Formalizing and Managing Trust

What is trust and how can we make use of trust information?

ibu radempa

2017-06-29

Avoid the control trap: *"Trust is good, control is better."*

Control is omnipresent, e.g.:

- surveillance; mutual policing
- regulation (cybernetic paradigm)
- computers controlling humans (illiteracy, AI)

Beyond denying corp. & gov. control, what can we do?

Self-control? Self-controlled spaces easily become social islands.

⇒Try patterns without control.

Real social networks are based on trust (social glue).

Formalizing
and
Managing
Trust

ibu radempa

Introduction
Building
trust
Concepts
Agents
On measures
Propagation
Topics
Managing
trust
information

Current socio-cultural background

The terrorism narrative enforces control over trust.

The market paradigm is in many ways opposite to trust:

- (over)simplification: assume independent exchange transactions
- discontinuity: every party can leave after a transaction
- anonymity: knowing the other party well is not required
- control: through contract-based penalty threats (backed by state authorities)

Visibility bias: Trust is private and thus less present in public discourse.

$\implies$ Relevance of trust not easy to see

Formalizing
and
Managing
Trust

ibu radempa

Introduction

Building
trust

Concepts

Agents

On measures

Propagation

Topics

Managing
trust
information

- [Building trust]
- ~~Creating conditions supporting the emergence of trust~~
- Formalizing trust
- ~~Reputation networks, e-Commerce~~
- Agents & decentralized trust propagation mechanisms
- Managing and using trust

Formalizing
and
Managing
Trust

ibu radempa

Introduction
Building
trust
Concepts
Agents
On measures
Propagation
Topics
Managing
trust
information

Some examples related to trust

- A trusts B to appear at an appointment on time
- A trusts B to help her in case she is ill
- A trusts B to treat personal information confidential
- A trusts B to not cooperate with secret service
- A trusts B that she/he ...
  - tells the truth
  - does what she says
  - does not talk badly about absent persons
  - does not change her mind frequently in basic matters
  - brings lent things back in time
  - is not severely affected by spectacular news and does not follow trends because others do (autonomy, resistance against manipulation)
  - is reliable in communication (answering phone calls, email, etc.)

Formalizing
and
Managing
Trust

ibu radempa

Introduction

Building
trust

Concepts

Agents

On measures

Propagation

Topics

Managing
trust
information

Examples of using trust information

- Can I trust this software package? (Does it what it claims? How secure is it?)
- Can I cooperate with you? (Are you knowledgeable and reliable?)
- Can I share confidential knowledge with you? (Will you keep it private?)
- Who has trusted you and what were their experiences?
- Show me all experts in $subject trusted by at least two of my friends (whom I trust w.r.t. $subject ...).
- Who of the people I trust does Alice trust?
- Show me evidence for (dis)trust in Bob.
- Let me review my trust statements older than 6 months.
- Which arbitration panels do we trust?

Formalizing
and
Managing
Trust

ibu radempa

Introduction
Building
trust
Concepts
Agents
On measures
Propagation
Topics
Managing
trust
information

Identity and trust are connected:

- In order to trust Alice's claims I have to know that I'm talking to Alice.
- In order to know that I'm talking to Alice I have to trust her not to give her (digital) identity to somebody else.

Identity and trust are sometimes mixed up:

- SSL certificate hierarchies certify identities
- GnuPG / PGP allows to build a web of identity

Read more on defining identity.

- vision of the trustee
  - experiences from past interactions
  - factual ability (availability, competence, expertness)
  - ability to be reliable: impeding conditions (disease, …)
  - motivations to be reliable: intentions, benevolence, goodwill, altruism, caring
  - motivations to lie/fool: competition, financial advantage, adverse goals
  - consistency of communication/behaviour
  - does the trustee express doubt where unsure?
- general or particular credibility of the trustee
  - reliability as information source
  - promises of the trustee w.r.t. future behaviour
- dependence on trustee
- sharing something (in theory and practice), value congruence

Some criteria for emergence trust (2)

- trustee's traits
  - fairness, integrity, loyalty, honesty, moral integrity
  - discreetness, confidentiality
  - openness, receptivity
- understanding the trustee
  - congruity
  - knowing the situation of the trustee (requires continuous communication)
  - knowing the trustee's needs and desires
- recommendations
- personal attraction
- autonomy
- experience of past interactions (few disappointments)
- stability

Formalizing
and
Managing
Trust

ibu radempa

Introduction

Building
trust

Concepts

Agents

On measures

Propagation

Topics

Managing
trust
information

Can I trust person X?

- How often do I communicate with X?
- How well do I know X's needs and desires?
- How well do I know X's limitations and faults?
- How often have I been dependent on X, with which impact, and what were the experiences? How often was I disappointed?
- How well does X understand what is important for me?
- How balanced is the general inter-dependency between me and X?
- For how long do I know X and will X remain in my social neighborhood?
- Do I know X's friends and conditions of living?
- ...

Formalizing
and
Managing
Trust

ibu radempa

Introduction
Building
trust
Concepts
Agents
On measures
Propagation
Topics
Managing
trust
information

Formalizing trust

- 1994 Stephen Marsh: *Diss. on computational trust*
- 2005 Lea Viljanen: *Towards an Ontology of Trust*
- 2005 Robert Demolombe: *Reasoning About Trust: A Formal Logical Framework*
- 2006 Huang, Fox: *An Ontology of Trust – Formal Semantics and Transitivity*
- 2009 Castelfranchi, Falcone, Lorini: *A Non-reductionist Approach to Trust*
- 2014 Ceolin, Nottamk, Fokkink, Maccatrozzo: *Towards the Definition of an Ontology for Trust in (Web) Data*

[...] a call to arms to the Trust researchers and developers to turn their thoughts to Untrust, Distrust and Mistrust, to consider them as unique entities in their own right, [...]

Marsh, Dibben 2005

Terminological confusion; informally:

Definition

(1) Trust is the expectation on one or more persons to act in a range of ways under a range of conditions.
(2) Trust is a psychological state comprising the intention to accept vulnerability based upon positive expectation of the intentions or behavior of another.
(3) Trust is the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party. (Cf. DOI:10.1145/1151454.1151499)

Trust models are simplifications which are formalizable. Usually:

$$Trust(\text{trustor}, \text{trustee}, \text{action}, \text{context})$$

Formalizing
and
Managing
Trust

ibu radempa

Introduction
Building
trust
Concepts
Agents
On measures
Propagation
Topics
Managing
trust
information

Demolombe: *Reasoning About Trust: A Formal Logical Framework*

p some proposition

$B_a p$: agent a believes that p is the case

$I_{a,b} p$: agent a has informed agent b that p is the case

Sincerity: The agent b is *sincere with regard to a for p* iff if b informs a about p then b believes p.

$$I_{b,a}p \rightarrow B_b p$$

Cooperativity: The agent b is *cooperative with regard to a for p* iff if b believes p then b informs a about p.

$$B_b p \rightarrow I_{b,a}p$$

Credibility: The agent b is *credible (or competent) about p* iff if b believes p then p is the case.

$$B_b p \rightarrow p$$

Vigilance: The agent b is *vigilant about p* iff if p is the case then b believes p.

$$p \rightarrow B_b p$$

Validity: The agent b is *valid with regard to a for p* iff if b informs a about p then p is the case.

$$I_{b,a} p \rightarrow p$$

Completeness: The agent b is *complete with regard to a for p* iff if p is the case then b informs a about p.

$$p \rightarrow I_{b,a} p$$

$K_a p$: agent a strongly believes that p is the case

$$Tsinc_{a,b}(p) \equiv K_a \left(I_{b,a}p \rightarrow B_b p \right)$$

$$Tcoop_{a,b}(p) \equiv K_a \left(B_b p \rightarrow I_{b,a}p \right)$$

$$Tcred_{a,b}(p) \equiv K_a \left(B_b p \rightarrow p \right)$$

$$Tvigi_{a,b}(p) \equiv K_a \left(p \rightarrow B_b p \right)$$

$$Tval_{a,b}(p) \equiv K_a \left(I_{b,a}p \rightarrow p \right)$$

$$Tcomp_{a,b}(p) \equiv K_a \left(p \rightarrow I_{b,a}p \right)$$

We have $Tsinc_{a,b}(p) \wedge Tcred_{a,b}(p) \rightarrow Tval_{a,b}(p)$ and $Tvigi_{a,b}(p) \wedge Tcoop_{a,b}(p) \rightarrow Tcomp_{a,b}(p)$.

Postulating suitable axioms Demolombe extends propositional logic in p to a modal logic of trust.

Huang, Fox: *An Ontology of Trust – Formal Semantics and Transitivity*

Employ situational calculus, "a logic language specfically designed for representing dynamically changing worlds": $holds(f(x), s)$ (predicate $f(x)$ holds in situation s)

"A fluent is a property (of the world) whose value is dependent on situations. In other words, a fluent dynamically changes when the situation does. The situation, in turn, changes when an action is performed by agent(s) in the world."

"A: the set of actions; S: the set of situations; F: the set of fluents; E: the set of entities; D: the set of domain objects."

Distinguish trust in performance and trust in belief.

Formalizing
and
Managing
Trust

ibu radempa

Introduction

Building
trust

Concepts

Agents

On measures

Propagation

Topics

Managing
trust
information

Trust definition (Huang, Fox)

believe(d, x) $\subseteq$ E $\times$ F : Entity d believes that thing x is true.

trust_p(d, e, x, k) $\subseteq$ E $\times$ E $\times$ F $\times$ F : In context k, trustor d trusts trustee e on thing x made by e. x is called expectancy.

trust_b(d, e, x, k) $\subseteq$ E $\times$ E $\times$ F $\times$ F : In context k, trustor d trusts trustee e on thing x that e believes.

has_p_tr(d, e, x, k) $\subseteq$ E $\times$ E $\times$ F $\times$ F : Trustor d has trust in performance type of inter-individual trust relationship with trustee e.

has_b_tr(d, e, x, k) $\subseteq$ E $\times$ E $\times$ F $\times$ F : Trustor d has trust in belief type of inter-individual trust relationship with trustee e.

made(x, d, k) $\subseteq$ F $\times$ E $\times$ F : Information x is made by entity d in context k.

memberOf(e, o) $\subseteq$ E $\times$ E : Entity e is a member of o, an organization entity or an entity group with a same set of characteristics.

Frame Castelfranchi, Falcone, Lorini use "state sequences":

"next time" operator X specifies facts that are true in the next state:

$$X\varphi \equiv \bigwedge_{i \in AGT, \alpha \in ACT} After_{i:\alpha}\varphi$$

BNF: $\varphi \equiv p \,|\, \neg\varphi \,|\, \varphi \vee \varphi \,|\, After_{i:\alpha}\varphi \,|\, [i:k]\varphi \,|\, Goal_i\varphi$

$p, q, \ldots$ are atomic formulas

$\alpha, \beta, \ldots$ are actions (ACT)

$i, j, \ldots$ are agents (AGT)

Examples:

p="Bob is alive", $After_{Alice:shoots\_into\_Bobs\_heart}$ p="Bob is dead"

$[i:k]\varphi$ means: φ is true in all worlds that according to agent i are possible at least degree k ($k \in \{1, \ldots, n\}$)

Core trust (depends on j's internal and external preconditions):

i distrusts j to ensure φ by performing α

$$CoreTrust(i, j, \alpha, \varphi) \equiv$$

$$Goal_i X\varphi \wedge Plaus_i \left(After_{j:\alpha}\varphi \wedge Int_j(\alpha) \wedge IPre(j, \alpha) \wedge EPre(j, \alpha) \right)$$

Generalized core trust (i trusts j to ensure φ):

$$CoreTrust(i, j, \varphi) \equiv \bigvee_{\alpha \in ACT} CoreTrust(i, j, \alpha, \varphi)$$

Distrust (i distrusts j to ensure φ by performing α):

$$Distrust(i, j, \alpha, \varphi) \equiv$$

$$Goal_i X\varphi \wedge Plaus_i \left(\neg After_{j:\alpha}\varphi \vee \neg Int_j(\alpha) \vee \neg IPre(j, \alpha) \vee \neg EPre(j, \alpha) \right)$$

Generalized distrust (i distrusts j to ensure φ):

$$Distrust(i, j, \varphi) \equiv \bigwedge_{\alpha \in ACT} Distrust(i, j, \alpha, \varphi)$$

Mistrust (i mistrusts j to ensure $\neg\varphi$ by performing α):

$$Mistrust(i, j, \alpha, \varphi) \equiv$$

$$Goal_i X\varphi \wedge Plaus_i \left(After_{j:\alpha}\neg\varphi \wedge Int_j(\alpha) \wedge IPre(j, \alpha) \wedge EPre(j, \alpha) \right)$$

Generalized mistrust (i mistrusts j to ensure $\neg\varphi$):

$$Mistrust(i, j, \varphi) \equiv \bigvee_{\alpha \in ACT} Mistrust(i, j, \alpha, \varphi)$$

Formalizing
and
Managing
Trust

ibu radempa

Introduction
Building
trust
Concepts
Agents
On measures
Propagation
Topics
Managing
trust
information

- relation between two agents (trusting, trusted)
- depends on particular expected bahaviour (expected action)
- depends on context (as seen by the trusting party)
- not necessarily symmetric
  (particular situation vs. general trust between friends)
- in general not transitive (which expectations can be propagated?)
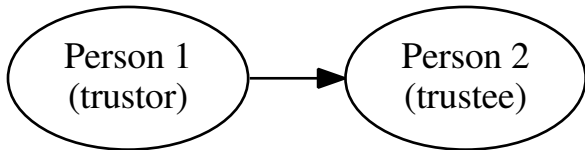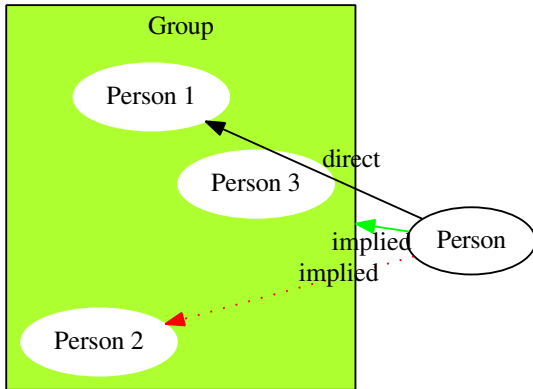
direct trust

Example: I trust a friend whom I know for years
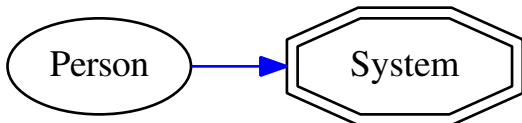
system trust

Examples: road user trusts bridge architect; patient trusts
physician; wikipedia

- Single humans fail; groups may fail less.
- A group usually declares its goals and may make promises.
- A group ususally can be referred to by its name and judgements usually are not considered private.
- Trustworthiness depends on the group's stability.

- internal (communication) structures
- goals / intentions
- norms
- cooperation, reviewing / checking (e.g., code review, quality control, ...)
- commitment (vs. arbitrariness in "social networks")

... and fails:

- popularity measures
- school marks

... because it's about subjects.

In today's internet the standard answer to the question of trust is *reputation*. Reputation is understood as a single value (often on a scale from 1 to 5 stars) attributed to an agent j. It is based on what (mostly untrusted) others say about the agent (whom they usually don't know).

Context and action are assumed to be normalized and the trustor is removed by averaging:

$$Trust(i, j, \alpha, c) \longmapsto Trust(j)$$

In local trust metrics the trustor is not removed.

A well know metric is the Advogato metric.

- Browser plugin: BEWARE!
- Term used since 1992 in PGP, cf. wikipedia
- In a web of trust, each user of the system can choose for herself whom she elects to trust, and whom not.
- analysis of the strong set in the PGP web of trust
- Rebooting Web-of-Trust: a new model of decentralized self-sovereign identity
- Mostly irrelevant: Domination of a centralized PKI

Assume Bob trusts Claire that she will repair his bicycle.

Can Alice trust Claire to repair her bicycle?
Yes, if:

- Claire is intending to perform this different action (different bibcycle).
- Bob thinks that this intent is plausible.
- Bob communicates this to Alice.
- Alice has a general trust (in believe) in Bob's statements.
- Alice thinks that the preconditions for Claire to repair the bicycle are fulfilled.

Trust propagation is the principle by which new trust relationships can be derived from pre-existing trust relationship.

There is also *trust fusion*, where two agents build a consenus of their trust in a 3rd agent. (See Jøsang, Marsh, Pope (2006).)

Group trust assertions by topic

Assume t2 is a subtopic of t1. Then

$$Trust(i, j, t1) \implies Trust(i, j, t2)$$

Formalizing
and
Managing
Trust

ibu radempa

Introduction
Building
trust
Concepts
Agents
On measures
Propagation
Topics
Managing
trust
information

Store, retrieve and combine

statements related to trust (expectations and/or trust actions)

between identifiable agents

in order to help with decisions whether to trust.

Public trust statements

Semantic web public approaches...

2002 A module for defining trust relationships in FOAF

2011 Web Of Trust RDF Ontology

Formalizing
and
Managing
Trust

ibu radempa

Introduction
Building
trust
Concepts
Agents
On measures
Propagation
Topics
Managing
trust
information

Privacy-agnostic trust propagation: FOAF

Example from Golbeck, *Trust Networks on the Semantic Web* (2003):

```
<Person rdf:ID="Bob">
    <mbox rdf:resource="mailto:joe@example.com"/>
    <trustsHighlyRe>
        <TrustsRegarding>
            <trustsPerson rdf:resource="#Dan"/>
            <trustsOnSubject
                rdf:resource="http://example.com/ont#Research"/>
        </TrustsRegarding>
    </trustsHighlyRe>
    <distrustsAbsolutelyRe>
        <TrustsRegarding>
            <trustsPerson rdf:resource="#Dan"/>
            <trustsOnSubject
                rdf:resource="http://example.com/ont#AutoRepair"/>
        </TrustsRegarding>
    </distrustsAbsolutelyRe>
</Person>
```

See also: http://xmlns.com/wot/0.1/

Formalizing
and
Managing
Trust

ibu radempa

Introduction
Building
trust
Concepts
Agents
On measures
Propagation
Topics
Managing
trust
information

Functionality for sharing trust (utopia)

Restrict the visibility of information on whom you trust in what respect:

- "You are allowed to obtain any trust statements about me."
- "Everybody can see trust statements regarding topic t about me."
- "My close friends can see all of my trustees."
- Any trustee can reject propagtion of the trust relation.
- "Show me my trust network for topic t."
- "Show me groups related to topic t and their goals/promises."
- "Show me trust actions for trustee j."

Requires a culture of confidentiality.

- A group may certify claims of its members.
- ... and identity of its members

Formalizing
and
Managing
Trust

ibu radempa

Introduction

Building
trust

Concepts

Agents

On measures

Propagation

Topics

Managing
trust
information

How to identify people making robust statements in a domain of knowledge?

Consider all interested in this domain of knowledge and let them decide.