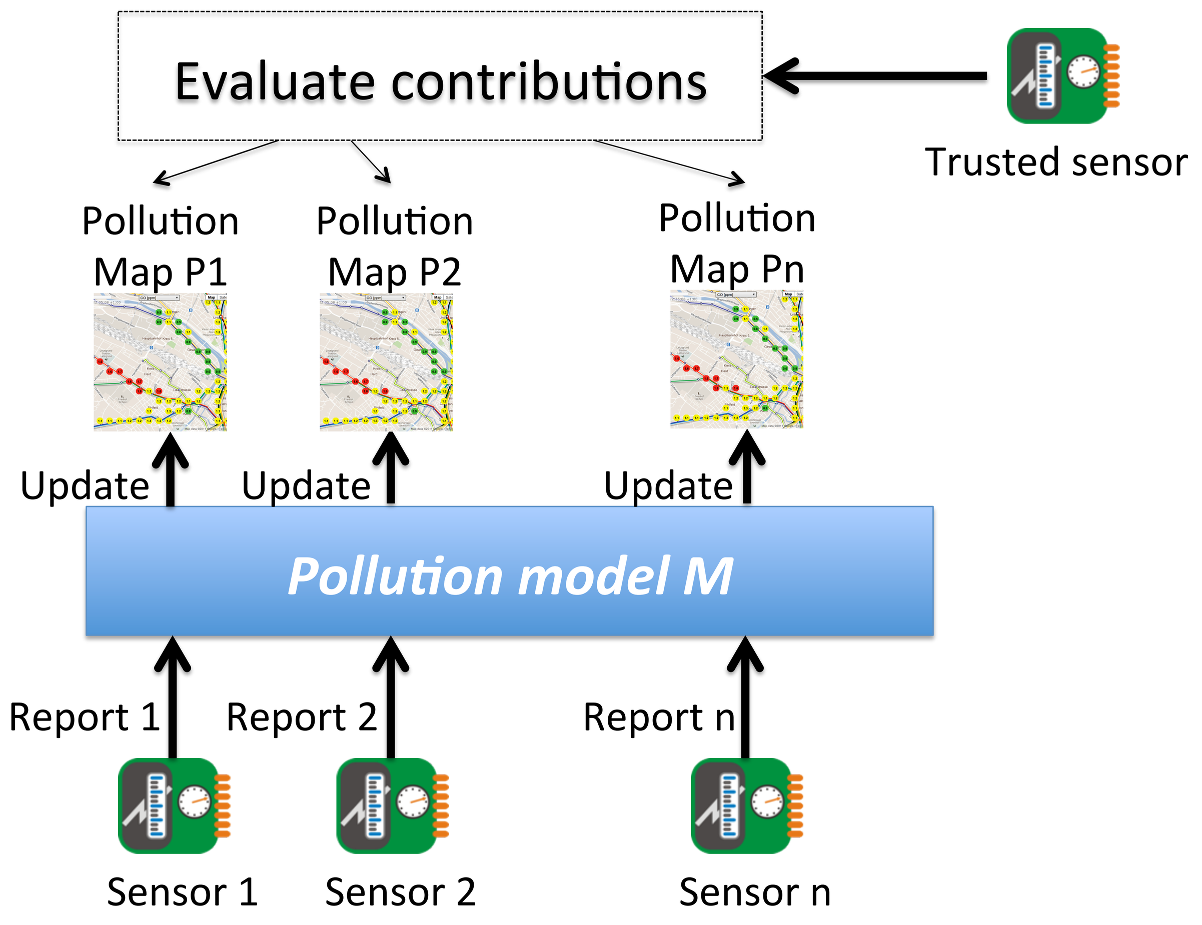
Limiting the Influence of Low Quality Information in Community Sensing

Introduction

A big issue in community sensing is that malicious agents can insert false information. This is usually addressed using reputation systems that estimate the credibility and punish misinformation. We present a novel reputation system that for the first time allows to bound the negative impact that malicious sensors can have on the learned outcome.

The Setting



The sensing scenario with *online information fusion*:

- Initially, the center has prior information about air pollution over an urban area.
- Crowd-sensors report their measurements sequentially. Each report *Y* is merged with the current pollution map *P* using pollution model *M*.
- When trusted sensor reports X_t, the center can evaluate the reports of crowd-sensors using a scoring function *S*. This corresponds to one period of sensing *t*.
- The crowd sensing process then continues in the same manner until the period $t = T$, which is called *sensing time*.

Goal: *Limit the overall negative influence that a sensor can have on the fused result*.

Our Approach: *Track the quality of reported information using a reputation system and discard information coming from sensors with low reputations. Reward a sensor based on its marginal contributions.*

Inefficiency Measures

Expected myopic impact – measures the influence of a sensors:

$$\Delta_s = \sum_{t=1}^{T} \Delta_{s,t} = \sum_{t=1}^{T} \Pr(update) \cdot [S(P_{s,t}^{new}, X_t) - S(P_{s,t}^{old}, X_t)]$$

Expected information loss – measures the amount of discarded information from non-malicious sensors:

$$IL_s = \sum_{t=1}^{T} (1 - \Pr(update)) \cdot [S(P_{s,t}^{new}, X_t) - S(P_{s,t}^{old}, X_t)]$$

Related Work

- *The influence limiter: Provably manipulation-resistant recommender system*, P. Resnick and R. Sami, 2007.
- *A robust reputation system for mobile ad-hoc networks*, S. Buchegger and J.-Y. L. Boudec, 2003.
- *The beta reputation system*, A. Josang and R. Ismail, 2002.

CS Influence Limiter (CSIL)

```
for t = 1 to t = T do:
    foreach Sensor s do:
        P_{s,t}^old = P;
        P_{s,t}^new = Update(P, Y_{s,t});
        if rand(0, 1) < ρ_{s,t}/(ρ_{s,t} + 1) then:
            P = P_{s,t}^new;

        When X_t is received:
            score_{s,t} = S(P_{s,t}^new, X_t) - S(P_{s,t}^old, X_t);
            ρ_{s,t} = ρ_{s,t} * (1 + 0.5 * score_{s,t});
```

Information fusion: stochastic and reputation dependent
Reputation update: exponential increase/decrease of reputations

Properties

Theorem 1: (*Query Complexity*) The number of queries to a black box model *M* of the CSIL algorithm in one time period *t* is O(*n*), where *n* is the number of reported values.

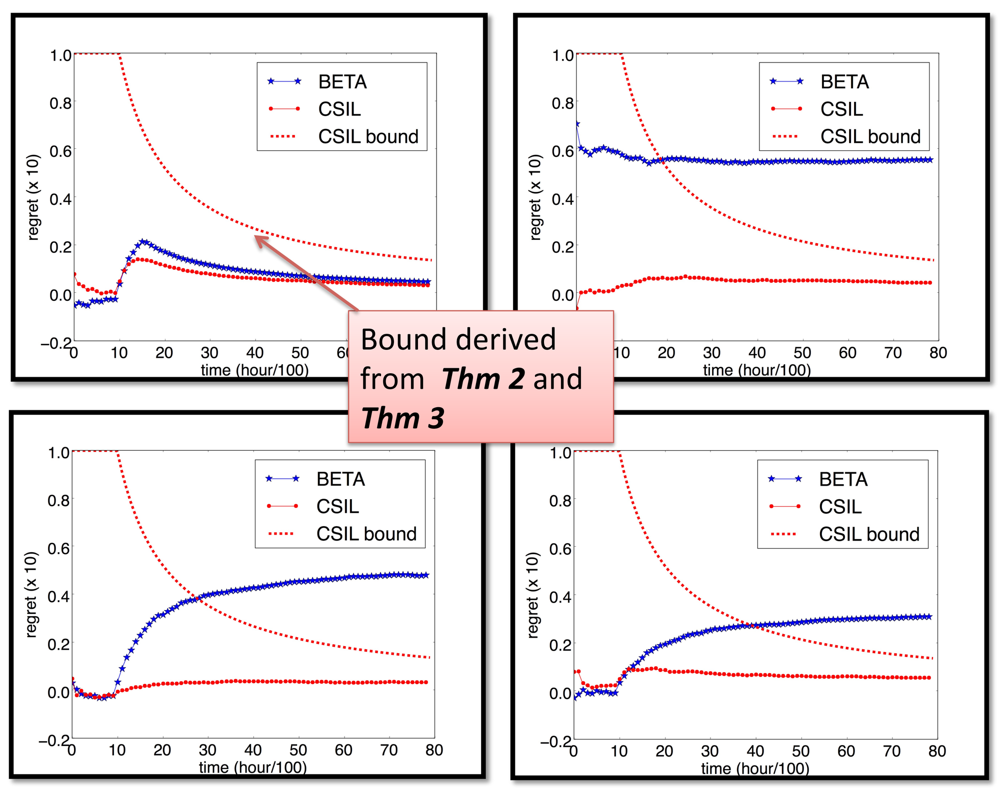
Theorem 2: (*Limited Damage*) The expected total myopic impact of sensor *s* is in the CSIL algorithm bounded from below by $\Delta_s > 2\rho_0$, where ρ_0 is the initial reputation of sensor *s*.

Theorem 3: (*Bounded Information Loss*) Informally, the expected information loss of the CSIL algorithm for potentially discarding the reports of an accurate sensor is bounded from above by a constant.

Theorem 4: (*Informed Reporting*) If a sensor *s* maximizes its expected score, then it also maximizes its expected impact.

Experimental Analysis

Baseline: Beta reputation system with trustworthiness determined by a fixed threshold.
Sensors: 25% honest; 75% malicious - 4 different misreporting strategies.
Quality measure: Average regret (over time) for not knowing which sensors are honest (the lower, the better).



Results: CSIL outperforms the baseline and satisfies the no-regret property.

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Artificial Intelligence Laboratory
Goran Radanovic and Boi Faltings
{firstname.lastname}@epfl.ch