Unfairness Correction in P2P Grids Based on Residue Number System of a Special Form

Mikhail Babenko, Nikolay Chervyakov
North-Caucasus Federal University
Stavropol, Russia
mgbabenko@ncfu.ru, k-fmf-primath@stavsu.ru

Andrei Tchernykh
CICESE Research Center
Ensenada, Mexico
chernykh@cicese.mx

Nikolay Kucherov, Maria Shabalina, Irina Vaschenko
North-Caucasus Federal University
Stavropol, Russia
{nkucherov, ikrisina}@ncfu.ru, mari.qwertyqwerty@mail.ru

Gleb Radchenko
South Ural State University,
Chelyabinsk, Russia
gleb.radchenko@susu.ru

Daniil Murga
King’s College London
London, United Kingdom
daniilmurga@gmail.com

Abstract—This paper addresses error correction codes approach in order to improve the performance of BOINC under uncertainty of users’ behavior. Redundant Residue Number System (RRNS) moduli set of the special form provides correction of user unfairness, reliability, decreased redundancy and load of the computing network. Error correction code in RRNS is improved by using error syndrome. It decreases the amount of computations required for data decoding up to 20/7 times compared to the projection methods. The proposed modification of the error syndrome allow to omit the assumption that clients are honest and reliable. The proposed approach decreases the execution time of the client programs asymptotically by 4 times.

Keywords— Uncertainty, BOINC, residue number system, error correction codes, replication, redundancy.

I. INTRODUCTION

A large class of problems in science and technology requires high-performance distributed computing to handle large amounts of data. One of the possible solutions to this problem is the high-performance grid computing based on the Peer-to-Peer (P2P) data distribution technique. Berkeley Open Infrastructure for Network Computing (BOINC) is a volunteer computing platform currently being successfully used for solving scientific problems in various fields such as: astrophysics, physics, cryptography, medicine, chemistry, etc.

However, there is a number of technical problems with this technology due to the following uncertainties [1]

1. Uncertainty of running time of the client program. It occurs due to internal and external system factors such as load of the computer, where the client program is executed, and communication bandwidth. To minimize the total running time of the tasks BOINC assigns the deadline for obtaining the results.

2. Loss of interest of the user to the project.

3. Uncertainty of getting the correct results returned by the client program. This uncertainty is either because of the dishonesty of the participants, when they can send faked result, or technical failures, attacks of hackers, or viruses.

An error in one of the subtasks can lead to a wrong result of the entire project. BOINC makes replications of the subtasks to solve this problem. By default, the number of replicates is five, and the result is correct, if it matches three or more results.

The classical replication technique has a number of shortcomings.

1. Redundancy leads to losses of computing power, reducing communication bandwidth, etc. For example, if the default replication value is used, the system performance drops by a factor of five with respect to the system without failures.

2. Delay in obtaining earned credits is increased, since it is provided according to the last client program that returns the result.

An alternative solution to this problem is to use error correction codes in the Residue Number System (RNS). Due to arithmetic properties, they allow results verification, error detection, localization and correction as well as parallel data processing.

To achieve the efficiency of error correction codes, we use numbers of a special form, for which the data encoding is performed in linear time.

Let us consider the following seven RRNS moduli set

\[
\{ 2^m - 2^{ \frac{m+1}{2} } + 1, \quad 2^m - 3, \quad 2^m - 1, \quad 2^m, \\
2^m + 3, \quad 2^m + 2^{ \frac{m+1}{2} } + 1 \},
\]

(1)

where \( m \) is an odd number. Four moduli define the dynamic range, and three ones are control moduli. Therefore, three errors in computation results can be detected.

The proposed scheme reduces data redundancy asymptotically by a factor of three. It significantly improves the system performance. However, only narrow class of problems can be solved efficiently. It is necessary to develop algorithms for error detection and localization with linear running time.

II. UNCERTAINTY IN BOINC

An extensive research has led to a general understanding of uncertainty issues in different fields ranging from physics, and computational biology to decision making in the economy and...
social sciences. However, an impact of uncertainty on P2P large scale computing systems is not studied.

The nature of uncertainties in BOINC is close to the uncertainties in cloud computing.

Tchernykh et al 2016 [2] study the impact of various types of uncertainties on the cloud computing resource provisioning. Particular attention is paid to the security and reliability of the distributed data processing.

Follow the classification of Tychinsky, 2006 [3], and Tchernykh et al., 2014 [9], we highlight the following uncertainties that significantly affect the data processing in BOINC: Retrospective uncertainty, Technical uncertainty, Uncertainty of restrictions, and Uncertainty of participants.

Retrospective uncertainty. Since the number of participants in the data processing projects in BOINC changes over time, there is no statistically valuable information on the past behavior of new users;

Technical uncertainty. BOINC has a probability do get results from one or several project participants.

Uncertainty of restrictions. It is caused by partial or complete ignorance of the operating conditions of the BOINC client programs. Since during the time interval, in which the calculations are performed, participants demands in computing capacity and load unpredictable varies;

Uncertainty of participants. It occurs in a situation of conflict between users and administrators of BOINC, when each one has his own preferences, incomplete, inaccurate information about the motives and behavior of the opposite party.

To minimize the influence of uncertainty on performance, a reliable synthesis based on the use of either a probabilistic approach or the worst-case approach are used. In this case, the uncertainty is perceived as a random variable or an interval variable.

The Monte Carlo simulation method allows obtaining reliability evaluation in the framework of optimization. Ali et al., 2003 [4] show that parallel and distributed systems need robustness to guarantee limited performance degradation despite fluctuations in the behavior of its component parts. Uncertainty can be estimated using standard deviation, differential entropy, etc. [5].

Schwiegelshohn et al., 2012 [6] proposed a model for rigid real-time planning. The practical application of algorithms is shown in Tchernykh et al., 2014 [7] and Barquet et al., 2013 [8], where the authors explore its applicability to high-performance computing, which allows it to be adapted for planning in BOINC.

The problem of non-anticipatory planning in P2P networks under uncertainty is discussed in detail in Tchernykh et al., 2014 [9]. The authors compare twenty algorithms from the point of view of Pareto dominance. The main idea of the approach is to select the replication threshold value and dynamically adapt it.

III. DATA PROCESSING IN RNS

RRNS is a non-positional number system, which is defined by the set of pairwise coprime numbers \( \{p_1, p_2, ..., p_n\} \) –RRNS moduli set. In order to detect, localize and correct errors, RRNS moduli set is divided in two subsets: first k moduli compose the dynamic range of the system, and the rest \( r = n - k \) are control moduli. In this case, the system can detect \( r \) errors.

A positive number \( X \) in RRNS, with a given moduli set, is represented as the tuple \( X = \{x_1, x_2, ..., x_n\} \), where \( x_i = [X]_{p_i} = X \mod p_i \) [8] for all \( i = 1, 2, ..., n \). This representation of a number is unique, if \( X \in [0, P - 1] \), where \( P = \prod_{i=1}^{k} p_i \) is the dynamic range of RRNS.

RRNS operations over big numbers are substituted by the operations over small numbers, which are the residues of big numbers divided by the coprime numbers from the moduli set \( p_1, p_2, ..., p_n \). Let

\[ A \equiv a_1 \pmod{p_1}, \ A \equiv a_2 \pmod{p_2}, ..., A \equiv a_n \pmod{p_n}. \]

For an integer \( A \), we have the tuple \( \{a_1, a_2, ..., a_n\} \) of least non-negative residues. According to Chinese remainder theorem (CRT), this mapping is bijective and allows error detection and correction, if \( A < \prod_{i=1}^{k} p_i \).

Fig. 1 shows the scheme of data processing in RNS.

![Fig. 1. Data processing in RNS](image)

The advantage of RRNS is the fact that addition, subtraction and multiplication can be done in parallel, and used as the basis for processing in BOINC. We formalize the parallel data processing in RRNS with the following expressions:

\[
A + B = \{a_1, a_2, ..., a_n\} + \{b_1, b_2, ..., b_n\} = \{[a_1 + b_1]_{p_1}, [a_2 + b_2]_{p_2}, ..., [a_n + b_n]_{p_n}\} \\
(2)
\]

\[
A \times B = \{a_1, a_2, ..., a_n\} \times \{b_1, b_2, ..., b_n\} = \{[a_1 \times b_1]_{p_1}, [a_2 \times b_2]_{p_2}, ..., [a_n \times b_n]_{p_n}\} \\
(3)
\]
In order to increase the reliability of the result, we add an additional mechanism based on the approximate computation of the rank of the number in RRNS, from the work [10]. This mechanism verifies the result of computations with smaller values, detects a range overflow, and efficiently compute the result of computations.

On the other hand, RRNS can be considered as a secret sharing scheme based on RRNS is a fully homomorphic cipher with respect to addition, subtraction and multiplication. These two properties of RRNS can be used to develop a secret sharing scheme based on RRNS is a fully homomorphic sharing scheme, in which each participant does not receive the result of computations. For modular multiplication, we compute the rank of the number in RRNS, from the work [10]. This additional mechanism based on the approximate computation of the module and the range of possible values of the sum of X and Y, the addition can be represented as follows:

\[ X + Y \bmod{2^m-k} = \begin{cases} X + Y, & \text{if } X + Y < 2^m - k, \\ X + Y - 2^m + k, & \text{if } X + Y \geq 2^m - k. \end{cases} \]

We compute \( R_1 = X + Y \) and \( R_2 = X + Y + k \), if m-th bit of \( R_2 \) is “0”, then “\( R_1 \)” is returned otherwise “\( R_2 - 2^m \)” “\( R_2 - 2^m \)” is computed by the following formula: \( R_2 - 2^m = R_1 \text{ AND}(2^m - 1) \). Modular subtraction is done in a similar manner with the following formula:

\[ X - Y \bmod{2^m-k} = \begin{cases} X - Y, & \text{if } X - Y \geq 0, \\ X - Y + 2^m - k, & \text{if } X - Y < 0. \end{cases} \]

For modular multiplication, we compute \( R = X \cdot Y \), where \( X,Y \in [0, 2^m - k] \) and \( R \in [0, (2^m - k - 1)^2] \), then \( R \) can be represented as follows \( R = 2^mR_{\bmod{2^m-1}} + R_{\mod{1}} \), therefore

\[ R_{\bmod{2^m-k}} = k \cdot R_{\bmod{2^m-1}} + R_{\mod{1}}. \]

Then, we consider how the (4) changes for different values of \( k \) proposed in the paper:

- \( k = 1 \), then (4) becomes \( R_{\bmod{2^m-1}} = R_{\bmod{2^m-1}} + R_{\mod{1}} \), where one modular addition is required.
- \( k = 3 \), then (4) becomes \( R_{\bmod{2^m-3}} = 3 \cdot R_{\bmod{2^m-1}} + R_{\mod{1}} \), where one modular addition and one modular multiplication by 2 are required.
- \( k = 2^{m+1} - 1 \), then (4) becomes

\[ R_{\bmod{2^m-2}} = \frac{2^{m+1} - 2^m}{2} \cdot R_{\bmod{2^m-1}} - R_{\bmod{2^m-1}} + R_{\mod{1}}. \]

where two modular additions, two modular subtractions and one modular multiplication by 2 are required.

- \( k = 2^{m+1} - 1 \). Because \( 2^m + 2^{m+1} \), then

\[ R_{\bmod{2^m+2}} = 2^2 R_{\bmod{2^m+1}}. \]
C. Computational Complexity Analysis

Let $T_A$, $T_S$, $T_M$ be the time required to add, subtract and multiply two $m$ bit operands respectively, then the time required to compute modular operations is shown in Table 1.

<table>
<thead>
<tr>
<th>Moduli</th>
<th>Addition</th>
<th>Subtraction</th>
<th>Multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^m - 1$</td>
<td>$2T_A$</td>
<td>$2T_S$</td>
<td>$T_M + 2T_A$</td>
</tr>
<tr>
<td>$2^m + 1$</td>
<td>$2T_A + T_S$</td>
<td>$T_M + 2T_A + 2T_S$</td>
<td></td>
</tr>
<tr>
<td>$2^m - 3$</td>
<td>$2T_A$</td>
<td>$2T_S$</td>
<td>$T_M + 5T_A$</td>
</tr>
<tr>
<td>$2^m + 3$</td>
<td>$2T_A + T_S$</td>
<td>$T_M + 4T_A + 5T_S$</td>
<td></td>
</tr>
<tr>
<td>$2^m - 2^{m+1} + 1$</td>
<td>$2T_A$</td>
<td>$2T_S$</td>
<td>$T_M + 3T_A + 3T_S$</td>
</tr>
<tr>
<td>$2^m + 2^{m+1} + 1$</td>
<td>$2T_A + T_S$</td>
<td>$T_M + T_A + T_S$</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 shows that the most time consuming is multiplication modulo $2^m + 3$. Considering the computational complexity of the multiplication algorithm, we can assume the time of modular operations is roughly equal for all moduli apart from $2^m$. Modular addition and subtraction take twice more time than usual addition and subtraction. Modular multiplication takes roughly the same time as usual multiplication.

V. ERROR LOCALIZATION AND CORRECTION CODE

Now, let us consider an example with seven participants.

If results are available from all seven client programs, then we can compute the result of applications using moduli

$R_{[2^{m-1}:0]} + R_{[2^{m-1}:0]} + R_{[2^{m-1}:0]} = R_{[2^{m-1}:0]} -$

$R_{[2^{m+1}:2m]} + R_{[2^{m+1}:2m]} = R_{[2^{m+1}:2m]} + R_{[2^{m+1}:2m]} + R_{[2^{m+1}:2m]} -$

$3 \cdot 2^{m+1} + 2)2^{m+2} + 1$. 

Two modular additions and one modular subtraction are required.

The class of tasks that can be solved in BOINC with RNS is related to a number of problems. These problems can be divided into two classes:

1. Operations in RNS. RNS with a non-positional number representation allows internal parallelism for addition/subtraction and multiplication with calculation in linear time. But RNS has reduced efficiency for comparisons, modular reductions, modular multiplications, sign determination, and overflow detection. Since the number conversion to a binary representation is not efficient, number approximations in RNS are used to determine the position of the number: core function, diagonal function, the approximate method, etc. As practice shows, solutions that combine several methods of approximation are efficient.

From (6a-6d), it follows that $R$ can be computed in linear time using CRT.

We check if all seven participants returned the true result. To do this, we check tree values $|R|_{[2^{m-3}:2]} \cdot r_2 = s_2, |R|_{[2^{m-1}:2]} \cdot r_4 = s_4, |R|_{[2^{m-1}:0]} \cdot r_6 = s_6$.

Without loss of generality, we assume that if $s_i \neq 0$, then $s_i = 1$.

Table II. ERROR LOCALIZATION AND CORRECTION

<table>
<thead>
<tr>
<th>#</th>
<th>$s_2$</th>
<th>$s_4$</th>
<th>$s_6$</th>
<th>Position of the error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>$r_6$</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>$r_4$</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>$r_6$</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>$r_2$</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>$r_2$, $r_6$</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>$r_2$, $r_4$</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Not defined</td>
</tr>
</tbody>
</table>

Table II shows that the most time consuming is multiplication modulo $2^m + 3$. Considering the computational complexity of the multiplication algorithm, we can assume the time of modular operations is roughly equal for all moduli apart from $2^m$. Modular addition and subtraction take twice more time than usual addition and subtraction. Modular multiplication takes roughly the same time as usual multiplication.

V. ERROR LOCALIZATION AND CORRECTION CODE

Now, let us consider an example with seven participants.

If results are available from all seven client programs, then we can compute the result of applications using moduli

$R_{[2^{m-1}:0]} + R_{[2^{m-1}:0]} + R_{[2^{m-1}:0]} = R_{[2^{m-1}:0]} -$

$R_{[2^{m+1}:2m]} + R_{[2^{m+1}:2m]} = R_{[2^{m+1}:2m]} + R_{[2^{m+1}:2m]} + R_{[2^{m+1}:2m]} -$

$3 \cdot 2^{m+1} + 2)2^{m+2} + 1$. 

Two modular additions and one modular subtraction are required.

The class of tasks that can be solved in BOINC with RNS is related to a number of problems. These problems can be divided into two classes:

1. Operations in RNS. RNS with a non-positional number representation allows internal parallelism for addition/subtraction and multiplication with calculation in linear time. But RNS has reduced efficiency for comparisons, modular reductions, modular multiplications, sign determination, and overflow detection. Since the number conversion to a binary representation is not efficient, number approximations in RNS are used to determine the position of the number: core function, diagonal function, the approximate method, etc. As practice shows, solutions that combine several methods of approximation are efficient.

Considering that at one time we can compute three sets of two errors, the total possible amount of errors is 42 [15-17]. The number of computations of $R$ with CRT is $\frac{42}{3} = 14$. Taking into consideration that projections method requires computation of 42 projections, the proposed method decreases the amount of computations by 3 times.

The proposed improvement excludes the requirement to have absolutely honest and reliable client programs.

VI. CHALLENGES

The class of tasks that can be solved in BOINC with RNS is related to a number of problems. These problems can be divided into two classes:

1. Operations in RNS. RNS with a non-positional number representation allows internal parallelism for addition/subtraction and multiplication with calculation in linear time. But RNS has reduced efficiency for comparisons, modular reductions, modular multiplications, sign determination, and overflow detection. Since the number conversion to a binary representation is not efficient, number approximations in RNS are used to determine the position of the number: core function, diagonal function, the approximate method, etc. As practice shows, solutions that combine several methods of approximation are efficient.
2. **Efficient implementation of basic algorithms in RNS.** The basis of the most computationally complex problems in physics, chemistry, ecology, biology, etc. have typical structures of linear algebra: parallel summation, scalar multiplication of vectors, triangular systems of equations, block-diagonal systems of equations, systems of equations by the Jordan method, difference schemes, etc. Hard problems to solve with RNS in BOINC are: block-diagonal systems of equations, systems of equations by the Jordan method. To solve these problems, the approach from HPC that divides the data into blocks can be used.

**VII. CONCLUSION**

Data encoding with RNS decreases the size of the operands asymptotically by the factor of four, which decreases the time of data processing by about 4 times. In order to improve fault tolerance under the conditions of uncertainty, we propose to use three control moduli set. Therefore, the redundancy of the proposed scheme is roughly 20/7, with the same level of security.

Error detection, localization and correction method is improved. It is shown that the proposed approach based on error syndrome does not require reliable computational nodes and honest users. It also decreases the number of computations compared to the method of projections.

In order to expand the class of problems that can be solved efficiently with RNS in BOINC it is required to find solutions of a number of problems. This will be the subject of future work. Moreover, we need a better understanding of user unfairness in real installations. It is another important issue to be addressed.

**ACKNOWLEDGMENT**

This work is partially supported by Russian Federation President Grant SP-1215.2016, and Act 211 Government of the Russian Federation, contract № 02.A03.21.0011.

**REFERENCES**


