Autonomous Storage Management for Personal Devices with PodBase

Ansley Post  Petr Kuznetsov
Juan Navarro Peter Druschel

MPI-SWS
TU Berlin/Deutsche Telekom Labs
TU Munich
Outline

• Motivation
• Problem Statement
• System: PodBase
• Evaluation
• Conclusion
Enterprise Data Management

Redundancy  Offline Storage  Offsite Storage

All professionally managed!
Personal Data Management

- Heterogeneous devices, connectivity, use cases
- User are inexperienced and reluctant administrators
Personal Data Management

• Use available storage/connectivity to replicate for durability
Personal Data Management

- Use available storage/connectivity to replicate for availability
Outline

• Motivation
• Problem Statement
• System: PodBase
• Evaluation
• Conclusion
PodBase

- Automated, transparent data management for personal devices
- OS and vendor neutral.
- Transparent replication for durability, availability
- Seeks to opportunistically and transparently exploit available resources while requiring minimal user attention
- Linear programming approach allows the system to adapt to changing conditions
System Goals

- On a set of intermittently connected personal devices:
- Opportunistically propagate information
- Replicate for availability and durability
  - Files should be k durable:
    \[
    \text{durability} = \min \{|d \in D : f \in \text{store-files}(d)| \}. f \in F
    \]
  - Files should be available on all devices where it might be useful
    \[
    \text{Availability} = \sum |\text{like-files}(d) \cap \text{store-files}(d)|. d \in D
    \]
Minimal User Interaction

- Add new device
- Report loss of device (optional)
- Restore data
- Low storage warning
- Device and type specific functionality through plug-ins
  - Archive data
  - Synchronization/reconciliation
Storage Devices

- Each storage device contains:
  - User files
  - Metadata
  - Replicas

- PodBase data is stored securely in device’s file system
Device Interaction

- Reconciles approximate global view
- Works even when devices are very small
AdapGate Replication

Actions

Copy A->B 50
Copy C->B 100
Delete ABC 3

Copy A->D 50
Copy A->C 100

Goals

Reconciled Metadata

LP Solver
Example: Automatic Sneakernets
Outline

• Motivation
• Problem Statement
• System: PodBase
• Evaluation
• Conclusion
Podbase: Implementation

• Implementation in Java
  – Small customization required per supported OS (Windows, OS X, Linux)

• Use off the shelf linear programming solver
Evaluation

• Performed both a controlled evaluation and two user studies
• Controlled evaluation validates basic functionality
• User study
  – 10 household deployed PodBase on a majority of the storage devices
  – System designed to unobtrusively provide availability and durability
User Study: Summary

• 10 Households, 25 devices, 30 days
Replication Results

- PodBase successfully replicates data without user attention
- Adaptive replication improves durability
Bandwidth Results

CDF

Time to Replicate Files (hours)

H. 1
H. 2
H. 4
H. 7
H. 9
Evaluation Summary

• PodBase transparently provides increased durability and availability for personal devices
  – Availability in the paper

• User study shows that the system is deployable and useful

• Adapts and takes advantage of free space and high bandwidth device connections
Conclusions

• Podbase: Automated storage management
  – Transparently increase the durability and availability of data
  – Uses free storage space and opportunistic connectivity

• Prototype evaluated in user study
Questions?