Technical Report CUCS-034-93

Solution of Ulam's Problem on Binary Search with Four Lies

Vincenzo Auletta* **Alberto Negro***

Giuseppe Parlati†

November 5, 1993

Abstract

In this paper we determine the minimal number of yes-no queries needed to find an unknown integer between 1 and 1000000 if at most four of the answers may be erroneous.

1 Introduction

S.M. Ulam, in his autobiography [9], suggested an interesting two–person search game which can be formalized as follows: a Responder chooses an element x in $\{1, 2, \ldots, 1000000\}$ unknown to a Questioner. The Questioner has to find it by asking queries of the form "$x \in Q$?", where Q is an arbitrary subset of $\{1, 2, \ldots, 1000000\}$. The Responder provides "yes" or "no" answers, some of which may be *erroneous* .

In [1], Berlekamp , studing Binary Simmetric Channels with Feedback, introduced two very useful concepts: the State of the problem when some questions have been already answered, and the Weight of a state. All the known results are fundamentally based on these concepts.

At present several solutions to Ulam's problem and its generalization are known.

In particular, Rivest and alt. [6] gave an asymptotically optimal solution to Ulam's problem in the continuous case (i.e. when the search space is the real interval [0,1]). Solutions to the original Ulam's problem were given by Pelc [5], Czyzowicz, Mundici and Pelc [2], Negro and Sereno [7]. They proved that the minimal number of questions to guess a number in the range $\{1, 2, \ldots, 1000000\}$ is 25, 29, 33 when up to 1, 2 or 3 errors are allowed respectively.

The generalized form of the problem, i.e. when the search space is $\{1, 2, \ldots, n\}$, $n \geq 2$ has also been studied.

In [5], Pelc proved that when at most 1 error is possible, q questions are sufficient iff either $n(q+1) \leq 2^q$ if n is even, or $n(q+1) - (q-1) \leq 2^q$ if n is odd. For the case with two errors, Czyzowicz and alt. [3] proved that, when n is a power of 2, q questions are sufficient

*Dipartimento di Informatica ed Applicazioni, Università di Salerno, 84081 Baronissi (SA) - Italy.

†Department of Computer Science, Columbia University, New York, N.Y. 10027 - USA.

iff $n(q^2+q+1) \leq 2^q$. For general n Guzicki [4] proved that q questions are sufficient iff $n(q^2+q+1) \leq 2^q$. Some exceptional values of n requires one more question. In [8] Negro and Sereno proved that when $n = 2^m$, $m \geq 0$, the minimum number of questions is the lower bound $q = min_i\{i: \; n\left(\binom{q}{2}\right) \leq 2^i\}$ [6] if $m = 4$ or $m \geq 6$. One more question is required in the other cases.

In this paper we prove that when up to 4 lies are allowed the solution of Ulam's problem is 37, the lower bound given in [6].

2 Notations and definitions

A game is considered between two players: the Questioner and the Responder. The Responder chooses an element $x \in \mathcal{N}$, where $\mathcal{N} = \{1, 2, \ldots, 1000000\}$, unknown to the Questioner who has to guess the element with queries of form "$x \in Q?$" where $Q \subset \mathcal{N}$. We want to find a Questioner's strategy for searching x using the minimal number of questions, when the Responder can lie at most four times.

Suppose the n-tuple $Q = Q_1, Q_2, \ldots, Q_n$ of yes-no questions has already been answered. The state of the Questioner knowledge can be summarized by the unique quintuple (A, B, C, D, E) of subsets of $\mathcal{N}$ with the following properties:

- $x \in A$ iff none of the answers is a lie;
- $x \in B$ iff exactly one of the answers is a lie;
- $x \in C$ iff exactly two of the answers are lies;
- $x \in D$ iff exactly three of the answers are lies.
- $x \in E$ iff exactly four of the answers are lies.

In the following we define a quintuple (A, B, C, D, E) as an Ulam set.

Now assume that the "$x \in Q?$" question is asked, where $Q = X \cup Y \cup Z \cup K \cup T$, and $X \subseteq A$, $Y \subseteq B$, $Z \subseteq C$, $K \subseteq D$, $T \subseteq E$. A positive answer trasforms (A, B, C, D, E) into the quintuple:

$(A, B, C, D, E)Q^{yes} = (A \cap Q, (B \cap Q) \cup (A \cap \overline{Q}), (C \cap Q) \cup (B \cap \overline{Q}), (D \cap Q) \cup (C \cap \overline{Q}), (E \cap Q) \cup (D \cap \overline{Q})$.

The state $(A, B, C, D, E)Q^{no}$ can be obtained in the same way, exchanging Q and $\overline{Q}$.

It can be easily seen that the components of an Ulam set are pairwise disjoint.

For the sake of clearness in the following we denote the sets by their cardinalities. In particular we denote an Ulam set as

$$U = (a, b, c, d, e)$$

where $a = |A|$, $b = |B|$, $c = |C|$, and $d = |D|$, $e = |E|$. Moreover if $Q = X \cup Y \cup Z \cup K \cup T$ is the set involved in the yes-no question we can say

$$Q = (x, y, z, k, t)$$

and

$$UQ^{yes} = (x, a - x + y, b - y + z, c - z + k, d - k + t),$$
$$UQ^{no} = (a - x, x + b - y, y + c - z, z + d - k, k + e - t).$$

Definition 1 *An Ulam set U is n-solvable iff*

- $n = 0$ *and* $(a + b + c + d + e) \leq 1$*;*
- *there is a yes-no question such that both UQ^{yes} and UQ^{no} are $(n-1)$-solvable.*

Following Berlekamp's idea [1], we define the weight of each state $U = (A, B, C, D, E)$ as follows:

Definition 2 *Let $U = (a, b, c, d, e)$ be the Ulam set when q questions remain to be asked. The* weight *of U is:*

$$w_q(U) = a\left(\binom{q}{4}\right) + b\left(\binom{q}{3}\right) + c\left(\binom{q}{2}\right) + d\left(\binom{q}{1}\right) + e\left(\binom{q}{0}\right)$$

where $\left(\binom{n}{m}\right) = \sum_{i=0}^{m} \binom{n}{i}$.

Definition 3 *For any state $U = (a, b, c, d, e)$, the character of U is defined as*

$$ch(U) = min_i\{i : w_i(U) \leq 2^i\}.$$

The character of a state represents a lower bound to the number of questions needed to solve that state. It is worth characterizing the states that can be solved optimally.

Definition 4 *The Ulam set $U = (a, b, c, d, e)$ is called* nice *iff the Questioner has a winning strategy in $ch(U)$ questions starting from U.*

Proposition 1 *Let $U = (a, b, c, d, e)$ be an Ulam set and $n \in \mathcal{N}$. If U is n-solvable then:*

1. *U is (n+1)-solvable;*
2. $2^n \geq w_n(U)$*;*
3. *if $U' = (a', b', c', d', e')$ is another Ulam set, and $a' \leq a$, $b' \leq b$, $c' \leq c$, $d' \leq d$, $e' \leq e$, then U' is solvable too.*

Proof.

1) By induction on n. If $n = 0$, then by definition $(a + b + c + d + e) \leq 1$. If we choose the question $Q = \mathcal{N}$ then $UQ^{yes} = (a, b, c, d, e)$, and $UQ^{no} = (0, a, b, c, d)$ which are both 0-solvable. For the induction step we can use the same technique.

2) By induction on n. The case $n = 0$ is trivial. Assume that U is (n+1)-solvable, and let $Q \subset \mathcal{N}$ be a yes-no question such that both UQ^{yes} and UQ^{no} are n-solvable. By the inductive hypothesis we have $2^n \geq w_n(UQ^{yes})$ and $2^n \geq w_n(UQ^{no})$. Since the Ulam sets are disjoint, and using $\left(\binom{n+1}{m}\right) = \left(\binom{n}{m}\right) + \left(\binom{n}{m-1}\right)$, we have $2^{n+1} \geq w_{n+1}(U)$.

3) The proof can be found in [2]. ∎

3 The main result

Proposition 2 *Let* $U = (0, 1, m, \binom{m}{2}, \binom{m}{3})$ *be an Ulam set with* $m \geq 6$. *Then* U *is nice.*

Proof. It can be found in [8]. ■

Proposition 3 *Let* $U_n = \left(0, 2^n, (8-n)2^n, \binom{8-n}{2}2^n, \binom{8-n}{3}2^n\right)$ *be an Ulam set, with* $0 \leq n \leq 8$. *Then* U_n *is* $(10+n)$*-solvable.*

Proof. By induction on n.
Induction basis: For $n = 0$ is true. In fact $U_0 = \left(0, 1, 8, \binom{8}{2}, \binom{8}{3}\right)$ is 10-solvable by proposition 2, because $ch(U_0) = 10$.
Induction step: let $n + 1 \leq 8$,

$$U_{n+1} = \left(0, 2^{n+1}, (8-(n+1))2^{n+1}, \binom{8-(n+1)}{2}2^{n+1}, \binom{8-(n+1)}{3}2^{n+1}\right).$$

If we set $Q = \left(0, 2^n, (8-n)2^n, \binom{8-n}{2}2^n, \binom{8-n}{3}2^n\right)$ then the state after a positive anwser is :

$$U_{n+1}Q^{yes} = \left(0, 2^n, (8-n)2^n, \binom{8-n}{2}2^n, \binom{8-n}{3}2^n\right).$$

Simmetrically we can obtain $U_{n+1}Q^{no} = U_{n+1}Q^{yes}$, exchanging Q with $\overline{Q}$.
Therefore, U_{n+1} is $(10 + (n+1))$-solvable, as required. ■

Proposition 4 *Let* $U_n = \left(0, 2^n, (12-n)2^n, \binom{12-n}{2}2^n, \binom{12-n}{3}2^n\right)$ *be an Ulam with* $0 \leq n \leq 12$. *Then* U_n *is* $(11+n)$*-solvable.*

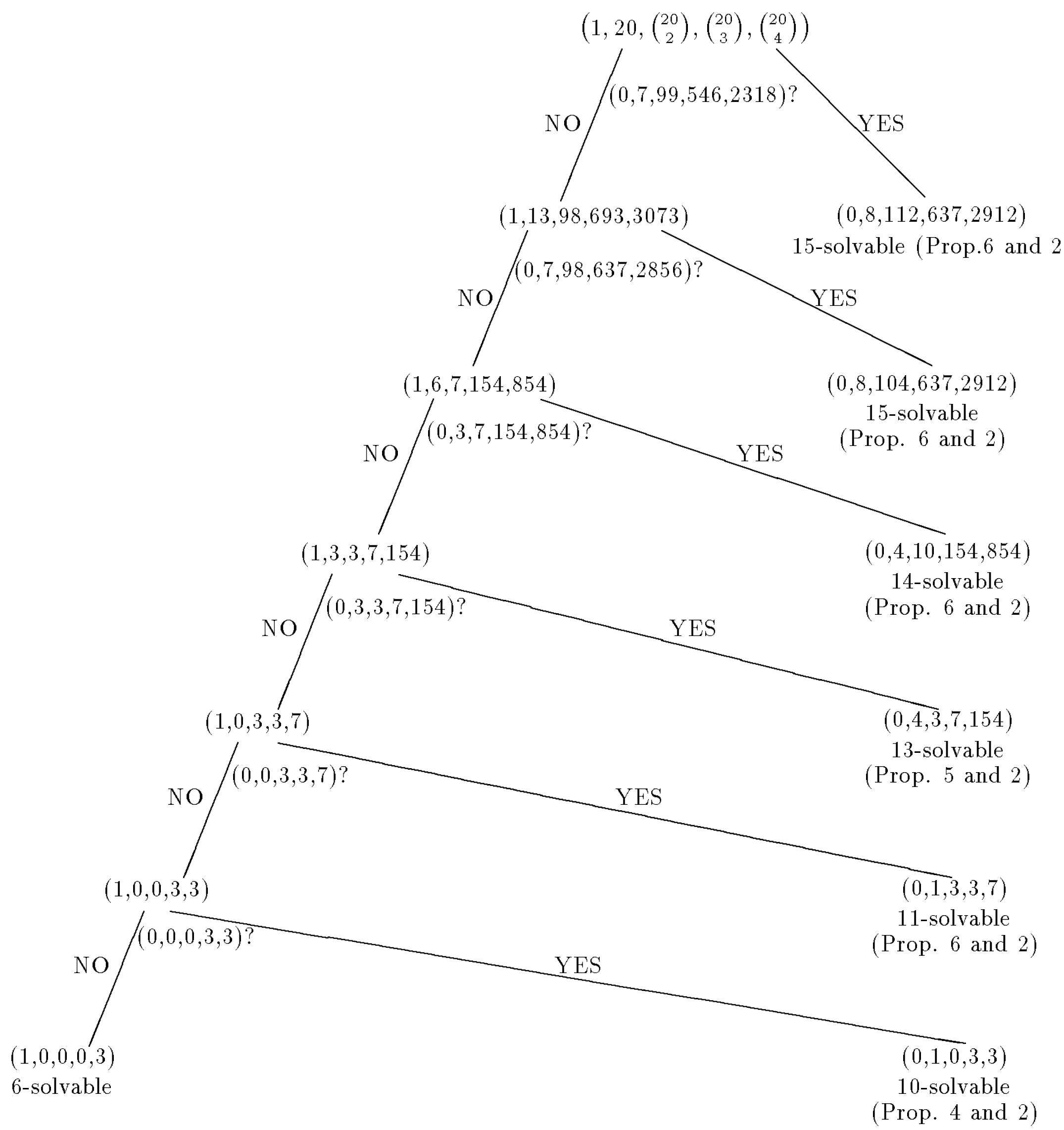
Proof. The proof is the same as in proposition 3.
In this case $U_0 = \left(0, 1, 12, \binom{12}{2}, \binom{12}{3}\right)$ that is 11-solvable by proposition 2 because $ch(U_0) = 11$. ■

Proposition 5 *Let* $U_n = \left(0, 2^n, (17-n)2^n, \binom{17-n}{2}2^n, \binom{17-n}{3}2^n\right)$ *be an Ulam set, with* $0 \leq n \leq 17$. *Then* U_n *is* $(12+n)$*-solvable.*

Proof. The proof is the same as in propositions 3 and 4.
In this case $U_0 = \left(0, 1, 17, \binom{17}{2}, \binom{17}{3}\right)$ that is 12-solvable by proposition 2 because $ch(U_0) = 12$. ■

Proposition 6 *Let* $\left(1, 20, \binom{20}{2}, \binom{20}{3}, \binom{20}{4}\right)$ *be an Ulam set.* U *is 17-solvable.*

Proof. The complete analysis of a Questioner's strategy is shown in the following figure. ■



Search tree for $(1, 20, \binom{20}{2}, \binom{20}{3}, \binom{20}{4})$.

Theorem 1 *Thirty-seven yes-no questions are sufficient to find an element $x \in \{1, 2, \ldots, 1000000\}$, if up to four lies are allowed.*

Proof. The theorem will be proved by showing that 37 questions are sufficient to find a number $x \in \{0, 1, \ldots, 2^{20} - 1\}$.

Suppose that $U = (2^{20}, 0, 0, 0)$. Let the first 20 questions be

$$Q_i = \left(2^{20-i}, (i-1)2^{20-i}, \binom{i-1}{2}2^{20-i}, \binom{i-1}{3}2^{20-i}, \binom{i-1}{4}2^{20-i}\right), \text{ for } i \leq 20.$$

The Ulam set resulting by the i-th question ($i \leq 20$) will be

$$U_i = \left(2^{20-i}, i2^{20-i}, \binom{i}{2}2^{20-i}, \binom{i}{3}2^{20-i}, \binom{i}{4}2^{20-i}\right).$$

In this phase of the algorithm $U_{i-1}Q_i^{yes} = U_{i-1}Q_i^{no}$.

After 20 questions the state $U_{20} = \left(1, 20, \binom{20}{2}, \binom{20}{3}, \binom{20}{4}\right)$ is reached, that is 17-solvable by proposition 6. Since by proposition 2 (case 2) $q < 37$ questions are not sufficient to find an integer in $\{1 \ldots 1000000\}$ when up to four answers may be erroneous, the proposed strategy is optimal with respect to the number of questions. ■

References

[1] E.R. Berlekamp, Block coding for the binary symmetric channel with noiseless, delayless, feedback, *in* "Error-Correcting Codes", pp. 61-85, Wiley, New York, 1968.

[2] J. Czyzowicz, D. Mundici, A. Pelc, Solution of Ulam's problem on binary search with two lies, *J. Combin. Theory Ser.A*, 49 (1988), 384-388.

[3] J. Czyzowicz, D. Mundici, and A. Pelc, Ulam's searching game with lies, *J. Combin. Theory Ser.A*, 50 (1989).

[4] W. Guzicki, Ulam's Searching Game with Two Lies, *J. Combin. Theory Ser.A*, 54 (1990).

[5] A. Pelc, Solution of Ulam's problem on searching with lie, *J. Combin. Theory Ser.A*, 44 (1987), 129-140.

[6] R.L. Rivest, A.R. Meyer, D.J. Kleitman, K. Winklmann, J. Spencer, Coping with errors in binary search procedures, *J. Comput. System Sci.* 20 (1980), 396-406.

[7] A. Negro, M. Sereno, "Solution of Ulam's problem on binary search with three lies", *J. Combin. Theory Ser.A*, Vol.59, n.1, January 1992, pp.149-154.

[8] A. Negro, M. Sereno, "An Ulam's searching game with three lies", Accepted to Advances in Mathematics.

[9] S.M. Ulam, "Adventures of a Mathematician", p.281, Scribner, New York, 1976.