Vulnerability Analysis Wireless Sensor Networks: Challenges and Solutions

Sajal K. Das

National Science Foundation, CISE/CNS

Center for Research in Wireless Mobility and Networking (CReWMaN)

E-mail: das@uta.edu http://crewman.uta.edu

Acknowledgements: *NSF, AFOSR*

Outline

- Wireless Sensor Networks (WSNs)
- Security Challenges: Need for Multi-level Approach
- Modeling Node Compromises: *Epidemic Theory*
- Secure Data Aggregation: *Reputation and Trust Model*
- Node Replication: *Sequential Hypothesis Testing*
- Revoking Compromised Nodes: *Key Management*
- Self-Correction: *Digital Watermarking*
- Conclusions

Wireless Sensor Network (WSN)

We live in a cyber-physical world that we need to understand, serve, and control

Computation

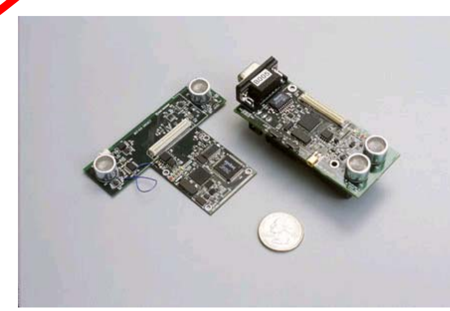
Sensory Data: A/D conversion, Compression, Filtering, Aggregation, Analysis

Communication (Wireless)

Broadcast sensory data, Dissemination, Routing

Control (Sensing / Actuation)

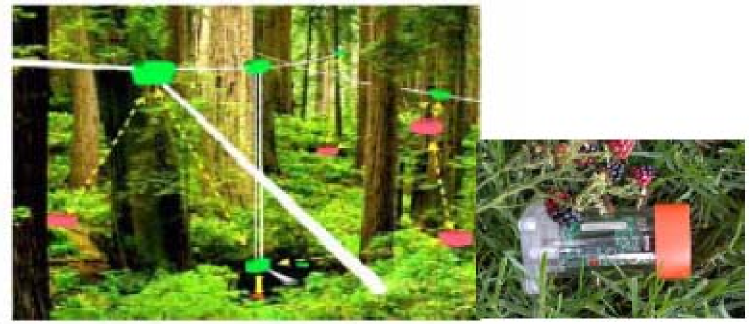
Sensing the physical world: temperature, humidity, pressure, light, velocity, sound, image



WSN Applications

Monitoring and Control

- Habitat
- Environment
- Ecosystem
- Agricultural
- Structural
- Traffic
- Manufacturing
- Health



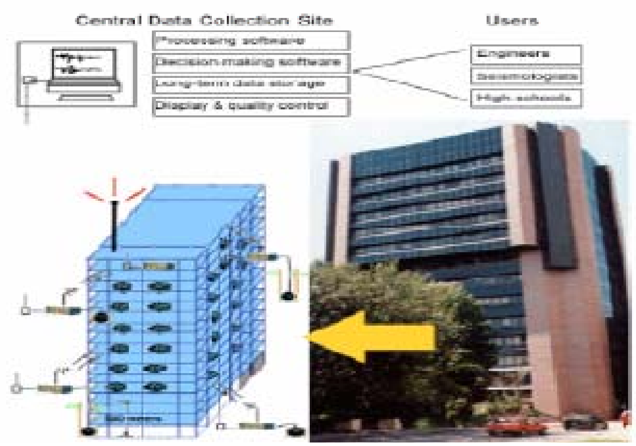
Ecosystems, Biocomplexity

ElderCare

Sensor Augmented Fire Response

Manufacturing

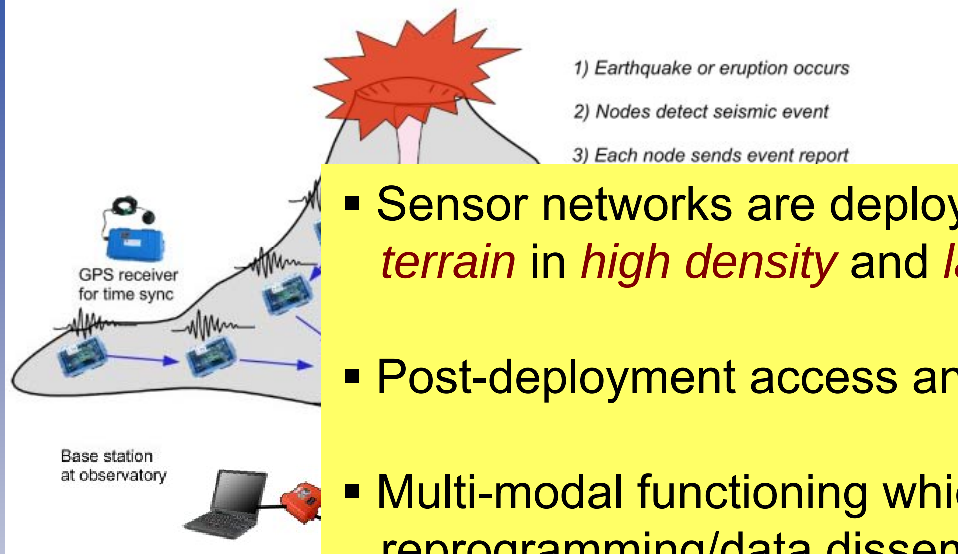
Seismic Structure Response



Security and Surveillance

- Infrastructure Security
- Border and Perimeter Control
- Target Tracking
- Intrusion Detection

WSN Applications



- Sensor networks are deployed in *unmonitored terrain* in *high density* and *large scale*
- Post-deployment access and control is an issue
- Multi-modal functioning which require reprogramming/data dissemination
  - Change applications
  - Troubleshoot
- Security is critical in many applications

SunSPOT

Wireless Sensing Devices embedded in our environment gathering data on physical phenomena !

Acknowledgment : http://www.eecs.harvard.edu/~mdw/proj/volcano/

Security Challenges in WSNs

Limited resources → Limited defense capability

– Energy (battery power), wireless bandwidth, computational power, storage, radio communication range (connectivity), sensing range (coverage)

– Public key too costly to authenticate packets with digital signatures and to disclose key with each packet

– Storing one-way chain of keys requires more memory and computation for message en-route nodes

Security Challenges in WSNs

Uncertain, unattended / hostile environment

- Uncertainty in sensing accuracy, wireless links, mobility, topology control, deployment (density), . . .
- Faulty prone nature vs. compromises (insider attacks)

Distributed control → No global knowledge

In-network processing: data fusion to exploit spatio-temporal redundancy → Loss of integrity, confidentiality

Multiple-attacking angles

→ Single level defense mechanism highly vulnerable

- Cryptographic technique is not the panacea

Threats to WSNs

Node Compromises / Replications and Intrusions

- Physical capture
- Sophisticated analysis: differential timing / energy analysis

Revealed Secrets

- Cryptographic keys, codes, commands, etc.

Enemy's Puppeteers

- Trojans in the network with full trust

Need for Multi-level Solution

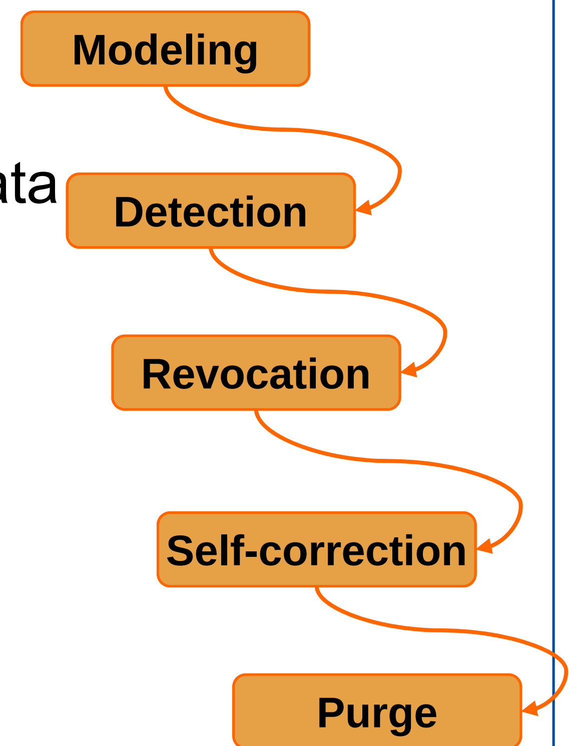
Attacks at multiple possible levels to be defended

– Model the propagation of node compromises
  E.g., trojan virus spreading

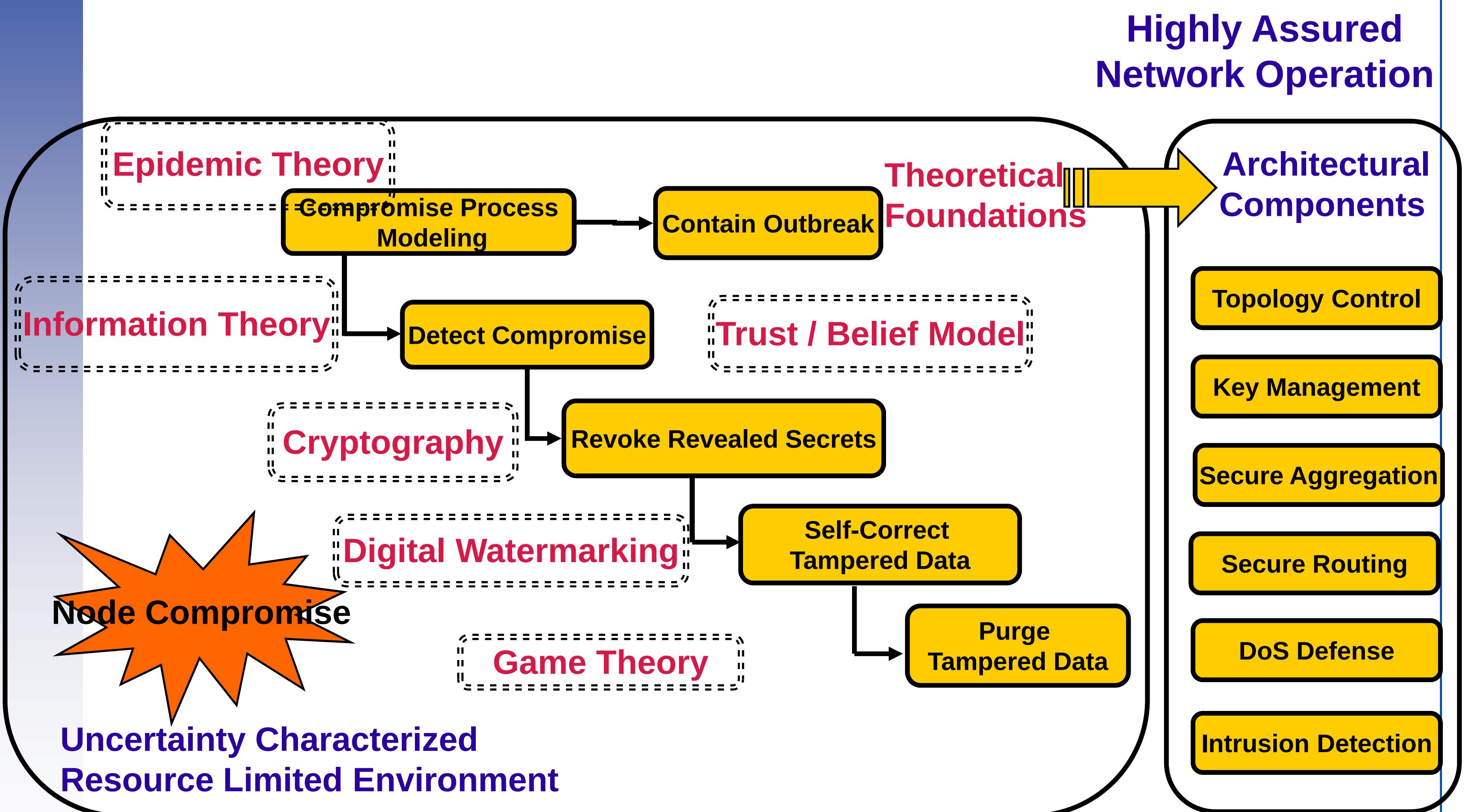
– Detect compromised nodes & forged data
  E.g., abnormal reports

– Revoke revealed secrets
  E.g., broadcast confidentiality

– Self-correct and purge false data
  E.g., average temperature calculation



Multi-Level, Integrative Framework



Relevant Publications

- W. Zhang, S. K. Das, and Y. Liu, “A **Trust** Based Framework for Secure Data Aggregation in Wireless Sensor Networks,” ***IEEE SECON 2006.***
- J.-W. Ho, M. Wright, and S. K. Das, “**ZoneTrust**: Fast Zone-Based Node **Compromise Detection and Revocation** in Sensor Networks Using Sequential Analysis," ***IEEE SRDS 2009.***
- J.-W. Ho, M. Wright and S. K. Das, “Fast **Detection of Replica Node Attacks** in Mobile Sensor Networks Using **Sequential Analysis**,” ***IEEE INFOCOM 2009.***
- P. De, Y. Liu, and S. K. Das, “An **Epidemic** Theoretic Framework for **Vulnerability Analysis** of Broadcast Protocols in Wireless Sensor Networks,” ***IEEE Transactions on Mobile Computing***, Vol. 8, No. 3, pp. 413-425, ***Mar 2009.***
- P. De, Y. Liu and S. K. Das, “Deployment Aware Modeling of **Compromise Spread** in Wireless Sensor Networks,” ***ACM Transactions on Sensor Networks, 2009.***
- J.-W. Ho, M. Wright, D. Liu, and S. K. Das, “Distributed **Detection of Replicas** with Deployment Knowledge in Wireless Sensor Networks,” ***Ad Hoc Networks, Aug 2009.***
- P. De, Y. Liu and S. K. Das, “Energy Efficient Reprogramming of a Swarm of Mobile Sensors,” ***IEEE Transactions on Mobile Computing, 2009.***
- W. Zhang, S. K. Das, Y. Liu, “**Secure Aggregation** in Wireless Sensor Networks: A **Digital Watermarking** Approach,” ***Pervasive and Mobile Computing, 2008.***

Modeling Compromises: Epidemic Defense

Premise: Node compromises in WSN broadcast protocols

– Capture node deployment, key distribution, topology

Research Objectives:

– Model and analyze spreading process of node compromises

– Characterize network-wide propagation rate and outbreak transition point of compromise process

– Study impact of infectivity duration of compromised nodes

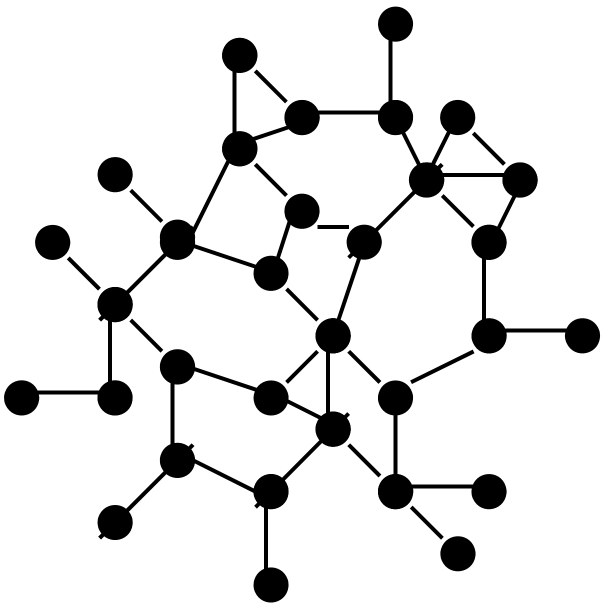
– Capture time dynamics of the spread

– Identify critical parameters to prevent outbreaks

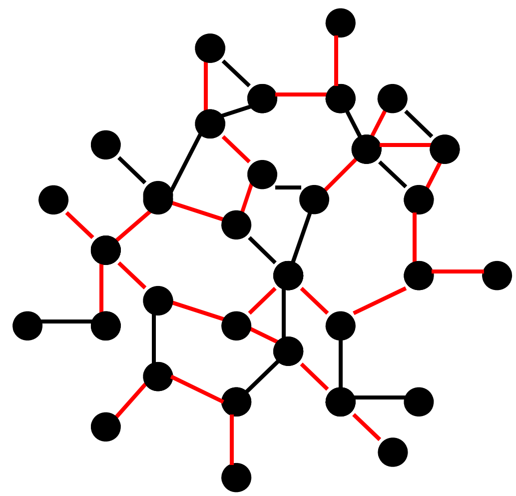
System Model

Random Pair-wise Key Pre-distribution

– A set of keys randomly chosen from a key pool



Physical Topology



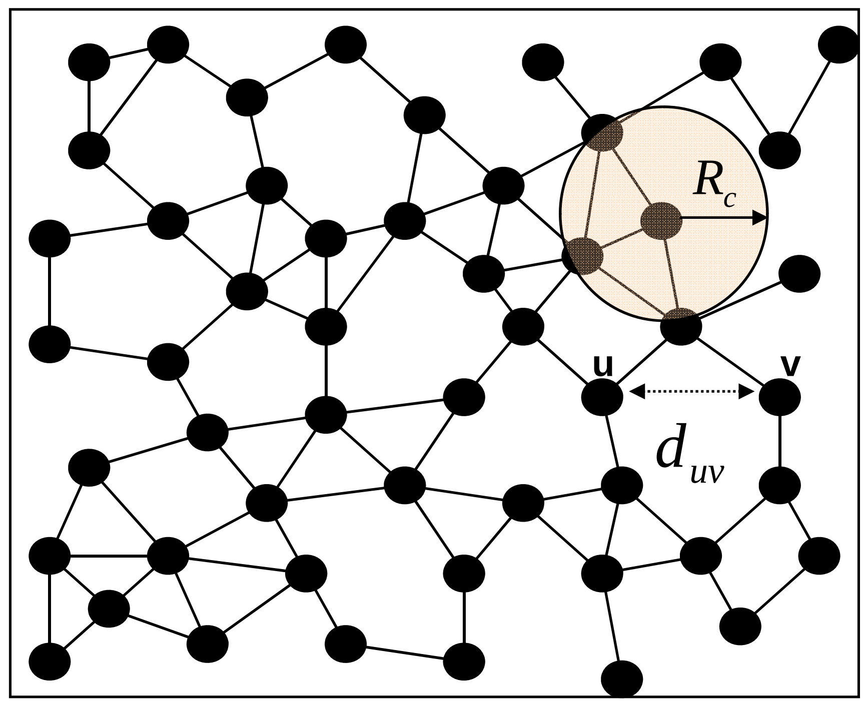
Virtual Key-Sharing Topology

Sensor Network Model

Modeled as undirected geometric random graph

- *N* nodes uniformly randomly distributed
- Unit Disk Model with transmission radius R_c
- $\alpha(d_{uv})$ is the probability of existence of an edge between nodes *u* and *v* at distance d_{uv}
- Node density $\sigma = \frac{N}{A}$

A = area of the terrain



Sensor Topology Model

$\rho = \frac{N}{R^2}$ denotes the node density of the network

N = total number of nodes, R = sensing radius

p = probability of existence of a physical link

$$p = \frac{r^2 \rho}{N}$$

r = average communication range between nodes

Probability for *l* nodes within communication range

$$p(l) = \binom{N}{l} p^l (1-p)^{N-l}$$

Sensor Topology Model

q = prob. of sharing pair-wise key between neighboring nodes

Probability of sharing at least one key with exactly k neighbors given l nodes within its range:

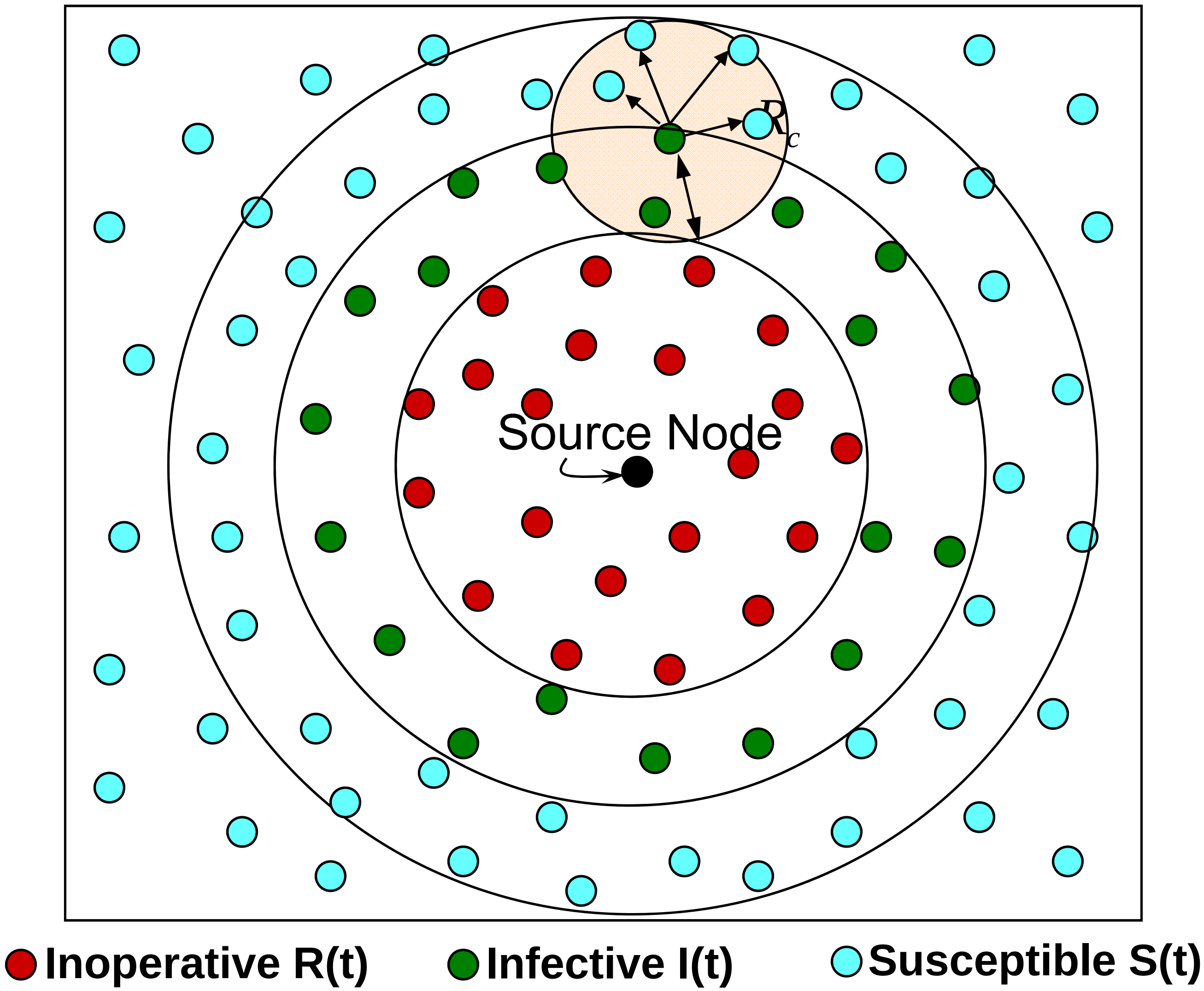
$$p(k|l) = \binom{l}{k} q^k (1-q)^{l-k}$$

Probability of having k neighbors sharing at least one key:

$$p(k) = \sum_{l=k}^{\infty} p(l)\, p(k|l)$$

$$p(k) = \sum_{l=k}^{\infty} \binom{N}{l} p^l (1-p)^{N-l} \binom{l}{k} q^k (1-q)^{l-k}$$

Infection Spread Model



S. K. Das

Epidemic Theoretic Framework

Model infection spread in a population of susceptibles

- Random graph based spatial model
- Differential equation based temporal model

Design *spread model* using network characteristics

- Local interactions based on transmission range
- Number of contacts determined by degree distribution of the key sharing network

Estimate the rate of infection (β) based on R_c

- Rate of communication paradigm of the broadcast protocol
- Infectivity potential (ρ) of the data

Epidemic Analysis

When nodes do not recover, *transmissibility* (T) is expressed only in terms of the *infection probability*, \beta

Node recovery is captured by expressing transmissibility as a function of average *duration of infectivity*, τ

$$1 - T = \lim_{\delta t \to 0} (1 - \beta \delta t)^{\tau / \delta t} \qquad T = 1 - e^{\beta \tau}$$

Average cluster size as epidemic attains outbreak proportions

$$s = 1 + \frac{T G_0^{'}(1)}{1 - T G_1^{'}(1)}$$

Average Epidemic size after outbreak results

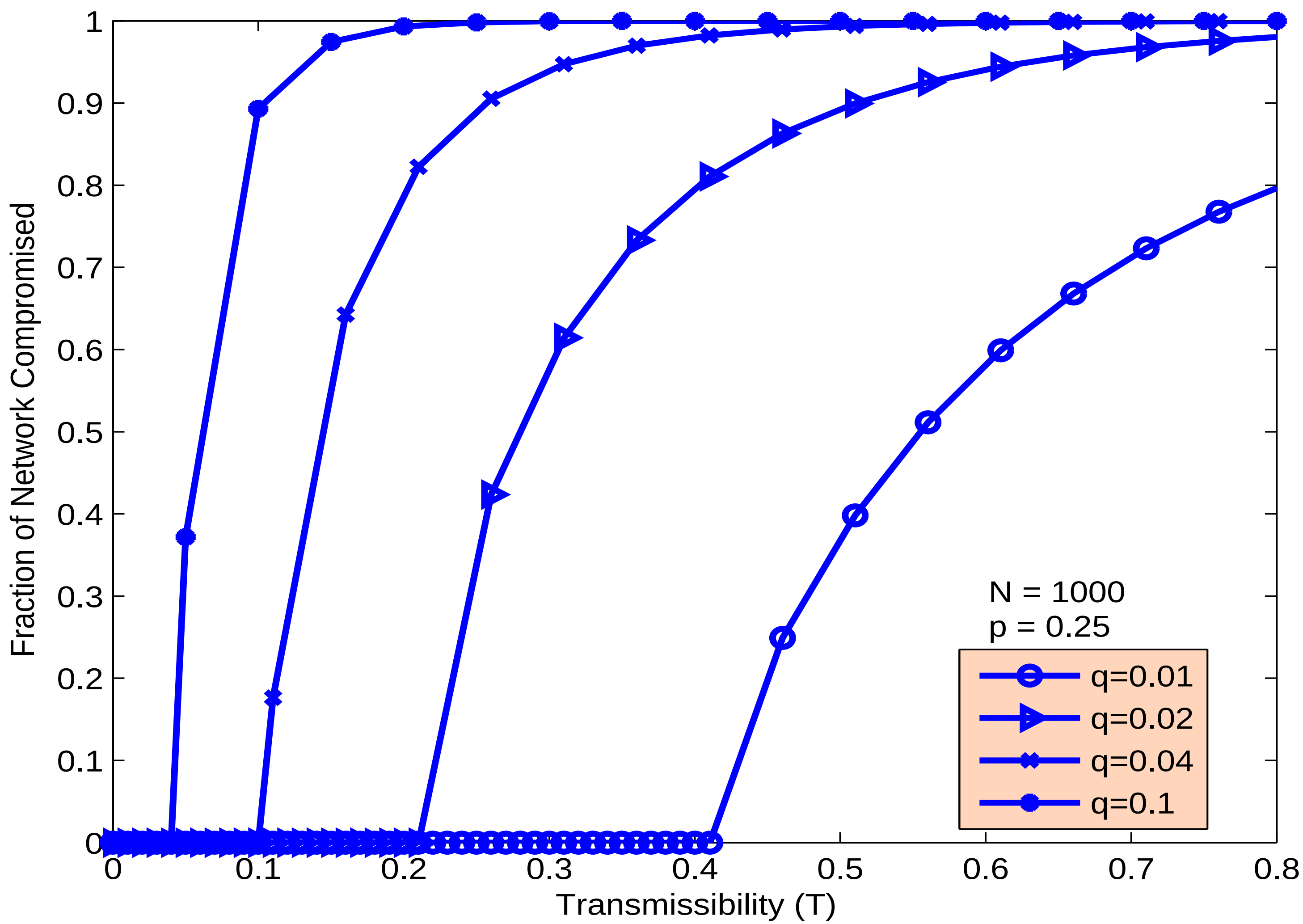
$$S = 1 - G_0(u)$$

$$u = G_1(u)$$

Epidemic Size with infection probability

q = prob. of sharing pair-wise key between neighboring nodes

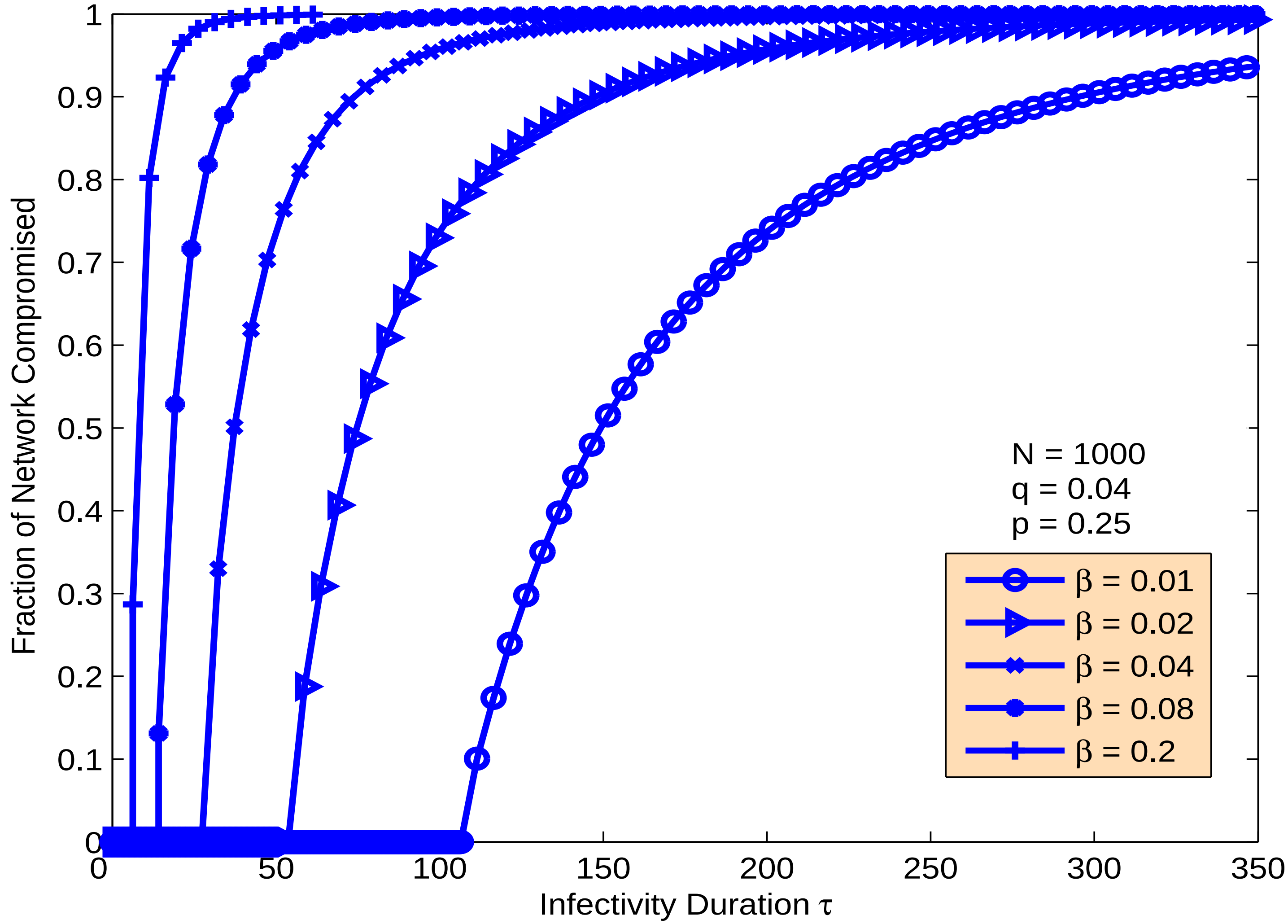
p = probability of existence of physical link



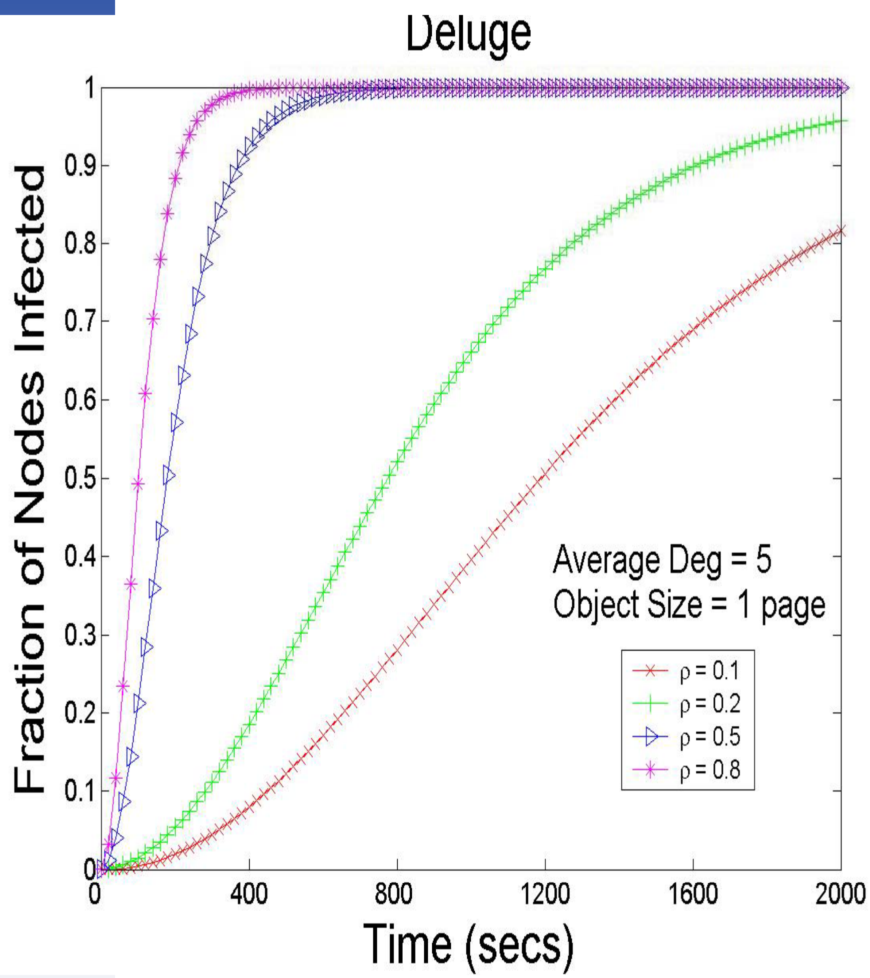
Epidemic Size with infectivity duration

q = prob. of sharing pair-wise key between neighboring nodes

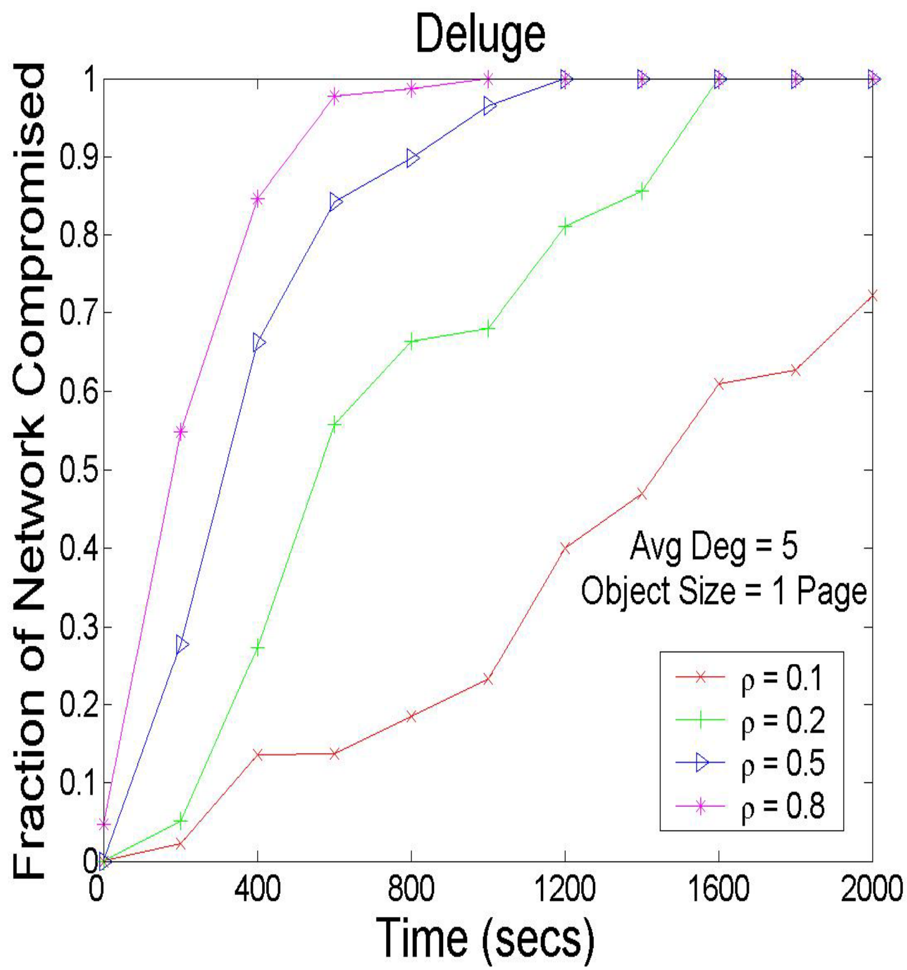
p = probability of existence of physical link



Deluge : Data Propagation Rate



Analytical



Simulation

Model captures rate of data propagation over dissemination protocols

Simulation Study

Capture the time dynamics of the spread of compromise

Observe duration and nature of gradual recovery process

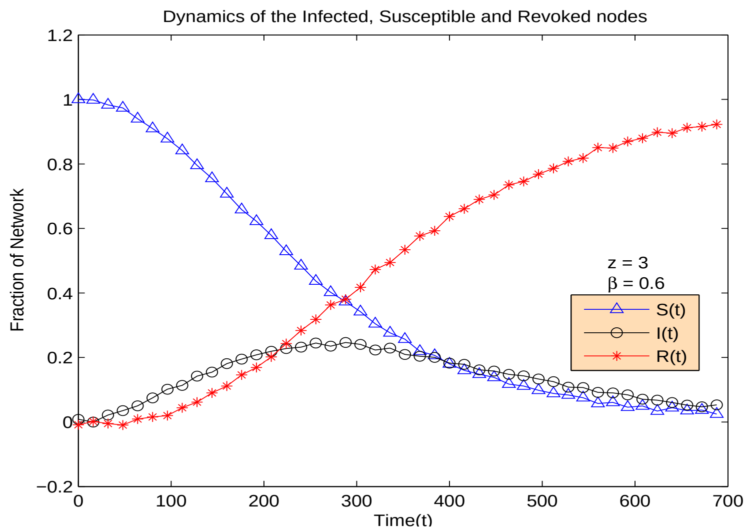
Observe effects of various network parameters

- Average node degree of key sharing network
- Average infection rate
- Average duration of infectivity

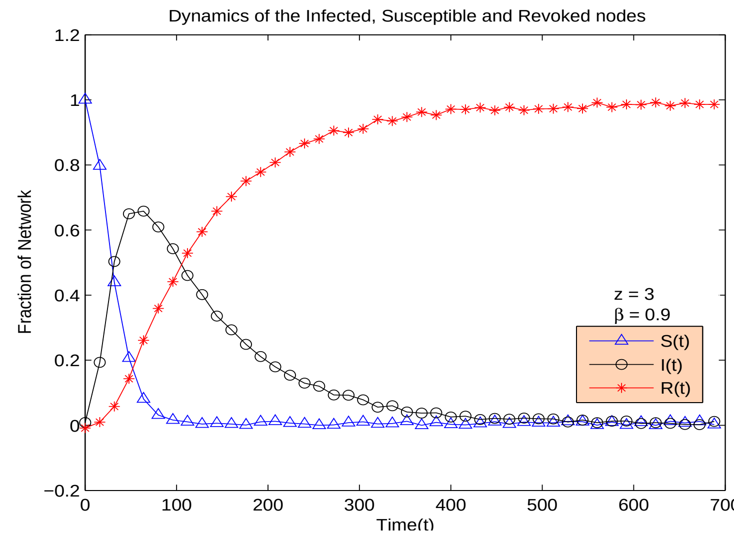
Simulation Results

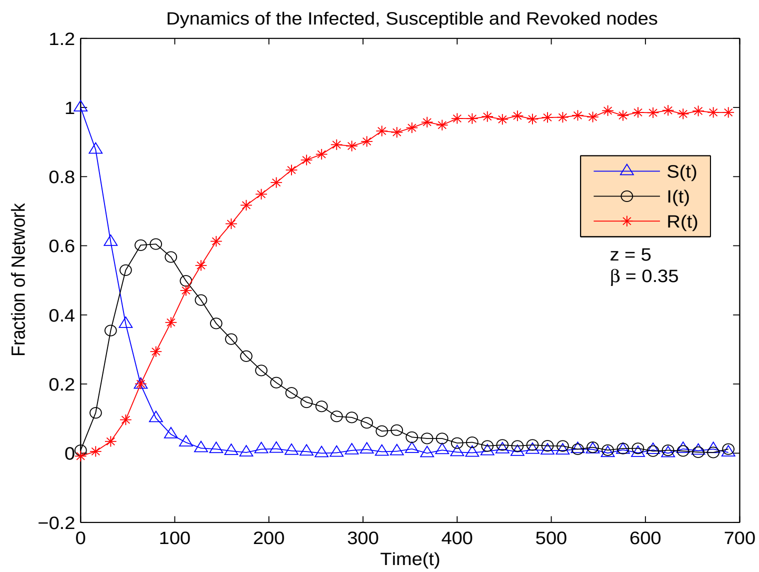
Under both scenarios – no node recovery and node recovery

$\tau = 30$

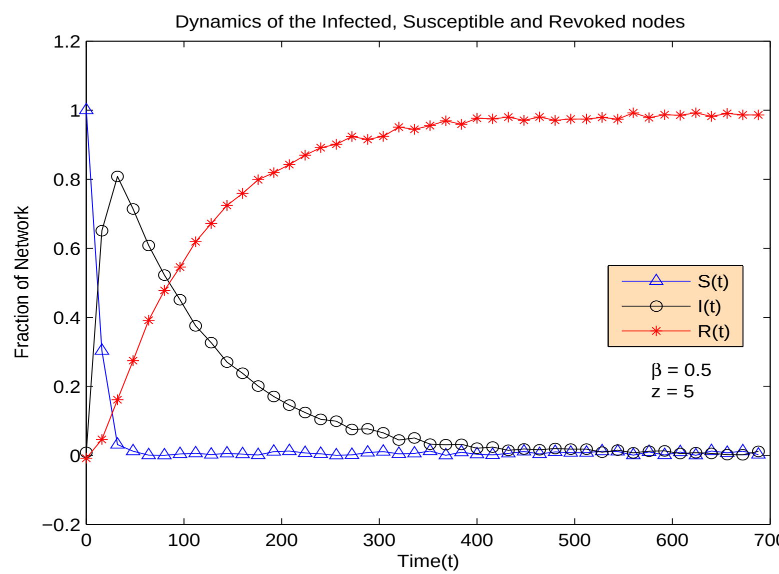


$\tau = 10$





$\tau = 10$



$\tau = 10$

Developing Belief / Trust Model

Premise: False data injection from compromised nodes

- Cryptographic techniques ineffective

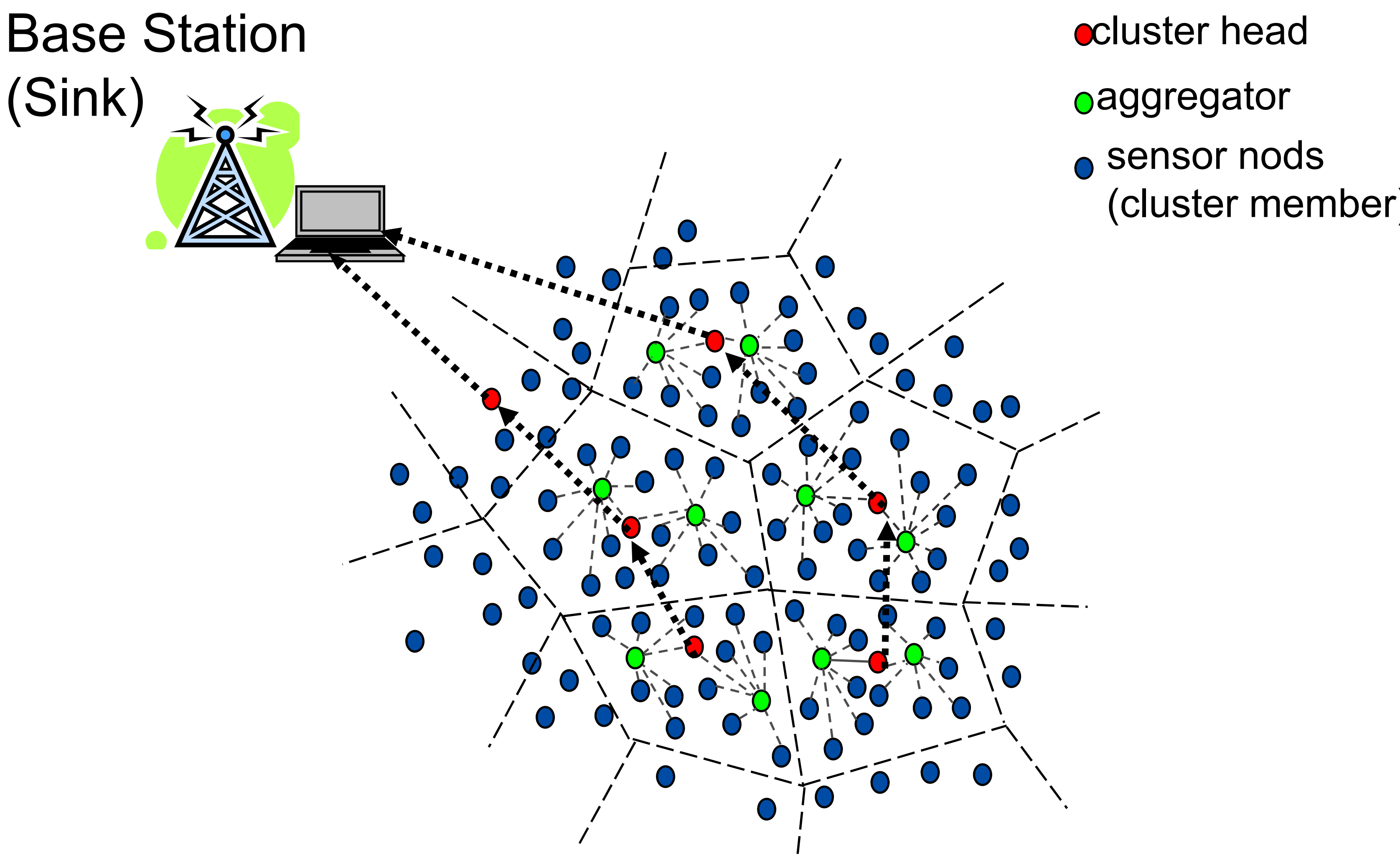
Objectives: Trust model to identify and purge false data. Reduce uncertainty in data aggregation / fusion.

Solution:

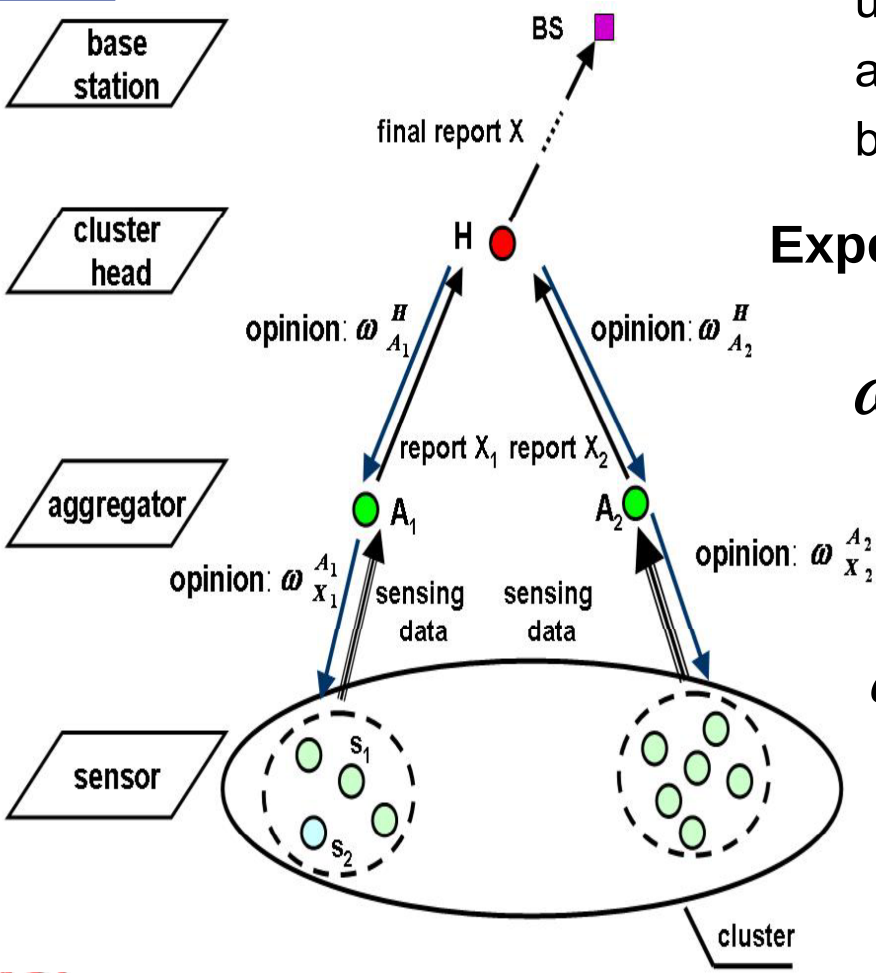
- Information theoretic (relative entropy) measure to quantify reputation / opinion of data, leading to higher confidence
  - Belief, disbelief, uncertainty, relative atomicity
- Josang's belief model to define and manage trust propagation through intermediate nodes along the route
- Identify malicious nodes by learning and outlier classification – purge false data to achieve secure aggregation

W. Zhang, S. K. Das and Y. Liu, "A Trust Based Framework for Secure Data Aggregation in Wireless Sensor Networks, *IEEE SECON*, Oct 2006.

Sensor Network Model



Josang's Belief Model



Opinion: ω = (b, d, u, a), b + d + u = 1

b: belief

d: disbelief

u: uncertain

a: relative atomicity

b, d, u, a $\in$ [0,1]

Expected Opinion: O = E(ω) = b + au

ω_A^H **: Cluster head's opinion about aggregator**

$$\omega_{A_1}^{H} = (0.95, 0.03, 0.02, 0.5)$$

$$O_{A_1}^{H} = 0.95 + 0.5 * 0.02 = 0.96$$

ω_X^A **: Aggregator's opinion about its report X**

$$\omega_X^{A_1} = (0.688, 0, 0.312, 0.9)$$

$$O_X^{A_1} = 0.688 + 0.9 * 0.312 = 0.969$$

Belief Propagation: Subjective Logic

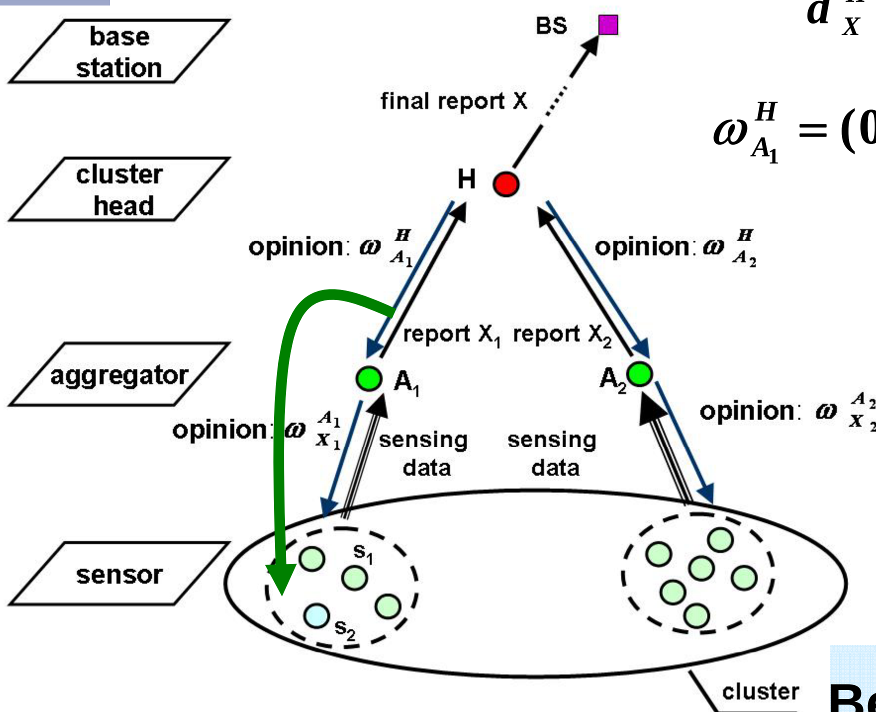
Belief discounting (recommendation)

Cluster head's opinion about X as a result of aggregator's opinion:

$$\omega_X^{H:A} = \omega_A^H \otimes \omega_X^A = (b_X^{H:A}, d_X^{H:A}, u_X^{H:A}, a_X^{H:A})$$

$$b_X^{H:A} = b_A^H * b_X^A \qquad u_X^{H:A} = d_A^H + u_A^H + b_A^H u_X^A$$

$$d_X^{H:A} = d_A^H * d_X^A \qquad a_X^{H:A} = a_X^A$$



$$\omega_{A_1}^H = (0.95, 0.03, 0.02, 0.5); \omega_X^{A_1} = (0.688, 0, 0.312, 0.9)$$

$$b_X^{H:A_1} = 0.95 * 0.688 = 0.654$$

$$d_X^{H:A_1} = 0$$

$$u_X^{H:A_1} = 0.03 + 0.02 + 0.95 * 0.312 = 0.364$$

$$a_X^{H:A_1} = 0.9$$

$$\omega_X^{H:A_1} = (0.654, 0, 0.364, 0.9)$$

Belief decreases, uncertainty increases

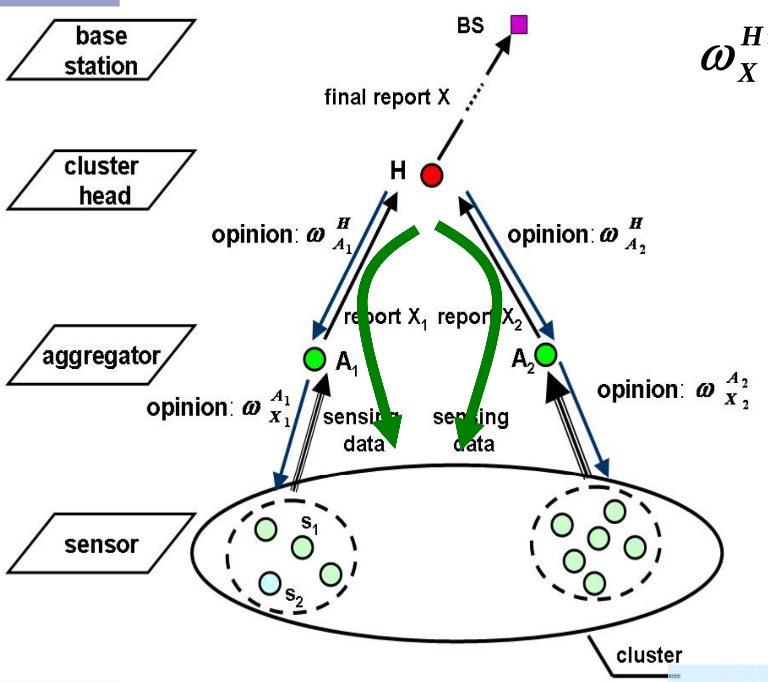
Belief Consensus

Cluster head's opinion about X via A_1: $\omega_X^{H:A_1} = (b_X^{H:A_1}, d_X^{H:A_1}, u_X^{H:A_1}, a_X^{H:A_1})$

Cluster head's opinion about X via A_2: $\omega_X^{H:A_2} = (b_X^{H:A_2}, d_X^{H:A_2}, u_X^{H:A_2}, a_X^{H:A_2})$

Cluster head's consensus opinion about X:

$$\omega_X^{H:A_1,H:A_2} = \omega_X^{H:A_1} \oplus \omega_X^{H:A_2} = (b_X^{H:A_1,H:A_2}, d_X^{H:A_1,H:A_2}, u_X^{H:A_1,H:A_2}, a_X^{H:A_1,H:A_2})$$



$$\omega_X^{H:A_1} = (0.654, 0, 0.346, 0.9); \omega_X^{H:A_2} = (0.368, 0, 0.632, 0.7)$$

$$b_X^{H:A_1,H:A_2} = \frac{0.654*0.632+0.368*0.346}{0.346+0.632-0.346*0.632} = 0.712$$

$$d_X^{H:A_1,H:A_2} = 0$$

$$u_X^{H:A_1,H:A_2} = \frac{0.346*0.632}{0.346+0.632-0.346*0.632} = 0.288$$

$$a_X^{H:A_1,H:A_2} = \frac{0.7*0.346+0.9*0.63-(0.7+0.9)*0.35*0.63}{0.35+0.63-2*0.35*0.63} = 0.85$$

$$\omega_X^{H:A_1,H:A_2} = (0.712, \quad 0, \quad 0.288, \quad 0.85)$$

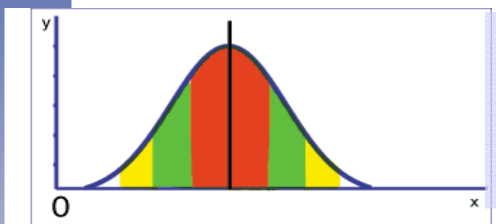
More evidences, belief in the result increases

S. K. Das

Aggregator: Compute Sensor Reputation

Outlier exclusion: Too far from median ➔ outlier

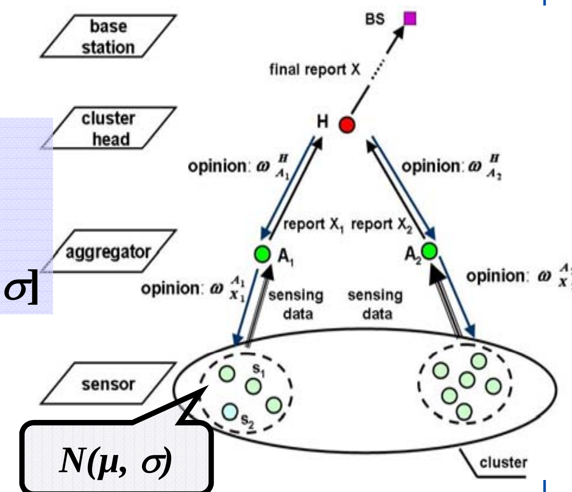
High density ➔ Normal distribution *N(μ, σ)*



Red: 68% of data within [$\mu - \sigma, \mu + \sigma$]

Green: 95% of data within [$\mu - 2\sigma, \mu + 2\sigma$]

Yellow: 99.7% of data within [$\mu - 3\sigma, \mu + 3\sigma$]



Each sampling independent

- Ideal node frequency: in long run, $\mathbf{Pr}(p_i \mid x_i \in [\bar{x} - \sigma, \bar{x} + \sigma]) = \mathbf{0.68}$
- Actual node frequency: $\mathbf{Pr}(q_i \mid x_i \in [\bar{x} - \sigma, \bar{x} + \sigma])$, learn from observation
- Measure difference in ideal and actual frequencies: **Kullback Leibler distance**

$$D(p \| q) = \sum p(x) \log \frac{p(x)}{q(x)};$$

p(x), q(x) prob. mass function for ideal/actual node freq.
D(.) also called relative entropy measure

Reputation: $$r = \frac{1}{1 + \sqrt{D}}$$

The shorter the distance, more trustworthy, higher reputation

Sensor Node's Reputation: Example

Two sensors, s_1 and s_2

– Time t_1: $f_{s_1}^{t_1} = 0.65\,,\ f_{s_2}^{t_1} = 0.63$

$$D(f_{s_1}^{t_1} \,\|\, f_{ideal}^{t_1}) = (1-0.65) * \log\frac{(1-0.65)}{(1-0.68)} + 0.65 * \log\frac{0.65}{0.68} = 0.0029$$

$$r(s_1^{t_1}) = \frac{1}{1+\sqrt{0.0029}} = 0.949$$

– Time t_2: $f_{s_1}^{t_2} = 0.68\,,\ f_{s_2}^{t_2} = 0.30$

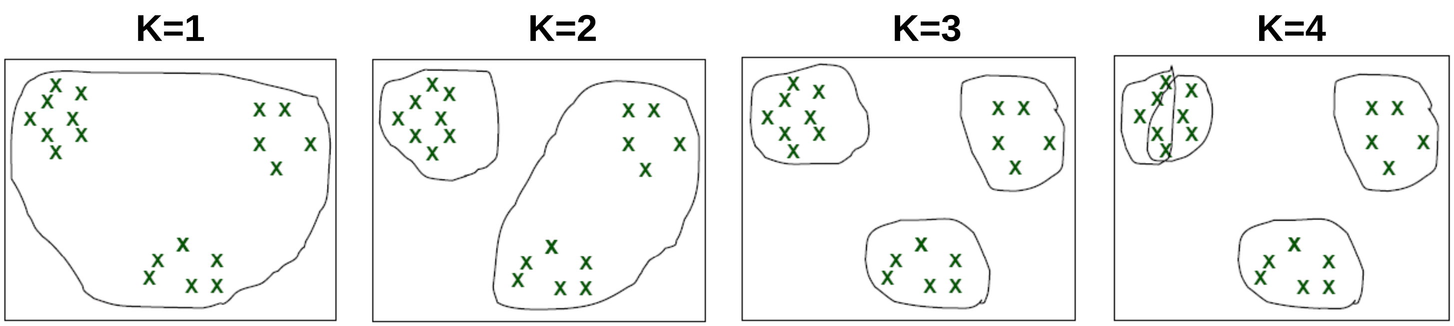
Time	Sensor node	Actual freq.	Ideal freq.	KL-distance	Reputation
t_1	s_1	0.65	0.68	0.0029	0.949
	s_2	0.63	0.68	0.0081	0.918
t_2	s_1	0.68	0.68	0	1
	s_2	0.30	0.68	0.436	0.602

Reputation changes with time based on behavior

Aggregator: Reputation Classification

Classify reputation to identify malicious nodes

- Traditional system: threshold based classification
- Online unsupervised learning, K-mean algorithm
- No prior K available, how to dynamically decide K?



Determining K

Ex:

Time	Sensor node	Reputation
t_1	s_1	0.949
	s_2	0.918
t_2	s_1	1
	s_2	0.602

1 group

2 groups

Aggregator: Opinion Formulation

$$\omega_X^A = (b_X^A, d_X^A, u_X^A, a_X^A)$$

Degree of trust in aggregation result

Trustworthy

Nodes whose data close to mean

Uncertain

Nodes whose data not close to mean

Uncertain nodes' reputation

- how much contribution to expected opinion?

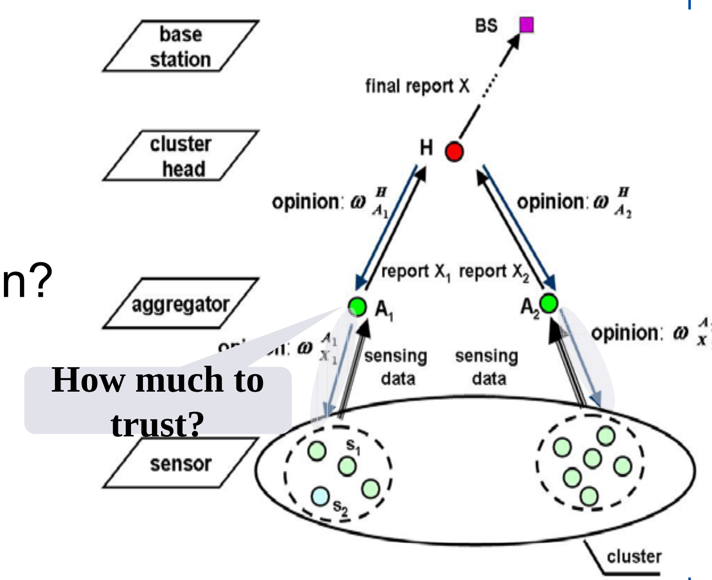
Formulation

belief: percentage in $(\overline{x} \pm \sigma)$

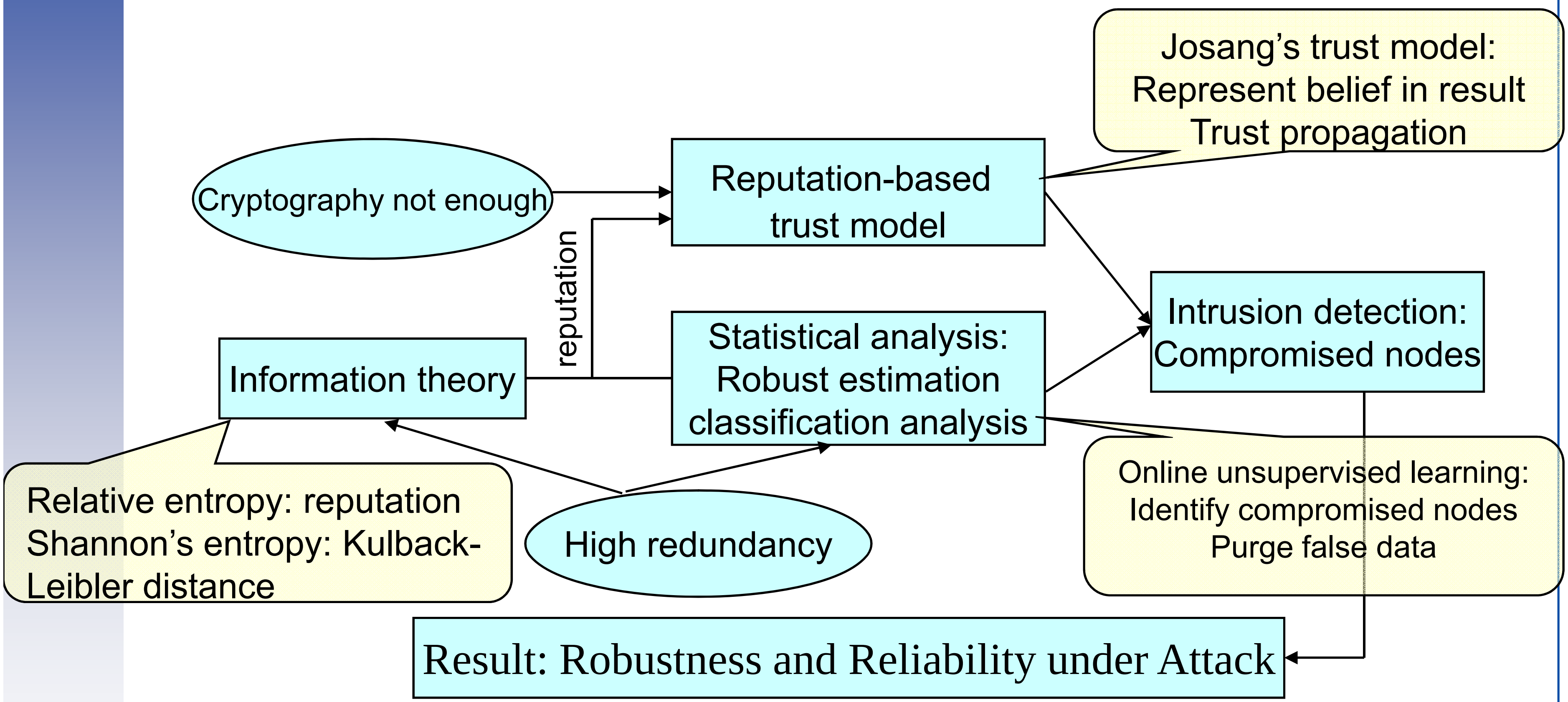
disbelief: 0 (after excluding outlier)

uncertain: percentage out of above range

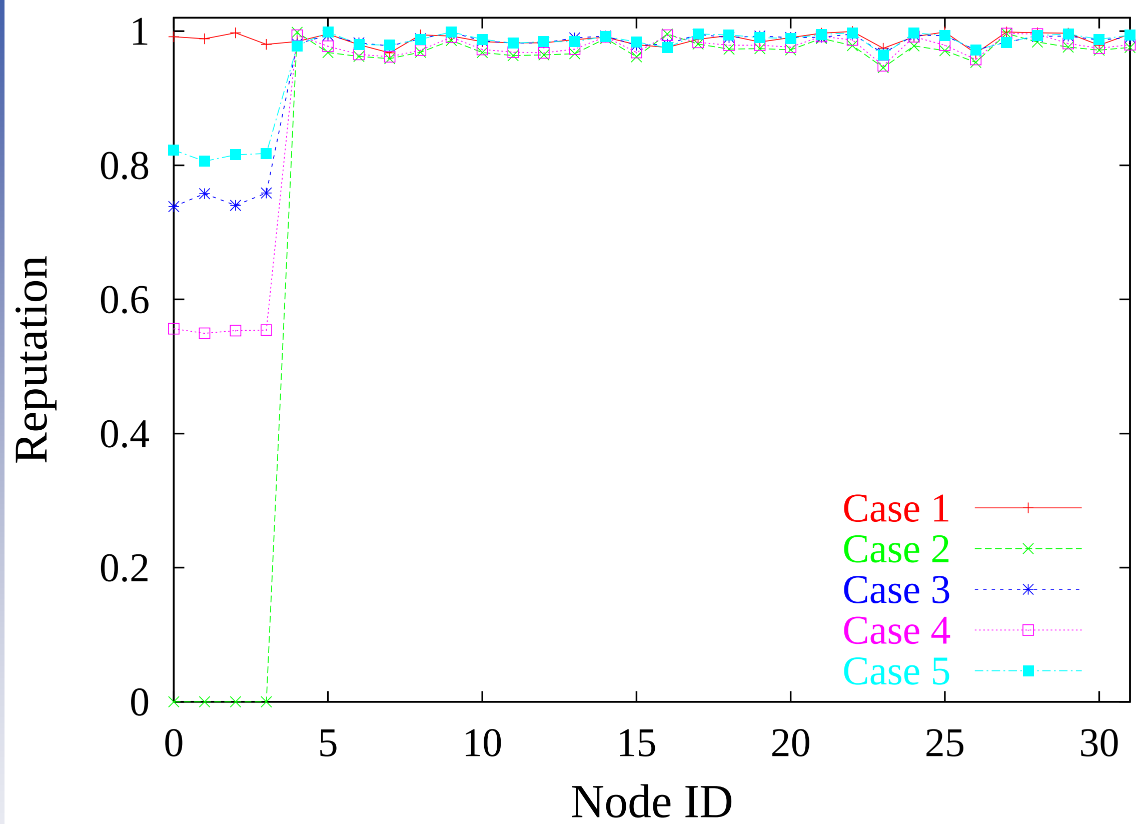
relative atomicity: reputation of nodes fall out the range



Trust Framework



Simulation Results: Reputation



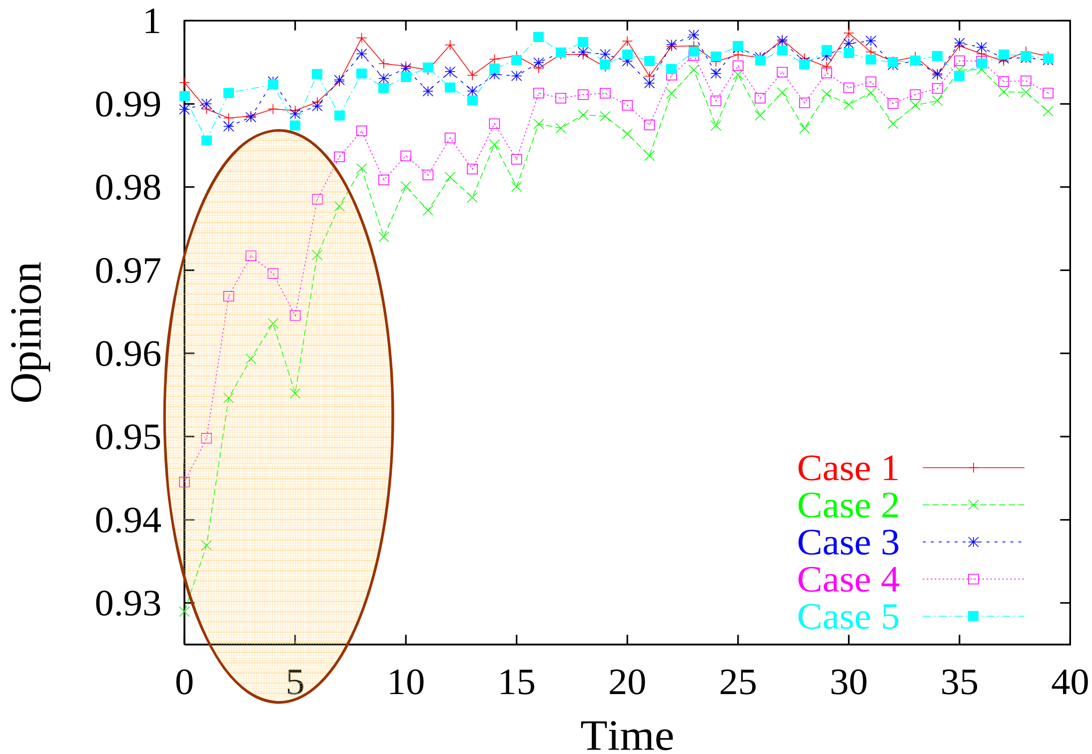
Case No.	Misbehaving time (%)	False data type
1	0	N/A
2	100	Obvious
3	100	Tricky
4	66	Obvious
5	66	Tricky

No malicious nodes, all nodes' reputation close to 1

Reputation of malicious nodes significantly lower than legitimate ones

Reputation of malicious nodes proportional to amount of true data they send

Simulation Result: Opinion



Test case

Case No.	Misbehaving time (%)	False data type
1	0	N/A
2	100	Obvious
3	100	Tricky
4	66	Obvious
5	66	Tricky

False data sneaking into aggregation (Cases 2, 4) may affect result ➔ "pollute" legitimate node's reputation

Low opinion or polluted reputation ➔ result from low reputation nodes

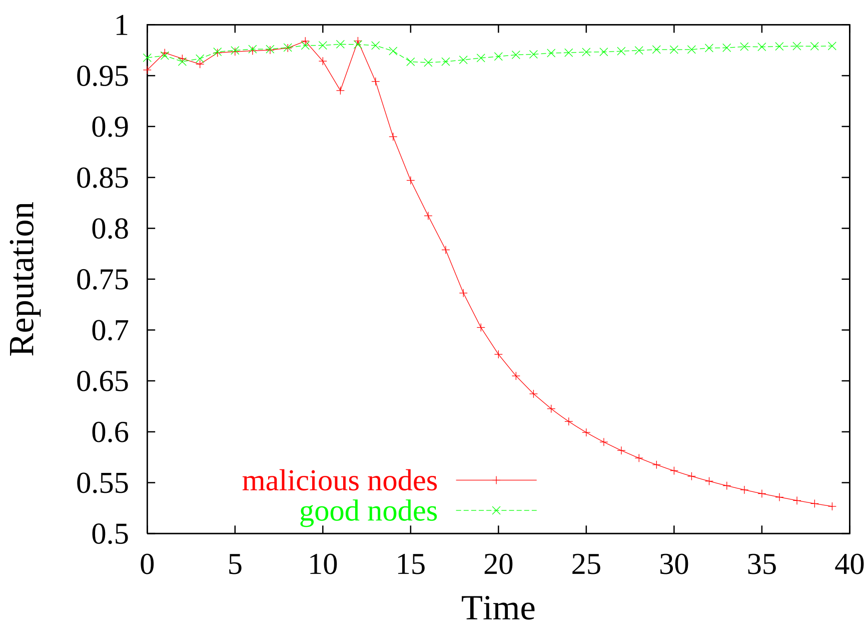
Detection/blocking malicious nodes ➔ opinion / confidence increases

Opinion correctly represents the belief in the result

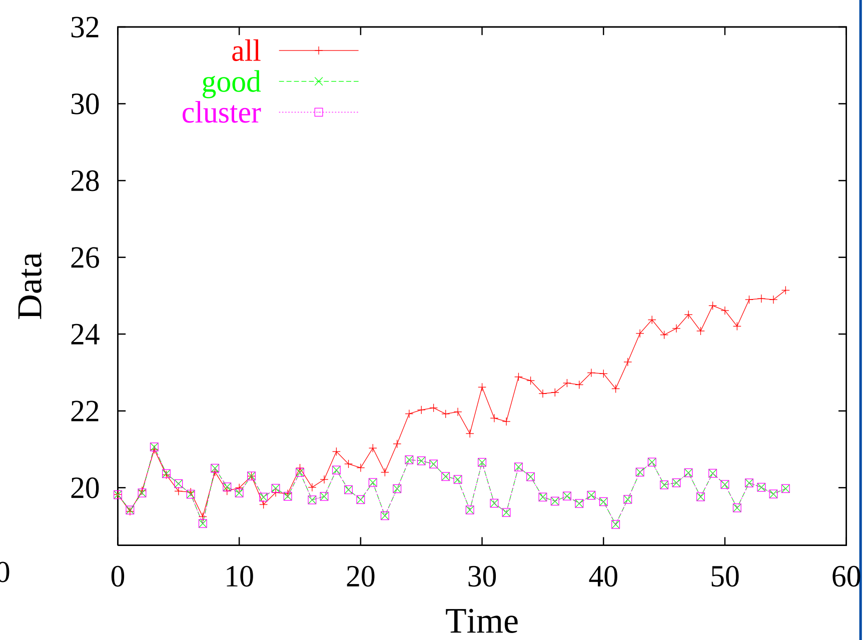
Cooperative Malicious Nodes (10%)

Scenario: Malicious nodes behave “good” at first 1/3 experiment, then they all send same data each time

Evolution of Reputation



Aggregation Result



Malicious nodes can be identified as long as they misbehave.

Aggregation result robust to cooperative malicious nodes of different fractions

Conclusions

- Integrated multi-level security framework in wireless sensor networks.

- Epidemic theory modeling to control spread of infected nodes and outbreak.

- Information theory-based reputation to detect intrusion of malicious nodes.

- Belief / trust model to ensure secure information aggregation by effectively filtering false data.

- Distributed key sharing and collaboration to revoke reveals secrets.

- Digital watermarking technique to self-correct compromised data.

Epilogue

"A *teacher* can never truly teach unless he is still learning himself. A lamp can never light another lamp unless it continues to burn its own flame. The teacher who has come to the end of his subject, who has no living traffic with his knowledge but merely repeats his lesson to his students, can only load their minds, he cannot quicken them".

Rabindranath Tagore (1861-1941)

(Indian Poet, Nobel Laureate,1913)

Thank You

http://crewman.uta.edu