

Bounding Bit Leakage

Peter Y A Ryan

Joint work with Arash Atashpendar and Bill Roscoe

APSIA: Applied Security and Information Assurance
University of Luxembourg

Maths/CS Lunch Seminar, 2014

- Information theory puzzle that arose in the context of Quantum Key Establishment protocols
- Typically random sampling of fresh session keys is used to estimate the noise/eavesdropping over the quantum channel.
- Problem arises from a modified scheme for error estimation proposed in [Ryan, 2013]
 - Sample set computed secretly by Anne and Bob based on the prior shared key
 - Compare the bits over an un-encrypted channel
 - We want to bound the information leakage for privacy amplification

Problem Statement

A bit string y of length n chosen at random from the space of all possible bit strings length n .

- Assume flat probability distribution over the n -bit strings
- From this, a subset S of $\{1, \dots, n\}$ of size k ($k \leq n$) chosen at random corresponding bits revealed at these indices in the long string
- Assume a flat distribution over the set of subsets of $\{1, \dots, n\}$ of size k for the resulting string x , thus every subset is equally probable.

Example, suppose that for $n = 12$ and $k = 4$ we have:

$$y = \langle 011000011001 \rangle$$

and we choose $S = \{2, 4, 5, 8\}$, then $x = 1001$.

Problem Statement

Cont'd

- If we reveal n and x but not S . how much information are we revealing about y ?
- Here we assume that the appropriate measure is defined in terms of the decrease in the Shannon entropy of the space of possible y strings given the observation of x .
- But other measures might be appropriate in some contexts (e.g. cryptographic), e.g. Renyi entropy, guessing entropy, min entropy etc. and indeed such measures may prove more tractable. For the purposes of this talk however we stick to Shannon.

Problem Statement

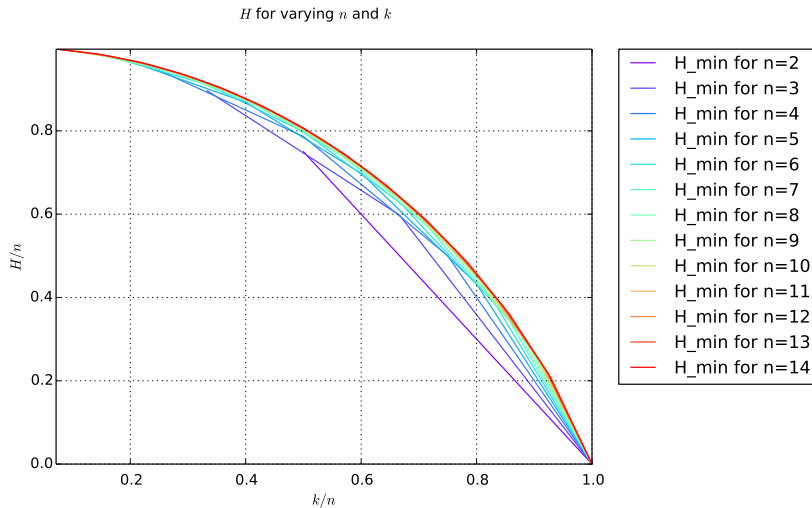
Cont'd

To illustrate:

- if you reveal 0 bits then obviously you reveal nothing about the full string.
- If you reveal just one bit ($k = 1$) and suppose that it is a 0, then essentially all you have revealed about the full string is that the all 1 string is not possible.
- At the other extreme, if you reveal all the bits ($k = n$) then obviously you reveal all n bits of the original string. For $k = n/2$ the leakage is approximately one bit. So intuitively the function starts off very shallow but then rises very fast as k approaches n .

Min Entropy plot for n and k

Minimum entropy values for varying n and k



Define $x = \text{Mask}(y, S)$ to mean that the string y filtered by the mask S gives the string x . Now we define the *compatible set*: given an x and n this is the set of y strings of length n that could project to x for some mask S .

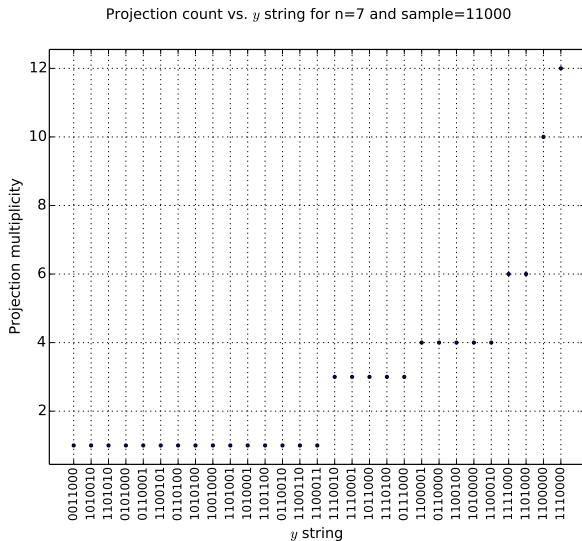
$$\Upsilon_{n,x} := \{y \mid |y| = n, \exists S \bullet \text{Mask}(y, S) = x\}$$

- Let $\pi_x(y)$ denote the number of distinct ways that y can project onto x .
- and $\mu_{n,x}$ the number of configurations for n and x , i.e. the number of pairs $\{y, S\}$ such that $\text{Mask}(y, S) = x$. It is easy to see that this is given:

$$\mu_{n,k} = \binom{n}{k} \cdot 2^{n-k}$$

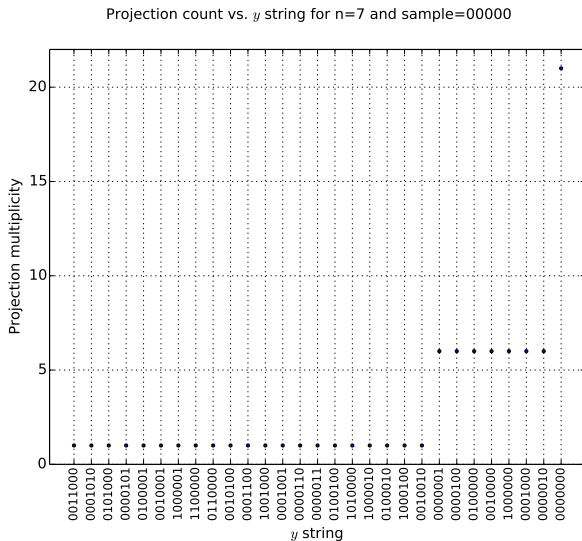
Projection Plot

Projection multiplicity vs. y string for $n = 7$ and $x = 11000$



Projection Plot

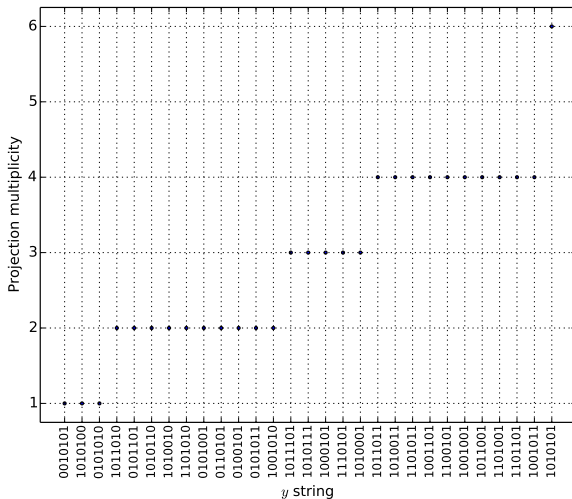
Projection multiplicity vs. y string for $n = 7$ and $x = 00000$



Projection Plot

Projection multiplicity vs. y string for $n = 7$ and $x = 10101$

Projection count vs. y string for $n=7$ and sample=10101



Theorem

Given n and x , the cardinality of $\Upsilon_{n,x}$ is purely a function of n and k , where $k = \text{length}(x)$, and independent of the exact x . Furthermore, the cardinality is given by:

$$\gamma_{n,k} := |\Upsilon_{n,x}| = \sum_{r=k}^n \binom{n}{r} \quad (1)$$

Cardinality of the Compatible Set

Proof

Proof:

$\gamma_{n,k}$ satisfies the following recursion:

$$\gamma_{n,k} = \gamma_{n-1,k} + \gamma_{n-1,k-1}$$

with $\gamma_{n,n} = 1$ and $\gamma_{n,0} = 2^n$.

- Partition the y strings into those that have a mask overlapping the first bit of y and those that don't.
- For the former we can enumerate them simply as the number of y strings of length $n - 1$ with ≥ 1 projections to the tail of x , i.e. x^* , i.e. $\gamma_{n-1,k-1}$.
- For the latter the number is just that of the set of $n - 1$ strings with ≥ 1 projection to x , which has length k , i.e. $\gamma_{n-1,k}$.

Observation

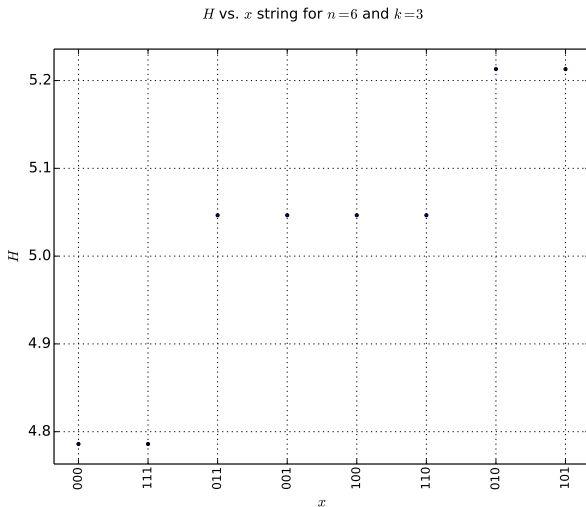
Maximum leakage

Now that we have established that γ is independent of the exact x string:

- We can just consider $x = 0^k$ (or $x = 1^k$). Now the formula is obvious: γ is just the sum of the number of y strings with n 0s, with $n - 1$ 0s etc down to k 0s.
- But this does yet give us the entropy: the pdf over Υ is not flat, in fact $\pi_x(y)$ can vary from 1 to $\binom{n}{k}$. The probability that $Y = y$ given the observation of a x is proportional to $\pi_x(y)$.
- We observe that the maximum leakage (minimum entropy of $\Upsilon_{n,x}$) is attained by $x = 0^k$ (or 1^k). The minimum leakage occurs with $x = 01010\dots$ or $x = 101010\dots$

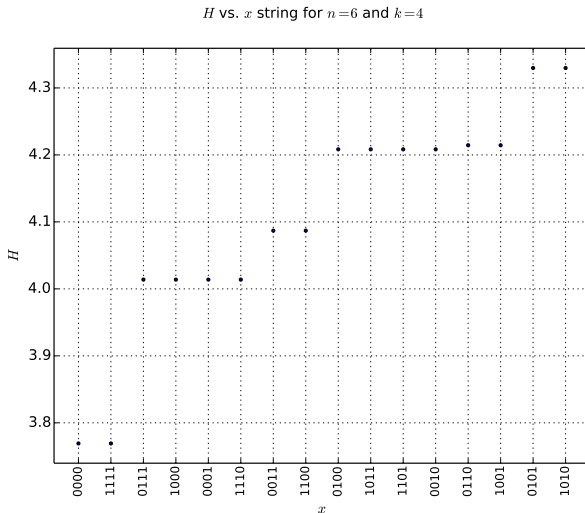
Entropy Plot

H vs. x for $n=6$ and $k=3$



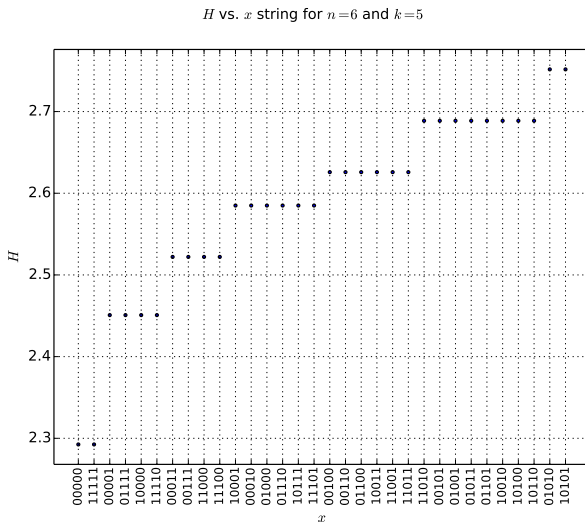
Entropy Plot

H vs. x for $n=6$ and $k=4$



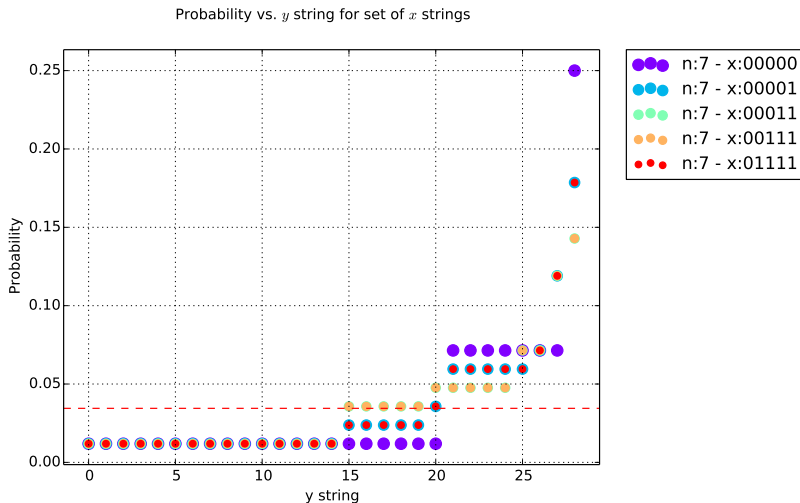
Entropy Plot

H vs. x for $n=6$ and $k=5$



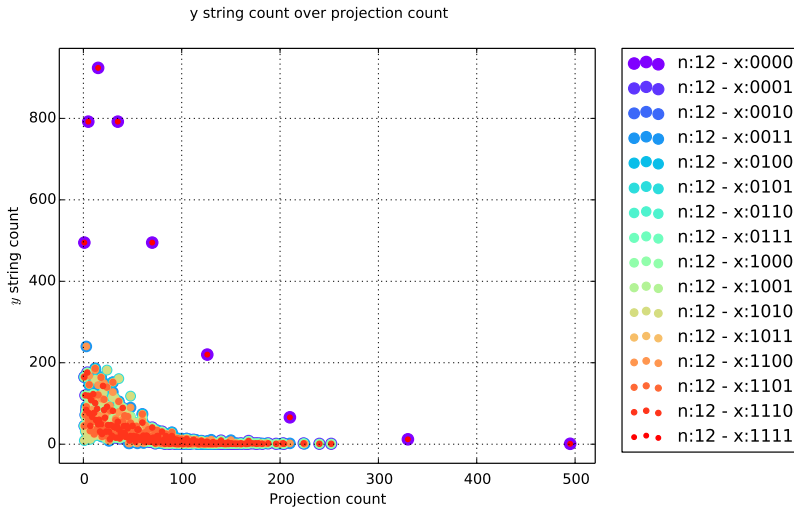
Probability Plot

Probability vs. y string



γ count vs. π Plot

$n = 12$



Minimum Entropy

Proof

Theorem

The minimum entropy is attained by $x = 0^k$ ($x = 1^k$).

Proof (sketch):

After some algebraic manipulation we can show:

$$H_{n,x} = A_{n,k}H_{n-1,x} + B_{n,k}H_{n-1,x^*} + C_{n,k}$$

Where $x^* := \text{Tail}(x)$, i.e. $\exists b \in \{0, 1\}$ such that $x = b \circ x^*$. Note: A, B and C are functions purely of n and k , and so this gives us the basis for an induction over n .

Minimum Entropy

Proof, cont'd

Base case: need to show:

$$H_{n,x} = 0, \text{ where } \textit{length}(x) = n$$

It is easy to show that

$$H_{3,<00>} \leq H_{3,<01>}$$

Theorem

The minimum entropy for given n and k is given by:

$$H_{n,k} = - \sum_{j=0}^{n-k} \binom{n}{j} \times \frac{\binom{n-j}{k}}{\mu_{n,k}} \times \log\left(\frac{\binom{n-j}{k}}{\mu_{n,k}}\right) \quad (2)$$

Proof: straightforward from the fact that with min corresponds to $x = 0^k (1^k)$.

Open questions:

- Prove max entropy attained by $\langle 010101\dots \rangle$ or $\langle 101010\dots \rangle$.
- Derive an expression for the expected leakage.
- Derive expressions for the asymptotics of the leakages.
- Get better intuitions into the structure of the Υ spaces.
- Investigate other measures of information.
- Better understand how the structure of the x strings influences the entropy.

References



Peter Y A Ryan and Bruce Christianson (2013)

Enhancements to Prepare-and-Measure Based QKD Protocols

Security Protocols XXI 123–133 Springer.



Arash Atashpendar and Peter Y A Ryan (2014)

QKD and Information Leakage Simulator

<http://www.qkdsimulator.com/>

Questions...