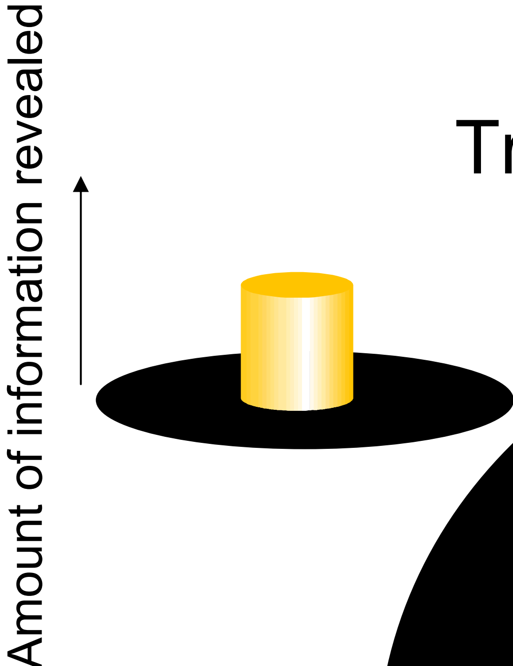
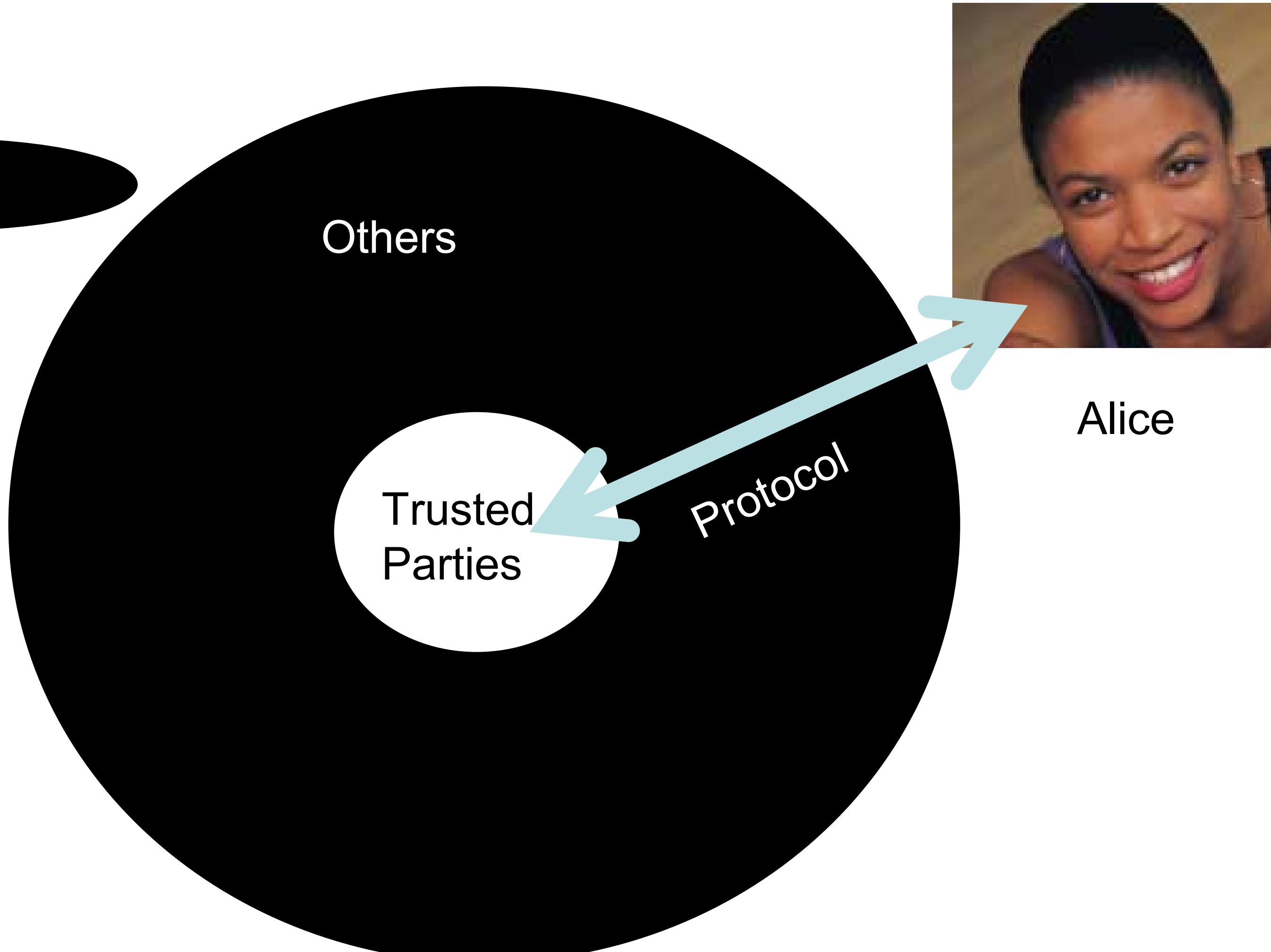
Towards a theory of variable privacy

Poorvi Vora

Hewlett-Packard Co.

Traditional security model





Traditional theory of security

Desirable protocols do not leak any information to non-trusted parties

<u>Information-theoretically perfect secrecy:</u>

a priori and *a posteriori* pdfs identical

− no information leakage to any adversary

<u>Computationally perfect secrecy:</u>

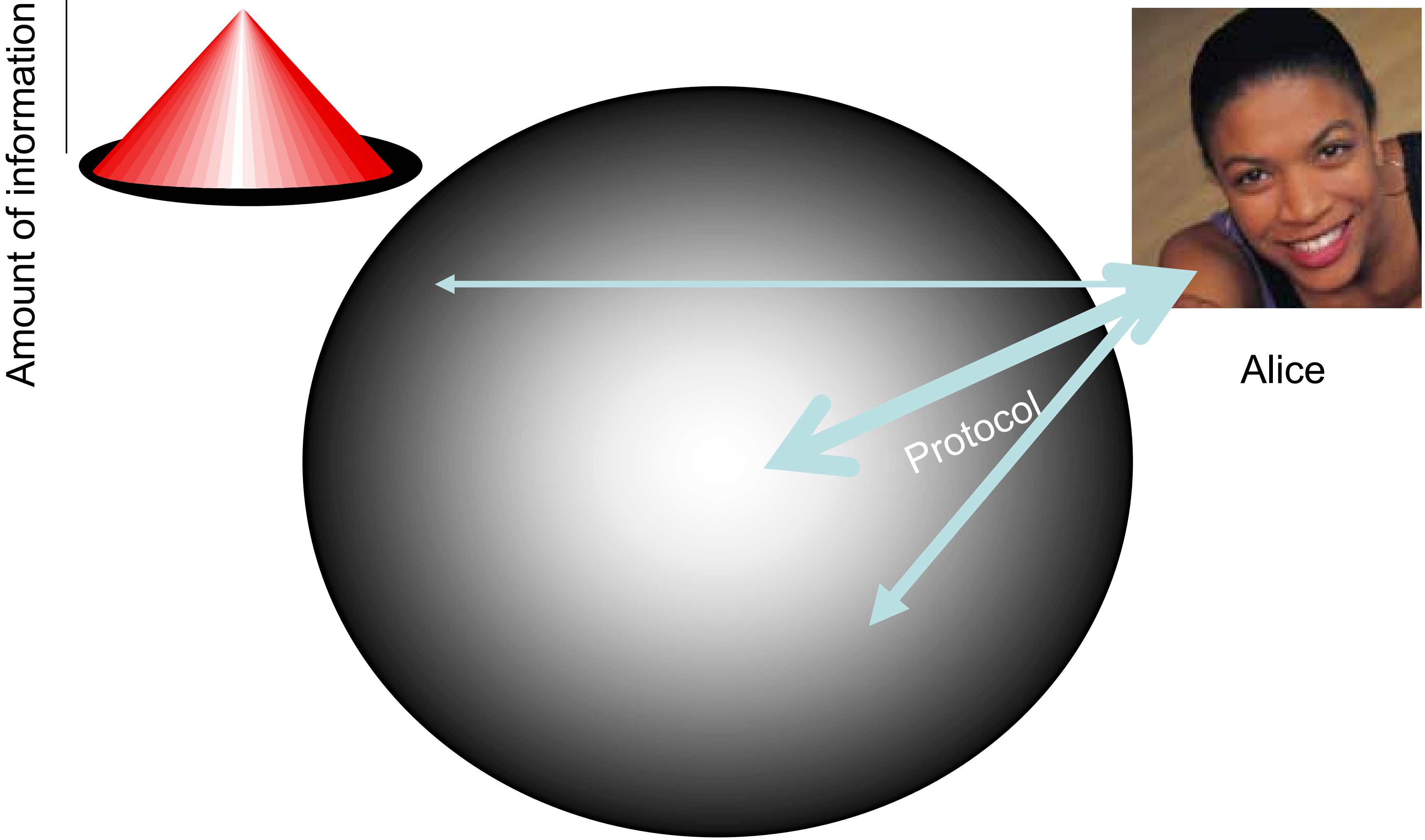
a probabilistic polynomial-time algorithm cannot distinguish between prior and posterior

− no information leakage to realistic adversary

Problems not addressed by perfectly secret protocols

- Need to leak statistics in:
  - Markets
  - Statistical databases
  - Collaborative filtering
- Need another model for communities
- There is an existing market for personal information
  - Safeway cards for 10% discount
  - Extra for unlisted phone numbers
  - Need an understanding of "amount of privacy" to study the value of privacy in this market

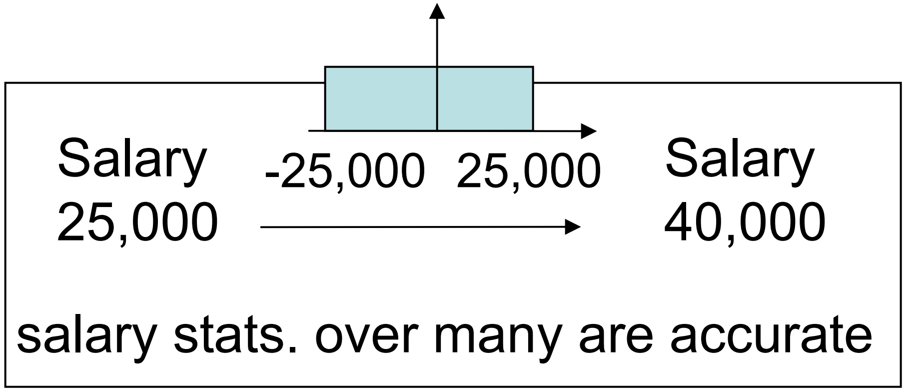
The privacy world



An intentionally non-perfect protocol

- Randomization (probabilistic perturbation of data)
  - provides statistics to data collector, privacy to individual
- Current Uses:
  - Public health surveys (20+ years)
  - Statistical database security (20+ years)
  - IBM application for personal privacy protection on data collection websites (6 months)
- Potential Use, with Alice's participation
  - Interaction with parties neither trusted nor untrusted (e.g. virtual communities)
  - Collaborative filtering with privacy
  - Negotiations

Randomization: continuous-valued



The output now decreases possible salary range:

15-65K

Randomization: binary-valued

HIV? —— P(truth) = 2/3 / P(lie) = 1/3 ——→ Yes

stats. over many are accurate

After the protocol, the possibilities are skewed

the answer is most likely to be correct

The statistical database security problem

- Data collector asks for:

$$f_i(x_1, x_2, x_3, \ldots) = A_i$$

- Can simultaneously solve above
- (perfect zk protocols do not leak additional information about x_i, but A_i are revealed; thus not a traditional cryptographic problem)
- If x_i perturbed each time, the equations are inconsistent

$$f_i(x_1+\Delta_{1i}, x_2 +\Delta 2_i, x_3 +\Delta_{3i}, \ldots) = A_i+ \Delta_i$$

- Security and attack characterization open problem for 20+ years; though many attempts (Denning, Adams, Duncan, … Landers).

Variable Privacy

Definition 1: "variable privacy" is the use of non-perfect protocols with Alice's participation in choice of protocol parameters

Natural consequence of the definition of privacy in a world that includes non-perfect protocols

Need a framework for “variable privacy”

- What is a measure of the privacy provided by randomization?
- Can it be related to the “security’’ of randomization?

Our privacy model

1. Alice and Bob determine a level of information leakage, P(Y|X)

2. Bob requests a data point X from Alice, she reveals Y according to P(Y|X)

3. Bob provides something to Alice in return

- Dishonest Bob can use the information leakage to find out more than Alice intended

- The cost to Dishonest Bob is a measure of protocol privacy

Would provide a framework for “variable privacy”, and an understanding of the security of randomization, an open problem for 20 years in statistical databases

Literature on information-theoretic measures of randomization (continuous-valued data)

D. Agrawal and C. Aggarwal (2001): *Mutual information*

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Measures change due to protocol and weights different probabilities differently

Problem: Dependent on pdf of X.

Natural fix: *Channel capacity*

$$C(X; Y) = \max_{p(x)} H(X) - H(X|Y) = \max_{p(y)} H(Y) - H(Y|X)$$

Related to protocol security?

Our approach - 1

Shannon's paper on secrecy:

- A protocol is perfect

$\Leftrightarrow$ the prior is identical to the posterior, i.e.

$\Leftrightarrow$ it is not a channel (or is a channel with zero capacity)

Randomization is generally not perfect

$\Leftrightarrow$ randomization is a channel with non-zero capacity;

(non-typical view of privacy/secrecy protocols)

Dishonest Bob wishes efficient communication over the channel

Protocol as channel

Protocol Input: The truth value of “X has HIV”

Output: Perturbed value of the bit.
Probabilities: of truth: 2/3, of lie: 1/3

Communication channel with probability of error 1/3

Our approach - 2

Yao's paper on computationally perfect secrecy:

– A protocol is computationally secret
$\Leftrightarrow$ prior and posterior computationally indistinguishable

Randomization is computationally imperfect

Computationally feasible attacks are (trivially) known to exist

Thus, their cost is important

Our approach - 3

All communication over the protocol-channel (including attacks) is governed by Shannon's theorems on communication in the presence of noise

We use the theorems to derive

- the complexity of attacks for arbitrarily small errors, and
- a corresponding privacy measure

We have not seen the connection between

- Shannon's work on secrecy and communication in the presence of noise anywhere else,
- though the connection between communication over a noiseless channel and secrecy has been published (Brassard and Giles)

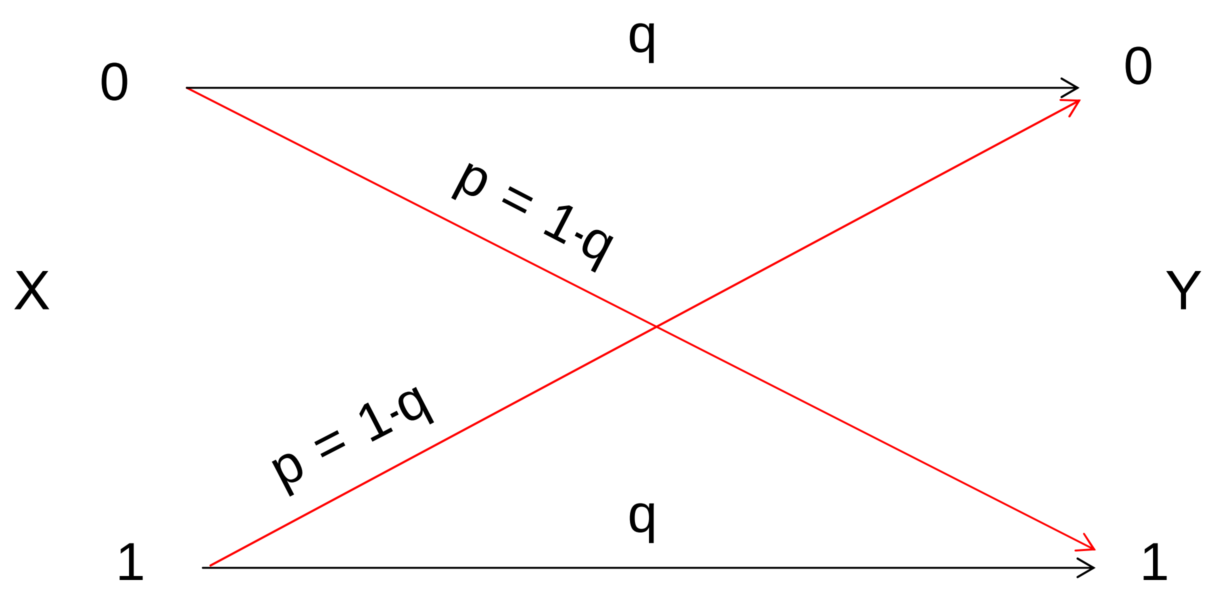
Formally: Protocols as communication channels

$$\varphi: \mathcal{X} \rightarrow \mathcal{Y}$$

$$\varphi(X) = Y$$

- $\mathcal{X}$ is the set of all possible values of user personal information, plaintext
- $\mathcal{Y}$ is the set of all possible values of observable information from a single instance of the protocol or the attack, ciphertext
- Unlike channels in communication theory, the purpose of φ is to limit communication of X.
- $\Phi = (\mathcal{X}, P(Y|X), \mathcal{Y})$

Binary Symmetric Randomization Protocol



$$\Phi = (\ \{0,1\},\ \{0,1\},\ P(Y|X))$$

$$P(Y \mid X) = q,\ Y = X;\ = 1\text{-}q,\ Y \neq X$$

Typical query sequence for attack

message bit 1: female?
message bit 2: over 40?

plaintext bit 1: Losing Calcium?
plaintext bit 2: Graying?
plaintext bit 3: Balding?
plaintext bit 4: Gaining weight?

Rate defined as
log(no of possible messages)/plaintext length

Rate (efficiency) of above attack = log (4)/4 = 0.5

PRP

Definition 2:

A *plaintext* is a string of bits each a function of bits in the database: $p = (f_1(a)_{a \in A_1 \subset D}, f_2(a)_{a \in A_2 \subset D}, \ldots f_n(a)_{a \in A_n \subset D})$

Definition 3:

A (M, n) *probabilistically-related plaintext* is a plaintext of length n having non-zero mutual information with M possible equal-length messages. Its rate is $\log_2 M/n$

$p = (p_1, p_2, \ldots p_n)$ a (M, n) PRP

$\Leftrightarrow \exists\ m = (m_1, m_2, \ldots m_k)$ such that $H(m|p) \neq H(m)$

(uncertainty in m decreases on knowing p)

Attacks on randomization - repeated plaintext

- An attack: asking the same question many times
- Can be thwarted by
  - never answering the same question twice, or
  - always answering it the same.
- Plaintext repetition:
  - corresponds to an error-correcting code word

    a a a a a a a

Error

Known that:

- tracker can reduce estimation error indefinitely
- by increasing the number of repeated plaintext bits indefinitely;

$$n \rightarrow \infty \Rightarrow \varepsilon^n \rightarrow 0$$

- and that this is the best he can do with repeated plaintext

$$\varepsilon^n \rightarrow 0 \Rightarrow \text{cost per message bit} = n/1 \rightarrow \infty$$

Is the following an attack?

- plaintext bit 1: “location = North”;
- plaintext bit 2: “virus X test = positive”;
- plaintext bit 3: “gender = male” AND “condition A = present”

If
(location = North) $\oplus$ (virus X test = positive)
$\Leftrightarrow$ (gender = male) AND (condition A = present)

Then: A3 = A1$\oplus$A2; check-sum bit

Not easily recognized as attack

DRP

Definition 4: A (M, n), $\log_2 M \leq n$ *deterministically-related plaintext* (DRP) is a plaintext of length n completely determined by the values of the corresponding message string from M possible equal-length message strings.

$p = (p_1, p_2, \ldots p_n)$ a (M, n) DRP
$\Leftrightarrow \exists\ m \in \mathcal{M}$, $m = (m_1, m_2, \ldots m_k)$ and Λ such that

$$p(\gamma) = \Lambda(m(\gamma))\ \forall \gamma \in \Gamma$$

$|\mathcal{M}| = M$

γ is an entity having property p, Γ the set of all entities

A DRP is a PRP

Attacks

Definition 5: An *(M, n) attack* for binary protocol Φ is:

- an (M, n) plaintext
- and an estimation map Ψ: $\Sigma^n \rightarrow \Sigma^k$ for
  - estimating the message m(γ) from
  - the ciphertext (randomized bits) Φ(p(γ)).

Its rate is $\log_2 M/n$.

Code

Definition 6: An *(M, n) code* for set of messages *M* and binary channel Φ is:

- A coding function f from *M* to code words of size n,

$$f: M \rightarrow \Sigma^n$$

- and a decoding function $g: \Sigma^n \rightarrow M$ for
  - estimating the message m from
  - the randomized bits $\Phi(f(m))$.

Its rate is $\log_2 M/n$.

Theorem 1

For a given set of message strings M, channel codes on set of messages M are DRP attacks and vice versa

- DRP attack is code:

A DRP attack on Φ consists of

  - estimation function Ψ and
  - DRP map Λ;

corresponds to code on channel Φ with

  - encoding function f = Λ,
  - decoding function g = Ψ

- Code is DRP attack:
  - encoding function Λ = f,
  - decoding function Ψ =g
  - f(m) is plaintext because f a function, and hence all bits of f(m) are functions of m too

Efficiency of attacks: repeated plaintext attack

- Definition 7: A *small error attack* is one in which $\varepsilon^n \rightarrow 0$ as $n \rightarrow \infty$

- Plaintext repetition:
  - corresponds to an error-correcting code word

    a a a a a a a
  - probability of error is monotonic decreasing with *n* for *n*-symbol code words
  - rate of code = 1/*n*
  - sacrifice rate for accuracy; rate of small error attack $\rightarrow 0$

- Are DRPs more efficient?

Reliable Attacks

Definition 8: A *reliable attack* of rate R is a small error attack of fixed rate R

Definition 9: A small error attack of asymptotic rate R_∞ is a small error attack with rate $\rightarrow R_\infty$

Do small error attacks of non-zero asymptotic rates exist?

Do reliable attacks exist?

Shannon (1948): Channel Coding Theorem

Codes exist for reliable transmission at all rates below capacity

A channel cannot transmit reliably at rates above capacity.

Theorem 2: *Existence of reliable DRP attacks*

Application of channel coding theorem requires a lemma,

- because channel coding defined for any set of messages,
- but DRP attacks only defined on equal-length messages

Lemma: If a sequence of $(2^{Rn}, n)$ codes with $\varepsilon^n \rightarrow 0$ exists, so does such a sequence on any sequence of messages $\{M_n\}$ of lengths $\{2^{Rn}\}$

Proof: Use the one-to-one correspondence between the messages, it preserves error and rate

Corollary: *Computability of reliable DRP attacks*

- Forney (1966): Existence of polynomial-time encodable and decodable Shannon codes
- Spielman (1995): Construction of linear-time encodable and decodable codes approaching Shannon codes

$\Rightarrow$ Corollary: Construction methods for linear time DRP attacks with k/n approaching C while $\varepsilon^n \to 0$

Converse of channel coding theorem and reliable DRP attacks

- Similarly, converse of channel coding theorem implies tight upper bound on rate of reliable DRP attacks

- But not enough: what about other attacks:
  - PRP attacks
  - small error attacks

Theorem 3

The asymptotic rate of a small error PRP attack is tightly bounded above by protocol capacity

$\log_2 M = nR_n = H(m_n) = H(m_n|\varphi(p_1),...\varphi(p_n)) + I(m_n; \varphi(p_1), \ldots\varphi(p_n))$

PRP attacks are not all channel codes, but
Fano's inequality holds even when p not a function of m:

$$H(m_n|\varphi(p_1),...\varphi(p_n)) \leq 1 + nR_n\varepsilon_n$$

Further, even when p not a function of m,

$$I(m_n; \varphi(p_1), \varphi(p_2), \ldots\varphi(p_n)) \leq \Sigma_i H(p_i) - \Sigma_i H(\varphi(p_i)|p_i) \leq nC$$

Hence,

$$nR_n \leq 1 + nR_n\varepsilon_n + nC$$

and,

$$\lim_{n \to\infty} R_n \leq \lim_{n \to\infty} (1/n + R_n\varepsilon_n + C)$$

$$\lim_{n \to\infty} \varepsilon^n = 0 \Rightarrow R_\infty \leq C$$

Theorem 4

The asymptotic length of plaintext per message for a stationary message sequence and a small error attack is tightly bound below by message entropy/protocol capacity

Use source-channel separation idea of source-channel coding theorem

- Bound can be achieved from above:

Given $\in$ and δ

- possible to find n such that n messages can be represented by at most

$$n(H(m) + \in C(\Phi)/2) \text{ bits}$$

With error at most $\delta/2$ (source coding theorem)

- Using a good code, possible to design a DRP attack with rate $C(\Phi) - \in C(\Phi)/2\ (H(m)/C(\Phi) + \in)$ and error at most $\delta/2$ (channel coding theorem)

$\Rightarrow \exists N$, s.t., $\forall n > N$, $\exists$ DRP with $\varepsilon^n < \delta$, plaintext length $\leq H(M)/C(\Phi) + \in$

Theorem 4

The asymptotic length of plaintext per message for a stationary message sequence and a small error attack is tightly bound below by message entropy/protocol capacity

- $H(\mathrm{M})/C(\Phi)$ is a lower bound:

Suppose possible that

given δ, $\exists$N, s.t., $\forall$n > N, $\exists$ PRP with:

- $\varepsilon^n < \delta$,
- plaintext length $\leq n(H(\mathrm{M})/C(\Phi) - \Delta) - \epsilon_n$, $\Delta > 0$

From Theorem 3, rate $< C(\Phi) + \epsilon_n$

$\Rightarrow$ Average message length $< H(\mathrm{M}) - \Delta C(\Phi) + \epsilon_n(H(\mathrm{M})/C(\Phi) - \Delta) - \epsilon_n(C(\Phi) + \epsilon_n)$

Violates source coding theorem

Proposed measure of privacy of randomization

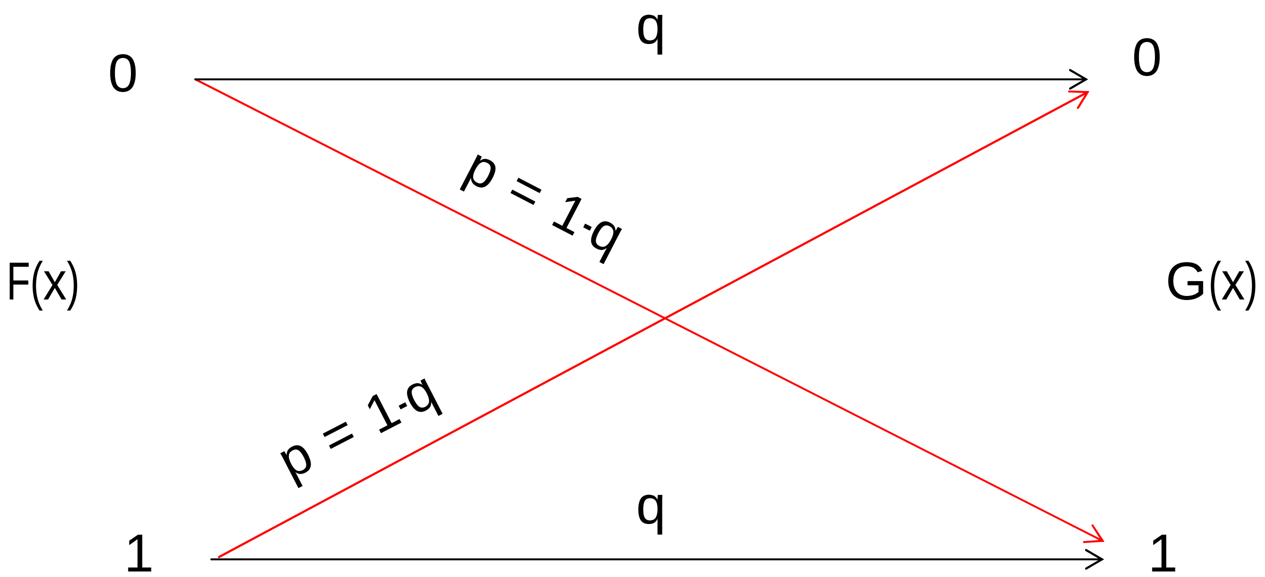
The privacy of randomization is the tight lower bound on the asymptotic length of plaintext per message, per bit of message entropy, for a stationary message sequence and a small error attack

Corollary: Privacy (Φ) = $1/\mathcal{C}(\Phi)$

Channel capacity is the appropriate protocol measure

- independent of input pdf
- provides weighting for different probability distributions
- is also security measure
- connects to a measure used in data mining (mutual information, Agrawal and Aggarwal, 2001)

Application: Binary Symmetric Protocol



$C = 1 + p\log_2 p + (1-p)\log_2(1-p)$

$= 0$ if $p = 0.5$;

$\approx 4\beta^2/\ln 2$ if $p = 0.5 \pm \beta$; $\beta << 0$

Application to binary randomization

Binary symmetric protocols for small bias β have channel capacity $O(\beta^2)$.

Corollary: Plaintext length required, per bit of message entropy, for a small error attack in the binary randomization protocol with small bias β is $O(1/\beta^2)$ and *independent of the error*

The privacy of binary randomization with small bias β is *$O(1/\beta^2)$*

We have shown that

Dishonest Bob can do better by increasing the number of points combined in a single query

i.e. there exist attacks for which

$$n \rightarrow \infty \Rightarrow \varepsilon_n \rightarrow 0$$

$$\varepsilon_n \rightarrow 0 \not\Rightarrow n/k \rightarrow \infty$$

There is a tight lower bound on the limit of n/k such that

$$n \rightarrow \infty \Rightarrow \varepsilon_n \rightarrow 0$$

i.e, $(n \rightarrow \infty \Rightarrow \varepsilon_n \rightarrow 0) \Rightarrow \lim n/k > 1/C$

Unlike other work

- Our bound is independent of error, i.e. there is a *finite* number of plaintext bits required per message bit for *arbitrarily small* error
- We connect security theory to statistical techniques for privacy protection
- We use Shannon's channel coding theorem to design exceptionally powerful attacks, and to bound their efficiency
  - (only the source coding theorem has been used so far for cryptography and for anonymous delivery)

The variable privacy big picture

- Alice can use randomization as a privacy protocol
  - designing the channel capacity
  - based on knowledge that error correcting codes are attacks
- Dishonest Bob cannot approach rates higher than channel capacity
- Randomization is a *game* between Alice and Bob
- In this world, *maximum privacy exists when Alice gets maximum benefit for a piece of revealed information*

Further questions, variable privacy

- What are best strategies for the user in different conditions in a variable privacy scenario?
- Are there structures that are protected, and structures that are revealed, with various randomization protocols?

Acknowledgements

Umesh Vazirani, UC Berkeley

- for the original suggestion to use randomization for privacy protection and economic valuation of privacy
- for spirited discussions
- for an observation leading to the definition of DRP attacks

Gadiel Seroussi, HPLabs.

Cormac Herley, Microsoft Research