A Robust Reputation System for P2P and Mobile Ad-hoc Networks

P2PEcon 2004

Sonja Buchegger, Jean-Yves Le Boudec

June 4, 2004

Presentation Outline

- Problem: Misbehavior in P2P and Mobile Ad-hoc Networks
- Proposed Solution: Misbehavior Detection and Reputation System
- Attacks on the Reputation System
- Performance Evaluation: Simulation Results
- Conclusions

Problem Statement

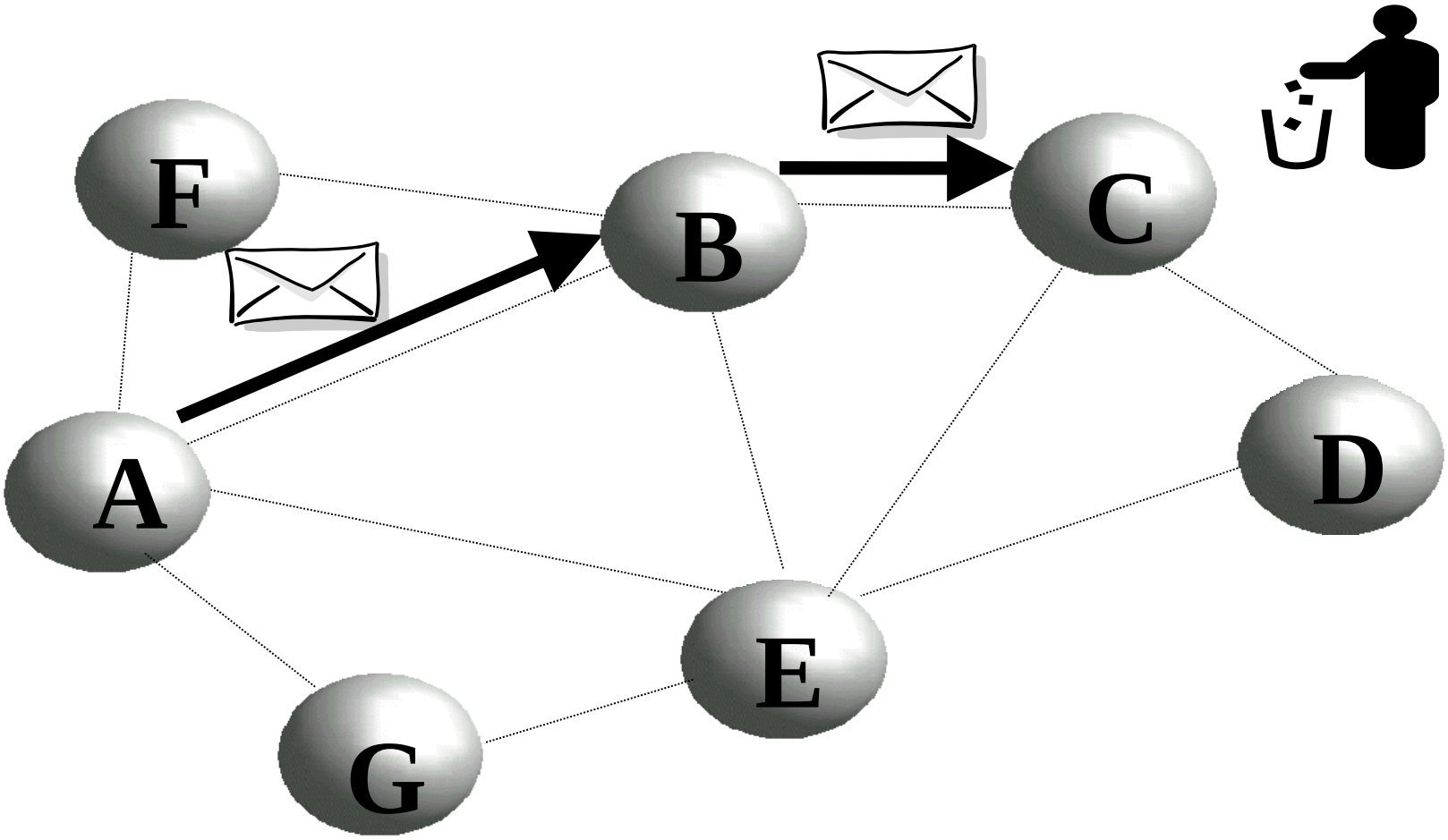
- How can we make a system work despite misbehavior?
- Example: Routing in mobile ad-hoc networks

  Misbehaving nodes decrease performance. They can be
  - Selfish
    - Example: No or incorrect forwarding
  - Malicious
    - Example: Route deviation
  - Faulty
    - Example: Repeating packets

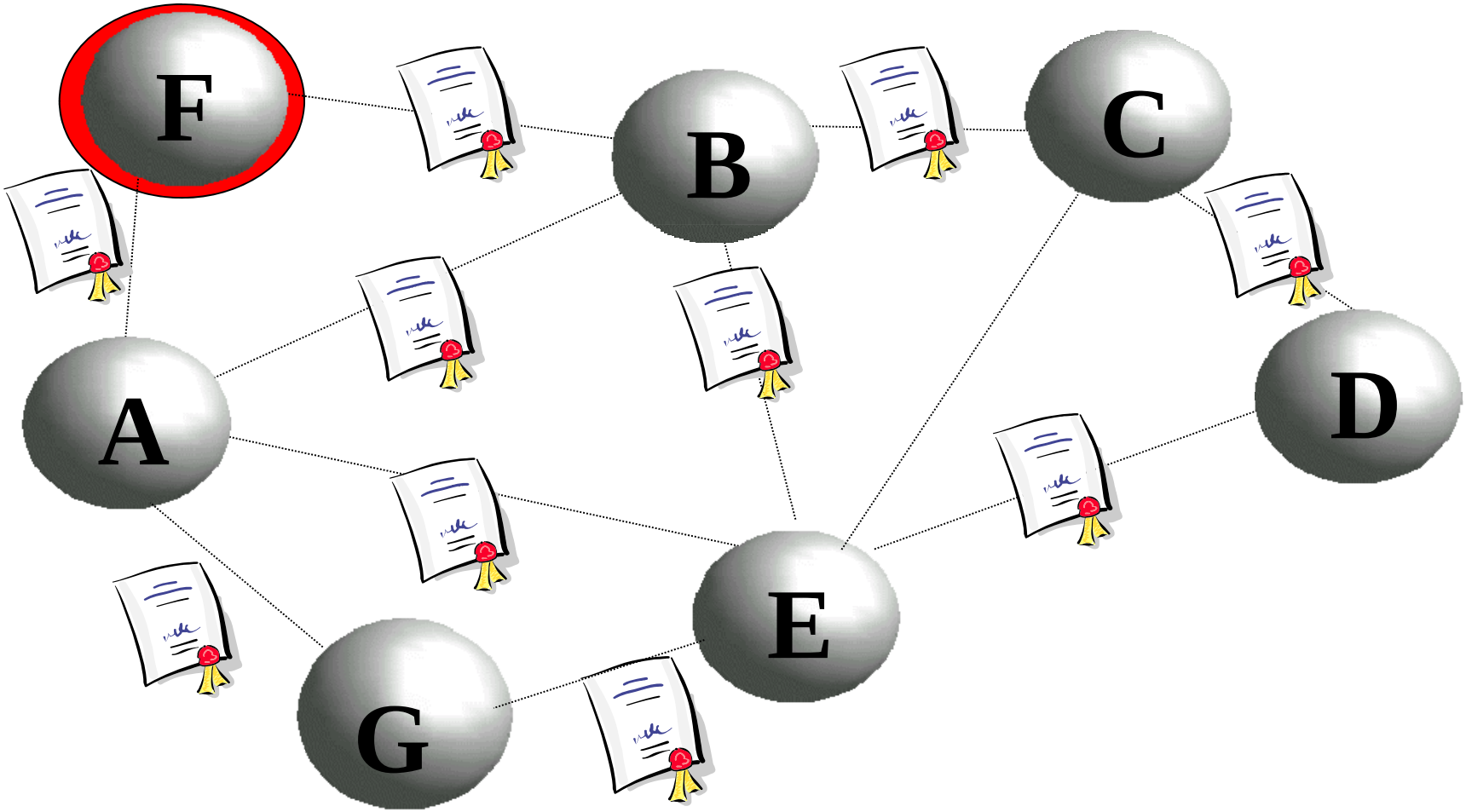
Solution Proposal: Misbehavior Detection and Reputation System

- Nodes monitor their neighbors and keep track of good and bad behavior, using a Bayesian approach
- Periodically they exchange this information.
- The information from others, if *compatible* or *trusted*, slightly modifies the view a node has about another. Compatible = passing a deviation test
- Nodes classify others as normal or misbehaving according to their protocol behavior, and as trustworthy or untrustworthy according to the compatibility of their information.
- Nodes classified as misbehaving are excluded.

Misbehavior



Publication



Reputation

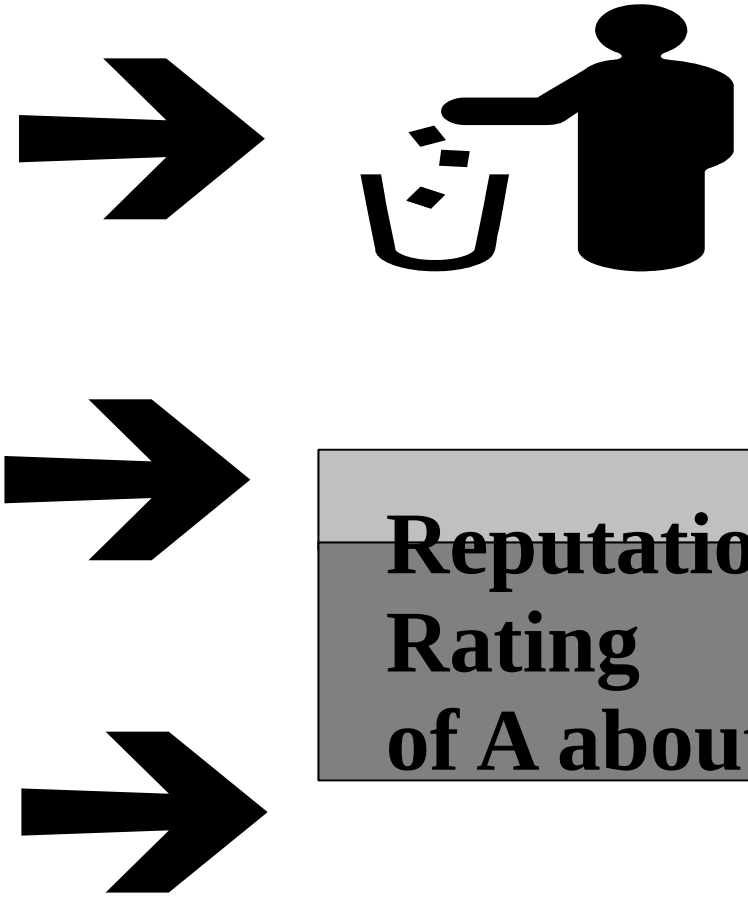
2nd Hand Info from F about C

2nd Hand Info from E about C

2nd Hand Info from B about C

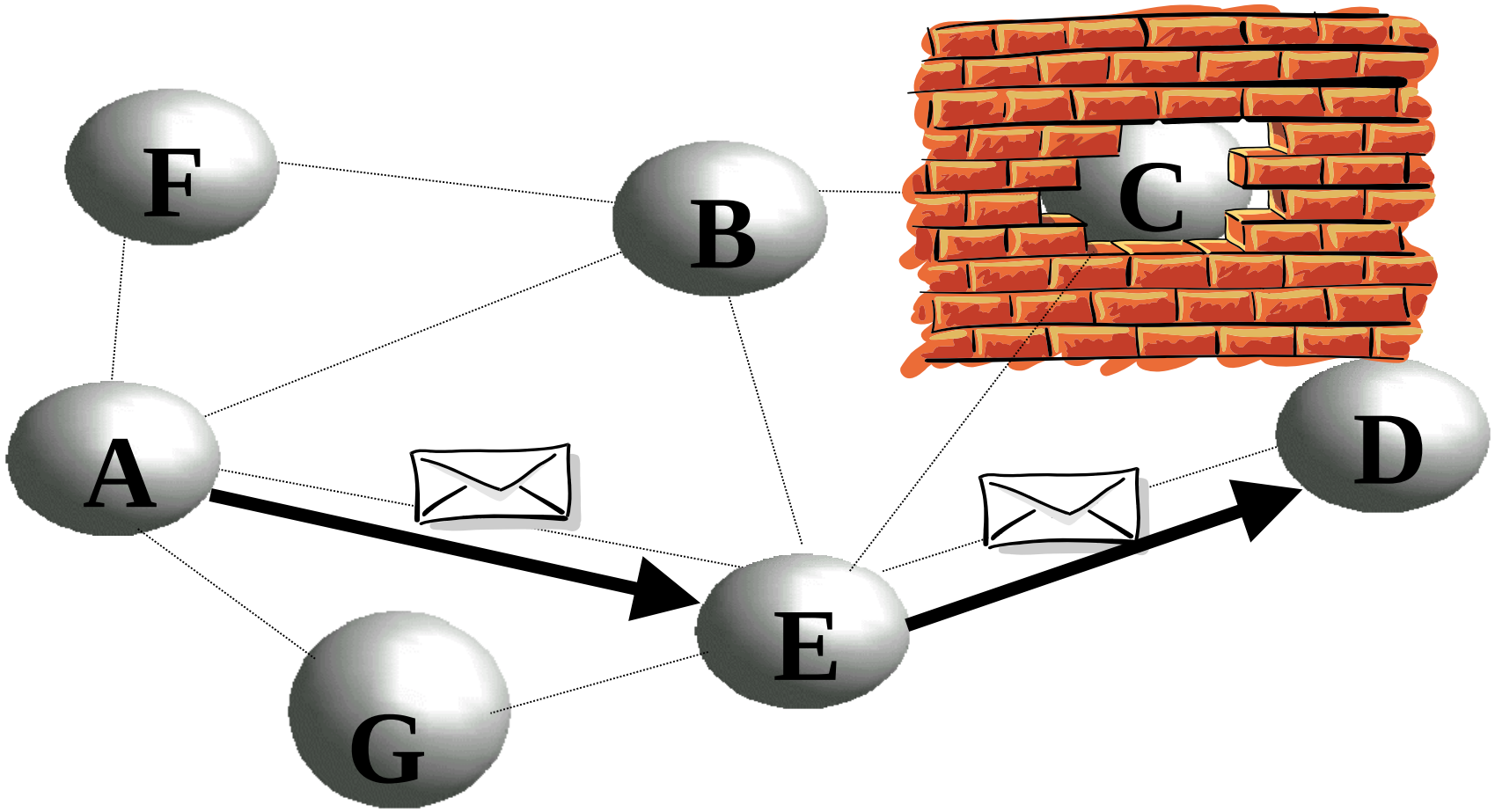
2nd Hand Info from G about C

Previous Rating of A about C



Reputation Rating of A about C

Isolation and Rerouting



Nodes Keep 3 Types of Ratings

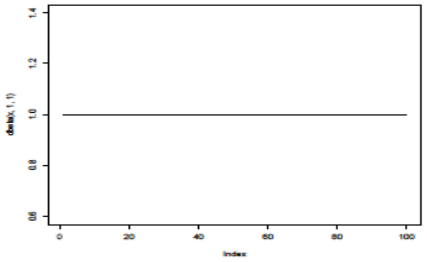
- *First hand* :
  - based on directly observed behavior, if any
- *Reputation* :
  - based on own and received reports
  - synthesis of first and second hand information
  - Reflect perceived behavior of other nodes as **protocol agents**
  - Used to classify others as **normal/misbehaving**
- *Trust* :
  - Based on received reports
  - Reflects perceived behavior of others as **reputation agents**
  - Used to classify others as **trustworthy/untrustworthy**

Bayesian Approach

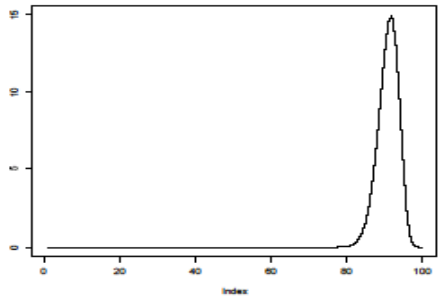
- The three types of records are based on the same Bayesian approach; we explain it for First Hand Observation
- Node i believes that node j misbehaves with probability θ
- θ is drawn from a B(α, β) distribution (= the prior probability)
- The prior B(α, β) is the record B(1,1)
- Modified by Bayesian inference + aging mechanism (fading). Let s = 0 or 1 (misbehave) be the result of one observation. Then (with u = 0.999)

$$\alpha := u\alpha + s$$
$$\beta := u\beta + (1 - s)$$

- Beta function is probability distribution of θ
- θ close to 0 is good, close to 1 is bad

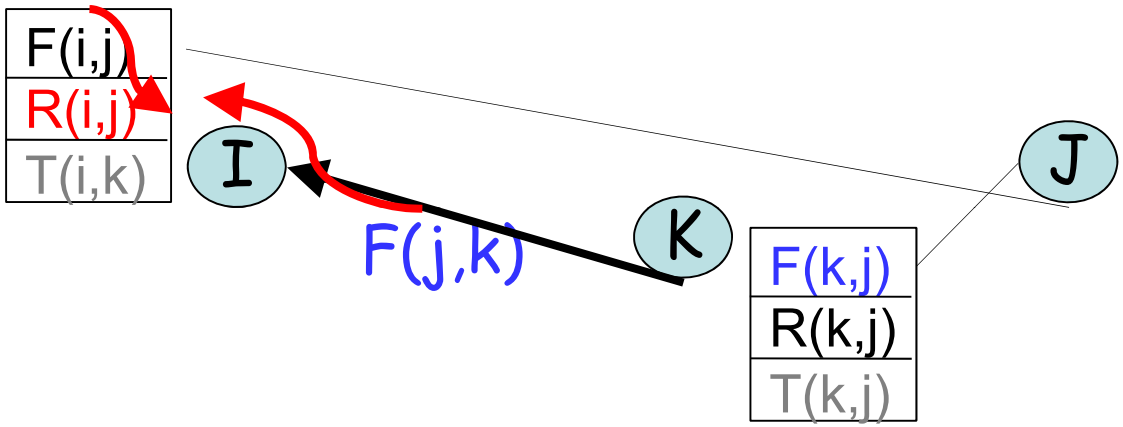


B(1,1)
Node i has no information on node j

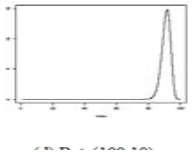


B(100,10)
Node i has conclusive information on node j

Reputation Ratings: First and Second-Hand



$F(j,k) = (\alpha, \beta) =$



- All nodes locally broadcast their first hand observations and send it to the source
- Node i uses own + received first-hand information to build its *reputation* record for node k

How Reputation is Made

- Node i makes direct observation on j
  - R(i,j) is updated according to

$$R_{i,j} = (\alpha', \beta')$$

$$\alpha' := u\alpha' + s$$

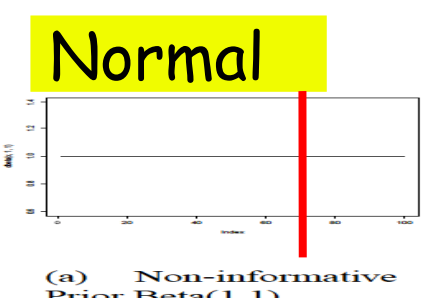
$$\beta' := u\beta' + (1 - s)$$

- Node i receives from node k the first hand observation report F(k,j)
  - Should received report F(k,j) be accepted (see later) ?
  - If no, node i does *not* update reputation recc
  - If yes, i updates reputation record R(i,j), usi

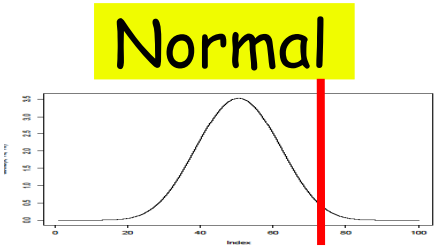
$$R_{i,j} := R_{i,j} + wF_{k,j}$$

- Node i classifies node j according to Bayesian control (with r = 0.75) as

$$\begin{cases} \text{Normal} & \text{if } \mathbb{E}(\text{Beta}(\alpha', \beta')) < r \\ \text{Misbehaving} & \text{if } \mathbb{E}(\text{Beta}(\alpha', \beta')) \geq r \end{cases}$$



(a) Non-informative Prior Beta(1,1)



(c) Beta(10,10)

Deciding Whether to Accept Received Report

- Node i receives from k the (alleged) report F(k,j)
  - **if** i thinks that k is *trustworthy* **then** F(k,j) is accepted
  - **else** i does a *deviation test:*
    if report F(k,j) deviates largely from R(i,j) **then** reject **else** accept

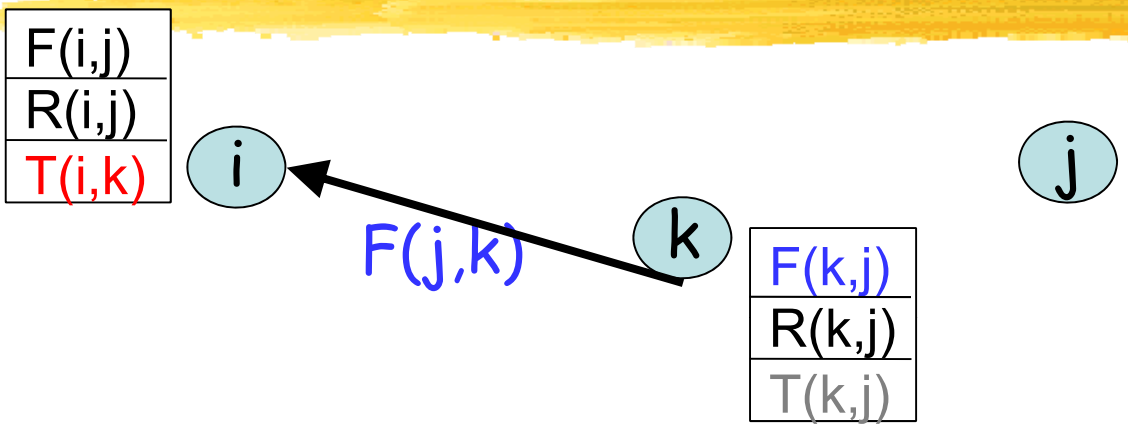
- Deviation Test is $|\mathbb{E}(\mathrm{Beta}(\alpha_F, \beta_F)) - \mathbb{E}(\mathrm{Beta}(\alpha, \beta))| \geq d$

  with d =0.5

Features of the Reputation System

- Only first hand information is propagated
- All information needs to be re-enforced otherwise it fades out (reputation fading), allows redemption
- Separate behaviour as
  - Protocol agent
  - Reputation agent

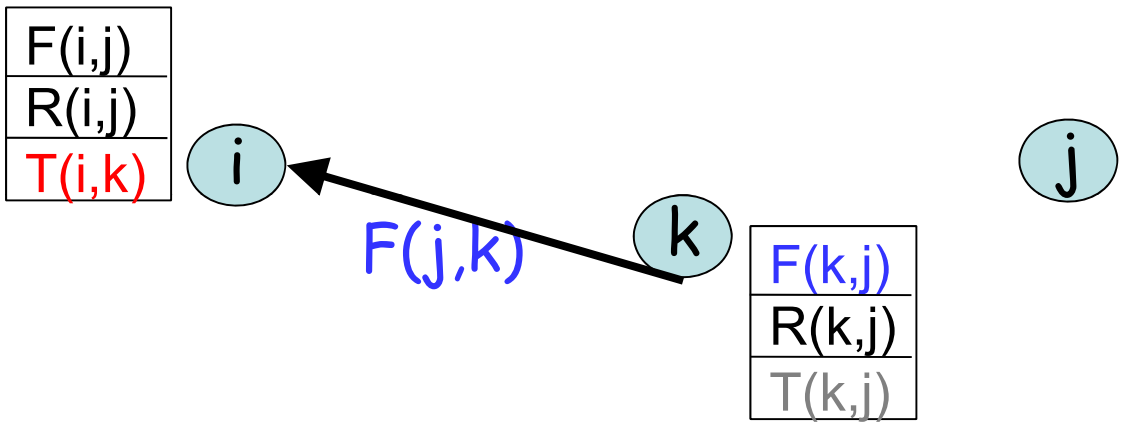
How Trust is Made



- Node i maintains a *trust rating* T(i,k): a rating of k as reputation agent
  - T(i,k) expresses how much k deviates from majority opinion, according to i
- Let T(i,k)=B(γ, δ); i decides that k is (with t =0.25 to 1)

$$\begin{cases} \text{Trustworthy} & \text{if } \mathbb{E}(\text{Beta}(\gamma,\delta)) < t \\ \text{Untrustworthy} & \text{if } \mathbb{E}(\text{Beta}(\gamma,\delta)) \geq t \end{cases}$$

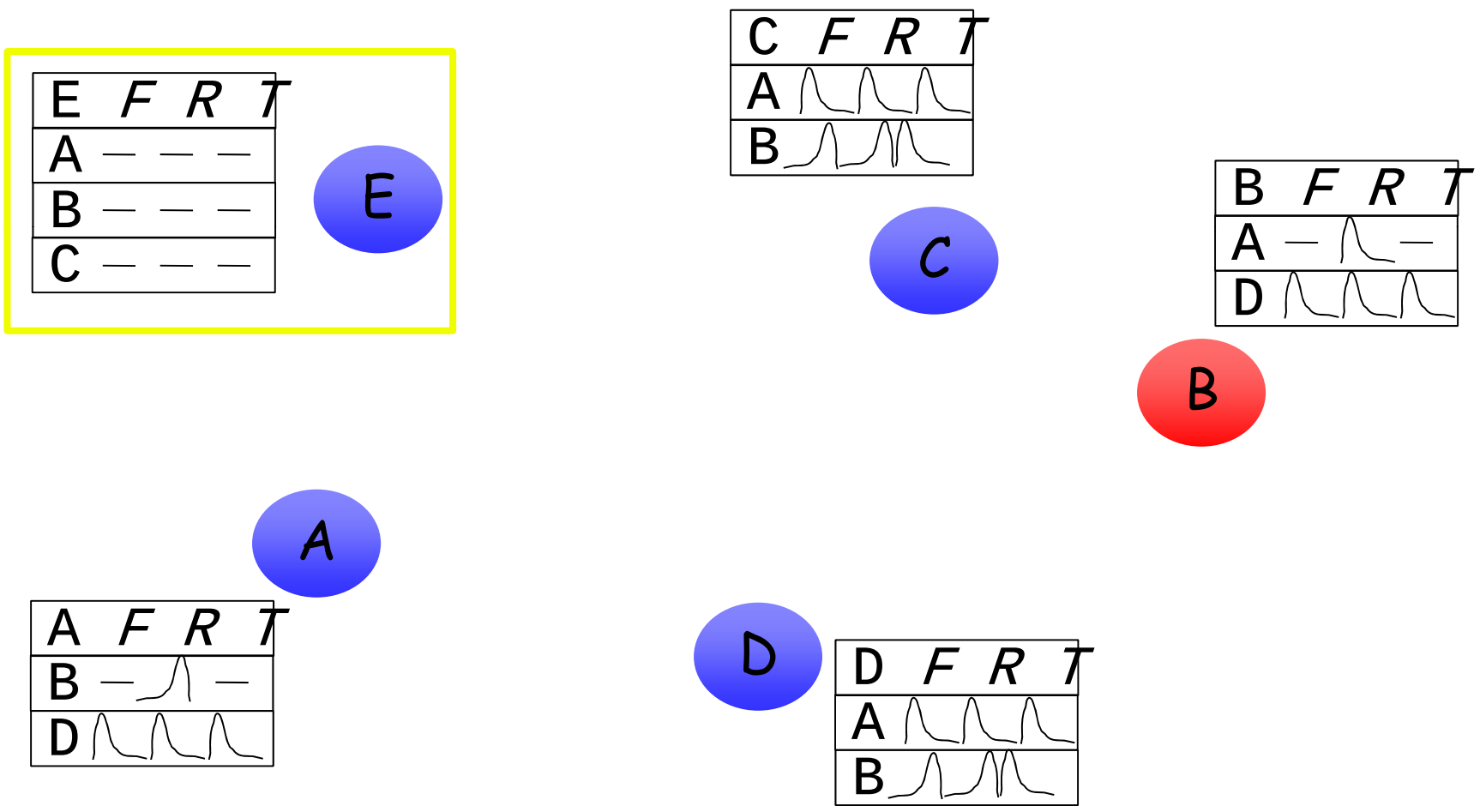
How Trust Rating is Computed



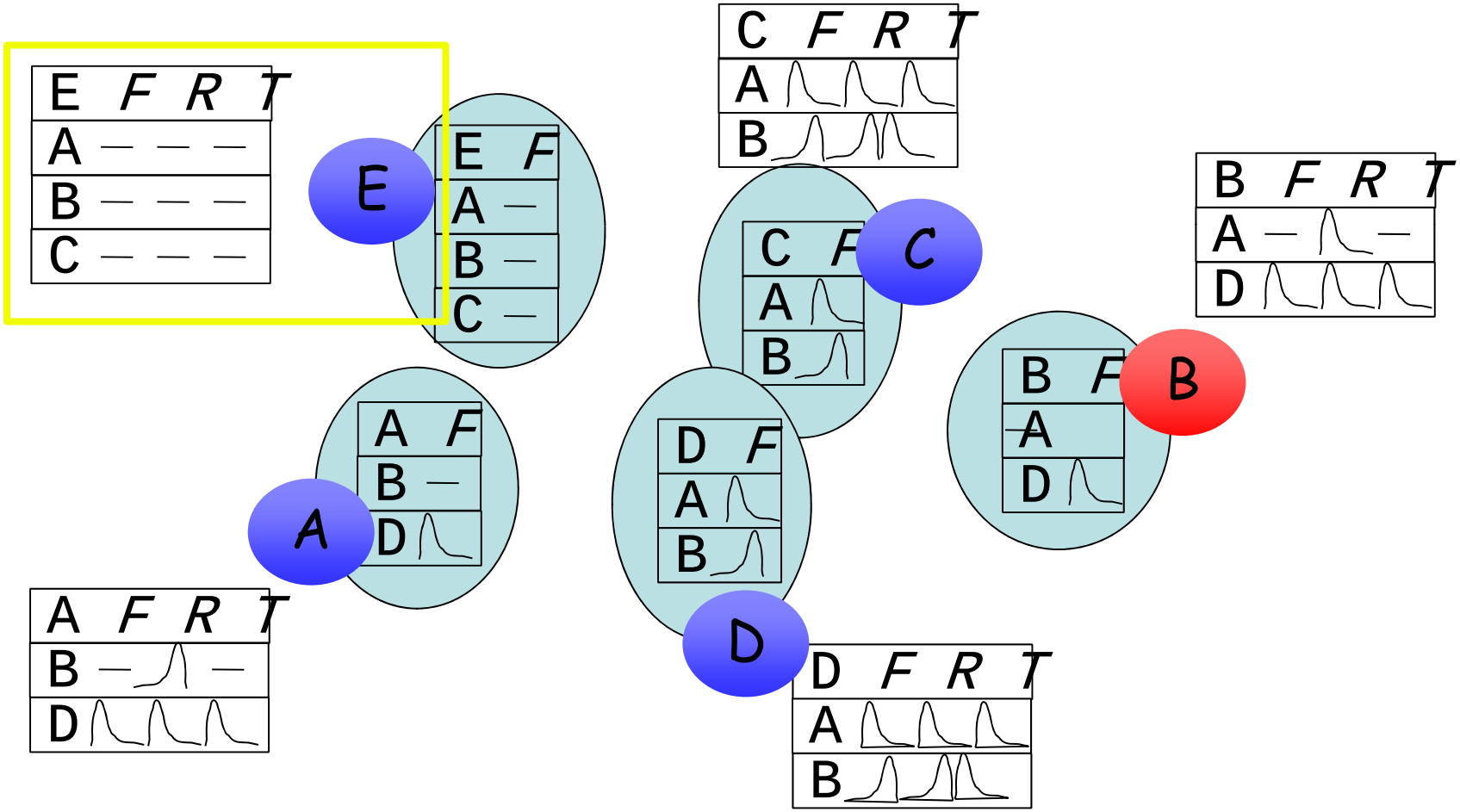
- Node i performs deviation test on received report; (s = 1 means the test is positive).
- T(i,k) = (♑ ♎ ♎) is updated accord (with v = 0.999)

$$\gamma := v\gamma + s$$
$$\delta := v\delta + (1 - s)$$

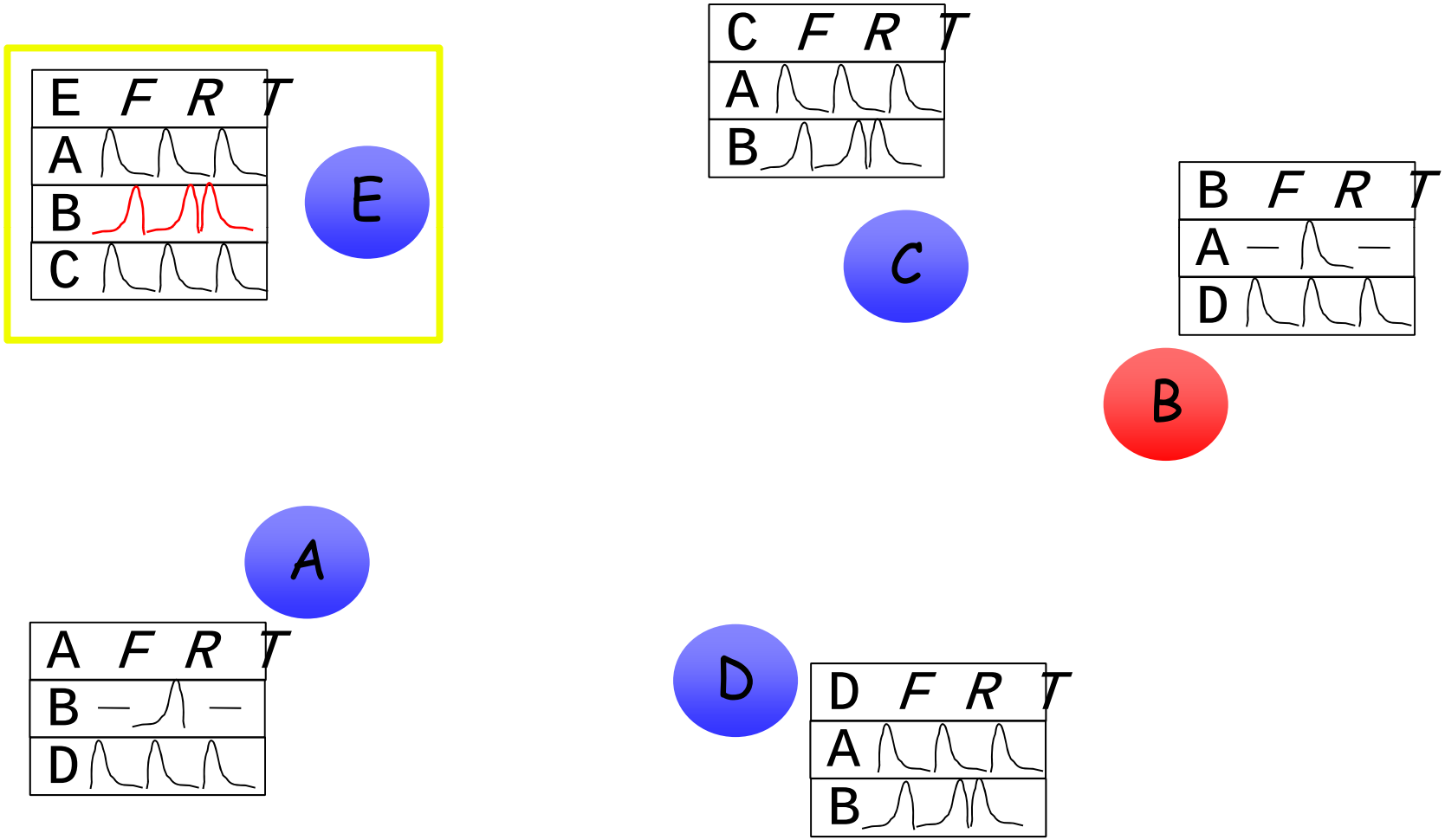
Example: E Joins Network, B misbehaves but does not lie



Nodes Publish First Hand Information



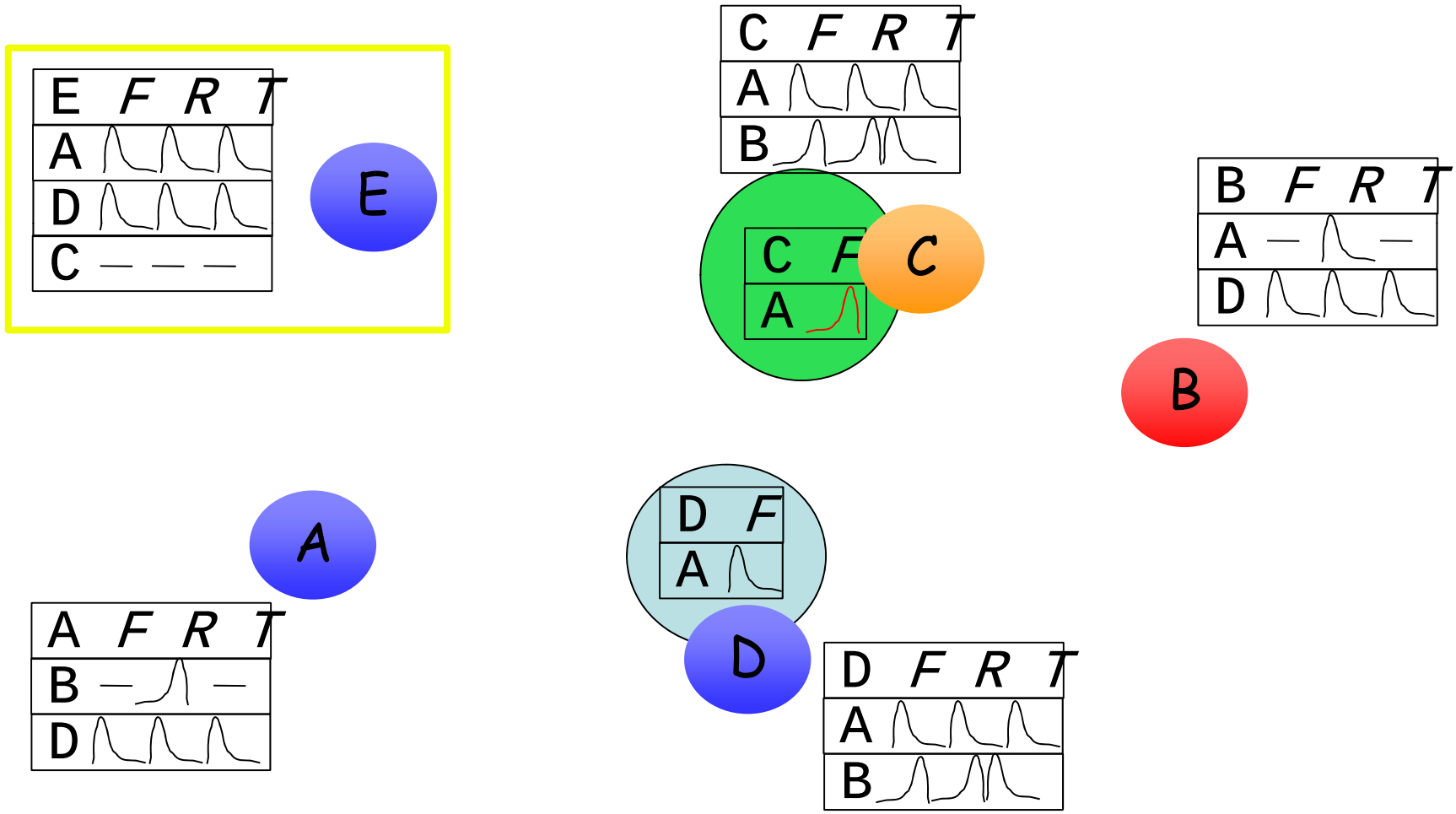
E Accepts Reports and Classifies B



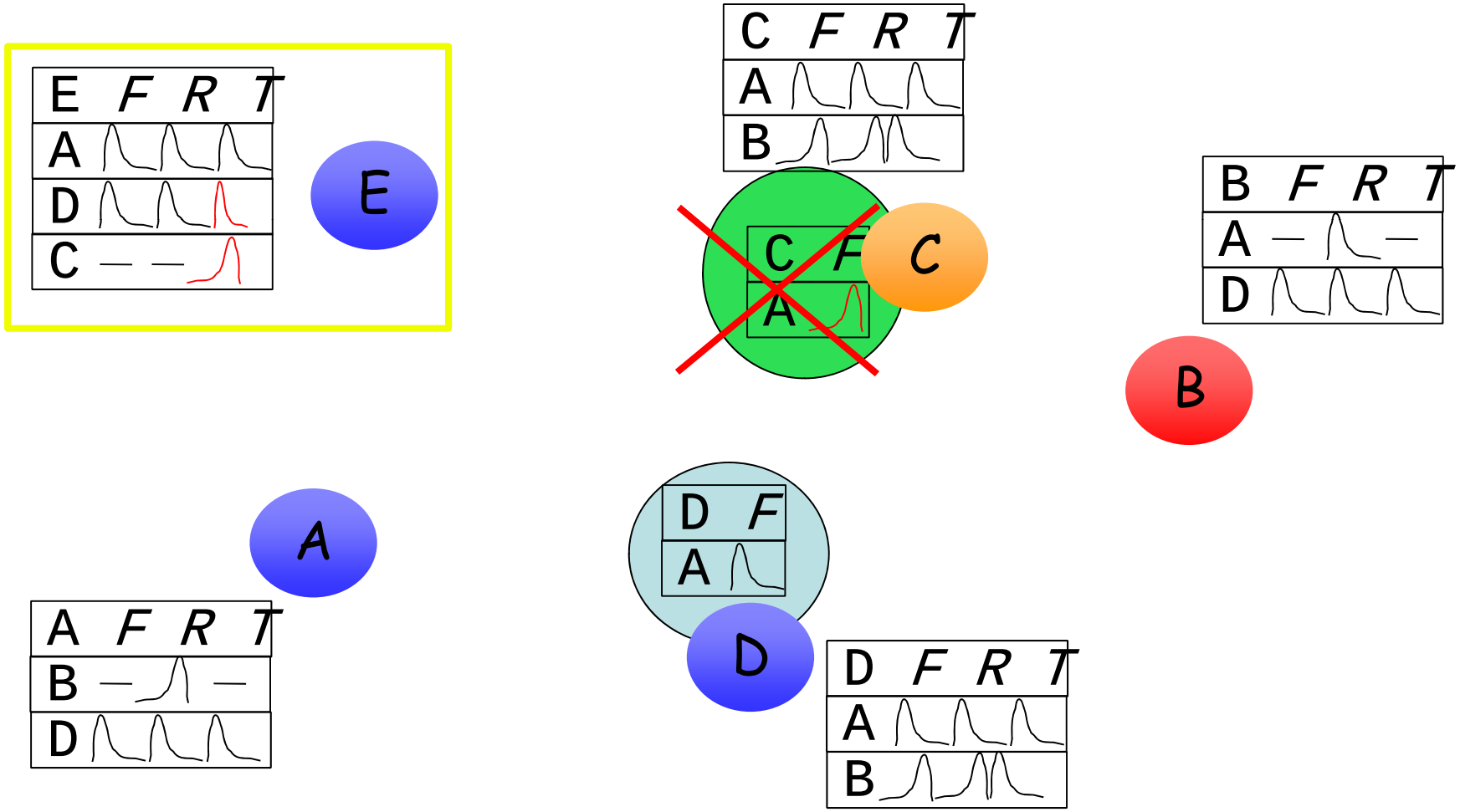
Dealing with Inconsistent Misbehavior

- Redemption
  - Excluded nodes cannot improve their rating (no observable behavior)
  - Reputation fading allows for redemption of
    - Misclassified nodes
    - Formerly faulty nodes
- Secondary Response
  - When excluded node re-enters the network
  - Lower thresholds for misbehavior tolerance
  - Faster classification and isolation

Attacking the Reputation System: Big Lies. Node C comes and lies



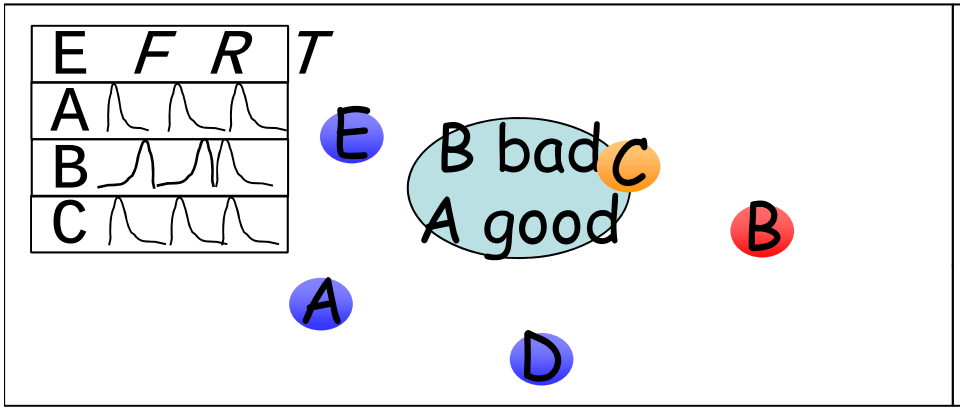
E does Deviation Test and Does Not Believe C



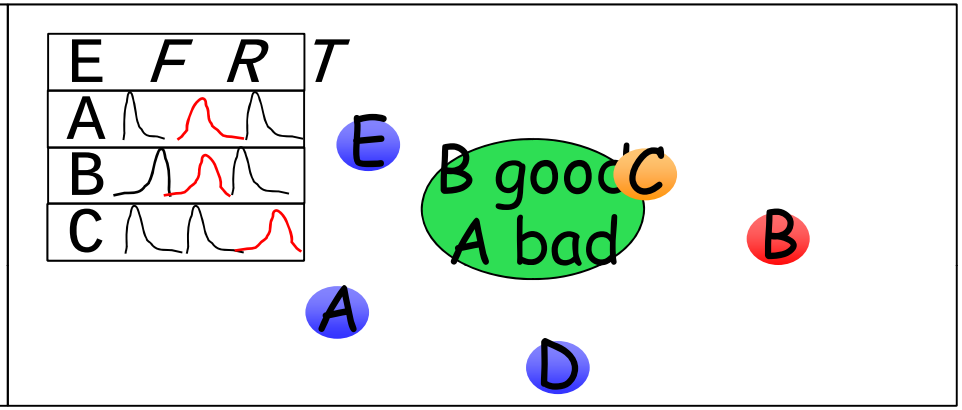
Attacking the Reputation System: Stealthy Lying

- Attackers lie a little bit to gradually worsen reputation
  - But the many little lies do not accumulate (reputation fading)
  - For a more effective attack, liars need to lie more
  - They will not be trusted anymore

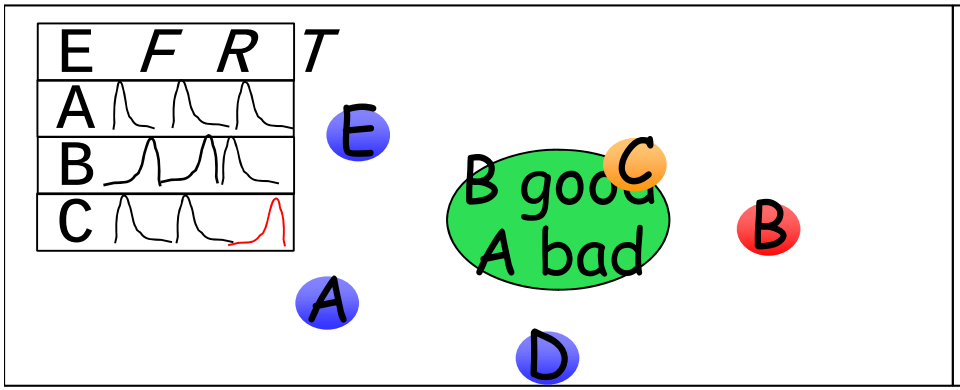
Attacking the Reputation System: Gain Trust and Then Lie



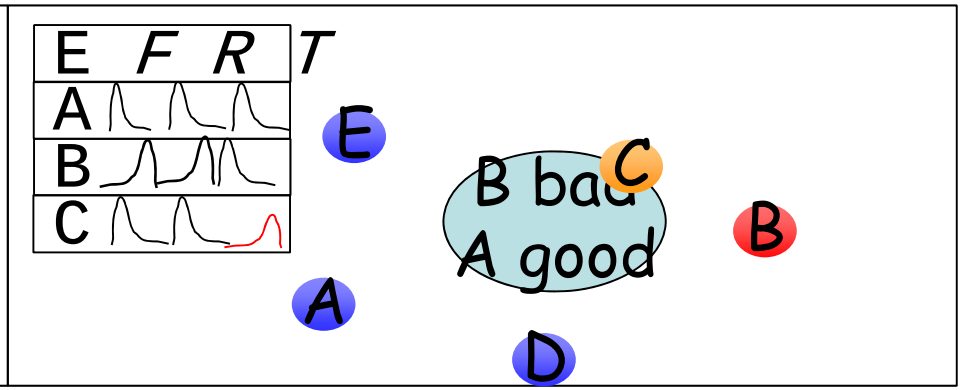
1. C says the truth to gain confidence; E trusts C



2. C says B good A bad; E believes C (and A and D) ang gets blurred picture on A and B. C's trust decreases

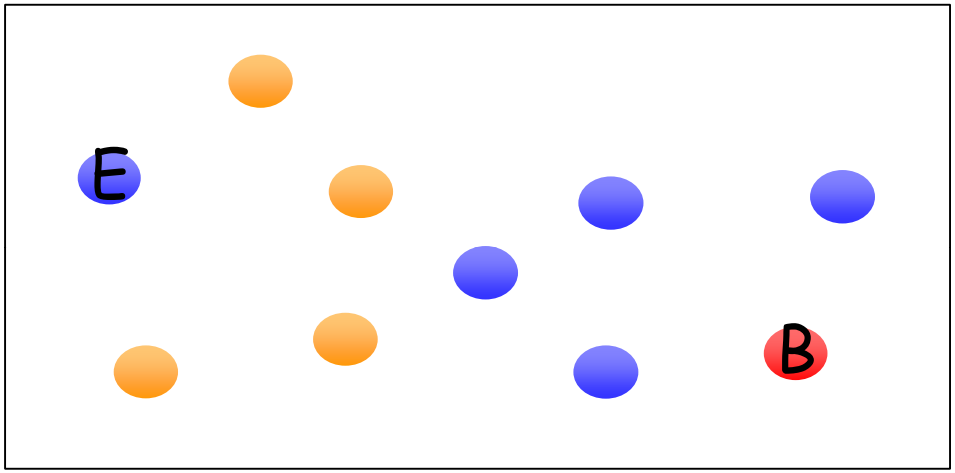


3. C continues to lie; E does not believe C



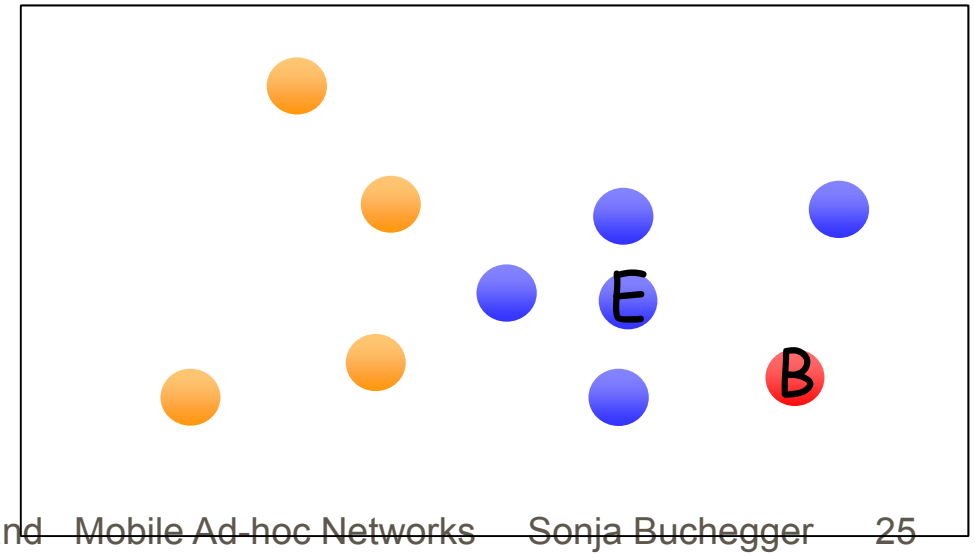
4. C says the truth again; trust becomes better.

Attacking the Reputation System: Brainwashing



1. E surrounded by liars.
No first hand information on B
E believes and trusts liars
« brainwashed »

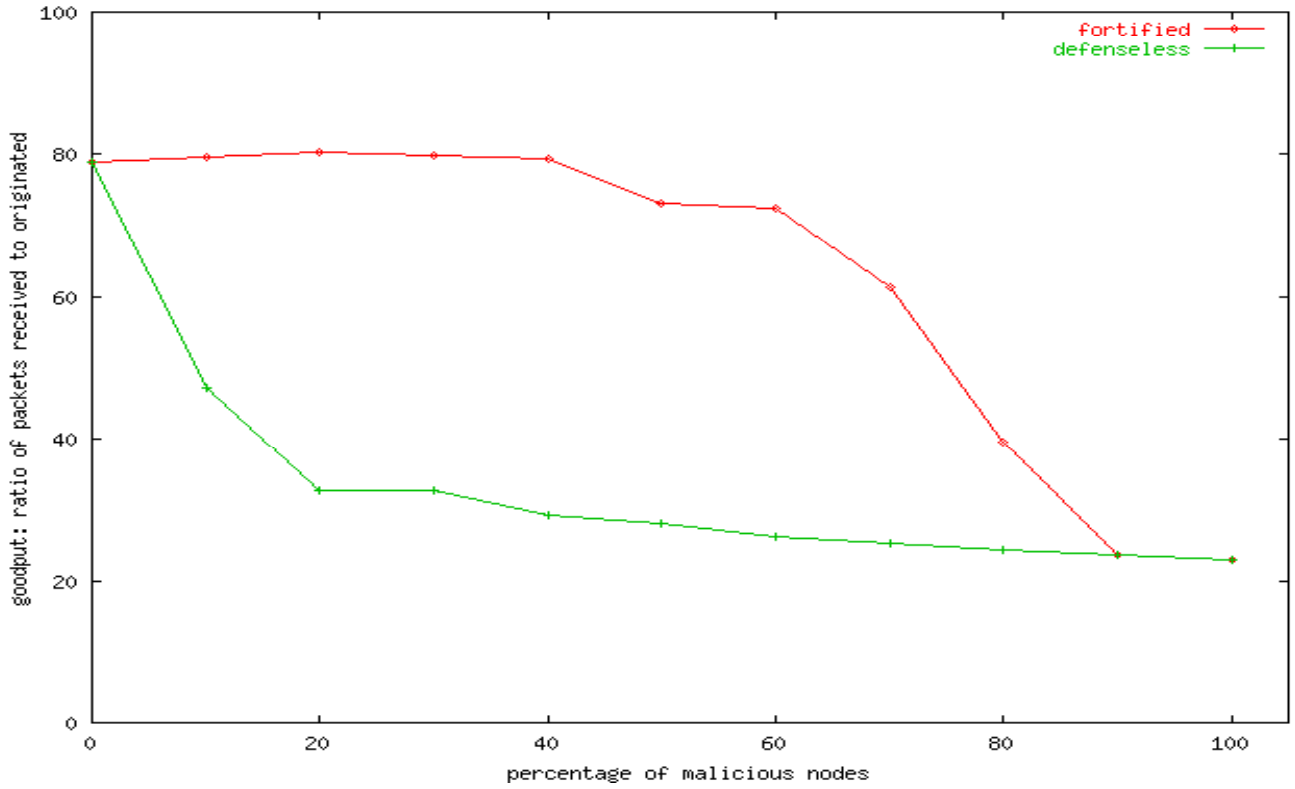
2. E moves to a healthier place. E does not believe the truth but does not update B's reputation either, which fades out. Trust in the liars also fades out. E has now no opinion on B and can start believing the truth.



Simulation using GloMoSim Dynamic Source Routing

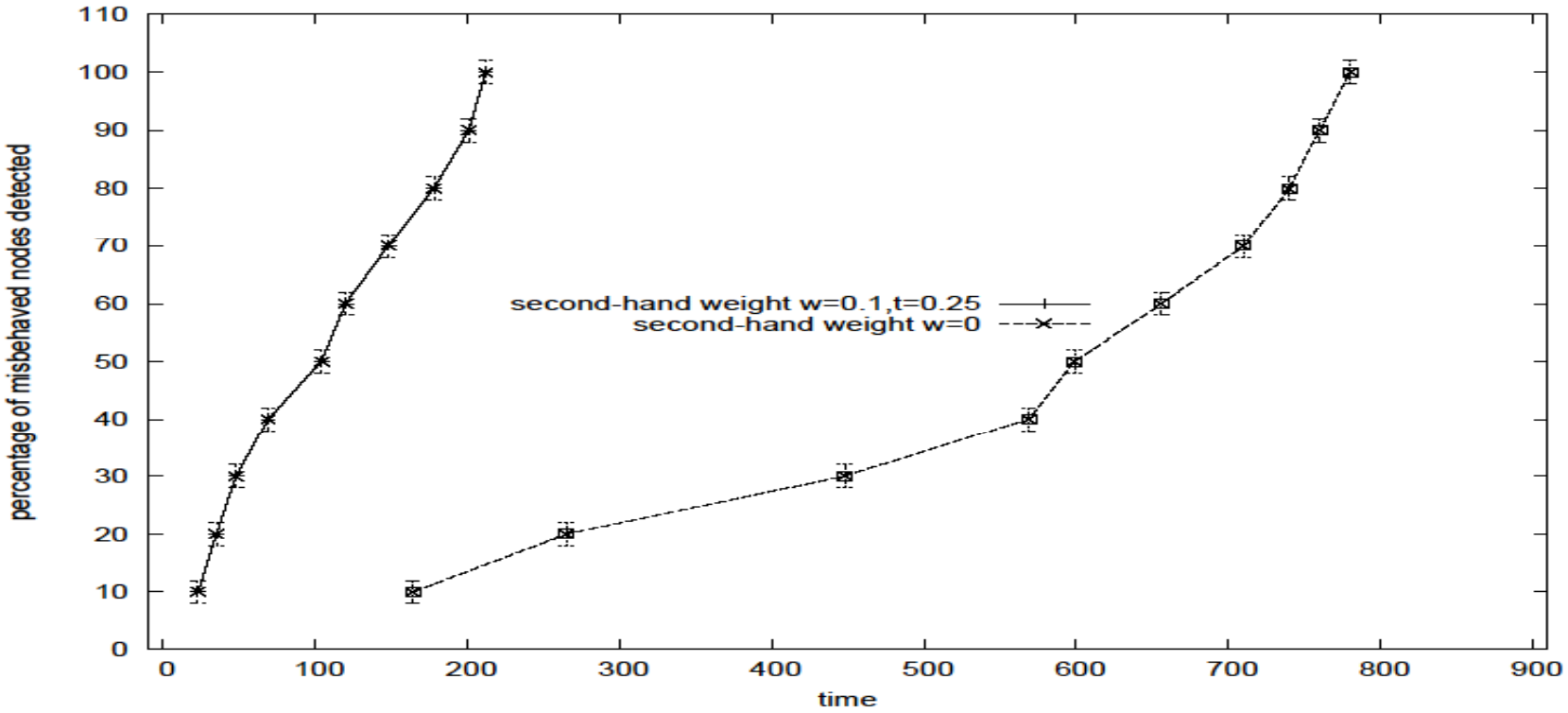
- Goal:
  - Find tolerable number of misbehaving nodes
- Metrics:
  - throughput, dropped messages
  - detection time
  - overhead (control messages)
- Scenarios:
  - infiltrated vs. benign network
  - defenseless vs. fortified network
- Misbehavior type:
  - no forwarding
  - lying

Throughput vs. % Misbehaving Nodes



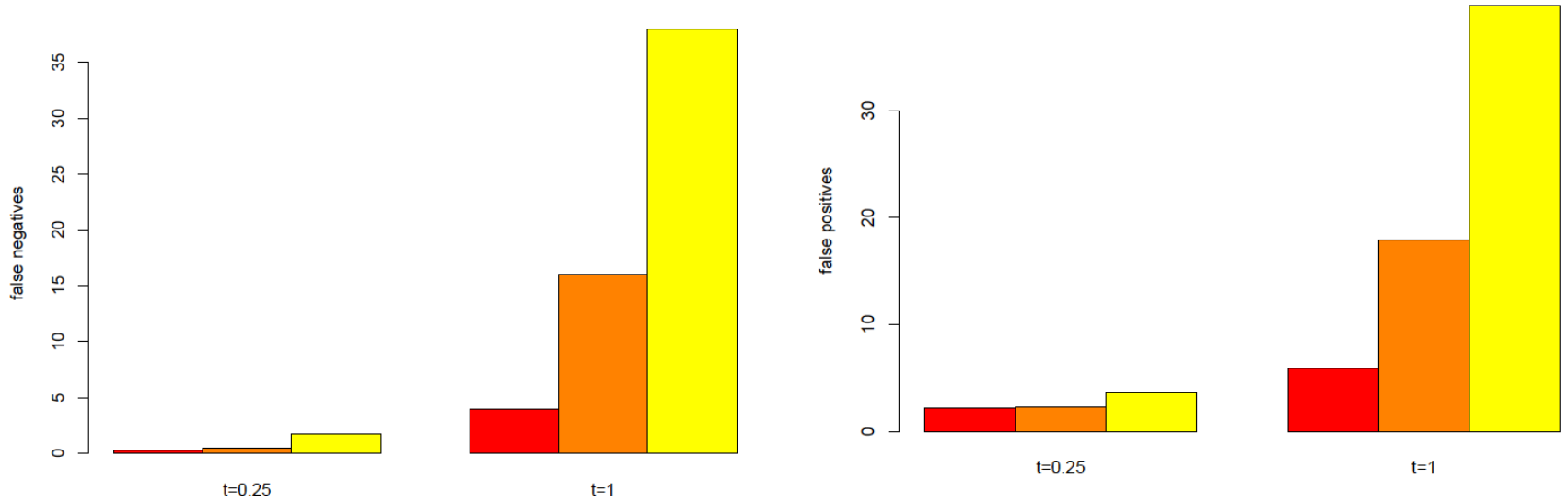
- Copes well with up to 50% misbehaving nodes (no forwarding attack)

Effect of Reputation: Detection Speed



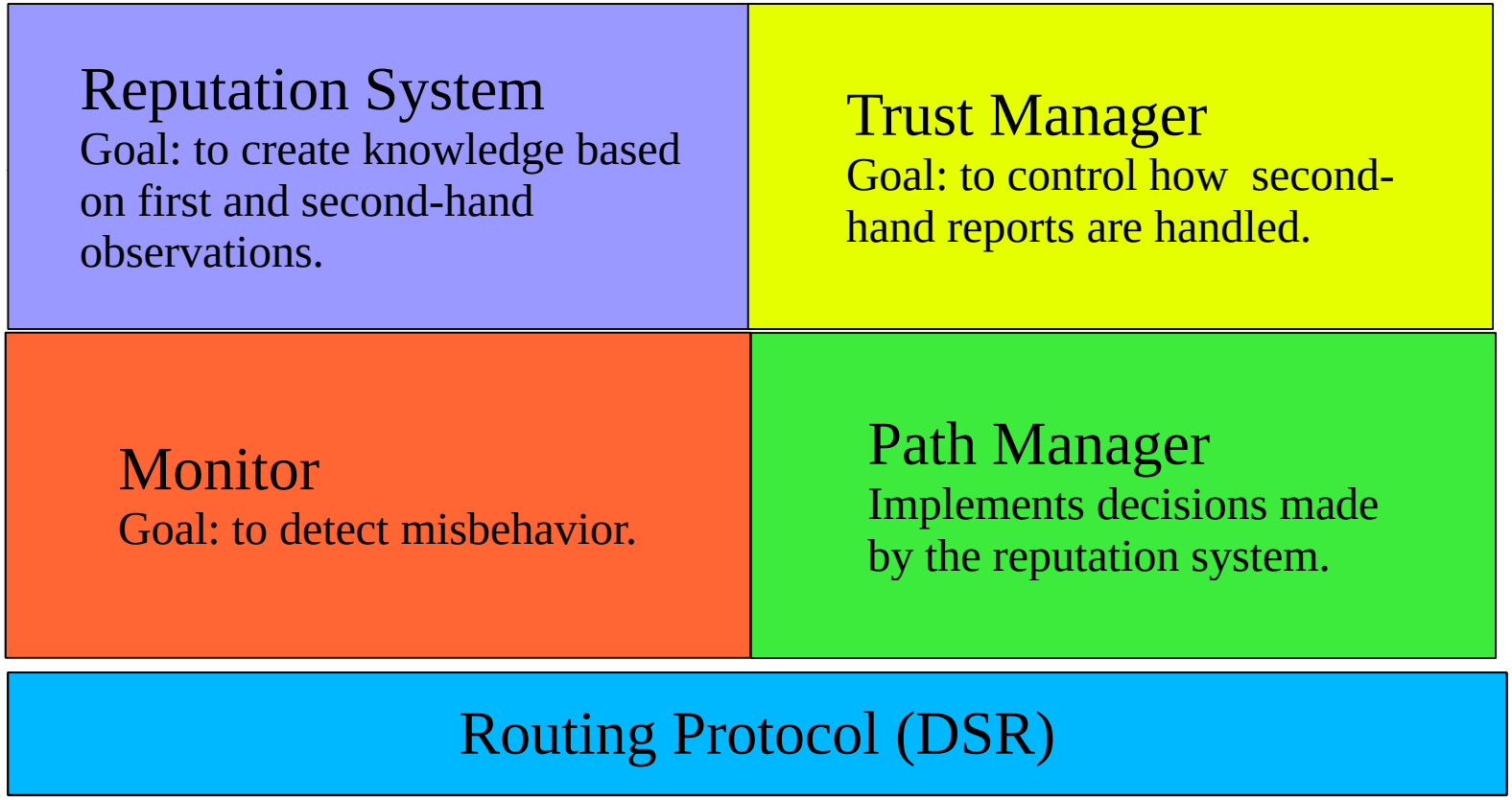
- Second-hand information decreases time to detect all misbehaving nodes.

Effect of Trust (False Pos./Neg.)



- Using trust reduces the number of false positives and false negatives.

Misbehavior Detection and Reputation System CONFIDANT Components/Applications



Future Work

- Application to P2P file sharing, anonymity systems
- Address identity problem
- More detailed attacker model and performance evaluation

Conclusions

- Have to cope with misbehavior
- Detection and reputation systems target a wider range of misbehavior than payment or cryptography approaches.
- Second-hand information helps detection.
- Our reputation systems makes use of it and is robust to spurious ratings of single attackers or small collusions and can recover from big collusions.

Questions?