Carnegie Mellon

CyLab
www.cylab.cmu.edu

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, Gaithrsburg, MD 2012
June 20, 2012

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

Outline

Carnegie Mellon
CyLab
www.cylab.cmu.edu

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**

	'60s	'70s	'80s	'90s	'00s	'10s
Example Technology Feature	TSS	PC, Servers LANs	email, IP connectivity	WWW, mobility	physical systems, smart phones	social networks
New Adversary	opportunistic user	man-in-the -middle attacks	viruses, worms, outsiders -> insiders	large-scale DDoS, untraceable attacks	infrastructure attacks, phone malware	network deception, large-scale scams

(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

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 perennially out-of-date assurances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)

(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 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

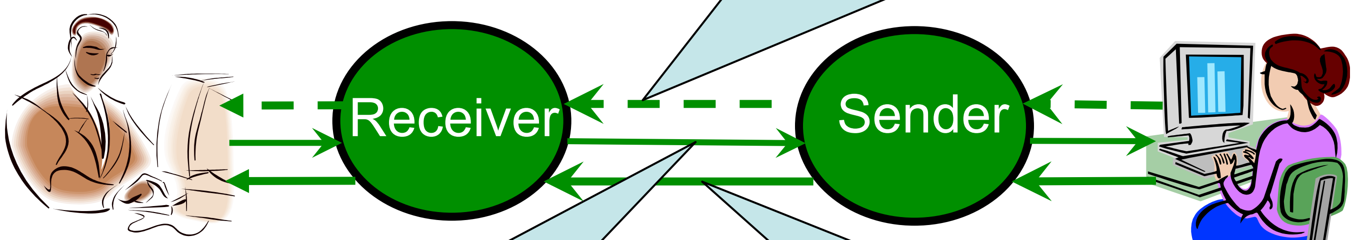
- **properties that hold *in the presence of an Adversary; e.g.,***
  - ***malware***
  - ***malicious insiders***

Interactive Trust Protocols

Carnegie Mellon

CyLab
www.cylab.cmu.edu

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

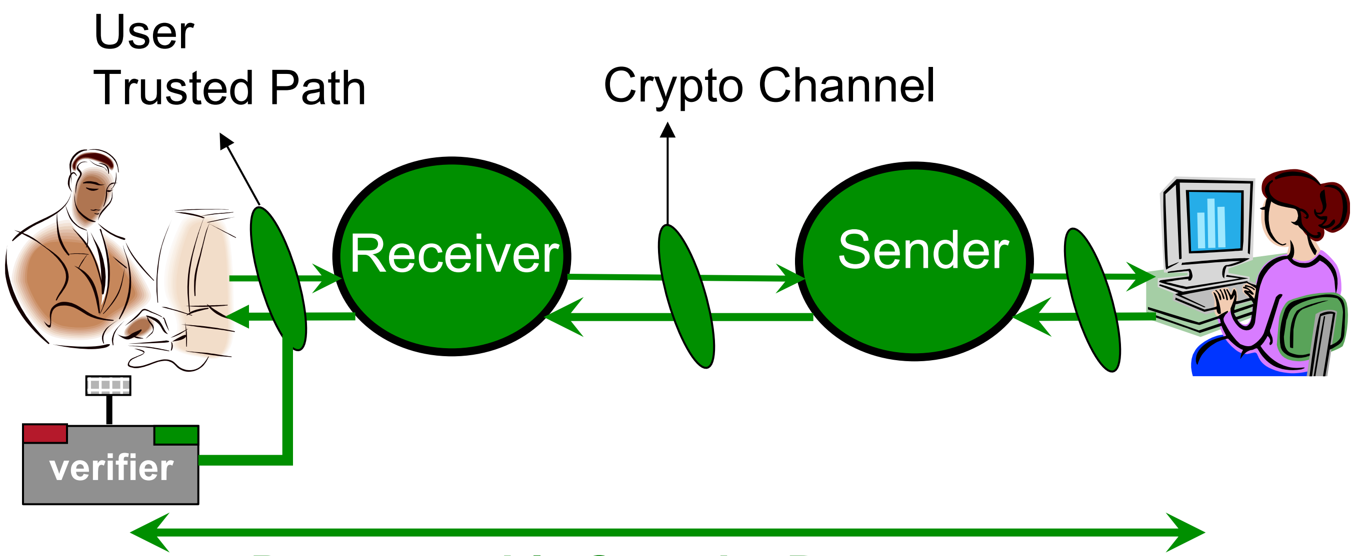


Either **Takes Specified Action** or **Rejects**

Provides Specified Service

?

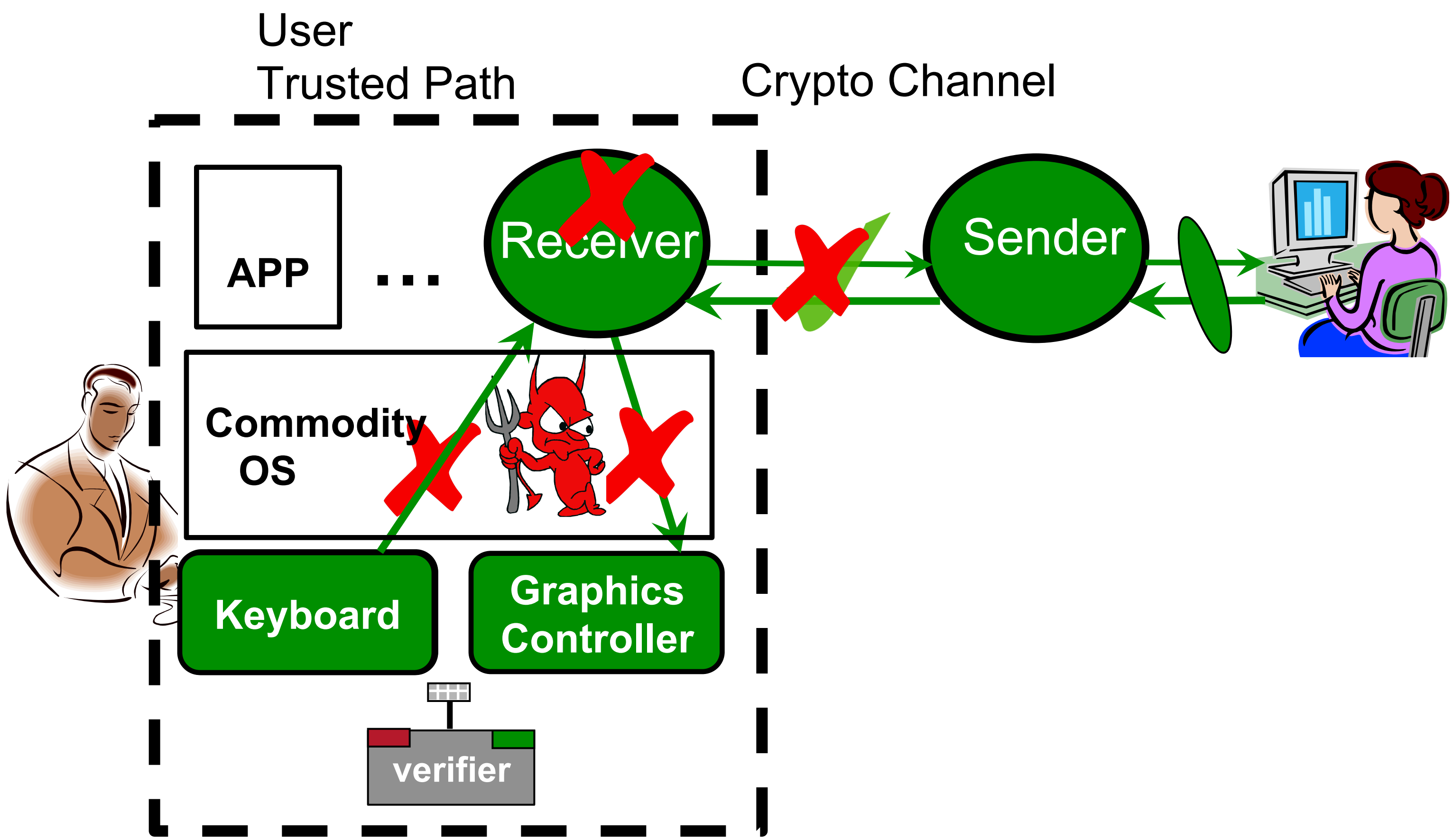
Am I talking to the Sender?



Demonstrable Security Property:

User-Verifiable End-to-End Trusted Path

Am I talking to the Sender?



Mouse Click: Accept {Sender, PK_{Sender}}?

Sources of Malware Today …

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

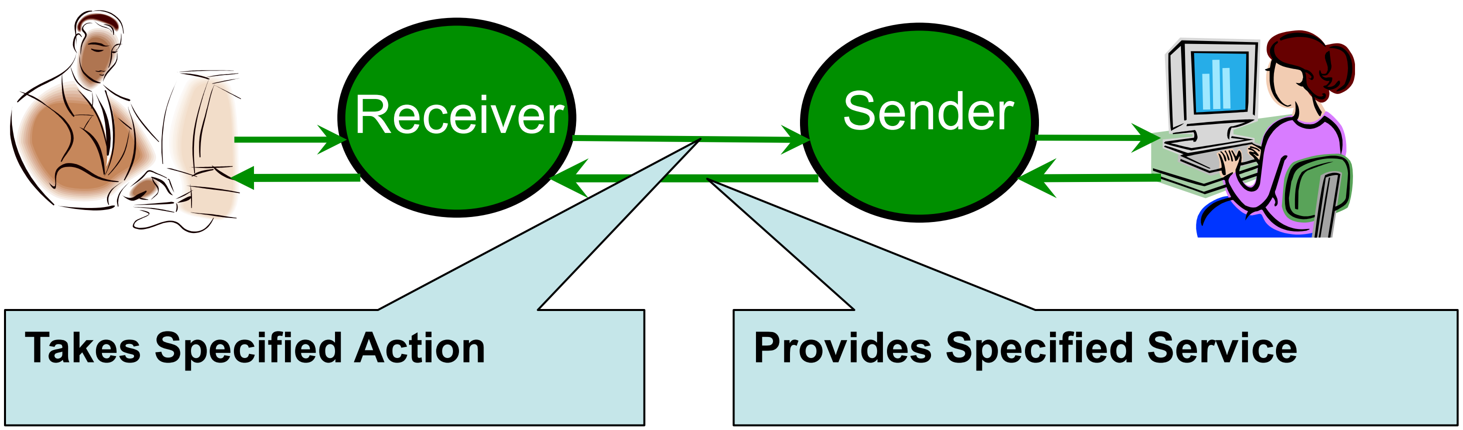
- ***Features*:**
  e.g., USB Drives, Network Drives; AutoRun/AutoPlay

- ***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

Value



Honest **or** ***Trustworthy*** **(TW) Behavior**

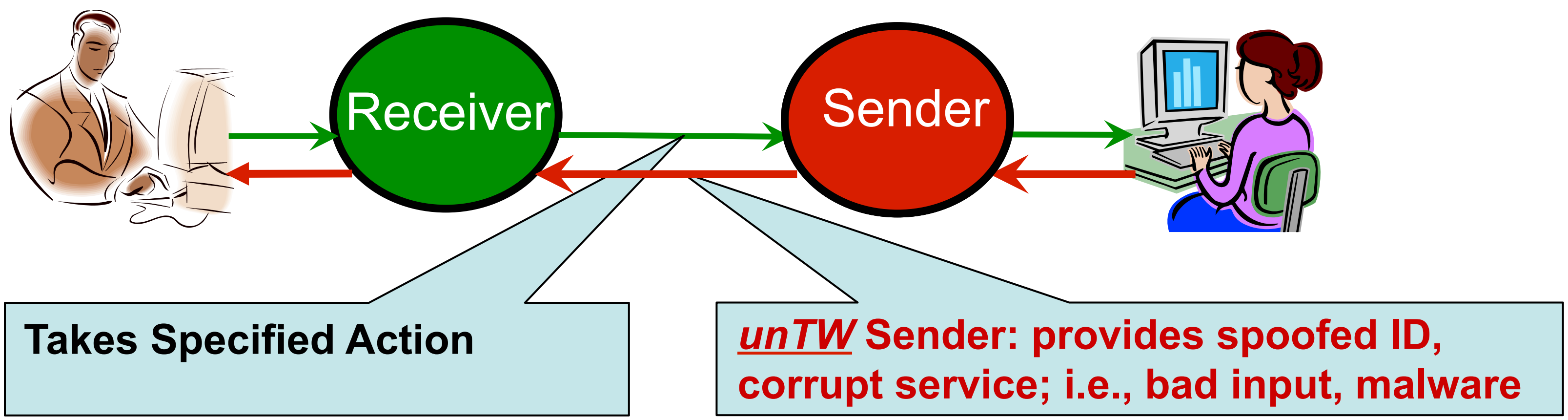
= compliance with the protocol specifications

- ***Both parties are TW => Both are <u>better off</u>*** **after session**
  Value to Receiver = $\mathbf{Tw_R > 0}$ and Value to Sender = $\mathbf{Tw_S > 0}$

- **Future sessions (Rational Receiver Takes Action again)**

Asymmetry

Carnegie Mellon
CyLab
www.cylab.cmu.edu



- ***unTW*** **Sender is** ***<u>better off</u>*** ***than TW Sender*** **and** ***TW*** **Receiver is** ***<u>worse off</u>*** **after session**

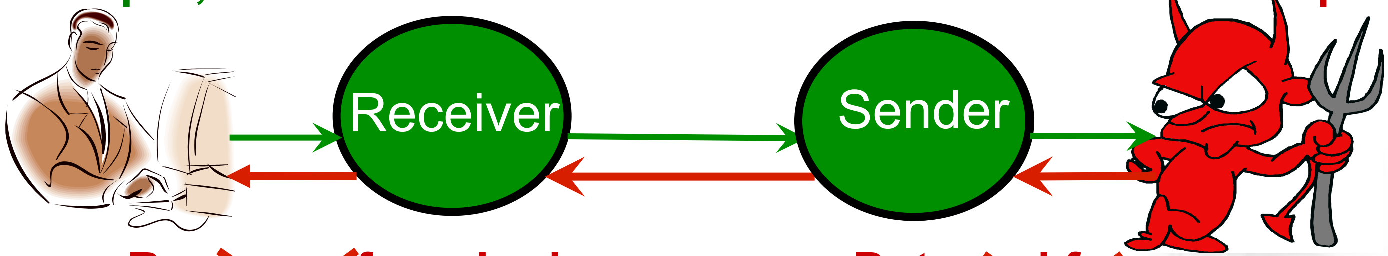
  $Gain_S = unTw_S - Tw_S > 0$ and $Loss_R > 0$

- ***unTW*** **Sender => No future sessions (Rational Receiver will "Reject")**

Asymmetry persists

Isolation ~~from~~ Sender:
bad input, malware discovered

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



Recovery from bad input, malware

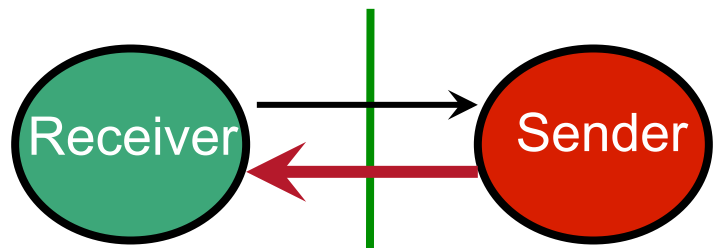
Deterred from sending a bad input, malware

	Machine? ✓	**Human?**	
	Correct System & Machine Code	Correct System & User Behavior	**TWness Evidence**
	——	Punishment	**Deterrence**

Completeness: Behavioral-Trust Primitives

Carnegie Mellon

CyLab
www.cylab.cmu.edu



Behavioral Trust:

- Beliefs in Sender's trustworthiness =>
- Preferences/Aversions
- Risk =>
- Betrayal =>

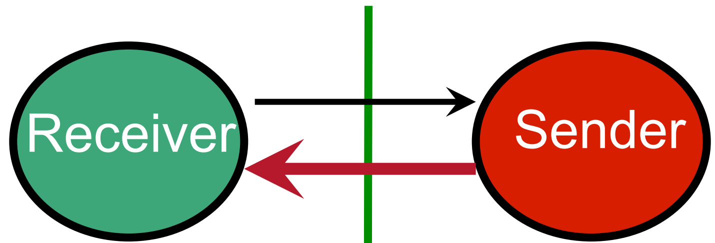
Primitives be Supported in Human and Computer Networks ?

Carnegie Mellon

CyLab

www.cylab.cmu.edu

Completeness: Behavioral-Trust Primitives



Behavioral Trust:

Networking (e-commerce) Practice

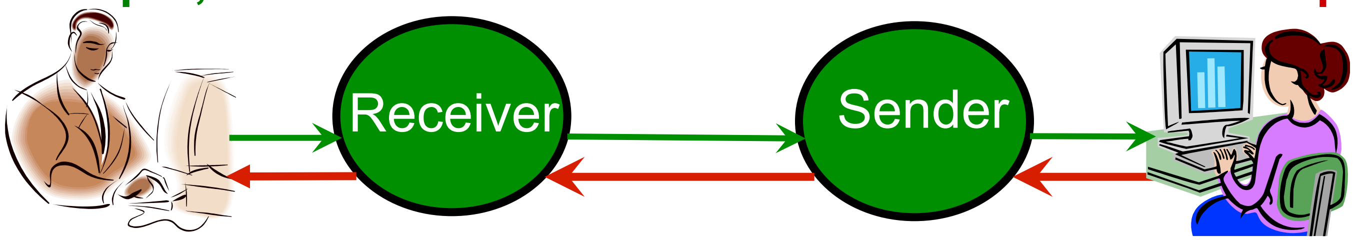
- **Beliefs in Sender's trustworthiness** => ***Trustworthiness evidence*** **=> *Correctness***
- **Preferences/Aversions**
- **Risk** => ***recovery*** from **Sender non-compliance**
- **Betrayal** => **deterrence =>"punishment"=>*Accountability***

we need: <= <=

Asymmetry persists

Isolation from Sender:
bad input, malware discovered

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



Recovery from bad input, malware

Deterred from sending a bad input, malware

*0% Isolation **and** 0%Trustworthiness <u>Evidence</u> **and** 0% Recovery **and** 0% Deterrence => 100% **Trust***

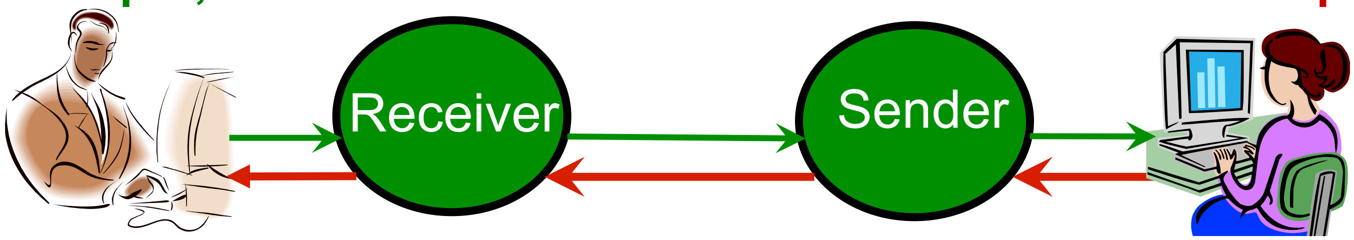
Is it *<u>ever</u>* Safe to Trust the Sender?

Yes, if *Trustworthy Behavior* is in *<u>Rational</u>* Sender's *interest*

Safety

Isolation ~~from~~ Sender:
bad input, malware discovered

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



~~Recovery~~ from bad input, malware

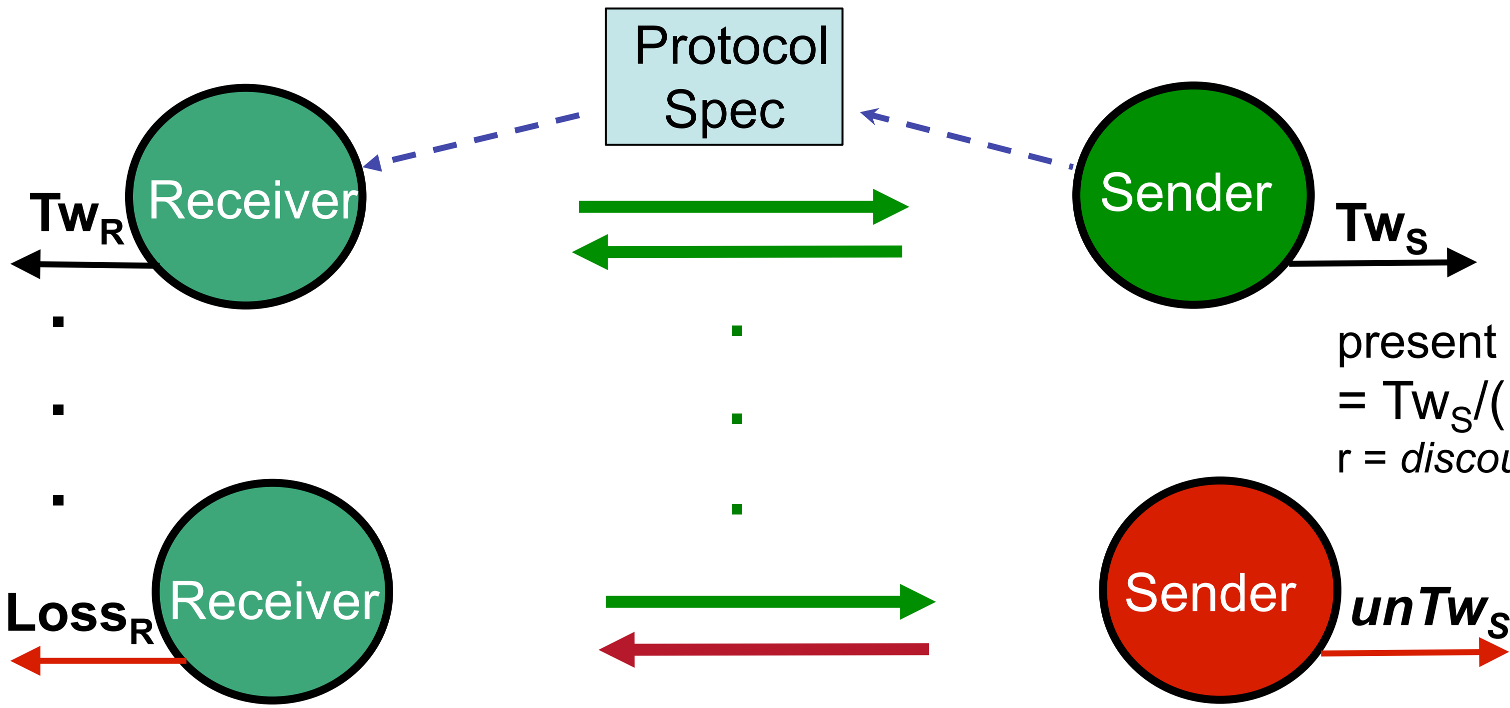
Deterred from sending a bad input, malware

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 session t
$= Tw_S/(1+r)^t$, t = 0, 1, 2…
r = *discount rate*

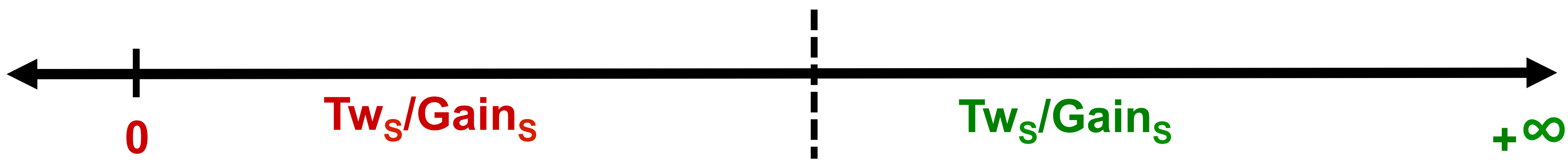
Present Value of all Future Sessions
$= Tw_S + Tw_S/(1+r) + Tw_S/(1+r)^2 + Tw_S/(1+r)^3 + \ldots = \mathbf{Tw_S(1+r)/r > unTw_S}$

- Trust: $r < Tw_S/(unTw_S - Tw_S) = Tw_S/Gain_S$

- no Trust: $r \geq Tw_S/Gain_S$

Safety

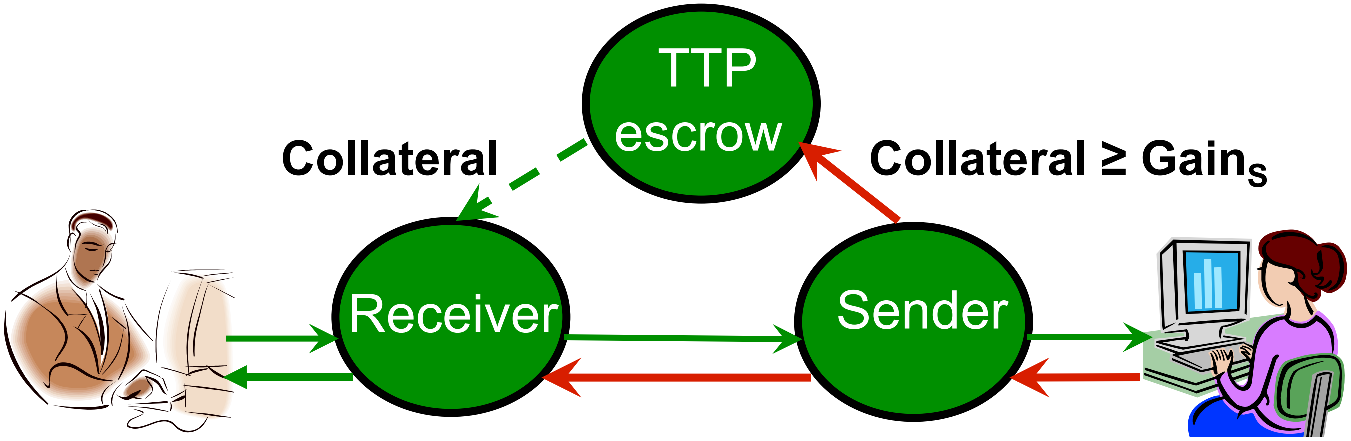
Problem:
r ?



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

- **$Tw_S/Gain_S$ -> + ∞ => Trust**
  - $Gain_S$ **-> 0**
  => *rational* Sender has **no incentive to be untrustworthy**

Role of Collateral: $Gain_S$ -> 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**: $Loss_R$ ≤ Collateral => Receiver can recover*

unacceptability: Receiver's $Loss_R$ > Collateral => Receiver could not recover
=> 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

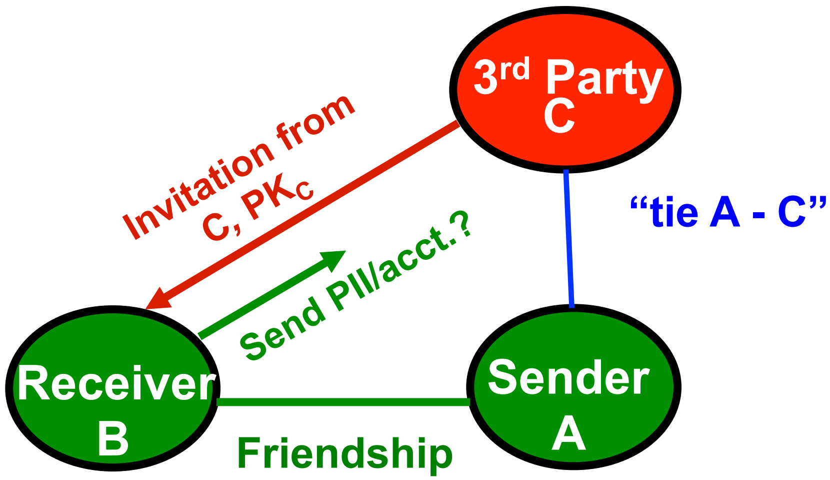
e.g., attributes:

- ***Identity***
- ***Certificates***
- ***Address/Location***
- ***Social Connections***
- ***Reputation/Credentials***

Carnegie Mellon

CyLab www.cylab.cmu.edu

Accepting an Attribute



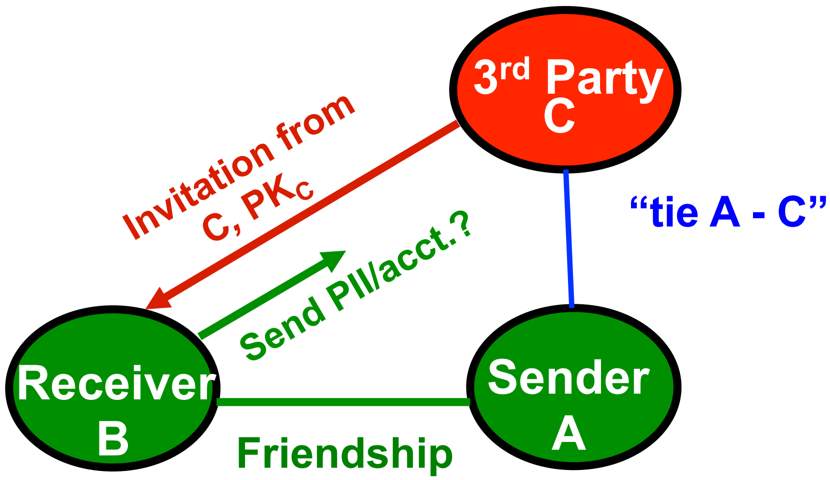
SC(A) @ B = social collateral of A at B

Friendship: a social relation

- *built-in* social collateral
- "street-level" punishment/sanction = *loss of future value*

Accepting an Attribute

Carnegie Mellon
CyLab
www.cylab.cmu.edu



SC(A) @ B

SC(C) @ A = assigned value by B to {tie A-C}^{SK_A} based on tie evidence and strength necessary for attribute auth; e.g., C, PK_C

Attribute Authenticity* => *evidence of tie to Sender

=> *strength of tie (social distance) to Sender*

(communication frequency, recency, reciprocity, length, common acquaintance)

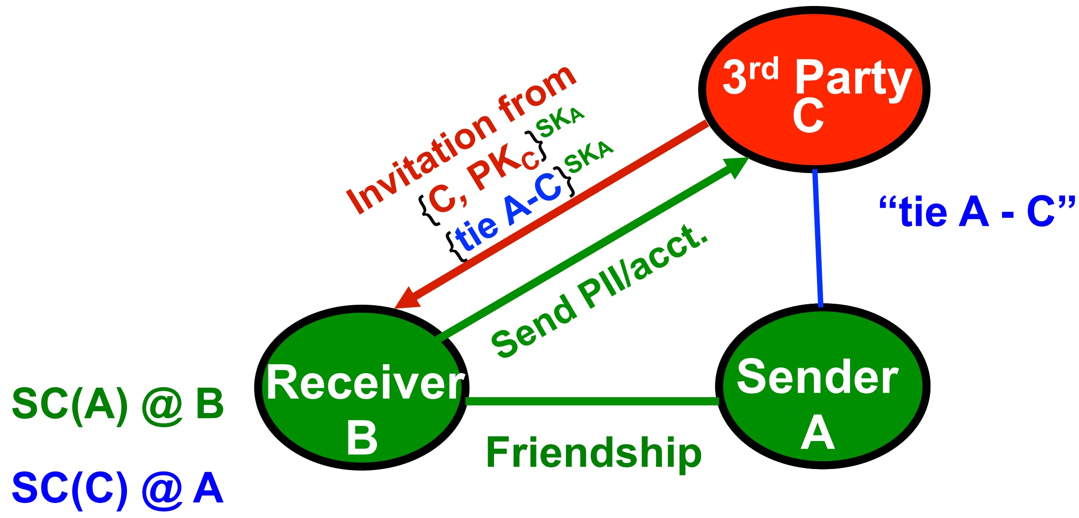
=>" street-level" punishment of 3rd party C (e.g., spoofed ID, false certificate)

=> loss of endorsement by Sender => loss of value at Receiver

Accepting an Attribute

Carnegie Mellon
CyLab
www.cylab.cmu.edu

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



Deterrence: SC(A) @ B – SC(C) @ A ≥ P,
where **P ≥ 0** measures **friend A's** ***net loss of collateral*** if $\{C, PK_C\}^{SK_A}$ is **false**

Acceptability: SC(C) @ A ≥ T_{Bapp},
where T_{Bapp} measures loss incurred by **B**'s application if $\{C, PK_C\}^{SK_A}$ is **false**

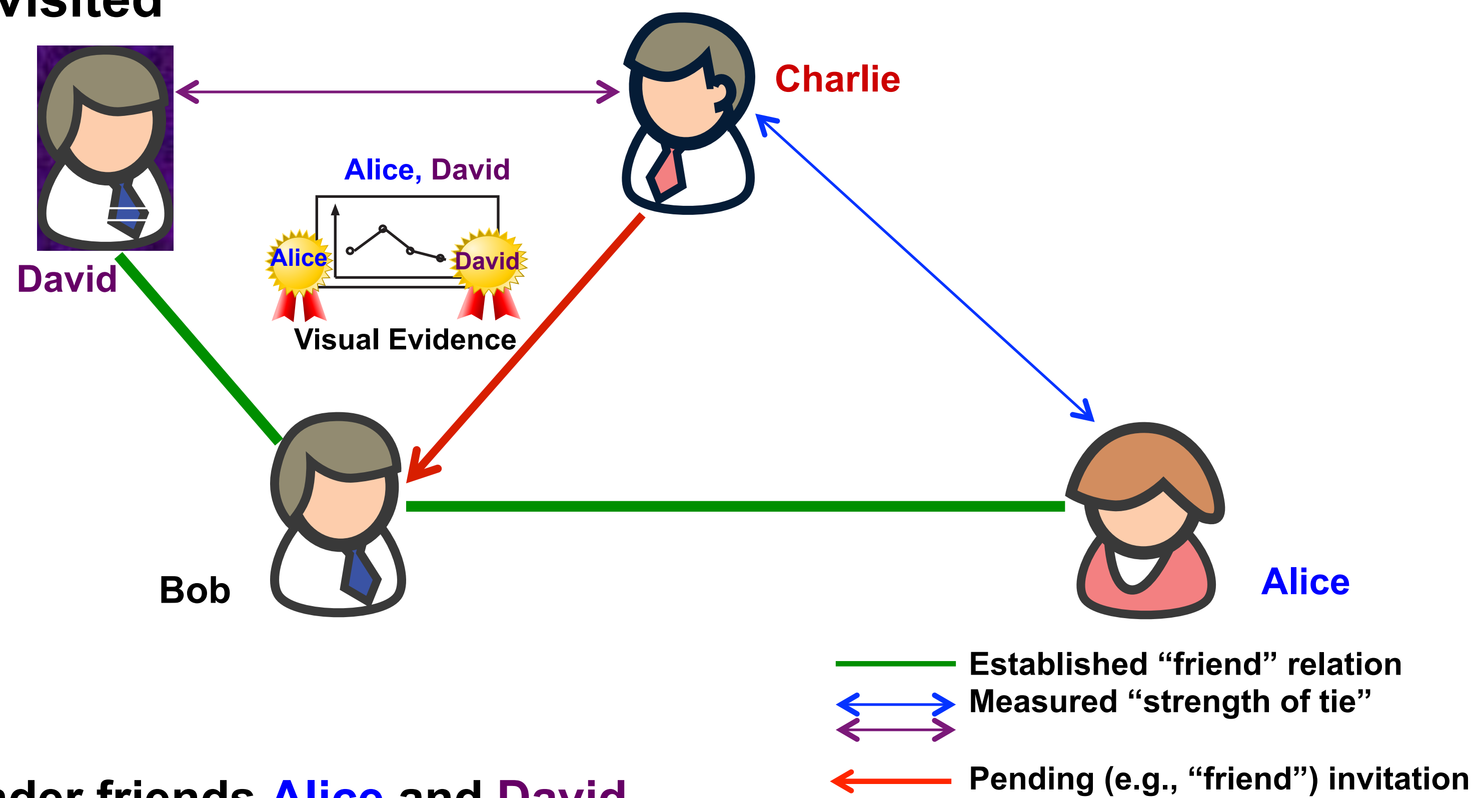
B accepts A's authentication of $\{C, PK_C\}^{SK_A}$

Visualization of "Tie Strength" Evidence

Visualizing Tie Strength

Carnegie Mellon
CyLab
www.cylab.cmu.edu

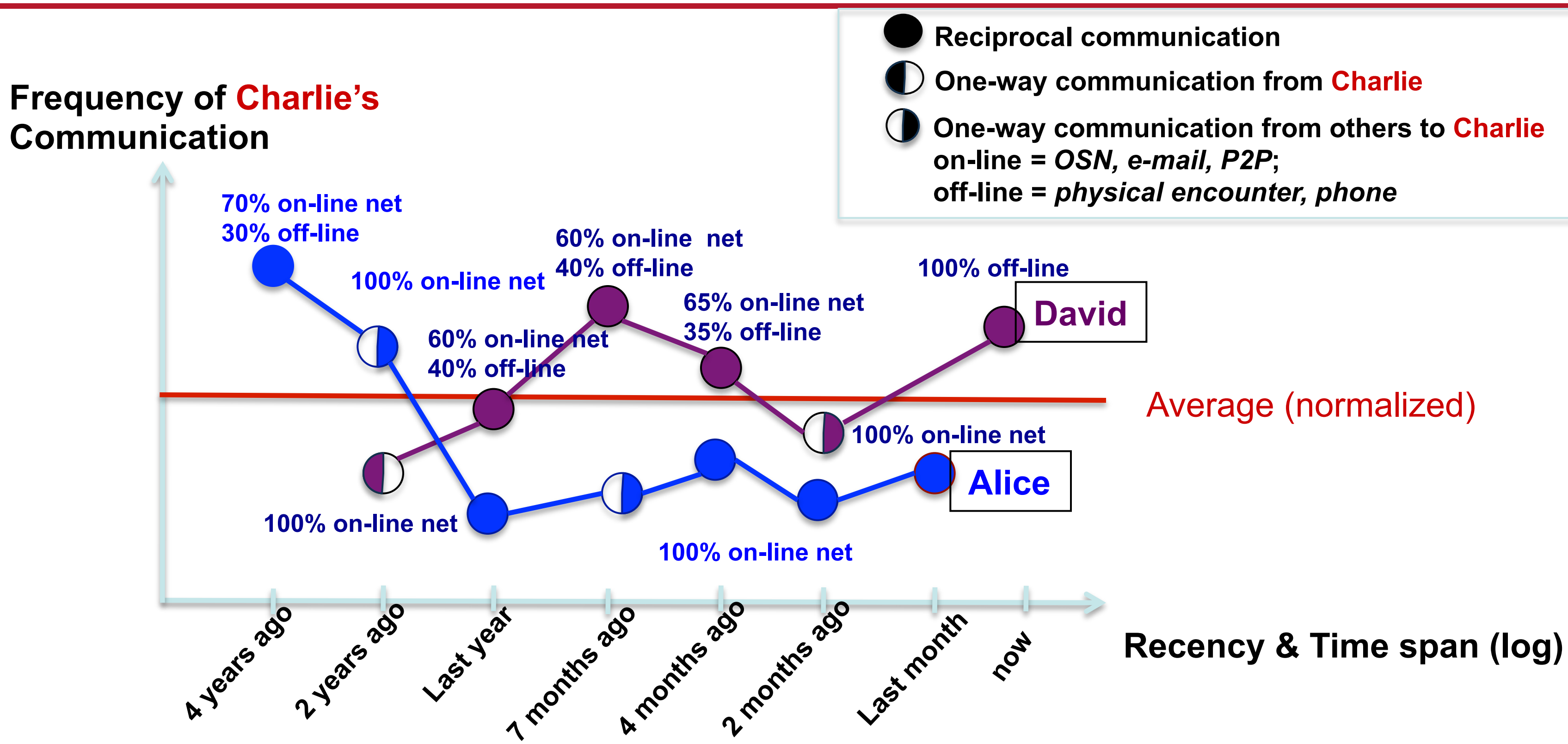
Example 2 Revisited



- 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?

Carnegie Mellon
CyLab
www.cylab.cmu.edu



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)

Usability: A Facebook Example

Carnegie Mellon
CyLab
www.cylab.cmu.edu

- **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**

Carnegie Mellon

CyLab
www.cylab.cmu.edu

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**