

# Using Reputation Systems to Reduce Behavior Uncertainty

© Original Artist

Reproduction-rights-obtainable from  
www.CartoonStock.com



"REMEMBER, JUNIOR, IT'S IMPORTANT TO AVOID  
ANY APPEARANCE OF WRONGDOING."

Sonja Buchegger  
Deutsche Telekom  
Laboratories  
[sonja@ieee.org](mailto:sonja@ieee.org)

Joint work with  
Jean-Yves Le Boudec  
EPFL

# Outline

- Reputation Systems (cf. Working group)
- Application example: MANET
- Our approach - CONFIDANT
  - Bayesian estimation
  - Fading
  - Second-hand information
  - Deviation test
  - Secondary response

# Purpose of Reputation

- Incentive for good behavior
  - Shadow of the future
- Prediction of future behavior based on past
  - Shadow of history
- Decision basis for choice of transaction partner and risk assessment, reduce uncertainty about others' behavior

You can't build a reputation on what you're going to do –Henry Ford

# Reputation System Tasks

- Gather information on transactions/behavior history
- Distribute information
- Aggregate into a reputation rating
- Cope with misbehavior, select for expected good behavior
- -> Decision based on rating, context, risk disposition, value, etc.

**Live in such a way that you would not be ashamed to sell your parrot to the town gossip.  
--Will Rogers (1879 - 1935)**

# Reputation System Infrastructure

## ■ Centralized

- EBay
- Amazon
- Consumer Reports

## ■ One reputation rating for each entity

## ■ Distributed

- Mobile Ad-hoc Networks
- Peer-to-Peer Networks

## ■ Local reputation ratings, may be conflicting

**Wise men learn by other men's mistakes, fools by their own.**  
**H.G. Wells**

# Design Issues/Problems

## ■ Identity:

- pseudonymity
- identity persistence
- newcomer's reputation

## ■ What is being rated:

- people vs. system
- overall vs. particular properties

**Why you say you no bunny rabbit when you have little powder-puff tail?**

**-- The Tasmanian Devil**

## ■ Spurious ratings:

- false accusations
- false praise (boosting)
- collusion
- retaliation
- fear of retaliation
- wrong observations

**I didn't do it! Nobody saw me do it!  
Can't prove anything!**

**-- Bart Simpson**

# Design Issues/Problems II

## ■ Redemption

- emphasis on past or recent behavior?
- fading
- secondary response
- influence of outliers
- speed of reputation increase/decrease

## ■ Incentives for Participation

- why contribute to the reputation system? (payment, altruism, expertise)
- esp. negative feedback

**"Honest criticism is hard to take, particularly from a relative, a friend, an acquaintance or a stranger."**

**-Franklin Jones**

# Design Issues/Problems III

- Weighting of Second-hand information
    - compared to own experience
    - likelihood of being true
    - trusted/respected source
    - upper bound on single influence
    - conflicting information
  - Causes for Misbehavior:
    - selfishness
    - malice
    - incompetence/faults
    - -> hard to distinguish
- Never attribute to malice that which can adequately be explained by stupidity.**  
-- R. Feynmann

**Nobody believes the official spokesman... but everybody trusts an unidentified source. --Ron Nesen**

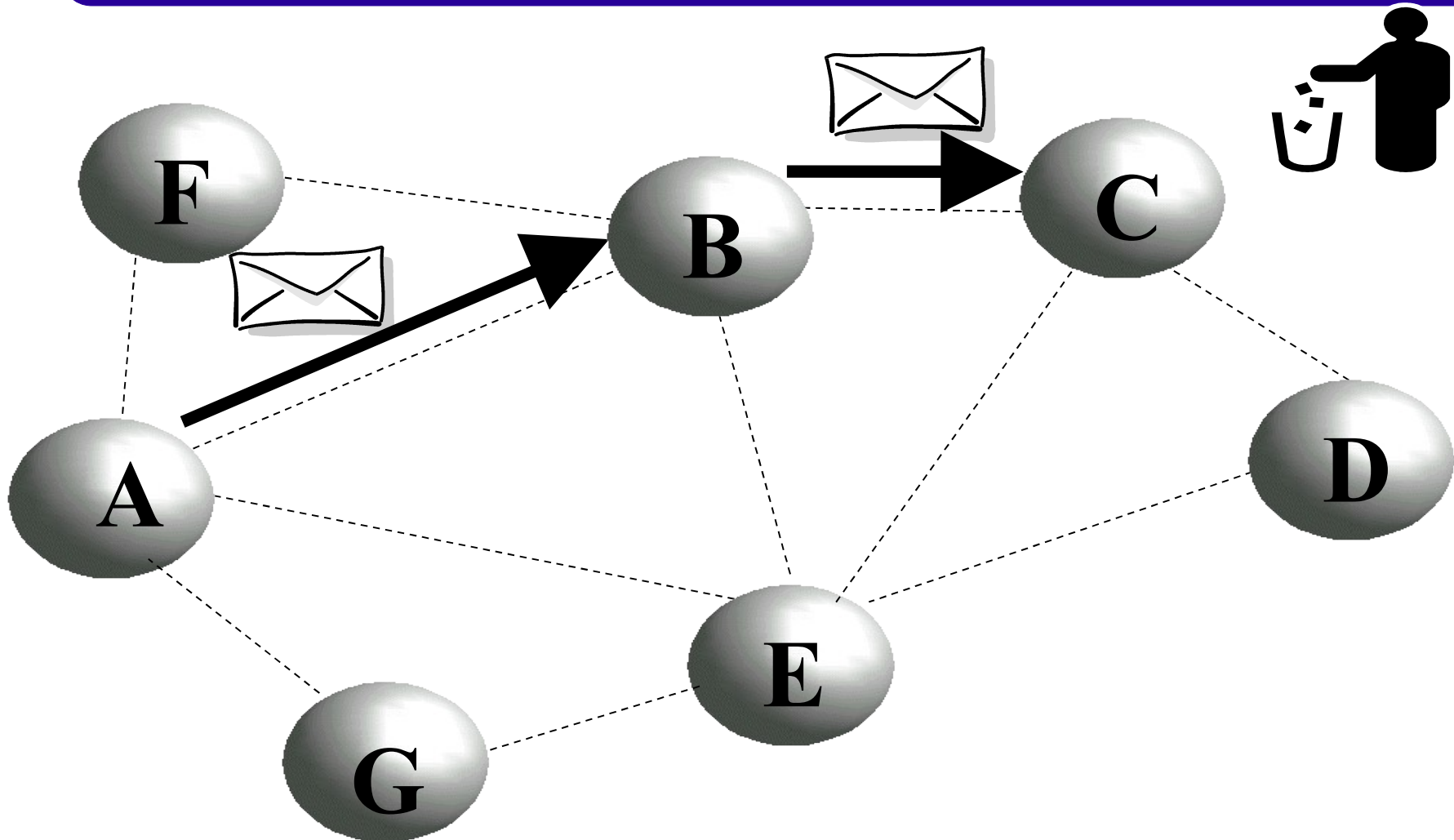
# Problem Statement

- How can we make a system work despite misbehavior?
- Example: Routing in mobile ad-hoc networks
- How to reduce uncertainty about nodes' behavior?
- Misbehaving nodes decrease performance. They can be
  - Selfish
    - Example: No or incorrect forwarding
  - Malicious
    - Example: Route deviation
  - Faulty
    - Example: Repeating packets

# Solution Proposal

- 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.

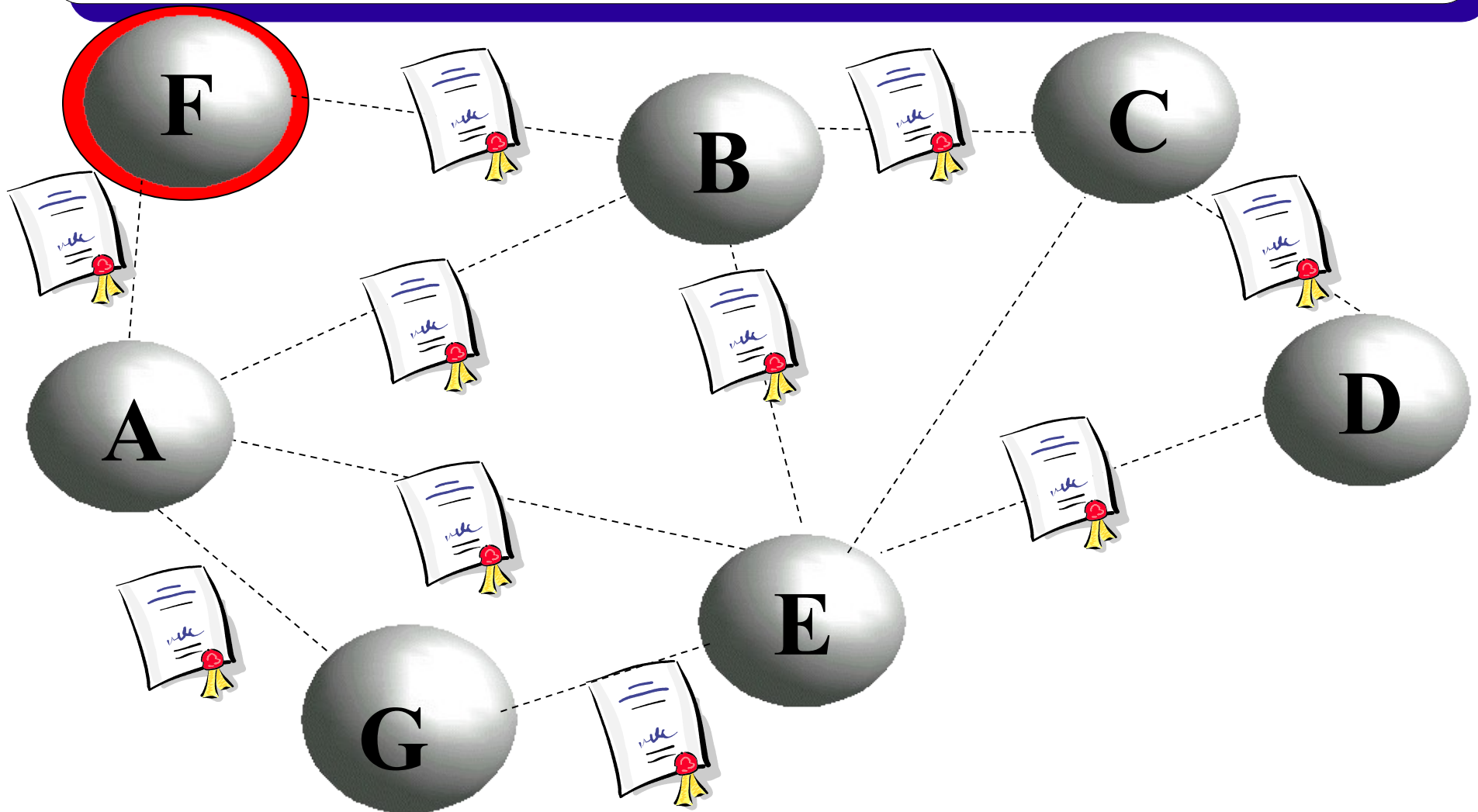
# MANET Misbehavior



If a thing's worth having, it's worth cheating for.

-- W.C. Fields

# Reputation Ratings Publication



# Reputation Rating

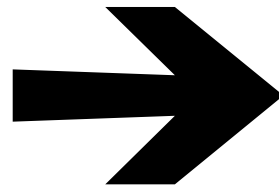
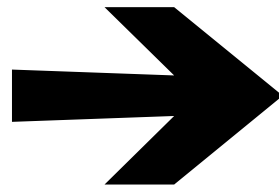
**2<sup>nd</sup> Hand Info  
from F about C**

**2<sup>nd</sup> Hand Info  
from E about C**

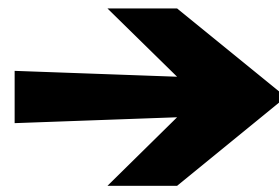
**2<sup>nd</sup> Hand Info  
from B about C**

**2<sup>nd</sup> Hand Info  
from G about C**

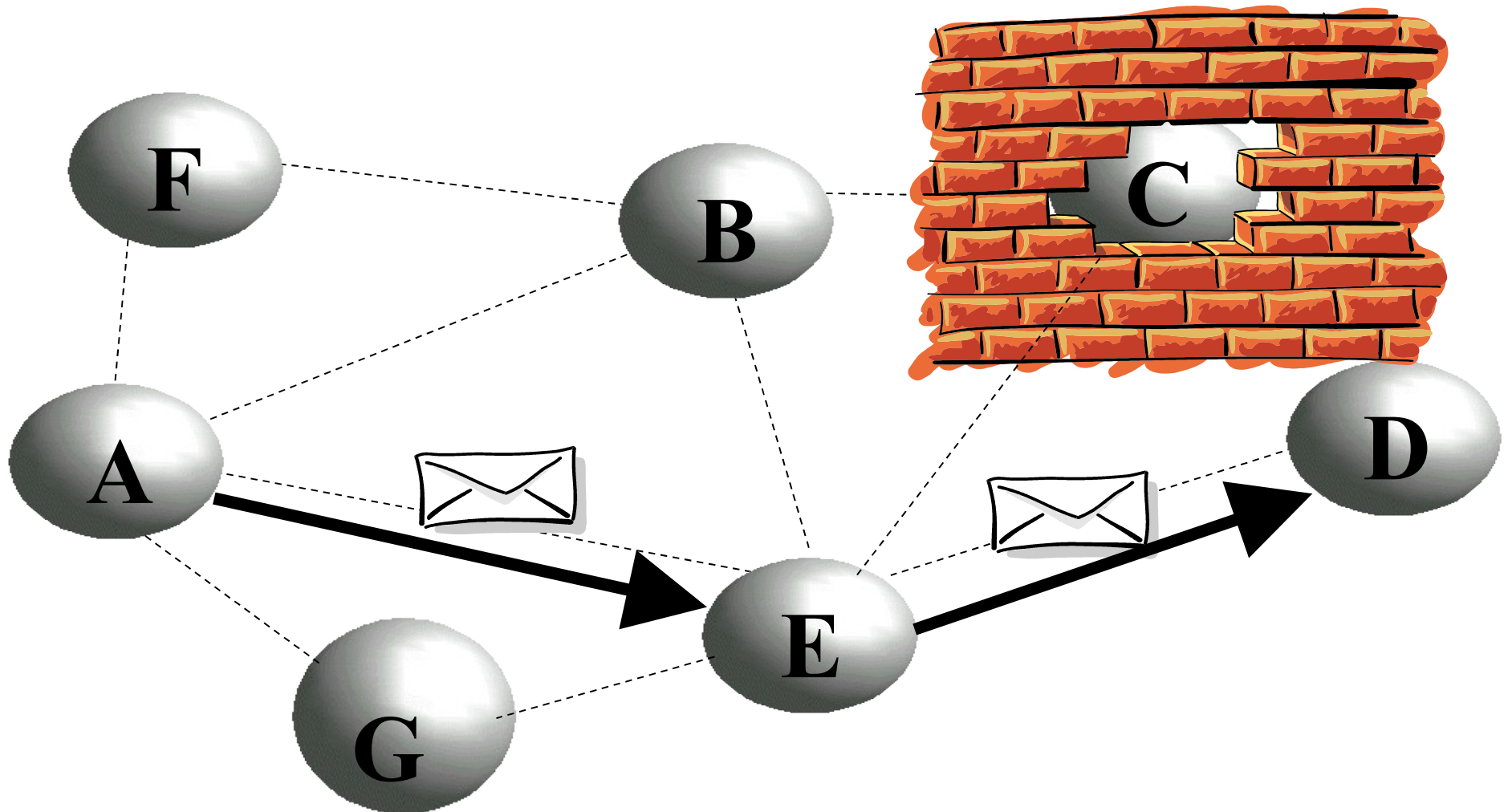
**Previous Rating  
of A about C**



**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
- Reflects 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**

**Our business is run on trust. We trust you will pay in advance.**

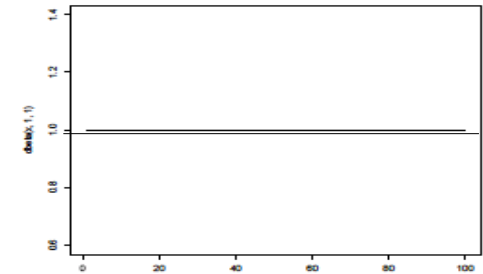
# Bayesian Approach

- 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  $\theta$
- $\theta$  is drawn from a  $B(\alpha, \beta)$  distribution (= the prior probability)
- The prior  $B(\alpha, \beta)$  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  $u = 0.999$ )

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

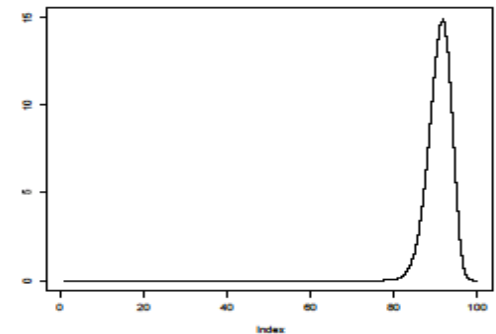
$$\beta := u\beta + (1 - s)$$

- Beta function is probability distribution of  $\theta$
- $\theta$  close to 0 is good, close to 1 is bad



$B(1,1)$

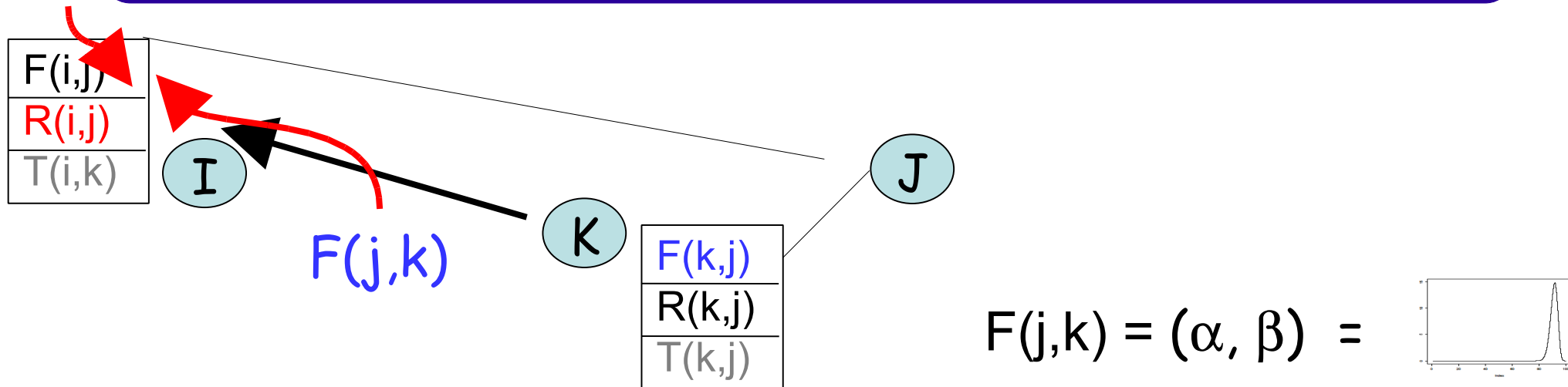
Node  $i$  has no information on node  $j$



$B(100,10)$

Node  $i$  has conclusive information on node  $j$

# Reputation Ratings: First and Second-Hand



- 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  $k$

**Experience is a good teacher, but she sends in terrific bills.**  
— Minna Antrim,

# How Reputation is Made

Node i makes **direct observation** on j

R(i,j) is updated according to

$$R_{i,j} = (\alpha', \beta') \quad \begin{aligned} \alpha' &:= u\alpha' + s \\ \beta' &:= u\beta' + (1 - s) \end{aligned}$$

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 record

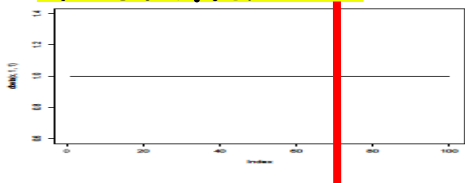
If yes, i updates reputation record R(i,j), using

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

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

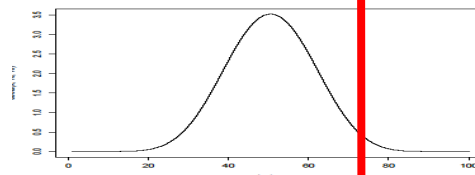
$$\begin{cases} \text{Normal} & \text{if } \mathbb{E}(\text{Beta}(\alpha', \beta')) < r \\ \text{Misbehaving} & \text{if } \mathbb{E}(\text{Beta}(\alpha', \beta')) \geq r \end{cases}$$

Normal



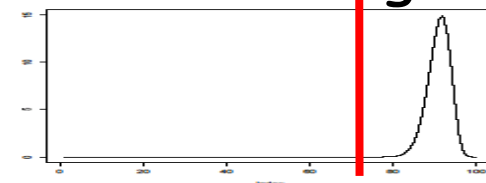
(a) Non-informative

Normal



(b) Beta(10, 10)

Misbehaving



# 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
- Secondary response (inspired by immune system)
- Separate behaviour as
  - Protocol agent
  - Reputation agent

# 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

# Purpose of CONFIDANT

- CONFIDANT detects misbehaving nodes
  - by means of observation or reports about several types of attacks
- 1. and thus allows nodes
  - to route around misbehaved nodes and
  - to isolate misbehaved nodes from the network, so that misbehavior
    - does not pay off,
    - cannot continue, and
    - routes are functional.

# Some Conclusions

- Careful design of a robust reputation system needed, many pitfalls
- Advances in identity persistence might help
- Can reduce uncertainty given some behavior consistency

