

Trust Management for Cyber-Physical Systems

Insup Lee

PRECISE Center

School of Engineering and Applied Science

University of Pennsylvania

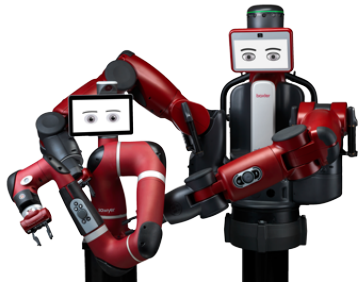
RSS 2016 Workshop – Social Trust in Autonomous Robots

June 19, 2016

Outline

- Introduction
 - The problem of trust in cyber-physical systems
- Previous related projects
 - Quantitative trust in federated networked systems
 - Diabetic patients' trust in insulin pumps
- Vision and open questions

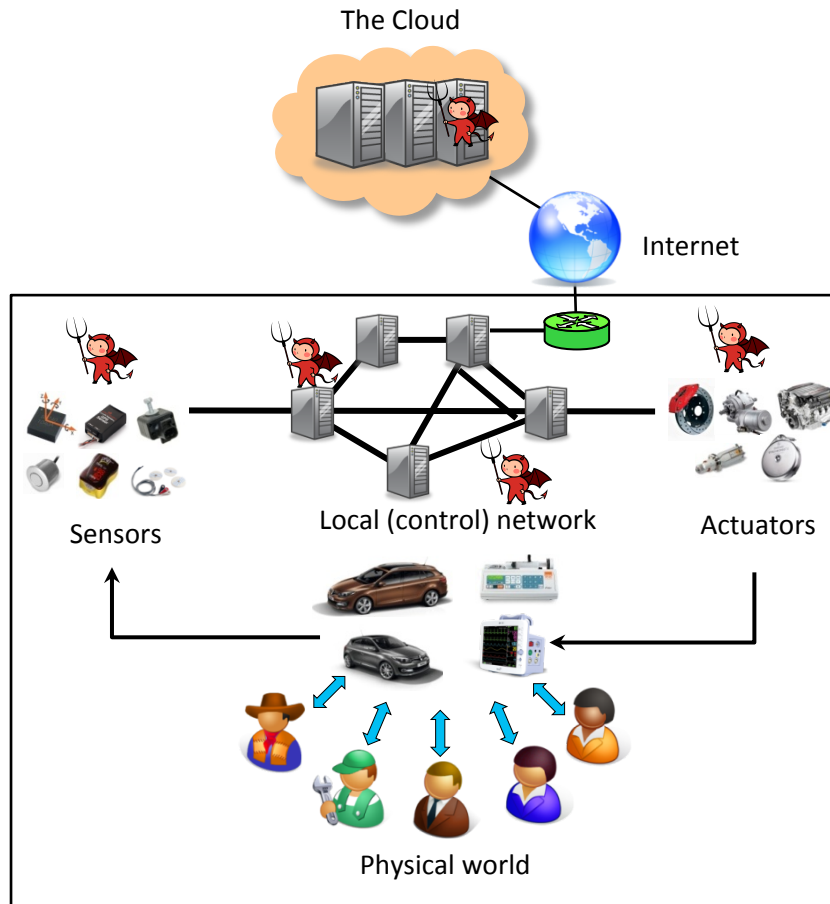
Cyber-Physical Systems (CPS)



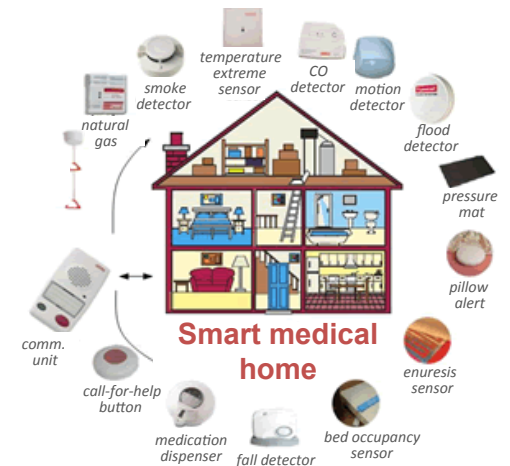
Autonomous robots



Medical devices



Internet-connected car



Smart medical home

Characteristics of CPS

- Pervasive computation, sensing and control
- Networked at multi- and extreme scales
- Dynamically reorganizing / reconfiguring
- Increasing degree of automation
- Dependable operation with potential requirements for high assurance of reliability, safety, security, and usability
- Human in/on the loop

The problem of trust in CPS?



Would You Trust a Robot Surgeon to Operate on You?

Precise and dexterous surgical robots may take over the operating room

-- IEEE Spectrum (June 2016)



Three-Quarters of Americans “Afraid” to Ride in a Self-Driving Vehicle

-- AAA Survey (March 2016)

Trust

- Dictionary: TRUST, -noun:
 - belief that someone or something is reliable, good, honest, effective, etc. [Merriam-Webster]
 - belief that somebody/something is good, sincere, honest, etc. and will not try to harm or trick you. [Oxford]
- Our Definition:
 - Trust is the expectation of an **entity** with respect to certain properties or actions of **another entity** under a specified **context** and **time**, considering the **risks, incentives,** and **historical information**.

Needs: trust management for CPS

- Increasing autonomy
 - How to assure trustworthy and reliable operation?
- Distributed, networked complex systems
 - Decentralized policies, dynamic environment
- Interacting with human operators
 - How does the system express its capability/intention to human operators?
- Social implications
 - Decision-making based on social rules, customs, laws, values, and ethics

Principal Questions

- Who/what to trust?
- How much to trust?
- How to interact accordingly?

Challenges

- What is the basis of trust?
- What is the appropriate notion of trust?
- How to establish trust?
- How to maintain/update trust?
- How to use (the level of) trust?

Research Issues: trust management for CPS

- Who is trusting whom
 - Human to human
 - Human to machine/automation
 - Machine to human
 - Machine to machine
- How to express trust?
- How to establish/evaluate trust?
- How to maintain trust?
- Multi facets (multiple factors contributing to trustworthiness)

Quantitative Trust for Federated Networked Systems

- The problem of TRUST
 - Decentralized policies
 - Dynamic environment, partial trust
 - Complex “trust” models (logic + reputation), in reality
- Applications
 - E-commerce systems
 - Service compositions
 - Reusing components/subsystem in complex DoD systems
 - Crowd-sourced development
 - Social Networks
 - Medical systems
 - Cloud computing

[QTM Project, 2007-2012]

Motivation

Trust in shared information / services

- DoD GIG (Global Information Grid)
- Wikipedia / Wikitrust
- Trusting google result

Trust in a single system / network (system of systems)

- Trusted Computing
- Safe on-line shopping, e-Commerce

Trust in social networks

- Facebook, Twitter
- Use social networking safely

If trust break...

- GiG and possibility of attackers modifying data
- Search engine poisoning

- Malware , virus, worm
- Router's error crashed the Internet

- Social engineering attack
- Revoking trust in companies

Goal and Scientific Challenges

Goal

- Develop the fundamental understanding of “trust” and its application to large complex federated systems

Scientific Challenges

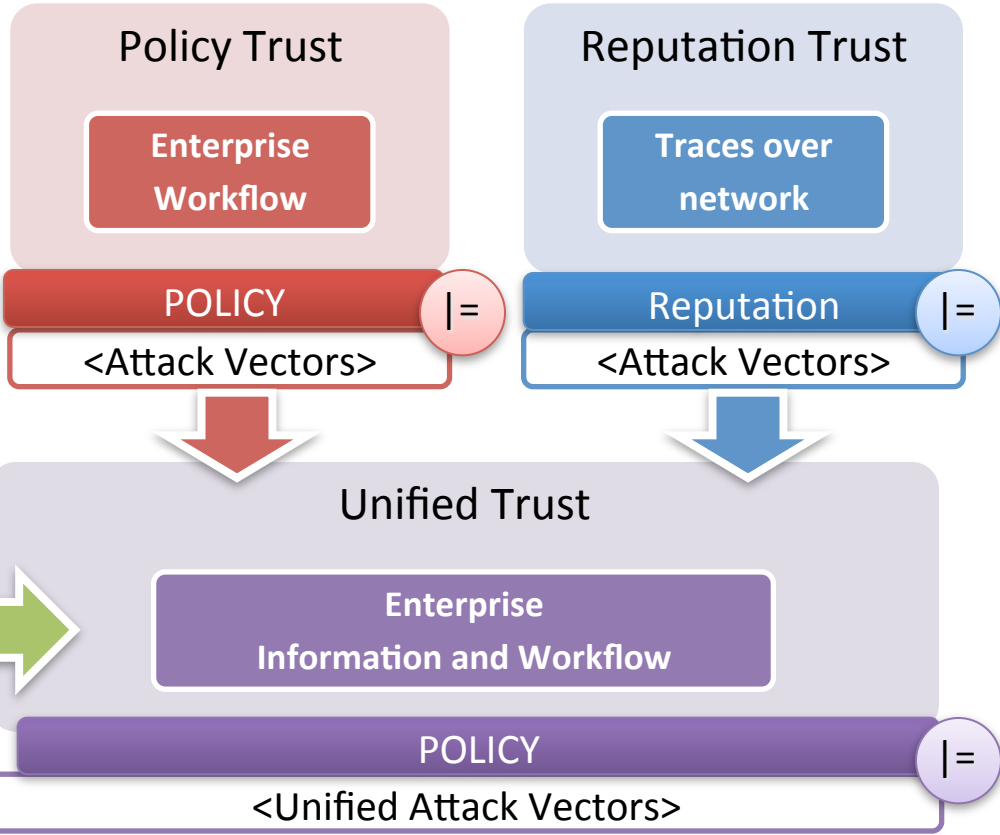
- Establishing trust under conditions of uncertainty
 - What is trust, when assumptions can change?
 - Trust metrics not tied to identify and resilient to attacks
 - Compositional Semantics of trust
 - Trust in anonymous communication networks
 - Trust in system of systems
 - Revocation of trust
- Making decisions based on trust
 - Understanding attack models
 - Bayesian techniques are needed to account for partial-trust

Our Approach

Given multiple trust management policies over different networks, determine a unified enterprise level framework and its ability to withstand disruptions.

Policy Trust Management
Policy also determines behavior toward companies (ally or not).

Reputation Trust Management
Interaction history of companies determines level of trust.



Trust Management

Policy-Based Trust Mgmt. (PTM)

- Effective for delegated credentials and access enforcement
- Can't handle uncertainty and partial information

Rep-Based Trust Mgmt. (RTM)

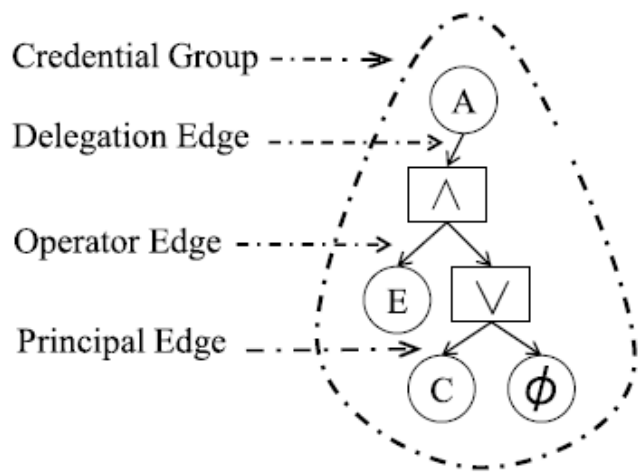
- Quantifies trust relationships
- No delegation (i.e., reputation non-transferable)
- No enforcement

QUANTITATIVE TRUST MANAGEMENT (QTM)

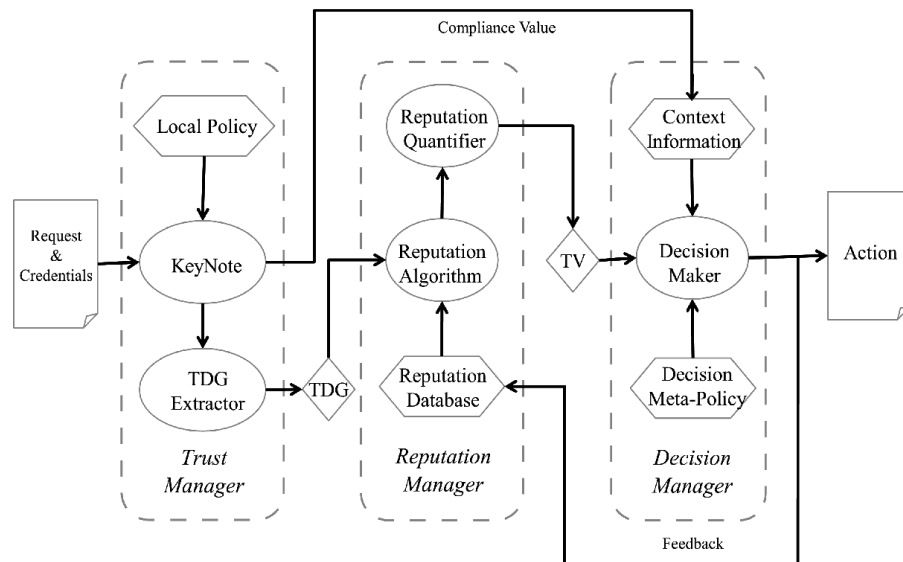
- Combine PTM and RTM
- Dynamic interpretation of authorization policies for access control decisions based on upon evolving reputations of the entities involved

Quantitative Trust Management (QTM)

- Quantitative Trust Management (QTM) provides a dynamic interpretation of authorization policies for access control decisions based on upon evolving reputations of the entities involved
- QuantTM is a QTM system that combines elements from RTM and PTM to create a novel method for trust evaluation
 - Describes the Trust Dependency Graph (TDG), a tree-encoding of policy-based trust relationships apt for reputation application
 - Reputations of not just **PRINCIPALS**, but also **DELEGATIONS** and **CREDENTIALS** are aggregated to arrive at a final value

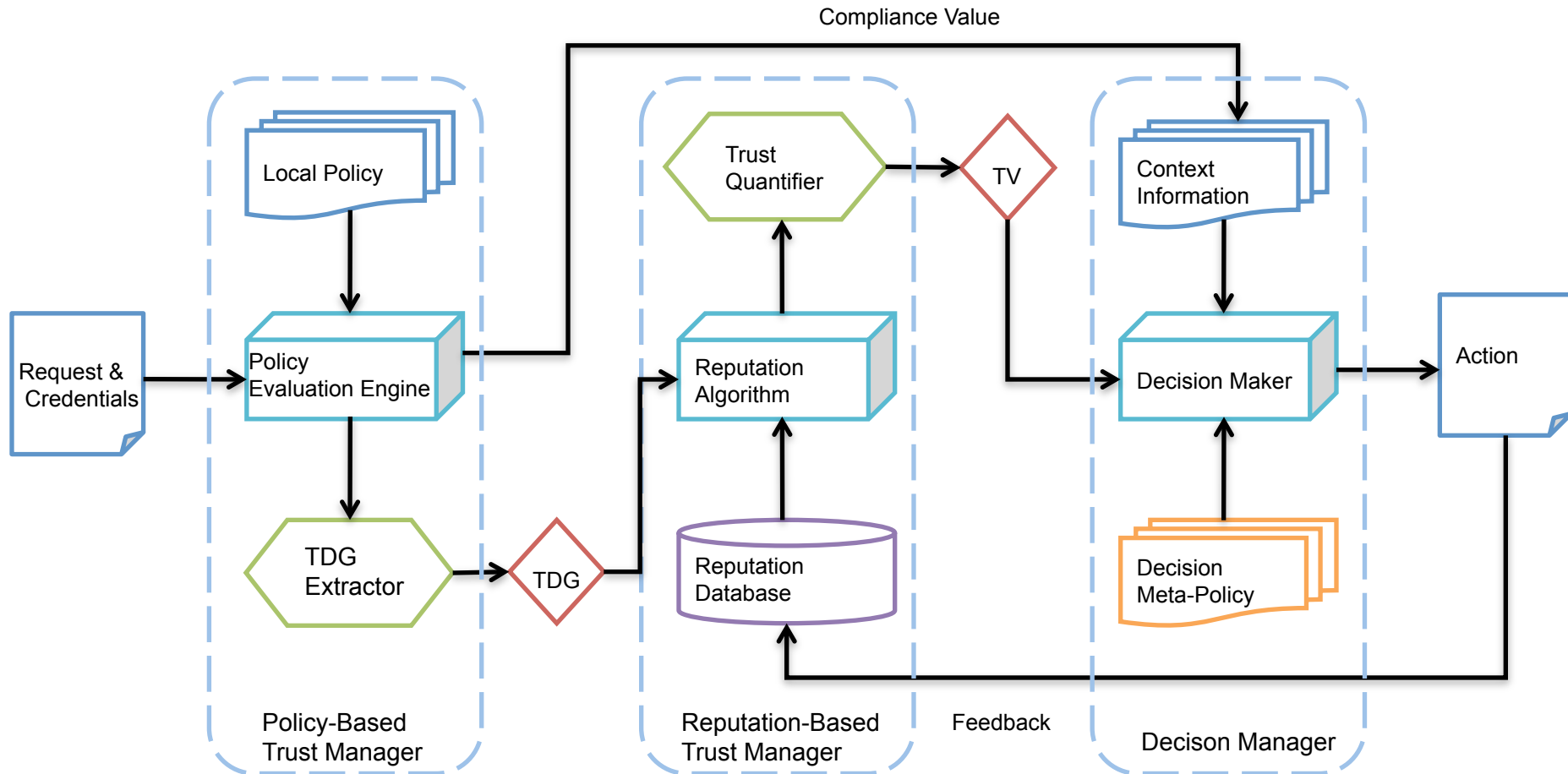


An Example TDG



The QuantTM Architecture

QTM Architecture

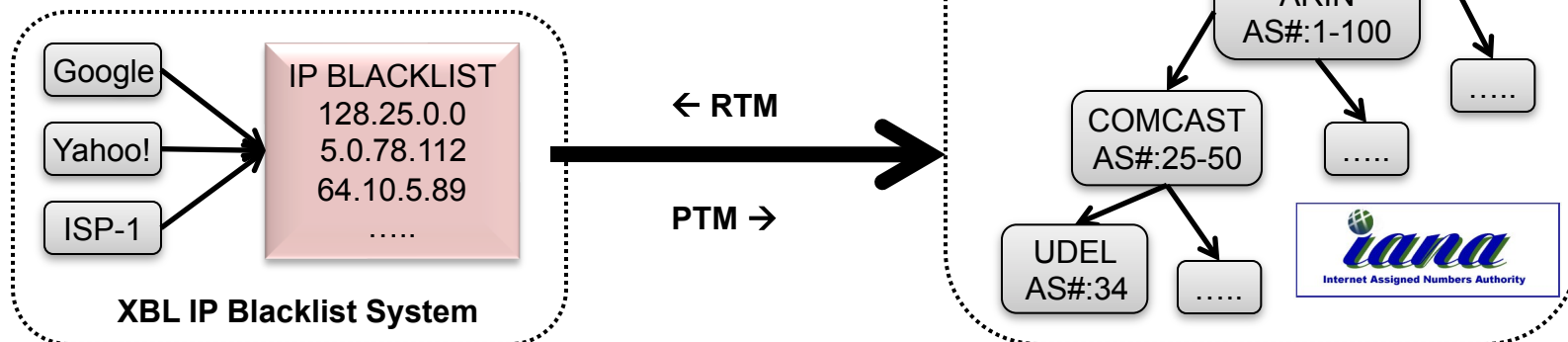


QTM Challenges

- What is the proper way to mathematically **combine reputations**?
 - Involves integration of logical/quantitative/probabilistic reasoning
 - Is there a means to agreeably synthesize distributed observations?
- How does a designer **counter malicious attacks using game-theory strategies**?
 - Attackers will try to exploit all features of a trust system
 - Are there techniques to provide *PROVABLE* resilience to attack?
- What **decision-theoretic approach** is appropriate for trust situations?
 - Huge parameter space a complicating factor
 - Bayesian techniques are needed to account for partial-trust
 - Can this integration provide a means for credential revocation?
- How does one **integrate** policy-based and reputation-based TM?
 - What are the relationships between latencies for authorization decisions and the number of nodes in a QTM-managed system?
 - What is the duration of inconsistent authorization information in a system as policy decisions are updated?
 - How stable and predictable are the behaviors of a large-scale QTM system in the face of rapid changes in factors such as policies, environment and reputations?

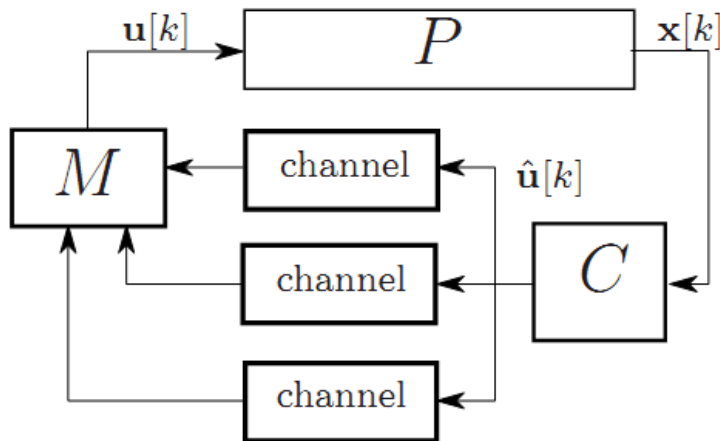
AS# Reputation (MURI)

- Potential unifying application of PTM, RTM, and QTM concepts
- Autonomous Systems (AS) are core Internet routing entities. AS act as hops (and ultimately hosts) in IP routing tables.
 - PTM: AS#'s (unique IDs) are delegated hierarchically, similarly with IP blocks. Attempts to secure BGP rely heavily on PTM-approaches, to ensure an AS actually owns the IP they are attempting to broadcast
 - RTM: An AS originates (hosts) some portion of IP space. Thus, we can use our 'IP Block Reputation' approach to associate reputation values with Autonomous Systems.
 - QTM: A synergy of the above techniques could additionally secure BGP, permit the discovery of spam-friendly ISPs, etc.



QTM-based Networked Control System

- Redundant communication channels are useful for networked control system as they provide the ability to tolerate faults and malicious behaviors occurred in networks.
- Observation:
 - With both faulty and lossy channels, a simple triple-modular redundancy scheme with majority voting may not be sufficient to maintain stability of a controlled plant



- Our Approach
 - Integrate a reputation manager with the networked control system to improve the decision making process and to enhance the overall stability of a plant being controlled.

Risk Control

- Key ideas:

- ① Relies on majority voting result, if it succeeds.
- ② Relies on channel's behavior model to select control input in other cases.
- ③ Applies bounded control inputs to control risk.

Require: Controller injects $-\mathbf{K}\mathbf{x}[k]$ and $\mathbf{u}^b[k]$ into each of the three channels. Both of these signals travel together in one packet. Let R_k denote the number of packets received by the manager M at time-step k .

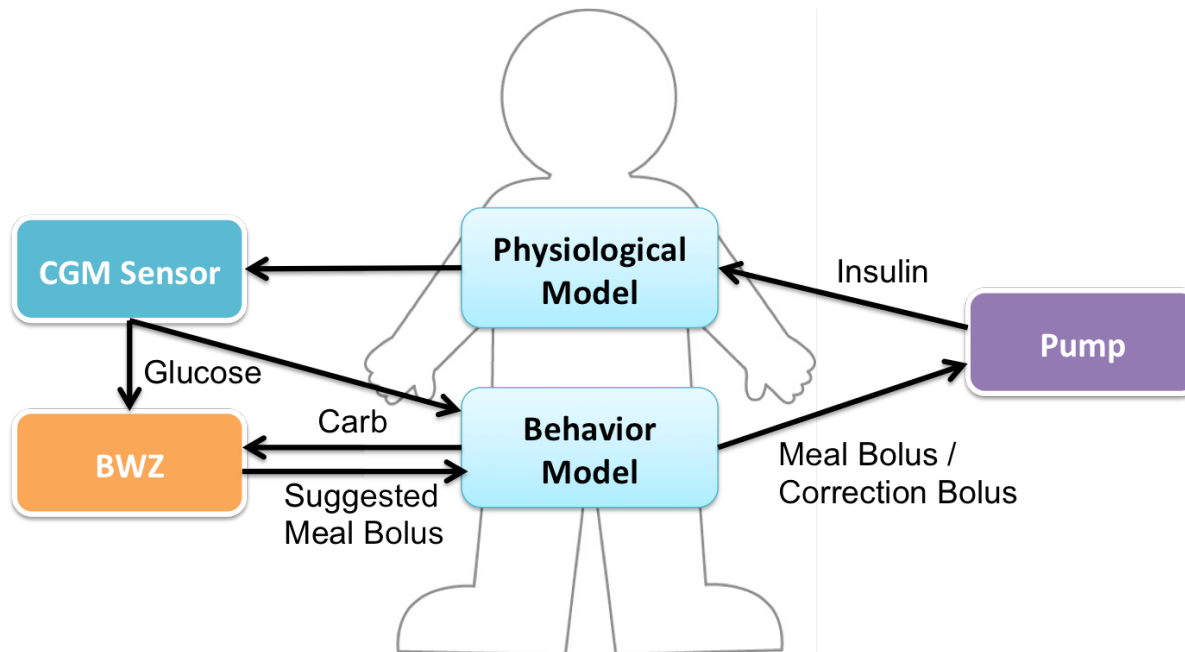
```
1: begin
2: if  $R_k = 3$  then
3:    $M$  applies the value  $-\mathbf{K}\mathbf{x}[k]$  specified by the majority of the packets, and
   updates the behavior model of all three channels accordingly
4: else if  $R_k = 2$  and they both match then
5:    $M$  applies  $-\mathbf{K}\mathbf{x}[k]$  and updates the behavior model of the two channels.
6: else if  $R_k = 2$  and they do not match then
7:   if  $T_i \geq \Theta$  and  $0 \leq T_j < \Theta$  then
8:      $M$  applies the value  $-\mathbf{K}\mathbf{x}[k]$  specified by channel  $c_i$ .
9:   else if  $T_i \geq \Theta$  and  $T_j \geq \Theta$  then
10:     $M$  randomly chooses one of the two channels and applies the input  $\mathbf{u}^b[k]$ 
    specified by it.
11:  else if  $T_i \geq \Theta$  and  $T_j = \phi$  then
12:     $M$  applies  $\mathbf{u}^b[k]$  specified by channel  $c_i$ .
13:  else if  $(T_i = \phi$  and  $T_j = \phi)$  or  $(T_i < \Theta$  and  $T_j < \Theta)$  then
14:     $M$  applies  $\mathbf{u}[k] = \mathbf{0}$ 
15:  end if
16: else if  $R_k = 1$  then
17:   if  $T_i \geq \Theta$  then
18:     $M$  applies  $\mathbf{u}^b[k]$  specified by channel  $c_i$ .
19:   else if  $T_i = \phi$  or  $T_i < \Theta$  then
20:     $M$  applies  $\mathbf{u}[k] = \mathbf{0}$ 
21:   end if
22: else if  $R_k = 0$  then
23:   RM applies  $\mathbf{u}[k] = \mathbf{0}$ 
24: end if
25: end
```

Outline

- Introduction
 - The problem of trust in cyber-physical systems
- Previous related projects
 - (Machine to machine) Quantitative trust in federated networked systems
 - (Human to machine) Diabetic patients' trust in insulin pumps
- Vision, challenges, and future work

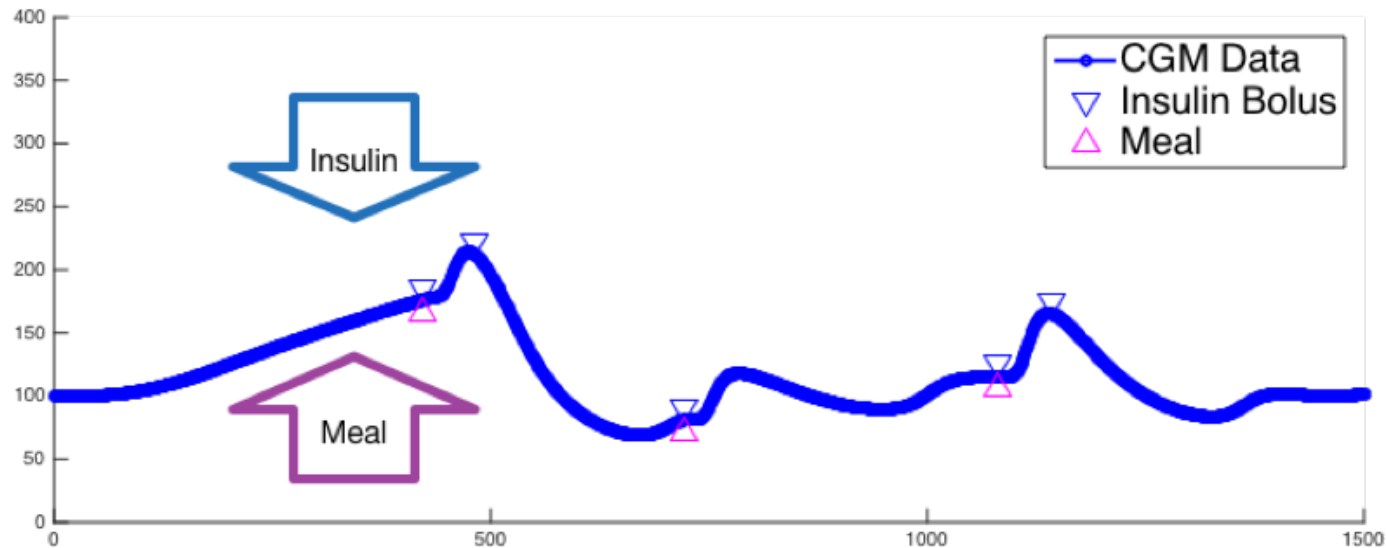
Type 1 Diabetes (T1D) on Insulin Pumps

- Sensor-augmented subcutaneous insulin therapy
 - 30% - 40% T1D patients in the US use insulin pumps
 - Requires user supervision
 - Critical needs for understanding the impact of insulin pumps on diabetic users, as highlighted in a American Association of Clinical Endocrinologists report

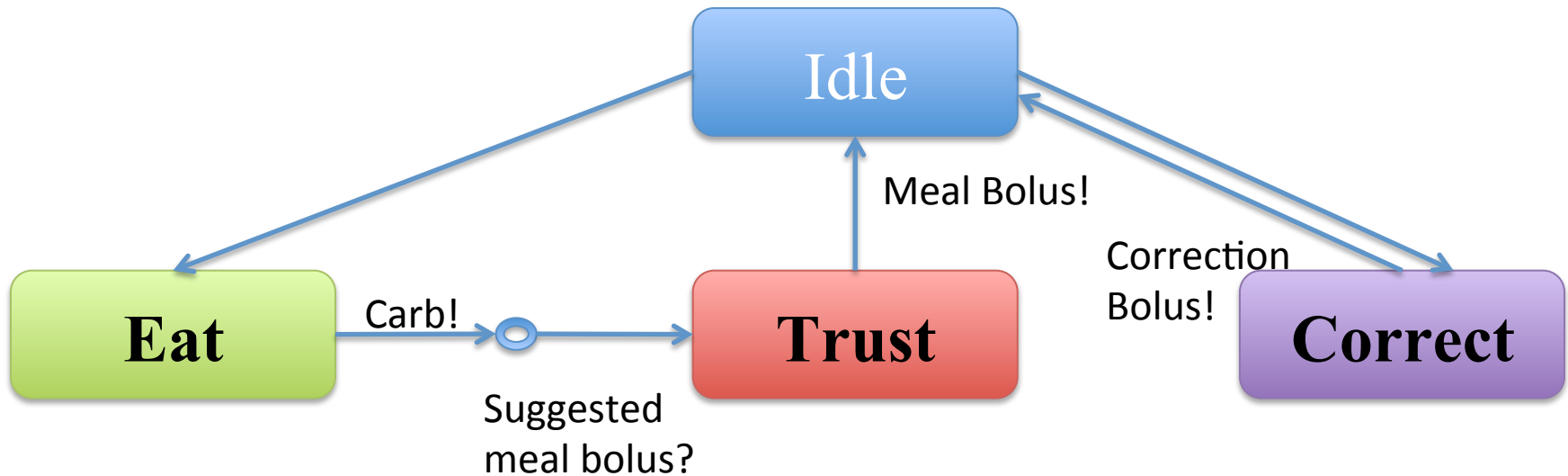


Clinical Dataset

- The dataset involves 55 T1D patients
 - Age 45.7 ± 15.3 , body weight 79.2 ± 21.9 kg
 - Average time duration 31 days
- Sensor-augmented insulin pump data
 - CGM readings, mealtimes & carb counts, pump suggested boluses, user-selected boluses



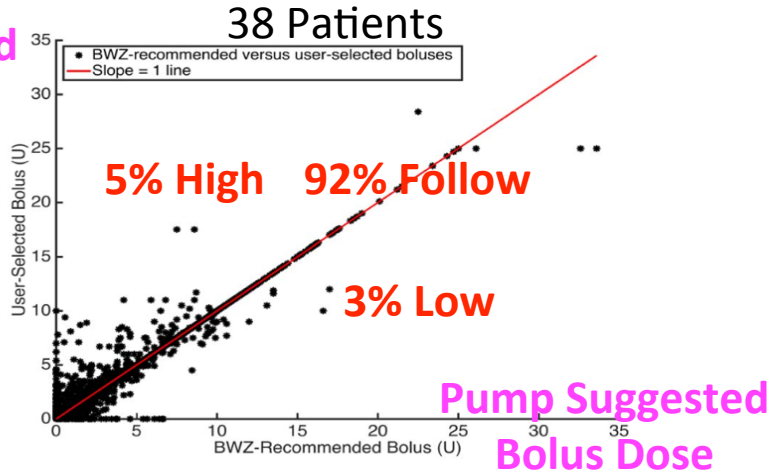
“Eat-Trust-Correct” Modeling Framework



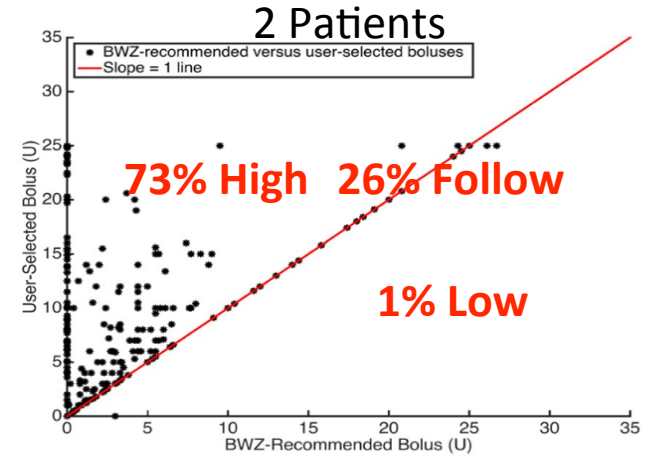
- **Eat**: how often the patient eats throughout a day, and how much carbohydrate he/she eats
- **Trust**: whether the patient follows the BWZ recommended bolus doses, and if not, how much dosage he/she adjusts
- **Correct**: how often the patient takes correction boluses and how much dosage he/she takes

Clustering of Patient Behavior Patterns: Trust

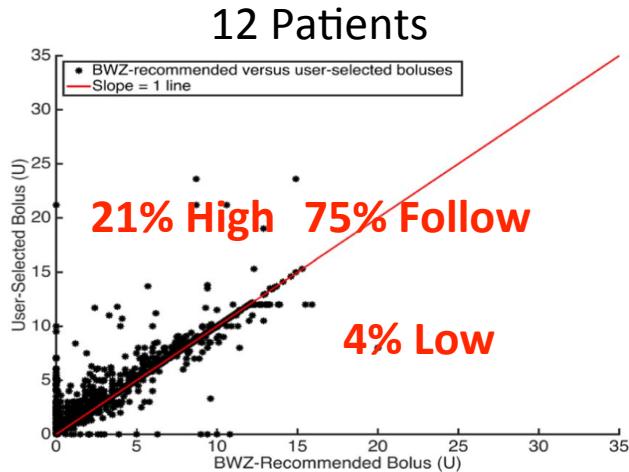
User Selected
Bolus Dose



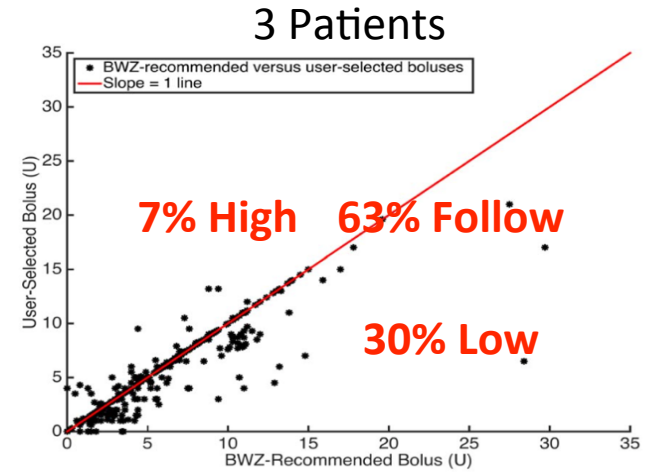
(a) Trust T1: high probability of following BWZ-recommended doses



(b) Trust T2: high probability of increasing BWZ-recommended doses



(c) Trust T3: moderate probability of increasing BWZ-recommended doses



(d) Trust T4: high probability of decreasing BWZ-recommended doses

Outline

- Introduction
 - The problem of trust in cyber-physical systems
- Previous related projects
 - Quantitative trust in federated networked systems
 - Diabetic patients' trust in insulin pumps
- Vision and open questions

Vision

- Decision Support based on Trust Management
 - State of Nature: Are the (other) CPS systems trustworthy?
 - Observations:
 - Prior behaviors of these systems;
 - Certifications by (partially-trusted) authorities
 - Loss Function: To be suitable defined
 - Decision Actions: to (or not to) depend on services provided by CPS systems

Desirable Properties

- Need composable, dynamically computable notion of trust.
- Trust should be quantitative, learned from history, and sensitive to context ... not absolute.
- Policy and Credentials should be formally specified and revocable.

Research Questions

- How to accumulate reputation/feedback? Locally or... should there be trusted authorities?
- How can (central/distributed) authorities monitor transactions to compute current trust levels without violating privacy?
- How do we compose trust values computed over time and from different components of a CPS system?
- What are appropriate Loss values? How sensitive is our decision procedure to the exact values.

Research questions

- How to build a unified framework for expressing, establishing and evaluating different types of trust among humans and machines in CPS?
- What is the right granularity for trust model?
- What data are the best indicators of trustworthiness?
- How do we continually monitor and modify the way we compute trust?

Thank You!
Questions?