Learning from Sensor Network Data
Matthias Keller, Jan Beutel, Andreas Meier, Roman Lim, Lothar Thiele
Computer Engineering and Networks Lab
Swiss Federal Institute of Technology (ETH) Zurich

PermaSense System Model

Deployment Site
(1) Node generates five packets every 2 minutes
(2) Forwards update packet travel time
(3) Generates packet stream
(4) Packet streaming
(5) Packet log
Base station
IP link
Local Storage Space
- Backup of all packets
- Outgoing packet backlog

(g) Calculation of global timestamp
(f) Packet log
(e) Packet streaming
(d) Update packet
(c) Data collection
(b) Forwarders

Analysis of Matterhorn Deployment Data

Data Analysis Methodology

Assumptions from System Design
- Increasing packet sequence numbers with overroll
- Correct ordering by timestamps is not guaranteed

Step 1: Data Preparation
- Merging of historically emerged data representations
- Data source combination
- Data cleaning

Step 2: Data Downsampling and Pre-Analysis
- Received packets per day
- Computation of sequence number gaps

Step 3: Analysis
- Detailed examination of unexpected system behaviour

End-to-End Data Delivery Performance

Analyzed Data Set
- 14 million unique packets from 15 nodes

1st Generation: Jul 08 - Aug 08
- Initial troubles stopped data collection for several days
- No packet logs

2nd Generation: Aug 08 - Nov 08
- Nov 6-19: Outage of base station and backend
- No packet logs from Sep 09 to Nov 20

3rd Generation: Nov 08 - May 09
- Nodes equipped with SD cards
- Jan 17-19: Backend server outage

Detailed Analysis of 3rd Generation Data

(1) Data collection stopped due to problems with base station and/or backend
(2) Performance degradation with steadily increasing number of packets lost per day

Performance Degradation: What Happened After 3.5 Months of Flawless Operation?

Loss Pattern Reveals Node Reset Problem
- Five subsequent packets lost in 99%
- Loss of five subsequent packets refers to node reset
- Similar pattern for all nodes

Successful Error Analysis
- Recovery of affected nodes
- Analysis of memory dumps taken after last node reset

Node Reset Cause
- Increasing execution time of task that reads and writes from and to SD memory card
- Software forced reset due to task queue overrun

Data Collection Does Not Stop at the Access Node
Design Data Structures for Usability
Long-term Effects Are Hard to Test and Predict