Robust Processing of Health Stream Data

Kathleen Ericson
Advisor: Shrideep Pallickara
Colorado State University
ericson@cs.colostate.edu

Motivation

Personal health sensors are more viable as devices become cheaper, smaller, and more mobile.

http://www.ubergizmo.com/2008/03/myecg-e3-80-portable-ecg/

Reduces costs to families
Reduces strain on health care system
Doctors are more well-informed
Supports independent lifestyles
Healthier living conditions.

http://www.cortechsolutions.com/Applications/Event-Related-Potentials/g-MOBIlab

Related Work

Patient monitoring via sensors has been performed by a variety of groups.

CASAS: Smart home in a box explores a lightweight smart home design which is affordable and easy to set up.

http://wsucasas.wordpress.com/news/smart-home-in-a-box-could-capture-community-behavior/

TigerPlace focuses on improving quality of life for seniors living in smart apartments

http://www.livescience.com/25872-aging-in-place-technology-elderly-independence-nsf-bts.html

Current work shows us that it is possible to merge data from multiple sensors to monitor patients. It has also shown us that doing so can improve overall patient health.

Gaps in Existing Work

Current research lacks scale. There are a limited number of participants, data is not processed in real time, and each user has a dedicated machine.

Dedicating a single machine to each user scales poorly and leads to wasted resources. Interleaving computations from multiple users on a particular cluster machine alleviates this.

Methodology

We use Granules, a stream processing framework, to support computations. Granules allows comptutations to enter a dormant state when data is unavailable.

Computations need to be robust to both failures and interference. Both need to be predicted early and accurately in order to minimize state and data loss in order to prevent user injury.

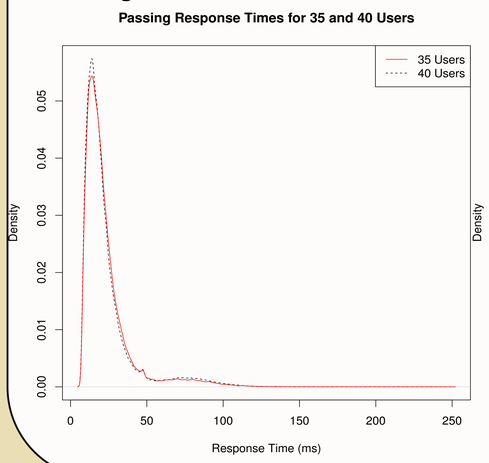
We achieve robustness to failures through replication, and track machine-liveness through gossiped heartbeats.

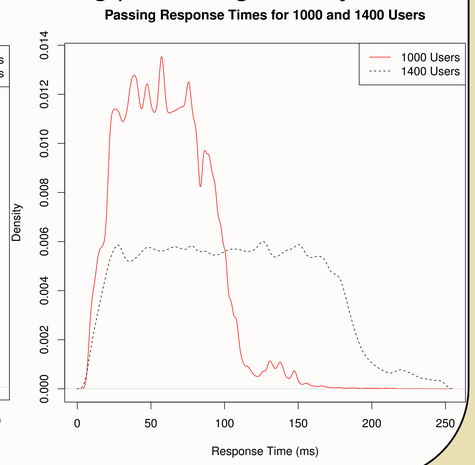
Robustness to Interference

Efficient scheduling is needed to meet processing guarantees. Optimal placements are an instance of the resource-constrained scheduling problem either NP-Hard or NP-Complete.

A robust framework needs to be able to guide placements in real time. We propose using statistical and ML algorithms to guide placements.

Round robin placements assume an equal number of each computation means an equal amount of work. It fails to take into account activation times: Computations which activate together have higher levels of interference, meaning processing is delayed.

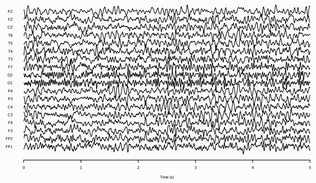


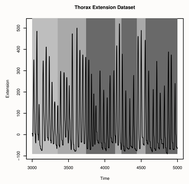


Data

Individual packets are small.
Packets have sub-second arrival rates.
Data must be processed at least as quickly as it is generated.
Bursty in amount of data received, and rate at which it is sent.

We have developed computations to work with electrocardiogram (ECG), electroencephalogram (EEG), and thorax extension datasets.





These computations all have different processing footprints, activation profiles, and resource requirements.

Expected Contributions

Distributed Stream Processing

Interference detection and avoidance
- *Computations are clustered based on resource requirements and activation schedule*
- *Computations in the same cluster are more likely to interfere*

Adaptive fault-tolerance scheme
- *Each computation has its own FT needs*
- *These should be able to change over time*

Solution to an NP-Hard problem without violating robustness constraints

Health Stream Processing

Support for general computation processing
- *Arbitraty computations developed in Java, R, Python, C, C++, C#.*

Meets stringent response time requirements
- *Violations may result in patient injury.*

Large-scale deployment and testing
- *Support hundreds of computations in a cluster environment.*

Societal Impact

Potential to reduce health care costs
- *Less time spent in the hospital*
- *Doctors can safely monitor larger numbers of patients*

Greater independence for elderly and those needing full-time care

Next Steps

Data arrival predictions and placement guidance
Isolation and adaptability to bursty data traffic

This research is supported by a grant from the US National Science Foundation's Computer Systems Research Program (CNS-1253908).

Project Website: http://granules.cs.colostate.edu/