annimmt.

Da die Wechselwirkungen des Systems bei einer Mobilisierung um einen großen Faktor zunehmen und die Effektivität der Verbreitungsalgorithmen stark von Art der Interaktion der einzelnen Systemteilnehmer abhängt, ist ein Modell, das die Charakteristika der auftretenden Szenarien widerspiegelt enorm hilfreich bei der Entwicklung neuer hochperformanter Algorithmen, die sich gegebenenfalls dynamisch an die vorherrschenden Bedingungen anpassen.

In diesem Abschnitt wird die Entwicklung eines Modells vorgestellt, das auf Daten beruht, die in einer Simulation gewonnen wurden, und einen einfachen SI-Algorithmus [JK99] verwendet. Hierbei wird analytisch untersucht, inwiefern sich die Knoten-Dichte auf die Informationsausbreitung auswirkt.

Bei der Simulation der Bewegung der Teilnehmer wird auf das random way point Modell [JB98] zurückgegriffen, bei dem nach einer Wartezeit ein zufälliges Ziel ausgewählt wird, auf das sich mit einer bestimmten Geschwindigkeit, gleich wahrscheinlich aus 0 bis max selektiert, zu bewegt wird, um anschließend dort erneut für eine gewisse Zeit zu verharren. Die Wartezeit wird uniform aus 0 bis max gewählt.

Die Simulation wird durch eine Java-Implementierung vollzogen, die eine Abstraktion des IEEE 802.11b MAC layer erzeugt.

Es werden folgende Parameter verwendet:

- Fläche 1000m x 1000m
- Anzahl der Knoten ist variabel
- Kommunikationsreichweite 75m
- Übertragungsrate 2048 Kbit/s
- Erkundungszeit uniform zwischen 2 und 3 Sekunden
- Nutzdatengröße 510 bytes
- Bewegungs-Geschwindigkeit uniform zwischen 3 und 70 km/h
- Wartezeit uniform zwischen 0 und 100 Sekunden

Die Infektionsrate ergibt sich aus der Knoten-Dichte und der Simulations-Parameter.

Folgende Ergebnisse wurden durch die Simulation erzielt:

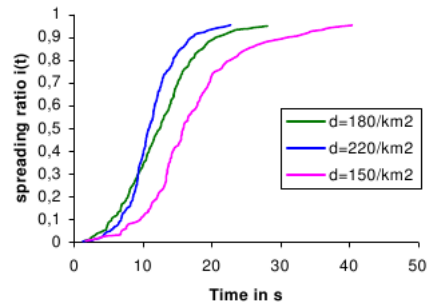


Abb. 3.10: Ausbreitung nach Zeit in Abhängigkeit der Knoten-Dichte d Quelle: [AK02]

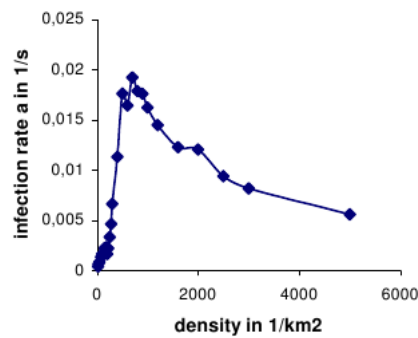


Abb. 3.11: Errechnete Infektionsrate in Abhängigkeit der Knoten-Dichte Quelle: [AK02]

Die Untersuchung zeigte, dass sowohl bei zu geringer Knoten-Dichte, die eine große Fragmentierung und Zerstreutheit erzeugt, als auch bei zu hoher Teilnehmer-Anzahl pro Fläche, die zu erhöhten Übertragungsstörungen führt, die Infektionsrate stark sinkt. Dies bestätigt bisherige Erkenntnisse [KS78], bei denen in einem statischen Ad-hoc-Netzwerk von optimal sechs Nachbarn ausgegangen wird.

[AK02]

4 Zusammenfassung und Ausblick

Zusammenfassend lässt sich sagen, dass sich epidemische Algorithmen als probates Mittel bei der Replikation von verteilten Datenbanken, zum Suchen und Auffinden von Knoten in P2P-Filesharing System und zum Routing in MANETs erwiesen haben und dort verbreitet zum praktischen Einsatz kommen, weil sie sich durch ihre gute Skalierung, Robustheit und Toleranz bei Teilausfällen von Knoten ausgezeichnet haben. Hierbei beschränken sich allerdings die eingesetzten Techniken oft auf das primitive Fluten (flooding).

In opportunistischen Netzen oder partitionierten MANETs gibt es erfolgversprechende Ansätze, die Partitionen durch Hyperflooding [KOV01] überwinden können, ohne Broadcast Stürme (broadcast storms) [SNS99] auszulösen und somit effektiv zur Informationsausbreitung in Sensornetzen beitragen können.

Die Verbreitung von Anwendungen auf modernen mobilen Endgeräten, um neuartigen innovative Dienste zur Verfügung zu stellen, lässt leider auf sich warten. So lässt sich durch den rapiden Anstieg der Verbreitung solcher Geräte und die vielseitigen möglichen Einsatzgebiete, in Hinblick auf Informationsverbreitung und Soziale Netze, ein Anstieg des Interesses an den epidemischen Techniken prophezeien, deren Entwicklung eine neue nie geahnte Informations-Ära einleiten werden.

Dabei gilt es, die Themen Datenschutz und -sicherheit, sowie Maßnahmen gegen Spam und Sabotage in einen speziellen Fokus zu stellen, um dem Missbrauch solcher Systeme ausreichend entgegen zu wirken.

Außerdem hat sich gezeigt, dass sich Algorithmen dynamisch an die vorliegenden Netzwerk-Topologie anpassen müssen, um nicht an Effektivität zu verlieren. Auch das diese Adaption keineswegs trivial ist, da das Erkennen vorliegender Netzstrukturen und das Vorhersagen möglicher Veränderungen oft an zu hohen Hürden scheitert, zeigt auf, dass auf dem Gebiet der epidemischen Informations-Dissemination noch reichlich Raum für Forschungsarbeit besteht.

Eventuell lassen sich Erkenntnisse aus verwandten Themen-Gebieten, wie der Evolutions- und Hirnforschung nutzen, um neuartige Herangehensweisen zu entwickeln, weil die Natur seit Milliarden von Jahren Epidemien, durchaus effektiv, zur Verteilung von genetischen Informationen nutzt und Nervenzellen im Gehirn als P2P-System verstanden werden können. Sich ergebende Synergieeffekte aus einer interdisziplinären Zusammenarbeit wären in diesem Bereich durchaus denkbar.

Hinzu kommt, dass Erfahrungen und Methoden aus der Informationsforschung helfen könnten, Daten zu priorisieren und deren Verteilung entsprechend unterschiedlich zu fördern. So ergibt sich durch die Verwendung von Ontologien vielleicht die Möglichkeit, den Wert einer Informationen für das System deren Teilnehmer zu ermitteln.

Literaturverzeichnis

- [AD03] Karl Aberer Anwitaman Datta, Manfred Hauswirth. Updates in highly unreliable, replicated peer-to-peer systems. Technical report, Swiss National Institute of Technology (EPFL), Lausanne, Switzerland, 2003.
- [AK02] Jing Tian Kurt Rothermel Abdelmajid Khelil, Christian Becker. An epidemic model for information diffusion in manets. Technical report, University of Stuttgart / IPVS, 2002.
- [AM92] R. M. Anderson and R. M. May. Infectious diseases of humans: Dynamics and control. Technical report, Oxford University Press, 1992.
- [BAP10] Michalis Faloutsos 3 Nicholas Valler 3 B. Aditya Prakash 1, Deepayan Chakrabarti 2 and Christos Faloutsos 1. Got the flu (or mumps)? check the eigenvalue! theorem for a wide range of virus propagation models (vpm) that include all virus propagation models in standard literature. Technical report, 1 Computer Science Department, Carnegie Mellon University, Pittsburgh 2 Yahoo! Research, Sunnyvale 3 Department of Computer Science and Engineering, University of California - Riverside, 2010.
- [BBC07] BBC. Mobile system promises free calls <http://news.bbc.co.uk/2/hi/technology/6987784.stm>. 2007.
- [Ber03a] Jem E. Berkes. Decentralized peer-to-peer network architecture: Gnutella and freenet. Technical report, University of Manitoba Winnipeg, Manitoba Canada, 2003.
- [Ber03b] Jem E. Berkes. Decentralized peer-to-peer network architecture: Gnutella and freenet. Technical report, 2003.
- [BV00] C. Schindelhauer R. Karp B. Vöcking, S. Shenker. Randomized rumor spreading. Technical report, International Computer Science Institute, Berkeley and University of California at Berkeley, Institut für Theoretische Informatik, Med. Universität zu Lübeck, Max-Planck-Institut für Informatik, 2000.
- [Cen10] Damon Centola. The spread of behavior in an online social network experiment <http://www.sciencemag.org/cgi/content/abstract/329/5996/1194>. Technical report, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, 2010.

- [CL05] Oliver P. Waldhorst Christoph Lindemann. Modeling epidemic information dissemination on mobile devices with finite buffers. Technical report, University of Dortmund Department of Computer Science, 2005.
- [CS04] Peter Mahlmann Christian Schindelbauer. Algorithmen für peer-to-peer-netzwerke. Technical report, Heinz Nixdorf Institut Fakultät EIM, Institut für Informatik Universität Paderborn, 2004.
- [Dem87] Dan Greene Carl Hauser Wes Irish John Larson Scott Shenker Howard Sturgis Dan Swinehart Doug Terry Demers, Alan. Epidemic algorithms for replicated database maintenance. Technical report, 1987.
- [DG02] Alec Woo David Culler Deborah Estrin Stephen Wicker Deepak Ganesan, Bhaskar Krishnamachari. An empirical study of epidemic algorithms in large scale multihop wireless networks. Technical report, Intel Research Berkeley, 2002.
- [Fou] Eclipse Foundation. The spatiotemporal epidemiological modeler (stem) <http://www.eclipse.org/stem/>.
- [Fra80] J. C. Frauenthal. *Mathematical Modeling in Epidemiology*. Springer-Verlag, 1980.
- [Gav09] Daniela Gavida. *Epidemic-style information dissemination in large-scale wireless networks*. PhD thesis, University of Amsterdam, 2009.
- [Gnua] *The Gnutella Protocol Specification v0.4*.
- [Gnub] *The Gnutella Protocol Specification v0.6*,
http://rfc-gnutella.sourceforge.net/src/rfc-0_6-draft.html.
- [Hei07] Andreas Heinemann. *Collaboration in Opportunistic Networks*. PhD thesis, Fachbereich Informatik der Technischen Universität Darmstadt, 2007.
- [HK07] Cecilia Mascolo Holger Kampffmeyer, Mirco Musolesi. Epitella: Improving the gnutella search algorithm through epidemic spreading models for complex networks. Technical report, Department of Computer Science, University College London, 2007.
- [Hu09] Liang Hu. *Mobile Peer-to-Peer Data Dissemination over Opportunistic Wireless Networks*. PhD thesis, Technical University of Denmark Lyngby, Denmark, 2009.
- [IC] Theodore W. Hong Oskar Sandberg & Brandon Wiley Ian Clarke, Scott G. Miller. *Protecting Free Expression Online with Freenet Freenet uses a decentralized P2P architecture to create an uncensorable and secure global information storage system*.

- [JB98] D. B. Johnson Y. C. Hu J. Broch, D. A. Maltz and Jorjeta Jetcheva. Performance comparison of multi-hop wireless ad hoc network routing protocols. Technical report, Appeared in the Proceedings of 4th Annual ACM/IEEE International Conference on Mobile Computing (MobiCom98), 1998.
- [JK99] Hari Balakrishnan Joanna Kulik, Wendi Rabiner. Adaptive protocols for information dissemination in wireless sensor networks. Technical report, Massachusetts Institute of Technology Cambridge, 1999.
- [Kan07a] Jussi Kangasharju. Opportunistic networks introduction. 2007.
- [Kan07b] Jussi Kangasharju. Opportunistic networks introduction. Technical report, Matemaattis-luonnontieteellinen tiedekunta University of Helsinki, 2007.
- [Kar00] Elias Koutsoupias Christos H. Papadimitrioi Scott Shenker Karp, Richard. Optimization problems in congestion control. Technical report, 41st Annual Symposium on Foundations of Computer Science (FOCS00), 2000.
- [Khe07] Abdelmajid Khelil. *A Generalized Broadcasting Technique for Mobile Ad Hoc Networks*. PhD thesis, Institut für Parallele und Verteilte Systeme (IPVS) u der Universität Stuttgart, 2007.
- [Kla06] Daniel Klan. Replikation in peer-to-peer systemen. Technical report, Fakultät für Informatik und Automatisierung, TU Ilmenau, 2006.
- [KOV01] G. Tsodik K. Obraczka and K. Viswanath. Pushing the limits of multicast in ad hoc networks. Technical report, USC Information Science Institute, 2001.
- [KS78] L. Kleinrock and J. Silvester. Optimum transmission radii for packet radio networks or why six is a magic number. Technical report, Department of Computer Science University of California at Los Angeles, 1978.
- [LL06a] IEEE Z. Huma Kamal Student Member IEEE Ajay Gupta Senior Member IEEE Leszek Lilien, Senior Member. Opportunistic networks: Challenges in specializing the p2p paradigm, 2006.
- [LL06b] Z. Huma Kamal Student Member Ajay Gupta Senior Member Leszek Lilien, Senior Member. Opportunistic networks: Challenges in specializing the p2p paradigm. Technical report, IEEE, 2006.
- [MA09] C.-P. Wang M. Abolhasan, B. Hagelstein J. Real-world performance of current proactive multi-hop mesh protocols. Technical report, University of Wollongong, 2009.
- [Mat] Roger Dingledine & Nick Mathewson. *Tor Protocol Specification*.
- [MM06] Cecilia Mascolo Mirco Musolesi. Controlled epidemic-style dissemination middleware for mobile ad hoc networks. Technical report, Dept. of Computer Science University College London Gower Street, London, 2006.

- [OLP] *One Laptop Per Child (OLPC)*,
<http://www.laptop.org>.
- [OMN] *OMNeT++ is an extensible, modular, component-based C++ simulation library and framework, primarily for building network simulators based on the eclipse framework* <http://omnetpp.org>
<http://www.youtube.com/watch?v=cFPJPVh3ouQ>.
- [PE] A.-M. Kermarrec L. Massoulie P.T. Eugster, R. Guerraoui. From epidemics to distributed computing. Technical report.
- [Pit87] Boris Pittel. On spreading a rumor. *siam journal on applied mathematics*. *SIAM Journal on Applied Mathematics*, 1987.
- [PKMS09] Ph.D. Professor Prof. Krishna Moorthy Sivalingam. Mobile ad hoc networks. Technical report, Dept. of CSE, IIT Madras, 2009.
- [PT05] Barry Brown Matthew Chalmers Scott Sherwood Paul Tennent, Malcolm Hall. Three applications for mobile epidemic algorithms. Technical report, Department of Computing Science University of Glasgow University Avenue, 2005.
- [RFC] *Optimized Link State Routing Protocol (OLSR)*
<http://tools.ietf.org/html/rfc3626>.
- [SNS99] Y.S. Chen S.Y. Ni, Y.C. Tseng and J.P. Sheu. Broadcast storm problem in a mobile ad hoc network. Technical report, Department of Computer Science and Information Engineering National Chiao-Tung University Taiwan, 1999.
- [Ste09] Peter Stechlinski. A study of infectious disease models with switching. Master's thesis, University of Waterloo, 2009.
- [TDH04] Peter Bleckmann Tim Daniel Hollerung. Epidemic algorithms. Technical report, Universität Paderborn, 2004.
- [WIK] *List of ad-hoc routing protocols from Wikipedia, the free encyclopedia*
https://secure.wikimedia.org/wikipedia/en/wiki/Ad_hoc_routing_protocol_list.
- [WK09] Javed Mostafa Weimao Ke. Strong ties vs. weak ties: Studying the clustering paradox for decentralized search. Technical report, Laboratory of Applied Informatics Research School of Information and Library Science & Translational and Clinical Sciences Institute University of North Carolina at Chapel Hill, 2009.

Abbildungsverzeichnis

2.1	Peer-Kommunikation im statischen Netz	
	A kann direkt mit B kommunizieren	2
2.2	Peer-Kommunikation im mobilenen Ad-hoc-Netz	
	A kommuniziert mit D über die Knoten B und C	3
2.3	Peer-Kommunikation in OPPNETs bzw. partitionierten MANETs	
	A möchte mit D kommunizieren	5
2.4	Peer-Kommunikation in OPPNETs bzw. partitionierten MANETs	
	Information wandert von A nach C und nach dem Partitionswechsel von C weiter zu F und B	6
2.5	Peer-Kommunikation in OPPNETs bzw. partitionierten MANETs	
	B wechselt die Partition und die Information erreicht schließlich D	6
2.6	Epidemiologie SIR-Zustands-Modell	9
2.7	Epidemiologie SEIR-Zustands-Modell	10
2.8	Epidemiologie MSIR-Zustands-Modell	10
2.9	Epidemiologie Träger-Zustands-Modell	11
2.10	Epidemiologie SIS-Zustands-Modell	11
2.11	Random Call - Push-Modell	12
2.12	Random Call - Pull-Modell	12
2.13	Random Call - Push & Pull-Modell	12
2.14	Informationsausbreitung im Push-Modell	
	Simulator: http://www.jonas-kress.de/epidemic/	13
2.15	Wachstumsphasen im Push-Modell Quelle: [CS04]	14
2.16	Informationsausbreitung im Pull-Modell	
	Simulator: http://www.jonas-kress.de/epidemic/	15
2.17	Wachstumsphasen im Pull-Modell Quelle: [CS04]	15
2.18	Informationsausbreitung im Push & Pull-Modell	
	Simulator: http://www.jonas-kress.de/epidemic/	16
2.19	Wachstumsphasen im Push & Pull-Modell Quelle: [CS04]	16
3.1	Schematik der Ausbreitung von Ping/Pong Deskriptoren im Gnutella- Protokoll	21
3.2	Schematik der Ausbreitung von Query/Hit Deskriptoren im Gnutella- Protokoll	21

3.3	Ausbreitung der Query Deskriptoren im Gnutella-Protokoll mit einer TTL von 5	22
3.4	Ausbreitung der Query Deskriptoren im Gnutella-Protokoll mit einer TTL von 5 und QRP	23
3.5	Ausbreitung Traffic-Control Nachrichten (TC) beim OLSR	26
3.6	Transmitter-Knoten mit Stromversorgung (916 Mhz Einkanal) Quelle: [DG02]	26
3.7	Ausbreitung der Epidemie mit 160 Knoten in einem 12x13 quadratischen Netz mit 2 Fuß Abstand Quelle: [DG02]	27
3.8	Paket-Verluste (CRC-Checksumme incorrect) mit zwei unterschiedlichen Sendeleistungen Quelle: [DG02]	28
3.9	Verteilung der bidirektionalen und asymmetrischen Verbindungen relativ zum Abstand Quelle: [DG02]	29
3.10	Ausbreitung nach Zeit in Abhängigkeit der Knoten-Dichte d Quelle: [AK02]	33
3.11	Errechnete Infektionsrate in Abhängigkeit der Knoten-Dichte Quelle: [AK02]	33