KOREA UNIVERSITY

mCerebrum: A Mobile Sensing Software Platform for Development and Validation of Digital Biomarkers and Interventions

Syed Monowar Hossain,
Timothy Hnat, Nazir Saleheen,
Nusrat Jahan Nasrin

Joseph Noor, Bo-Jhang Ho,
Tyson Condie, Mani Srivastava

Santosh Kumar,
University of Memphis

Sensor Network | 컴퓨터공학과 / **박휘재**

INDEX

mCerebrum: A Mobile Sensing Software Platform for Development and Validation of Digital Biomarkers and Interventions

1. Introduction
2. Related Work
3. Overview of mCerebrum
4. Sense
5. Analyze
6. Act
7. Energy Estimation & Management
8. Lessons Learned
9. Limitation and Future work
10. Conclusion

Introduction

- 수 많은 앱을 통해서 건강 연구에 혁신을 가져왔다.
- 건강, wellness, 일상의 행동 등 다양한 요인을 연구하기 위해 새로운 바이오 마커에 의존도가 높아지고 있다.
- System Requirements Base
  - 이전의 기술로 발견할 수 없었던 데이터를 발견할 수 있도록 하기 위해 raw data를 수집해야 함
  - 처리 방법에 따라 새로운 bio-marker를 얻을 수 있음
    - 자이로 스코프를 이용한 손동작을 추적해 흡연 여부, 식사, 음주 등을 측정 가능
  - 누적되는 데이터 처리를 위해 휴대폰 어플리케이션 플랫폼 이용해야 함

Introduction

- System Requirements
  1. 새로운 센서를 플러그인 가능한 웨어러블 센서의 동시 연결 지원
  2. 데이터의 손실 없이 대량의 데이터 수집 가능
  3. 빠르게 증가하는 센서 데이터를 안정적으로 저장
  4. 수신 데이터를 신속하게 분석하고 batching 오류를 신속하게 수정
  5. 스트리밍 센서 데이터에 대한 감지 분석 파이프라인
  6. 멀티 센서 바이오 마커를 계산할 수 있도록 여러 센서가 스트리밍 데이터를 공유
  7. 어플리케이션은 범용적이어야 하고 확장 가능해야 함
  8. 휴대폰의 계산 용량이나 배터리 수명에 영향을 주지 않으며 동시 연산 지원하도록 구조적 확장 가능
  9. 실험 진행 시 참가자의 방해가 되지 않아야 함

Introduction

- Our Contributions
  - mCerebrum
    - 앞의 9가지 요구사항을 충족 / 오픈소스 / 일반화, 재사용 가능한 플랫폼

    - 하루 7000만개 이상의 샘플 데이터 수집, 실시간 처리를 위해 데이터 교환 아키텍쳐(DataKit) 사용
    - 복잡한 연산 부담 줄이기 위해 연산 재사용, micro-batching 지원
    - Storage Architecture : **Pebbles**
      - 저장된 데이터와 빠르게 증가하는 데이터(2GB/day)에 대해 유연
      - 통신에 효율적인 공통 데이터 이용
      - 빠르게 증가하는 센서 데이터 저장과 관련된 대기 시간 및 오버헤드를 줄임

Related Work

- mCerebrum의 범주
  1. 웨어러블 제품의 공급 업체가 만든 **데이터 수집 시스템** (Fitbit, Withings, Polar and Garmin)
     - 일반적으로 모바일 어플리케이션 제공
     - 다른 어플리케이션 및 서비스에 대한 클라우드 서비스 시각화(Dashboard)
     - SDK를 통해 튜닝이 가능하지만 타 제품에 노출되는 서비스는 없음
  2. 소프트웨어 플랫폼
     - 소프트웨어 생태계 주요 플레이어 (e.g. Apple, Google, MS)가 제공하는 풍부한 서비스
     - 광범위한 건강 관련 센서 데이터 수집
     - 수집 속도에 있어서는 약간의 차이가 있음
  - 다른 유사 시스템이 있으나 데이터 품질, bio-marker 계산, 개인정보 관리, 알림 등 기능 부족
  - 대부분 고밀도 센서를 구성하지만 저속 수집이고 스마트폰의 로컬 스토리지를 이용함

    - mCerebrum ▶ 고성능 데이터 수집, 포괄적인 기능 및 센서 지원 세트 제공

Related Work

- 기존 소프트웨어 플랫폼과 비교
  - GoogleFit
    - ResearchKit을 Android용으로 포팅한 시스템
    - 데이터 품질 평가, 바이오 마커 계산,개인정보 관리 지원 없는 데이터 수집
  - HealthKit
    - 광범위한 건강 관련 센서 데이터 수집
    - 스마트폰 기반의 저장 엔진 제공
    - 저속 데이터를 대상으로 함
    - 바이오 마커 계산, 데이터 품질 평가, 개인정보 관리 등 지원 못함

Framework	Sensor-support	High Throughput	Scalable Storage	Data Quality	Computation Sharing	Data Sharing	Sense-Analyze -Act Support	Burden Aware
GoogleFit	Internal, Bluetooth				✓	✓	✓	
HealthKit	Internal, Bluetooth				✓	✓	✓	
OpenDataKit	Bluetooth, USB	✓						
AWARE	Internal				✓	✓	✓	
mCerebrum	Internal, Bluetooth, Bluetooth-LE, ANT+	✓	✓	✓	✓	✓	✓	✓

Table 1: Comparison of different extensible mHealth frameworks that support external sensors

Related Work

- 기존 소프트웨어 플랫폼과 비교
  - Open DataKit(ODK)
    - Bluetooth, USB 기반 웨어러블 센서에서 고속 센서 데이터 수집 지원 초기 플랫폼
    - 드라이버 서비스를 플러그인으로 추가해 외부 장치 데이터를 바이너리 스트림 수집
    - 응용 프로그램에서 사용할 수 있도록 인코딩
  - AWARE
    - 휴대 전화에 부착된 센서(자이로스코프, 가속도), 휴대 전화를 활용한 활동 (캘린더, 통화, 메시지) 및 설문지 등을 통해 데이터 수집
    - 외부 센서 플러그인 지원하지만 자체 데이터 저장소를 공유, 관리해야 함
      - 고속 센서 데이터, 전체 처리, CPU 사용에 대한 Storage 확장성 제한
  - 이 외에도 Commcare, Paco 등 여러 서비스 언급

Framework	Sensor-support	High Throughput	Scalable Storage	Data Quality	Computation Sharing	Data Sharing	Sense-Analyze -Act Support	Burden Aware
GoogleFit	Internal, Bluetooth				✓	✓	✓	
HealthKit	Internal, Bluetooth				✓	✓	✓	
OpenDataKit	Bluetooth, USB	✓						
AWARE	Internal				✓	✓	✓	
mCerebrum	Internal, Bluetooth, Bluetooth-LE, ANT+	✓	✓	✓	✓	✓	✓	✓

Table 1: Comparison of different extensible mHealth frameworks that support external sensors

Related Work

- 기존 소프트웨어 플랫폼과 비교
  - mCerebrum
    - 효율적인 애플리케이션 간 통신
    - 확장 가능한 스토리지 아키텍쳐
    - 중앙 집중화 된 스토리지
    - 고속 데이터 수집
    - 실시간 생체 데이터 계산
    - 사용자 프롬프트, 개인 정보 관리
    - 실제 연구 진행을 통한 사용자 피드백 반영

Framework	Sensor-support	High Throughput	Scalable Storage	Data Quality	Computation Sharing	Data Sharing	Sense-Analyze -Act Support	Burden Aware
GoogleFit	Internal, Bluetooth				✓	✓	✓	
HealthKit	Internal, Bluetooth				✓	✓	✓	
OpenDataKit	Bluetooth, USB	✓						
AWARE	Internal				✓	✓	✓	
mCerebrum	Internal, Bluetooth, Bluetooth-LE, ANT+	✓	✓	✓	✓	✓	✓	✓

Table 1: Comparison of different extensible mHealth frameworks that support external sensors

Overview

- mCerebrum은 독립적으로도 동작하고, 일부만 동작하기도 함
- 이 작업은 휴대 전화에서 동작하는데 초점을 둠

- Architecture
  - 구성 요소 및 통신 설계
- Real-Life deployments
  - 실제 현실에 배포
- Key feature
  - mCerebrum의 주요 기능

Overview

- Architecture
  - 고속 데이터 스트리밍 데이터 수집 목표 유연한 계층 구조 구축
    - 데이터의 로깅
    - 실시간 처리 및 개입

1. 데이터 수집 (센서 데이터, 휴대폰 데이터)
2. 무선 송수신 기를 통해 수집
3. 데이터 품질 조정
4. 개인정보 조정
5. 기계 학습 모델
6. 바이오 마커 생성
7. 사용자/수신자를 위한 리포트 생성 후 장치 저장소에 저장
8. 저장소 관리를 위한 클라우드 공유
9. 연구진에 전달

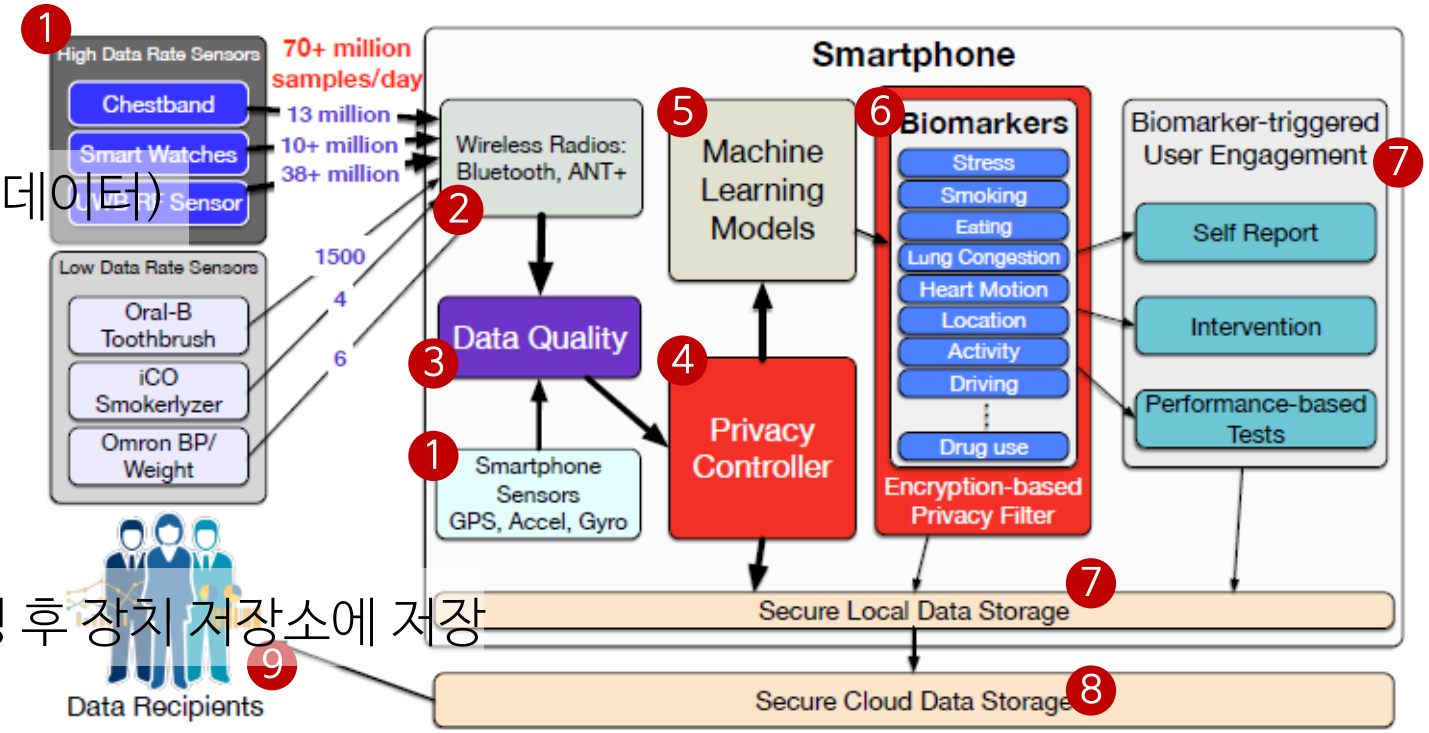


Figure 1: mCerebrum overview illustrating key components such as ingestion of high-frequency sensor data, real-time data quality assessment and biomarker computation.

KOREA UNIVERSITY

Overview

- Architecture
  - 통신 인터페이스, 스마트폰 센서, 웨어러블 센서 지원
  - mCerebrum 플랫폼 간의 인터페이스를 제공하는 데이터 소스
  - 영구 데이터 저장 및 중간 결과 라우팅 인터페이스
  - 장기 실행 응용 프로그램이 실행되기 위한 데이터 수신, 처리 지원
  - 참여자 인터페이스 및 모든 기능 구현

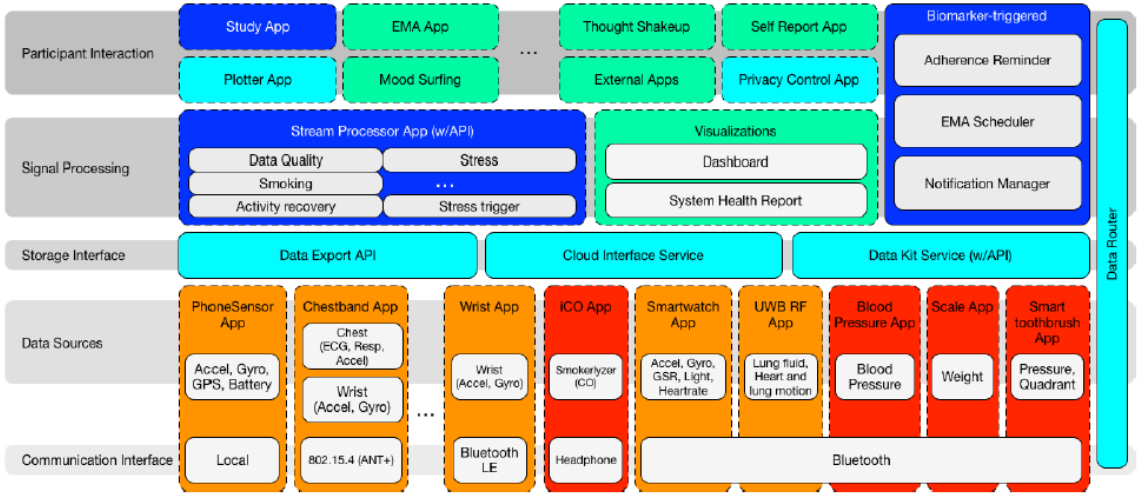


Figure 2: mCerebrum's architecture consists of 5 layers: (1) Communication, (2) Data Sources, (3) Storage, (4) Signal Processing, and (4) Participant Interaction, all connected through a data router. The colors indicate different categorizations of applications — red for high-rate and orange for lower-rate sensors, blue indicates core components of mCerebrum, cyan represents system services, and green represents user-centric applications.

Application	Description
DataKit	Handles routing, privacy, and storage
DataKitAPI	API library for apps to use DataKit
Plotter	Real-time data visualizer
Privacy Controller	Allows the participant to suspend data collection and EMA prompting
Utilities	Common helper functions
Phone	Integrates the smartphone sensors
Chestband	Data collection from ANT+ sensor suite
Wrist	BLE wrist-worn motion capture device
iCO	Carbon Monoxide sensor support
Smartwatch	Bluetooth 4 connected watch
UWB RF	BLE chest sensor for measuring heart function and lung fluid
Blood Pressure	BLE-connected blood pressure cuff
Weight	BLE-connected weight scale
Smart Toothbrush	BLE-connected smart toothbrush
Stream Processor	Provides real-time computation of biomarkers (e.g. stress, smoking, etc.)
Mood Surfing	A custom built stress reduction app
Thought Shakeup	A custom built stress reduction app
Medication	Medication adherence compliance app and reminder system
Self Report	Customizable self-report prompts
EMA	Customizable EMA delivery application
Study	Main study interface; provides application management for all other apps
EMA/EMI Scheduler	Customizable scheduler for delivering user prompts based on biomarkers
Adherence Reminder	A scheduler for episodic data collection
Notification Manager	Gatekeeper for all user prompts

Table 2: Overview of mCerebrum apps and libraries

github.com/MD2Korg

Overview

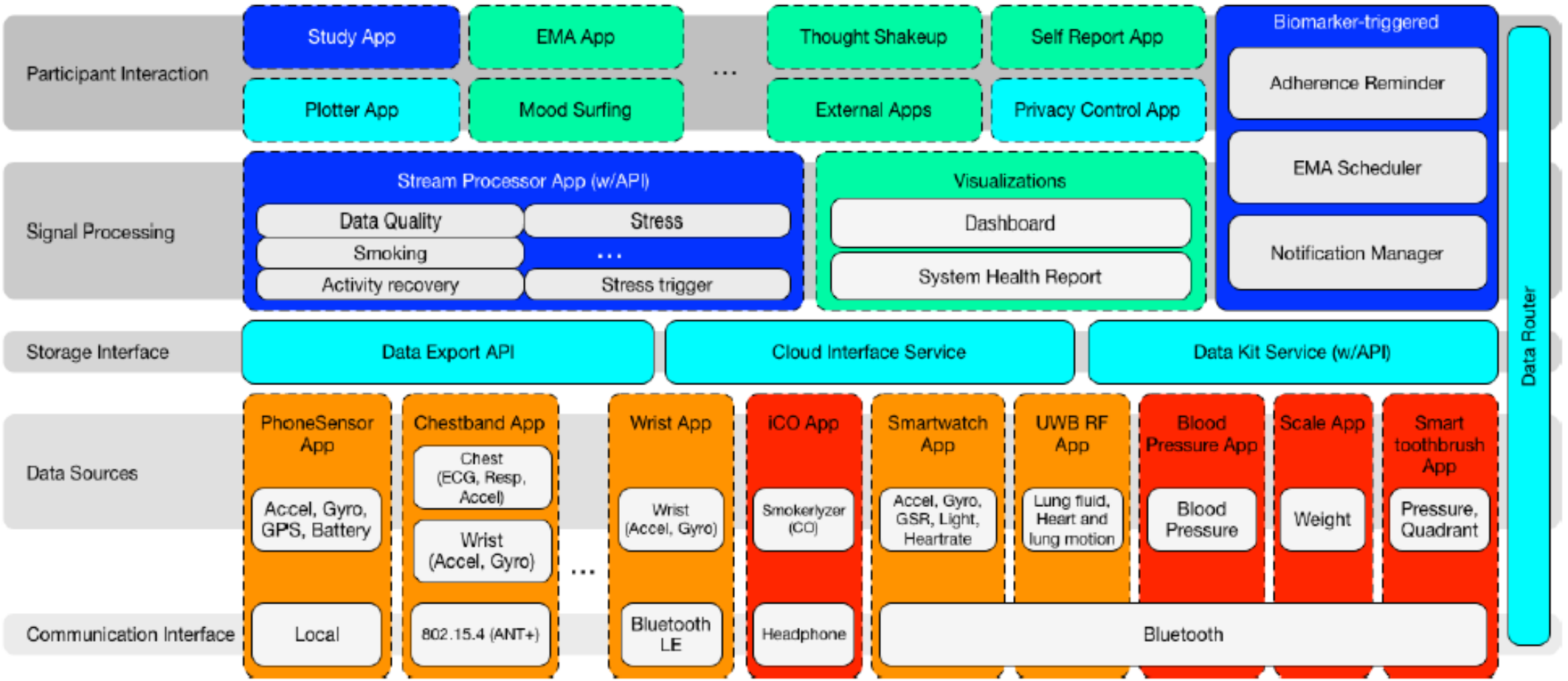


Figure 2: mCerebrum's architecture consists of 5 layers: (1) Communication, (2) Data Sources, (3) Storage, (4) Signal Processing, and (4) Participant Interaction, all connected through a data router. The colors indicate different categorizations of applications — red for high-rate and orange for lower-rate sensors, blue indicates core components of mCerebrum, cyan represents system services, and green represents user-centric applications.

Overview

- Real-Life deployments
  - 독립적인 6개의 연장 연구를 진행하는 5년간의 경험을 통해 설계됨
  - 현재 10개의 서로 다른 현장에 배치 될 준비 중
- 이 연구는 2,251명의 특수 참가자와 106,806명의 일반 참가자의 high-frequency 센서 데이터를 통해 실시
- 순수 생성, 처리, 저장되는 데이터는 100TB 이상
  - 4조 7천 억 개의 데이터 포인트 추정

ID	Study	Users	Person-Days	Samples (Billions)
1	Smoking & Eating	225	3,150	136
2	Smoking	300	4,200	182
3	Smoking	300	4,200	182
4	Smoking & fMRI	90	1,260	55
5	Smoking & Stress	24	336	15
6	Heart Failure	225	6,750	224
7	Oral Health	162	29,160	968
8	Cocaine Use	25	350	18
9	Behavior Change	100	1400	58
10	Job Performance	800	56,000	2,891
	Total	2,251	106,806	4,729

Table 3: mCerebrum powered studies span 10 sites – Northwestern, Rice, Utah, Vermont, Moffitt, Ohio State, UCLA, Johns Hopkins, Dartmouth, and Minnesota.

Overview

- Real-Life deployments
- Smoking Result (1-5)
  - 담배를 끊기 전 4일, 담배를 끊은 후 10일간 센서 착용
  - 스트레스 및 흡연의 감지에 대한 응답으로 EMA(Ecological Momentary Assessment) 프롬프트 생성 포함
  - 참여자가 직접 표시한 하루의 시작과 종료 사이 4시간마다 전달되어야 함
  - EMA는 사용자가 운전이나 운동같은 활동을 하지 않을 때 EMA가 생성
- Heart failure study (6)
  - 병원의 재 입원 값을 예측하는 마커를 찾는 것
  - 퇴원한 환자의 30일 활동, 식사, 약물 투여, 혈압 등

ID	Study	Users	Person-Days	Samples (Billions)
1	Smoking & Eating	225	3,150	136
2	Smoking	300	4,200	182
3	Smoking	300	4,200	182
4	Smoking & fMRI	90	1,260	55
5	Smoking & Stress	24	336	15
6	Heart Failure	225	6,750	224
7	Oral Health	162	29,160	968
8	Cocaine Use	25	350	18
9	Behavior Change	100	1400	58
10	Job Performance	800	56,000	2,891
	Total	**2,251**	**106,806**	**4,729**

Table 3: mCerebrum powered studies span 10 sites – Northwestern, Rice, Utah, Vermont, Moffitt, Ohio State, UCLA, Johns Hopkins, Dartmouth, and Minnesota.

Overview

- Real-Life deployments
- Oral health study (7)
  - 프라그 수준의 변화를 예측하는 센서 마커
  - 6개월간 칫솔질이나 치실을 추적
- Cocaine study (8)
  - 코카인 사용 시기 감지
  - 활동적인 코카인 사용자는 2주간 센서를 착용

ID	Study	Users	Person-Days	Samples (Billions)
1	Smoking & Eating	225	3,150	136
2	Smoking	300	4,200	182
3	Smoking	300	4,200	182
4	Smoking & fMRI	90	1,260	55
5	Smoking & Stress	24	336	15
6	Heart Failure	225	6,750	224
7	Oral Health	162	29,160	968
8	Cocaine Use	25	350	18
9	Behavior Change	100	1400	58
10	Job Performance	800	56,000	2,891
	Total	2,251	106,806	4,729

Table 3: mCerebrum powered studies span 10 sites – Northwestern, Rice, Utah, Vermont, Moffitt, Ohio State, UCLA, Johns Hopkins, Dartmouth, and Minnesota.

Overview

- Real-Life deployments
- Behavior change study (9)
  - 14일간 흡연, 비만 참가자의 행동 개선을 위한 스트레스 중재 효과 연구
- Job performance study(10)
  - 10주간 센서 착용을 통해 직무 수행 지표를 발견

- 두 손목에 센서와 스마트폰 센서의 지속적인 데이터 수집
- 모든 수행은 가슴 밴드를 통해 데이터 수집 포함
- 스트레스 및 흡연 감지에 대한 응답
  - EMA 생성에 응용

- IRB 승인을 받아 수행



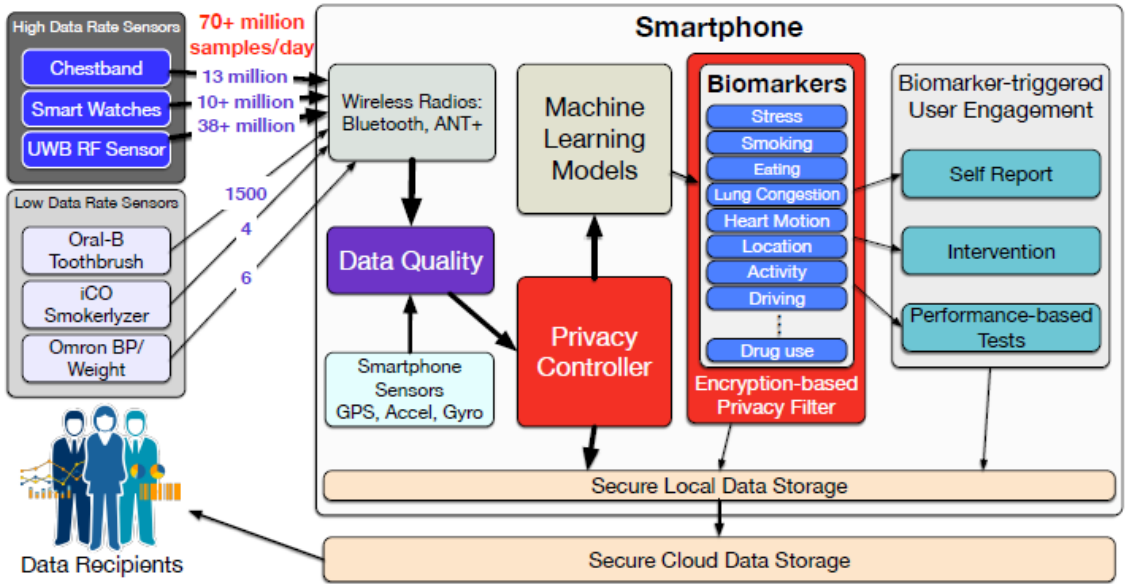
ID	Study	Users	Person-Days	Samples (Billions)
1	Smoking & Eating	225	3,150	136
2	Smoking	300	4,200	182
3	Smoking	300	4,200	182
4	Smoking & fMRI	90	1,260	55
5	Smoking & Stress	24	336	15
6	Heart Failure	225	6,750	224
7	Oral Health	162	29,160	968
8	Cocaine Use	25	350	18
9	Behavior Change	100	1400	58
10	Job Performance	800	56,000	2,891
	Total	2,251	106,806	4,729

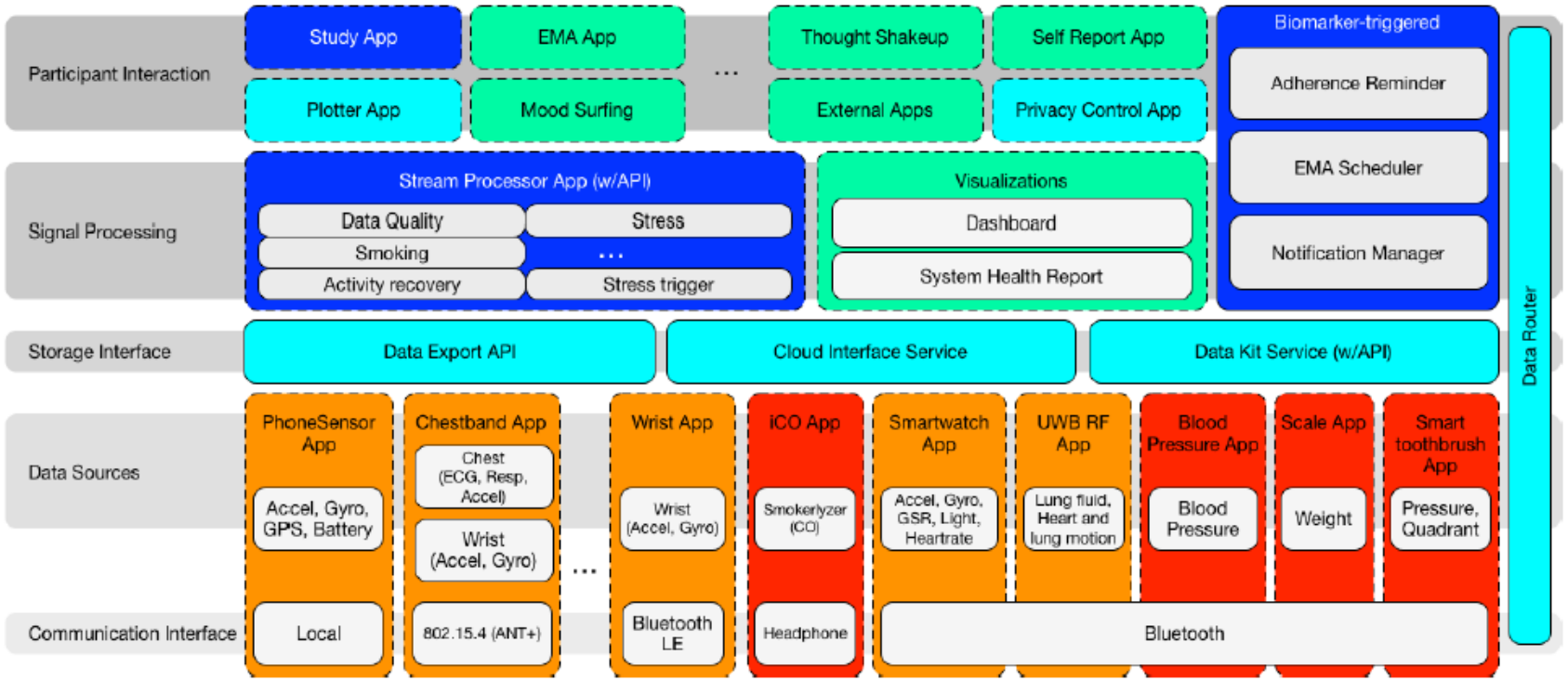
Table 3: mCerebrum powered studies span 10 sites – Northwestern, Rice, Utah, Vermont, Moffitt, Ohio State, UCLA, Johns Hopkins, Dartmouth, and Minnesota.

Overview

- Key features of mCerebrum

1. **다양한** mHealth **바이오 마커** 및 **센서 트리거**를 지원하는 **범용 플랫폼**으로 설계
2. 여러 센서의 **고속 스트리밍 데이터**에 대한 감지, 분석 작업의 **전체 파이프라인 통합**
3. **확장 가능한 스토리지**와 함께 데이터 품질 및 다중 센서를 이용한 바이오 마커의 **동시 실시간 계산**을 지원


High Data Rate Sensors
Chestband
Smart Watches
UWB RF Sensor
70+ million samples/day
13 million
10+ million
38+ million
Low Data Rate Sensors
Oral-B Toothbrush
iCO Smokerlyzer
Omron BP/ Weight
1500
4
6
Smartphone
Wireless Radios: Bluetooth, ANT+
Machine Learning Models
Biomarkers
Stress
Smoking
Eating
Lung Congestion
Heart Motion
Location
Activity
Driving
Drug use
Encryption-based Privacy Filter
Biomarker-triggered User Engagement
Self Report
Intervention
Performance-based Tests
Data Quality
Smartphone Sensors GPS, Accel, Gyro
Privacy Controller
Secure Local Data Storage
Secure Cloud Data Storage
Data Recipients


Participant Interaction
Study App
EMA App
Thought Shakeup
Self Report App
Plotter App
Mood Surfing
External Apps
Privacy Control App
Biomarker-triggered
Adherence Reminder
EMA Scheduler
Notification Manager
Signal Processing
Stream Processor App (w/API)
Data Quality
Stress
Smoking
Activity recovery
Stress trigger
Visualizations
Dashboard
System Health Report
Storage Interface
Data Export API
Cloud Interface Service
Data Kit Service (w/API)
Data Router
Data Sources
PhoneSensor App
Accel, Gyro, GPS, Battery
Chestband App
Chest (ECG, Resp, Accel)
Wrist (Accel, Gyro)
Wrist App
Wrist (Accel, Gyro)
iCO App
Smokerlyzer (CO)
Smartwatch App
Accel, Gyro, GSR, Light, Heartrate
UWB RF App
Lung fluid, Heart and lung motion
Blood Pressure App
Blood Pressure
Scale App
Weight
Smart toothbrush App
Pressure, Quadrant
Communication Interface
Local
802.15.4 (ANT+)
Bluetooth LE
Headphone
Bluetooth

Sense – resilient data collection, sharing, and storage

- Sensing layer
  - 여러 센서의 스트리밍 데이터를 안정적으로 수집, 저장 및 공유하는 역할

  - Major challenge
  - 현재와 미래의 데이터 유형 및 관련 메타 데이터를 수용할 수 있는 유연한 표현 제공
    - 무선 장치를 통해 연결된 여러 센서의 수신 데이터 속도를 처리할 수 있는 높은 처리량
  - 높은 처리량을 유지하면서 여러 소스 및 수신자간에 들어오는 데이터를 효율적으로 공유
  - 빠르게 증가하는 데이터에 직면하여 쿼리 응답을 유지하는 스토리지 지원

1. 빠르게 증가하는 데이터를 처리하는 DataKit
2. 시스템 내에서 수집되고 저장되는 데이터의 양을 최대화 하는데 필요한 기능을 다루는 확장 가능한 **스토리지 설계**

Sense – resilient data collection, sharing, and storage

- DataKit
  - 여러 동시 소스로부터 데이터를 빠르게 수집하고, 데이터 소스와 싱크 어플리케이션간 데이터를 효율적으로 다 - 대 - 다 공유 할 수 있도록 설계

  - mCerebrum의 데이터 모델은 2개의 개념을 기반으로 한다

1. Data point : Tuple consisting of a timestamp and value
   - 모든 배열 변형 중 하나로 구성될 수 있음
   - 대부분의 데이터를 기본 유형으로 제한함으로써 JSON 인코딩을 통해 복잡한 데이터 유형 허용
2. Data stream : Uniquely identifiable collection of data points

  - Flexible and efficient communication
    - DataKitAPI를 통해 통신 메커니즘 제공
    - Publish-Subscribe mechanism

Sense – resilient data collection, sharing, and storage

- DataKit
  - Flexible and efficient communication – DataKitAPI를 통해 통신 메커니즘 제공
    - Publish – Subscribe 메커니즘
      - 비동기 메시징 패러다임
      - 수신자가 지정되어 있지 않고 정해진 범주에 따라 구독을 신청한 수신자에게 전달
    - 구독자 – Callback, subscribe, unsubscribe 메커니즘이 존재
    - 발행자 – Registration, unregistration 방법이 존재

➢ 코드 내부 **복잡성 줄어들어** 효율적인 데이터 처리 및 라우팅 보장

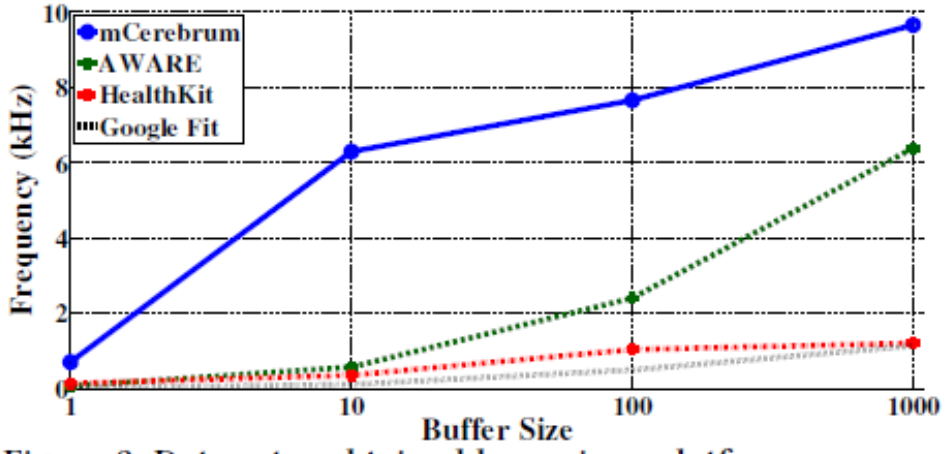
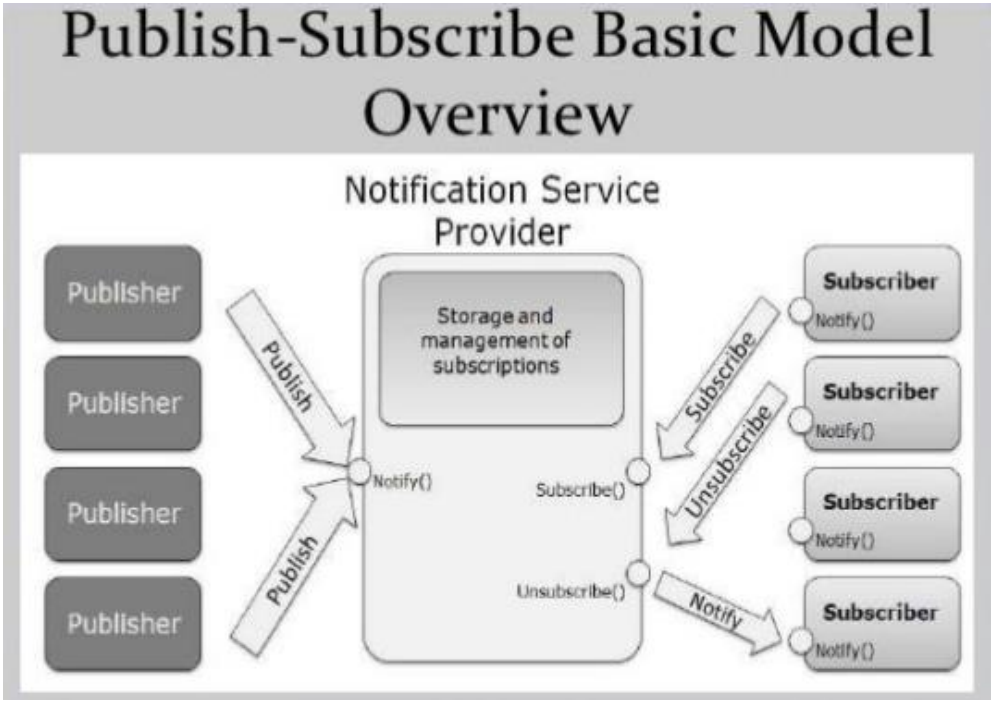


Figure 3: Data rates obtained by various platforms as number of samples (with a common timestamp) are written.



Sense – resilient data collection, sharing, and storage

- DataKit
  - Handling Data Representation Diversity
    - 다양한 외부 및 내부 센서 지원
    - 무선 기술 및 API 구현에 따라 데이터가 single time stamp와 관련된 블록에 도달하거나 샘플에 개별적으로 time stamp를 적용
      - Time stamp란 둘 이상의 장치에 시간을 비교하거나 계산할 때 쓰임
      - 저장된 파일이 생성되거나 변경된 시간을 나타내기도 함
  - Resilient Communication Management
    - 일괄 처리나 스트리밍 모드로 작동
    - 스트리밍 모드는 장치를 잠시 분리해도 데이터 손실이 생길 수 있어 연결의 유지가 필요
      - 흡연 측정 알고리즘은 5초의 손목 운동 데이터를 탐지하여 바이오 마커를 생성하는데 통신이 실패하면 누락될 수 있음
      - Back-off 메커니즘
      - 데이터 손실 최소화 지침을 사용자에게 제공

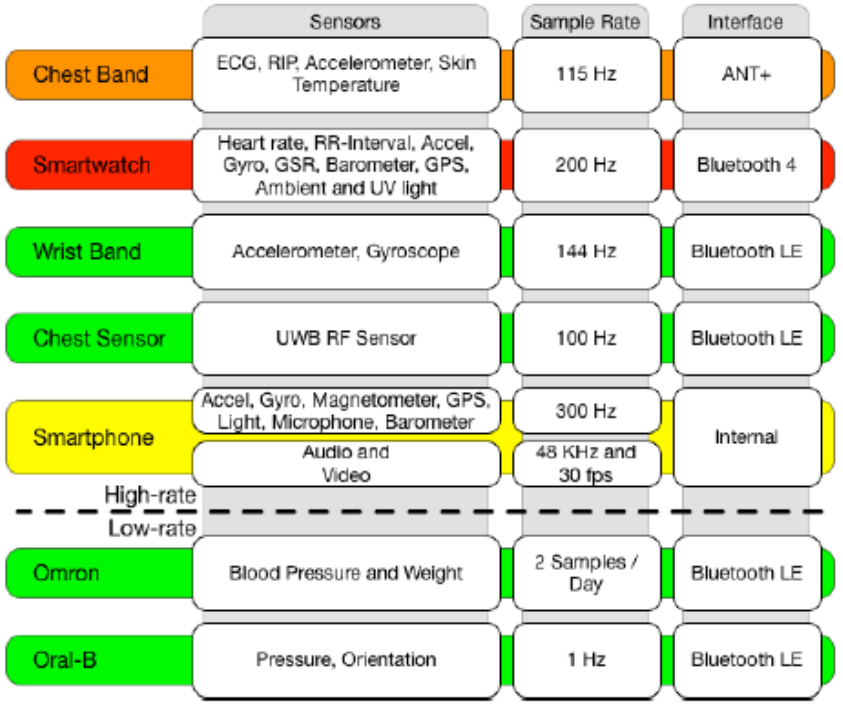


Figure 4: mCerebrum supports sensors ranging from 2 samples/day to 300 Hz per device including: BLE (green), Bluetooth 4.0 (red), ANT+ (orange), and internal (yellow). Additionally, it support short audio and video clips with a high data rate storage mechanism.

Sense – resilient data collection, sharing, and storage

- DataKit
  - Handling Large Data Objects
    1. 데이터를 큰 덩어리로 분할하고 개별적으로 Datakit으로 보냄
       - TCP 통신과 비슷하게 덩어리가 재조합됨
    2. 어플리케이션과 DataKit간의 보안 파일 공유 접근 방식 사용
       - FileProvider를 통해 [contents://Uri]를 생성해 데이터 공유
       - Contents://는 어플리케이션 개발에 이용되는 객체나 리스트 전달방식
       - 대용량 데이터 객체 처리 기본 매커니즘이 됨

Sense - resilient data collection, sharing, and storage

- Scalable Storage of High-rate Sensor Data
  - SQLite는 Android나 iOS같은 모바일 장치의 실제 데이터 스토어
  - High-frequency raw sensor data stream을 저장하는데 적합하지 않음

  - RocksDB와 같이 개발중인 로그 구조의 스토리지 시스템
    - RDBMS (Relational Database Management System) - 관계형 데이터 베이스 관리 시스템
  - Pebbles
    - 사용자 지정 log-structured storage layer
    - 어플리케이션이 데이터 보관을 위해 클라우드로 데이터를 로드할 수 있음
    - 데이터 스트림을 식별하는데 timestamp와 각 항목에 대한 경량 인덱스를 유지

Sense – resilient data collection, sharing, and storage

- Scalable Storage of High-rate Sensor Data
  - Pebbles와 다양한 SQLite, SQLite Cluster(AWARE에서 사용) 를 이용한 최대 처리량 보여줌
  - mCerebrum과 같은 raw data writing size에서는 Pebbles가 최적화 되어있음
  - 데이터의 크기가 작아도 CPU에서 연산량이 따라주지 못해 병목현상 발생
    - 그럼에도 불구하고 Pebbles는 충분히 큰 연산 처리량 보여줌
  - 다른 storage 방식은 sequential write 수행 능력이 떨어져 성능 저하가 발생

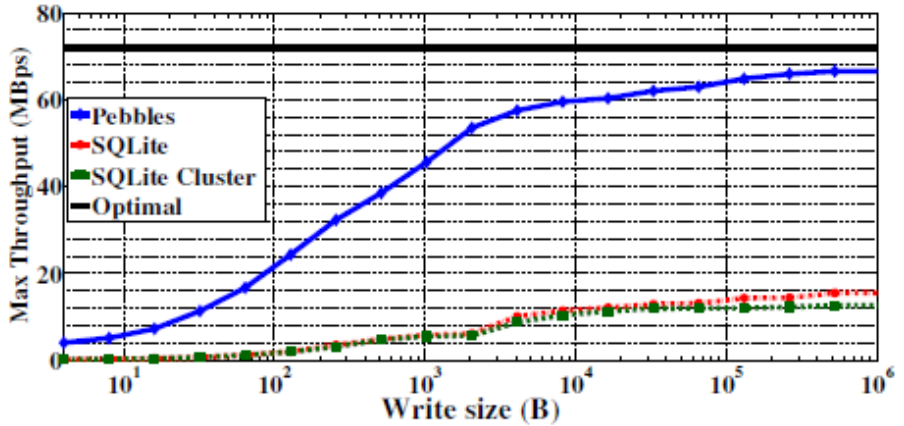


Figure 5: Maximum write throughput with increasing write size. Pebbles achieves 92% of the optimal write throughput while SQLite and SQLite cluster (used in AWARE [7]) achieve 22% and 18% respectively at their steady states while writing large blocks of data.

Analyze - concurrent computation of multi-sensor biomarkers

- 수집된 high-rate 센서 데이터를 처리하여 다른 앱에서 사용할 수 있는 기능 및 바이오 마커를 계산
- The Challenge
  - 수용 가능한 데이터의 질을 위해 데이터를 real-time으로 정제하고, 수백가지 기능을 계산한 뒤, 모든 바이오 마커의 기계학습 모델에 적용
  - 리소스가 제한된 휴대폰의 CPU나 메모리를 이용

1. 데이터와 계산을 재사용 할 수 있는 방법
2. 센서 데이터 분석을 지원하기 위해 시스템 과부화 관리 기술 탐색, 평가
3. mCerebrum 전체에서 실시간 계산 및 기능 바이오 마커 공유 구현

Analyze - concurrent computation of multi-sensor biomarkers

- Data and Computation Reuse
  - 통신효율성이 충분하지 않기 때문에 가능한 많은 데이터와 계산을 재사용해야 함
  - 플랫폼의 다양한 프로세싱 구성 요소가 **중간 결과** 또는 raw data에서 다시 계산하는 대신, 처리된 스트림을 활용할 수 있는 DataKit 버스에 있는 **바이오 마커를 계산할 때** 생김

  - Supporting Onboard Sense-Analyze-Act
    - 휴대폰에서 감지 분석 작업을 로컬로 활성화 하기 위해 **micro-batch / batch / on-demand** 데이터 처리 지원
    - 데이터 수신 즉시 처리해야 하여 micro-batch latency = 1초
    - 스트레스나 흡연 바이오 마커 계산을 위해 60초의 batch 대기 시간 사용

Analyze - concurrent computation of multi-sensor biomarkers

- Handling System Overload
  - Android는 사용자에게 제공되는 품질을 자동으로 관리
  - 시스템 리소스 부족하면 선택적으로 애플리케이션 제거하고 메모리에서도 제거

  - 응용프로그램 kill을 방지하기 위해 세가지 별도 메커니즘 채택
    1. 중요한 어플리케이션 서비스는 foreground process로 선언
    2. mCerebrum 커널이 관련 어플리케이션의 감시 시스템처럼 작동
    3. 내부 로컬 메모리에 영구 복사본을 유지하고 어플리케이션 재 부팅시 시작되어야 함

Analyze - concurrent computation of multi-sensor biomarkers

- Handling System Overload
  - Micro-batching to Control Communication Load
    - 실시간으로 데이터 공유하고 처리하면 시스템과 통신 부하 증가됨
    - 시스템이 처리할 수 있는 데이터 frequency 및 이와 관련된 CPU 사용량에 대한 micro-batching 처리 대기시간의 영향을 보여줌

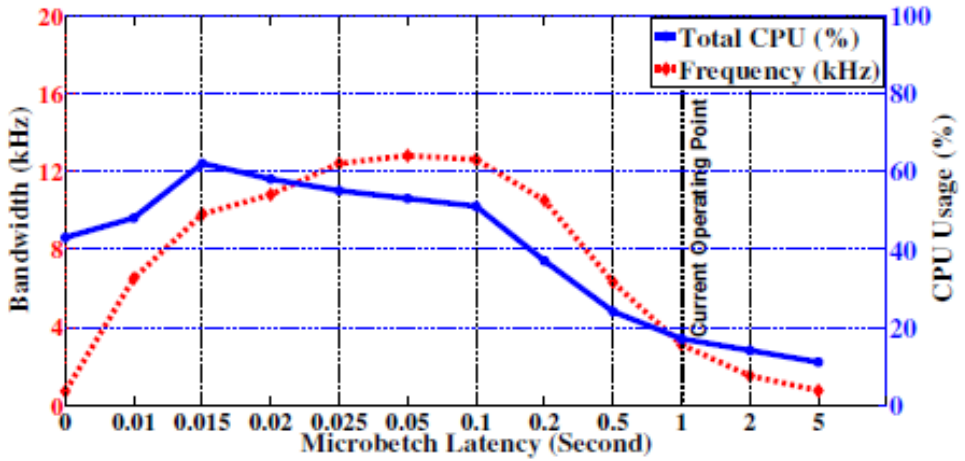


Figure 6: The effect of micro-batch latency on DataKit's communication bandwidth and CPU usage. We observe that with no latency, communication bandwidth is limited by bandwidth limits of IPC, while at higher latency, bandwidth is limited by the buffer size limits of IPC.

Analyze - concurrent computation of multi-sensor biomarkers

- Handling System Overload
  - Micro-batching to Control Communication Load
    - mCerebrum과 AWARE의 프레임 워크 비교
    - 센서 샘플링 속도를 1에서 150Hz로 변경
    - 스마트폰 내장 가속도계, 자이로스코프, 자력계 센서만 활용
    - 비교 결과 CPU 사용량이 8.4배 낮음을 확인 - micro-batching과 블록 저장 기능 결합으로 인한 이점

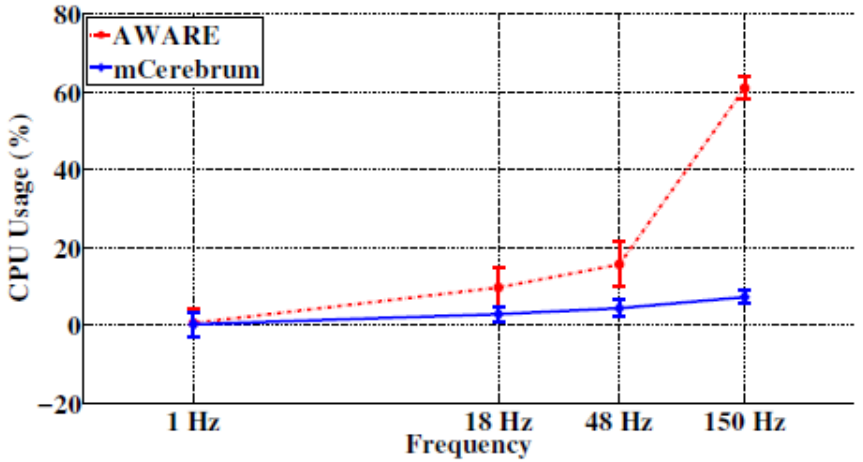


Figure 7: mCerebrum is compared with the AWARE framework by adjusting the total sample rate of the accelerometer, gyroscope, and magnetometers between 18 and 150 hertz per sensor. mCerebrum's benefits by up to 8.4 times lower CPU load when recording at 150 hertz)

Analyze - concurrent computation of multi-sensor biomarkers

- Handling System Overload
  - Effect of Buffer Size on System Load

	Iteration time			
	30s	60s	120s	300s
Computation time (s/min)	2.67	2.48	2.33	2.23
Memory load (MB)	64.26	73.71	80.49	100.38

Table 4: CPU time (normalized to 60 seconds) decreases as the buffer size is increased from 30 to 300 seconds. 0.34 seconds (17%) can be saved through buffering; however, it comes with an increase in memory load (156%).

- 버퍼링 데이터와 시스템의 계산 및 메모리 로드간 균형
- 바이오 마커 계산 파이프 라인 실행 후 30초~300초 사이 버퍼링 데이터 제공
- [adb shell dumpsys meminfo] 2Hz로 실행
  - Android의 ram 사용량을 기록하는 명령어
- 각 Iteration time 별로 20분씩 실행

- 시스템 과부화를 제한하며 수용 가능한 대기 시간을 제공하는 1분 동작지점 사용
- 하드웨어 프로필 향상으로 인한 작동지점 변경

Analyze - concurrent computation of multi-sensor biomarkers

- Stream Processor: Real-time Computation and Sharing of Feature and Biomarkers
  - 스트림 프로세서 모듈은 여러 윈도우 기반 신호 처리 파이프라인에 적합한 버퍼, 추정기 제공
  - Stream Processor의 자원 사용을 제한하는 여러 설계상 절충점
    1. 스마트폰 CPU가 더 나은 에너지 효율을 위해 버스트 모드로 작동하게 함
       - 데이터 재 처리 양을 제한하기 위해 매 60초마다 데이터 수신하는 batch 매커니즘으로 처리
    2. 개발자가 명시적으로 기록 보존을 구성하지 않는 항 현현재의 계산에 대해 RAM에 데이터 보관됨
    3. 알고리즘은 센서 소스를 공유해 계산을 재사용할 수 있기 때문에 일반적으로 파이프 라인으로 구현
       - i.e.심전도와 호흡을 통해 생긴 바이오 마커는 스트레스 / 손의 움직임과 특정 호흡을 통해 생긴 바이오 마커는 흡연

Analyze - concurrent computation of multi-sensor biomarkers

- Stream Processor: Real-time Computation and Sharing of Feature and Biomarkers
  - 본 파이프라인 모델은 SVM(Support vector machine)기반
    - 기계학습 분야중 하나로 패턴 인식, 자료 분석을 위한 지도 학습 모델
    - 두 카페고리 중 하나에 속한 데이터 집합이 주어졌을 때 두 공간을 경계로 표현
  - 신호 처리 알고리즘 실행에 평균 14.87 ± 4.12초 소요
  - 실시간 로컬 계산을 할 때 수명의 변화가 큼
    - (?) 논문에서 말하는 13%가 무엇을 의미하는지 모르겠음
  - CPU 로드 제한을 위한 Future work 필요

Module Set	Current	Lifetime
Baseline Power	17.13 mA	151.78 hrs
+ Wrist Sensor 1	96.28 mA	27.00 hrs
+ Wrist Sensor 2	107.16 mA	24.26 hrs
+ Biomarker Computation	151.05 mA	17.21 hrs
+ Decision Making	160.42 mA	16.21 hrs
+ EMA/EMI (25 minutes)	187.23 mA	13.89 hrs
+ Location Service	273.77 mA	9.50 hrs
+ Phone Accelerometer	284.33 mA	9.14 hrs
+ Phone Gyroscope	287.67 mA	9.04 hrs
+ Chest Sensor	327.46 mA	7.94 hrs

Table 5: Impacts on expected (2600 mAh) bettery life by incrementally adding sensing and processing capabilities.

Analyze - concurrent computation of multi-sensor biomarkers

- Quantifying the Benefits of Computation Reuse - A Case Study
  - Computation reuse의 효과 분석을 위해 어플리케이션을 두개로 나눔
    1. 흡연, 스트레스, 활동 및 식사 감지 어플리케이션
    2. 바이오 마커 계산 어플리케이션
  - 어플리케이션 두개 동시 실행하여 CPU 및 메모리 로드 정하고 컴퓨팅 공유 활성화
  - 바이오 마커 생성을 위해 겹치는 데이터는 재사용이 가능해 CPU 및 메모리 사용률이 낮아짐

➢ 이와 같이 재사용을 통해 계산 시간 27% 감축, 메모리 47% 감축



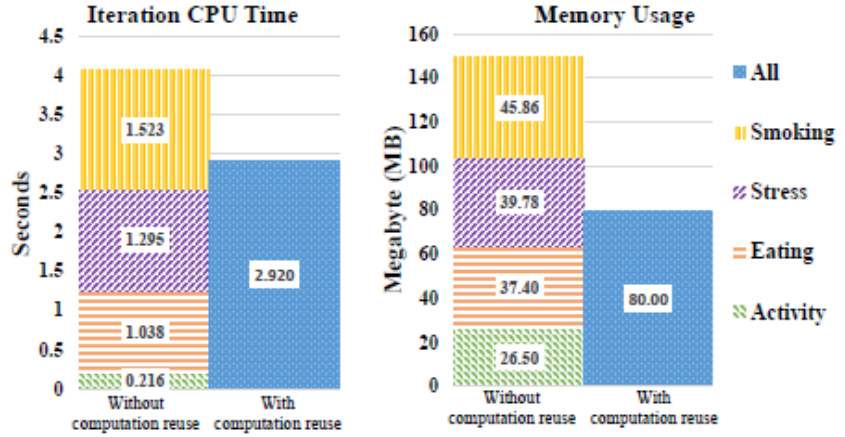


Figure 9: The effects of computation reuse on CPU and memory efficiency. The first columns show the CPU and memory usage when computing four biomarkers without sharing computation. A reduction of 27% CPU time and 47% memory is achieved through reuse, as shown in the second columns.

Act - burden-and context-aware interactions with participants

- 센서 데이터와 함께 참가자의 직업 입력도 연구 조사에 수집
- 참가자 상호작용 유형
  1. Voluntary - 즉각적인 상호작용을 통해 시스템 EMA, EMI 제공
  2. Prompt - 병과 관련된 센서 데이터 수집, 약물 복용 상기
  3. Glanceable - 데이터 상호작용을 GUI를 통해 업데이트

- The Challenge
  - 참가자 부담을 줄이기 위해 모든 출처에서 생성된 바이오 마커 조정
  - 자체보고서를 작성하기 전에 우수한 품질의 센서 데이터 사용하도록 함

- 연구 요구사항 설명
- 참가자 부담 및 상황 인식을 위한 관리자 디자인

Act - burden-and context-aware interactions with participants

- EMA/EMI Scheduling Requirements
  - 시간이나 보고된 바이오 마커에 따라 무작위로 유도될 수 있음
  - 참자가의 순간 데이터를 수집

1. EMA 스케쥴러가 포함된 연구는 다음을 포함할 수 있다:
   - 특정 시간에 진행되는 평가 i.e. 기상 직후, 숙면 직전
   - 평가의 시간이 지정된 창 내에서 임의로 생성되는 평가 i.e. 식후 30분 이내
   - 자체 보고 이벤트에 대한 응답 i.e. 흡연 빈도수 증가 설문
   - 이벤트 유발 평가 i.e. 흡연 바이오 마커 생성
2. 계산된 데이터 스트림 기반으로 조건부 연산 지원
   - 운전중인 사용자에게 전달되지 않음
   - 센서가 올바르게 착용되었는지 확인
   - 배터리 충전 상태 확인 후 충전 유도 평가
3. EMA나 EMI가 유사한 내용이 전달되지 않도록 함
4. Prompt 사용자 군은 전달 대기 시간을 최대로 늘림
5. 충분한 분포 보장을 위해 시간별로 사용자 군을 나눔
6. 사용자가 프라이버시 통제를 진행하면 EMA는 전달되지 않음
7. EMA나 EMI가 표시되는 원하는 시간대를 설정할 수 있음

Act - burden-and context-aware interactions with participants

- Burden-and Context-aware EMA/EMI Scheduling
  - EMA나 EMI를 전달하기 위해 구성 파일 변경만으로 유연한 스케줄링 메커니즘 구현

현재 선택 가능 항목
- 부담 제한 조건
- 임의 이벤트 트리거 입력
- 개인정보 제약 조건
- 시작일 및 종료일
- 다양한 시간 작업
- 센서 데이터 품질
- 배터리 상태

➢ 일련의 제약 조건으로 정의

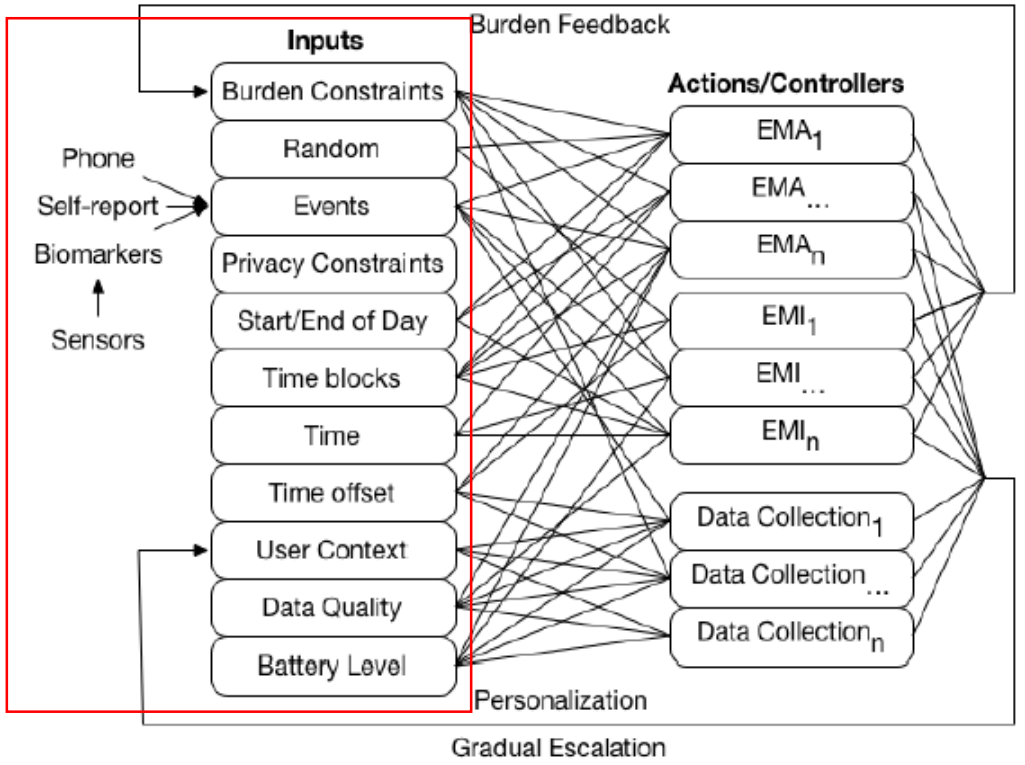


Figure 10: A bipartite graph design of EMA/EMI scheduler. Boxes on left side show the inputs to the actions/controllers on the right, which prompt the participants. Feedback is accomplished by examining the conditions surrounding the participant response and passed back to key input blocks.

EMA 스케줄러
- 트리거 해야 하는 시간을 예측
- 이후 다음 이벤트 처리를 위해 input으로 돌아감
- 점차 참가자의 행동 요청에 적극적 이어짐

타 연구에서 EMA/EMI 스케줄러는 매일 평균 처리 시간을 0.18초로 하루 평균 55건의 시도가 이루어짐

Energy estimation & management

- 스마트폰 배터리는 센서의 raw data 수집시 지속적으로 방전
- 다양한 구성으로 전력 측정을 수행하며, 배터리 수명에 대한 감지 분석 단계의 영향을 보여주어 AWARE와 비교

BLE 이용 시 상당 전류 소비

두 번째 웨어러블 추가되어도 처음만큼 심하지 않음

2.5분 1회 기준 하루 10회

Module Set	Current	Lifetime
Baseline Power	17.13 mA	151.78 hrs
+ Wrist Sensor 1	96.28 mA	27.00 hrs
+ Wrist Sensor 2	107.16 mA	24.26 hrs
+ Biomarker Computation	151.05 mA	17.21 hrs
+ Decision Making	160.42 mA	16.21 hrs
+ EMA/EMI (25 minutes)	187.23 mA	13.89 hrs
+ Location Service	273.77 mA	9.50 hrs
+ Phone Accelerometer	284.33 mA	9.14 hrs
+ Phone Gyroscope	287.67 mA	9.04 hrs
+ Chest Sensor	327.46 mA	7.94 hrs

Table 5: Impacts on expected (2600 mAh) bettery life by incrementally adding sensing and processing capabilities.

- **총 에너지 소비**
  - 기본 작동 9%
  - 감지 48%
  - 연산소비 28%
  - 사용자 상호작용 15%

KOREA UNIVERSITY

Energy estimation & management

- mCerebrum과 AWARE 프레임워크 비교

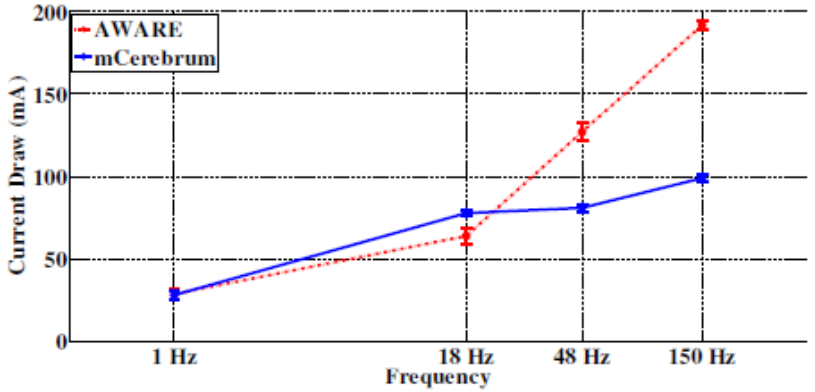


Figure 11: mCerebrum is compared with the AWARE framework by adjusting the total sample rate of the accelerometer, gyroscope, and magnetometers between 18 and 150 hertz per sensor. mCerebrum's benefit on power consumption is improved as the sampling rate increases, thus pushing the CPU into higher-power states.

- 20hz 미만의 센싱 속도에서 mCerebrum과 AWARE 프레임워크 비교
- 데이터 속도가 임계 값을 초과하면 데이터 속도가 아무리 늘어도 에너지 소비량은 더 낮음
- 150hz인 경우 거의 2배 차이 발생

Lessons learned

- System design process
  - 실제 연구를 개발하고 실행하는데 불확실성으로 인해 반복적
  - 버그 및 기능 사이트 설정, 지난 1년간 약 1000개의 요청 해결
  - 소프트웨어 변경 사항을 논의하기 위한 포럼은 참여적 design process로 이어짐

- Raw-data의 저장으로 인해 데이터 손상되고 느려지는 문제

➢Pebbles를 발달시키게 됨

- System 과부하 문제로 Android의 자동 어플리케이션 종료

➢Mini-batching 진행

- 이외에도 데이터 산출 극대화를 위한 mDebugger 개발

Limitation and futurework

- 외부 웨어러블 센서의 소프트웨어와 클라우드 플랫폼은 생략됨
- GUI나 UX에 대한 설명을 추가로 작성하려고 함

- 현재 mCerebrum의 한계점
  1. 현재 AOS에서만 구동, iOS는 백그라운드 어플리케이션 실행 방식이 달라 재설계가 필요
  2. 선택적 샘플링, 전력 인식 감지 또는 시스템 수명 예측을 위한 적응형 전원 관리 기술 필요
  3. Raw-data 중 신체 데이터는 프라이버시 문제를 야기하여 개인 정보 보호 기간 만료 후 생체 데이터 은행처럼 저장할 수 없음
  4. 센서 분류기의 계단식 활성화/비활성화 지원 제공
  5. Pebbles는 현재 시간, 위치 기반 데이터를 지원하지 않음

- mCerebrum 플랫폼은 전원관리, 추론 또는 사용자 참여 등 커뮤니티 형성이 가능한 생태계를 제공

Conclusion

- 모바일 및 웨어러블 센서의 미래는 새로운 바이오 마커를 발견 여부
- mCerebrum은 모바일 운영체제에 기본 지원이 없는 상태에서 데이터를 고속으로 수집, 바이오 마커를 찾아냄
- 실제 우리 생활에 적용되기 위해서는 android 등과 같은 모바일 운영 체제의 기본 지원에서 부터 시작해야 함

- 다른 플랫폼도 mCerebrum과 같은 데이터 수집을 지원하지만, 이 데이터를 BigData화 하는데 오래 걸림
- mHealth 센서 데이터를 공유하기 위해서는 더 광범위한 작업이 필요,
  - ➢ 이를 바이오 마커로 변환
- mHealth 센서 데이터를 위한 메타 데이터 자동 생성시스템 개발되면 mCerebrum같은 플랫폼의 유용성 증폭 가능

Critique

- Advantage
- 다양한 알고리즘을 적용해 처리 속도를 빠르게 했다는 점
- 데이터 트레이닝 시간과 데이터 양이 타 논문들과 비교하면 아주 많은편
- 실험의 규모를 보았을 때 결코 짧은 시간 계획된 연구 같지 않음
- ~~개인정보 보호법까지 신경 쓰는 섬세함..~~

- Desired point
- 어플리케이션의 UI를 일부라도 노출해줬으면 이해가 빠를 것 같음
- 사용자가 착용했다고 하는 하드웨어에 대해서 보여지지 않아 아쉬움
- 센서를 통해서 얻어낸 결과가 실제 결과와 어느정도 맞는지 측정하는 실험도 있으면 좋을 것 같음

KOREA UNIVERSITY

Thank you

mCerebrum: A Mobile Sensing Software Platform for Development and Validation of Digital Biomarkers and Interventions

Sensor Network | 컴퓨터공학과 / **박휘재**