KHAP: Keyed Hard AI Problems for Securing Human Interfaces

Jeff King

André dos Santos

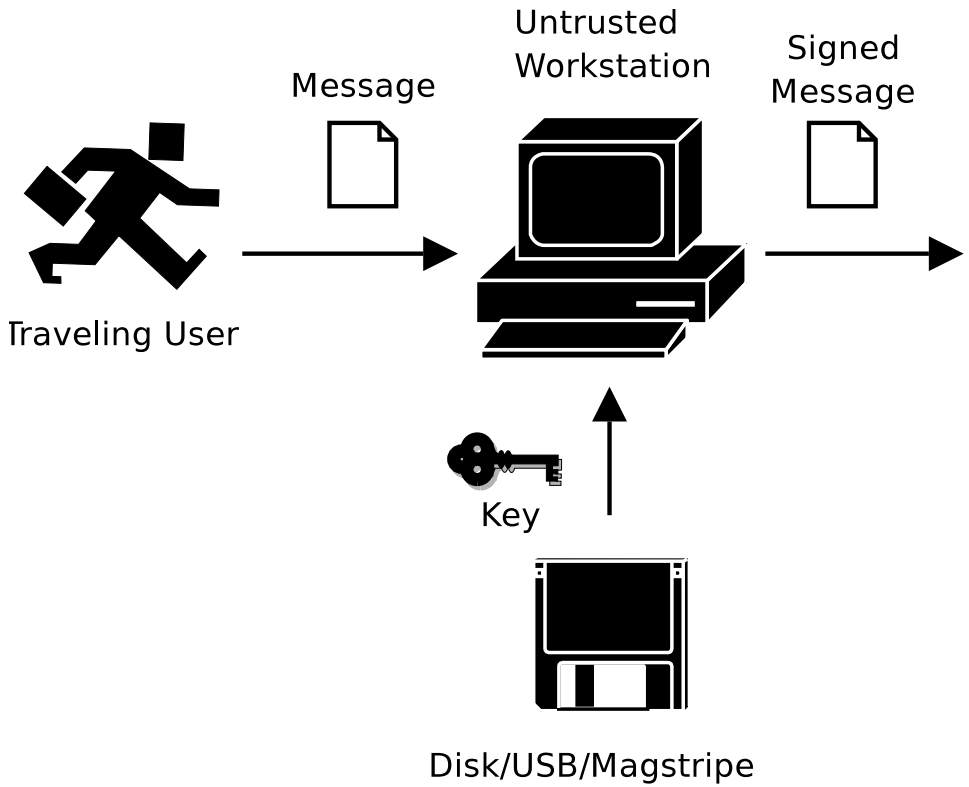
Chaoting Xuan

Georgia Institute of Technology

Outline

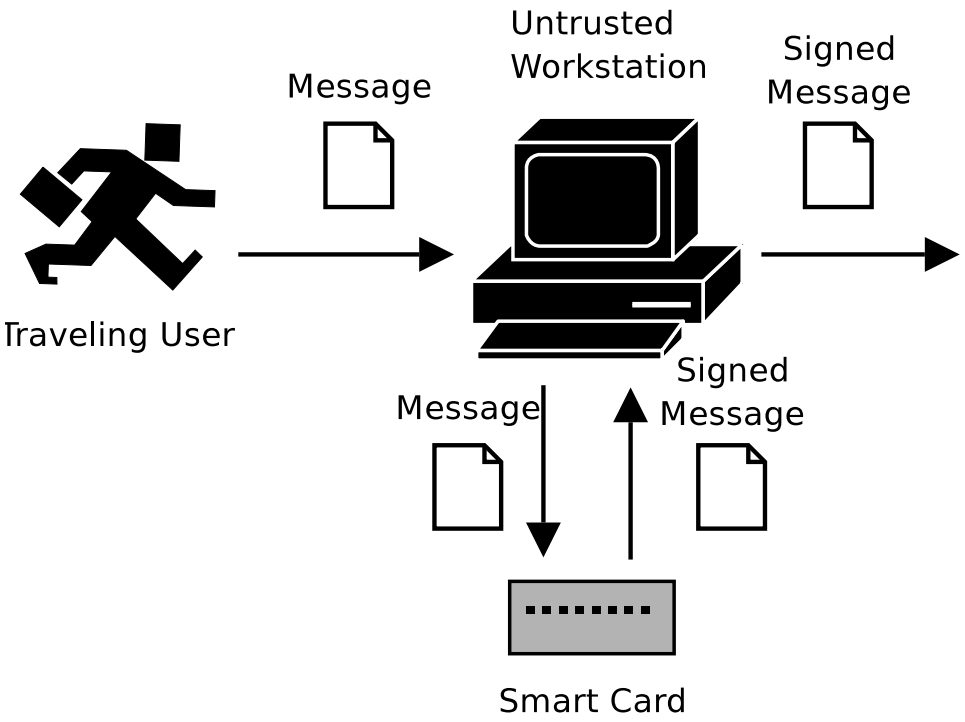
- Problem Definition
- Hard AI Problems in Security
- Keyed Hard AI Problems
- Example Problems
- Protocols
- Attacks
- Conclusions

Signing a Message



- Applications: electronic vote, point-of-sale, document signing
- Problem: Key Compromise

Signing a Message – Better



- Solved: No key compromise
- Problem: what message was signed?

Hard AI Problems

- Informally, something that humans can do easily but computers can't.
- More formally (von Ahn, 2003)
  - S – a set of problem instances
  - f – a function mapping instances to answers
  - For human H, $H(x) = f(x)$ with high probability
  - Security parameters – (δ, τ)-hard
  - For any algorithm A running in time τ, $Pr[A(x) = f(x)] \leq \delta$
- CAPTCHA – Completely Automated Turing Test to Tell Computers and Humans Apart
- Generate random message, transform it, ask human to repeat it

Transformation Problems

- Subset of hard AI problems that transform a message
- m – message
- $T(m)$ – problem instance from message
- $f(T(m)) = m$ – solving the problem returns original message
- Example: distort text of message so that only humans can read it
- Previous CAPTCHA is a transformation problem

KHAP: Keyed Hard AI Problems

- A transformation problem that includes a shared secret key
- Instances generated with different keys are **distinguishable**
- Computers can't steal keys from messages
- Formalisms (this paper):
  - $H_d(m, m')$ – human distinguishes between messages with different keys
  - $|k - k'|$ – difference between two keys (quantifiable?)
  - Security parameters – $(\alpha, \varepsilon, \tau)$-hard
  - Given $|k - k'| > \varepsilon$, $Pr[H_d(m, m')] \geq \alpha$
  - For any algorithm A, $Pr[H_d(m, A(m', m))] \geq \alpha$
- Problem: How to quantify keys and distinguishability?
- Solution: Empirical testing?

Example Keyed Problems

- Speech Synthesis
  - Transform text message into audible speech
  - Human retrieves message by listening
  - Key is vocal synthesis parameters (randomly select at key generation)
  - Human recognizes specific voice
  - Random noise on audio track makes automated analysis difficult
- Handwriting
  - Transform text message into image of handwriting
  - Human reads message
  - Key is characteristics of handwriting
  - Human recognizes handwriting
  - Is α (probability of distinguishing) high enough?

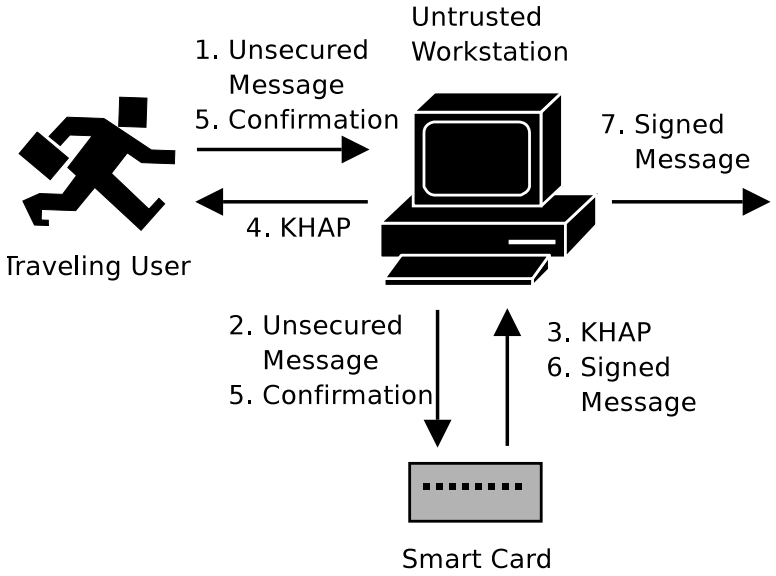
3-D Keyed Transformation

- Render text and objects in a 3-D scene to 2-D image (raytrace)
- Randomize parameters (lighting, position, rotation, size, colors)
- Human can read text from 2-D image
- Key is appearance of objects
- Human looks for particular objects in scene
- Scene is hard to modify in a meaningful way (shadows, reflections, finding objects)
- Provide authenticity (presence of keys) and integrity (modifications can be detected by human)

Example 3-D Instance

Protocols

- Original problem: securely interacting with a trusted platform through untrusted terminal
- Message from trusted platform to human: encode in KHAP
- Keyed transformation makes it hard to forge or tamper with messages
- Message from human to trusted platform: harder!



What to do in an Emergency

- Only human can detect cheating by terminal. What to do?
- Complain out-of-band
  - Point-of-sale tries to overcharge
  - Complain to customer service and remove charge
- Disconnect trusted platform
  - Platform waits N seconds before signing
  - User disconnects device if cheating is detected; otherwise, wait.
- Confirmation word
  - Platform waits for one-time secret message, human types message
  - Human carries pre-arranged message list?
  - Embed confirmation word in KHAP message

Attacks

- Forge KHAP message
  - Must guess key if no messages have been seen
  - Cannot extract key from message (definition of KHAP)
- Modify KHAP message
  - Hard because of AI problem domain
- Replay old messages
  - Solution: don’t send duplicate messages (bad user-friendliness)
  - Solution: change key periodically (bad user-friendliness?)
  - Solution: connect confirmation to key (confirmation word in KHAP)
- Implicit confirmation (human must disconnect to cancel)
  - Solutions: reverse (not intuitive) or connect only when used (annoying)

Conclusions

- Approach is general (mobile device, network, etc)
- Secure
  - Security depends on AI problem parameters
  - Advances in AI break problems (as factoring breaks RSA)
- Easy to use
  - Avoid computation, memory aids: ask humans to do what they do best
  - Some problems are intuitive (e.g., recognizing voice)
- Issues/Future Work
  - Human attackers
  - Performance on mobile devices
  - User-friendly key changing
  - Find and analyze more schemes!

Questions?