



Doctoral Thesis

Data storage, transfers and communication in personal clouds

Author(s):

Ucan, Ercan

Publication Date:

2014

Permanent Link:

<https://doi.org/10.3929/ethz-a-010111223> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

DISS. ETH NO. 21762

**DATA STORAGE, TRANSFERS AND COMMUNICATION
IN PERSONAL CLOUDS**

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

ERCAN UCAN

Master of Science in Computer Science, University of Illinois at
Urbana-Champaign, USA

January 19, 1983

citizen of Turkey

accepted on the recommendation of

Prof. Dr. Timothy Roscoe, examiner
Prof. Dr. Gustavo Alonso, co-examiner
Dr. Anne-Marie Kermarrec, co-examiner

2014

Abstract

A modern user of technology today owns a collection of devices, typically fewer than ten, such as smart-phones, tablets, laptops, PCs, and file servers at home or in the office. With the emergence of cloud computing, the user can also add virtual machines and associated storage into their computation and communication environments by renting these resources on demand from cloud providers such as Amazon EC2.

A *personal cloud* in this dissertation refers to the aforementioned ensemble of personal devices owned by a single user, where the devices interact in a peer-to-peer fashion in combination with the acquirable resources from the cloud. This dissertation addresses the problems of data storage, replication, transfers and communication in personal cloud platforms. The main motivation for doing the research presented in this dissertation as opposed to a research on building central cloud services is the fact that such an approach to personal data storage and replication is likely to offer a greater degree of privacy. Moreover, this approach is more resilient in the face of a provider becoming insolvent, or the victim of a large-scale compromise of private data.

Anzere Personal Cloud addresses the problem of data storage and replication in personal clouds. It proposes a personal data storage system that aims to preserve a user's growing body of personal data by replication, and that makes it selectively available according to the user's applications and needs, specified as a set of declarative replication policies. Anzere demonstrates that the range of allowable policies (*i.e.*, the expressivity of the policy language) can be widened over previous systems without sacrificing scalability. It proposes a way in which the replication policies for personal data can be written independently of specific devices and can even result in the system acquiring and releasing virtual resources on-demand. Moreover, in order to preserve policy goals, Anzere can react to changes in the environment such as failures or network outages.

Following up on Anzere, Dexferizer tackles the problem of optimizing data transfers in personal clouds as there are a number of resources that can be considered scarce in such environments when data transfers are considered. Firstly, bandwidth may be limited. Secondly, energy (*i.e.*, battery life) is a valuable resource that should be used carefully. Thirdly, money is also a constraining resource

in these environments. Dexferizer proposes a transfer optimization framework, which employs a declarative approach whereby high-level requests are made on a *transfer service*. This service then schedules data transmission and late-binds the mechanism used for transferring an object, the paths taken by the data through the network, and the origin nodes of the data itself. Furthermore, clients of Dexferizer can write *transfer policies* in order to constrain how objects are transferred (such as avoiding public links or intermediate nodes, or transferring only at certain times). They can also choose the optimization metrics while planning the transfer (such as monetary cost, latency, etc.) as well as the relative importance of particular object types from a transfer perspective.

The third problem explored in the dissertation is how to provide the users of personal clouds with means to make sense of the detailed and complicated underlying network information and low-level network measurements. This helps them to further optimize the data transfers within their storage and replication systems. This work is designed to integrate with the Dexferizer framework. It introduces the idea of data transfer estimators which employ existing and new statistical models to estimate the cost (such as transfer time, money, energy consumption) of data transfers.

The rest of the work presented in this dissertation is about communication in personal cloud platforms. The first half of the work tackles the problem of how to perform intra-personal cloud routing: the routing within a single personal cloud. The work proposes a novel approach inspired from declarative networking, combining detailed representations of the specific network elements with concise and executable descriptions of routing algorithms. The work extends declarative networking with the discovery and representation of the network resources, and with using constraint logic programming in order to enable flexible optimization of routes based on various performance metrics. The second half of the work tackles the inter-personal cloud routing problem: routing between instances of personal clouds owned by different users. This work first presents and motivates the problem, and then proposes an approach that allows users to flexibly and selectively advertise network and routing information to one another and to optimize the routes for various metrics such as latency, bandwidth, monetary cost and energy.

Kurzfassung

Heutzutage besitzt jeder Mensch ein Anzahl, normalerweise bis zu 10, verschiedener programmierbarer Geräte, zum Beispiel Smartphones, Tablets, Laptops, Computer und Server zu Hause oder auf der Arbeit. Durch die Entstehung von Cloud Computing können zusätzliche virtuelle Geräte und entsprechender Speicherplatz hinzugefügt werden indem man von Anbietern wie zum Beispiel Amazon EC2 entsprechende Optionen erwirbt.

Eine *Personal Cloud* in dieser Dissertation beschreibt diese Ansammlung von Geräten die genau einem Nutzer gehören. Die Geräte des Nutzers sind hierbei über ein Peer-to-Peer Netzwerk miteinander verbunden und können Ressourcen über die Cloud beziehen. In dieser Arbeit wird beschrieben wie Datenhaltung, Datenreplikation, Datenübertragungen und Kommunikation in solchen Personal Clouds gehandhabt werden kann. Dies wird motiviert durch den erhöhten Datenschutz in Personal Clouds im Vergleich zu herkömmlichen, zentralisierten Cloud-Systemen. Ausserdem wird durch Personal Clouds der Nutzer besser vor einer etwaigen Insolvenz und kompromittierenden Datenübergriffen des Anbieters geschützt.

Um das Problem der Datenhaltung und -replikation in Personal Clouds zu untersuchen, wurde die Anzere Personal Cloud entwickelt. Es bietet dem Nutzer die Möglichkeit seine Daten durch Replikation konsistent zu halten und entsprechend seinen Anwendungen und Geräten nutzbar zu machen. Die Nutzungsanweisungen werden in deklarativen Regeln festgehalten. In Anzere können diese Regeln beliebig ausdrucksstark gemacht werden ohne Skalierbarkeit einzubüssen. Es bietet die Möglichkeit, Replikationsanweisungen unabhängig von den Geräten zu machen, wobei das System ausserdem virtuelle Ressourcen automatisch allokiert und aufgeben kann. Änderungen in der Systemumgebung, wie zum Beispiel Stromausfälle, können durch Anzere dynamisch kompensiert werden um die Replikationsziele des Nutzers zu verwirklichen.

Auf Anzere aufbauend ist Dexferizer ein System welches die Datenübertragungen für den Nutzer optimiert, welche aufgrund von Ressourcenmangel oftmals nicht ausreichend mit herkömmlichen Strategien realisiert werden können. Beispiele für solche limitierten Ressourcen sind Bandweite oder Batterielebensdauer. Dexferizer implementiert ein Optimierungs-Framework welches durch einen "Transfer Service" gemachte deklarative Anfragen des Nutzers umsetzt. Dieser Service

plant die Datenübertragungen sowie den Netzwerkpfad der Nutzerdaten. Nutzer von Dexferizer können diese Datenübertragungen ausserdem spezifizieren durch Übertragungsregeln, zum Beispiel dass während einer Übertragung öffentliche Knotenpunkte vermieden werden sollten oder dass nur zu bestimmten Zeiten Daten übertragen werden. Weitere absolute Parameter sind die Optimierungsmetriken, die intern im System verwendet werden, wie die entstehenden Kosten oder Latenz. Zusätzlich kann die Wichtigkeit der Datenobjekte relativ zueinander beschrieben werden um Übertragungen entsprechend zu optimieren.

Das dritte Problem, welches in dieser Dissertation bearbeitet wird, ist wie die Nutzer von Personal Clouds Sinn und erweiterten Nutzen aus den verfügbaren Netzwerk-Informationen und -Messungen gewinnen können. Dies hilft ihnen ihre Datenübertragungen innerhalb ihrer Datenhaltungs- und Datenreplikationssysteme weiter zu verbessern. Das System, welches diese Konzepte umsetzt, kann direkt in Dexferizer integriert werden. Es ermöglicht die Verwendung von existierenden und neuen Statistiken um die Kosten (z.B. übertragungszeit, monetäre Ressourcen, Energieverbrauch) von Datenübertragungen besser abschätzen zu können.

Zuletzt beschäftigt sich diese Dissertation mit der Kommunikation in und zwischen Personal Clouds. Das erste Problem hierbei ist Routing innerhalb einer Personal Cloud: Um dieses Problem zu lösen entwickeln wir einen Ansatz, der Netzwerkelemente intern repräsentiert und mit präzisen und ausführbaren Beschreibungen von Routing-Algorithmen vereinigt. Er erweitert deklarative Ansätze indem Netzwerkressourcen automatisch entdeckt und repräsentiert werden und indem logisches Programmieren dazu verwendet wird um verschiedene Routen anhand von unterschiedlichen Performance-Metriken zu optimieren. Das zweite Problem, welches in diesem Kontext bearbeitet wird, ist das Problem Daten zwischen verschiedenen Personal Clouds von verschiedenen Nutzern zu verteilen. Hierfür entwickeln wir Methoden, die es ermöglichen, Netzwerk- und Routinginformationen miteinander auszutauschen und entsprechend von unterschiedlichen Metriken (Latenz, Bandweite, Energieverbrauch etc.) zu optimieren.