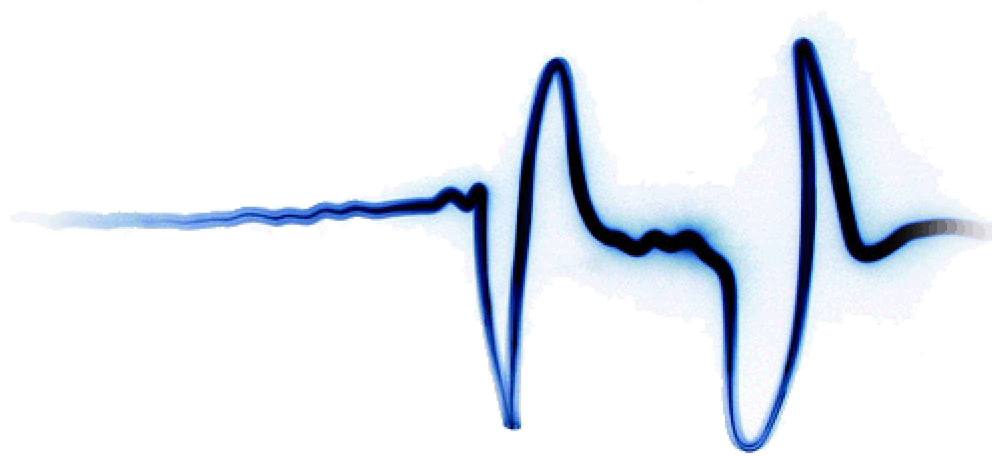
Privacy Preserving (ECG) Signal Quality Evaluation
(Extended version for VIPP meeting)

Riccardo Lazzeretti, Jorge Guajardo, Mauro Barni

R. Lazzeretti J. Guajardo, M. Barni *Privacy Preserving ECG Quality Evaluation*
In 14th ACM Workshop on Multimedia and Security, MM&SEC, 2012

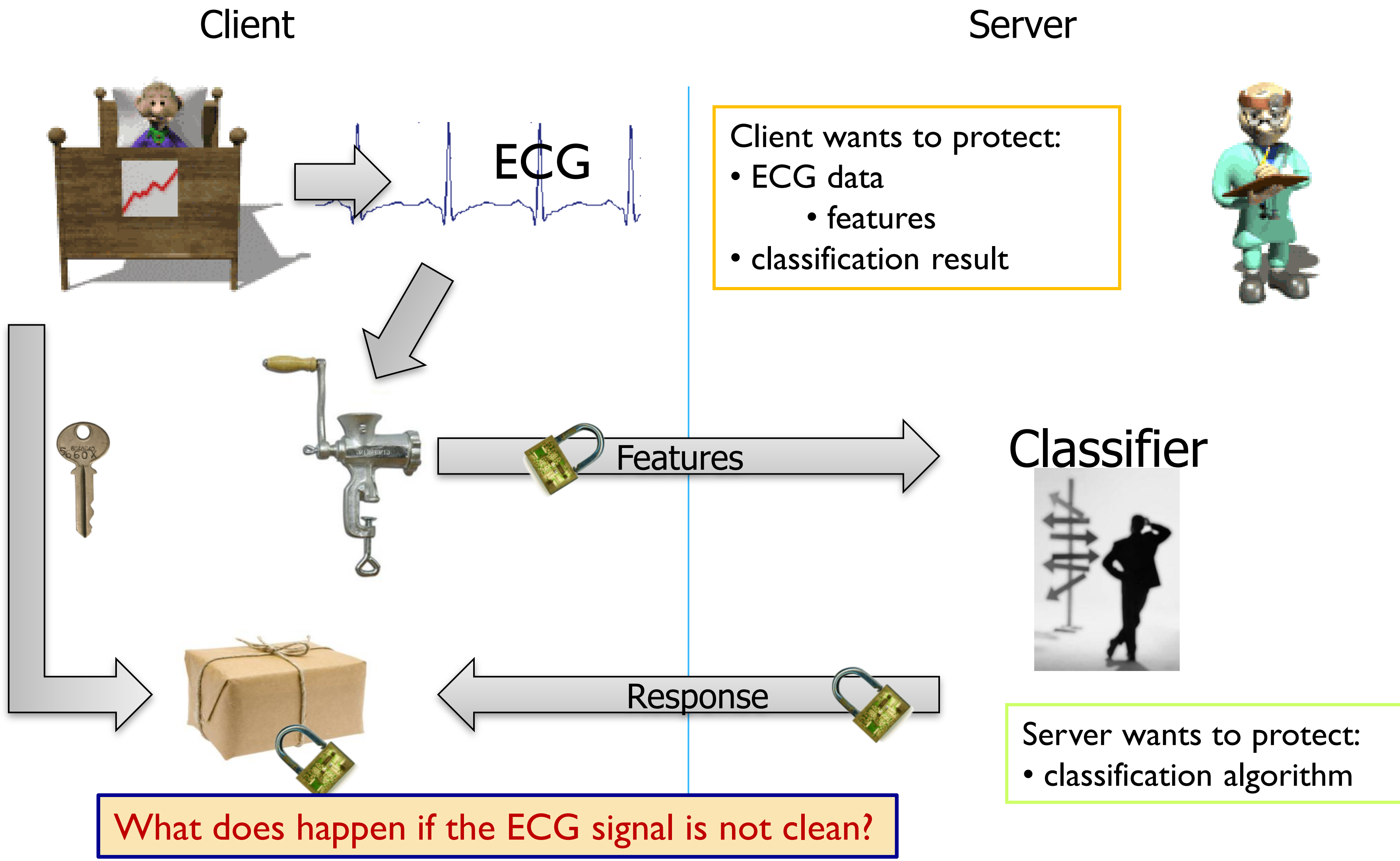
Outline

- Introduction to the problem
- Cryptographic primitives
- Proposed protocol
- Complexity
- Accuracy
- Conclusions



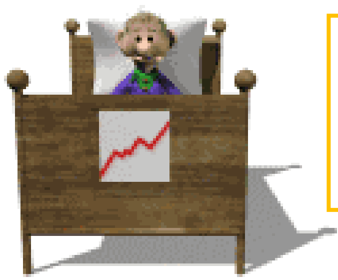
Prior art: remote ECG classification

[BFL+11]M. Barni, P. Failla, R. Lazzeretti, A. Sadeghi, and T. Schneider. Privacy-preserving ecg classication with branching programs and neural networks. IEEE Transactions on Information Forensics and Security, 2011.



Current research: Quality Evaluation

Client



- Prevent the server from obtaining the data of the patient

Server

- Keep the details of the quality evaluation algorithm secret
- Change algorithm depending on the analysis
- Application-based quality definition

Connect the electrods

Record the signal

Encrypted Quality Evaluation

Is the quality sufficient?

No

Yes

Data Analysis

So we propose …

- Privacy-preserving protocol for signal quality evaluation
  - Easy to be implemented in the encrypted domain
  - Preserves the privacy of the patient
  - Protects the server parameters
- Difficult problem even in the plain domain
  - ECG signal can be affected by
    - Power line interference
    - Baseline wander
    - Muscle movement
    - **Electrode contact noise**
  - No reference available

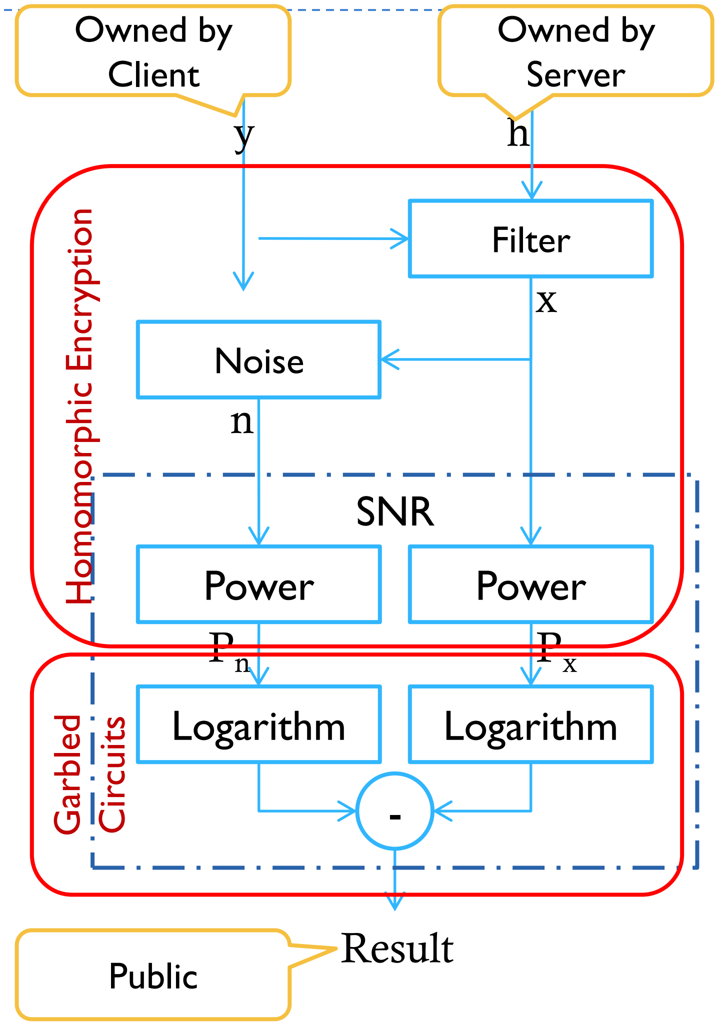
Cryptographic primitives

	Homomorphic Encryption $[\![a+b]\!] = [\![a]\!][\![b]\!]$ $[\![c \cdot a]\!] = [\![a]\!]^c$	**Garbled Circuit with Free-Xor**
Permits to compute:	•Linear operations (no interaction) •Products and square values (interaction)	Any function that can be represented by a boolean circuit (interaction)
Encryption scheme:	Asymmetric (Paillier)	Symmetric (xor with hash function)
Data representation:	Encryption of integer numbers (1024 bits)	Encryption of each bit (80 bits)
Dependence on data size:	Quite independent	Highly related
Computation complexity:	High	Small
Communication complexity:	Small (only if interaction is required)	High
Suitable for:	Sums, products, square values, filtering, linear transformations	Linear operations having data represented with few bits, any function that can not be computed by Homomorphic Encryption

Hybrid Protocol

Previous work [BGL10]

- Noisiness of ECG signal
  - fc=20Hz
  - Desired characteristics of the linear filter:
    - Integer coefficients
    - Small number of coefficients
    - Small number of bits used to represent the coefficients
  - The filter is considered private property of the server
- Estimation of noise as difference between the recorded and the filtered signal
- **SNR measured in this way -> basis for quality evaluation**



[BGL10] M. Barni, J. Guajardo, and R. Lazzeretti. Privacy preserving evaluation of signal quality with application to ECG analysis. In IEEE International Workshop on Information Forensics and Security (WIFS), 2010.

Working in the Encrypted Domain

Plain Domain		Encrypted Domain
Algorithm works on floating point numbers (REALS)	Quantization →	Algorithm works on integer numbers (INTEGERS)

- Original ECG data coming from MIT-BIH database
  - 10 bits for the magnitude and 1 for the sign
- Necessity to design a good filter with
  - minimum number of integer coefficients
  - coefficients represented with the minimum number of bits
- The filtered integer signal will be amplified by a factor k
- Bitsize of filtered signal and values obtained during processing can be obtained under worst-case analysis

- Security under semi-honest adversaries in the standard model

Precomputation by HE

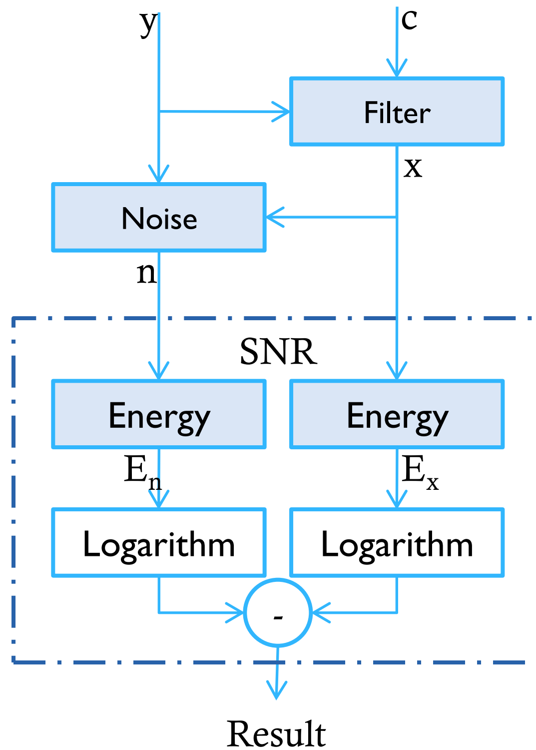
- Signal Filtering and Noise Computation

$$[\![x_i]\!] = [\![c_0 y_i + \sum_{j=1}^{order} c_j (y_{i-j} + y_{i+j})]\!]$$

$$[\![n_i]\!] = [\![x_i - k\ y_i]\!]$$

- Energy evaluated instead of Power

$$E_x = \sum_i x_i^2 \qquad E_n = \sum_i n_i^2$$



client

server

Obfuscation

$$[\![x_i^{ob}]\!] = [\![x_i + r_i^x]\!]$$

$$[\![x_i^{ob}]\!]\ [\![\sum_i (2 x_i r_i^x)]\!]$$

Energy computation

$$E_x^{ob} = \sum_i (x_i^{ob})^2 =$$

$$\sum_i x_i^2 + \cancel{\sum_i (2 x_i r_i^x)} + \sum_i (r_i^x)^2$$

Can be computed by server and transmitted, after obfuscation, to the client, to be removed

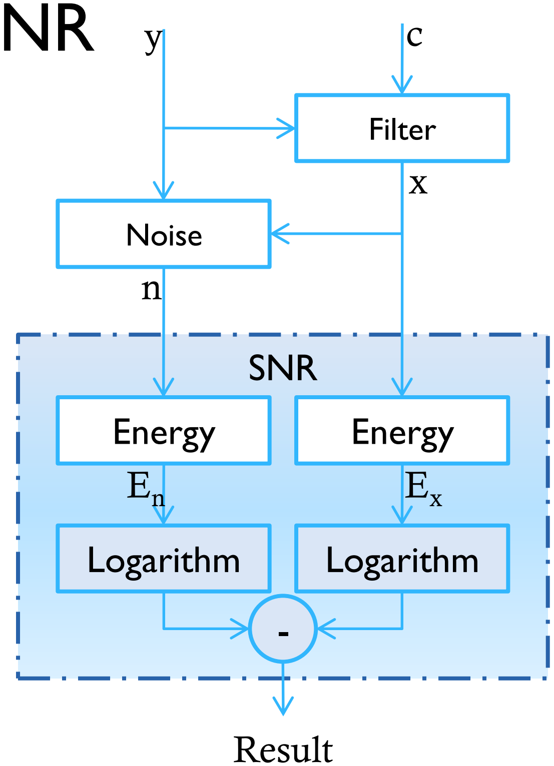
SNR Computation by GC

- The client evaluates a GC to obtain the SNR

$$SNR = 10\log_{10}\frac{P_x}{P_n} = \frac{10}{\log_2 10}\log_2\frac{E_x}{E_n}$$

It is an amplification factor that can be omitted

$$SNR = \log_2\frac{E_x}{E_n} = \log_2 E_x - \log_2 E_n$$



- Client inputs obfuscated energy to GC
- Server inputs total obfuscation to GC
- GC removes obfuscation, compute logarithm and subtraction

Logarithm Computation by GC

- Integer $\log_2$ evaluation $b = \begin{cases} \lfloor \log_2 a \rfloor + 1 & \text{if } a > 0 \\ 0 & \text{if } a = 0 \end{cases}$
  - minimum number of binary digits necessary to represent the number
- Being ℓ the number of bits used to represent a we apply the following protocol:

$$b_\ell = a_\ell$$
$$\text{for } i = \ell - 1 \text{ downto } 1$$
$$b_i = b_{i+1} \vee a_i$$
$$\text{endfor}$$
$$\log_2(a) = \sum_{i=1}^{\ell} b_i$$

$\ell - 1$ non-XOR gates

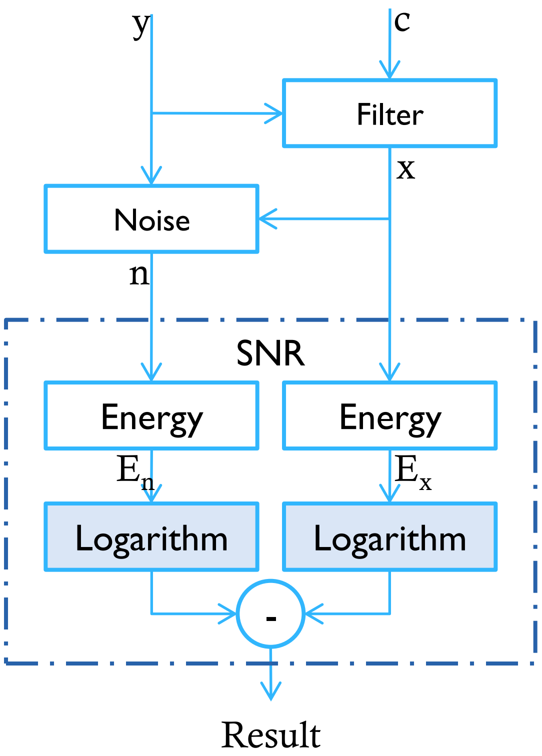
COUNTER circuit

Example:

$a = 00001011$

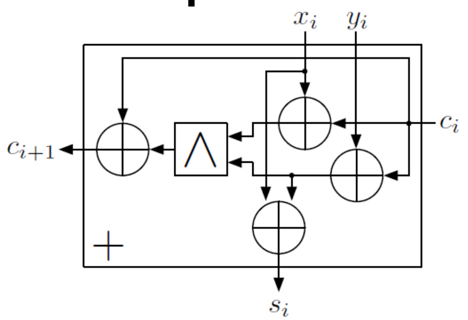
$b = 00001111$

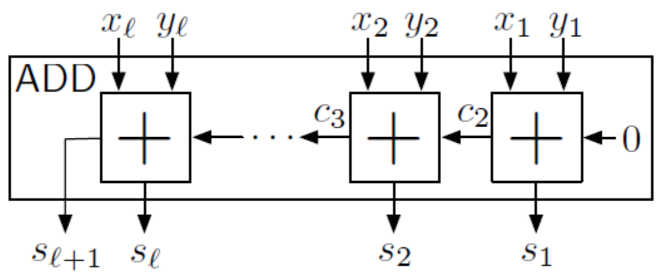
$log_2(a) = \log_2(b) = 4$



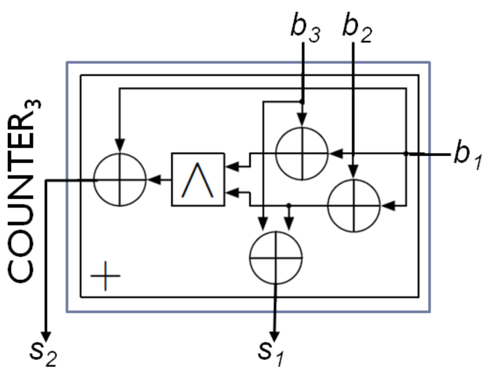
COUNTER Circuit

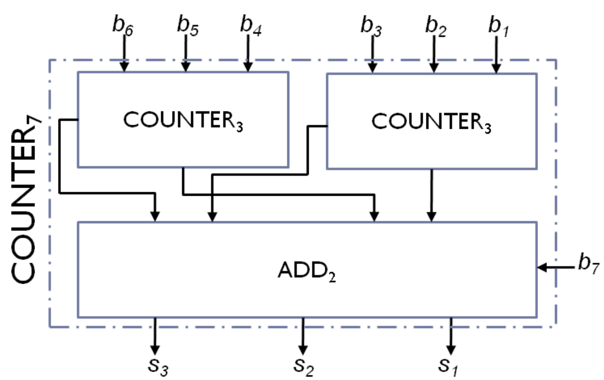
- Developed recursively by using adders blocks

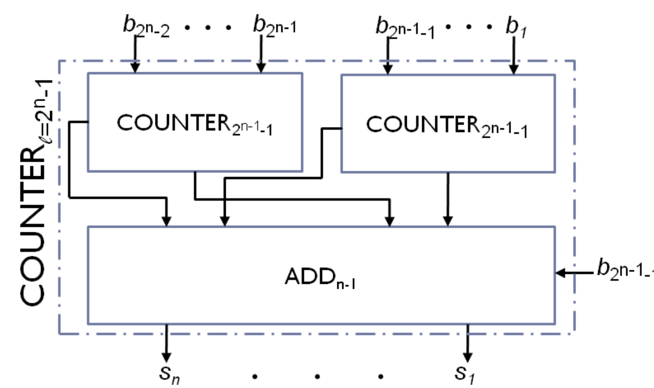




- COUNTER:



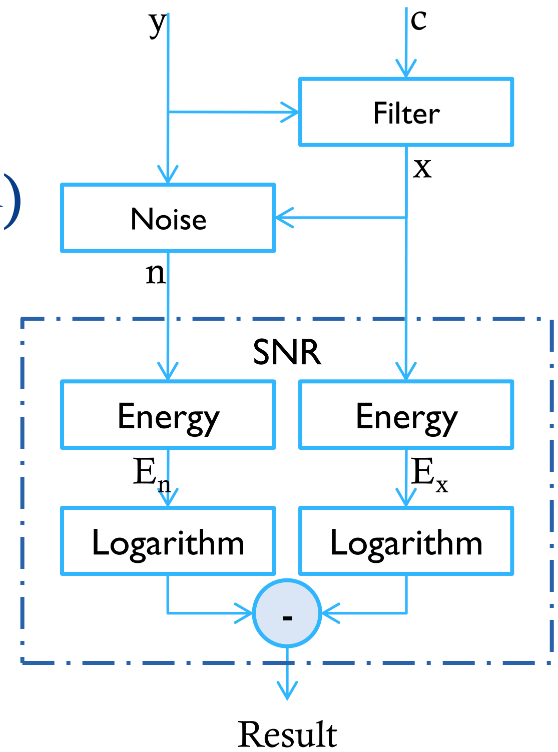




- $\mathrm{COUNTER}_k$ can be developed by optimizing $\mathrm{COUNTER}_{\ell=2^n-1}$

Obtaining the SNR

- Final result obtained by using Subtractor Circuit
  - Result=COUNTER(b^{E_x}) - COUNTER(b^{E_n})
- Optimization:
  - |Result|=COUNTER($b^{E_x} \oplus b^{E_n}$)
  - sign(Result)=$b^{E_x} < b^{E_n}$



- Example:

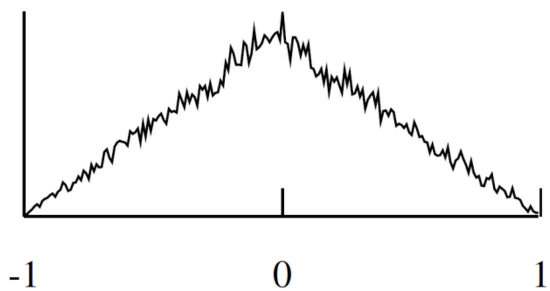
$$
\begin{array}{rr}
b^{E_x} = 00011111 & \log_2 E_x = 5 \\
b^{E_n} = 00000011 & \log_2 E_n = 2 \\
\hline
b^{E_x} \oplus b^{E_n} = 00011100 & \log_2 E_x - \log_2 E_n = 3
\end{array}
$$

Error Analysis

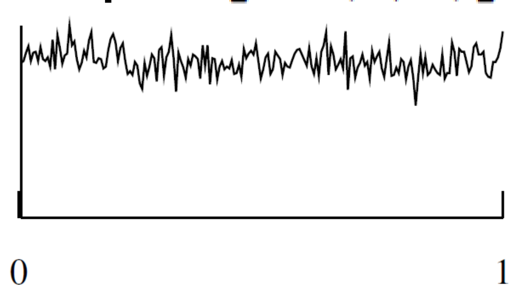
- Considering $a, b \in \mathbb{N}$

$$\begin{aligned}\epsilon_{tot} &= |\log(a/b) - \lfloor \log_2 a \rfloor + \lfloor \log_2 b \rfloor| \\ &= |\log(a/b) - \log_2 a + \epsilon_a + \log_2 b - \epsilon_b| = |\epsilon_a - \epsilon_b| < 1\end{aligned}$$

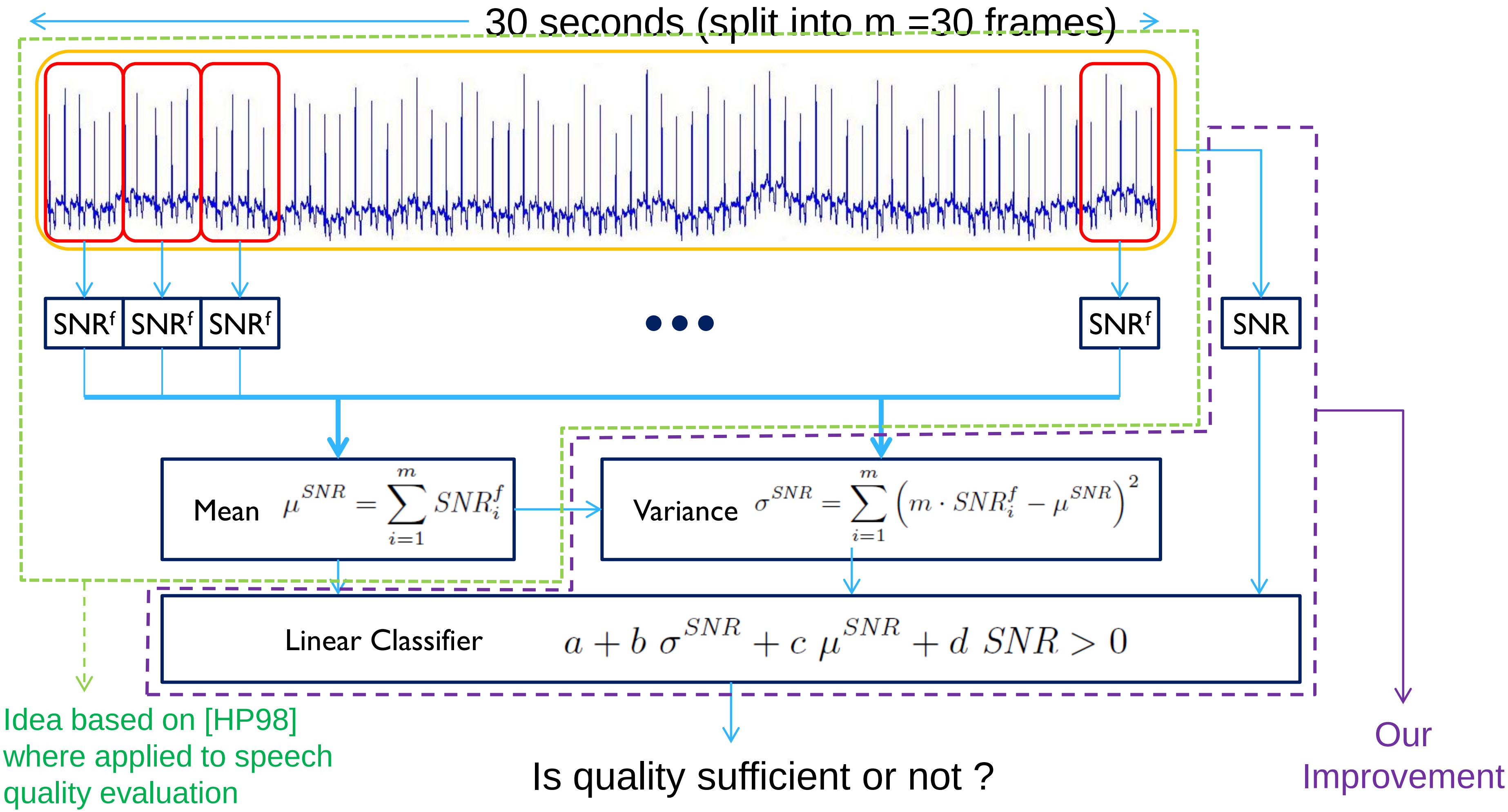
- Error Histogram obtained from practical tests:



- What if we could compute $\lfloor \log(a/b) \rfloor$



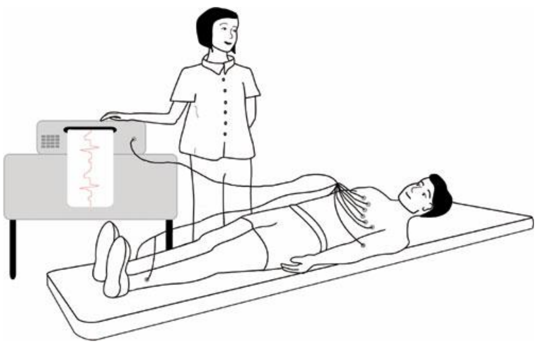
Quality evaluation



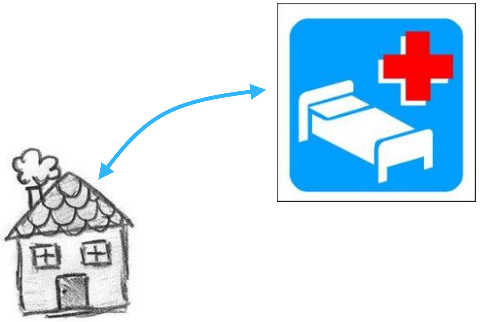
[HP98] J. Hansen and B. Pellom. An effective quality evaluation protocol for speech enhancement algorithms. In Fifth International Conference on Spoken Language Processing. Citeseer, 1998.

Practical use

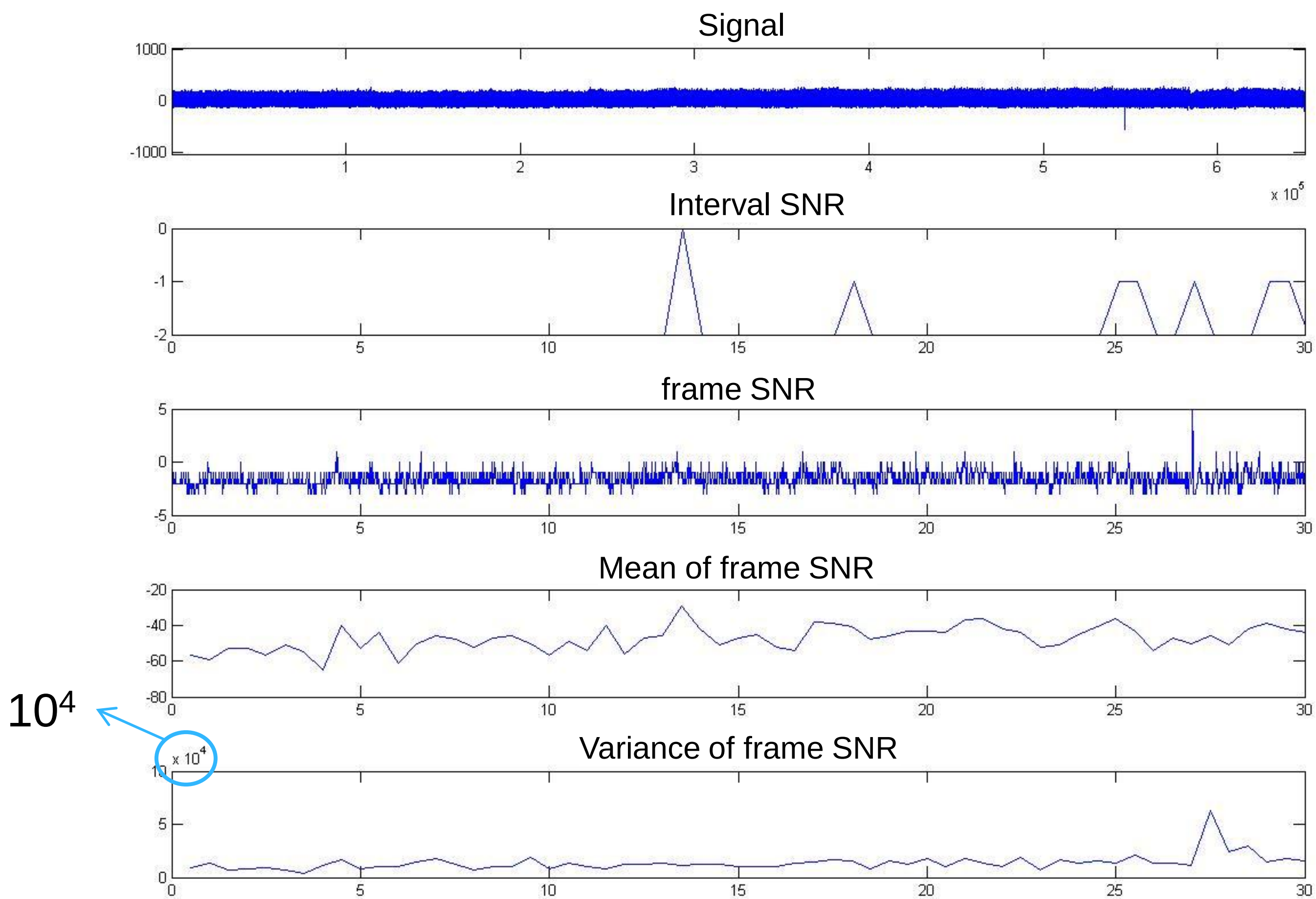
- Training phase
  - An expert assists a non-expert user in recording clean and noisy signals
  - Clean signal and noise signals are separed
  - Classifier trained
  - Classification parameters are a property of the service provider
    - This prevents that the patient uses them with other products



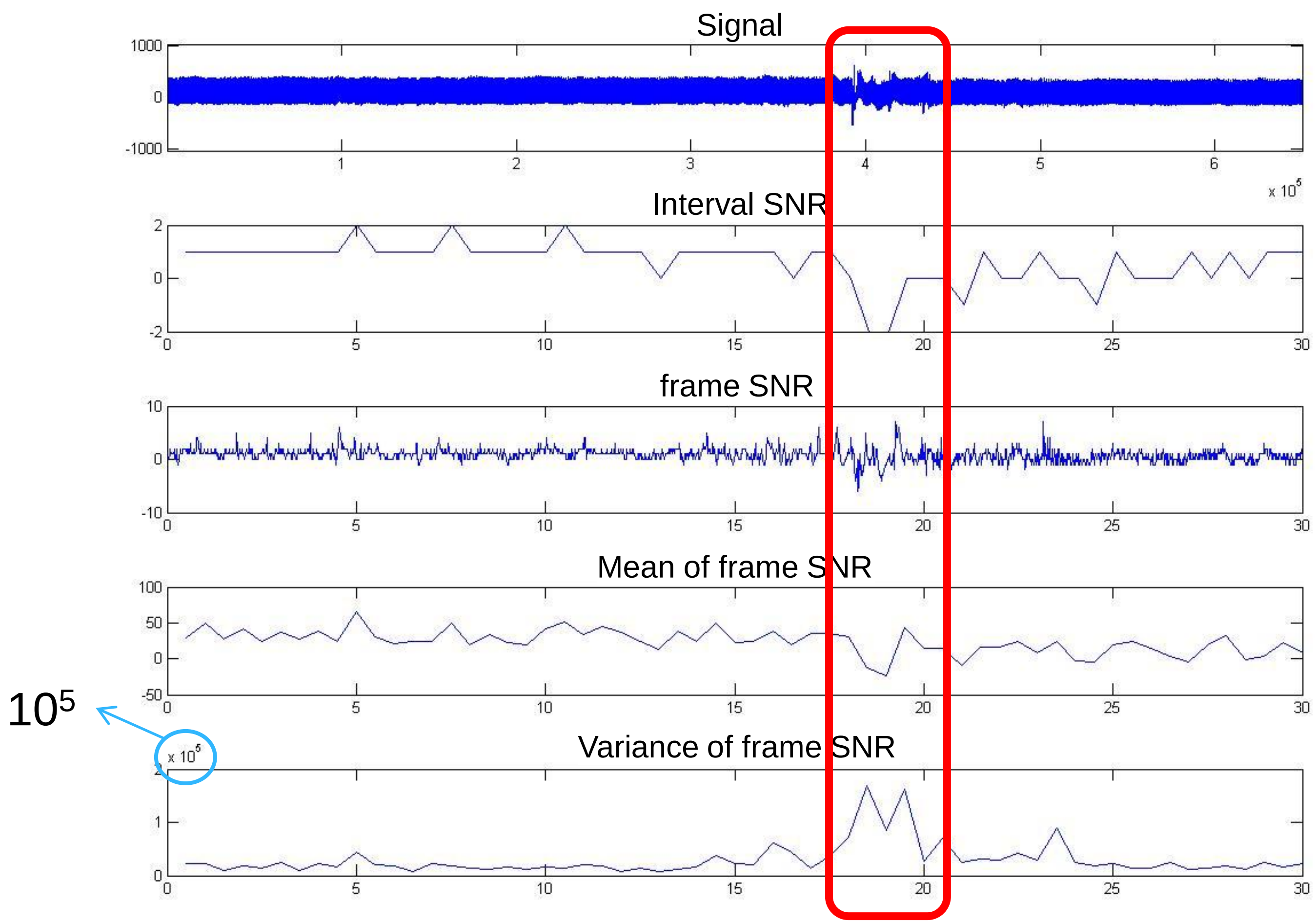
- Telecare analysis:
  - The patient applies the electrodes at home and runs the secure protocol



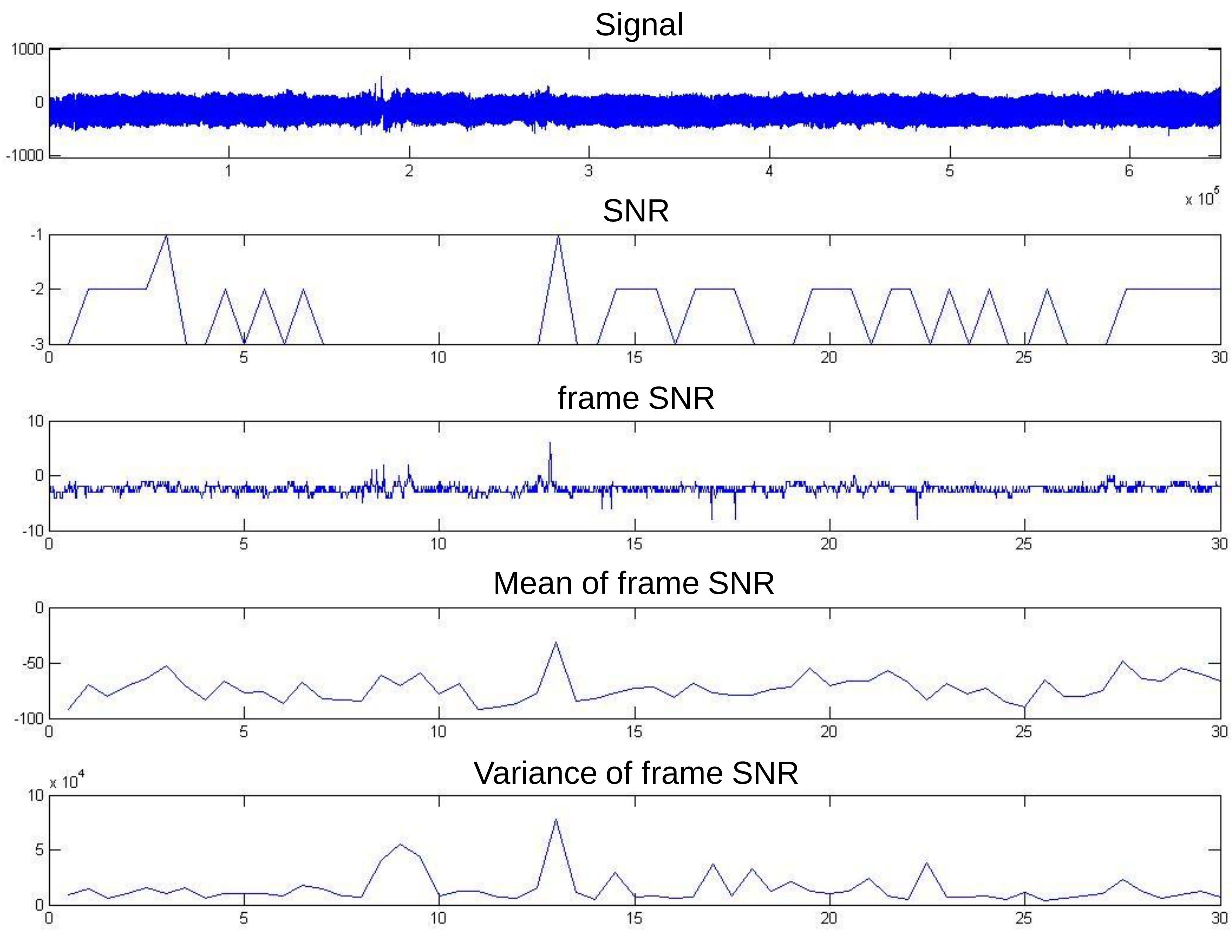
Clean Signal



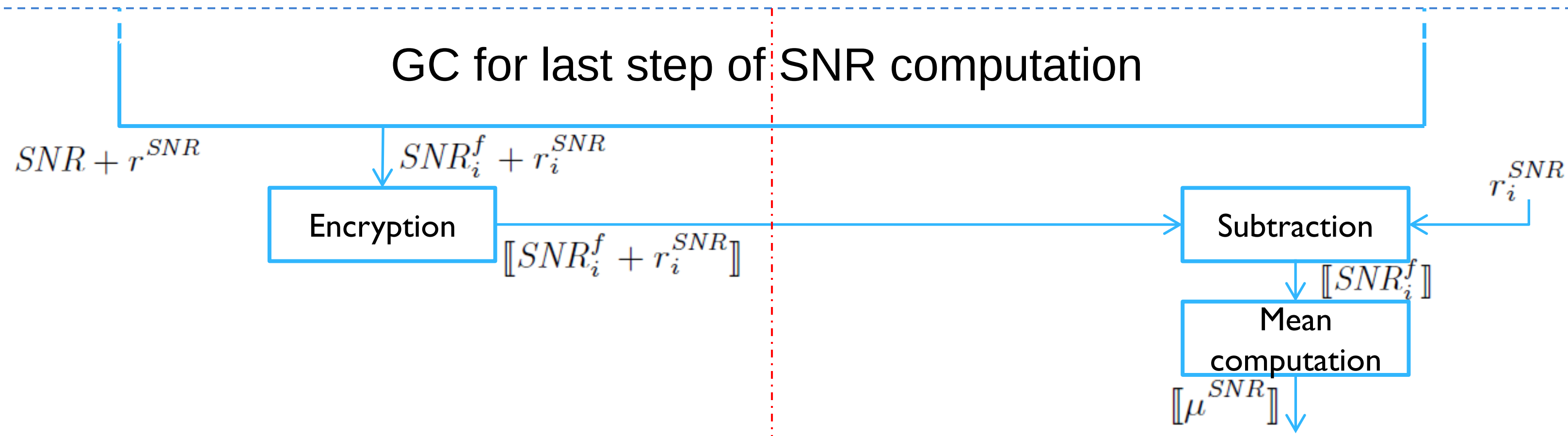
Signal with noise (ex. 1)



Signal with noise (ex. 2)



Quality classification protocol (HE)

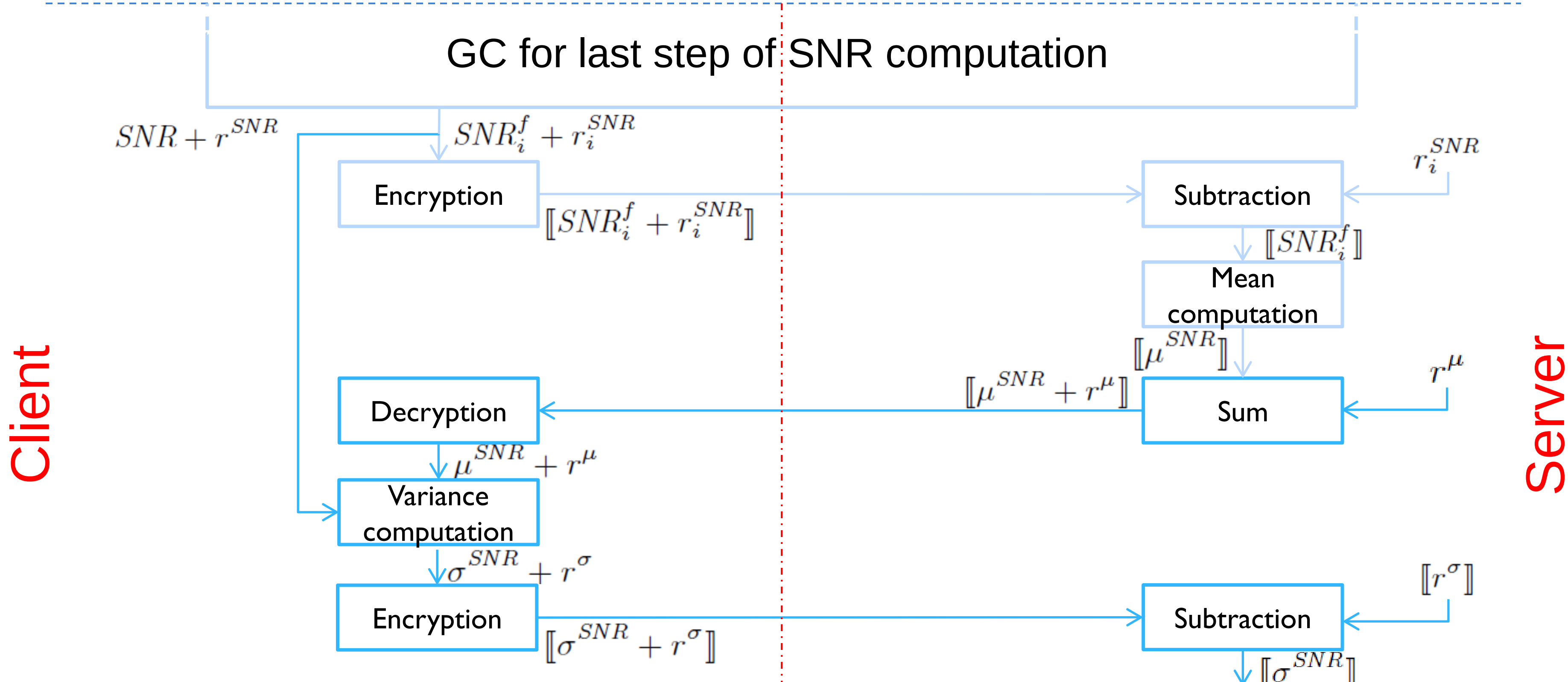


Client

$$[\![\mu^{SNR}]\!] = [\![\sum_i SNR_i^f]\!] = \prod_i [\![SNR_i^f]\!]$$

Server

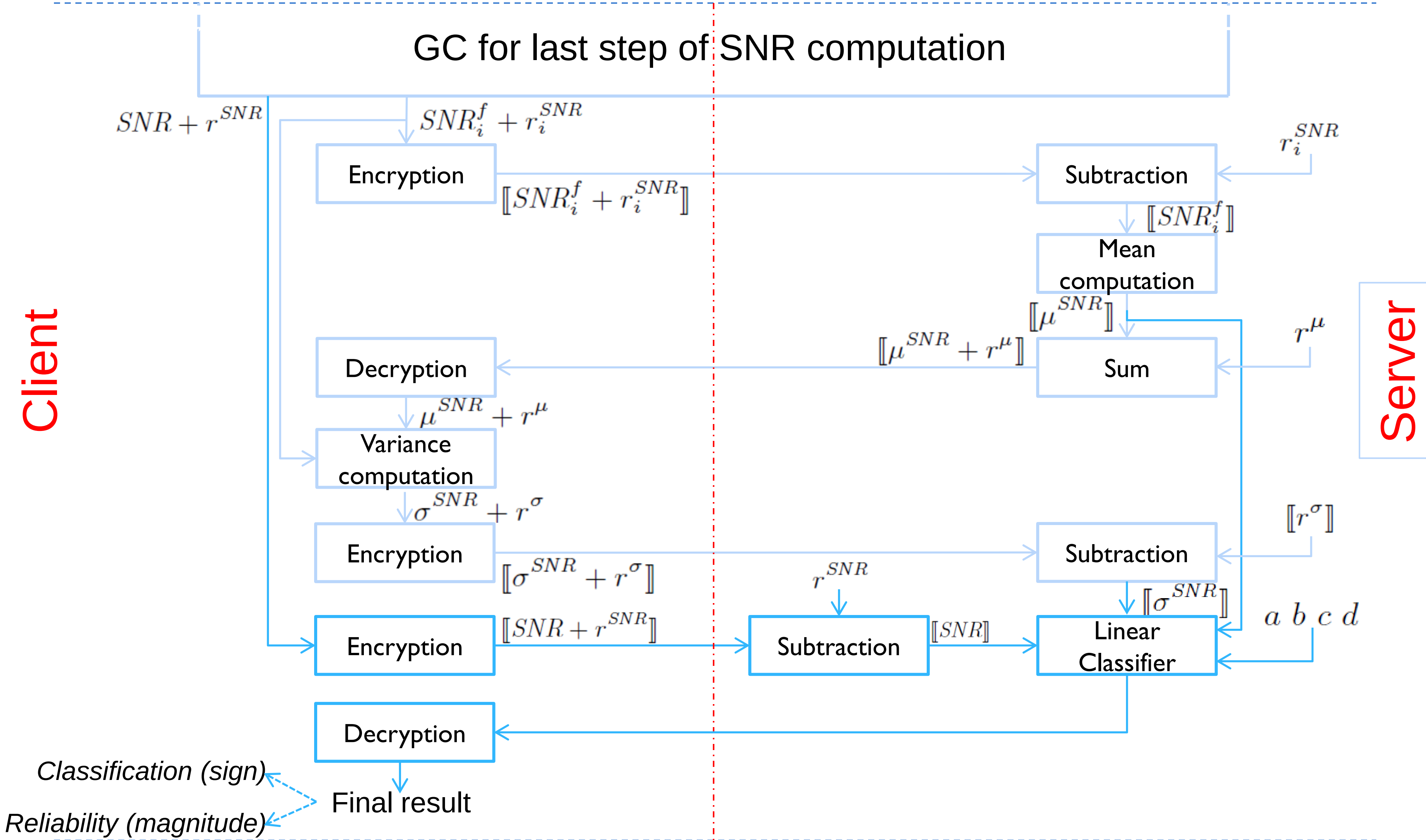
Quality classification protocol (HE)



$$\sigma^{SNR} + r^{\sigma} = \sum_i \left(m(SNR_i^f + r_{SNR_i}) - (\mu^{SNR} + r_{\mu}) \right)^2$$

$$[\![r^{\sigma}]\!] = [\![\sum_i (mr_{SNR_i} - r_{\mu})^2 - 2\sum_i (mSNR_i^f - \mu^{SNR})(mr_{SNR_i} - r_{\mu})]\!]$$
$$= [\![\sum_i (mr_{SNR_i} - r_{\mu})^2]\!] \prod_i ([\![SNR_i^f]\!][\![\mu^{SNR}]\!]^{-1})^{-2(mr_{SNR_i} - r_{\mu})}$$

Quality classification protocol (HE)



Complexity

- Worst case analysis:

Variable	Maximum value	Magnitude bitlength
Original sample	1,023	10
Filter coefficient	8	4
Filtered sample	114,576	17
Noise sample	180,048	18
Frame Energy	$\sim 1.17 \cdot 10^{13}$	44
Frame SNR	44	6
Signal Energy	$\sim 3.50 \cdot 10^{14}$	49
Signal SNR	49	6
SNR mean*	1,320	11
SNR variance*	52,272,000	26

* To avoid division:
- mean amplified by m
- variance amplified by m^3

- Communication (bits):

	Offline	Online
HE	0	26,626,048
Circuit	3,402,240	0
Client secret transmission	914,560	438,080
Server secret transmission	432,320	0
Total	**4,749,120**	**27,064,128**

	Offline	Online
Frame SNRs	4,584,000	26,980,864
SNR	165,120	15,680
SNR mean	0	61,440
SNR variance	0	4,096
Linear classifier	0	2,048
Total	**4,749,120**	**27,064,128**

Accuracy tests

- Signals coming from Physiobank MIT-BIH Arrhythmia database
- ECG signals are divided in 30 seconds intervals
  - Each interval is labelled as clean or noisy
- Additional noise signals are created by adding simulated electrode contact noise
  - To the whole (30 seconds) interval
  - To a smaller portion
- Tests performed for each signal:
  - Clean vs noisy
  - Clean vs Simulated noise added to the whole interval
  - Clean vs Simulated noise added to a portion of the interval
- Compared classifiers based on
  - SNR of whole interval
  - Mean SNR
  - Variance of SNR
  - Linear classifier

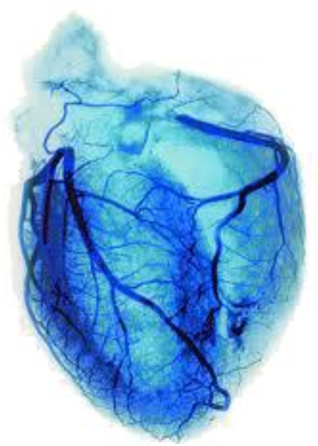
Accuracy results

- Training
  - 60% of clean intervals and 60% of noisy intervals, randomly chosen
  - A threshold estimated for each signal (minimum error probability)
- Testing
  - ECG intervals not used for training

	intSNR	intMean	intVariance	Linear Classifer
Clean/noisy	0.732	0.706	0.815	0.849
Clean/added	0.823	0.836	0.800	0.836
Clean/partially added	0.666	0.669	0.672	0.737

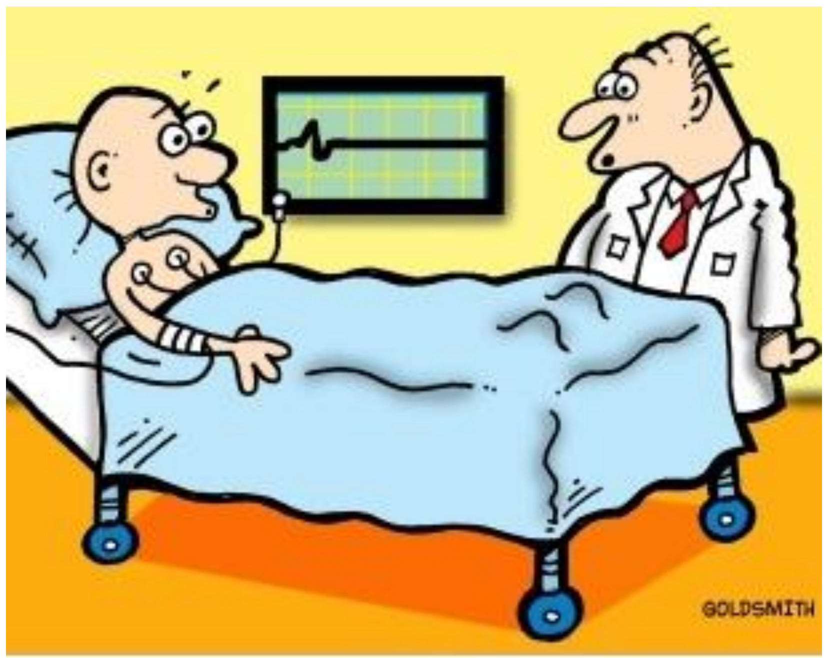
Conclusions

- Proposed a protocol to evaluate the quality of an ECG signal in remote health monitoring applications
- Easy to implementin the encrypted domain
  - Hybrid protocol
  - Online transmission of 3.4 Mbytes of data
- More than 84% correct classification rate on signals of the MIT Arrhythmia database
- Track for the future:
  - Packing during filtering (Bianchi's paper)
  - Change of the secret key owner in HE subsection
  - Replace worst case with statistical analysis
  - Evaluation of computational complexity



Thanks for your attention

Questions?



If the ECG isn't broken then we have problem