



BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

Masterarbeit

Medieninformatik

Fachbereich VI – Informatik und Medien

Untersuchung der Konfliktmanagementstrategien verschiedener offlinefähiger Systeme

Berlin, den 24. März 2018

Autorin:

Jacoba BRANDNER

Matrikelnummer:

833753

Betreuer:

Herr Prof. Dr. Hartmut SCHIRMACHER

Gutachterin:

Frau Prof. Dr. Petra SAUER

Abstract

In dieser Arbeit wird ...

Abstract

This work includes ...

INHALT

1 Einführung	5
1.1 Motivation	5
1.2 Zielstellung	5
2 Bestehende offlinefähige Systeme / Konzepte	6
2.1 Kollaborative Software	6
2.1.1 Google Docs	6
2.1.2 Google Wave	6
2.1.3 Dropbox / 'Clouds'	6
2.1.4 Kollaborative Editoren	6
2.2 Offline-First Frameworks/Bibliotheken	7
2.2.1 redux? react-native? webpack?	7
2.2.2 Realm	7
2.2.3 Datomic? (Closure)	7
2.2.4 redux-offline	7
2.2.5 redux-persist	7
2.2.6 react-native-offline	7
2.2.7 offline-plugin für webpack	7
2.2.8 hoodie	8
3 Grundlagen	9
3.1 Konflikte	9
3.2 Offline First	9
3.2.1 Anforderungen an Offline-First / PWA?	10
3.2.2 Konflikte	10
3.3 Progressive Web Apps?	10
3.3.1 ServiceWorker?	10
3.3.2 localForage AsyncStorage?	10
3.3.3 IndexedDB?	10
3.4 Kollaboration in verteilten Systemen	10
3.4.1 Last-Write-Wins (LWW) Blockieren?	10
3.4.2 Operational Transformation	11
3.4.3 Conflict-free replicated data type	13

3.5 CouchDB/PouchDB	14
3.5.1 CouchDB	14
3.5.2 PouchDB	15
4 Weiß noch keine Überschrift	16
Abkürzungen	17
Glossar	18
Abbildungsverzeichnis	19
Literaturverzeichnis	19
Anhang	22

1 EINFÜHRUNG

We live in a disconnected & battery powered world, but our technology and best practices are a leftover from the always connected & steadily powered past. [off]

Heutzutage besitzen mehr als fünf Milliarden Menschen ein Mobiltelefon und drei Milliarden haben Zugang zum Internet [Ban16].

langsame Verbindungen, Unterbrechungen. Auch bei 3G und 4G ist die Latenz schrecklich (Lie-Fi?) -> Offline-First Apps können eine bessere User experience bieten.

1.1 MOTIVATION

Ich möchte eine offlinefähige (mobile?) Anwendung entwickeln und stelle mir folgende Fragen.

Welche Software/ Framework benutze ich dazu?

Auf was muss ich bei der Auswahl achten?

Was erwarte ich von einer offline fähigen App?

(funktioniert und kein Datenverlust) -> Synchronisation und Konfliktmanagement

1.2 ZIELSTELLUNG

Wichtig: Offline nutzbar ohne Datenverlust.

Untersuchung des Verhaltens bei Konflikten (verursacht durch paralleles Arbeiten ohne Internetverbindung).

Wie leicht/schwer ist es zu implementieren?

2 BESTEHENDE OFFLINEFÄHIGE SYSTEME / KONZEPTE

2.1 KOLLABORATIVE SOFTWARE

Kann ich benutzen um kollaborativ zu arbeiten

2.1.1 GOOGLE DOCS

(benutzt OT)

2.1.2 GOOGLE WAVE

(benutzt OT)

2.1.3 DROPBOX / 'CLOUDS'

2.1.4 KOLLABORATIVE EDITOREN

Wiki: https://en.wikipedia.org/wiki/Collaborative_real-time_editor

<https://atom.io/packages/covalent>

<https://atom.io/packages/firepad>

Markdown: <https://hackmd.io/>

LaTeX: <https://www.sharelatex.com/>

Online editor: <http://etherpad.org/> (OT)

Mockingbird (tool for creating wireframes): <https://gomockingbird.com/home> (OT)

marvelapp? <https://marvelapp.com/collaboration/>

-> Zahl steigend, (nachdem Google die Drive Realtime API veröffentlicht hat, die auf Operational Transformation (OT) basiert und es *third-party Apps* ermöglicht, dieselbe Zusammenarbeit wie Google Docs zu verwenden)

plus wachsende Anzahl von offen zur Verfügung gestellter Bibliotheken und Frameworks die es ermöglichen, offlinefähige Anwendungen zu programmieren. Siehe Kapitel 2.2

2.2 OFFLINE-FIRST FRAMEWORKS/BIBLIOTHEKEN

Ich möchte aber auch eigenständig Software entwickeln die man vielleicht nicht nur zum Arbeiten nehmen kann, sondern auch um Quatsch zu machen wie Katzengifs zu teilen.

2.2.1 REDUX? REACT-NATIVE? WEBPACK?

Progressive Web App (PWA)

2.2.2 REALM

2.2.3 DATOMIC? (CLOSURE)

[rea]

2.2.4 REDUX-OFFLINE

github [reda] hackernoon [red17]

2.2.5 REDUX-PERSIST

localStorage. github [redb] medium [San16]

2.2.6 REACT-NATIVE-OFFLINE

github [Acu] medium [Acu17]

2.2.7 OFFLINE-PLUGIN FÜR WEBPACK

github [Sto] dev [web17]

2.2.8 HOODIE

[hoo]

3 GRUNDLAGEN

Software, mit der ein freigegebenes Dokument zusammen mit anderen über das Internet bearbeitet werden kann, kann überaus wertvoll sein wenn man beispielsweise in einem Team arbeitet. Heutzutage gibt es viele webbasierte Software für simultanes kollaboratives Editieren (*Textdokumente, Tabellenkalkulationen, Präsentationen, Quellcode*). Im Kapitel 2 werden Produkte und Frameworks vorgestellt, mit deren Hilfe man solche Software erstellen kann.

Lassen Sie uns einen Moment nehmen, um genauer zu definieren, was wir unter dem Begriff kollaboratives Echtzeit-Editieren verstehen. Wir möchten, dass mehrere Personen, die an verschiedenen Computern arbeiten, jederzeit Änderungen an einem auf einem Server gehosteten Dokument vornehmen können. Diese Änderungen werden sofort mit den anderen KollegInnen synchronisiert. Kein Client sollte vor einer Änderung mit dem Server oder einem anderen Client kommunizieren müssen. Insbesondere ist es nicht erforderlich, eine Sperre vom Server zu erhalten um ein Dokument zu bearbeiten und gleichzeitige Editierungen können auftreten. Nachdem alle Änderungen synchronisiert wurden, sollte jeder Client das exakt gleiche Dokument sehen.

3.1 KONFLIKTE

Verteilte Systeme: Das ist ein mächtiger Begriff für viele Ideen und Konzepten, aber es läuft in der Regel darauf hinaus: Da sind zwei oder mehr Computer, die durch ein Netzwerk verbunden sind und es wird versucht, dass einige der Daten auf beiden Computern gleich aussehen.

Zwei Geräte, ein Server, über Netzwerk verbunden.

Einfachster Fall:

3.2 OFFLINE FIRST

1. Separate Apps from Data
2. Deliver App Code (and make it cachable) appcache & ServiceWorkers

3. Save Data Offline localStrocare / localForage, IndexedDB, other Wrapper
4. Detect Connectivity navigator.onLine (Lie-fi)
5. Sync Data - Build upon existing solutions – CouchDB/PouchDB | remoteStorage

3.2.1 ANFORDERUNGEN AN OFFLINE-FIRST / PWA?

1. Lese- und Schreibfähigkeit
2. Sync im Hintergrund
3. Cache sollte in der Lage sein App-Aktualisierungen zu “überstehen”

3.2.2 KONFLIKTE

3.3 PROGRESSIVE WEB APPS?

3.3.1 SERVICEWORKER?

3.3.2 LOCALFORAGE || ASYNCSTORAGE?

3.3.3 INDEXEDDB?

3.4 KOLLABORATION IN VERTEILTEN SYSTEMEN

Es stellt sich heraus, dass die Implementierung dieser Art von Echtzeit-Zusammenarbeit alles andere als trivial ist. Im Folgenden werden die drei Strategien OT, Conflict-free replicated data type (CRDT) und Last-Write-Wins (LWW) vorgestellt plus CouchDBs Peplikationsmodell.

3.4.1 LAST-WRITE-WINS (LWW) || BLOCKIEREN?

NoSQL bietet dies, indem es einen Zeitstempel von irgendeiner Art beibehält, der ihnen hilft, zu entscheiden, welcher Schreibvorgang zuletzt kam. Datenbanken wie DynamoDB¹ oder Cassandra² verwenden LWW, um Schreibvorgänge zu verarbeiten. Bei die-

¹https://aws.amazon.com/de/dynamodb/faqs/What_is_a_readwrite_capacity_unit

²<https://docs.datastax.com/en/cassandra/3.0/cassandra/dml/dmlConfigConsistency.html>

sen Leuten erfordern stark konsistente Schreibvorgänge das Schreiben in ein Quorum von Shards, wodurch Sie mehr Geld kosten.

3.4.2 OPERATIONAL TRANSFORMATION

OT ist eine weit verbreitete Technologie zur Unterstützung von Funktionalitäten in Kollaborativer Software. Sie stammt aus einer im Jahre 1989 veröffentlichten Forschungsarbeit und wurde ursprünglich nur für die gemeinsame Bearbeitung von Klartext-Dokumenten entwickelt [EG89]. Später ermöglichte weitere Forschung OT durch Unterstützung von *Sperrungen, Konfliktlösungen, Benachrichtigungen, Bearbeitung von Baumstrukturierten Dokumenten, ...* zu verbessern und erweitern. Im Jahr 2009: Google Wave, Google Docs

Es wird das Problem untersucht, dass OT in einer idealen Umgebung löst und dadurch zu einem funktionierendem Algorithmus gelangt.

Ziel: Mehrere BenutzerInnen können gleichzeitig an einem Dokument arbeiten, sehen Änderungen der anderen in Echtzeit (live), ohne dass einer Verzögerung durch die Netzwerklatenz verursacht wird. Gleichzeitig auftretende Mehrfachänderungen sollen nicht zu unterschiedlichen Dokumentenzuständen führen.

FUNKTIONSWEISE

Kollaborative Systeme, die OT verwenden, benutzen normalerweise den replizierten Dokumentenspeicher. Das heißt jeder **Client** verfügt über eine eigene Kopie des Dokuments. Jede Änderung an einem freigegebenen Dokument wird als Operation dargestellt. Operationen sind Repräsentationen von Änderungen an einem Dokument. (Beispielsweise: Füge 'Hello world!' an Position 0 in das Textdokument ein). Eine Operation zeichnet im Wesentlichen den Unterschied zwischen einer und der nachfolgenden Version eines Dokuments auf. Die Anwendung einer Operation auf das aktuelle Dokument führt zu einem neuen Dokumentstatus. Die Operationen erfolgen auf lokalen Kopie und die Änderungen werden an alle anderen **Clients** weitergegeben. Wenn ein **Client** die Änderungen von einem anderen **Client** empfängt, werden die Änderungen normalerweise **vor** ihrer Ausführung transformiert. Die Transformation stellt sicher, dass anwendungsabhängige Konsistenzkriterien (Invarianten) von allen Standorten gepflegt werden.

Es gibt die Operationen Einfügen

Das Einfügen besteht aus dem eingefügten Text und dessen Position im Dokument (`insert('h', 0)`). Für die Position kann ein Koordinatensystem ermittelt werden (Zeilennummer: Position in Zeile oder einfacher: Dokument wie eine Folge von Zeichen behandeln, also einfach einen nullbasierten Index vergeben.)

und Löschen

Löschen(5,6) = löscht 5 Zeichen, beginnend bei Position 6. Mehr benötigt man nicht, denn update = delete & insert

Um gleichzeitige Operationen zu behandeln, gibt es eine Funktion (normalerweise Transform genannt), die zwei Operationen übernimmt, die auf denselben Dokumentstatus angewendet wurden (aber auf verschiedenen Clients). Daraus wird eine neue Operation berechnet, die nach der zweiten Operation angewendet werden kann. Diese behält die erste beabsichtigte Änderung der Operation.

Des Weiteren unterstützt OT Operationen wie update, point, lock.

Beispiel : Benutzer A fügt an Position 12 das Zeichen 'A' ein Benutzer B fügt am Anfang des Dokuments ein 'B' ein. Die konkurrierenden Operationen sind daher Einfügen (12, 'A') und Einfügen (0, 'B'). Wenn wir die Operation von B einfach an Client A senden und dort anwenden würden, gäbe es ein Problem. Aber wenn die Operation von A an B gesendet, und angewandt wird nachdem Operation B angewandt wurde ist, würde das Zeichen 'A' eine Position zu weit links von der korrekten Position eingefügt werden. Dokumentstatus A und Dokumentstatus B sind nicht identisch.

Daher muss A's `insert(12, 'A')` gegen die Operation von B transformiert werden. So wird berücksichtigt, dass B ein Zeichen vor der Position 12 eingefügt hat (die die Operation `insert(13, 'A')` erzeugt.)

Diese neue Operation kann auf Dokument B nach B's Operation angewandt werden. Die Grundidee von OT besteht darin, die Parameter einer Editieroperation gemäß den Auswirkungen **zuvor ausgeführter** konkurrierender Operationen anzupassen, so dass die transformierte Operation die korrekte Wirkung erzielen und die Dokumentenkonsistenz aufrechterhalten kann.

STRUKTUR?

Transformations- 1. Kontrolle, 2. Eigenschaften, Bedingungen, 3. Funktionen
Vorteile der Trennung?

KRITIK

False-Tie puzzle? *A Generic Operation Transformation Scheme for Consistency Maintenance in Real-time Cooperative Editing Systems* und *Achieving convergence, causality-preservation, and intention-preservation in real-time cooperative editing systems*

Während der klassische OT-Ansatz, Operationen durch ihre Versätze im Text zu definieren, einfach und natürlich zu sein scheint, werfen real verteilte Systeme ernsthafte Pro-

bleme auf. Nämlich, dass sich die Operationen mit endlicher Geschwindigkeit fortpflanzen, die Zustände der TeilnehmerInnen sind oft verschieden, so dass die resultierenden Kombinationen von Zuständen und Operationen extrem schwer vorherzusehen und zu verstehen sind. Wie Li und Li es ausdrückten: "Aufgrund der Notwendigkeit, eine komplizierte Fallabdeckung in Betracht zu ziehen, sind formale Beweise sehr kompliziert und fehleranfällig, selbst für OT-Algorithmen, die nur zwei charakteristische Primitive behandeln (Einfügen und Löschen)" [LL10].

Damit OT funktioniert, muss jede einzelne Änderung an den Daten erfasst werden: "Einen Schnappschuss des Zustands zu erhalten, ist normalerweise trivial, aber das Erfassen von Bearbeitungen ist eine ganz andere Sache. [...] Der Reichtum moderner Benutzerschnittstellen kann dies problematisch machen, besonders in einer browserbasierten Umgebung" [Fra09].

KONSITENZMODELLE??? CC MODELL, CCI, CSM, CA

3.4.3 CONFLICT-FREE REPLICATED DATA TYPE

Die Idee von CRDTs ist, dass jeder "Typ"(wie ein Einkaufswagen) mit Intelligenz handelt, um Konflikte automatisch zu lösen.

CRDTs sind Objekte, die ohne teure Synchronisation aktualisiert werden können. Sie konvergieren schließlich, wenn alle gleichzeitigen Aktualisierungen kommutativ³ sind und wenn alle Aktualisierungen schließlich von jeder Replik ausgeführt werden [SPBZ11a]. Um diese Garantien zu geben, müssen diese Objekte bestimmte Kriterien erfüllen, welche im Folgenden beschrieben werden [SPBZ11b].

ZUSTANDBASIERTER ANSATZ

Wenn ein Replikat ein Update von einem Client empfängt, aktualisiert es zuerst seinen lokalen Status und dann, einige Zeit später, seinen **vollständigen Status**. So sendet jedes Replikat gelegentlich seinen vollständigen Status an ein anderes Replikat im System. Um ein Replikat, das den Status eines anderen Replikats empfängt, wendet eine **Zusammenführungsfunktion** (merge) an, um den empfangenen Status mit dem lokalen Status zusammenzuführen. Entsprechend sendet dieses Replikat gelegentlich auch seinen Status an ein anderes Replikat, sodass jedes Update schließlich alle Replikate im System erreicht.

³Unverändert bei Vertauschen der Operanden

OPERATIONSBASIERTER ANSATZ (OT?)

Bei diesem Ansatz sendet ein Replikant seinen vollständigen Status (kann groß sein) nicht an ein anderes Replikant. Stattdessen sendet es nur den **Aktualisierungsvorgang** an **alle** anderen Replikate im System und erwartet von ihnen, dass sie das Update auf sich anwenden.

Da es sich um einen Sendevorgang handelt, wenn zwei Updates u_1 und u_2 , bei einem Replikant i angewendet werden und diese Updates an zwei Replikate r_1 und r_2 gesendet werden, können diese Updates in unterschiedlicher Reihenfolge bei diesen Replikaten ankommen. r_1 kann sie in der Reihenfolge u_1, u_2 empfangen, während bei r_2 die Updates in umgekehrter Reihenfolge (u_2, u_1) ankommen können. Sind die Aktualisierungen **kommutativ**, können die Repliken zusammengeführt werden, egal in welcher Reihenfolge die Updates bei ihnen ankommen - der resultierende Zustand ist derselbe. In diesem Modell wird ein Objekt, für das alle gleichzeitigen Aktualisierungen kommutativ sind, CmRDT (commutative replicated data type - kommutativ replizierter Datentyp) genannt.

Beispiel:...

CRDTs befassen sich mit einem interessanten und grundlegendem Problem in verteilten Systemen, haben jedoch eine wichtige Einschränkung: "Da ein CRDT konstruktionsbedingt keinen Konsens verwendet, hat der Ansatz starke Einschränkungen; Dennoch sind einige interessante und nicht-triviale CRDTs bekannt- [SPBZ11b]. Die Einschränkung ist, dass die CRDT-Adresse nur einen Teil des Problemraums betrifft, da nicht alle möglichen Aktualisierungsoperationen kommutativ sind und daher nicht alle Probleme in CRDTs umgewandelt werden können. Auf der anderen Seite können CRDTs für einige Arten von Anwendungen durchaus nützlich sein, da sie eine nette Abstraktion zur Implementierung replizierter verteilter Systeme bieten und gleichzeitig theoretische Konsistenzgarantien bieten.

3.5 COUCHDB/POUCHDB

3.5.1 COUCHDB

Vektoruhr⁴ test test

content addressable versions: Idee: Nimm den Objektkinhalt (content) , und jag ihn durch eine Hashfunktion

⁴https://en.wikipedia.org/wiki/Vector_clock

3.5.2 POUCHDB

4 WEISS NOCH KEINE ÜBERSCHRIFT

- was verspricht redux-offline
- Wie funktioniert redux-offline? Welche Strategie zur Konfliktlösung wird verwendet?
- was könnte daran problematisch sein?
- Wie hoch ist der Implementierungsaufwand?
- Dasselbe für Couch & Pouch

Implementierungsaufwand: Anforderungen an Simulator? Konzeption? usw.

ABKÜRZUNGEN

CRDT Conflict-free replicated data type

LWW Last-Write-Wins

OT Operational Transformation

PWA Progressive Web App

GLOSSAR

Hashfunktion

TODO: ist eine Abbildung, die eine große Eingabemenge (die Schlüssel) auf eine kleinere Zielmenge (die Hashwerte) abbildet

Kollaborativ

Als kollaborative Software oder kollaboratives System wird eine Software zur Unterstützung der computergestützten Zusammenarbeit in einer Gruppe über zeitliche und/oder räumliche Distanz hinweg bezeichnet

Netzwerklatenz

Die Wartezeit, die im Netzwerk verbraucht wird bevor eine Kommunikation beginnen kann, wird als Netzwerklatenz oder nur Latenz bezeichnet

Progressive Web App

Oder "fortschrittliche Web App" eine mobil nutzbare Webseite, erstellt mit den Webstandards, die als Symbiose aus einer nativen, mobilen Anwendung und einer responsiven Webseite beschrieben werden kann. Die Idee dahinter ist, dass Apps zukünftig nicht mehr über einen App Store, sondern über den Browser installiert werden kann

ABBILDUNGSVERZEICHNIS

LITERATURVERZEICHNIS

- [Acu] ACUÑA, Raúl G.: *react-native-offline*. <https://github.com/rauliyohmc/react-native-offline>,. – Zugriff: 12.04.2018 2.2.6
- [Acu17] ACUÑA, Raúl G.: Your React Native Offline Tool Belt. In: *Medium* (2017), 6. – Zugriff: 12.04.2018 2.2.6
- [Ban16] BANK, World: World Development Report 2016: Digital Dividends / International Bank for Reconstruction and Development / The World Bank. Washington DC, 2016. – Research Report. – doi: doi:10.1596/978-1-4648-0671-1 1
- [EG89] ELLIS, C. A. ; GIBBS, S. J.: Concurrency Control in Groupware Systems. In: *SIGMOD Rec.* 18 (1989), jun, Nr. 2, 399–407. <http://dx.doi.org/10.1145/66926.66963>. – DOI 10.1145/66926.66963. – ISSN 0163–5808 3.4.2
- [Fra09] FRASER, Neil: Differential Synchronization. Version: Jan 2009. <https://neil.fraser.name/writing/sync/eng047-fraser.pdf>. 2009. – Research Report. – 8 S. 3.4.2
- [hoo] *Hoodie – The Offline First Backend*. <http://hood.ie>,. – Zugriff: 12.04.2018 2.2.8
- [LL10] LI, Du ; LI, Rui: An Admissibility-Based Operational Transformation Framework for Collaborative Editing Systems. In: *Comput. Supported Coop. Work* 19 (2010), feb, Nr. 1, 1–43. <http://dx.doi.org/10.1007/s10606-009-9103-1>. – DOI 10.1007/s10606-009-9103-1. – ISSN 0925–9724 3.4.2
- [off] *Offline First*. <http://offlinefirst.org/>,. – Zugriff: 12.04.2018 1
- [rea] *realm – The new standard in data synchronization*. <https://realm.io/>,. – Zugriff: 12.04.2018 2.2.3
- [reda] *redux-offline*. <https://github.com/redux-offline/redux-offline>,. – Zugriff: 12.04.2018 2.2.4
- [redb] *react-native-offline*. <https://github.com/rt2zz/redux-persist>,. – Zugriff: 12.04.2018 2.2.5

- [red17] Introducing Redux Offline: Offline-First Architecture for Progressive Web Applications and React Native. In: *HACKERnoon* (2017), 3. – Zugriff: 12.04.2018 2.2.4
- [San16] SANFORD, Clark: Persistence is Key: Using Redux-Persist to Store Your State in LocalStorage. In: *Medium* (2016), 12. – Zugriff: 12.04.2018 2.2.5
- [SPBZ11a] SHAPIRO, Marc ; PREGUIÇA, Nuno ; BAQUERO, Carlos ; ZAWIRSKI, Marek: A comprehensive study of Convergent and Commutative Replicated Data Types / Inria – Centre Paris-Rocquencourt ; INRIA. Version: Jan 2011. <https://hal.inria.fr/inria-00555588>. 2011 (RR-7506). – Research Report. – 50 S. 3.4.3
- [SPBZ11b] SHAPIRO, Marc ; PREGUIÇA, Nuno ; BAQUERO, Carlos ; ZAWIRSKI, Marek: Conflict-free Replicated Data Types. Version: Jul 2011. <https://hal.inria.fr/inria-00609399>. 2011 (RR-7687). – Research Report. – 18 S. 3.4.3, 3.4.3
- [Sto] STOLYAR, Arthur: *offline-plugin*. <https://github.com/NekR/offline-plugin>, . – Zugriff: 12.04.2018 2.2.7
- [web17] *Easy Offline First Apps With Webpacks Offline Plugin*. <https://dev.to/kayis/easy-offline-first-apps-with-webpacks-offline-plugin>, 2017. – Zugriff: 12.04.2018 2.2.7

ANHANG

EIDESSTATTLICHE ERKLÄRUNG

CD-INHALT

Auf der beigefügten CD befinden sich

- Die schriftliche Ausarbeitung dieser Masterrarbeit im PDF-Format
- Das erstellte Projekt