Covert Communication Over Additive-Noise Channels

Cécile BOUETTE[1], Laura LUZZI[1], Ligong WANG[2]

[1]Laboratoire ETIS-UMR 8051, CY Cergy Paris Université, ENSEA, CNRS

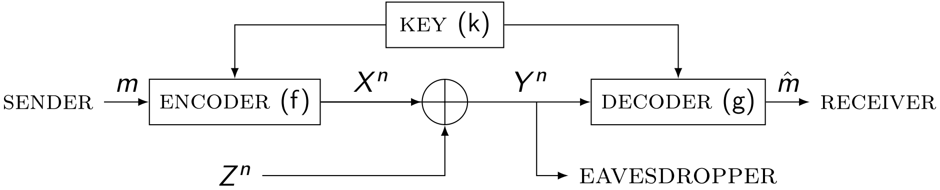
[2]Department of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zurich, Switzerland

https://arxiv.org/abs/2402.16475

7$^{\text{th}}$ of June 2024 - France PhD Information Theory Workshop at Télécom Paris

Équipes Traitement
de l'Information
et Systèmes

Covert communication setup



A code $\mathcal{C} = (f, g)$ of length n for message set $\mathcal{M}$ and key set $\mathcal{K}$ consists of an encoder $f : \mathcal{M} \times \mathcal{K} \to \mathbb{R}^n, (m, k) \mapsto x^n$ and a decoder $g : \mathbb{R}^n \times \mathcal{K} \to \mathcal{M}, (y^n, k) \mapsto \hat{m}$.

Covertness constraint

For some given $\Delta > 0$,

$$\mathbb{D}(P_{Y^n} || P_{Z^n}) \leq \Delta.$$

Fundamental limit of covert communication

- The framework of covert communication was introduced by [Bash, Goeckel, and Towsley, 2012].
- [Wang, Wornell, and Zheng, 2016], [Bloch, 2016]: fundamental asymptotic limits for discrete memoryless channels and memoryless Gaussian channels.

Square-root law

It is not possible to achieve a positive rate of communication.

The maximum amount of information that can be transmitted reliably and covertly over n channel uses scales like $\sqrt{n}$.

Covert communication scaling constant

Given $\epsilon > 0$, we denote by $A_n(\Delta, \epsilon)$ the maximum of $\ln|\mathcal{M}|$ for which there exists a random code C of length n that satisfies the covertness condition, and whose average probability of decoding error is at most ϵ.

$$L \triangleq \lim_{\epsilon \downarrow 0} \varliminf_{n \to \infty} \frac{A_n(\Delta, \epsilon)}{\sqrt{n\Delta}}.$$

Theorem [Wang, Wornell, and Zheng, 2016]

For AWGN channels, $L = 1\sqrt{\text{nat}}$ irrespectively of the noise variance.

Theorem [Bouette, Luzzi, and Wang, 2023]

For Gaussian channels with memory, $L = 1\sqrt{\text{nat}}$ irrespectively of the noise covariance.

Motivation and main contributions

Motivation: in many scenarios, the noise is not Gaussian. In particular, in networks with interference, it can be heavy-tailed [Clavier et al., 2021].

Goal

Characterize L for memoryless additive channels with general noise distributions.

- Under mild integrability assumptions on the noise PDF, we show that the square-root scaling constant is upper-bounded by a simple expression.
- Under some additional assumptions, the upper bound is tight.
- We provide upper bounds on the key length.

A general upper bound on L

Integrability assumptions

We suppose the noise is i.i.d. with PDF p_Z, and assume there exists $\zeta \in (0,1)$ s.t.

$$\int_{\mathbb{R}} p_Z(z)\left(\ln(p_Z(z))\right)^4 \mathrm{d}z < \infty$$

$$\int_{\mathbb{R}} p_Z(z)^{\zeta} \mathrm{d}z < \infty$$

$$\int_{\mathbb{R}} p_Z(z)^{\zeta}\left(\ln(p_Z(z))\right)^4 \mathrm{d}z < \infty.$$

Theorem: converse

Under the previous integrability conditions, $L \leq \sqrt{2}\sqrt{\mathrm{Var}\left[\ln(p_Z(Z))\right]}$.

Sketch of the converse proof 1/4

Idea: Characterize the distribution P_Y that maximizes $h(Y)$ for a given $D(P_Y||P_Z)$.

Lemma 1

Consider the random variable $\tilde{Z}$ with density

$$p_{\tilde{Z}}(z) = \alpha p_Z(z)^{1-\lambda} \qquad \text{where } \alpha = \left(\int_{\mathbb{R}} p_Z(z)^{1-\lambda} \mathrm{d}z\right)^{-1}.$$

Then for any random variable Y:

$$\mathbb{D}(P_Y||P_Z) \leq \mathbb{D}(P_{\tilde{Z}}||P_Z) \implies h(Y) \leq h(\tilde{Z})$$

with equality if and only if Y follows the same distribution as $\tilde{Z}$.

Sketch of the converse proof 2/4

- There exists a sequence $\{\gamma_n\}$ such that

$$\lim_{n\to\infty} \gamma_n = 0$$

and the random variables $\{\tilde{Z}_n\}$ with PDFs defined as

$$p_{\tilde{Z}_n}(\tilde{z}) = \alpha_n \cdot p_Z(\tilde{z})^{1-\gamma_n}, \qquad \text{where } \alpha_n = \left(\int_{\mathbb{R}} p_Z(z)^{1-\gamma_n} \mathrm{d}z\right)^{-1}$$

satisfy

$$\mathbb{D}(P_{\tilde{Z}_n} \| P_Z) = \frac{\Delta}{n}.$$

- Using Taylor expansions we find

$$\gamma_n = \sqrt{\frac{2}{\mathrm{Var}\left[\ln(p_Z(Z))\right]}}\sqrt{\frac{\Delta}{n}} + o\left(\frac{1}{\sqrt{n}}\right).$$

Sketch of the converse proof 3/4

- Take any covert random code C of length n.
- Let $P_{\bar{Y}}$ denote the average output distribution over all possible keys, a uniformly drawn message, and the n channel uses.
  From the covertness constraint, using the chain rule and convexity of KL divergence,

$$\mathbb{D}\left(P_{\bar{Y}} \| P_Z\right) \leq \frac{\Delta}{n} = \mathbb{D}(P_{\tilde{Z}_n} \| P_Z).$$

- Lemma 1 implies

$$h\left(\bar{Y}\right) \leq h(\tilde{Z}_n).$$

Sketch of the converse proof 4/4

- Let ϵ_n be the average error probability over the random codebook.
- By averaging over the random code and using Fano's inequality:

$$\ln|\mathcal{M}|(1-\epsilon_n) - 1 \leq nI(\bar{X};\bar{Y}) = n\left(h\left(\bar{Y}\right) - h(Z)\right) \leq n\left(h(\tilde{Z}_n) - h(Z)\right).$$

- Knowing γ_n, we find

$$h(\tilde{Z}_n) - h(Z) = \sqrt{2}\sqrt{\mathrm{Var}\left[\ln(p_Z(Z))\right]}\sqrt{\frac{\Delta}{n}} + o\left(\frac{1}{\sqrt{n}}\right).$$

- Recalling $L \triangleq \lim_{\epsilon\downarrow 0} \varliminf_{n\to\infty} \frac{A_n(\Delta,\epsilon)}{\sqrt{n\Delta}}$ and taking $n \to +\infty$:

$$L \leq \sqrt{2}\sqrt{\mathrm{Var}\left[\ln(p_Z(Z))\right]}.$$

Tightness of the upper bound

Assumption 1

- p_Z is bounded,
- $z \mapsto p_Z(z) \ln(p_Z(z))$ is uniformly continuous,
- $\exists \xi \in (0,1)$ such that, for all $\gamma \in [0,\xi)$, there exists a random variable X independent of $Z \sim p_Z$ such that the PDF of $X + Z$ is $p_{\tilde{Z}}$ given by Lemma 1.

Theorem: achievability

Under the previous integrability conditions and Assumption 1:

$$L = \sqrt{2}\sqrt{\mathrm{Var}\left[\ln(p_Z(Z))\right]}.$$

Sketch of the achievability proof 1/3

- Fix $\chi \in \left(1, \frac{3}{2}\right)$. For each n, let $\tilde{Z}_n$ have the PDF in Lemma 1, with the choice

$$\gamma_n = \sqrt{\frac{2}{\mathrm{Var}\left[\ln(p_Z(Z))\right]}\left(\frac{\Delta}{n} - \frac{1}{n^\chi}\right)}.$$

- The existence of X_n s.t. $X_n + Z = \tilde{Z}_n$ is guaranteed by Assumption 1.
- We generate a random codebook C by picking every codeword i.i.d. $\sim P_{X_n}$.
- We check that the covertness constraint is satisfied:

$$\begin{aligned}\mathbb{E}_{\mathsf{C}}[\mathbb{D}(P_{Y^n|\mathsf{C}} \| P_{Z^n})] &= \mathbb{E}_{\mathsf{C}}\left[\mathbb{D}\left(P_{Y^n|\mathsf{C}} \middle\| P_{\tilde{Z}_n}^{\times n}\right)\right] + \mathbb{D}\left(P_{\tilde{Z}_n}^{\times n} \middle\| P_{Z^n}\right).\\ &= \mathbb{E}_{\mathsf{C}}\left[\mathbb{D}\left(P_{Y^n|\mathsf{C}} \middle\| P_{\tilde{Z}_n}^{\times n}\right)\right] + \Delta - n^{1-\chi} + O\left(\frac{1}{\sqrt{n}}\right) \leq \Delta\end{aligned}$$

- We assume the key is sufficiently long to have $\mathbb{E}_{\mathsf{C}}\left[\mathbb{D}\left(P_{Y^n|\mathsf{C}} \middle\| P_{\tilde{Z}_n}^{\times n}\right)\right]$ close to 0.

Sketch of the achievability proof 2/3

Lemma 2

Consider p_Z satisfying the integrability conditions and Assumption 1.

Then $P_{\tilde{Z}_n}$ converges weakly to P_Z, and P_{X_n} converges weakly to the Dirac distribution.

Proof idea of Lemma 2

For any bounded continuous function f on $\mathbb{R}$:

$$\left|\mathbb{E}\left[f(\tilde{Z}_n)\right] - \mathbb{E}[f(Z)]\right| \leq \|f\|_\infty \int_{\mathbb{R}} \left|p_{\tilde{Z}_n}(z) - p_Z(z)\right| \mathrm{d}z \leq \|f\|_\infty \sqrt{2\mathbb{D}\left(P_{\tilde{Z}_n} \| P_Z\right)} \to 0.$$

- Weak convergence of P_{X_n} towards the Dirac distribution follows by Lévy's convergence theorem.

Sketch of the achievability proof 3/3

- From [Verdú and Han, 1994], we know there exists a sequence of codes with vanishing error probabilities such that:

$$\lim_{n\to\infty} \frac{\ln|\mathcal{M}|}{\sqrt{n}} \geq \mathbb{P}\text{-}\liminf_{n\to\infty} \frac{i_{X^n,Y^n}(X^n,Y^n)}{\sqrt{n}}.$$

- With Lemma 2, we show $\mathrm{var}\left(\frac{1}{\sqrt{n}} i_{X^n,Y^n}(X^n,Y^n)\right) \xrightarrow[n\to+\infty]{} 0.$
- Using Chebyshev's inequality, we prove

$$\lim_{n\to\infty} \frac{\ln|\mathcal{M}|}{\sqrt{n}} \geq \mathbb{P}\text{-}\liminf_{n\to\infty} \frac{1}{\sqrt{n}} i_{X^n,Y^n}(X^n,Y^n) = \lim_{n\to\infty} \frac{I(X^n;Y^n)}{\sqrt{n}}.$$

- Since $P_{\tilde{Z}}$ maximizes the entropy, we have the desired result of achievability:

$$L = \sqrt{2}\sqrt{\mathrm{Var}[\ln(p_Z(Z))]}.$$

Example 1/3: exponential noise

$$p_Z(z) = \lambda e^{-\lambda z}, \qquad \lambda > 0.$$

We have

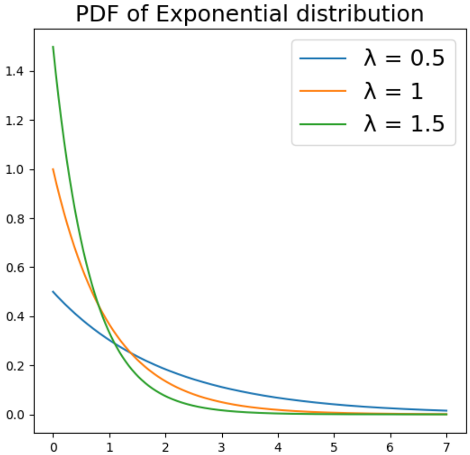
$$p_{\tilde{Z}_n}(\tilde{z}) = (1-\gamma_n)\lambda e^{-(1-\gamma_n)\lambda \tilde{z}}, \gamma_n = O\left(\frac{1}{\sqrt{n}}\right).$$

Assumption 1 holds [Verdú, 1996]:

$$p_{X_n}(x) = \gamma_n(1-\gamma_n)\lambda e^{-(1-\gamma_n)\lambda x} + (1-\gamma_n)\delta_0(x),$$

Finally

$$L = \sqrt{2}.$$



Example 2/3: generalized Gaussian noise

[Nadarajah, 2005]: $p_Z(z) = \frac{c_p}{\sigma} e^{-\frac{|z|^p}{2\sigma^p}}, \quad z \in \mathbb{R},$

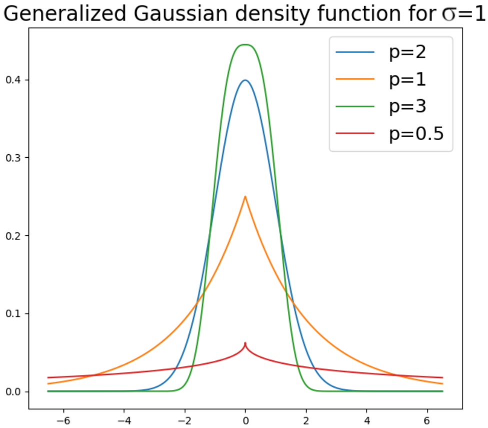
$$\text{where } c_p = \frac{p}{2^{\frac{p+1}{p}}\Gamma(\frac{1}{p})}, \text{ and } p, \sigma > 0,$$

with $\Gamma(\cdot)$ denoting the gamma function.

[Dytso et al., 2017]: Assumption 1 holds for $p \in (0, 1]$ or $p = 2$ then

[Bouette, Luzzi, and Wang, 2023]: $L \leq \sqrt{\frac{2}{p}}$,

with equality if $p \in (0, 1]$ or $p = 2$.



Example 3/3: generalized gamma noise

[Stacy, 1962]: $r, \beta, \sigma > 0$

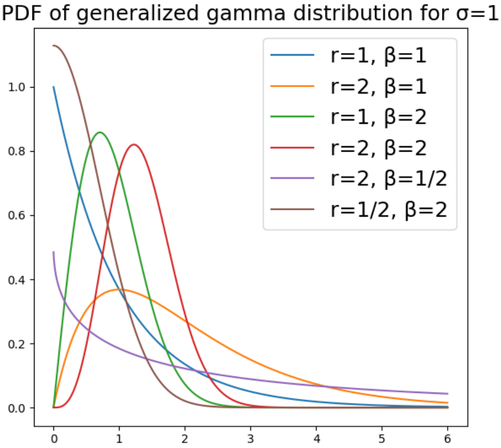
$$p_Z(z) = \frac{\beta}{\Gamma(r)\sigma^{\beta r}} z^{\beta r - 1} e^{-\left(\frac{z}{\sigma}\right)^{\beta}} \qquad z \in \mathbb{R}^+,$$

where $\Gamma(\cdot)$ denotes the gamma function.

$$L \leq \sqrt{2}\sqrt{\left(r - \frac{1}{\beta}\right)^2 \psi^{(1)}(r) - r + \frac{2}{\beta}},$$

with $\psi^{(1)}(\cdot)$ denoting the first derivative of the digamma function.

Remark: L does not depend on σ.



Bounds on the key length

Motivation: Up to now, we have assumed that an arbitrarily long key is shared between Alice and Bob. We now show that a finite key is sufficient.

Proposition 1

If p_Z satisfies the integrability conditions as well as Assumption 1, then there exists a sequence of codes that asymptotically achieves the optimal scaling factor L of Theorem 2 with key lengths satisfying

$$\ln|\mathcal{K}| = O(n).$$

Proposition 2

For P_Z being a Gaussian or exponential distribution, it can be strengthened to

$$\ln|\mathcal{K}| = o(\sqrt{n}).$$

Sketch of proof

- We consider the previous random codebook. The covertness condition requires

$$\mathbb{E}_{\mathsf{C}}[\mathbb{D}(P_{Y^n|\mathsf{C}}\|P_{Z^n})] = \mathbb{E}_{\mathsf{C}}\left[\mathbb{D}\left(P_{Y^n|\mathsf{C}}\middle\|P_{\tilde{Z}_n}^{\times n}\right)\right] + \mathbb{D}\left(P_{\tilde{Z}_n}^{\times n}\middle\|P_{Z^n}\right) \leq \Delta.$$

  We characterize a sufficient key length such that $\mathbb{E}_{\mathsf{C}}\left[\mathbb{D}\left(P_{Y^n|\mathsf{C}}\middle\|P_{\tilde{Z}_n}^{\times n}\right)\right] \to 0$.

- Channel resolvability bound of [Hayashi and Matsumoto, 2016]: for $\rho \in (0,1]$,

$$\mathbb{E}_{\mathsf{C}}\left[\mathbb{D}\left(P_{Y^n|\mathsf{C}}\|P_{\tilde{Z}_n}^{\times n}\right)\right] \leq \frac{1}{\rho}\ln\left(1 + e^{-\rho\ln|\mathcal{K}|-\rho\ln|\mathcal{M}|+n\Psi(\rho|P_{Y|X},P_X)}\right),$$

$$\text{where } \Psi(\rho|P_{Y|X},P_X) = \ln\left(\mathbb{E}\left[\left(\frac{p_{Y|X}(Y|X)}{p_Y(Y)}\right)^{\rho}\right]\right).$$

- We show that
  - Ψ is bounded.
  - for Gaussian and Exponential noise: $n\Psi(\rho|P_{Y|X},P_X) = \rho\ln|\mathcal{M}| + o\left(\sqrt{n}\right)$.

Conclusions and open problems

- Under integrability conditions on the noise PDF

$$L \leq \sqrt{2}\sqrt{\mathrm{Var}\left[\ln(p_Z(Z))\right]},$$

with equality for many noise distributions.

- A sufficient key length is $\ln|\mathcal{K}| = O(n)$ and can be reduced to $\ln|\mathcal{K}| = o(\sqrt{n})$ when the noise is Gaussian or Exponential.

Open problems

- cases where the legitimate receiver and eavesdropper have different channels.
- more general additive channels with memory.

Bibliography I

E. Abbe and M. Ye. Reed-Muller codes polarize. *IEEE Transactions on Information Theory*, 66(12):7311–7332, 2020. doi: 10.1109/TIT.2020.3023487.

B. Bash, D. Goeckel, and D. Towsley. Limits of reliable communication with low probability of detection on AWGN channels. *IEEE Journal on Selected Areas in Communications*, 31, 02 2012. doi: 10.1109/JSAC.2013.130923.

M. R. Bloch. Covert communication over noisy channels: A resolvability perspective. *IEEE Transactions on Information Theory*, 62(5):2334–2354, 2016. doi: 10.1109/TIT.2016.2530089.

A. Dytso, R. Bustin, H. V. Poor, and S. Shamai. Analytical properties of generalized Gaussian distributions. *Journal of Statistical Distributions and Applications*, 2017. doi: 10.1186/s40488-018-0088-5.

G. Frèche, M. R. Bloch, and M. Barret. Polar codes for covert communications over asynchronous discrete memoryless channels. *Entropy*, 20(1):3, 2017. doi: 10.3390/e20010003.

H. Hassani, S. Kudekar, O. Ordentlich, Y. Polyanskiy, and R. Urbanke. Almost optimal scaling of Reed-Muller codes on BEC and BSC channels. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 311–315, 2018. doi: 10.1109/ISIT.2018.8437453.

Bibliography II

M. Hayashi and R. Matsumoto. Secure multiplex coding with dependent and non-uniform multiple messages. *IEEE Trans. Inform. Theory*, 62(5):2355–2409, May 2016. doi: 10.1109/TIT.2016.2530088.

I. A. Kadampot, M. Tahmasbi, and M. R. Bloch. Codes for covert communication over additive white gaussian noise channels. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 977–981, 2019. doi: 10.1109/ISIT.2019.8849662.

S. Nadarajah. A generalized normal distribution. *Journal of Applied statistics*, 32(7):685–694, 2005. doi: 10.1080/02664760500079464.

E. W. Stacy. A generalization of the gamma distribution. *Ann. Math. Stat.*, 33(3):1187–1192, September 1962. doi: 10.18187/pjsor.v14i1.1692.

M. Tahmasbi and M. R. Bloch. First-and second-order asymptotics in covert communication. *IEEE Transactions on Information Theory*, 65(4):2190–2212, 2018. doi: 10.48550/arXiv.1703.01362.

S. Verdú. The exponential distribution in information theory. *Problems Inform. Transmission*, 32(1):86–95, March 1996. ISSN 0555-2923.

S. Verdú and T. S. Han. A general formula for channel capacity. *IEEE Transactions on Information Theory*, 40 (4):1147–1157, 1994. doi: 10.1109/18.335960.

Bibliography III

L. Wang, G. W. Wornell, and L. Zheng. Fundamental limits of communication with low probability of detection. *IEEE Transactions on Information Theory*, 62(6):3493–3503, 2016. doi: 10.1109/TIT.2016.2548471.