Cryptographical Properties of Ising Spin Systems

Yoshiyuki Kabashima,[1] Tatsuto Murayama,[1] and David Saad[2]
[1]*Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Yokohama 2268502, Japan*
[2]*The Neural Computing Research Group, Aston University, Birmingham B4 7ET, United Kingdom*

(Received 19 August 1999)

The relation between Ising spin systems and public-key cryptography is investigated using methods of statistical physics. The insight gained from the analysis is used for devising a matrix-based cryptosystem whereby the ciphertext comprises products of the original message bits; these are selected by employing two predetermined randomly constructed sparse matrices. The ciphertext is decrypted using methods of belief propagation. The analyzed properties of the suggested cryptosystem show robustness against various attacks and competitive performance to modern cyptographical methods.

PACS numbers: 89.70.+c, 03.67.Dd, 05.50.+q, 89.80.+h

Public-key cryptography plays an important role in many aspects of modern information transmission, for instance, in the areas of electronic commerce and internet-based communication. It enables the service provider to distribute a public key which may be used to encrypt messages in a manner that can be decrypted only by the service provider. The ongoing search for safer and more efficient cryptosystems produced many useful methods over the years such as RSA (by Rivest, Shamir, and Adleman), elliptic curves, and the McEliece cryptosystem, to name but a few.

In this Letter, we employ methods of statistical physics to study a specific cryptosystem, somewhat similar to the one presented by McEliece [1]. These methods enable one to study the typical performance of the suggested cryptosystem, to assess its robustness against attacks, and to select optimal parameters.

The main motivation for the suggested cryptosystem comes from previous studies of Gallager-type error-correcting codes [2–4] and their physical properties [5,6]. The analysis exposes a significantly different behavior for the two-matrix based codes (such as the MN code [3]) and single-matrix codes [4], which may be exploited for constructing an efficient cryptosystem.

In the suggested cryptosystem, a plaintext represented by an N dimensional Boolean vector $\boldsymbol{\xi} \in (0,1)^N$ is encrypted to the M dimensional Boolean ciphertext $\boldsymbol{J}$ using a predetermined Boolean matrix G, of dimensionality $M \times N$, and a corrupting M dimensional vector $\boldsymbol{\zeta}$, whose elements are 1 with probability p and 0 otherwise, in the following manner:

$$\boldsymbol{J} = G\boldsymbol{\xi} + \boldsymbol{\zeta}\,, \qquad (1)$$

where all operations are (mod 2). The matrix G and the probability p constitute the public key; the corrupting vector $\boldsymbol{\zeta}$ is chosen at the transmitting end. The matrix G, which is at the heart of the encryption/decryption process is constructed by choosing two randomly selected sparse matrices A and B of dimensionality $M \times N$ and $M \times M$, respectively, defining

$$G = B^{-1}A \qquad (\mathrm{mod}\ 2)\,.$$

The matrices A and B are generally characterized by K and L nonzero unit elements per row and C and L per column, respectively; all other elements are set to zero. The finite, usually small, numbers K, C, and L define a particular cryptosystem; both matrices are known only to the authorized receiver. Suitable choices of probability p will depend on the maximal achievable rate for the particular cryptosystem as discussed below.

The authorized user may decrypt the received ciphertext $\boldsymbol{J}$ by taking the (mod 2) product $B\boldsymbol{J} = A\boldsymbol{\xi} + B\boldsymbol{\zeta}$. Solving the equation

$$A\boldsymbol{S} + B\boldsymbol{\tau} = A\boldsymbol{\xi} + B\boldsymbol{\zeta} \qquad (\mathrm{mod}\ 2) \qquad (2)$$

is generally computationally hard. However, decryption can be carried out for particular choices of K and L via the iterative methods of belief propagation (BP) [3], where pseudoposterior probabilities for the decrypted message bits, $P(S_i = 1|\boldsymbol{J})$ $1 \le i \le N$ (and similarly for $\boldsymbol{\tau}$), are calculated by solving iteratively a set of coupled equations for the conditional probabilities of the ciphertext bits given the plaintext and vice versa. For details of the method used and the explicit equations see [3].

The unauthorized receiver, on the other hand, faces the task of decrypting the ciphertext $\boldsymbol{J}$ knowing only G and p. The straightforward attempt to try all possible $\boldsymbol{\zeta}$ constructions is clearly doomed, provided that p is not vanishingly small, giving rise to only a few corrupted bits; decomposing G to the matrices A and B is known to be a computationally hard problem [7], even if the values of K, C, and L are known. Another approach to study the problem is to exploit the similarity between the task at hand and the error-correcting model suggested by Sourlas [4], which we will discuss below.

The treatment so far was completely general. We will now make use of insight gained from our analysis of

Gallager-type [5] and Sourlas [6] error-correcting codes to suggest a specific cyptosystem construction and to assess its performance and capabilities. The method used in both analyses [5,6] is based on mapping the problem onto an Ising spin system Hamiltonian, in the manner discovered by Sourlas [4], which enables one to analyze typical properties of such systems.

To facilitate the mapping we employ binary representations (±1) of the dynamical variables $\boldsymbol{S}$ and $\boldsymbol{\tau}$, the vectors $\boldsymbol{J}$, $\boldsymbol{\zeta}$, and $\boldsymbol{\xi}$, and the matrices A, B, and G, rather than the Boolean $(0,1)$ ones.

The *binary* ciphertext $\boldsymbol{J}$ is generated by taking products of the relevant binary plaintext message bits $J_{\langle i_1,i_2,\dots\rangle} = \xi_{i_1}\xi_{i_2}\dots\zeta_{\langle i_1,i_2,\dots\rangle}$, where the indices $i_1, i_2, \dots$ correspond to the nonzero elements of $B^{-1}A$, and $\zeta_{\langle i_1,i_2\dots\rangle}$ is the corresponding element of the corrupting vector (the indices $\langle i_1, i_2\dots\rangle$ correspond to the specific choice made for each ciphertext bit). As we use statistical mechanics techniques, we consider both plaintext (N) and ciphertext (M) dimensionalities to be infinite, keeping the ratio between them N/M finite. Using the thermodynamic limit is quite natural here as most transmitted ciphertexts are long and finite-size corrections are likely to be small.

An authorized user may use the matrix B to obtain Eq. (2). To explore the system's capabilities one examines the Gibbs distribution, based on the Hamiltonian

$$\begin{aligned}\mathcal{H} = &\sum_{\langle i_1,\dots,i_K;j_1,\dots,j_L\rangle} \mathcal{D}_{\langle i_1,\dots,i_K;j_1,\dots,j_L\rangle} \\ &\times \delta[-1; J_{\langle i_1,\dots,i_K;j_1,\dots,j_L\rangle} S_{i_1}\dots S_{i_K}\tau_{j_1}\dots\tau_{j_L}] \\ &- \frac{F_s}{\beta}\sum_{i=1}^{N} S_i - \frac{F_\tau}{\beta}\sum_{j=1}^{M}\tau_j\,. \qquad (3)\end{aligned}$$

The tensor product $\mathcal{D}_{\langle i_1,\dots,i_K;j_1,\dots,j_L\rangle}\mathcal{J}_{\langle i_1,\dots,i_K;j_1,\dots,j_L\rangle}$, where $\mathcal{J}_{\langle i_1,\dots,j_L\rangle} = \xi_{i_1}\xi_{i_2}\dots\xi_{i_K}\zeta_{j_1}\zeta_{j_2}\dots\zeta_{j_L}$, is the binary equivalent of $A\boldsymbol{\xi} + B\boldsymbol{\zeta}$, treating both signal ($\boldsymbol{S}$ and index i) and the corrupting noise vector ($\boldsymbol{\tau}$ and index j) simultaneously. Elements of the sparse connectivity tensor $\mathcal{D}_{\langle i_1,\dots,j_L\rangle}$ take the value 1 if the corresponding indices of both signal and noise are chosen (i.e., if all corresponding elements of the matrices A and B are 1) and 0 otherwise; it has C unit elements per i index and L per j index, representing the system's degree of connectivity. The δ function provides 1 if the selected sites' product $S_{i_1}\dots S_{i_K}\tau_{j_1}\dots\tau_{j_L}$ is in disagreement with the corresponding element $\mathcal{J}_{\langle i_1\dots,j_L\rangle}$, recording an error, and 0 otherwise. Notice that this term is not frustrated, and can therefore vanish at sufficiently low temperatures ($T = 1/\beta \to 0$), imposing the restriction of Eq. (2), while the last two terms, scaled with β, survive. The additive fields F_s and F_τ are introduced to represent our prior knowledge of the signal and noise distributions, respectively.

The random selection of elements in $\mathcal{D}$ introduces disorder to the system which is treated via methods of statistical physics. More specifically, we calculate the partition function $\mathcal{Z}(\mathcal{D}, \boldsymbol{J}) = \mathrm{Tr}_{\{S,\tau\}}\exp[-\beta\mathcal{H}]$, which is then averaged over the disorder and the statistical properties of the plaintext and noise, using the replica method [5,8], to obtain the related free energy $\mathcal{F} = -\langle \ln\mathcal{Z}\rangle_{\xi,\zeta,\mathcal{D}}$. The overlap between the plaintext and the dynamical vector $m = \frac{1}{N}\sum_{i=1}^{N}\xi_i S_i$ will serve as a measure for the decryption success.

Studying this free energy for the case of $K = L = 2$, and in the context of error-correcting codes [5], indicates the existence of paramagnetic and ferromagnetic solutions depicted in the inset of Fig. 1. For corruption probabilities $p > p_s$ one obtains either a dominant paramagnetic solution or a mixture of ferromagnetic ($m = \pm1$) and paramagnetic ($m = 0$) solutions as shown in the inset; thin and thick lines correspond to higher and lower free energies, respectively; dashed lines represent unstable solutions. Lines between the $m = \pm1$ and $m = 0$ axes correspond to suboptimal ferromagnetic solutions. Reliable decryption may be obtained only for $p < p_s$, which corresponds to a spinodal point, where a unique ferromagnetic solution emerges at $m = 1$ (plus a mirror solution at $m = -1$).

The most striking result is the division of the complete space of $\boldsymbol{S}$ and $\boldsymbol{\tau}$ values to two basins of attraction for the ferromagnetic solutions, for $p < p_s$, implying convergence from *any* initialization of the BP equations. Critical corruption rate values for $M/N = 2$ were obtained from the analysis and via BP solutions as shown in Fig. 1, in comparison to the rate obtainable from Shannon's channel capacity [9] (solid line). The priors assumed for both

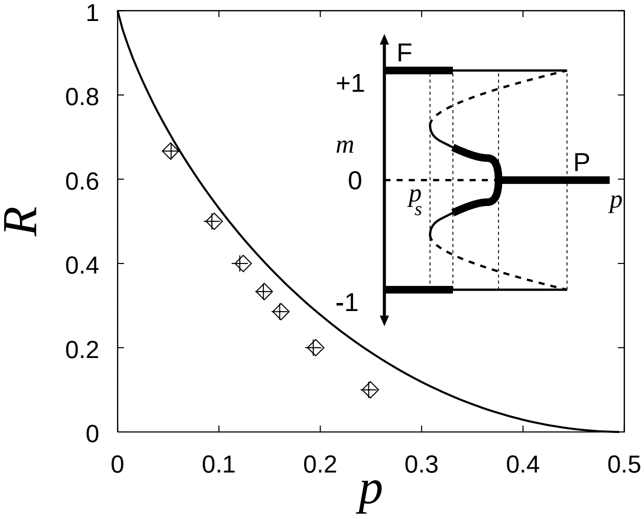

FIG. 1. Critical transmission rate as a function of the corruption rate p for an unbiased ciphertext. Numerical solutions (of the analytically obtained equations, $\Diamond$) and BP iterative solutions (of system size $N = 10^4$, $+$), were averaged over 10 different initial conditions of almost zero magnetization with error bars much smaller than the symbol size. Inset: The ferromagnetic (F) (optimal/suboptimal) and paramagnetic (P) solutions as functions of p; thick and thin lines denote stable solutions of lower and higher free energies, respectively; dashed lines correspond to unstable solutions.

the plaintext (unbiased in this case, $F_s = 0$) and the corrupting vector $[F_\tau = (1/2)\ln[(1-p)/p]]$ correspond to Nishimori's condition [10], which is equivalent to having the correct prior within the Bayesian framework [11].

The initial conditions for the BP-based decryption were chosen almost at random, with a very slight bias of $\mathcal{O}(10^{-12})$ in the initial magnetization, corresponding to typical statistical fluctuation for a system size of 10^{24}. Cryptosystems with other K and L values, e.g., $K, L \geq 3$, seem to offer optimal performance in terms of the corruption rate they accommodate theoretically, but suffer from a decreasingly small basin of attraction as K and L increase. The coexistence of stable ferromagnetic and paramagnetic solutions implies that the system will converge to the undesired paramagnetic solution [5] from most initial conditions which are typically of close-to-zero magnetization. It may still be possible to use successfully specific matrices with higher K and L values (such as in [12]); however, these cannot be justified theoretically and there is no clear advantage in using them.

To conclude, for the authorized user, the $K = L = 2$ cryptosystem offers a guaranteed convergence to the plaintext solution, in the thermodynamic limit $N \rightarrow \infty$, as long as the corruption process has a probability below p_s. The main consequence of finite plaintexts would be a decrease in the allowed corruption rate with little impact on the decoding success.

The task facing the unauthorized user, i.e., finding the plaintext from Eq. (1), was investigated via similar methods by considering the Hamiltonian

$$\mathcal{H} = -\sum_{\langle i_1,\ldots,i_{K'}\rangle} \mathcal{G}_{\langle i_1,\ldots,i_{K'}\rangle} J_{\langle i_1,\ldots,i_{K'}\rangle} S_{i_1}\ldots S_{i_{K'}} - \frac{F_s}{\beta}\sum_{k=1}^{N} S_k\,,$$

using Nishimori's temperature $\beta = (1/2)\ln[(1-p)/p]$. The number of plaintext bits in each product is denoted K', $\boldsymbol{S}$ is the N dimensional binary vector of dynamical variables, and $\mathcal{G}$ is a dense tensor with C' unit elements per index (setting the rest of the elements to zero) and is the binary equivalent of the Boolean matrix G [6]. The latter, together with the statistical properties of the corrupting vector $\boldsymbol{\zeta}$, constitutes the public key used to determine the ciphertext $\boldsymbol{J}$. The last term on the right is required in the case of sparse or biased messages and will require assigning a certain value to the additive field F_s.

The matrix G generated in the case of $K = L = 2$ is dense and has a certain distribution of unit elements per row. The fraction of rows with a low [finite, not of $\mathcal{O}(N)$] number of unit elements vanishes as N increases, allowing one to approximate this scenario by the diluted random energy model [13] studied in [6] where $K', C' \rightarrow \infty$ while keeping the ratio C'/K' finite.

To investigate the typical properties of this (frustrated) model, we calculate again the partition function and the free energy by averaging over the randomness in choosing the plaintext, the corrupting vector, and the choice of the random matrix G (being generated by a product of two sparse random matrices). To assess the likelihood of obtaining spin-glass/ferromagnetic solutions, we calculated the free-energy landscape (per plaintext bit f) as a function of overlap m. This can be carried out straightforwardly using the analysis of [5], and provides the energy landscape shown in Fig. 2. This has the structure of a golf course with a relatively flat area around the one-step replica symmetry breaking (frozen) spin-glass solution and a very deep but extremely narrow area, of $\mathcal{O}(1/N)$, around the ferromagnetic solution. To validate the use of the random energy model we also added numerical data (+, with error bars), obtained by BP, which are consistent with the theoretical results.

This free-energy landscape may be related directly to the marginal posterior $P(S_i = 1|\boldsymbol{J})$ $1 \leq i \leq N$ and is therefore indicative of the difficulties in obtaining ferromagnetic solutions when the starting point for the search is not infinitesimally close to the original plaintext (which is clearly highly unlikely). It is plausible that any local search method, starting at some distance from the ferromagnetic solution, will fail to produce the original plaintext. Similarly, any probabilistic method, such as simulated annealing, will require an exponentially long time for converging to the $m = 1$ solution. Numerical studies of similar energy landscapes show that the time required increases exponentially with the system size [14].

Most attacks on this cryptosystem, by an unauthorized user, will face the same difficulty: without explicit knowledge of the current plaintext and/or the decomposition of G to the matrices A and B it will require an exponentially

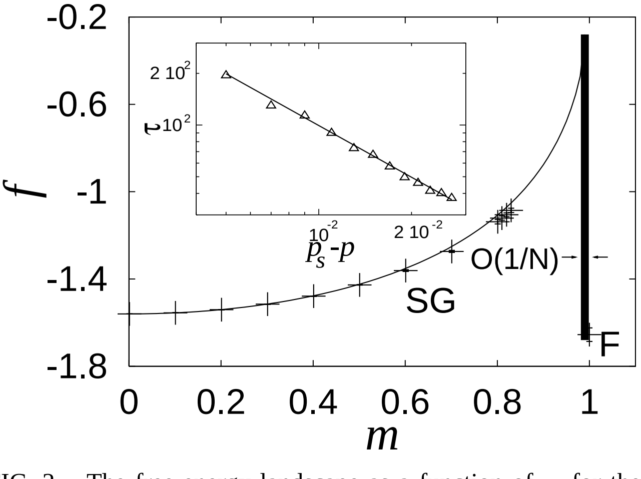

FIG. 2. The free-energy landscape as a function of m for the transmission rate $N/M = 1/2$ and flip rate $p = 0.08$; theoretical values are represented by the solid line; numerical data, obtained by BP ($N = 200$) and averaged over 10 different initial conditions, are represented by symbols (+). The landscape is deep and narrow [of width $\mathcal{O}(1/N)$] at $m = 1$ and rather flat elsewhere. Inset: scattered plot of mean decryption times, τ. The optimal fitting of straight lines through the data provides another indication for the divergence of decryption time for corruption rate close to $p_s = 0.953 \pm 5$ (in this example).

long time to decipher a specific ciphertext. Partial or complete knowledge of the ciphertext and/or plaintext as well as partial knowledge of the matrix B [while $\mathcal{O}(N)$ of the elements remain unknown] will not be helpful for decomposing G which will still require an exponentially long time to carry out.

We will consider here two attacks on specific plaintexts with partial knowledge of the corrupting vector $\boldsymbol{\zeta}$ or of the matrix B. In the first case, knowing $p_a M$ of the pM corrupting bits may allow one to subtract the approximated vector $\hat{\boldsymbol{\zeta}}$ from the ciphertext and take the product of G^{-1} and the resulting ciphertext. This attack is similar to the task facing an unauthorized user with a reduced corruption rate of $(p - p_a)$. For any nonvanishing difference between p_a and p, deciphering the transmitted message remains a difficult task.

A second attack is that whereby the matrix B is known to some degree; for instance, the location of a fraction of the unit elements, say $1 - \rho$ is known. From Eq. (2) one can identify the absent information as having a higher effective corruption level of $p + g(\rho)$, where $g(\cdot)$ is some nontrivial function that depends on the actual scenario. To secure the transmission one may work sufficiently close to the critical corruption level p_s such that the additional effective noise ρ will bring the system beyond the critical corruption rate and into the paramagnetic/spin-glass regime. However, the drawbacks of working *very close* to p_s are twofold: First, average decryption times using BP methods (τ) will diverge proportionally to $1/(p_s - p)$ as demonstrated in the inset of Fig. 2. Second, finite-size effects are expected to be larger close to p_s, which practically means that the system may not converge to the attractive optimal solutions in some cases.

We will end this presentation with a short discussion on the advantages and drawbacks of the suggested method in comparison with existing techniques. First, we point out the differences between this method and the McEliece cryptosystem. The latter is based on Goppa codes and is limited to relative low corruption levels. These may allow for decrypting the ciphertext using (many) estimates of the corruption vector. Our suggestion allows for a significant corruption level, thus increasing the security of the cryptosystem. In addition, the suggested construction, $K = L = 2$, is not discussed in the information theory literature (e.g., in [3]) which typically prefers higher K, L value systems. Second, in comparison to RSA where decryption takes $\mathcal{O}(N^3)$ operations, our method requires only $\mathcal{O}(N)$ operations, multiplied by the number of BP iterations (which is typically smaller than 100 for most system sizes examined except very close p_s). Encryption costs are of $\mathcal{O}(N^2)$ (as in RSA) while the inversion of the matrix B is carried out only once and requires $O(N^3)$ operations.

The two obvious drawbacks of our method are the following: (1) the transmission of the public key, which is a dense matrix of dimensionality $M \times N$. However, as public-key transmission is carried out only once for each user we do not expect it to be of great significance. (2) The ciphertext to plaintext bit ratio is greater than one to allow for corruption, in contrast to RSA where it equals 1. Choosing the N/M ratio is in the hands of the user and is directly related to the security level required; we therefore do not expect it to be problematic as the increased transmission time is compensated by a very fast decryption.

We examine the typical performance of a new cryptosystem, based on insight gained from our previous studies, by mapping it onto an Ising spin system; this complements the information theory approach which focuses on rigorous worst-case bounds. We show that authorized decryption is fast and simple while unauthorized decryption requires a prohibitively long time. Important aspects that are yet to be investigated include finite-size effects and methods for alleviating the drawbacks of the new method.

Support by JSPS-RFTF (Y. K.), The Royal Society, and EPSRC-GR/L52093 (D. S.) is acknowledged. We thank Manfred Opper and Hidetoshi Nishimori for reading the manuscript.

[1] R. McEliece, JPL-Caltech, California, DSN Progress Report No. 42-44, 1978, p. 114.

[2] R. G. Gallager, *Low Density Parity Check Codes,* Research Monograph Series Vol. 21 (MIT Press, Cambridge, 1963); IRE Trans. Inf. Theory **IT-8**, 21 (1962).

[3] D. J. C. MacKay, IEEE Trans. Inf. Theory **45**, 399 (1999).

[4] N. Sourlas, Nature (London) **339**, 693 (1989).

[5] Y. Kabashima, T. Murayama, and D. Saad, Phys. Rev. Lett. **84**, 1355 (2000).

[6] Y. Kabashima and D. Saad, Europhys. Lett. **45**, 97 (1999).

[7] M. R. Garey and D. S. Johnson, *Computers and Intractability* (Freeman, San Francisco, 1979), p. 251.

[8] K. Y. M. Wong and D. Sherrington, J. Phys. A **20**, L793 (1987).

[9] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).

[10] H. Nishimori, Prog. Theor. Phys. **66**, 1169 (1981).

[11] N. Sourlas, Europhys. Lett. **25**, 159 (1994).

[12] I. Kanter and D. Saad, Phys. Rev. Lett. **83**, 2660 (1999).

[13] B. Derrida, Phys. Rev. B **24**, 2613 (1981).

[14] E. Marinari, G. Parisi, and F. Ritort, J. Phys. A **27**, 7615 (1994); **27**, 7647 (1994); E. Marinari and F. Ritort, J. Phys. A **27**, 7669 (1994).