Belief in Information Flow

By M. R. Clarkson, A. C. Myers and F. B. Schneider

Sari Haj Hussein[1]

APSIA Breakfast Talk
[1]Interdisciplinary Center for Security, Reliability and Trust
University of Luxembourg

2011-06-22

- Information flow analysis determines the amount of information that is leaked about a program's secret inputs during the execution of that program
- Information flow security establishes bounds on information leakage

- Information flow analysis determines the amount of information that is leaked about a program's secret inputs during the execution of that program
- Information flow security establishes bounds on information leakage

- Qualitative techniques prohibit flow from a program's secret inputs to its public outputs
  $\rightsquigarrow$ Some programs do not function correctly
- Quantitative techniques allow information flow at a certain rate
  $\rightsquigarrow$ At most k bits leak per program execution

- Qualitative techniques prohibit flow from a program's secret inputs to its public outputs
  $\rightsquigarrow$ Some programs do not function correctly

- Quantitative techniques allow information flow at a certain rate
  $\rightsquigarrow$ At most k bits leak per program execution

- Treat a program's execution as a channel for transmitting messages
  $\rightsquigarrow$ Compute the capacity of this channel
- Set bounds on the values of the entropy of input distributions
- Assume that the program input values are independently and uniformly chosen
- Fix a probability distribution on a program's secret inputs
  $\rightsquigarrow$ Clarkson is here

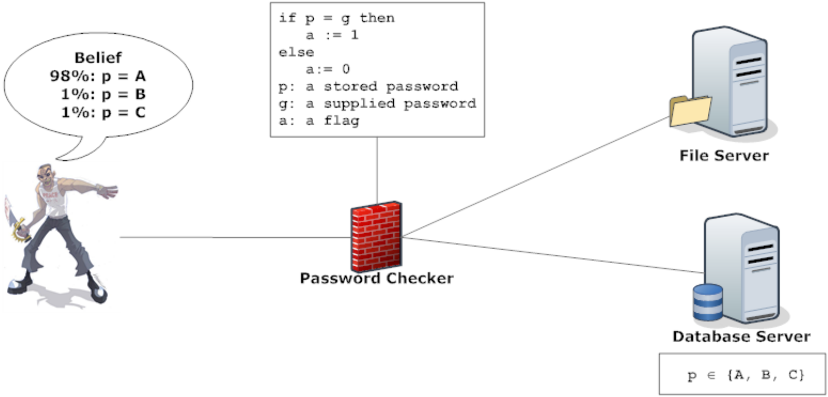
- Treat a program's execution as a channel for transmitting messages
  $\leadsto$ Compute the capacity of this channel
- Set bounds on the values of the entropy of input distributions
- Assume that the program input values are independently and uniformly chosen
- Fix a probability distribution on a program's secret inputs
  $\leadsto$ Clarkson is here

- Treat a program's execution as a channel for transmitting messages
  $\rightsquigarrow$ Compute the capacity of this channel

- Set bounds on the values of the entropy of input distributions
- Assume that the program input values are independently and uniformly chosen
- Fix a probability distribution on a program's secret inputs
  $\rightsquigarrow$ Clarkson is here

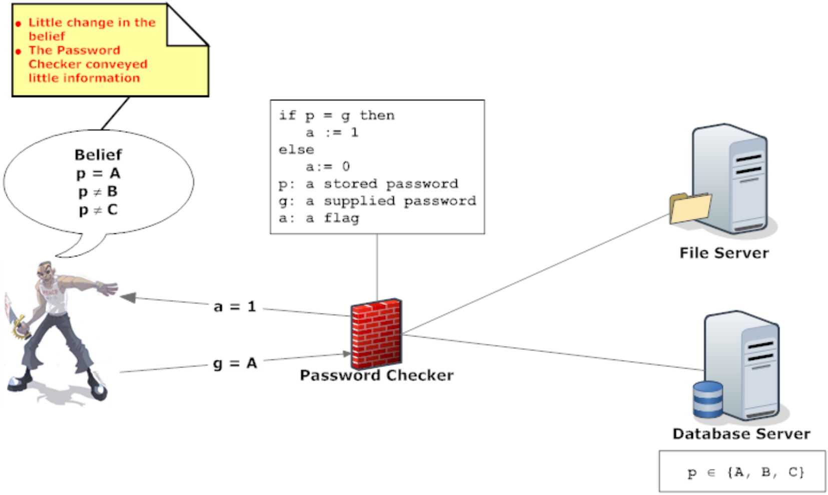
- Treat a program's execution as a channel for transmitting messages
  $\rightsquigarrow$ Compute the capacity of this channel

- Set bounds on the values of the entropy of input distributions
- Assume that the program input values are independently and uniformly chosen
- Fix a probability distribution on a program's secret inputs
  $\rightsquigarrow$ Clarkson is here

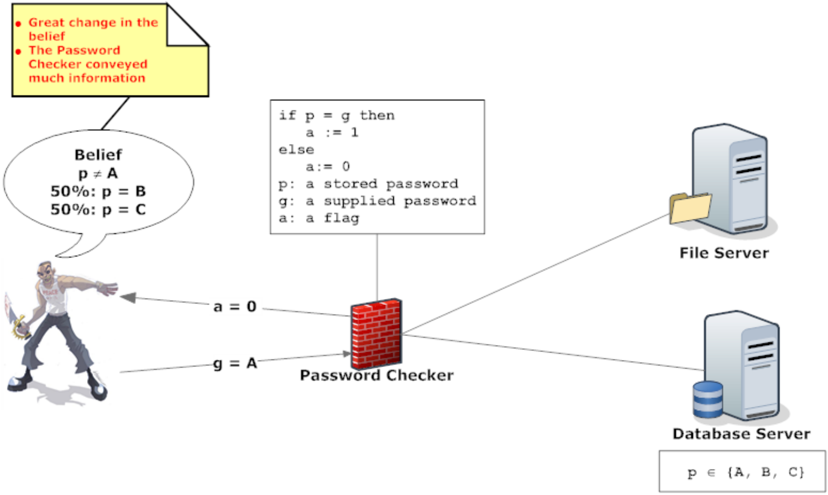


Initial State

Experiment 1



Experiment 1 $\rightsquigarrow$ p is A



Experiment 2 $\rightsquigarrow$ p is C

The Uncertainty Reduction Principle

- A measure of information flow proposed by Denning in the eighties
  - $\uparrow$ in uncertainty $\rightsquigarrow$ information has flowed
  - $\downarrow$ in uncertainty $\rightsquigarrow$ information has not flowed

- This principle is unsuitable when input distributions represent attacker beliefs
  - In the initial state $\rightsquigarrow$ attacker is almost certain
  - After experiment 2 $\rightsquigarrow$ attacker is somewhat uncertain
  - This is $\uparrow$ in uncertainty $\rightsquigarrow$ information has not flowed
  - Untrue!

The Uncertainty Reduction Principle

- A measure of information flow proposed by Denning in the eighties
  - $\uparrow$ in uncertainty $\rightsquigarrow$ information has flowed
  - $\downarrow$ in uncertainty $\rightsquigarrow$ information has not flowed

- This principle is unsuitable when input distributions represent attacker beliefs
  - In the initial state $\rightsquigarrow$ attacker is almost certain
  - After experiment 2 $\rightsquigarrow$ attacker is somewhat uncertain
  - This is $\uparrow$ in uncertainty $\rightsquigarrow$ information has not flowed
  - Untrue!

The Proposed Principle

- Information flow corresponds to an improvement in the accuracy of an attacker's belief
  - $\uparrow$ in accuracy $\rightsquigarrow$ attacker was informed $\rightsquigarrow$ information has flowed
  - $\downarrow$ in accuracy $\rightsquigarrow$ attacker was misinformed $\rightsquigarrow$ information has not flowed

- Based on this principle, we need to devise a measure for information flow...

The Proposed Principle

- Information flow corresponds to an improvement in the accuracy of an attacker's belief
  - $\uparrow$ in accuracy $\rightsquigarrow$ attacker was informed $\rightsquigarrow$ information has flowed
  - $\downarrow$ in accuracy $\rightsquigarrow$ attacker was misinformed $\rightsquigarrow$ information has not flowed

- Based on this principle, we need to devise a measure for information flow...

- We suppose 4 sets
  - *Var* set of variables
  - *Val* set of values
  - *State* set of program states
  - *Dist* set of distributions
- A state $\sigma \in State$ is an assignment in $Var \to Val$
- A distribution $\delta \in Dist$ is an assignment in $State \to \mathbb{R}^+$
- A state mass $\dot{\sigma}$ is a probability distribution that maps σ to 1
- With a program S, we use the function $[S] : State \to Dist$

- We suppose 4 sets
  - *Var* set of variables
  - *Val* set of values
  - *State* set of program states
  - *Dist* set of distributions
- A state $\sigma \in \mathit{State}$ is an assignment in $\mathit{Var} \to \mathit{Val}$
- A distribution $\delta \in \mathit{Dist}$ is an assignment in $\mathit{State} \to \mathbb{R}^+$
- A state mass $\dot{\sigma}$ is a probability distribution that maps σ to 1
- With a program S, we use the function $[S] : \mathit{State} \to \mathit{Dist}$

- We use confidentiality labels to identify secret data
  - $L \rightsquigarrow$ low-confidentiality public data
  - $H \rightsquigarrow$ high-confidentiality secret data

- $\sigma \upharpoonright L \rightsquigarrow$ low projection of the state $\sigma \rightsquigarrow$ the part of σ visible to the attacker
- $\sigma \upharpoonright H \rightsquigarrow$ high projection of the state $\sigma \rightsquigarrow$ the part of σ not visible to the attacker
- $x_L \rightsquigarrow$ a variable that contains low information
- $x_H \rightsquigarrow$ a variable that contains high information

- We use confidentiality labels to identify secret data
  - $L \rightsquigarrow$ low-confidentiality public data
  - $H \rightsquigarrow$ high-confidentiality secret data

- $\sigma \upharpoonright L \rightsquigarrow$ low projection of the state $\sigma \rightsquigarrow$ the part of σ visible to the attacker
- $\sigma \upharpoonright H \rightsquigarrow$ high projection of the state $\sigma \rightsquigarrow$ the part of σ not visible to the attacker
- $x_L \rightsquigarrow$ a variable that contains low information
- $x_H \rightsquigarrow$ a variable that contains high information

- We use confidentiality labels to identify secret data
  - $L \rightsquigarrow$ low-confidentiality public data
  - $H \rightsquigarrow$ high-confidentiality secret data

- $\sigma \upharpoonright L \rightsquigarrow$ low projection of the state $\sigma \rightsquigarrow$ the part of σ visible to the attacker
- $\sigma \upharpoonright H \rightsquigarrow$ high projection of the state $\sigma \rightsquigarrow$ the part of σ not visible to the attacker
- $x_L \rightsquigarrow$ a variable that contains low information
- $x_H \rightsquigarrow$ a variable that contains high information

- PWC: if $p_H = g_L$ then $a_L := 1$ else $a_L := 0$
- The attacker chooses a pre-belief $b_H = (0.98, 0.01, 0.01)$
- The system chooses $\sigma_H = (p \rightarrow A)$
- The attacker chooses $\sigma_L = (g \rightarrow A, a \rightarrow 0)$

- The input to PWC is $\dot{\sigma_L} \otimes \dot{\sigma_H}$
- PWC executes once
- The output is a frequency distribution $\delta' = [PWC](\dot{\sigma_L} \otimes \dot{\sigma_H})$ from which one state is chosen $\sigma' = (p \rightarrow A, g \rightarrow A, a \rightarrow 1)$
- The attacker observes $o = \sigma' \upharpoonright L = (g \rightarrow A, a \rightarrow 1)$

- PWC: if $p_H = g_L$ then $a_L := 1$ else $a_L := 0$
- The attacker chooses a pre-belief $b_H = (0.98, 0.01, 0.01)$
- The system chooses $\sigma_H = (p \rightarrow A)$
- The attacker chooses $\sigma_L = (g \rightarrow A, a \rightarrow 0)$

- The input to PWC is $\dot{\sigma_L} \otimes \dot{\sigma_H}$
- PWC executes once
- The output is a frequency distribution $\delta' = [PWC](\dot{\sigma_L} \otimes \dot{\sigma_H})$ from which one state is chosen $\sigma' = (p \rightarrow A, g \rightarrow A, a \rightarrow 1)$
- The attacker observes $o = \sigma' \upharpoonright L = (g \rightarrow A, a \rightarrow 1)$

- The attacker generates a prediction of getting authenticated $\delta'_A = [PWC](\dot{\sigma}_L \otimes b_H)$
- To incorporate the information in o, The attacker conditions $\delta'_A|o$

p	g	a	δ'_A	$\delta'_A\|o$
A	A	0	0	0
A	A	1	0.98	1
B	A	0	0.01	0
B	A	1	0	0
C	A	0	0.01	0
C	A	1	0	0

- The attacker projects on the high state to obtain her post-belief $b'_H = (\delta'_A|o) \upharpoonright H = (1, 0, 0)$
- This matches with the informal reasoning

The Proposed Principle

- Information flow corresponds to an improvement in the accuracy of an attacker's belief
  - $\uparrow$ in accuracy $\rightsquigarrow$ attacker was informed $\rightsquigarrow$ information has flowed
  - $\downarrow$ in accuracy $\rightsquigarrow$ attacker was misinformed $\rightsquigarrow$ information has not flowed

- How can we use that?...

The Proposed Principle

- Information flow corresponds to an improvement in the accuracy of an attacker's belief
  - $\uparrow$ in accuracy $\rightsquigarrow$ attacker was informed $\rightsquigarrow$ information has flowed
  - $\downarrow$ in accuracy $\rightsquigarrow$ attacker was misinformed $\rightsquigarrow$ information has not flowed

- How can we use that?...

The Proposed Measure

- Information flow corresponds to an improvement in the accuracy of an attacker's belief
  - Accuracy of attacker's pre-belief b_H is $D(b_H \rightarrow \dot{\sigma_H})$ (Kullback–Leibler divergence)
  - Accuracy of attacker's post-belief b'_H is $D(b'_H \rightarrow \dot{\sigma_H})$
  - $\Delta = D(b_H \rightarrow \dot{\sigma_H}) - D(b'_H \rightarrow \dot{\sigma_H})$
  - $= \dot{\sigma}_H \bullet log \frac{\dot{\sigma}_H}{b_H(\sigma_H)} - \dot{\sigma}_H \bullet log \frac{\dot{\sigma}_H}{b'_H(\sigma_H)}$ (Kullback–Leibler)
  - $= 1 \bullet log \frac{1}{b_H(\sigma_H)} - 1 \bullet log \frac{1}{b'_H(\sigma_H)}$ (Definition of state mass)
  - $= -log b_H(\sigma_H) + log b'_H(\sigma_H)$
  - $= -log b_H(\sigma_H) + log b_H(\sigma_H) \bullet \frac{\delta_S(o)}{\delta_A(o)}$ (proved in the paper)
  - $= -log b_H(\sigma_H) + log b_H(\sigma_H) + log \frac{\delta_S(o)}{\delta_A(o)}$
  - $= -log \delta_A(o) + log \delta_S(o)$
  - $= -log Pr_{\delta_A}(o) + log Pr_{\delta_S}(o)$
  - $= I_{\delta_A}(o) - I_{\delta_S}(o)$ (a result from information theory)

The Proposed Measure

- Information flow corresponds to an improvement in the accuracy of an attacker's belief
  - Accuracy of attacker's pre-belief b_H is $D(b_H \to \dot{\sigma_H})$ (Kullback–Leibler divergence)
  - Accuracy of attacker's post-belief b'_H is $D(b'_H \to \dot{\sigma_H})$
  - $\Delta = D(b_H \to \dot{\sigma_H}) - D(b'_H \to \dot{\sigma_H})$
  - $= \dot{\sigma}_H \bullet log \frac{\dot{\sigma}_H}{b_H(\sigma_H)} - \dot{\sigma}_H \bullet log \frac{\dot{\sigma}_H}{b'_H(\sigma_H)}$ (Kullback–Leibler)
  - $= 1 \bullet log \frac{1}{b_H(\sigma_H)} - 1 \bullet log \frac{1}{b'_H(\sigma_H)}$ (Definition of state mass)
  - $= -log b_H(\sigma_H) + log b'_H(\sigma_H)$
  - $= -log b_H(\sigma_H) + log b_H(\sigma_H) \bullet \frac{\delta_S(o)}{\delta_A(o)}$ (proved in the paper)
  - $= -log b_H(\sigma_H) + log b_H(\sigma_H) + log \frac{\delta_S(o)}{\delta_A(o)}$
  - $= -log \delta_A(o) + log \delta_S(o)$
  - $= -log Pr_{\delta_A}(o) + log Pr_{\delta_S}(o)$
  - $= I_{\delta_A}(o) - I_{\delta_S}(o)$ (a result from information theory)

The Proposed Measure

- Information flow corresponds to an improvement in the accuracy of an attacker's belief
  - Accuracy of attacker's pre-belief b_H is $D(b_H \to \dot{\sigma_H})$ (Kullback–Leibler divergence)
  - Accuracy of attacker's post-belief $b_H^{'}$ is $D(b_H^{'} \to \dot{\sigma_H})$

  - $\Delta = D(b_H \to \dot{\sigma_H}) - D(b_H^{'} \to \dot{\sigma_H})$
  - $= \dot{\sigma}_H \bullet log \frac{\dot{\sigma}_H}{b_H(\sigma_H)} - \dot{\sigma}_H \bullet log \frac{\dot{\sigma}_H}{b_H^{'}(\sigma_H)}$ (Kullback–Leibler)
  - $= 1 \bullet log \frac{1}{b_H(\sigma_H)} - 1 \bullet log \frac{1}{b_H^{'}(\sigma_H)}$ (Definition of state mass)
  - $= -log b_H(\sigma_H) + log b_H^{'}(\sigma_H)$
  - $= -log b_H(\sigma_H) + log b_H(\sigma_H) \bullet \frac{\delta_S(o)}{\delta_A(o)}$ (proved in the paper)
  - $= -log b_H(\sigma_H) + log b_H(\sigma_H) + log \frac{\delta_S(o)}{\delta_A(o)}$
  - $= -log \delta_A(o) + log \delta_S(o)$
  - $= -log Pr_{\delta_A}(o) + log Pr_{\delta_S}(o)$
  - $= I_{\delta_A}(o) - I_{\delta_S}(o)$ (a result from information theory)

Flow in Experiment 1

- $\Delta_1 = -log Pr_{\delta_A}(o_1) + log Pr_{\delta_S}(o_1) = -log 0.98 + log 1 =$ 0.0291 bit.

Flow in Experiment 2

- $\Delta_2 = -log Pr_{\delta_A}(o_2) + log Pr_{\delta_S}(o_2) = -log 0.02 + log 1 =$ 5.6439 bit.

- Thus the flow in Experiment 2 is larger than it is in Experiment 1
- Again, this matches with the informal reasoning

Flow in Experiment 1

- $\Delta_1 = -logPr_{\delta_A}(o_1) + logPr_{\delta_S}(o_1) = -log0.98 + log1 = 0.0291$ bit.

Flow in Experiment 2

- $\Delta_2 = -logPr_{\delta_A}(o_2) + logPr_{\delta_S}(o_2) = -log0.02 + log1 = 5.6439$ bit.

- Thus the flow in Experiment 2 is larger than it is in Experiment 1
- Again, this matches with the informal reasoning

Flow in Experiment 1

- $\Delta_1 = -logPr_{\delta_A}(o_1) + logPr_{\delta_S}(o_1) = -log0.98 + log1 = 0.0291$ bit.

Flow in Experiment 2

- $\Delta_2 = -logPr_{\delta_A}(o_2) + logPr_{\delta_S}(o_2) = -log0.02 + log1 = 5.6439$ bit.

- Thus the flow in Experiment 2 is larger than it is in Experiment 1
- Again, this matches with the informal reasoning

- The Uncertainty Reduction Principle cannot satisfactorily explain information flow when input distributions represent attacker beliefs
- Accuracy is the appropriate measure for information flow in the presence of attacker beliefs

Thank you!