

Human Errors in Security Protocols

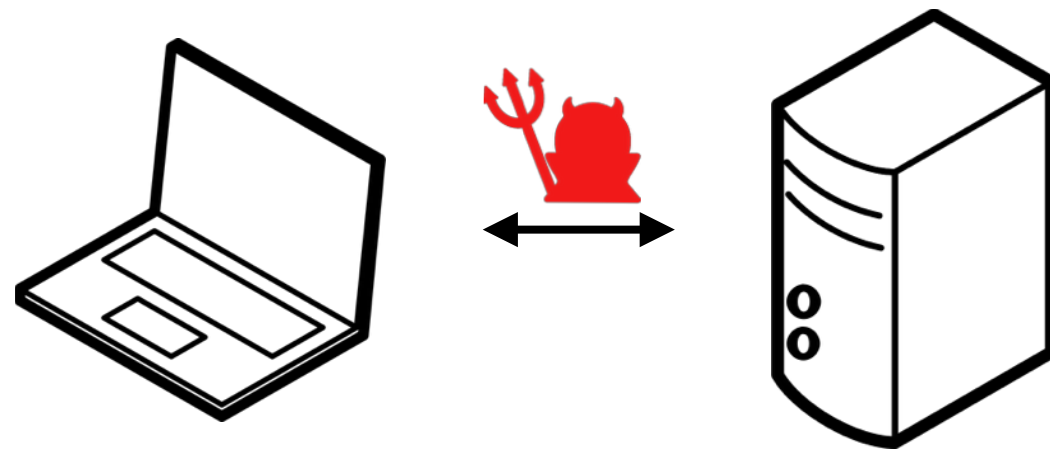
David Basin

joint work with Saša Radomirović and Lara Schmid

Institute of Information Security

ETH zürich

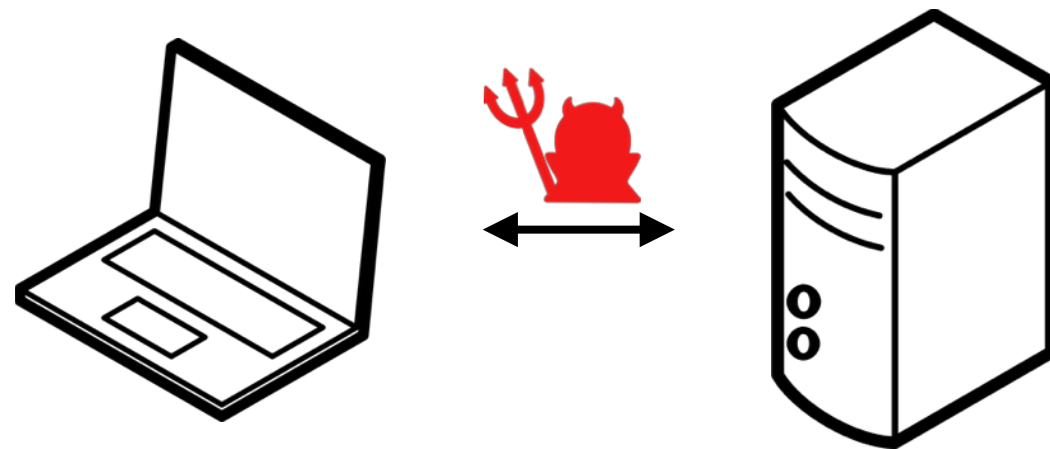
Recap Security Protocols



We have defined

- the **Dolev-Yao** adversary
- **communicating agents** that follow a **role specification**
- and the **security properties** desired

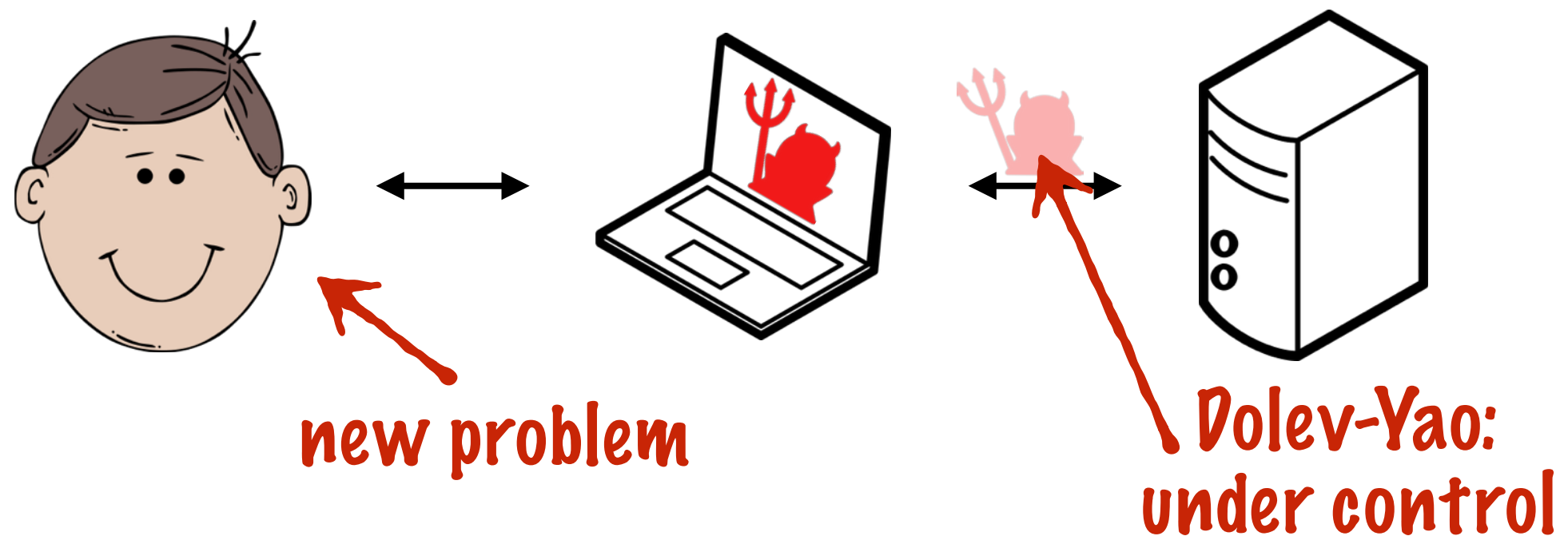
What we have not seen



We glossed over that

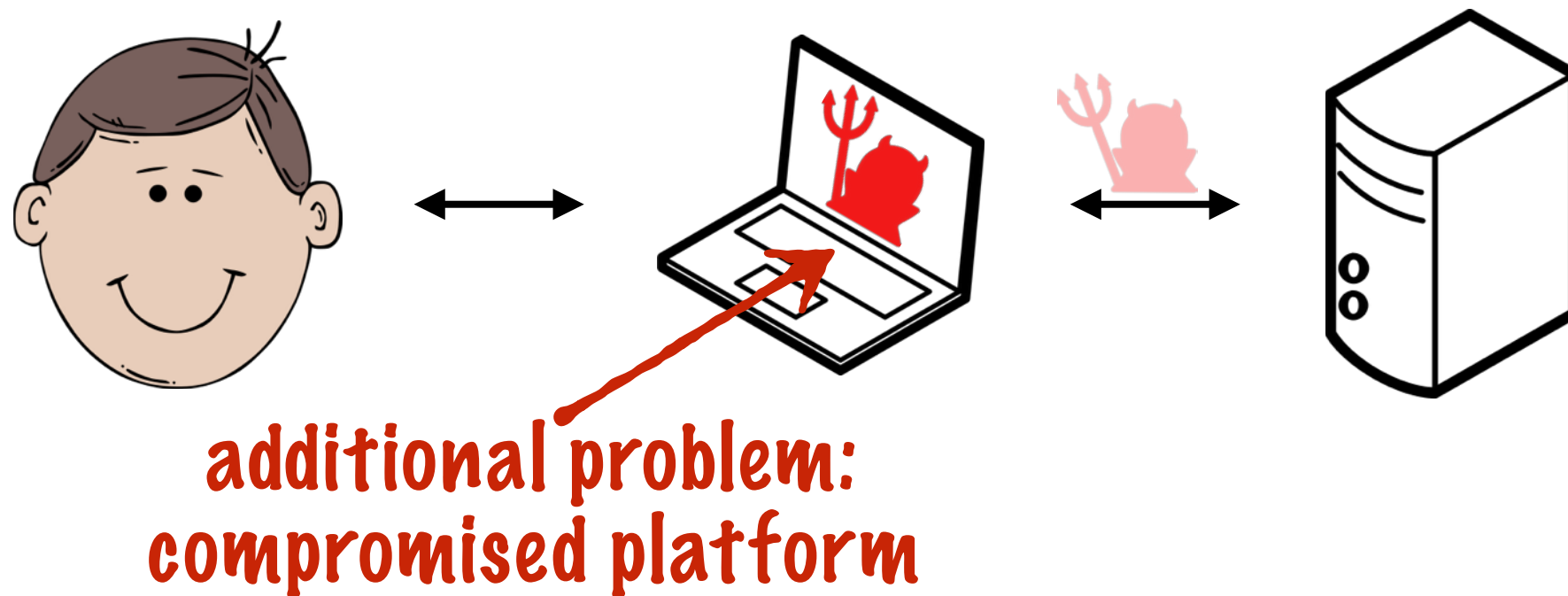
- a **human** is one of the communicating parties,
- humans have **limited** computational abilities, and
- they are **error-prone**.

How can we achieve secure communication between a **human** and a remote server?



- Examples: Online Banking, Internet Voting, Electronic Tax Returns, ...
- How do we model and reason about interaction between humans and computers?

How can we achieve secure communication between a **human** and a remote server?



- If platform is compromised: no useful secure communication is possible.
- A trusted device is necessary.



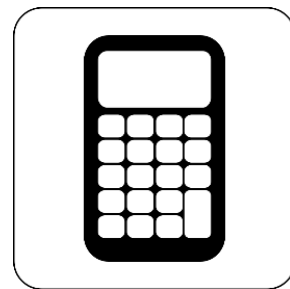
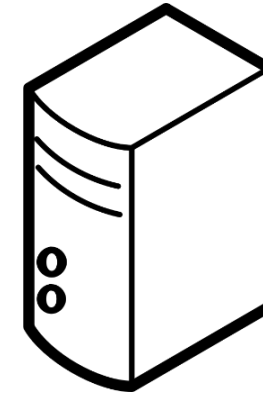
Human H



Platform P



Server S



Device D

**For which kinds of devices is
secure communication possible?**

(A Complete Characterization of Secure Human-Server
Communication, CSF 2015)

Focus in this talk on human errors

(Modeling Human Errors in Security Protocols, CSF 2016)

Possible “devices”:

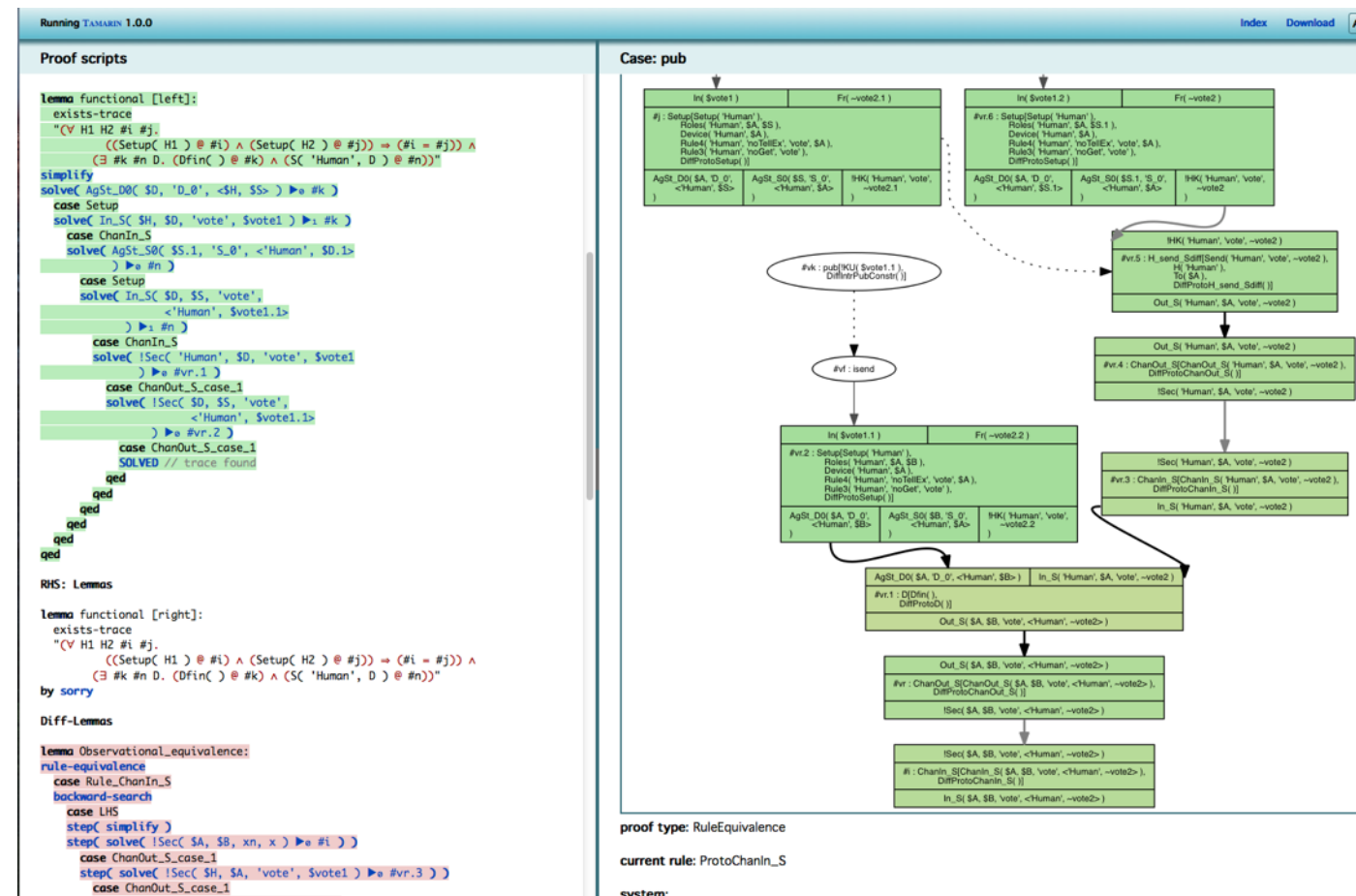


Overview

- 1. Security protocol model**
2. Modelling Human Error
3. Applications

Security Protocol Model — Tamarin

- Symbolic formal model specified using **multiset rewriting**
- **Dolev-Yao adversary** controlling communication network.
- Possible executions modeled by **traces**
- **Tool support**



Protocol Specification Example

Alice & Bob specification

$A: \text{fresh}(n)$
 $A \rightarrow B: n$

Role specification of A

Protocol rules

$[\text{Fr}(n)] \longrightarrow [\text{AgSt}(n)]$

$[\text{AgSt}(n)] \xrightarrow{S(n)} [\text{Out}(n)]$

$[\text{In}(n)] \xrightarrow{R(n)} []$

Adversary rules (simplified)

$[\text{Out}(n)] \xrightarrow{K(n)} [!K(n)]$

$[!K(n)] \longrightarrow [\text{In}(n)]$

$[!K(n), !K(m)] \longrightarrow [!K(\text{pair}(n,m))]$

$[] \longrightarrow [!K(\$x)] \quad (\$x: \text{public term})$

...

Fresh rule

$[] \longrightarrow [\text{Fr}(n)]$

Protocol Execution Example

State	Term Rewriting Rule	Instantiation	Trace
[]	$[] \longrightarrow [\text{Fr}(n)]$	Specified rules: $[] \longrightarrow [\text{Fr}(n)]$ $[\text{Fr}(n)] \longrightarrow [\text{AgSt}(n)]$ $[\text{AgSt}(n)] \longrightarrow [\text{Out}(n)]$ $[\text{Out}(n)] \longrightarrow [!K(n)]$ $[!K(n)] \longrightarrow [\text{In}(n)]$...	
[Fr(~1)]	$[\text{Fr}(n)] \longrightarrow [\text{AgSt}(n)]$		
[AgSt(~1)]	$[\text{AgSt}(n)] \xrightarrow{S(n)} [\text{Out}(n)]$		S(~1)
[Out(~1)]	$[\text{Out}(n)] \xrightarrow{K(n)} [!K(n)]$		
[!K(~1)]	$[!K(n)] \longrightarrow [\text{In}(n)]$	$[!K(\sim 1)] \longrightarrow [\text{In}(\sim 1)]$	K(~1)
[!K(~1) , In(~1)]	$[\text{In}(n)] \xrightarrow{R(n)} []$	$[\text{In}(n)] \xrightarrow{R(\sim 1)} []$	R(~1)

Communication Channels

Authentic $\bullet \rightarrow \circ$, confidential $\circ \rightarrow \bullet$, and secure $\bullet \rightarrow \bullet$ channel rules are used to restrict capabilities of Dolev-Yao adversary.

Example: Confidential channel rules

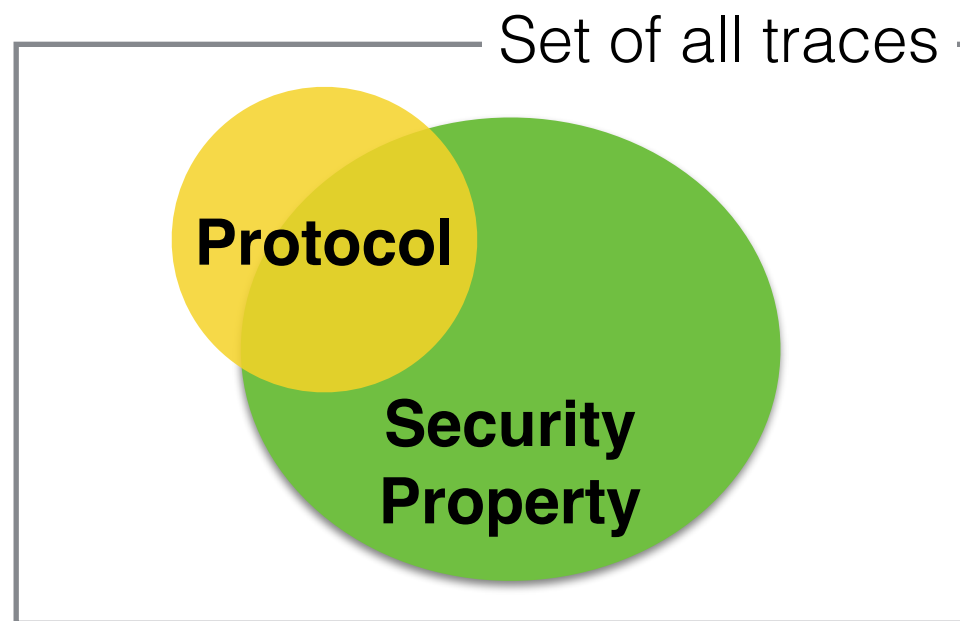
$[\text{SndC}(\$A, \$B, m)] \longrightarrow [!\text{Conf}(\$B, m)]$

**\$ sign: public term.
Agent names are
public knowledge.**

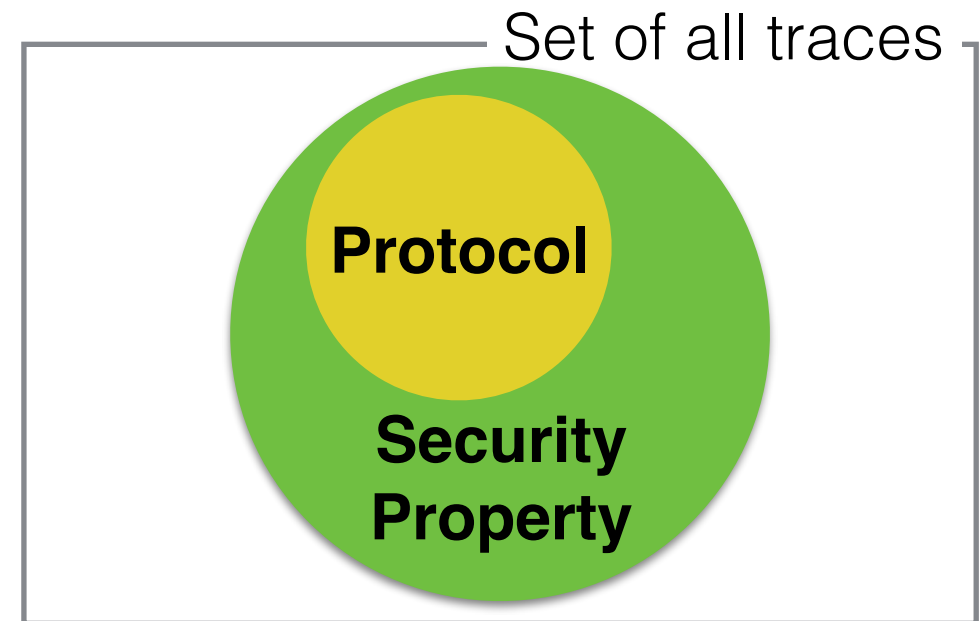
$[!\text{Conf}(\$B, m), !K(\$A)] \longrightarrow [\text{RcvC}(\$A, \$B, m)]$

$[!K(<\$A, \$B, m>)] \longrightarrow [\text{RcvC}(\$A, \$B, m)]$

Security Properties



Protocol does not satisfy security property.



Protocol satisfies security property.

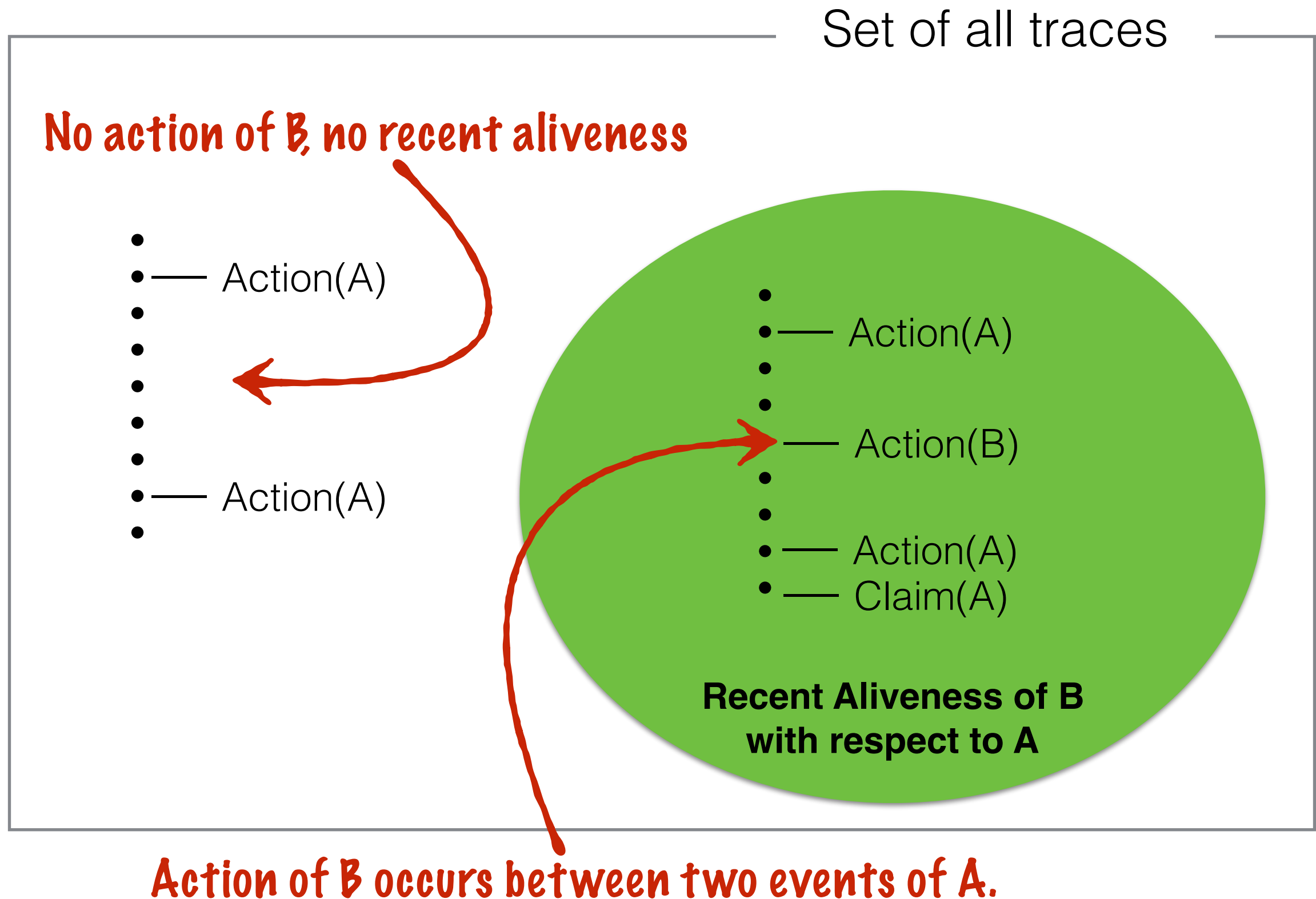
Confidentiality

If m is claimed to be secret, then the adversary does not learn m .

Set of traces:

$$\forall m \#i \#j. \text{Secret}(m)@i \Rightarrow \text{not } K(m)@j$$

Authentication Properties: Recent Aliveness



Authentication Properties

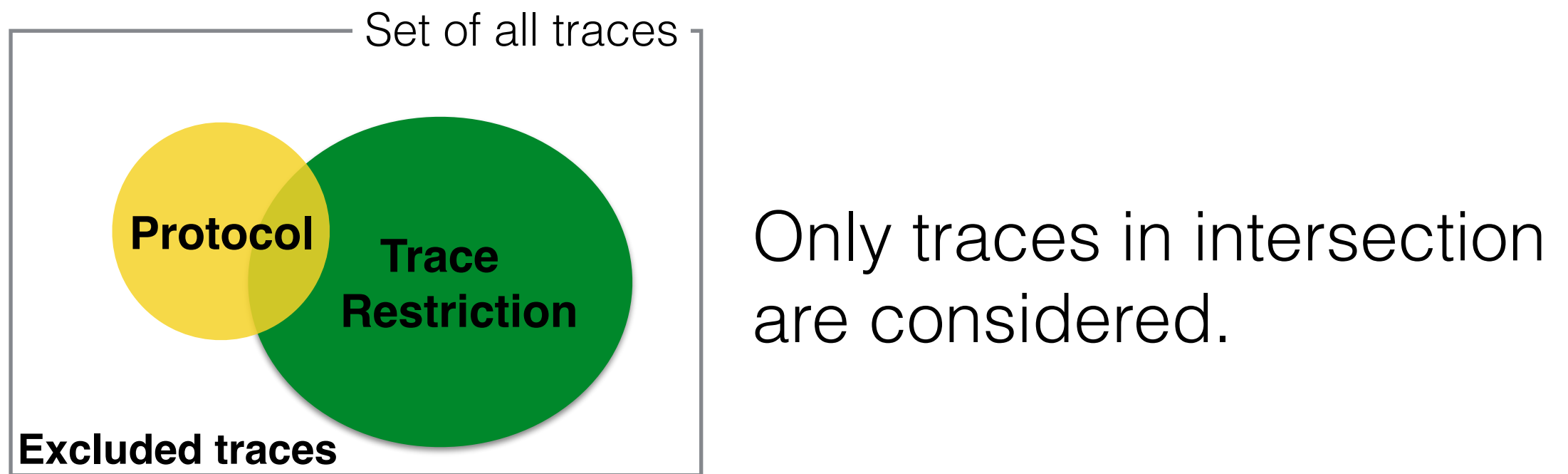
Entity Authentication: Recent aliveness of an entity H , with respect to verifier (remote server S).

Device Authentication: Recent aliveness of a device D . We generally assume exclusive access of human H to D .

Message Authentication: If verifier claims that H has sent m , then H has indeed sent m .

Trace Restrictions

Exclude traces that violate the specification.

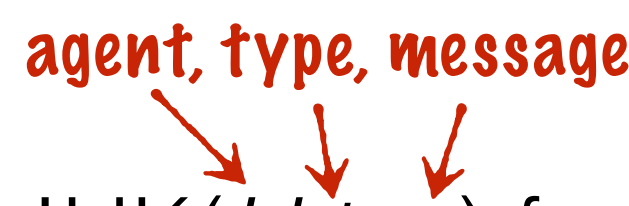


Example: A trusted agent was not previously dishonest.

Set of traces:

$$\forall A \#i \#j. (\text{Trusted}(A)@i \wedge \text{Dishonest}(A)@j) \Rightarrow i < j$$

Modelling Humans

- Humans can **communicate** over provided interfaces.
- Human **knowledge** is modelled with $!HK(H, t, m)$ facts.
E.g.: $!HK(H, 'pw', p)$ means human H knows password p .
The diagram shows the text "agent, type, message" in red above the formula $!HK(H, t, m)$. Three red arrows point from "agent" to H , from "type" to t , and from "message" to m .
- Humans can **concatenate** and **split** messages:
$$[!HK(H, t_1, m_1), !HK(H, t_2, m_2)] \longrightarrow [!HK(H, \langle t_1, t_2 \rangle, \langle m_1, m_2 \rangle)]$$
$$[!HK(H, \langle t_1, t_2 \rangle, \langle m_1, m_2 \rangle)] \longrightarrow [!HK(H, t_1, m_1), !HK(H, t_2, m_2)]$$

(simplified rules)

Overview

1. Security protocol model
- 2. Modelling Human Error**
3. Applications

Modelling Human Error

- Users **don't know protocol** specifications
- **Mistakes are made**, even experts slip up
- We are susceptible to **social engineering**
- So how should we analyze security of systems in view of human errors?



Definition

A **human error** in a protocol execution is **any deviation** of a human from his or her **role specification**.

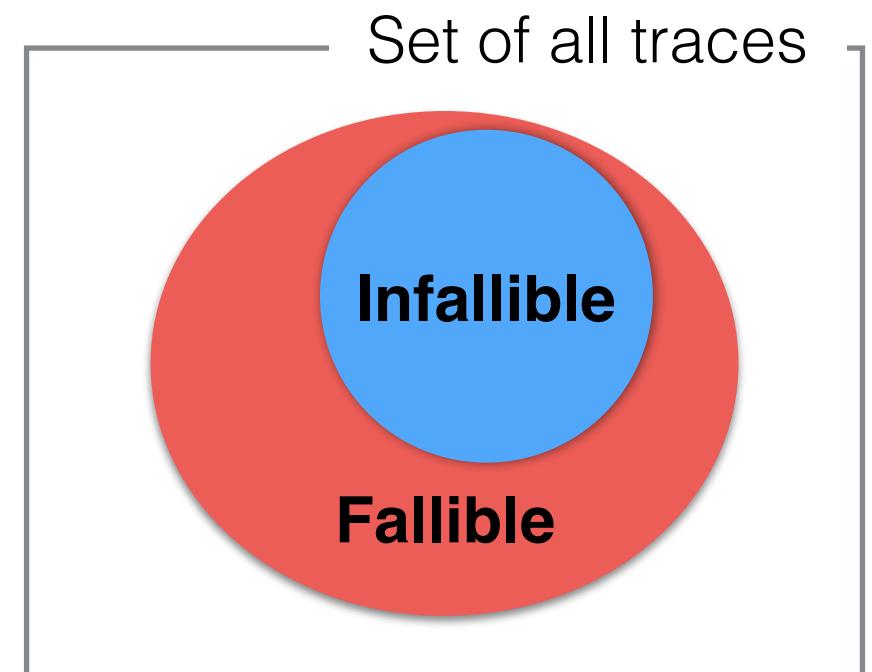
Two Classes of Human Error

- Distinguish between slips and lapses by **skilled users** and mistakes by **inexperienced users**.
- Model slips and lapses: Allow an **infallible agent** to make a small number of mistakes.
- Model rule-based behaviour: Allow for arbitrary behaviour of an **untrained agent** up to a few simple rules (guidelines).



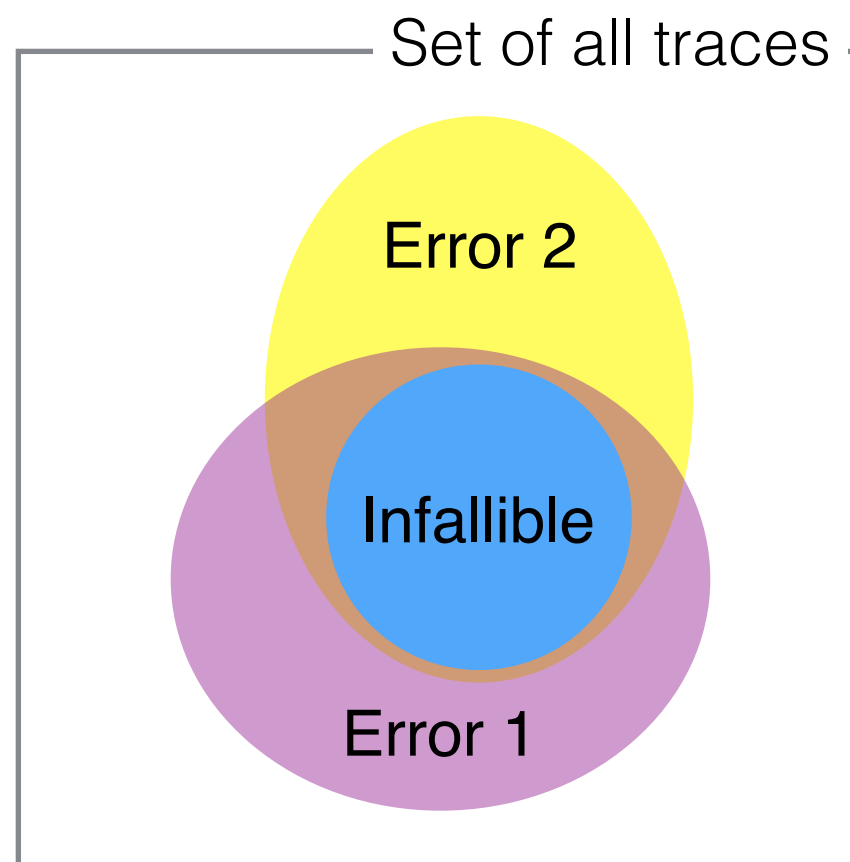
Infallible vs Fallible Humans

- **Infallible** human follows protocol specification.
- **Fallible** human *may* deviate from protocol specification.
- Fallible humans give rise to **more system behaviours** than the infallible human.

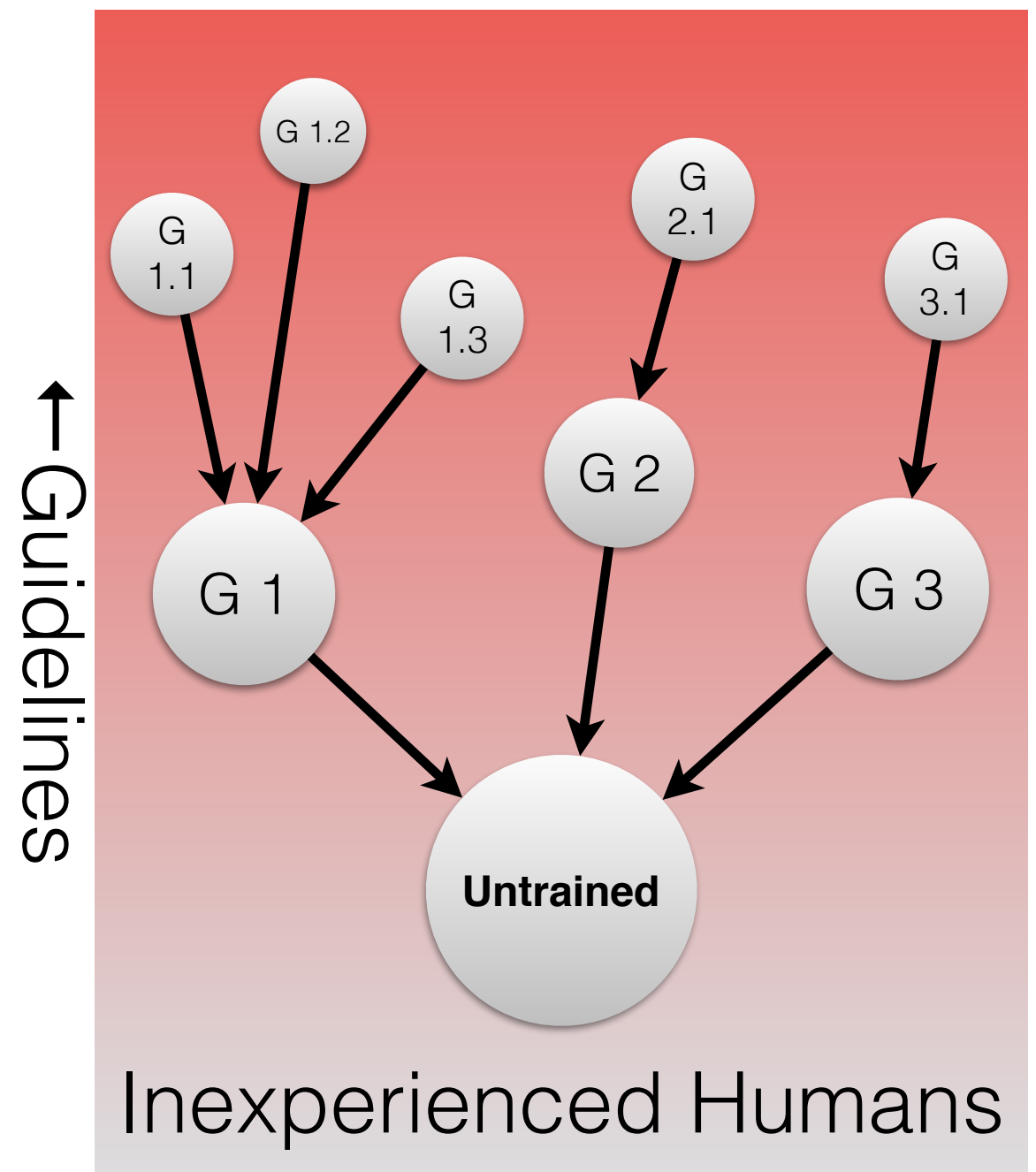
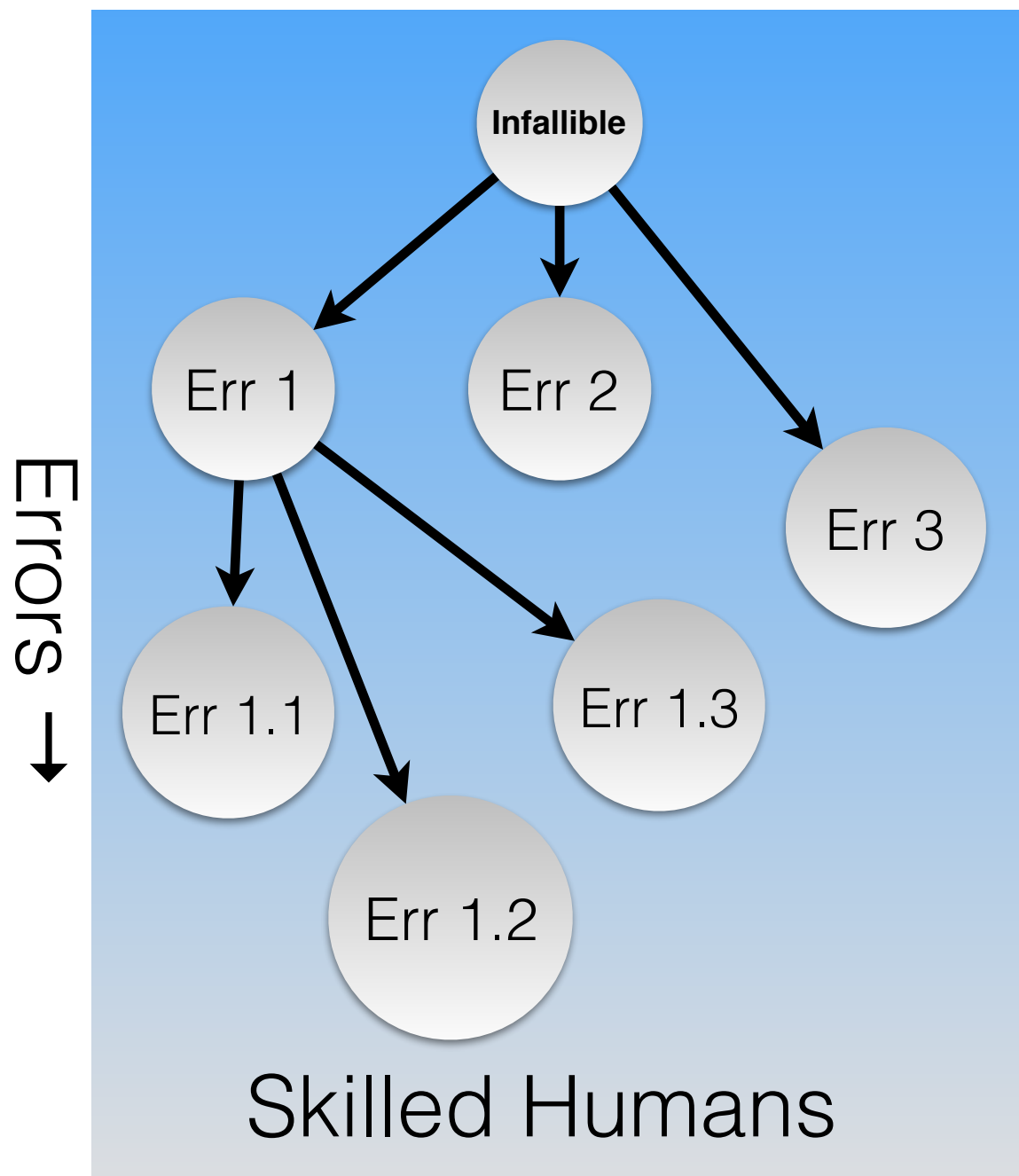


Comparing Specific Errors

Partial order of human errors by comparing sets of induced traces.

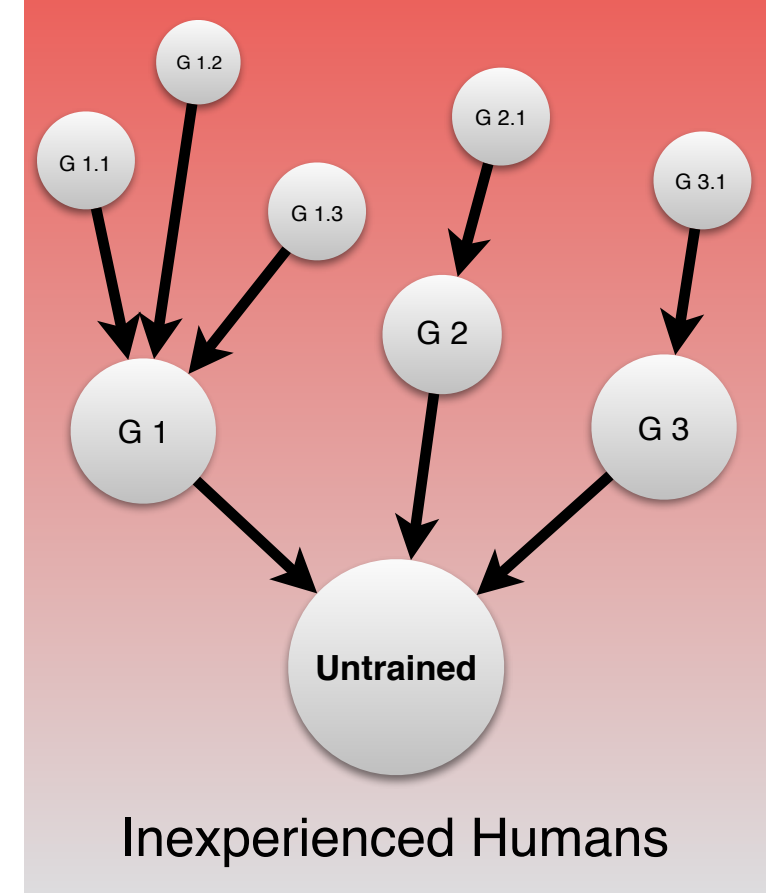


Two Classes of Human Error



Arrows indicate trace-set containment
(node at arrowhead contains more behaviors than node at tail)

Untrained Humans



We focus on this class

- They are **ignorant of** and **may deviate arbitrarily** from protocol specification.
- They accept any message received and send any message requested.

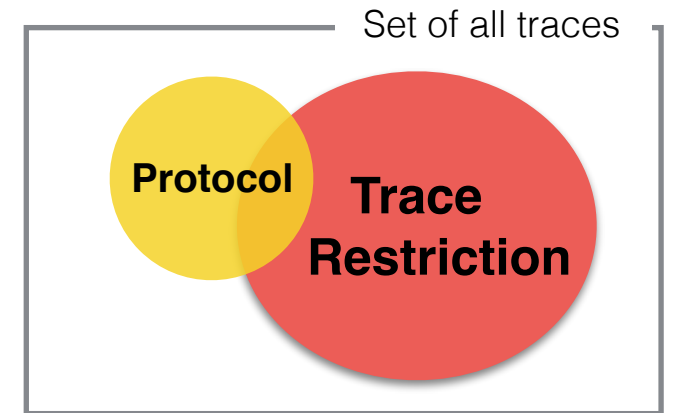
$$[\text{In}(\langle tag, msg \rangle)] \longrightarrow [!HK(H, tag, msg)]$$

$$[!HK(H, tag, msg)] \longrightarrow [\text{Out}(\langle tag, msg \rangle)]$$

(Trace labels omitted.)

- But they can be trained, given guidelines!

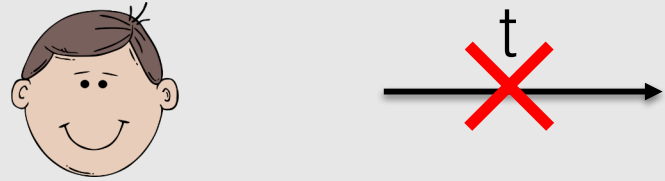
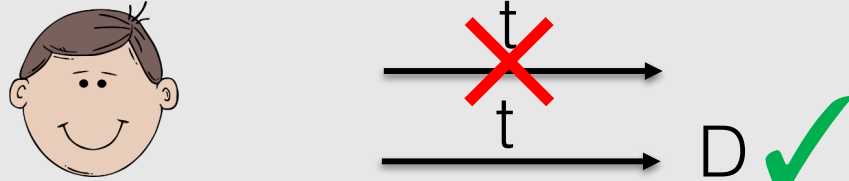
Guidelines



Guidelines are modelled by trace restrictions.



Exemplary Guidelines I

NoTell(H, t)	
NoTellExcept(H, t, D)	

- **NoTell**(H, tag):

$\forall m \#i \#j. \text{NoTell}(H, tag)@i \Rightarrow \text{not Snd}(H, \langle tag, m \rangle)@j$

Human H does not send information of type tag to anyone.

E.g.: Never reveal your private key.

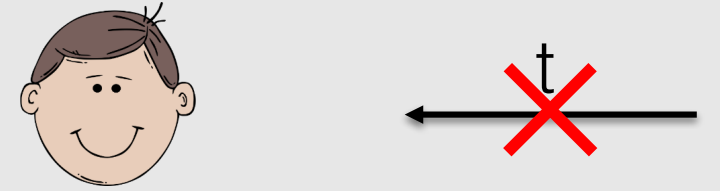
- **NoTellExcept**(H, tag, D):

Human H does not send information of type tag to anyone except D .

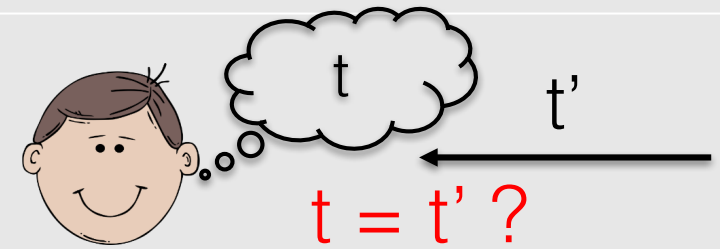
E.g.: Only enter your password into your own device.

Exemplary Guidelines II

NoGet(H, t)



ICompare(H, t)



- **NoGet**(H, tag): Human H rejects information of type tag from everyone.
E.g.: Never click on links in emails.
- **ICompare**(H, tag): Human H always compares received information of type tag with information in his initial knowledge.
E.g.: Always check the website's URL.

Concrete example — ebanking

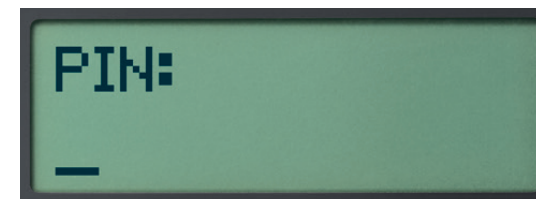
Login with the Access Card and card reader

Access the desired online service via ubs.com/online and initiate the login process (self-authorization).

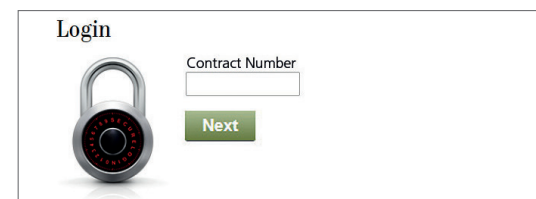
1. Activate the card reader by inserting the Access Card.



2. Enter your PIN and press **OK**.

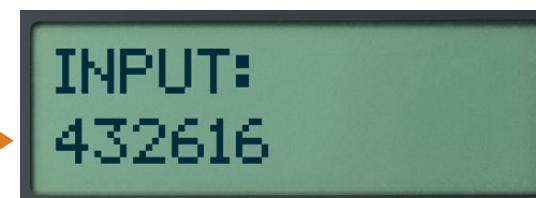
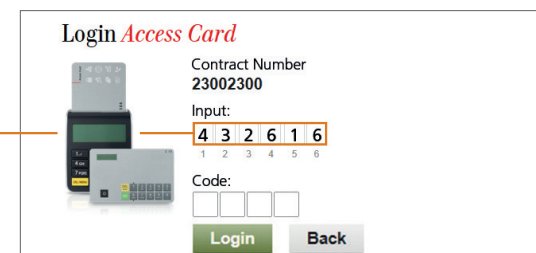


3. Enter your contract number on the login page and click **Next**.

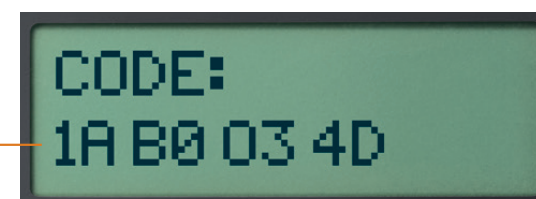


4. Enter the six-digit code displayed on the login page into the card reader and press **OK**.

Security note: The login number displayed by UBS **always has six digits**. If it has fewer digits, this could be a case of attempted fraud. Contact the support team as soon as possible in this case.



5. Enter the eight-digit code from the card reader on the login page and click **Login**.

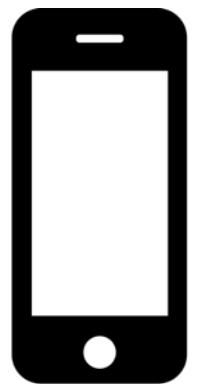


Overview

1. Security protocol model
2. Modelling Human Error
- 3. Applications**

Phone-based Authentication

- **Cronto**: Scan a code on platform, decrypted by mobile device, enter code + password on platform
- **Google 2-step**: login/password + SMS
- **MP-Auth**: Enter password into mobile device
- **One-time passwords over SMS**: single-factor authentication
- **Phoolproof**: choose server on device, device-server communication, then enter password on the platform
- **Sound-Proof**: ambient noise recorded by platform and mobile



MP-Auth* without Human Role

D knows: H , $pk(S)$, pw

S knows: H , $sk(S)$, pw

S: $fresh(r_S)$

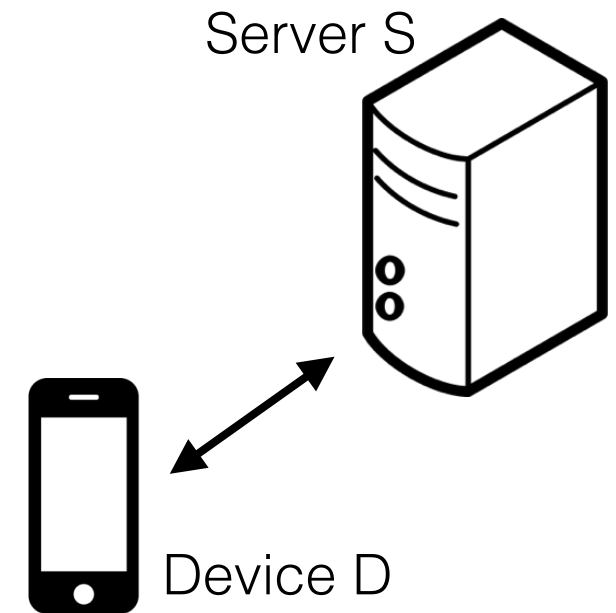
S \rightarrow D: S, r_S *fresh session, no replay*

D: $fresh(r_D)$ *only S can read this*

D \rightarrow S: $\{r_D\}_{pk(S)}$, $\{h(r_S), H, pw\}_{h(r_S, r_D)}$ *shared key: only D and S can compute this*

S \rightarrow D: $\{h(r_D)\}_{h(r_S, r_D)}$ *S must have sent this*

D must have sent this



satisfies confidentiality & authenticity of $h(r_S, r_D)$

- (*) Mohammad Mannan and Paul C. van Oorschot. Leveraging personal devices for stronger password authentication from untrusted computers. *Journal of Computer Security*, 19(4):703–750, 2011.

MP-Auth without Human Role

D knows: H , $pk(S)$, pw
S knows: H , $sk(S)$, pw

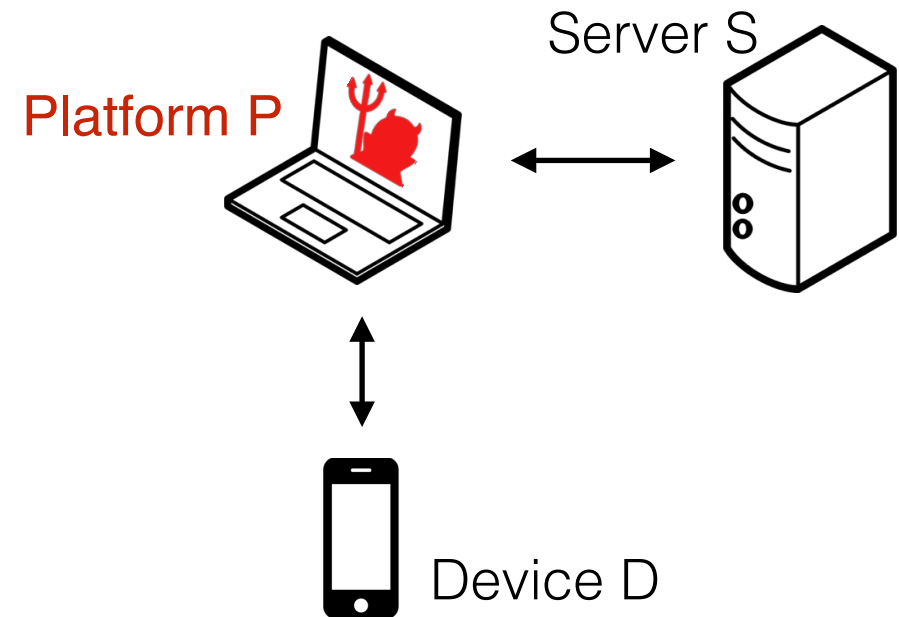
S: $fresh(r_S)$

$S \rightarrow P \rightarrow D$: S, r_S

D: $fresh(r_D)$

$D \rightarrow P \rightarrow S$: $\{r_D\}_{pk(S)}, \{h(r_S), H, pw\}_{h(r_S, r_D)}$

$S \rightarrow P \rightarrow D$: $\{h(r_D)\}_{h(r_S, r_D)}$



D: trusted device, **P**: untrusted platform 

MP-Auth

H knows: D, P, S, pw

D knows: H, pk(S)

S knows: H, sk(S), pw

H \rightarrow P: S

P \rightarrow S: 'start'

S \rightarrow P \rightarrow D: fresh(r_S) . S, r_S

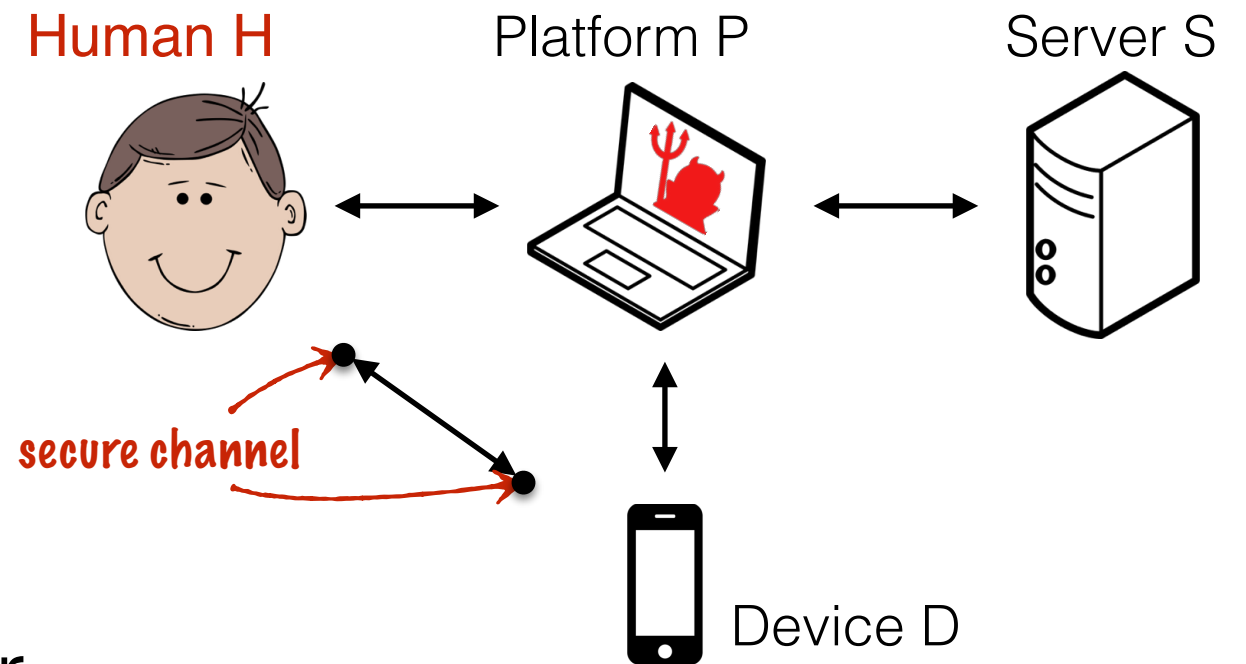
D $\bullet \rightarrow \bullet$ H: S

H $\bullet \rightarrow \bullet$ D: pw, H

D \rightarrow P \rightarrow S: fresh(r_D) . $\{r_D\}_{pk(S)}$, $\{h(r_S), H, pw\}_{h(r_S, r_D)}$

S \rightarrow P \rightarrow D: $\{h(r_D)\}_{h(r_S, r_D)}$

D $\bullet \rightarrow \bullet$ H: 'success'



MP-Auth

H knows: D, P, S, pw

D knows: H, pk(S)

S knows: H, sk(S), pw

H \rightarrow S: 'start'

S \rightarrow D: fresh(r_S) . S, r_S

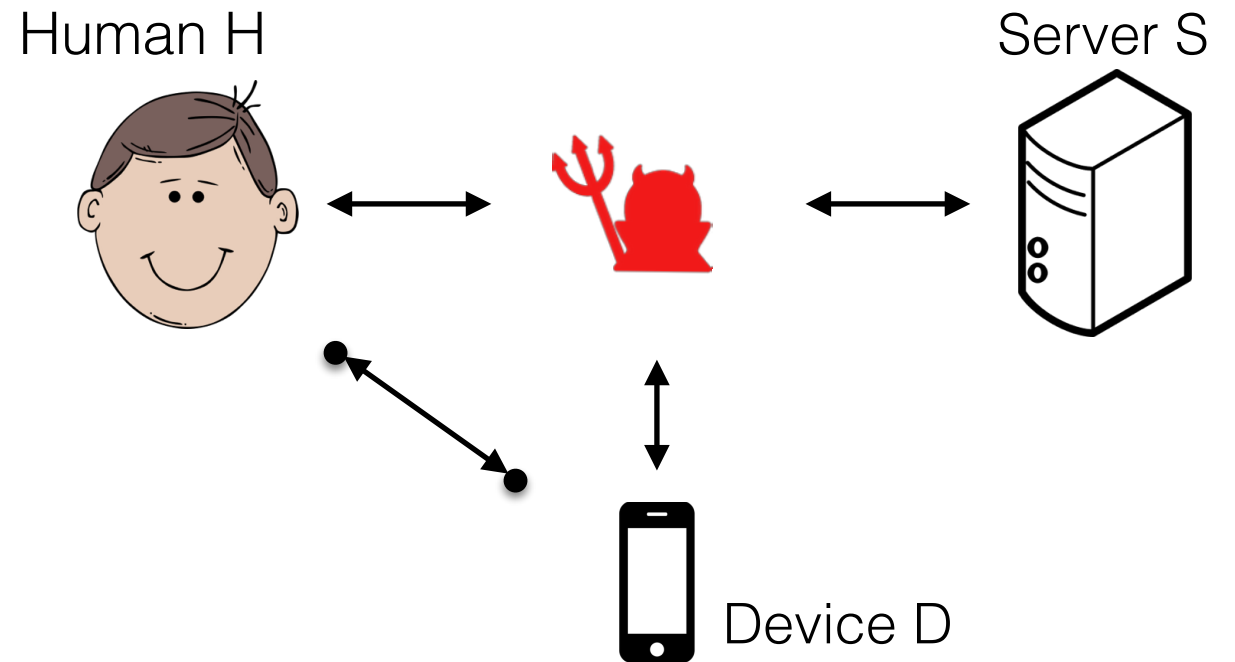
D $\bullet \rightarrow \bullet$ H: S

H $\bullet \rightarrow \bullet$ D: pw, H

D \rightarrow S: fresh(r_D) . $\{r_D\}_{pk(S)}$, $\{h(r_S), H, pw\}_{h(r_S, r_D)}$

S \rightarrow D: $\{h(r_D)\}_{h(r_S, r_D)}$

D $\bullet \rightarrow \bullet$ H: 'success'



Modelling untrusted platform with insecure channels.

Comparison: Phone-based MP-Auth Analysis Authentication Protocols

	Entity Authentication			Device Authentication		
	Infallible	Untrained	With Guidelines	Infallible	Untrained	With Guidelines
MP-Auth	✓	✗	✓	✓	✗	✓

Guideline:
NoTellExcept(H, 'pw', 'D')

Guideline:
NoTellExcept(H, 'pw', 'D')

Adversary impersonates H and D to server,
after untrained H enters password on corrupted platform.

Entity Authentication vs Message Authentication

- Both are important.
 - E.g., message (origin) authentication used to authenticate transactions in online banking.
- Some entity authentication protocols can be extended for message authentication
 - Extensions not always possible or straightforward

MP-Auth Message Authentication

H knows: D, P, S, m

D knows: H, pk(S), S, k ← derived from shared key established in login protocol

S knows: sk(S), H, k

H → S: m “please wire 10€ to account #123”

S → D: fresh(r_s) . {m, r_s } $_k$ confirm: 10€ to #123

D •→• H: m transfer 10€ to #123 ?

H •→• D: ‘ok’

D → S: {h(m, r_s)} $_k$ confirmed: 10€ to #123

replay protection

MP-Auth Message Authentication Analysis

- MP-Auth MA with **infallible human**



- MP-Auth MA with **untrained human**

H presses OK without reading display, confirms message *m* sent by adversary.



- **Guidelines** NoTell, NoTellExcept, NoGet, and ICompare **insufficient** to prevent attack.



Improved MP-Auth Message Authentication

H knows: D, P, S, m

D knows: H, pk(S), S, k

S knows: sk(S), H, k

$H \rightarrow S: m$

$S \rightarrow D: \text{fresh}(r_s) . \{m, r_s\}_k$

$D \bullet \rightarrow \bullet H: \text{fresh}(vc) . vc, m$ **H must read display**

$H \bullet \rightarrow \bullet D: vc$ **in order to proceed**

$D \rightarrow S: \{h(m, r_s)\}_k$

Satisfies message authentication with human following **ICompare** guideline.

Google-2-step

with message authentication

$H : \text{knows}(P, D, S, \text{pw}, m, \text{idH})$

$D : \text{knows}(H)$

$S : \text{knows}(H, \text{pw}, D, \text{idH})$

$H \rightarrow P : S, \text{idH}, \text{pw}, m$ ← enters name/password
+ message to authenticate

$P \circ \rightarrow \bullet S : \text{idH}, m$

$S \circ \rightarrow \bullet D : \text{fresh}(c). c, m$

$D \bullet \rightarrow \bullet H : c, m$ ← gets code on device (SMS)

$H \rightarrow P : S, c$ ← code entered on platform

$P \rightarrow S : c, \text{pw}, m$ ← and forwarded to server

Authenticity of m from H to S ?

Authenticity in Google-2-step

For an infallible human: **verified**.

For a fallible human: **falsified**.

Human does not know he has to compare
message on phone with the m that he sent.

For a human with rule $/Compare(H, 'm')$: **verified**.

Comparison: Message Authentication

	Infallible	Untrained	With Guidelines
Cronto MA	✓	×	✓
Google 2-Step*	✓	×	✓
OTP over SMS*	✓	×	✓
MP-Auth VC	✓	×	✓
MP-Auth MA	✓	×	×
Phoolproof*	✓	✓	
Sound-Proof	×		

Guideline: $ICompare(H, 'm')$

* Our extension based on HISP design guidelines.

Conclusion

- First **formal model of human errors** in security protocols, providing **systematic approach** for reasoning about human errors
- **Applications** to authentication protocols:
 - Finding **attacks** arising from human errors.
 - Identifying protocol techniques that provide **effective protection** against various mistakes.
 - Ranking protocols WRT their robustness to human errors

Future Work

- What are good **guidelines**?
- Verify protocols in **combination** of **skilled** and **untrained** human error models.
- **Apply** the model to improve security in the real world:
 - **Improve** system and protocol design.
 - **Identify** critical user actions that must be monitored.
 - Identify critical concepts to **teach** to untrained users.

Literature

- **Modeling Human Errors in Security Protocols**
D.B., Sasa Radomirovic, Lara Schmid, CSF 2016.
- **A Complete Characterisation of Secure Human-Server Communication**
D.B., Sasa Radomirovic, Michael Schläpfer, CSF 2015.

Details

Skilled Humans

- Skilled humans **follow protocol specification**, may make a small number of mistakes (slips & lapses).
- **Slips & lapses**: Inattentiveness, routine behaviour in an unusual situation. E.g, clicking “OK” w/o reading an alert.
- Modelled by adding **failure rules** to protocol model.

Specifying Skilled Human Role

Skilled Human H follows protocol specification, keeps state information: $\text{AgSt}(H, \text{step}, \text{knownTerms})$

Pattern for **receiving** messages:

$$[\text{AgSt}(H, s_1, k), \text{Rcv}(H, \langle t, m \rangle)] \longrightarrow$$
$$[\text{!HK}(H, t, m), \text{AgSt}(H, s_2, \langle k, m \rangle)]$$

Pattern for **sending** messages:

$$[\text{AgSt}(H, s_1, \langle k, m \rangle), \text{!HK}(H, t, m)] \longrightarrow$$
$$[\text{Snd}(H, \langle t, m \rangle), \text{AgSt}(H, s_2, \langle k, m \rangle)]$$

(Trace labels omitted.)

Example of a Failure Rule (Skilled Human Error)

Message confusion: Human H intends to send message m_1 , sends instead message m_2 .

$$[\text{Snd}(H, \langle t_1, m_1 \rangle), !\text{HK}(H, t_2, m_2), \text{Fail}(H, \text{'msc'})] \longrightarrow$$
$$[\text{Snd}(H, \langle t_2, m_2 \rangle)]$$

Fail fact: allows control over type and number of errors.

(Trace labels omitted.)

Related Work

- Beckert and Beuster (2006), Rukšėnas et al. (2008) formally model humans and human error in *human-machine interfaces*.
- Their models correspond to our **skilled human** approach, but capture only *finite scenarios*.
- We model human error in *unbounded protocol executions*.
- A set of failure rules for **skilled human agents** in security protocols are given by Schläpfer (2016).
- Our **untrained human** approach is new.

HISP Channel Assumptions

Authentic Channel:

$$\begin{aligned} [\text{Snd}_A(A, B, m)] \multimap [\text{Snd}_A(A, B, m)] &\rightarrow [\text{!Auth}(A, m), \text{Out}(\langle A, B, m \rangle)] \\ [\text{!Auth}(A, m), \text{In}(B)] \multimap [\text{Rcv}_A(A, B, m)] &\rightarrow [\text{Rcv}_A(A, B, m)] \end{aligned}$$

Confidential Channel:

$$\begin{aligned} [\text{Snd}_C(A, B, m)] \multimap [\text{Snd}_C(A, B, m)] &\rightarrow [\text{!Conf}(B, m)] \\ [\text{!Conf}(B, m), \text{In}(A)] \multimap [\text{Rcv}_C(A, B, m)] &\rightarrow [\text{Rcv}_C(A, B, m)] \\ [\text{In}(\langle A, B, m \rangle)] \multimap [\text{Rcv}_C(A, B, m)] &\rightarrow [\text{Rcv}_C(A, B, m)] \end{aligned}$$

Secure Channel:

$$\begin{aligned} [\text{Snd}_S(A, B, m)] \multimap [\text{Snd}_S(A, B, m)] &\rightarrow [\text{!Sec}(A, B, m)] \\ [\text{!Sec}(A, B, m)] \multimap [\text{Rcv}_S(A, B, m)] &\rightarrow [\text{Rcv}_S(A, B, m)] \end{aligned}$$