Wireless Network Security
Spring 2016

Patrick Tague

Class #14 – Trust and Reputation

Reminders

- No class on March 3
  - I will have some time available if you want to have a skype/hangout about hw#3
- HW#3 due on March 3
  - I will be reachable if you have questions
- HW#4 goes out soon

Class #14

- Evaluating trust in networked systems

- Network reputation systems

What are Trust and Reputation?

- Trust:
  - Subjective expectation of an agent receiving positive outcomes from another in a specific context
- Reputation:
  - Global perception of an agent's trustworthiness in a system
- Why do we care about these issues?

What does that mean?

- Trusting claims made by other devices/users about identity, services, events, etc.
- Trusting others to correctly manage data and services
- Trusting others to behave as expected/promised
- Trusting others to be fair / not greedy
- And so on...

Trust in the Internet

- The Internet uses a centralized or hierarchical trust model based on identify certification
  - A certificate authority attests the identity and trustworthiness of individuals/groups by issuing a signed/certified public key
    - CA claims "**X** is identifiable and trustworthy"
    - **X** provides signed certificate from CA to **Y**
    - Transitive trust: CA → **X**, **X** → **Y** ==> CA → **Y**
  - This type of model also provides a notion of accountability

Trust Challenges

- In MANET, the biggest challenge is lack of a centralized authority, so nobody to act as a CA
  - How to distribute and approximate the CA trust model?
  - Is there a different model that works as well/better?
- In mesh and WSN, latency and trusted paths are major challenges
  - How to bootstrap a secure/trusted path to the CA?
- In DTN, latency is a huge problem

Let's go into more detail about how to model and measure trust

Individual Trust Models

- Each agent evaluates its trust of another
  - Combines direct and indirect observations
  - Includes past behavior
- Trust is an opinion
  - It can be expressed / shared, modified, changed, etc.

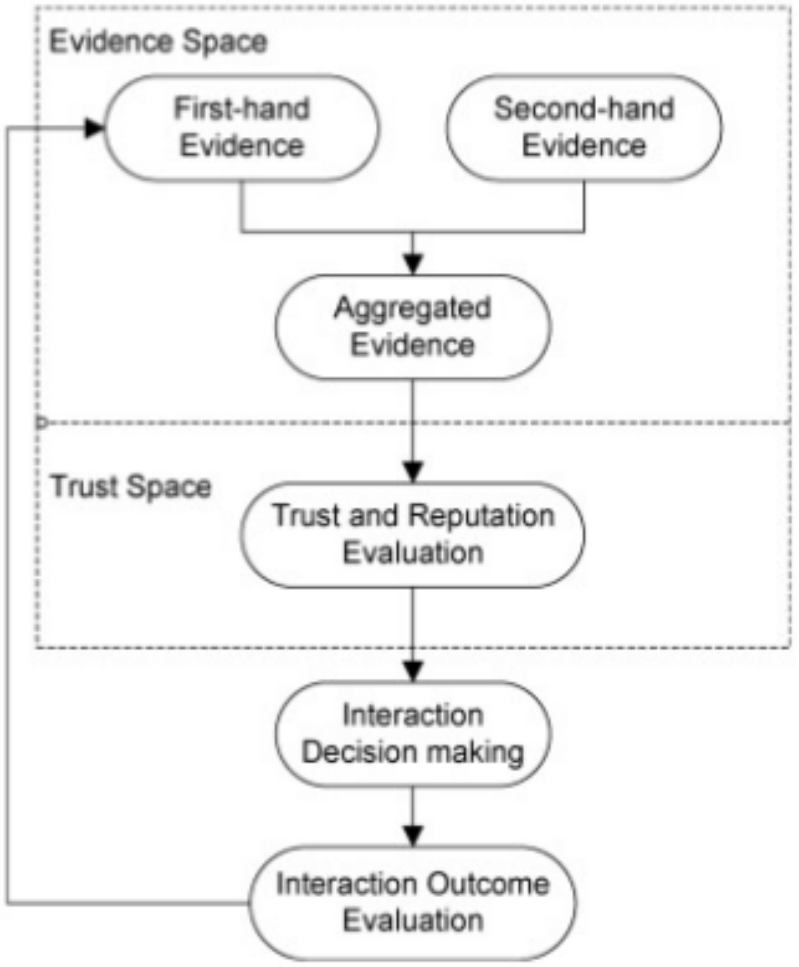
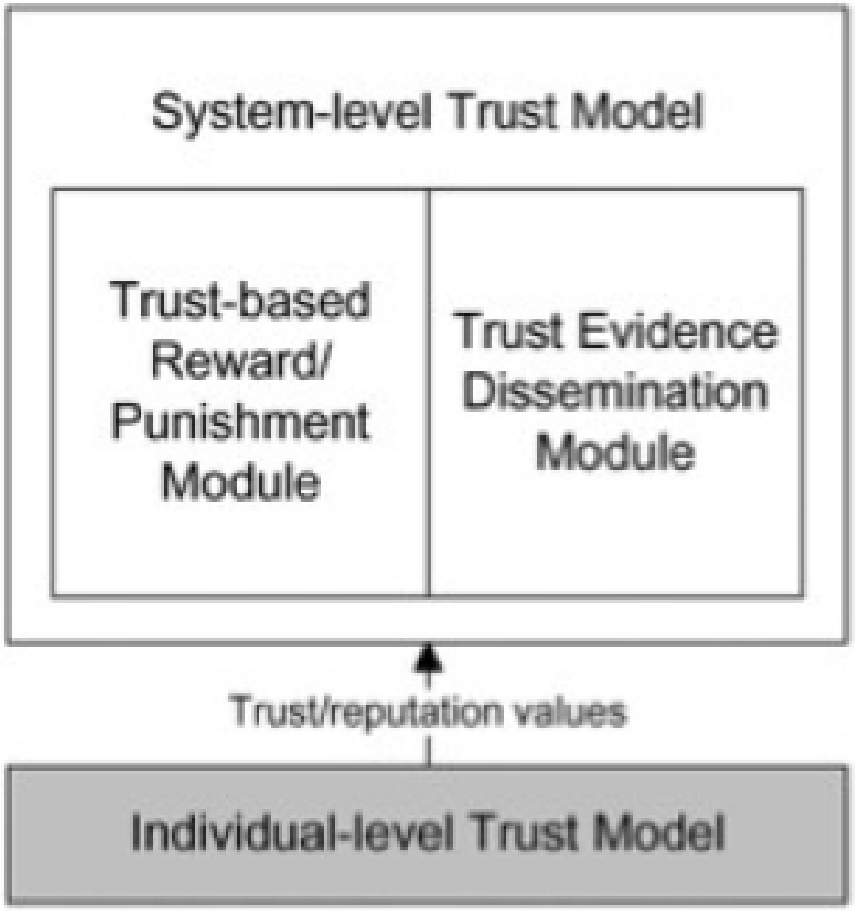


Figure from [Yu et al., Proc. IEEE, 2010]

Reputation, aka System Trust

- Reputation is a globally shared view of trust in an agent
  - Sort of an aggregate of individual trust values
  - Allows for consistent action wrt non-cooperative agents



Trust Issues

- How to initialize / bootstrap trust?
  - Ex: I'm evaluating **X**, but I've never met them before (and none of my contacts have met them before)
- How to weigh past vs. current events?
  - Ex: **X** was uncooperative two weeks ago but nice since then
- How to weigh direct vs. indirect observations?
  - Ex: **X** cooperated with my neighbor (supposedly) but not with me

Trust Issues

- How to map events to a trust metric?
  - Ex: **X** cooperated 9 times and refused once
- How to capture the dynamics of trust?
  - Ex: [**X** cooperated 9x and refused 1x]

    vs.

    [**X** cooperated 4x, refused 1x, cooperated 5x]
- How to use trust metrics once evaluated?

Common Trust Themes

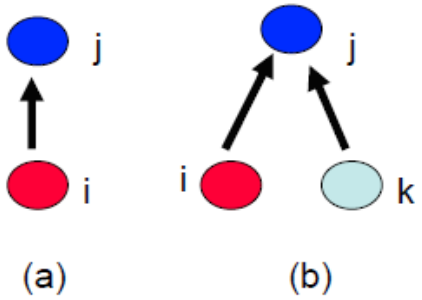
- Most techniques for evaluating trust use some common concepts
  - Trust is difficult to build but easy to lose
  - Importance of past events decays over time
  - Trust should be somewhat robust to "natural" events
    - E.g., can incorporate uncertainty or confidence
  - The trust mechanism itself should be robust to misbehavior
    - More on this in a bit...

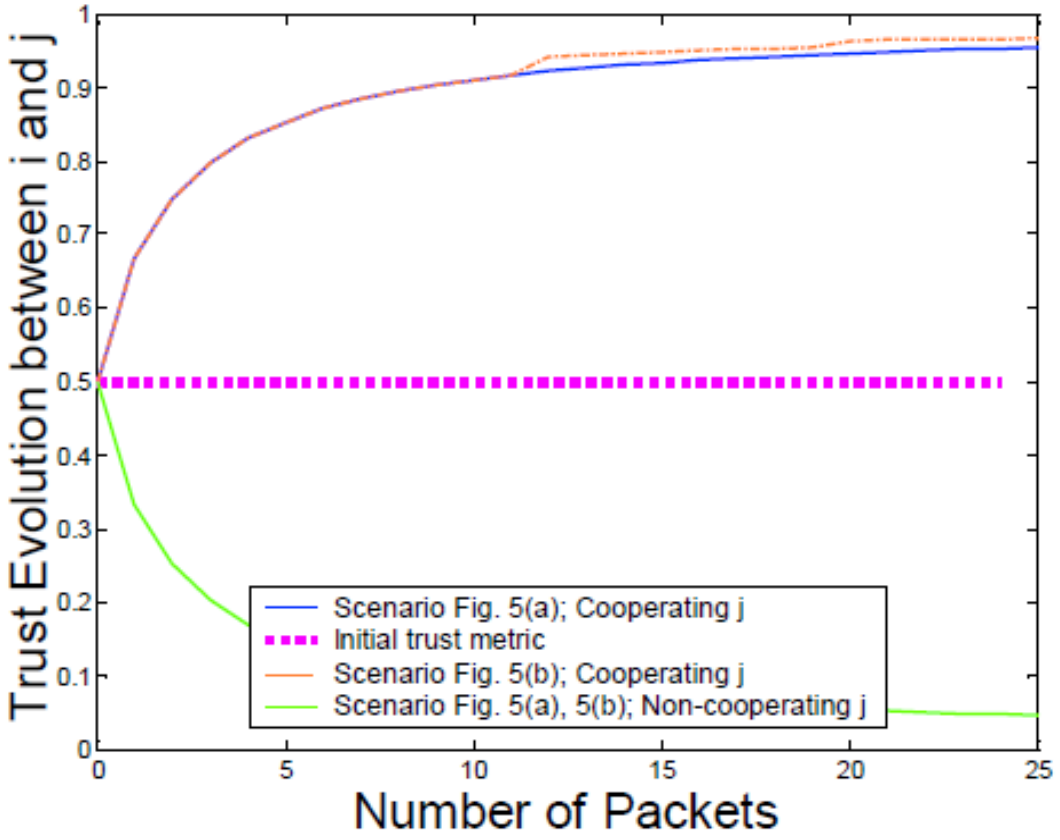
Trust Metric Dynamics

- Various approaches using different evaluation policies, such as:
  - For each positive/negative action, add/subtract a constant to/from the trust value
  - For each positive action, add a constant; for each negative action, multiply by a constant fraction
  - For each positive action, add a constant; for each negative action, drop to the lower boundary (0 or -1)

Example

- From [Ganeriwal & Srivastava, 2004]





What about attacks on the trust/reputation system itself?

Trust/Reputation Attacks

- Attack model:
  - Attacker is an active insider, can cooperate/comply or choose to misbehave
  - Motivated by selfish/unfair or malicious intent
  - Can work alone or collude with others

- In general, one of three goals:
  - Falsely increasing trust values (itself or friend)
  - Falsely decreasing trust values (attack target)
  - Denial of service

Self-Promotion Attack

- **Goal:** obtain a higher trust among neighbors and/or reputation in the system
- **Means:** fabricate positive feedback or modify reputation values in transit, possibly at the expense of others
- **Assumptions:** (i) reputation system is based on positive feedback, (ii) mechanism is exploitable

Whitewashing Attack

- **Goal:** quickly repair a trust/reputation value after selfish/malicious action is performed

- **Means:** after abuse, re-enter or exploit the system to reset the trust values to default or previous state, possibly in combination with other attacks

- **Assumptions:** (i) reputation system may need negative feedback, (ii) mechanism is exploitable

Slandering Attack

- **Goal:** falsely decrease the trust/reputation value of other actor(s)

- **Means:** as the name suggests, spread false opinions of the other actor, often through negative feedback which can be very damaging

- **Assumptions:** (i) reputation system needs negative feedback, (ii) mechanism is exploitable, (iii) may require collusion

Orchestrated Attacks

- **Goal:** multiple attackers collude to force the system into a particular desired state

- **Means:** combine promotion, whitewashing, and slander as needed for specific goal
  - Ex: oscillation attack – divide into teams, ½ does slander and ½ does promotion, switching occasionally

- **Assumptions:** (i) collusion, (ii) whatever assumptions required for component attacks

DoS Attacks

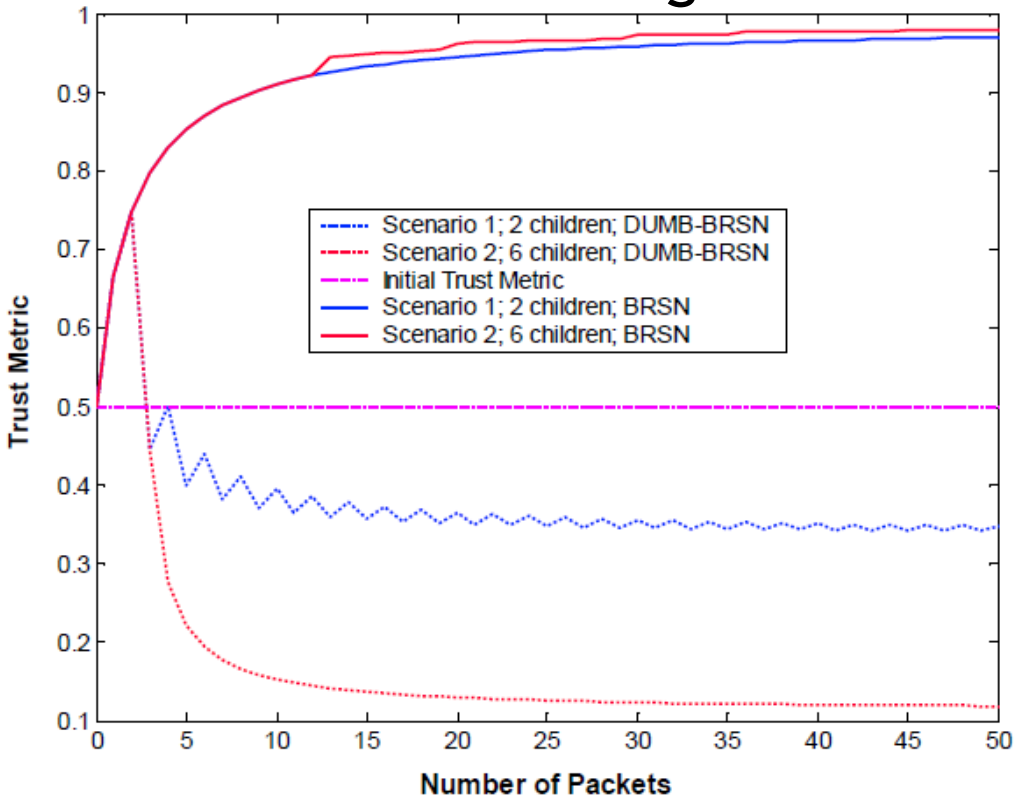
- **Goal:** prevent computation and dissemination of trust/reputation values, denying any supported protocol/application

- **Means:** overloading the system or blocking messages in some way (typically through some existing form of DoS attack)
  - Ex: flood reputation updates so nobody can process them all; jam/drop reputation update messages

- **Assumptions:** (i) sufficient resources, (ii) collusion as in DDoS

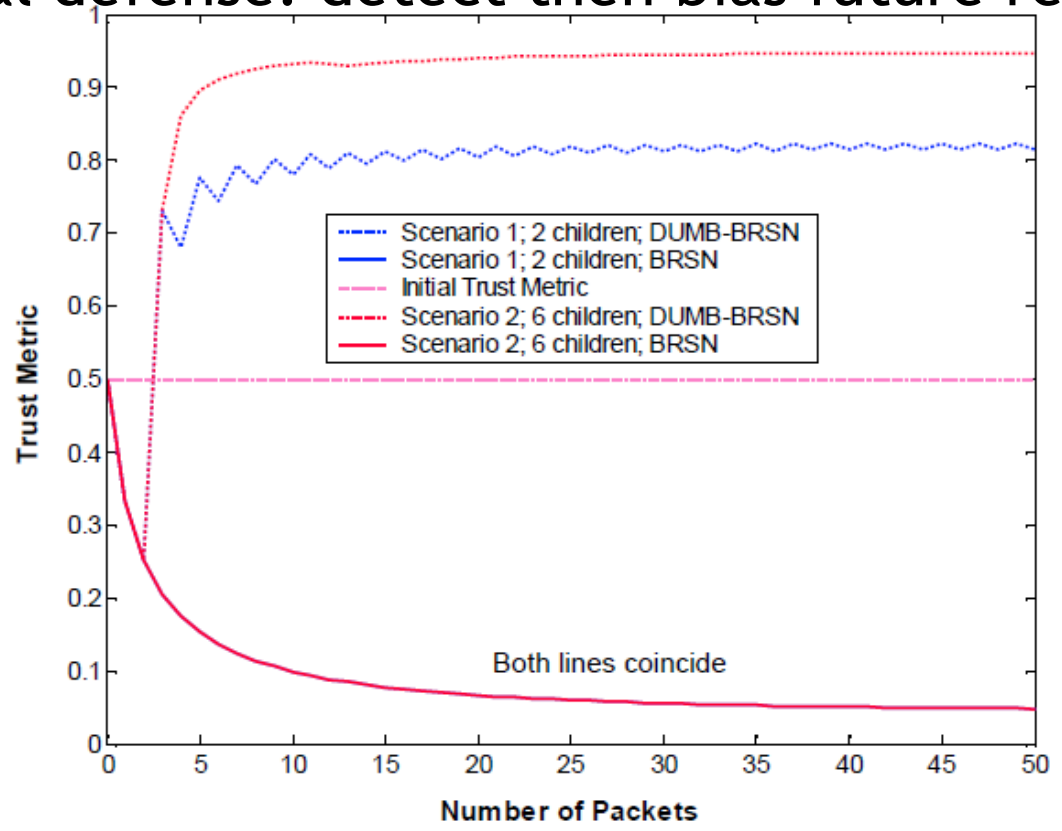
Attacks on Reputation

- Malicious or unfair negative reputation
  - "Bad-mouthing attacks" [Ganeriwal & Srivastava, 2004]
  - Potential defense: eliminate negative feedback



Attacks on Reputation

- Malicious or unfair positive reputation
  - "Ballot-stuffing attacks" [Ganeriwal & Srivastava, 2004]
  - Potential defense: detect then bias future results



How to defend against these attacks on trust/reputation?

Defense of Sybil Attacks

- Problem: many of the above attacks are due to Sybil-like behavior (multiple identities per node)
  - Allows each attacker to present multiple opinions

- Potential defense:
  - Centralized or distributed identity management, potentially binding the ID to the device, address, or other static parameter
  - IDs can also be based on social “web of trust”

Mitigating False Rumors

- Problem: attackers can fabricate false rumors to alter reputation computation

- Defense #1 (mitigating generation):
  - Bind reports using cryptographic protection such as a digital signature (for accountability)

- Defense #2 (mitigating spreading):
  - Filter out reports that don't match others using voting or consistency with direct observations
  - Don't forward any reports that are inconsistent

Mitigating Short-Term Abuse

- Problem: attackers can misbehave for a relatively short time then play nice (or reset with a new ID) to restore reputation

- Potential defenses:
  - New actors start with low reputation and need to build up before getting service
  - Enforce strict penalties on misbehavior with slow rate of improvement

Mitigating DoS Attacks

- Problem: DoS attacks on trust dissemination and update can prevent reputation building

- Potential defenses:
  - Distribute dissemination/update tasks over multiple actors for diversity
  - Employ common network reliability and DoS mitigation strategies
    - ACK/NACKs, multi-path routing, gossip mechanisms, error-correcting codes, etc.

How do trust and reputation systems apply to networking?

Trust-Based Networking

- Network nodes can be selective about communication and networking decisions by using trust-based policies
  - A node can decide to locally communicate only with nodes that it trusts above a threshold τ
  - A node can construct/select routing paths using an aggregate path trust/reputation metric

Watchdog & Pathrater

[Marti et al., 2000]

- Watchdog monitors forwarding by overhearing subsequent transmissions

  - If A→B and B→C, then A can listen to and analyze B's forwarding behavior

- Pathrater uses observed statistics to choose which paths are most reliable
  - Can be helpful in route selection for source routing, e.g. DSR

Issues with Watchdog

- Unsurprisingly, watchdog+pathrater is not robust to misbehavior or network error on its own
  - Collisions, fading, asymmetric links, and many other events are treated as misbehavior

- Not robust to many types of attacks
  - Slander/framing attacks by a watchdog affect the aggregated path rates

CORE

[Michiardi and Molva, CMS 2002]

- CORE combines direct and indirect trust values wrt a collection of different functions
  - E.g., forwarding, route discovery, network management, location management, etc.
  - Builds on the Watchdog mechanism using a requester-provider model

CORE Requests

- When a requester makes a service request:
  - Watchdog monitors the request and reply
  - Provider accepts request only if reputation value of requester is high enough
  - Watchdog can update provider of reputation value if it changes, e.g. if requester is DoS-ing provider

- CORE prevents some attacks by only allowing positive reports to propagate; negative reports only go 1 hop

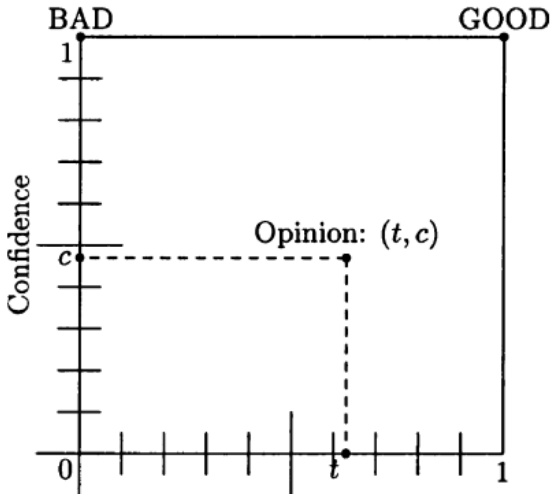
CORE Limitations

- CORE adopts all the limitations of the watchdog
- Not scalable, as a watchdog is needed in every neighborhood
  - Also, watchdogs need global (or at least E2E) info
- Mobility can break CORE

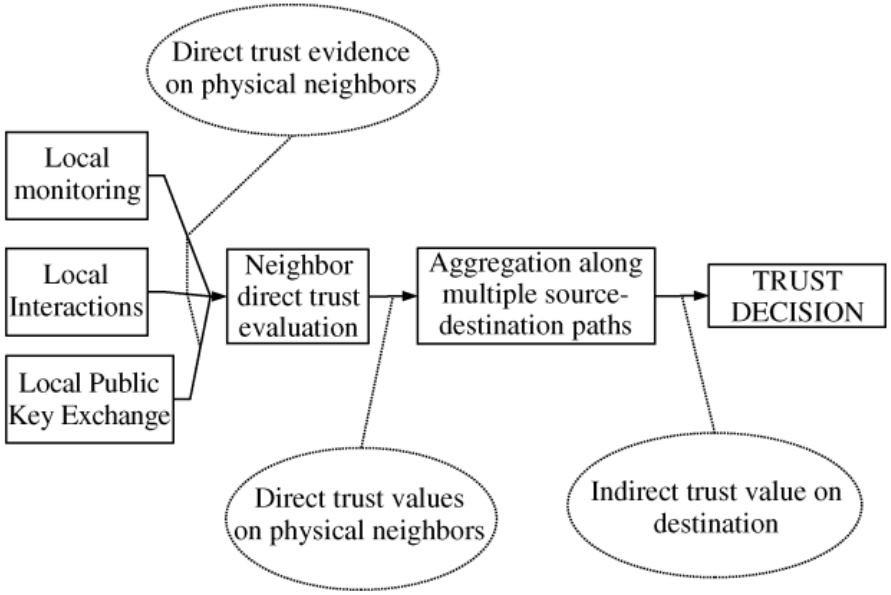
Trusted MANET Routing

[Theodorakopoulos and Baras, JSAC 2006]

- Routers report trust of each next hop, including confidence in the estimate



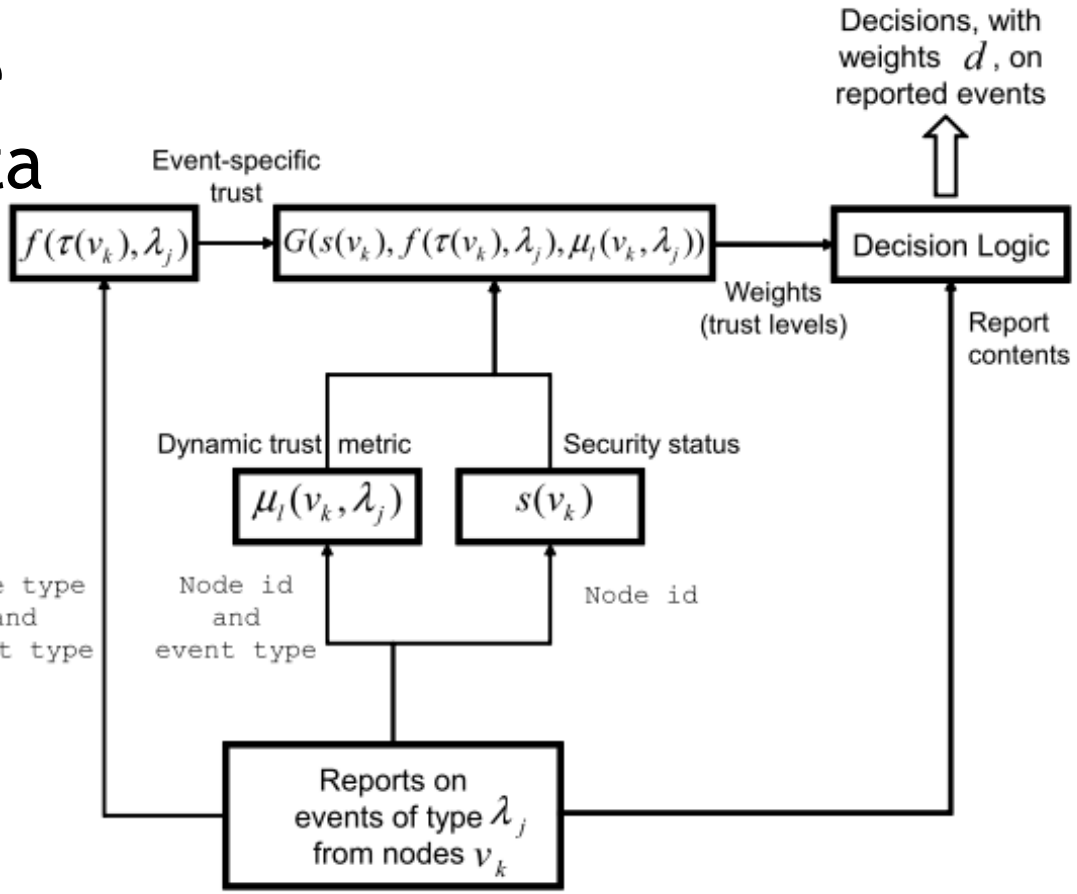
- Source computes an aggregate over each of several paths to inform path selection



Data-Centric Trust in VANETs

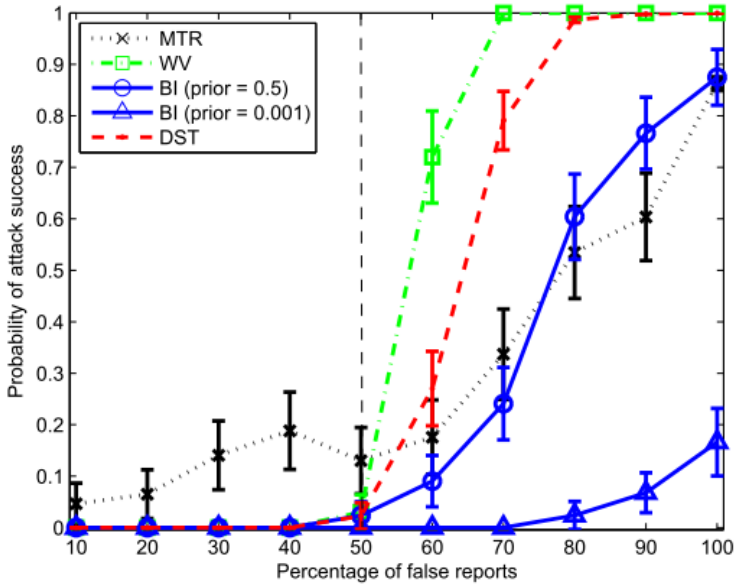
[Raya et al., Infocom 2008]

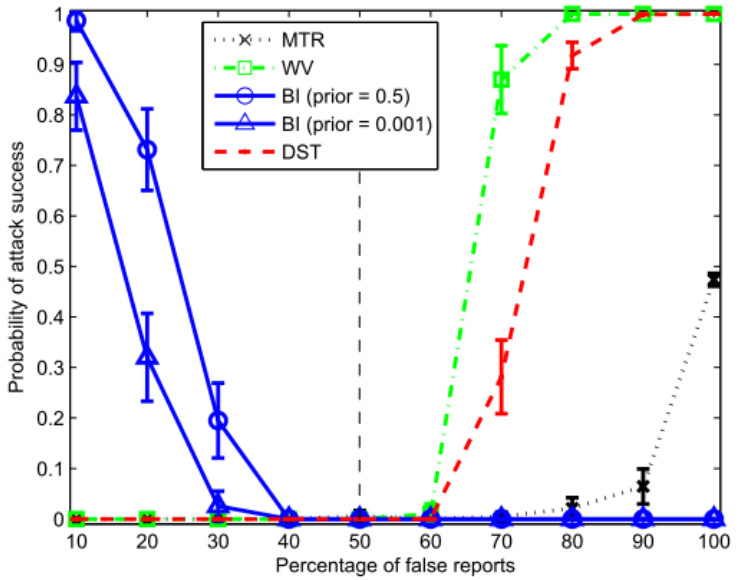
- In highly dynamic environments like VANETs, trust can be computed on the data instead of the actors
  - Including dynamic factors such as location and time
  - All relevant factors can be individually weighted to come up with a trust value in each piece of data



Data Trust Evaluation

- Paper provides a framework for evaluating the trust levels using a number of different statistical methods
  - Different techniques provide resilience against certain types of attacks





Open Questions

- How to choose the right type of reputation dynamics for a given system/task/data type?
- How to detect the events that cause reputation to increase and decrease?
- How to mitigate the effects of detection error?
- Is reputation effective given the well-known attacks (slander, lying, etc.)?

Mar 3, 8, 10:

NO CLASS – SPRING BREAK!

Mar 15:

Wireless Transport Security