CS 599: Formal Methods in Security and Privacy

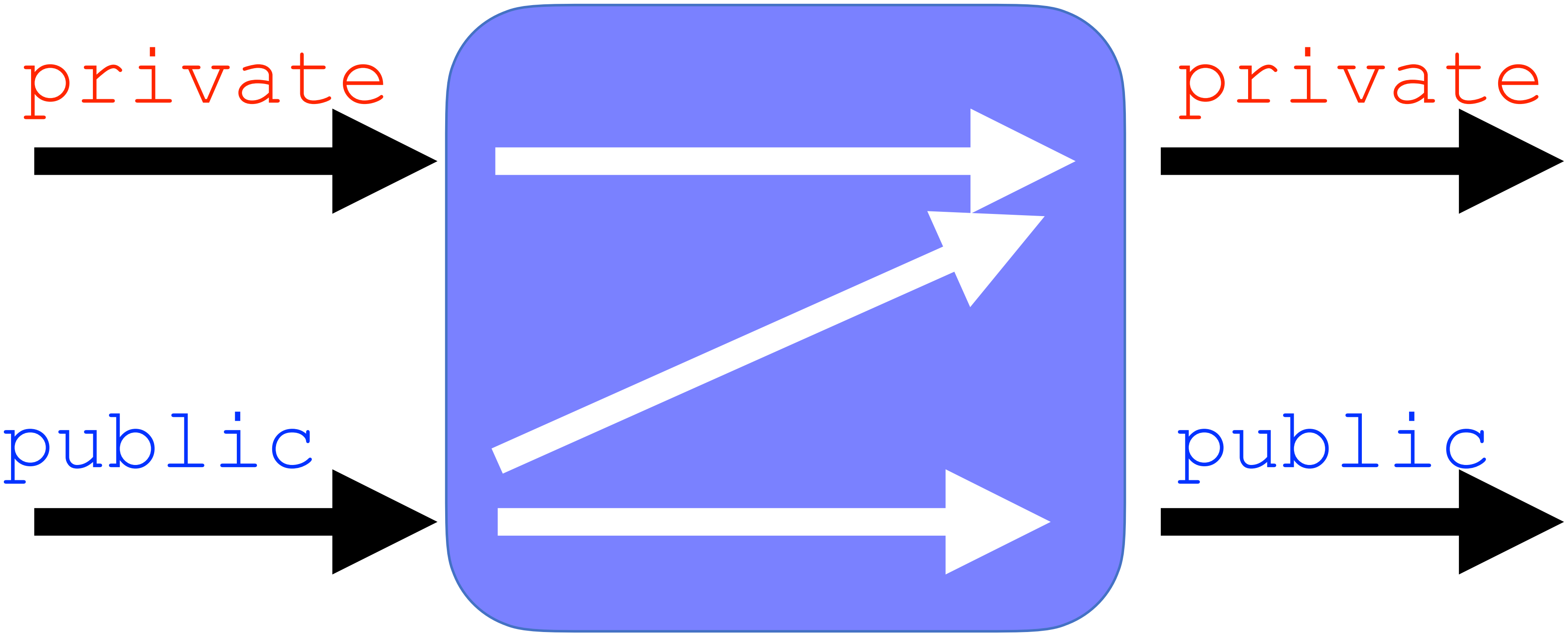
Quantitative Information Flow

Marco Gaboardi
gaboardi@bu.edu

Alley Stoughton
stough@bu.edu

Information Flow Control

We want to guarantee that confidential information do not flow in what is considered nonconfidential.



Comparing strings

```
s1:public
s2:private
r:private
i:public

proc Compare (s1:list[n] bool,s2:list[n] bool)
i:=0;
r:=0;
while i<n do
 if not(s1[i]=s2[i]) then
    r:=1
 i:=i+1

: n>0 /\ =low ⇒ =low
```

Comparing strings

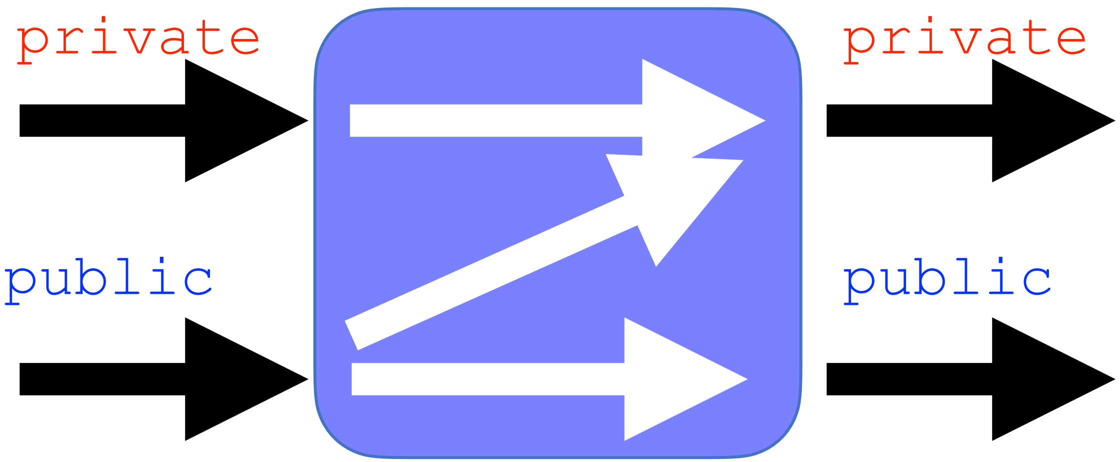
```
s1:public
s2:private
r:private
i:public

proc Compare (s1:list[n] bool,s2:list[n] bool)
i:=0;
r:=0;
while i<n do
 if not(s1[I]=s2[i]) then
     r:=1;
     i:=n-1;
 i:=i+1

: n>0 /\ =low ⇒ =low
```

Releasing the mean of Some Data

```
Mean(d : private data) : public real
   i:=0;
   s:=0;
   while (i<size(d))
       s:=s + d[i]
       i:=i+1;
   return (s/i)
```



(ε,δ)-Differential Privacy

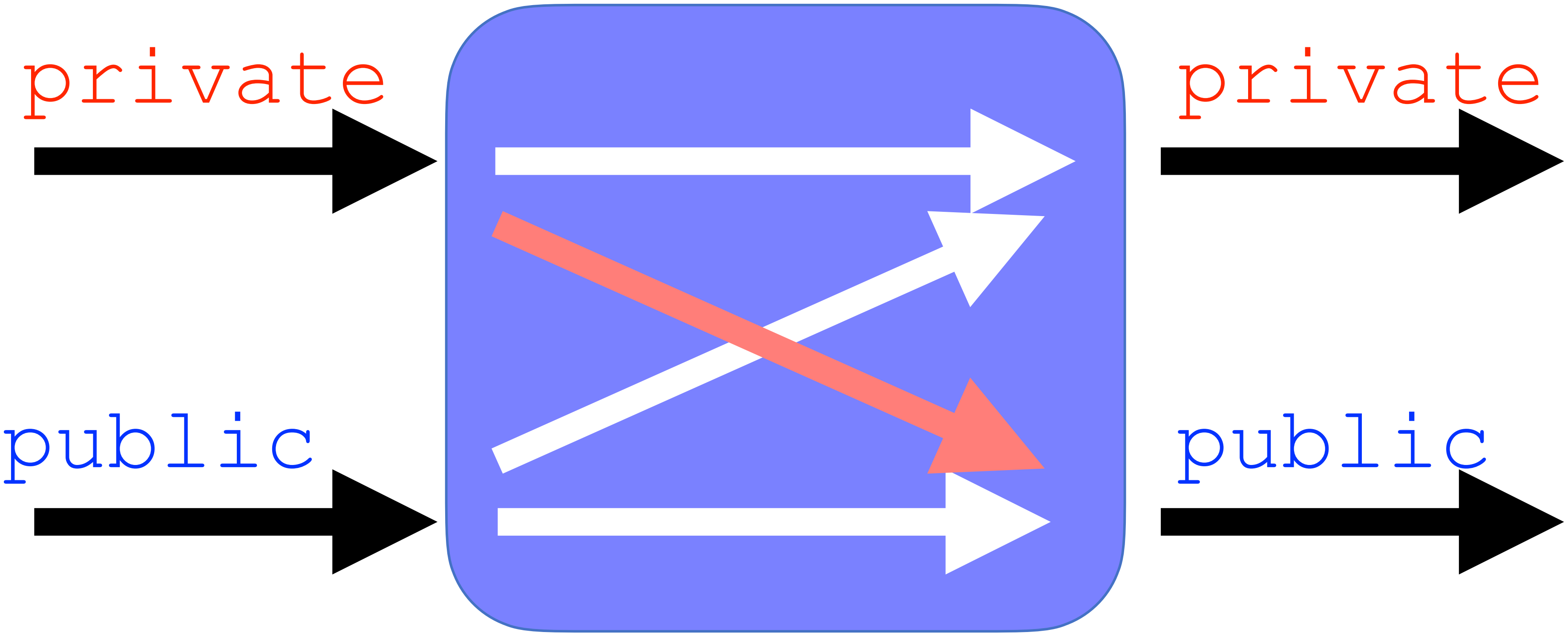
Definition

Given $\varepsilon,\delta \geq 0$, a probabilistic query $Q: X^n \rightarrow R$ is (ε,δ)-differentially private iff for all adjacent database b_1, b_2 and for every $S \subseteq R$:

$$\Pr[Q(b_1) \in S] \leq \exp(\varepsilon)\Pr[Q(b_2) \in S] + \delta$$

Quantitative Information Flow Control

We want to quantify the confidential information that leaks in what is considered nonconfidential.



Quantitative Information Flow Control

Quantitative information flow has been used for:

- Analyzing distributed protocols and scheme,
- Analyzing side-channel vulnerabilities and preventions.
- Analyzing crypto protocols,
- Analyze election protocols
- Analyze differential privacy mechanisms
- …

Information Security and Cryptography

Mário S. Alvim
Konstantinos Chatzikokolakis
Annabelle McIver · Carroll Morgan
Catuscia Palamidessi · Geoffrey Smith

The Science of Quantitative Information Flow

Springer

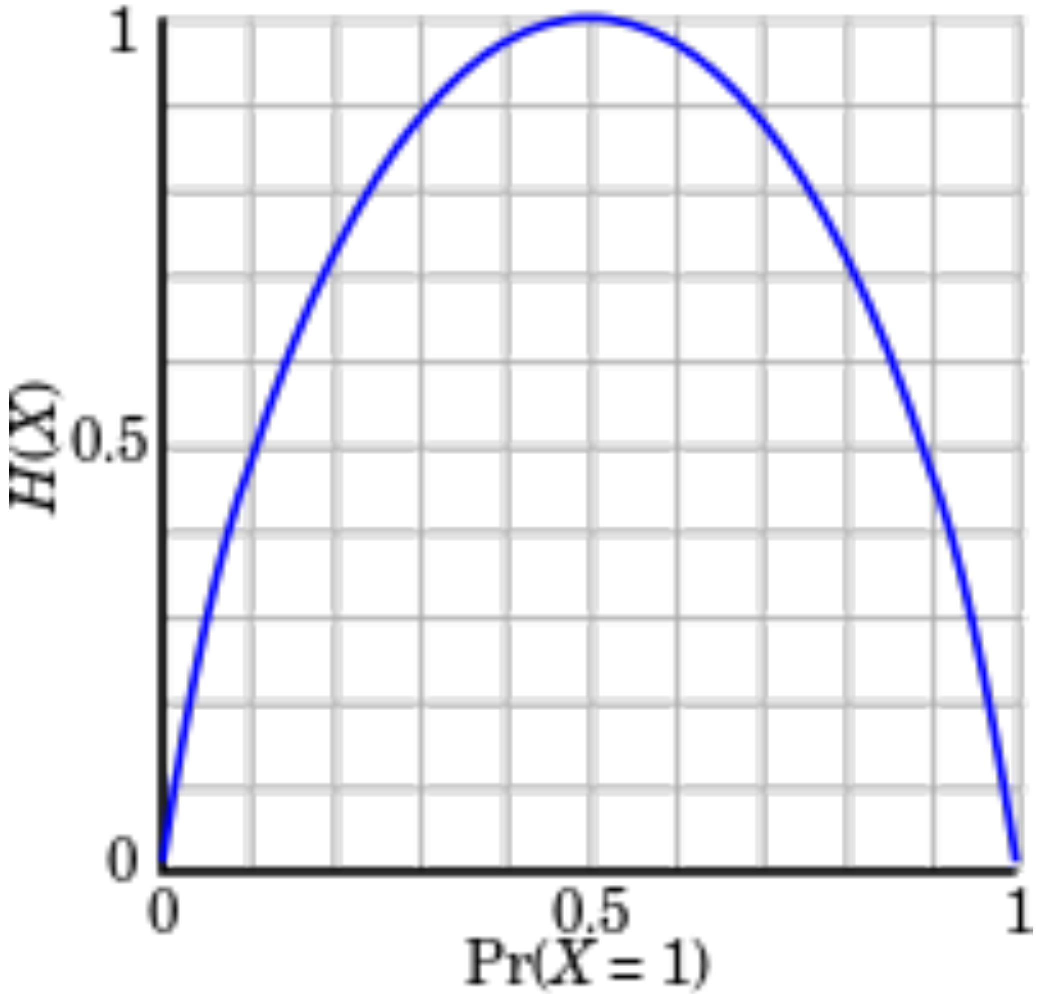
How do we quantify information (leakage)?

Shannon Entropy

$$H(X) = \sum_{x \in \mathcal{X}} \Pr[X = x] \log(\frac{1}{\Pr[X = x]}) = \mathbb{E}\left[\log(\frac{1}{\Pr[X = x]})\right]$$

- uncertainty about X
- expected amount of information gain by observing the value of the random variable,
- average number of bits required to transmit X optimally

Shannon Entropy of coins



Conditional Entropy

$$H(X \mid Y) = \sum_{y} \Pr[Y = y] \cdot \sum_{x \in \mathcal{X}} \Pr[X = x \mid Y = y] \log\left(\frac{1}{\Pr[X = x \mid Y = y]}\right)$$

- uncertainty about X given Y

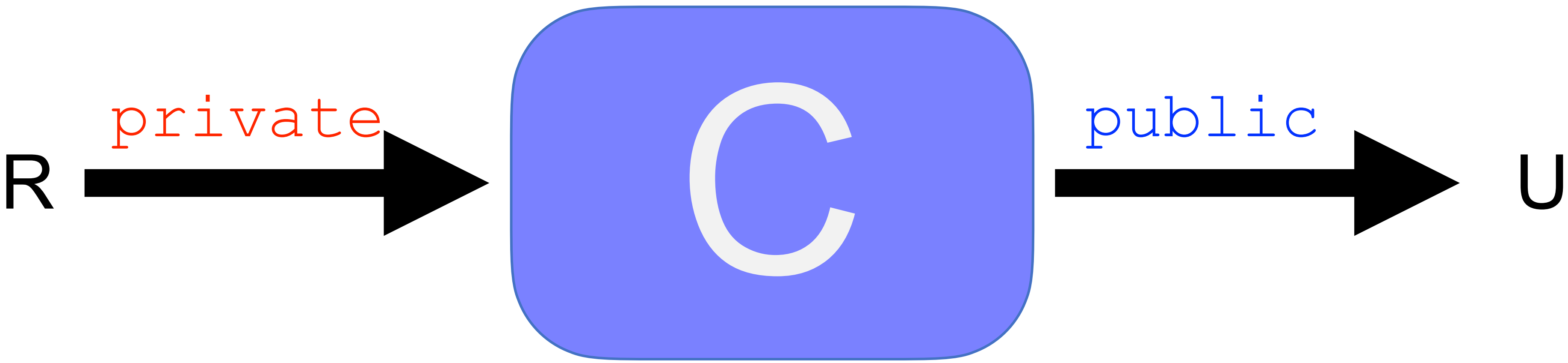
Mutual Information

$$I(X;Y) = H(X) - H(X \mid Y)$$

- amount of information shared between X and Y

How can we use these measures for QIF?

A concrete setting



Guessing Game

- The adversary has some prior π_R on R and it updates it after seeing U.

Information leakage

Information leaked =

initial uncertainty - remaining uncertainty

- Which could be

$$\texttt{Leakage}(U) = H(R) - H(R \mid U)$$

- This is the mutual information between R and U

Conditional Entropy

$$H(X \mid Y) = \sum_{y} \Pr[Y = y] \cdot \sum_{x \in \mathscr{X}} \Pr[X = x \mid Y = y] \log(\frac{1}{\Pr[X = x \mid Y = y]})$$

- If C is constant H(R|U)=1.
- If C is non constant and deterministic H(R|U)=0, so:

$$\mathtt{Leakage}(U) = H(R)$$

Example

- Assume that R is a uniformly-distributed 32bit integer

Program	Leakage(U)	H(R)	H(R|U)
U:=0	0	32	32
U:=R	32	32	0
U:= R && 11111	5	32	27

Properties

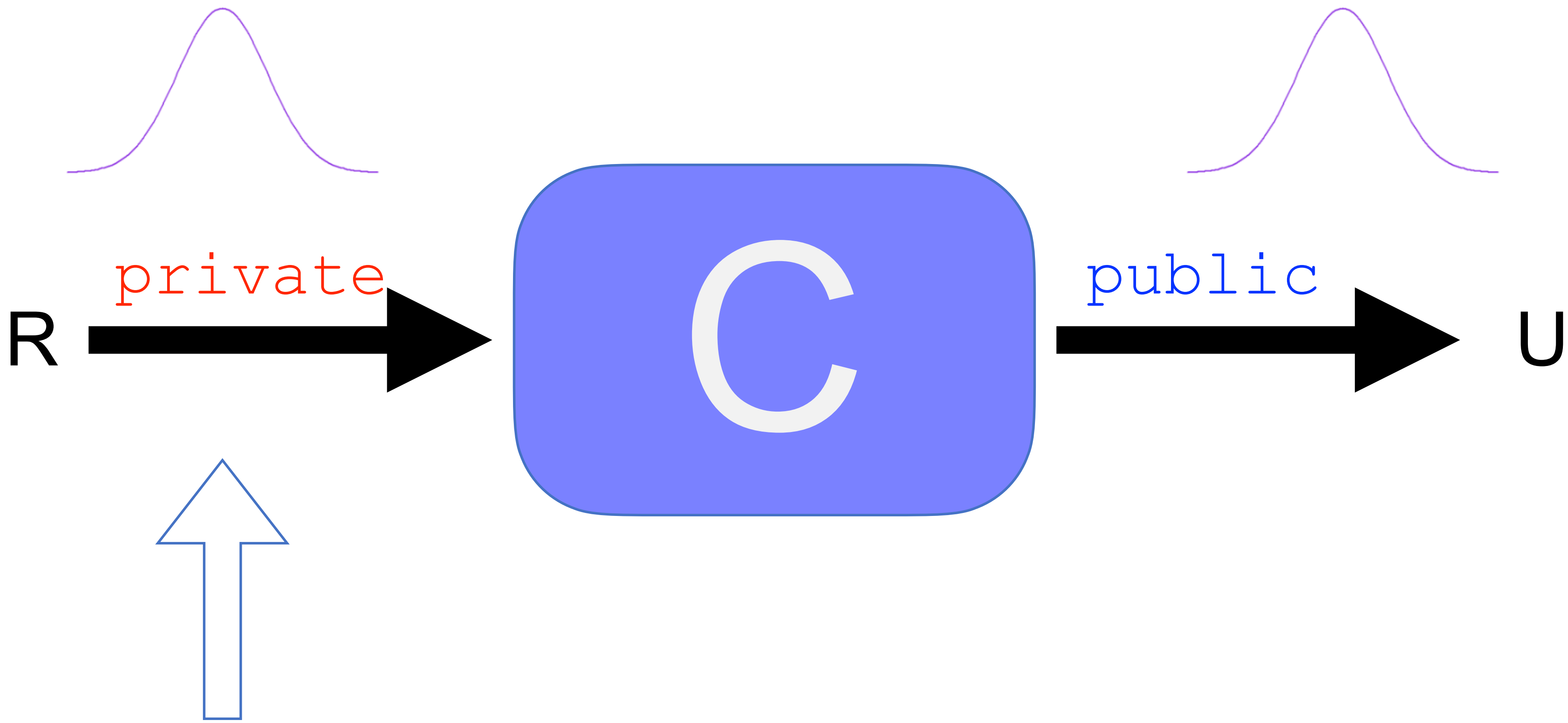
- If C is deterministic we have Leakage (U)=0 iff C satisfies non-interference
- We have $G(R|U) \geq 2^{H(R|U)-2}+1$ where

$$G(X|Y) = \sum_i i \cdot \Pr[X = x_i | Y = y]$$

Is the conditional guessing entropy, i.e. the expected number of guesses needed to guess X given Y. (We assume the probabilities to be in non-decreasing order).

Is Shannon entropy the only measure?

Guessing Game



Let’s focus on the prior

Shannon Entropy

$$H(X) = \sum_{x \in \mathcal{X}} \Pr[X = x] \log\left(\frac{1}{\Pr[X = x]}\right) = \mathbb{E}\left[\log\left(\frac{1}{\Pr[X = x]}\right)\right]$$

- A point distribution has Shannon entropy 0
- A uniform distribution of n values has Shannon entropy log(n).

Shannon Entropy

We could think that:

"If a secret X has distribution π, then an adversary's probability of guessing the value of X correctly in one try is at most $2^{-H(\pi)}$"

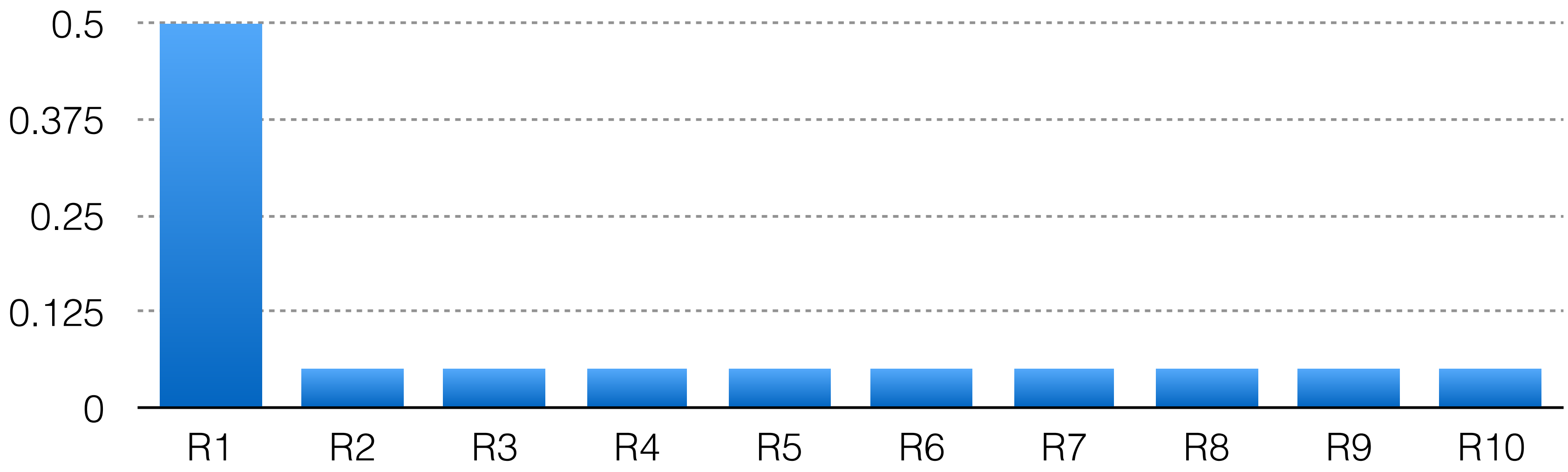
- This is false. E.g. for this distribution $H(\pi) \sim 2.44$, and $2^{-H(\pi)} \sim 0.18$

Shannon Entropy

We could think that:

> "If a secret X has distribution π, then an adversary's probability of guessing the value of X correctly in one try is at most $2^{-H(\pi)}$"

- This is false. E.g. for this distribution $H(\pi)$~2.44, and $2^{-H(\pi)} \sim 0.18$



Bayes Vulnerability

$$V(X) = \max_{x \in \mathcal{X}} \Pr[X = x]$$

- In our case it is the max probability assigned by the prior π_R.
- Best choice for a rational adversary to guess the secret in one try.

Bayes Vulnerability examples

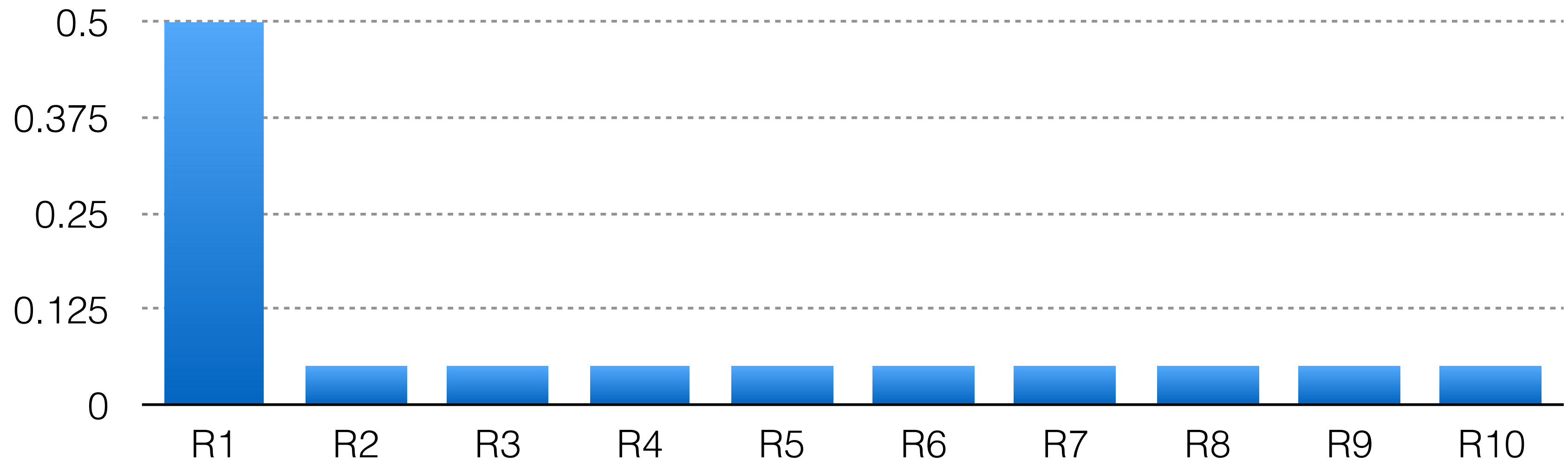
$$V(X) = \max_{x \in \mathcal{X}} \Pr[X = x]$$

- Consider π_R to be a uniform distribution over n outcomes. Then, $V(\pi_R)=1/n$
- Consider π_R to be the following distribution again, we have $V(\pi_R)=.5$

Bayes Vulnerability examples

$$V(X) = \max_{x \in \mathcal{X}} \Pr[X = x]$$

- Consider π_R to be a uniform distribution over n outcomes. Then, $V(\pi_R)=1/n$
- Consider π_R to be the following distribution again, we have $V(\pi_R)=.5$



How do we quantify information leakage?

In Easycrypt

- Look at how to guarantee trace-based noninterference.
- Look at how to guarantee side-channel free noninterference.
- Look at the relations between self-composition and relational logic.

Not related to Easycrypt

- Look at type systems for non-interference.
- Look at other methods for relational reasoning
- Look at declassification