Net Trust: User-Centered Detection of Pharming, Phishing and Fraud
L Jean Camp
www.ljean.com
How to inform individual assessments of trustworthiness of a potential online transaction.
Design for Trust

• Start with human trust behaviors

• Trust
  – Used for simplification
  – Encompasses discrete technical problems
    • privacy, integrity, data security
  – Embeds discrete policy problems
    • business behavior, customer service, quality of goods, privacy
Human vs. Computer Trust

• Computers
  – Process data
  – Store data
  – Transmit data
  – Distinguish
    • atomicity, privacy, availability,

• Humans
  – Understand context
  – Evaluate uncertainty
  – Make lumping decisions based on context

• Begin with the human as the basis of the design
  – Examine human interactions
  – Signal humans using pre-existing social capital
Net Trust Goals

• Detect fraud
  – Notification
    • Warning the user
  – Prevention
    • Refuse to connect or require coping the url
  – Remediation
    • Connection to a remediation service
Trust and Context

Resource Verification
Resources are often fairly easy to identify as “good” or “bad” in physical realms
Trust and Context

Identity Verification
Traditional mechanisms to communicate trustworthiness.
Signaling Requires Malicious Party to Cooperate
Social Ratings Don’t Depend on Third Parties

Alice is friend’s with Bob, and trusts Bob’s opinion.

Bob likes this website.

Alice has never heard of this website, but will inform her trust decision with Bob’s experience with website.
Net Trust Reputations

• Do not require explicit user action, but are created by observations of user behaviors.
• Variables underlying the ratings are neither under control of attackers nor subject to highly parallel attacks.
• The ratings integrate personal, social network, and centralized information sources.
• The identity of the participants in a social network used for ratings are known only to members of that social network
Done & Working

- Ratings Engine
  - Implicit ratings (history-based)
  - Explicit ratings (manual interaction), comments
  - Local evaluation with age threshold adjustment
- Toolbar UI
  - Correct updates; coherent over tabs & windows
- Social Network
  - Manual email invitation and buddy ID entry
  - Self-enforcement of rating partition over personas
- Synchronization
  - Local ratings storage
  - Immediate server read/write on persona load/unload
Views
Security & Privacy Properties

- Sybil attack resistance
- Web scripting resistance
- Server authentication (anti-spoofing)
- Write authentication for peer records
- NT ID to email address commitment
- NT ID deniability ("That’s not my ID")
- Linking resistance (NT ID and personal info)
- Social network confidentiality
Short Term Objectives

• Synchronization (protecting social network)
  – Time delays for server access on persona change
  – Anonymous server access via Tor

• Third-Party rating assurance
  – Net Trust Certificate Authority
  – Signed rating lists

• Social Network
  – Mandatory history partition over multiple personas
  – Invite automation & validation
Longer Term Initiatives

• Expand rating sets for client-side pharming detection
  – Include server IP address & certs in history
• Blend rating sets across social networks
  – Deter unauthorized sharing of NT IDs
  – Improves ID deniability
  – Improves information diffusion
  – Enable server intersection attack on social network
• Narrative risk communication
  – Rich warnings: cartoons, video, animation