On the Foundations of Trust in networks of Humans and Computers*

Virgil D. Gligor
Carnegie Mellon University
Pittsburgh, PA 15213
gligor@cmu.edu

SERE
NIST, Gaithersburg, MD 2012
June 20, 2012

*joint work with Jeannette Wing, Tiffany Hyun-Jin Kim and Adrian Perrig
Outline

1. What are Trustworthy Systems?
   - (in)security axioms

2. Interactive Trust Protocols on Trustworthy Systems
   - necessary conditions: value, asymmetry, safety

3. Role of Collateral in Interactive Trust Protocols
   - advantages of social (“street-level”) collateral

4. An Example: Street-Level Semantics for Attribute Authentication
   - semantics and visualization

5. Summary and Future Research
   - why trust? why interactive protocols? why street-level?
   - systems, deception and scams, machine learning, trust networks
## (In)Security Axioms

1. There will always (> 15 years) be
   - bugs/features & human “errors” that will lead to security vulnerabilities
   - adversaries (e.g., malware, insiders) willing and able to exploit them

<table>
<thead>
<tr>
<th>Example Technology Feature</th>
<th>‘60s</th>
<th>‘70s</th>
<th>‘80s</th>
<th>‘90s</th>
<th>‘00s</th>
<th>‘10s</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Adversary</td>
<td>opportunistic user</td>
<td>PC, Servers LANs</td>
<td>email, IP connectivity</td>
<td>WWW, mobility</td>
<td>physical systems, smart phones</td>
<td>social networks</td>
</tr>
</tbody>
</table>

- ‘60s: TSS
- ‘70s: opportunistic user, man-in-the-middle attacks
- ‘80s: viruses, worms, outsiders -> insiders
- ‘90s: large-scale DDoS, untraceable attacks
- ‘00s: infrastructure attacks, phone malware
- ‘10s: network deception, large-scale scams
1. There will always (> 15 years) be
   - bugs/features & human “errors” that will lead to security vulnerabilities
   - adversaries (e.g., malware, insiders) willing and able to exploit them

2. There will always be rapid innovation in IT, and it will always lead to low-assurance systems
   - frequent updates of system configurations
     => perennially out-of-date assurances
     (e.g., “high assurance is always available when you no longer need it”)
   - systems comprising components of diverse provenance
     => non-uniform assurances and more attack surfaces
     (e.g., “lemon” apps always will drive high-assurance apps out of the market)
1. There will always (> 15 years) be
   - bugs/features & human “errors” that will lead to security vulnerabilities
   - adversaries (e.g., malware, insiders) willing and able to exploit them in the Internet

2. There will always be rapid innovation in IT, and it will always lead to low-assurance systems
   => frequent updates of system configurations
   => perennially out-of-date assurances
   (e.g., “high assurance is always available when you no longer need it”)
   ⇒ systems comprising components of diverse provenance
   ⇒ non-uniform assurances (“toxic” components?) & more attack surfaces
   (e.g., “lemons” always will drive high-assurance apps out of the market)

3. There will always be
   - large, complex systems whose security is not fully understood by most users

   “in software, only [module] giants survive….” [Lampson, ICSE, 1999]

   “security is fractal: every part is as complex as the whole” [Lampson, CACM 2009]
What are Trustworthy Systems, then?

Systems with Demonstrable Security Properties despite Axiomatic Insecurity of their Commodity Computing Platforms

• properties that hold in the presence of an Adversary; e.g.,
  • malware
  • malicious insiders
Interactive Trust Protocols

Specification: Take Action (e.g., “invest, click, send PII/acct, send problem”) => Receive Service

Either Takes Specified Action or Rejects

Provides Specified Service

Provided Specified Service

Either Takes Specified Action or Rejects

Receiver

Sender

?
Am I talking to the Sender?

Demonstrable Security Property:
User-Verifiable End-to-End Trusted Path
Am I talking to the Sender?

Mouse Click: Accept \{\text{Sender, } PK_{\text{Sender}}\}?
Sources of Malware Today ...

- **Non-Uniform Assurances:**
  - e.g., unpatched systems -> exploits based on buffer overflows, XSS, etc.

- **Features:**
  - e.g., USB Drives, Network Drives; AutoRun/AutoPlay

- **Large Software Systems:**
  - e.g., Microsoft Office (e.g., .ppt, .doc, .xls), Adobe .pdf ...
    viral file infection

- **Human Errors**
  - e.g., social engineering, scams, deception via e-mail, P2P sharing, social networks

Most of Today’s Problems will not Disappear any Time Soon
Honest or Trustworthy (TW) Behavior
= compliance with the protocol specifications

- Both parties are TW => Both are better off after session
Value to Receiver = $T_{WR} > 0$ and Value to Sender = $T_{WS} > 0$

- Future sessions (Rational Receiver Takes Action again)
Asymmetry

• \( unTW \) Sender is **better off** than \( TW \) Sender and
  \( TW \) Receiver is **worse off** after session

  \[
  \text{Gain}_S = unT_{wS} - T_{wS} > 0 \quad \text{and} \quad \text{Loss}_R > 0
  \]

• \( unTW \) Sender => No future sessions (Rational Receiver will “Reject”)

\( \text{unTW} \) Sender: provides spoofed ID, corrupt service; i.e., bad input, malware
Asymmetry persists

Isolation from Sender: bad input, malware discovered

Recovery from bad input, malware

Trustworthiness Evidence (past)
never sends a bad input, malware

Deterred from sending a bad input, malware

Machine? Human?

Correct System & Machine Code
Correct System & User Behavior

TWness Evidence

Evidence

Punishment
Deterrence

Isolation from Sender: bad input, malware discovered

Recovery from bad input, malware

Trustworthiness Evidence (past)
never sends a bad input, malware

Deterred from sending a bad input, malware

Machine? Human?

Correct System & Machine Code
Correct System & User Behavior

TWness Evidence

Evidence

Punishment
Deterrence
Completeness: Behavioral-Trust Primitives

- Beliefs in Sender’s trustworthiness
- Preferences/Aversions
  - Risk
- Betrayal

Primitives be Supported in Human and Computer Networks?
Networking (e-commerce) Practice

Completeness: Behavioral-Trust Primitives

Receiver — Sender

- Beliefs in Sender’s trustworthiness
- Preferences/Aversions
  - Risk
- Betrayal

Trustworthiness evidence

recovery from Sender non-compliance

deterrence => “punishment” => Accountability

we need: <=
Asymmetry persists

Isolation from Sender: bad input, malware discovered

Trustworthiness Evidence (past) never sends a bad input, malware

0% Isolation and 0% Trustworthiness Evidence and 0% Recovery and 0% Deterrence => 100% Trust

Is it ever Safe to Trust the Sender?

Yes, if Trustworthy Behavior is in Rational Sender’s interest
Safety

Isolation from Sender:
bad input, malware discovered

Trustworthiness Evidence (past)
ever sends a bad input, malware

Recovery from bad input, malware

Deterred from sending a bad input, malware

Trust (Belief in Rational Sender’s Trustworthy Behavior)
=> Sender’s Present Value of all Future Sessions > unTw_S

<=>

Sender’s discount rate = r < Tw_S / Gain_S
Safety

Present Value of all Future Sessions

\[ \text{Present Value of all Future Sessions} = \frac{\text{Tw}_S}{(1+r)} + \frac{\text{Tw}_S}{(1+r)^2} + \frac{\text{Tw}_S}{(1+r)^3} + \ldots = \frac{\text{Tw}_S(1+r)}{r} > \frac{\text{unTw}_S}{\text{Gain}_S} \]

- **Trust:** \( r < \frac{\text{Tw}_S}{(\text{unTw}_S - \text{Tw}_S)} = \frac{\text{Tw}_S}{\text{Gain}_S} \)
- **No Trust:** \( r \geq \frac{\text{Tw}_S}{\text{Gain}_S} \)
Problem:

$r$ ?

- $Tw_S/Gain_S \rightarrow 0 \Rightarrow$ no Trust
  - $Gain_S >> Tw_S \Rightarrow unTw_S/Tw_S >> 2$
  - $\Rightarrow$ few future sessions if any $\Rightarrow$ no trust
    - e.g., possible scams, insider attacks

- $Tw_S/Gain_S \rightarrow +\infty \Rightarrow$ Trust
  - $Gain_S \rightarrow 0$
  - $\Rightarrow$ rational Sender has no incentive to be untrustworthy
Role of Collateral: $\text{Gain}_S \rightarrow 0$

(-) **Non-starter**: Sender has to post Collateral (for all Receivers)

(-) **Trusted Third Party**: a bootstrapping challenge

(+)** Deterrence**: rational Sender has no incentive to be unTW

(+/-)** Acceptability**: $\text{Loss}_R \leq \text{Collateral} \Rightarrow \text{Receiver can recover}$

unacceptability: Receiver’s $\text{Loss}_R > \text{Collateral} \Rightarrow \text{Receiver could not recover}$

$\Rightarrow \text{Protocol would not start}$
Role of Social (“Street-level”) Collateral

+ Social Collateral: a Sender-Receiver Social Relation exists
e.g., friend, relative, classmate, co-worker, boss, co-conspirator…
=> (high) present value of future cooperation/sessions

=> Trust protocol always starts

+ A Trusted Third Party is unnecessary

**Deterrence Hypothesis**: Loss of Social Relations
(i.e., loss of social collateral) deters more than the Law
- some support in Hu et al., CACM, vol. 64, no. 6, 2011]

+ Deterrence:
  - Sender’s **loss of social collateral** reduces **asymmetry** of Trust protocol

+ Acceptability:
  - the greater Receiver’s **exposure** to **Loss**, the higher **Social Collateral**
Street-Level Semantics for Attribute Authentication

e.g., attributes:
- Identity
- Certificates
- Address/Location
- Social Connections
- Reputation/Credentials
Accepting an Attribute

Friendship: a social relation
- built-in social collateral
- “street-level” punishment/sanction = loss of future value
Accepting an Attribute

Attribute Authenticity => evidence of tie to Sender

=> strength of tie (social distance) to Sender
(communication frequency, recency, reciprocity, length, common acquaintance)

=>” street-level” punishment of 3rd party C (e.g., spoofed ID, false certificate)

=> loss of endorsement by Sender => loss of value at Receiver
Accepting an Attribute

Example 1: Accepting a 3rd Party Attribute (Certificate) signed by a Friend

Deterrence: \[ SC(A) @ B - SC(C) @ A \geq P, \]
where \( P \geq 0 \) measures friend A’s net loss of collateral if \( \{C, PK_C\}^{SK_A} \) is false

Acceptability: \[ SC(C) @ A \geq T_{Bapp}, \]
where \( T_{Bapp} \) measures loss incurred by B’s application if \( \{C, PK_C\}^{SK_A} \) is false

B accepts A’s authentication of \( \{C, PK_C\}^{SK_A} \)
Visualization of “Tie Strength” Evidence
Example 2 Revisited

- Bob has Sender friends Alice and David
- Bob receives an “invitation” from 3rd Party Charlie
  - Charlie’s “invitation” contains endorsed visual ‘tie strength’ evidence
- Bob accepts Charlie’s “invitation” based on the social collateral it assigns to the “tie strength” between Alice and Charlie and David and Charlie
What Does Bob see?

Visualized parameters:
- Frequency of communication (y axis)
- Length of relationship (x axis)
- Reciprocity of communication (circles)
- Selected mutual friends (individual graphs: Alice, David)
- Recency of interaction (leftmost points on x axis)

Average (normalized)

Recency & Time span (log)

Reciprocal communication
- One-way communication from Charlie
- One-way communication from others to Charlie

on-line = OSN, e-mail, P2P;
off-line = physical encounter, phone
Usability: A Facebook Example

- Mechanical Turk-based user study result: 93 participants
  - 84.9% understood “tie strength” on our graph
  - 90.3% would **not** accept “invitations” below the average communication frequency
  - 60.2% felt in control of their privacy in confirming “strength of ties”
  - 82.8% mentioned that our authentication application was easy to use
  - 88.2% indicated that our visual evidence was useful
  - 83.8% indicated that they would use our application before accepting “invitations”
Why Trust?

1. Trust Correlates with Wealth
   - countries where people _trust more_ have _higher GDP_
   - measured trust: _surveys_ (e.g., German Socio-Economic Panel, US General Social Survey, World Value Survey)

2. Network Interpretation
   - _new_ trust relations => larger pool of services, more cooperation, “network effect,” increased competition, _productivity_, innovation, markets and ultimately _economic development/wealth_

3. New Focus For Security Research
   - _past_: most security researchers have been merchants of fear! We’re good at it!
   - _future_: security infrastructures that promote _new_ trust relations (and cooperation)
     Safety Analogy:
     air breaks in railcars (1896), automated railways signals and stops (1882)
     => safe increase in train speeds, railroad commerce, economic opportunities
   - _goal_: seek _security mechanisms that create new value, not just prevent losses_
Future Research

1. Systems – Other roots of trust: software roots of trust
   - TPM are not useful for device controllers and power-challenged devices
   - explore security mechanisms without secrets
   - “simplify” provably complex (e.g., crypto) problems by using valid trust assumptions

2. Understand on-line deception and scams
   - initial work by Stajano and Wilson
   - interactive scams have trust-protocols w/ failed safety conditions

3. Explore machine learning techniques for scam detection
   - other areas than intrusion detection; e.g., advice to users
   - insider attacks explained

4. Trust Networks
   - explore social collateral and relations for deterrence and recovery