Evaluating Trustworthiness through Monitoring

the foot, the horse and the elephant

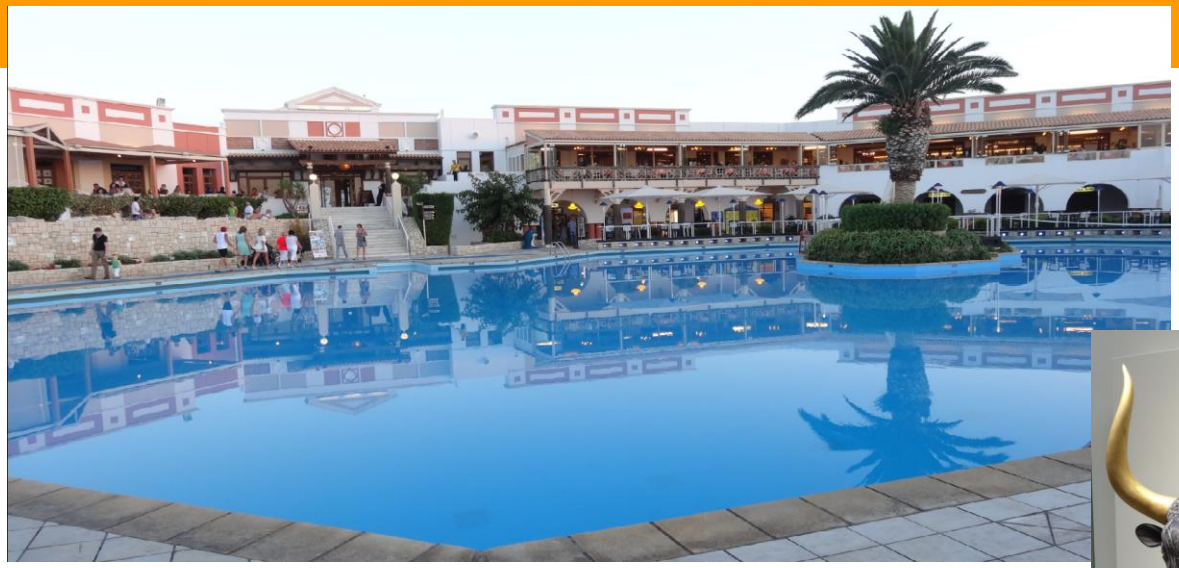
7th International Conference on
Trust & Trustworthy computing

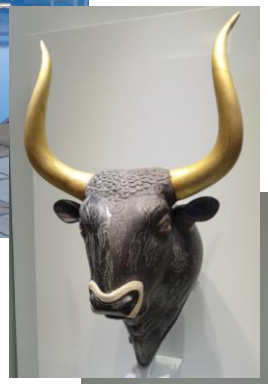
Johan Lukkien, Vinh Bui, Richard Verhoeven
Eindhoven University, Computer Science
System Architecture and Networking

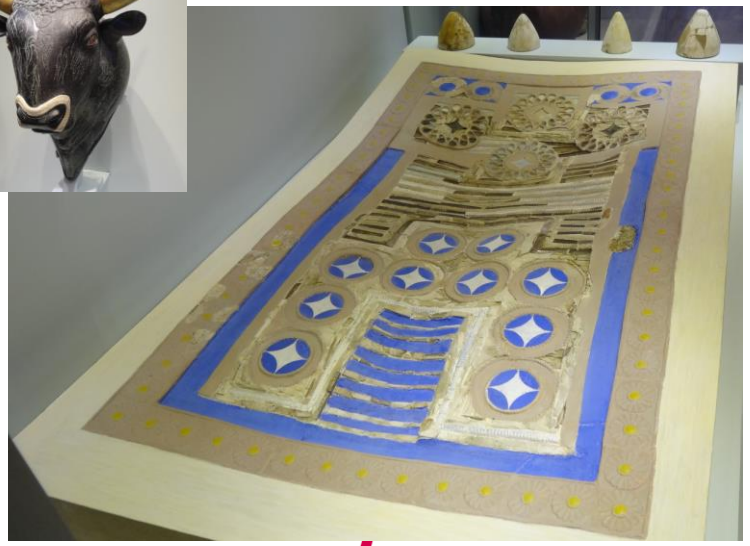
TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

Some hard work….







Technische Universiteit Eindhoven University of Technology

….on Crete

Evaluating Trustworthiness through Monitoring

the foot, the horse and the elephant

7th International Conference on
Trust & Trustworthy computing

Johan Lukkien, Vinh Bui, Richard Verhoeven
Eindhoven University, Computer Science
System Architecture and Networking

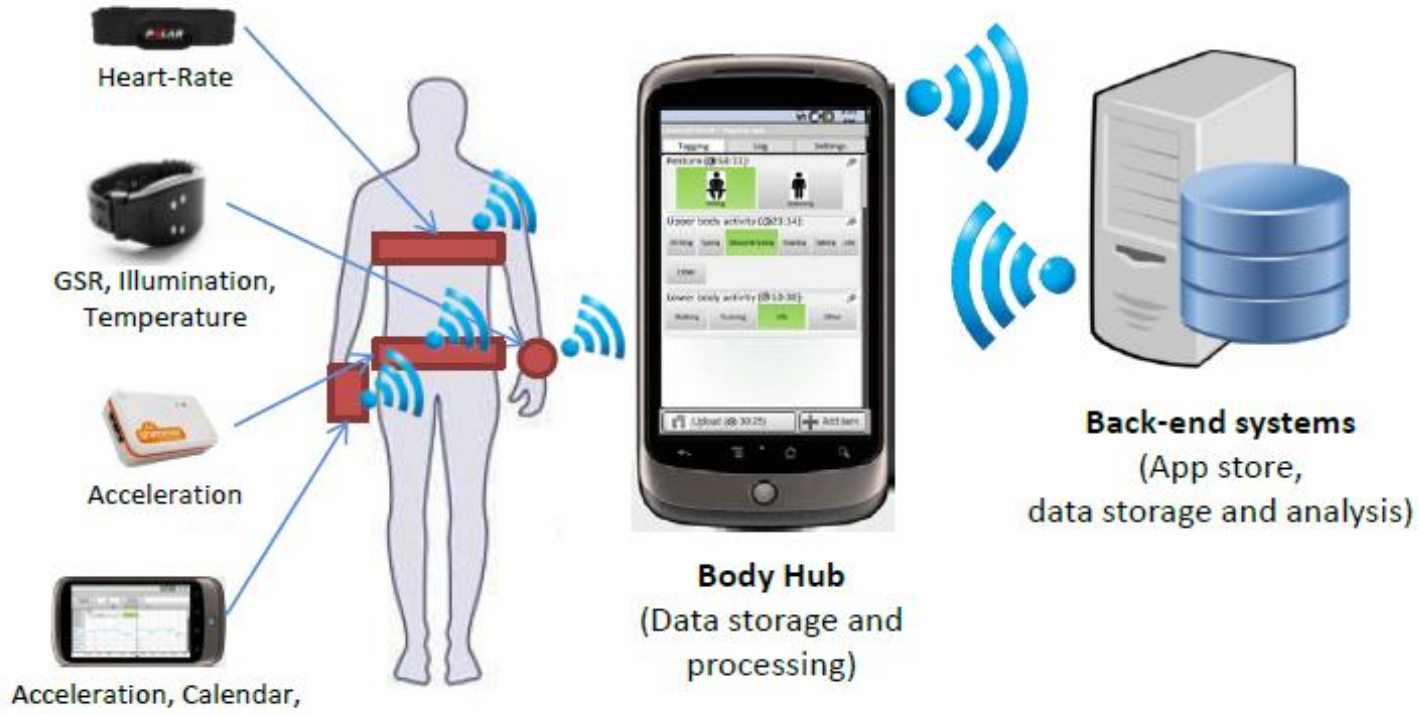
TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

Outline

- Motivation and problem statement
- Monitoring chain
- Monitoring models and requirements
- Examples
- Conclusion

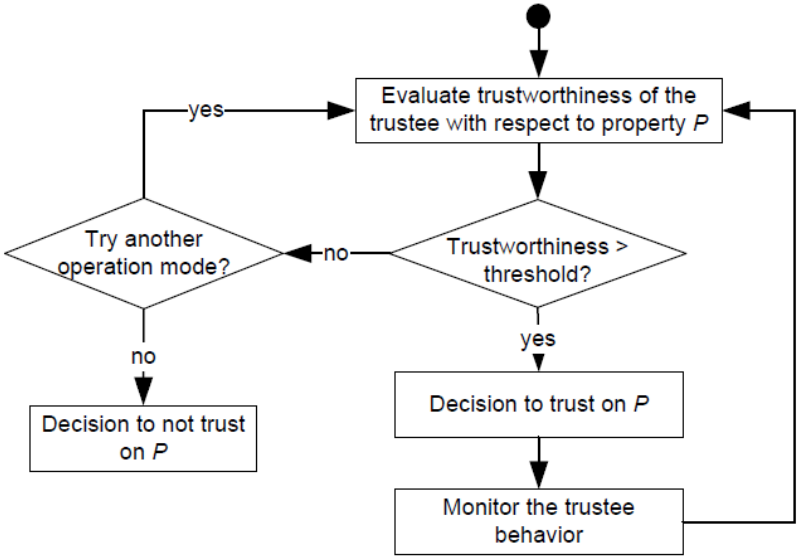
Motivation: VITRUVIUS project



- Trustworthy Body Sensor platform

Trustworthiness concerns

- *Doctor*
  - quality of measurements
  - dependability of the system
    - e.g. how long it will run
- *Wearer, owner*
  - protection, control of personal data
- *System management*
  - maintain trustworthiness under changes
    - new software components
    - upload data to backend
  - manage resources



Global Vitruvius Monitoring & Decision scheme

Problem statement

- Design an integrated *Monitoring and Decision method* for maintaining trustworthiness
  - deal with *subjectivism* and *uncertainty*
    - local, partial knowledge of a *Trustor* about a *Trustee*
    - obtained directly through monitoring or indirectly through referral
  - generic
    - aim at a light-weight ‘kernel’
    - parameterized by monitoring and decision parameters, and primitives

Monitoring, trust statements

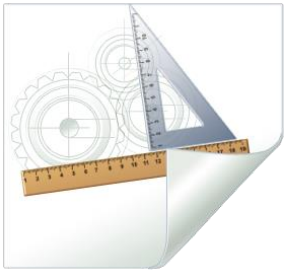
- Define *indicators* that together address a particular trustworthiness concern
  - e.g. levels of resource use, volume of communication, connections with remote machines, pattern of sensor reading, frequency of location checking, …
- Define a *monitoring state* and criteria for this state to be safe
- Model: a Trustor ***A*** develops trust in a trust statement ***P = (T, p, c, t)***
  - ***T****:* trustee
  - ***p****:* statement / predicate about an underlying indicator
    - can refer to any property that admits monitoring, e.g. ‘memory usage < 90%’
  - ***c****:* context
  - ***t****:* time (duration that *p* holds)

Monitoring and decision chain for a trust statement *P*

World

Measurement

Observation about *P*



Update rule

Evidence about *P*

Opinion Generation

Opinion about *P*

Extraction

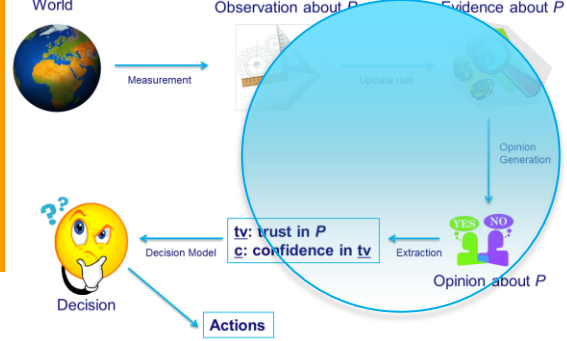
tv: trust in *P*
c: confidence in tv

Decision Model

Decision

Actions

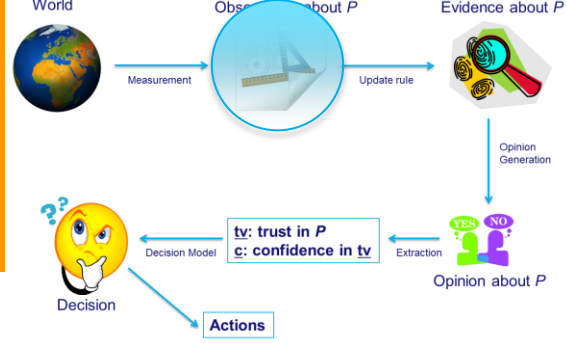
Monitoring requirements



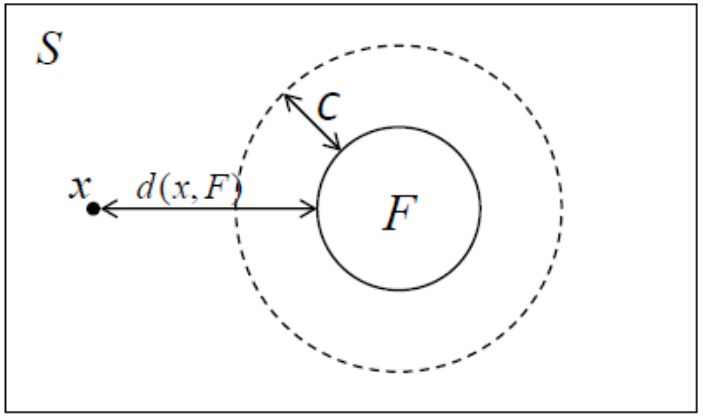
- **`Soundness'**
  - a long series of success (failures) leads to trust and confidence going to 1 (0) (or at least non-decreasing (-increasing))
  - relatively frequent failure leads to low trust
  - confidence decreases upon missing observations

- **Expressive power**
  - take *time* into account: more recent observations (could) count more
  - parameterization, to achieve particular behavior:
    - slow increase, quick decrease of trust (the foot and the horse)
    - do not forget failures in the past (the elephant)

Observations



- An observation is the outcome of a measurement *x*
- We consider a *success, failure* outcome: *(s(x), f(x))*
  - *f(x) = 1-s(x)*
- *Discrete*: *(1,0)* or *(0,1)*, i.e. truth of a statement
  - e.g. 'the message has been delivered'
- *Continuous*: success represents distance to a failed state (generalization of the discrete case)
  - e.g. *x* is fraction of memory usage, *F* = 90%, *C* = 10%



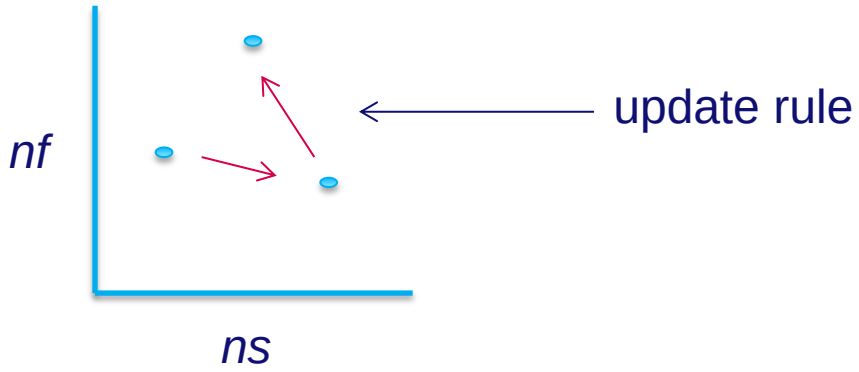
$$s(x) = \begin{cases} (\frac{d(x,F)}{C})^{\alpha}, & \text{if } d(x,F) < C \\ 1, & otherwise \end{cases}$$

Evidence and update rules

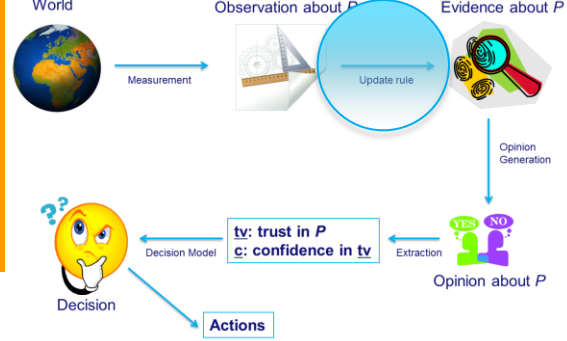
- *Evidence*: representation that captures the history of observations

- *Update rules*: adjust the evidence (periodically) based on
  - an observation
  - (optionally) a timer: 'unknown' observation leading to a decay of evidence

- We consider a representation with two variables: *ns* and *nf*
  - representing positive (success) and negative (failure) aspects



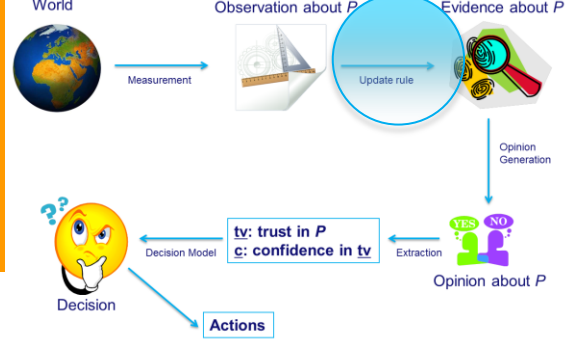
Update rules



- *'Classical' update rule* (from subjective logic):
  - discrete: (*s*,*f*) = (0,1) or (1,0)
  - add up all positive and negative contributions
    - upon success: $ns \leftarrow ns + 1$
    - upon failure: $nf \leftarrow nf + 1$
  - problem:
    - always symmetric
    - effect of new observations vanishing

- Introduce parameters: $0 < \delta, \gamma, \zeta \leq 1$
  - modeling reduction of evidence components upon new observations

Proposed update rules

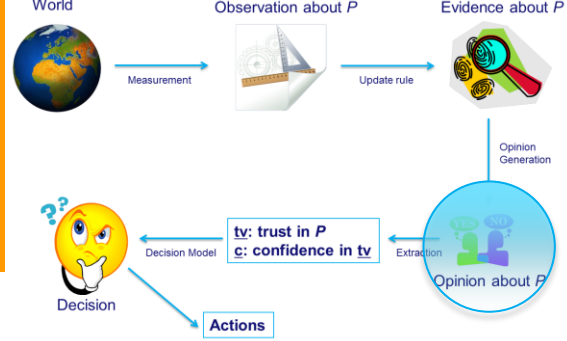


- Observation: *(s, f =1-s):*
$$ns \leftarrow \delta^f \cdot ns + s$$
$$nf \leftarrow \gamma^s \cdot nf + f$$
- Outcome *unknown:*
$$ns \leftarrow \zeta \cdot ns;\ nf \leftarrow \zeta \cdot nf$$

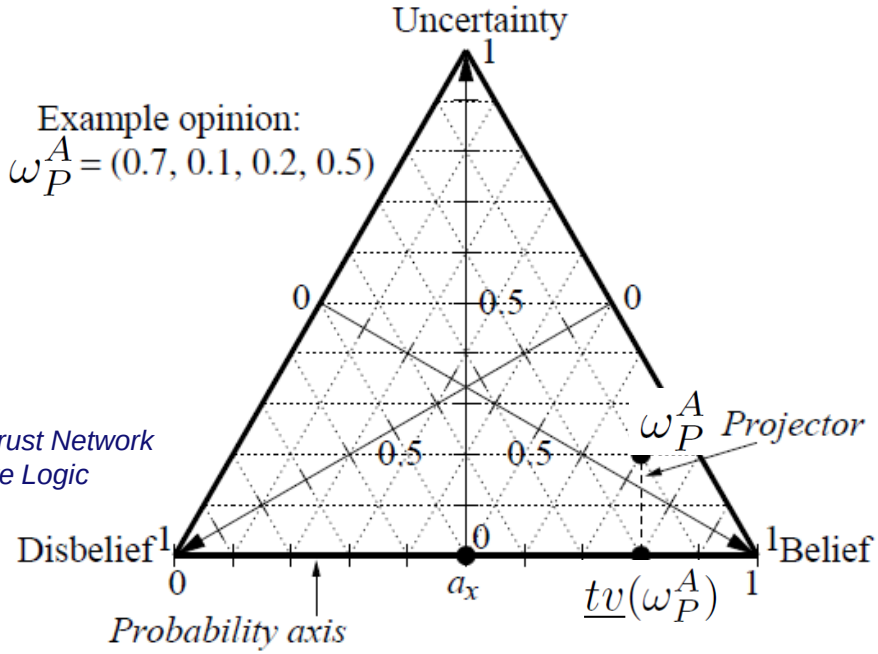
- Discrete case:
  - upon success:
$$ns \leftarrow ns + 1;\ nf \leftarrow \gamma \cdot nf$$
  - upon failure:
$$nf \leftarrow nf + 1;\ ns \leftarrow \delta \cdot ns$$

- 'Classical': discrete, with $\gamma = \delta = \zeta = 1$

Opinions



- Following the concepts of *subjective logic (*cf. Audun Jøsang (’97, ‘06))
- Belief model: an *opinion*, consisting of a tuple of *belief*, *disbelief* and *uncertainty,* and a ‘*base rate*’
  - $\omega_P^A = (b, d, u, a)$: opinion of a trustor *A* about a trust statement *P*
  - base rate represents belief taken from uncertainty



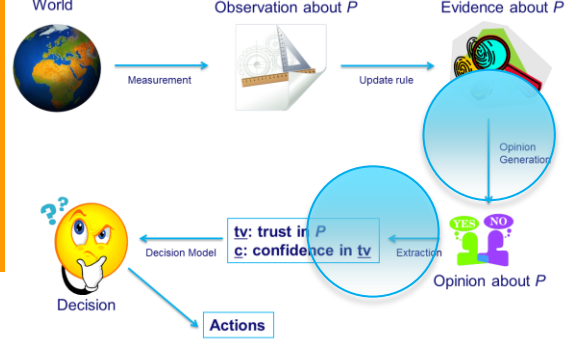
$$b + d + u = 1$$

$$0 \leq b, d, u \leq 1$$

$$\underline{tv}(\omega) = \omega.b + \omega.u \cdot \omega.a$$

Jøsang et al. (2006), *Trust Network Analysis with Subjective Logic*

Opinions, evidence and trust



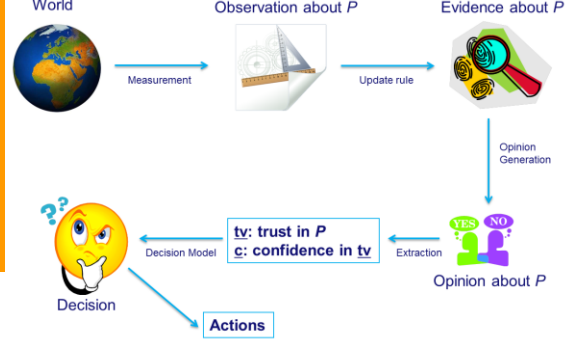
- An opinion is computed from evidence (Jøsang):

$$b = \frac{ns}{ns + nf + \epsilon}; \quad d = \frac{nf}{ns + nf + \epsilon}; \quad u = \frac{\epsilon}{ns + nf + \epsilon}.$$

- ϵ represents an initial uncertainty
- base rate *a* remains free to choose by the trustor

- a trust value is derived from an opinion: $\underline{tv}(\omega) = \omega.b + \omega.u \cdot \omega.a$
- a confidence gives the degree of certainty: $\underline{c}(\omega) = 1 - \omega.u$

Example (discrete)



success: $ns \leftarrow ns + 1;\ nf \leftarrow \gamma \cdot nf$

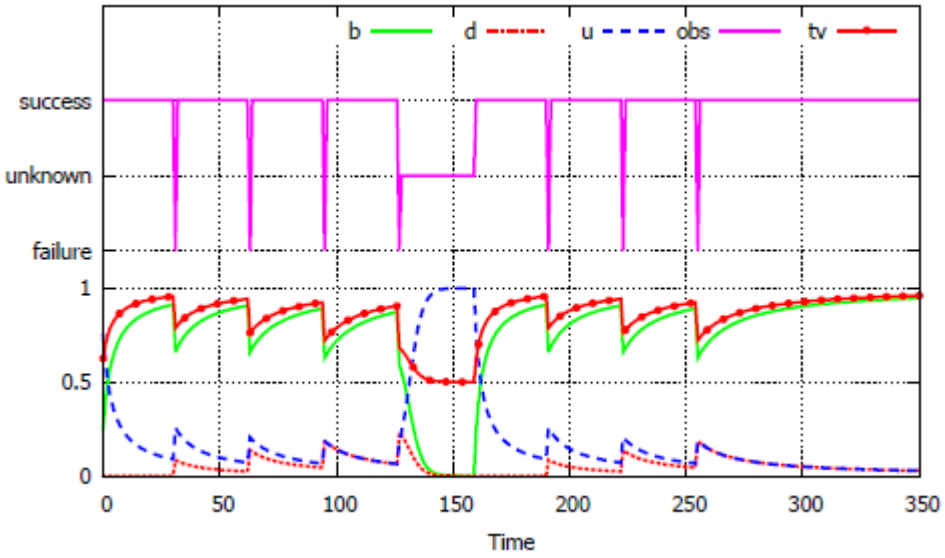
failure: $nf \leftarrow nf + 1;\ ns \leftarrow \delta \cdot ns$

unknown: $ns \leftarrow \zeta \cdot ns;\ nf \leftarrow \zeta \cdot nf$

discrete input

rapid decrease upon failure

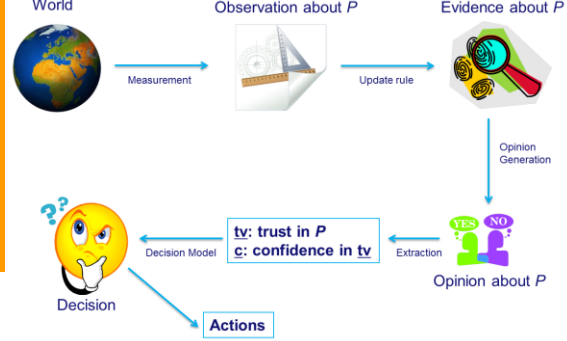
low confidence around150



recovery increasingly slow (failure is only 'forgotten' through decay)

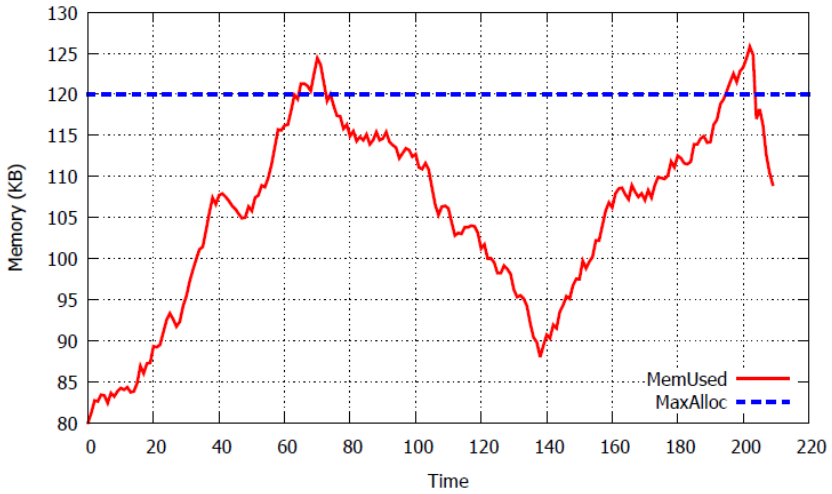
$\gamma = 1.0, \delta = 0.25, \zeta = 0.7$

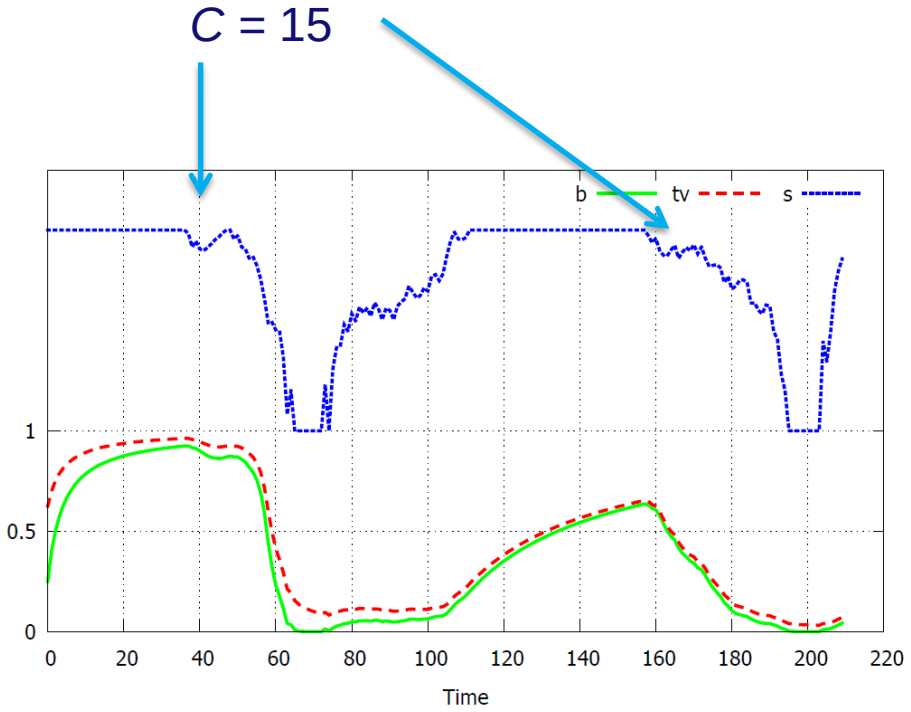
Example (continuous)



$$nf \leftarrow \gamma^s \cdot nf + f$$

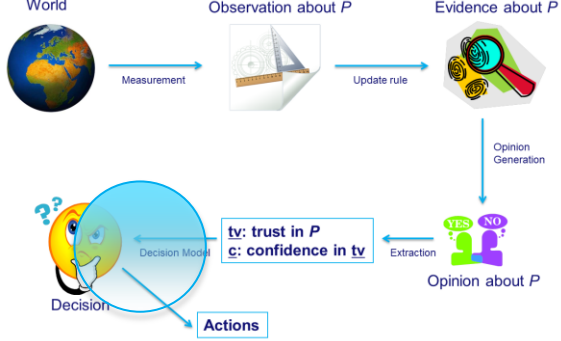
$$ns \leftarrow \delta^f \cdot ns + s$$





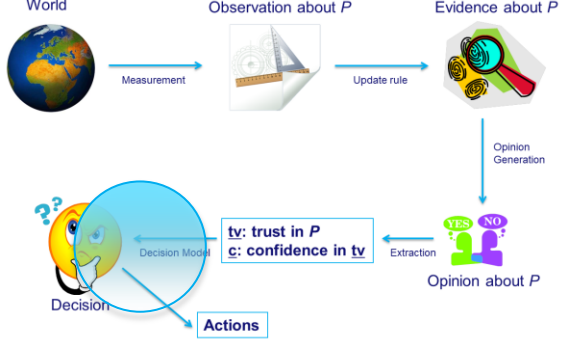
$$\gamma = 1.0, \delta = 0.25, \zeta = 0.7$$

Decision model



- Use weighted sums of trust values (and confidences) and compare with a threshold
- Weights allow to trade trust values of different properties
- Thresholds capture a context
  - e.g. importance (= threshold) of a privacy property goes down upon a medical condition

Decision model



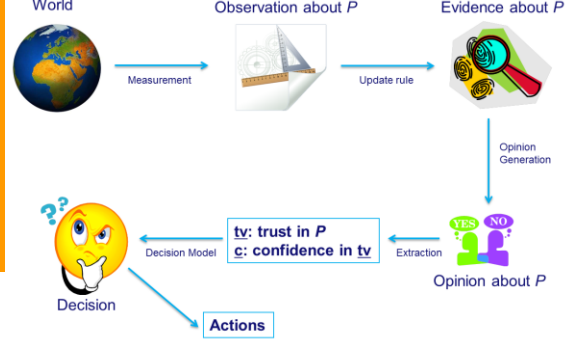
- Maintain trust and confidence values for *n* different trust statements
  - vectors of length *n*: **tv, cv**

- Specify *m* weight vectors of length *n* for both trust and confidence
  - *m* x *n* matrices: **TW, CW**
- Specify *m* thresholds for both
  - vectors of length *m:* **tThr, cThr**

- Safe state:

$$\mathbf{TW\ tv \geq tThr}$$
$$\mathbf{CW\ cv \geq cThr}$$

Example (model)

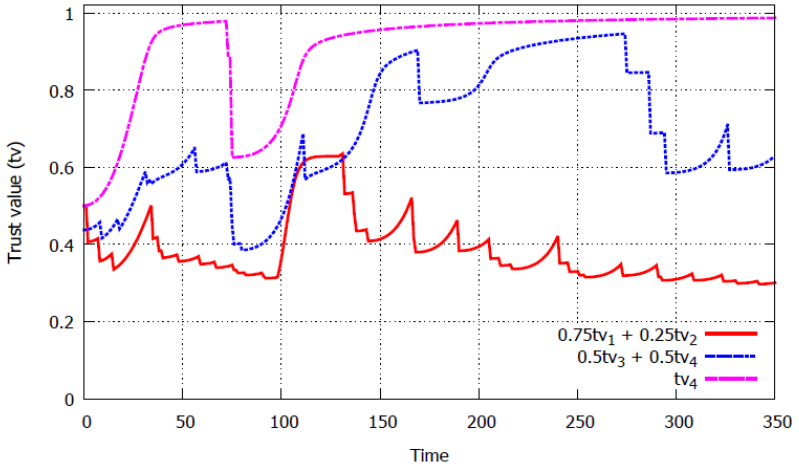


- Monitoring 4 properties, leading to $tv_1, \ldots, tv_4$

p_1: CPU usage is less than 60%.
p_2: Memory usage is less than 40%.
p_3: Network bandwidth usage is less than 25%.
p_4: The degree of privacy is larger than 90%.

- Decision model

$$0.75tv_1 + 0.25tv_2 \geq 0.9$$
$$0.5tv_3 + 0.5tv_4 \geq 0.6$$
$$tv_4 \geq 0.8$$



Related work (see references)

- Monitoring in Sensor Networks for specific goals
  - e.g. reliability of packet forwarding, malicious node detection
- Monitoring for system health as occurs in, e.g., device driver subsystems
- Intrusion detection
- Various applications of subjective logic, based on the simple update rules
- Mode-based resource management

Conclusion & further work

- Proposed a general M&D framework based on monitoring
  - decay
  - uncertainty
  - decision model

- Further work
  - consistency between evidence composition and opinion composition
  - the shape of trust increase should not be convex (‘the foot’)
  - machine learning of cues
  - elaborate on the role of context

- **Questions?**