On Iterated Logarithm Solutions to Identification Protocols

Hervé Chabanne*†
*Institut Télécom
Télécom ParisTech
Paris, France

Gérard Cohen*

Bruno Kindarji*†
† Sagem Sécurité
Osny, France

***Abstract*—We show in this paper that when some constraints are relaxed in the League Problem, we can obtain a communicating scheme in $1 + \lceil \log\log\log n \rceil$ bits exchanged where n is the number of teams involved in the league. This contrasts with previous results due mainly to Ortlitsky. The price we have to pay is that, with a vanishing error probability, a wrong decision is made. We also give some extensions and applications to our setting.**

I. Introduction

In 1990, Orlitsky introduced the following *League Problem* [7]. There are n soccer teams. Agata knows that Palerme and Juventus played against each other. Benedetto hears the name of the winning team, but not Agata. Unfortunately, he did not get the name of the loser. The League Problem asks for how many bits must Benedetto and Agata exchange in the worst case for Agata to find out who won? Whenever two interactions are allowed, there is a solution where Agata sends the index of the first bit in which the name of the two teams differ and Benedetto responds with this bit. One can prove that $\lceil \log\log n \rceil + 1$ bits are optimal [7], even in the case where more than two messages are exchanged.

We here recall a different construction wich used identification codes to solve this League Problem. Identification codes have been introduced by Ahlswede and Dueck in [1] to permit to Agata to know if Benedetto sends a message indicating that a particular team (for instance, say Palerme, Agata's favorite team) has won. These codes demand less bits than the traditional transmission codes which convey more general messages of the type "What team did Benedetto send?". With these identification codes, Agata can make two types of mistakes. First, she can believe that the team identified was not the one sent by Benedetto when it was. Second, she can conclude that this team was the one sent by Benedetto when it was not.

Coming back to our League Problem, if we now allow Agata to sometimes take wrong decisions as it is the case when using identification codes, we can exhibit as the main result of this paper a solution where only $\log\log\log n$ bits are needed.

II. The League Problem: a Specific Case of Two-Way Communications

A. Two-Way Communications

Orlitsky [8] explored many aspects of communication between two players. Player A knows X, player B knows Y, and we want to design a communication protocol between A and B such that, at the end of the protocol, A knows $f(X, Y)$ where f is a given function. The goal of this protocol is for A and B to send as few messages as possible.

This problem is pretty generic and obviously depends on the function f. We here suppose that n is a parameter for the length of the data X, Y (for example, Y is an element of $\{0,1\}^n$), and we look for – asymptotic – optimal communication protocols. An upper bound on the number of bits to be sent is $\log n$, as it suffices that B sends Y to A for A to be able to compute the result. We focus here on a specific problem.

As stated in the introduction, we explore the different existing possibilities for the League Problem, before exhibiting a new solution which, while allowing some errors on the result, outperforms the previously existing solutions.

B. Problem Statement

In a well-known league, n teams $t_1, \ldots, t_n$ are competing, until the final match where team t_α and team t_β are to play against each other. Agata knows t_α and t_β, but misses the result of the game. Benedetto knows who is the winner t, but not who was his opponent. How can Agata and Benedetto communicate so that Agata gets the result without using the channel more than is necessary?

We assume that the channel between Agata (A) and Benedetto (B) is two-way and noiseless, so that each sent bit is correctly received. We also assume that the ordering of the teams is known and shared between the two partners. In the following, $\log$ denotes the binary logarithm.

C. Practical Solutions

The trivial solution, without any interaction, is for Benedetto to send the name of the winning team to Agata. This takes $\lceil \log n \rceil$ bits to transmit, and is optimal in the lossless case (if Benedetto can transmit his message in k bits, in a lossless way, there is an injection between $\{0,1\}^k$ and $\{1, \ldots, n\}$; thus $2^k \geq n$).

If we allow interaction between A and B, then Orlitsky showed [7] a solution in $O(\log\log n)$ bits. First, A sends the position where the bit strings representing t_α and t_β differ - which takes $\lceil \log\log n \rceil$ bits, then B replies with the actual value of this bit – thus a $1 + \lceil \log\log n \rceil$-long solution. This solution is also shown to be optimal.

The problem widens if we allow some error to be made within a controlled probability. Let λ be the probability of the event "after the communication, A is mistaken about the winning team". This is referred to in [9] as "the ϵ-randomized model". The case $\lambda = 0$ leads to the previous results; we show in the following that $\lambda > 0$ leads to new interesting results.

D. Existing Bounds

The optimal solutions of this problem, as stated in Section II-A, satisfy some strict boundaries showed in [7]. Reusing the notations employed in that article, we note $C_m(Y|X)$ the *m-message complexity of Y knowing X*, *i.e.* the minimal number of bits required to transmit Y to a person who knows X, with m messages sent over the channel. Here, m is a natural number ($m \geq 1$). $C_m(Y|X)$ is a decreasing sequence, whose limit is noted $C_\infty(Y|X)$.

Note that $C_m(Y|X)$ only denotes the case where A knows without any doubt Y at the end of the protocol. In the case where A knows Y with probability $1 - \lambda$, the corresponding quantity is noted $C_m^\lambda(Y|X)$.

With these notations, several bounds can be found in [7], among which we highlight the following two:

- (1) $C_\infty(Y|X) \geq \lceil \log C_1(Y|X) \rceil + 1$ with equality in the case of the League Problem;
- (2) $C_1^\lambda(Y|X) \leq 4C_\infty(Y|X) + 2\log\frac{1}{\lambda}$

This shows that the League Problem is well studied in the case of an exact result. Our work aims at improving the second bound in that specific case, in showing that allowing vanishing errors in the result enables to reduce the communication cost by a logarithmic factor. Moreover, we derive an inequality similar to the first one in the error case.

III. Identification codes

A. Definition

Informally speaking, an identification is a data representation which enables a receiver Bob to know, within a given error probability, if Alice sent a message $i \in \{1, \ldots, N\}$, or not. To be more specific, the following definition is commonly adopted.

Definition 1 (Identification Code, [1]): Let $\mathcal{X}, \mathcal{Y}$ be two alphabets, and W^n a channel from $\mathcal{X}^n$ to $\mathcal{Y}^n$. A $(n, N, \lambda_1, \lambda_2)$-identification code from $\mathcal{X}$ to $\mathcal{Y}$ is given by a family $\{(Q(\cdot|i), \mathcal{D}_i)\}_{i\in\{1,\ldots,N\}}$ where:

- $Q(\cdot|i)$ is a probability mass function over $\mathcal{X}^n$, that encodes i,
- $D_i \subset \mathcal{Y}^n$ is the decoding set,
- λ_1 is the first-kind error rate, with $\lambda_1 \geq \sum_{x^n\in\mathcal{X}^n} Q(x^n|i)W^n(\overline{D_i}|x^n)$
- λ_2 is the second-kind error rate, with $\lambda_2 \geq \sum_{x^n\in\mathcal{X}^n} Q(x^n|j)W^n(D_i|x^n)$

for all $i, j \in \{1, \ldots, N\}$ such that $i \neq j$.

The first-kind error rate denotes the probability for a transmitted message not to be identified, and the second-kind error rate is the probability for a transmitted message to be falsely identified.

The relevant rate to consider in such a case is the *Identification Rate*, defined as $R_{ID} = \frac{1}{n}\log\log N$. The following theorem was shown in [1]:

Theorem 1 ([1]): Let κ be the capacity of the channel W. Let $\epsilon > 0$.

- For each $0 < \lambda_1, \lambda_2 \leq 1$, there exist n, N and an $(n, N, \lambda_1, \lambda_2)$-identification code such that $\frac{1}{n}\log\log N \geq \kappa - \epsilon$;
- If there exists an $(n, N, \lambda_1, \lambda_2)$-identification code with $\lambda_1, \lambda_2 \leq 2^{-n\epsilon}$, then the rate of this code is such that $\frac{1}{n}\log\log N \leq \kappa$.

This theorem basically shows that for a given channel, the transmission capacity is the same as the identification capacity.

B. Constructing Identification Codes

There exist few constructions of identification codes. [1] use constant-size sets as a general frame-work for identification codes. This idea was then applied by [10], [5], in constructions using constant-size codes as an instance of [1]. Another construction, based on prime numbers, is given in [2]. Finally, [6] designs an identification code based on Reed-Solomon codes, thus showing that it is possible to design such an ID-code thanks to the minimal distance of an error-correcting code.

C. Using Identification Codes to solve the League Problem

A first way of solving this problem is to use Identification Codes. Instead of going through the two-round communications, B directly sends an identification tag for the winning team. As A must choose between two teams, she must check whether the received tag is identifying t_α of t_β.

To successfully achieve this goal, A and B agree beforehand on an $(m, n, \lambda_1, \lambda_2)$-identification code, where n is the number of teams and m the number of bits to be transmitted. As A knows t_α and t_β, she sets her target on $\mathcal{D}_{t_\alpha}$, then listens to B. Then B picks a message x^m according to $Q(\cdot|t)$ and sends it to A, who checks whether $x^m \in \mathcal{D}_{t_\alpha}$.

To evaluate the error probability of such a construction, consider the following: either $t = t_\alpha$, or $t = t_\beta$. In the first case, the probability for A not to read t_α in x^m is $Q(\overline{\mathcal{D}_{t_\alpha}}|t_\alpha)$, which is smaller than λ_1. In the second case, the probability for A to read t_α anyways is $Q(\mathcal{D}_{t_\alpha}|t_\beta)$, which is smaller than λ_2. The overall error probability is thus $\lambda \leq \frac{\lambda_1+\lambda_2}{2}$.

Note that, according to Theorem 1, there exist identification codes such that m is about $\frac{1}{\kappa}\log\log n$. We therefore obtain a communication protocol for the League Problem in $O(\log\log n)$ bits.

IV. Achieving a Triple-Log League Solution

We now allow two-way communications between A and B. In the errorless case, this reduced the communication

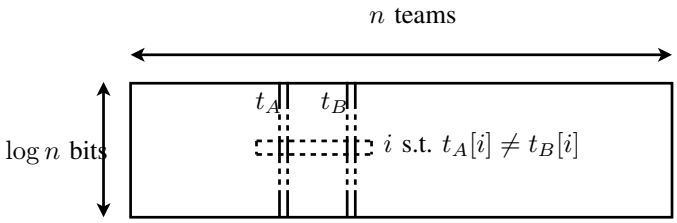

Figure 1. Lossless representation of n teams, and resulting two-way communication

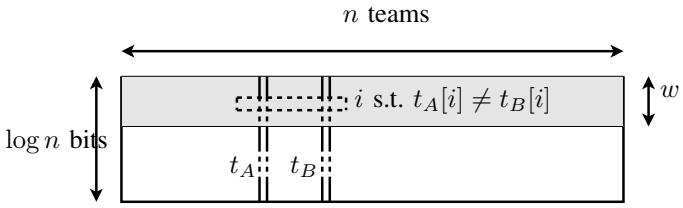

Figure 2. Representation of n teams on $w < \log n$ bits.

complexity from $O(\log n)$ to $O(\log \log n)$. We here show that if we allow errors, we reduce the communication complexity from $O(\log \log n)$ to $O(\log \log \log n)$.

A. Going one Step Further

Our proposal starts with the original protocol from Orlitsky. To achieve the optimal two-way communication, [7] represents the set of all teams according to an entropic coding, as is illustrated in Figure 1.

If we wish to achieve communication while enabling a (small) error probability, it suffices to relax the representation of Figure 1, and reduce the number of bits needed to represent each team from $\log n$ to w, see Figure 2. In doing so, A needs only send $\log w$ bits to B. Let h be an "entropic" hash function from $\{1, \ldots, n\}$ to $\{0,1\}^w$; for example, $h(x)$ takes the first w bits of the representation of x in $\lceil \log n \rceil$ bits.

Indeed, as n elements are represented with a set of 2^w elements, for each w-long bit string, there will be $\frac{n}{2^w}$ elements which have the same representation. However, as we only wish to distinguish between any two elements, the probability for t_α and t_β to have the same representation is $\frac{1}{2^w}$.

From this fact, we deduce the probability for the protocol to fail, *i.e.* the probability for A not to correctly guess t between $\{t_\alpha, t_\beta\}$:

$$\Pr[\text{fail}] = \Pr[\text{fail}|h(t_\alpha) = h(t_\beta)] + \Pr[\text{fail}|h(t_\alpha) \neq h(t_\beta)]$$

As the protocol is always successful when t_α and t_β have different representations, we find the probability of error to be $\Pr[\text{fail}] = \frac{1}{2^{w+1}}$.

B. Triple-log Solution with Vanishing Error Probability

In the specific case where $w = \lceil \log \log n \rceil$, the protocol takes an overall length of $\lceil \log \log \log n \rceil + 1$ bits of communications with $\Pr[\text{fail}] = \frac{1}{2 \log n}$.

This error probability might seem non-negligible if stated under those terms - as $\frac{1}{\log x}$ slowly converges to 0. However, we emphasize the fact that this enables to solve the League Problem with a huge number of competing teams with very small communication between A and B.

Another way of stating this is that by sending $m+1$ bits over a channel, it is possible to identify $2^{2^{2^m}}$ teams, with an error probability of only $\frac{1}{2^{2^m+1}}$, which is actually negligible.

This result also beats the bound of Inequality (2): $C_1^\lambda(Y|X) \leq 4C_\infty(Y|X) + 2\log\frac{1}{\lambda}$. In this case, the upper bound is equal to $6\lceil \log \log n \rceil + 4$, which is still greater than $\lceil \log \log \log n \rceil + 1$.

C. Revisiting the Two-Way Communication Paradigm

Using the notations introduced in Section II-D, we here show that the triple-log result for C_∞^λ obtained is coherent with the double-log communication result for C_1^λ. This is shown in the following theorem:

Theorem 2: For all (X, Y) pairs, for all $0 < \lambda < 1$, the following inequality holds:

$$C_2^\lambda(Y|X) \geq \left\lceil \log C_1^\lambda(Y|X) \right\rceil$$

Proof. It is similar to that of Inequality (1): we formalize a two-way protocol in this fashion:

- A sends a – possibly randomized – message $\sigma_A = \mathsf{createMessage}(X)$,
- B receives σ_A and replies with $\sigma_B = \mathsf{reply}(Y, \sigma_A)$, which can also be randomized;
- A receives σ_B and deduces $Y' = \mathsf{deduce}(X, \sigma_A, \sigma_B)$ such that $\Pr[Y' = Y] \geq 1 - \lambda$.

Assume that messages σ_A have (maximal) length l_A and messages σ_B have (maximal) length l_B. In this case, $f_B = \mathsf{reply}(Y, \cdot)$ is a function from $\{0,1\}^{l_A}$ to $\{0,1\}^{l_B}$, which enables to determine Y with probability $1 - \lambda$.

The graph of f_B is the set of all (σ_A, σ_B) such that $f_B(\sigma_A) = \sigma_B$, thus a subset of $\{0,1\}^{l_A} \times \{0,1\}^{l_B}$, and can be represented as a subset of $\{0,1\}^{l_A + l_B}$, *i.e.* an element of $\{0,1\}^{2^{l_A+l_B}}$.

In order to transform a two-ways protocol into a one-way protocol, it suffices for B to fully send his function reply, which he can do in $2^{l_A+l_B}$ bits.

Then, A can compute σ_A using createMessage, apply it to f_B, and deduce Y' such that $\Pr[Y' = Y] \geq 1 - \lambda$.

This shows that if there exists a protocol with 2-message complexity C, then there exist a protocol with 1-message complexity 2^C; thus $C_2^\lambda(Y|X) \geq \left\lceil \log C_1^\lambda(Y|X) \right\rceil$. □

Remark 1: For the sake of clarity we dealt with protocols that only use 2 messages, but in fact our theorem can easily be extended to any number of messages greater than 2, using a function f_B which depends not only on σ_A, but on all previously sent messages. This shows that $C_\infty^\lambda(Y|X) \geq \left\lceil \log C_1^\lambda(Y|X) \right\rceil$.

As we showed that it is possible, using identification codes, to solve the League Problem with errors with a one-way communication cost of $\lceil \log \log(n) \rceil + 1$, applying Theorem 2 shows that our result, namely a two-ways protocol for the League Problem with communication cost $\lceil \log \log \log n \rceil + 1$, is coherent.

D. Double-log One-Way Solution with Vanishing Error Probability Without Identification Codes

The result of the previous section incites us to apply the proof of Theorem 2 in order to find an efficient solution for the League Problem, in only one message.

Actually, instead of sending the graph of the function f_B as previously defined, it suffices to send, in an equivalent way, the first $w = \lceil \log \log n \rceil$ bits of the winning team t.

As the receiver A has only the choice between two teams, she fails exactly when both teams have the same $\lceil \log \log n \rceil$ first bits. This happens with probability $2^{-\lceil \log \log n \rceil} \approx \frac{1}{\log n}$.

This shows that the League Problem has a trivial one-way solution in $\lceil \log \log n \rceil$ with vanishing error-probability.

V. Unlocking Possible Extensions with Coding Theory

A. Communicating over a Noisy Channel

The problem of communicating over a noisy channel was introduced by Shannon and is well-known. Given two alphabets $\mathcal{X}$ and $\mathcal{Y}$, a channel from $\mathcal{X}^m$ to $\mathcal{Y}^m$ is a mass function $W : \mathcal{X}^m \times \mathcal{Y}^m \to [0, 1]$ which defines the output of a message x^m. The channel is noisy if W cannot be represented as the identity function.

Transmitting information over such a channel is always possible at a given rate if this rate is lower than the capacity of the channel $\kappa(W)$. This means in practice that in order to transmit k bits of information, one must send at least $m = k/R$ bits where $R < \kappa$.

Finding the optimal data structure to communicate over a noisy channel is still an open problem, however beyond the scope of this paper. In the following, we shall assume that the channel noise is overcome by classical coding techniques, and thus focus only on the problem of the information to transmit.

B. A League Problem with More than 2 Competing Teams

Consider a generalization of the initial problem, where Agata misses the result of the game between t_α and t_β, to the following: In the universe of the n teams competing, the final round involved $s + 1 \geq 2$ teams. How can now A get from B the identity of the winner? A trivial solution is to call $\binom{s+1}{2}$ times the initial ($s = 1$) protocol. One can however get a linear (in s) solution by making use of *separating* codes (see [3]), defined as follows:

Definition 2: Let Q be an alphabet of size q, s, u integers. A subset $C \subset Q^m$ is (s, u)-*separating* if for any two disjoint subsets S, U of C with $|S| = s$, $|U| = u$, there is some coordinate $i \in \{1, \ldots, m\}$ such that for any $x \in S$ and any $y \in U$, we have $x_i \neq y_i$.

We only need here a specialization to the case $q = 2, u = 1$.

There exist asymptotic families of $(s, 1)$-separating codes with rate $R_s > 0$. An existential proof is easy to come up with; for constructions, one can resort to algebraic geometry codes on large alphabets (see, *e.g.* [11]) and then concatenate to get binary codes. We do not elaborate on this topic here, since we only need to achieve a non zero rate R_s for our purpose.

The idea is the following: encode $n = 2^{R_s m}$ binary sequences (teams) on m bits using such a code: then, for any ordered $(s + 1)$-subset of teams $(t_{i_1}, ...t_{i_{s+1}})$, there exists an index $j \in \{1, \ldots, m\}$ such that the j-th bit of t_{i_1} is 0 and all others t_i's have a 1, or the opposite. When A asks B for this bit, she identifies t_{i_1}; calling this protocol at most $s+1$ times is enough.

C. Real-Life Application

In this section, we give an example of application which makes use of our results. A Smartdust [4] is a network of small micro-electromechanical systems equipped with wireless communications. Imagine that a cloud of n Smartdust is released over a geographical zone. Some sensors are installed in this zone. During a kind of system setup, the sensors collect the identities of the different specks of Smartdust in their area of listening. We assume that each sensor possesses at most $s + 1 < n$ specks in their area of listening. Using Section V-B, we encode each identifier t_i on $m = \lceil \frac{1}{R_s} \log n \rceil$ bits.

After that, the sensors periodically want to verify if a given speck of Smartdust is still working. We can imagine that sensors have to reduce the length of communications to a minimum; for instance to save needed energy of transmission.

To test the liveness of an element noted $t_e \in \{t_{i_1}, \ldots, t_{i_{s+1}}\}$, using Section V-B, sensors compute the index $j \in \{1, \ldots, m\}$ such that $t_e[j]$ is different from the other $t[j]$ for $t \in \{t_{i_1}, \ldots, t_{i_{s+1}}\}$. They then broadcast a message of type: $(j, t_e[j])$ where j is encoded over $\log m$ bits. Note that the total size of the message is $1 + \lceil \log \frac{\log n}{R_s} \rceil$. Each node of Smartdust which receives the message checks whether it is the one which has to answer to the sensor. In this case, it emits an acknowledgement sequence.

VI. Conclusion

In this paper, we show that allowing vanishing errors into the determination of the results can save an extra log factor in the communication cost of the League Problem. More generally, denoting by $\log^{(i)} n$ the i-th iterated logarithm, a straightforward extension of the results in Section IV. B yields that, if we code in length $w = \lceil \log^{(i)} n \rceil$, then the overall protocol length will be in $\lceil \log^{(i+1)} n \rceil$ and the error-probability less than $1/\log^{(i-1)} n$.

In the future, we will investigate the relationships between identification codes and interactive communications. In particular, we will further study the applications of our results to pervasive systems such as RFID tags or Smartdust.

References

[1] R. Ahlswede and G. Dueck. Identification via channels. *Information Theory, IEEE Transactions on*, 35(1):15–29, Jan 1989.

[2] R. Ahlswede and B. Verboven. On identification via multiway channels with feedback. *Information Theory, IEEE Transactions on*, 37(6):1519–1526, Nov 1991.

[3] G.D. Cohen and H.G. Schaathun. Upper bounds on separating codes. *Information Theory, IEEE Transactions on*, 50(6):1291–1294, June 2004.

[4] Joseph M. Kahn, Y Howard Katz, and Kristofer S. J. Pister. Emerging challenges: Mobile networking for Smart dust. *Journal of Communications and Networks*, 2:188–196, 2000.

[5] K. Kurosawa and T. Yoshida. Strongly universal hashing and identification codes via channels. *Information Theory, IEEE Transactions on*, 45(6):2091–2095, Sep 1999.

[6] Pierre Moulin and Ralf Koetter. A framework for the design of good watermark identification codes. In Edward J. Delp III and Ping Wah Wong, editors, *SPIE*, volume 6072, page 60721H. SPIE, 2006.

[7] A. Orlitsky. Worst-case interactive communication. i. two messages are almost optimal. *Information Theory, IEEE Transactions on*, 36(5):1111–1126, Sep 1990.

[8] Alon Orlitsky. Worst-case interactive communication - ii: Two messages are not optimal. *IEEE Transactions on Information Theory*, 37(4):995–1005, 1991.

[9] King F. Pang and Abbas El Gamal. Communication complexity of computing the Hamming distance. *SIAM Journal on Computing*, 15(4):932–947, 1986.

[10] S. Verdu and V.K. Wei. Explicit construction of optimal constant-weight codes for identification via channels. *Information Theory, IEEE Transactions on*, 39(1):30–36, Jan 1993.

[11] Chaoping Xing. Asymptotic bounds on frameproof codes. *Information Theory, IEEE Transactions on*, 48(11):2991–2995, Nov 2002.