

On the security of key distribution based on Johnson-Nyquist noise

Charles H. Bennett, C. Jess Riedel
IBM Watson Research Center, Yorktown Heights, NY, USA
(Dated: June 7, 2021)

We point out that arguments for the security of Kish's noise-based cryptographic protocol have relied on an unphysical no-wave limit, which if taken seriously would prevent any correlation from developing between the users. We introduce a noiseless version of the protocol, also having illusory security in the no-wave limit, to show that noise and thermodynamics play no essential role. Then we prove generally that classical electromagnetic protocols cannot establish a secret key between two parties separated by a spacetime region perfectly monitored by an eavesdropper. We note that the original protocol of Kish is vulnerable to passive time-correlation attacks even in the quasi-static limit. Finally we show that protocols of this type can be secure in practice against an eavesdropper with noisy monitoring equipment. In this case the security is a straightforward consequence of Maurer and Wolf's discovery that key can be distilled by public discussion from correlated random variables in a wide range of situations where the eavesdropper's noise is at least partly independent from the users' noise.

Quantum key distribution [1] boasts unconditional security even in the presence of realistic noise [2], and the techniques have matured enough that small commercial implementations have been explored. However, the practical difficulty of manipulating individual quantum states has prompted some investigation into purely classical schemes which might be able to achieve similar ends. In particular, Kish has proposed a strictly classical protocol (Kirchoff Law-Johnson Noise or KLJN) on an insecure transmission line using Johnson-Nyquist noise in resistors [3] (figure 1). He and collaborators claim that the idealized model of this protocol is unconditionally secure and that the prospects of real-world implementations are promising [4]. Intuitively, this seems dubious; information passing between two parties must travel as an electromagnetic wave over the transmission line, and a potential eavesdropper should be able to perfectly measure a classical signal (regardless of its noisy origins) with arbitrary precision and with arbitrarily little disturbance.

First, we note that the security of the KLJN protocol is only claimed in the limit where electromagnetic modes of finite wavelength are neglected [3]. In other words, currents and voltages are assumed to be quasi-static. Kish has argued that this obviates any need to analyze waves moving along the transmission line [5] and he agrees intercepted waves would compromise security [6]. We believe this no-wave limit is inappropriate and nonphysical for analyzing communication protocols (even as a mathematical idealization) because if propagating waves are excluded there is no way for information to get from Alice's side of the circuit to influence Bob's side, or vice versa. Unfortunately this no-wave limit has formed the basis of subsequent claims that the idealized protocol is unconditionally secure [4, 5] in response to criticism [7].

To illustrate our critique, we depict a simplified analog of the KLJN protocol in figure 2 which is similarly secure against passive attack in the no-wave limit. Since the

concepts of temperature and noise have been eliminated, they play no fundamental role in the protocol, undermining the recent claim of Kish et al. that the security of the KLJN protocol follows from the 2nd law of thermodynamics [4]. This was already suggested by the earlier observation in reference [3] that artificial noise generators were as good as true Johnson-Nyquist noise from resistors.

Kish has argued that many cryptographic schemes (such as quantum key distribution) were initially only



FIG. 1. **Basic KLJN protocol.** Alice and Bob are at opposite ends of an ideal transmission line, with Eve somewhere in the middle. At each clock cycle Alice and Bob decide independently and randomly whether to terminate their end of the line with a low resistance (L) or a high resistance (H). The Johnson-Nyquist noise from these thermal resistances R_L and R_H at temperature T can be modeled as an ideal resistor in series with a voltage source undergoing Gaussian fluctuations. After transients resulting from this switching have died down, the mean square noise voltage will be the same at any point between Alice and Bob, having one of three values: low (for the LL combination), high (for the HH combination) and intermediate (for LH or HL). Alice and Bob discard data from the HH and LL events, and keep the HL and LH data. The kept events are candidates for a secret key because, knowing their own switch settings, they can distinguish HL from LH, but Eve cannot do so from the mean square noise voltage on the transmission line.

analyzed as mathematical idealizations, and that the usefulness of the KLJN protocol can only be assessed with a detailed mathematical analysis of real-world inefficiencies [5, 8]. Although we are sympathetic to the general idea that physical robustness of information systems is very important and often non-obvious—witness fault-tolerant quantum computation [9]—we emphasize that quantum key distribution *has* been shown to be robust with imperfect components against very general attacks [2] while exploiting an unphysical limit seems to be essential to the KLJN protocol. We can be more precise as follows.

Suppose that Eve makes a continuous measurement of the electromagnetic field on a thin cross-section of the transmission line (including the grounding wire) between times $t = 0$ and $t = T$, denoted by the variable Z . This variable contains a complete history of the fields inside the cross-section during the interval of length T . Since this is a purely classical analysis, the measurement can be ideal. We can either (1) use the exact (or ‘microscopic’) version of Maxwell’s equations and imagine that this cross-sectional plane passes between all atoms in the transmission line or (2) use the macroscopic version of Maxwell’s equations and take the transmission line to be a continuous medium. Either way, Eve can decompose the fields into orthogonal components Z_A and Z_B , describing the waves heading toward Alice and Bob, respectively.

Let Y be a variable that describes everything on Bob’s side of Eve’s location during the same time interval, including waves traveling toward him, away from him, and



FIG. 2. **Noiseless KLJN protocol.** The wires and voltage sources are taken to be ideal, with zero thermal noise. For each clock cycle, Alice randomly chooses either “L-first” or “H-first”. If the former, she connects the L side of her circuit for the first half of the clock interval, and the H side for the second half. (If the latter, she does the reverse.) Bob does likewise. If the choices are the same, then for half of the clock interval the voltage on the transmission line will be zero and that clock cycle is discarded. If the choices are different, then the voltage is V_0 for the entire interval, regardless of who chose “L-first”. (The transmission line can be momentarily re-grounded between half-clock cycles to ensure it has zero voltage when neither party connects a battery.) The choice is known to both Alice and Bob, which now forms their key bit, but is unconditionally secure from an eavesdropper according to the standards used to claim security of the original KLJN protocol in figure 1. Thermodynamics and noise do not play a role.

all of his equipment and memory. Let X be the same for Alice. X , Y , and $Z = (Z_A, Z_B)$ are random variables characterizing the distribution of possible histories when the protocol is run. (This captures, for example, the outcome of any coin flips Alice and Bob perform, as well as possible noise in the components.) The initial data on the $t = 0$ surfaces of X and Y must be independent variables—otherwise, Alice and Bob could simply use them to construct a secret key—but X , Y , Z_A , and Z_B will be generally all be correlated.

The deterministic and locally causal structure of Maxwell’s equations ensures that Y can only be influenced by X via the intermediary Z_A . This can be expressed mathematically as

$$H(X|Z_A) = H(X|Z) = H(X|Z, Y). \quad (1)$$

The conditional information $H(S|T)$ denotes the uncertainty in S remaining when T is known. Equation (1) just says that, once Z_A is known, nothing further about X is learned from measuring Z_B or Y . This can equivalently be stated as a restriction on the joint probability density:

$$p(x, y, z) = \int dz p(z) p(x|z_A) p(y|z_B), \quad (2)$$

where $z = (z_A, z_B)$ and where $p(z)$, $p(x|z_A)$, and $p(y|z_B)$ are defined in the usual way from $p(x, y, z)$. Intuitively, X only “knows” about Z_A and so is completely specified by $p(x|z_A)$.

It follows that $X \rightarrow Z \rightarrow Y$ forms a Markov chain and that the conditional mutual information between X and Y , conditional on Z , vanishes:

$$I(X:Y|Z) = H(X|Z) - H(X|Z, Y) = 0. \quad (3)$$

Under this condition Alice and Bob, even with the help of public discussion, cannot establish a key that is secret from Eve [11]. That is, the distillable key rate is zero.

This conclusion depends on Eve’s ability to accurately measure the time history of the electromagnetic field, not just its instantaneous spectrum. Thus, while the steady state mean square noise voltage in the original KLJN protocol (figure 1) does not allow Eve to distinguish between the LH and HL settings of Alice’s and Bob’s resistors [3], she can distinguish them using either (a) transient waves created by the switching action before the steady state is established, or (b) time correlations in the steady-state distribution of traveling waves resulting from the fluctuations that give rise to Johnson-Nyquist noise. For example Bob’s resistor affects the phase and amplitude correlations between a right-traveling wave at time t and its left-traveling echo at time $t + \Delta$, where Δ is the transit time from Eve to Bob and back, with the echo vanishing only if the resistor is perfectly impedance matched to his end of the line.

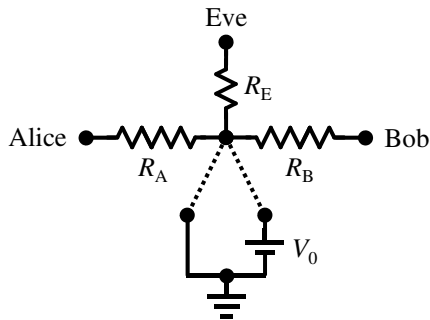


FIG. 3. **Key distillation from Johnson-Nyquist noise.**

Consider a static lumped circuit with resistors at finite temperature connected as in the figure. The switch for the central node is flipped randomly between ground and a (tiny) voltage V_0 . Alice, Bob, and Eve see this through noisy thermal resistors R_A , R_B , and R_E , each of which applies an independent Gaussian offset to the voltage at the central node. Even if Eve's noise is less than Alice's and Bob's (e.g. if $R_E < R_A$, $R_E < R_B$, with everyone at the same temperature), Alice and Bob can use advantage distillation [10] to distill a secret key via 2-way public communication provided R_E is strictly positive. However if Eve's measurement of the fluctuating voltage at the central node is noiseless ($R_E = 0$), Alice's and Bob's conditional mutual information, conditioned on Eve's observation, is zero, and no secret key can be distilled.

Our noiseless protocol in figure 2 with ideal components has no fluctuations and is not open to this passive steady-state attack, but of course it could be broken by observing the transients as the voltage was being raised on the transmission line from one side or the other. These transients are precisely the waves which carry the information between Alice and Bob. However, if Eve graciously promises not to observe the transients, she could still learn the key by an *active* steady-state attack in which she would place a very high-resistance shunt between her node and ground, and monitor the direction of current flow into it. Of course Alice and Bob could try to detect this weak leakage current also, and abort the protocol if they found it. The result would be an unstable arms race, won by whichever side had the more sensitive ammeter, not the sort of robustness reasonably expected of a practical cryptosystem.

It has been claimed that low-pass filters added by Alice and Bob in an attempt to enforce the no-wave limit would prevent attacks on propagating waves in KLJN systems [4, 5]. This claim has not been quantified, and in any case filters added by Alice and Bob do not avoid our general theorem (3) about classical electromagnetic signaling; any frequency components traveling toward Bob which are removed by Alice's low-pass filter before entering the public section of the transmission line are unavailable to both Eve and Bob.

We emphasize that so long as Eve makes perfect measurements, equation (3) holds regardless of whether X , Y , and Z are continuous or discrete, whether Alice

and Bob perform deterministic or probabilistically programmed actions specified by the protocol, and whether the equipment and transmission lines have loss, noise, filters, or even memory. Therefore, the problems of finite resistance in the transmission line [7] or temperature differences between Alice and Bob [12] are moot. Attacks based on these imperfections become important only when the unphysical no-wave limit is accepted, which as we noted would also prevent key agreement by preventing any information from passing between Alice and Bob in the first place.

Of course, none of this shows that ideas contained in the KLJN protocol cannot find useful application. There are many classical cryptographic protocols whose security rests on some assumed limitation on the class of attacks available to the eavesdropper. For example, the last two decades have seen interest in the use of synced chaotic lasers for transmitting messages which cannot be decoded by an eavesdropper using simple measurement strategies. An early example relevant to key generation from thermal noise is the scenario illustrated in figure 3, where Alice, Eve, and Bob listen to the same random binary source through three independent Gaussian channels, yielding variables A , B , and E . Maurer showed (reference [10] section V) that Alice and Bob can distill secret key at a positive rate even if their channels are more noisy than Eve's by using two-way public communication to collaboratively agree on a subset of their data on which their noise is less than Eve's, even though it is greater than Eve's on the raw data as a whole. The key rate obtainable from this "advantage distillation" technique depends on the noisiness of Eve's measurements, being upper bounded by the conditional mutual information $I(A; B|E)$, which vanishes in the limit $R_E \rightarrow 0$. Advantage distillation has been further generalized and developed by Maurer and Wolf [11]. For quantitative studies of key rates obtainable with discrete and Gaussian channels see [13] and [14].

In conclusion, we have shown that protocols of the KLJN type do not use thermodynamics in an essential way, and that they can be completely defeated by an adversary able to perfectly monitor the time-dependent fields in the region between Alice and Bob, as classical electromagnetic theory would allow. On the other hand, if in practice Eve's monitoring can be guaranteed to be finitely noisy, such protocols may yield secure key, at a rate not exceeding the conditional mutual information, using Maurer's technique of advantage distillation.

We thank John Smolin and Graeme Smith for extremely helpful discussions. This work was supported in part by the John Templeton Foundation, grant number 21484.

-
- [1] C. Bennett, G. Brassard, *et al.*, in *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, Vol. 175 (Bangalore, India, 1984); P. W. Shor and J. Preskill, *Physical Review Letters* **85**, 441 (2000); N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, *Reviews of modern physics* **74**, 145 (2002).
 - [2] D. Mayers, *Journal of the ACM (JACM)* **48**, 351 (2001); D. Gottesman, H.-K. Lo, N. Lutkenhaus, and J. Preskill, in *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on* (IEEE, 2004) p. 136; H.-K. Lo, M. Curty, and B. Qi, *Physical Review Letters* **108**, 130503 (2012); S. L. Braunstein and S. Pirandola, *ibid.* **108**, 130502 (2012).
 - [3] L. B. Kish, *Physics Letters A* **352**, 178 (2006).
 - [4] L. B. Kish, in *Classical, Semi-classical and Quantum Noise*, edited by L. Cohen, H. Poor, and M. Scully (Springer, 2011); L. B. Kish and F. Peper, *IEICE Transactions on Communications* **95**, 1501 (2012).
 - [5] L. B. Kish, *Physics Letters A* **359**, 741 (2006).
 - [6] L. B. Kish, “Why wave-based classical communicators can never be totally secure?” (2007), arXiv:physics/0610102.
 - [7] J. Scheuer and A. Yariv, *Physics Letters A* **359**, 737 (2006).
 - [8] L. B. Kish, *Fluctuation and Noise Letters* **6**, 37 (2006).
 - [9] P. Shor, in *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on* (IEEE, 1996) pp. 56–65.
 - [10] U. Maurer, *Information Theory, IEEE Transactions on* **39**, 733 (1993).
 - [11] U. Maurer and S. Wolf, *Information Theory, IEEE Transactions on* **45**, 499 (1999).
 - [12] F. Hao, *IEEE Information Security* **153**, 142 (2006).
 - [13] M. J. Gander and U. M. Maurer, in *Information Theory, 1994. Proceedings., 1994 IEEE International Symposium on* (IEEE, 1994) p. 351.
 - [14] M. Naito, S. Watanabe, R. Matsumoto, and T. Uye-matsu, *IEICE transactions on fundamentals of electronics, communications and computer sciences* **92**, 525 (2009).