# Testing the Properties of Large Quasigroups

Eliška Ochodková, Jiří Dvorský, Václav Snášel
*Department of Computer Science*
*Technical University of Ostrava*
*17. listopadu 15, Ostrava - Poruba*
*Czech Republic*
{*eliska.ochodkova, jiri.dvorsky, vaclav.snasel*}*@vsb.cz*

Ajith Abraham
*Norwegian Center of Excellence*
*Norwegian University of Science and Technology*
*O.S. Bragstads plass 2E, NO-7491 Trondheim*
*Norway*
*ajith.abraham@ieee.org*

*Abstract*—**With the growing importance of data security a growing effort to find new approaches to the cryptographic algorithms designs appears. One of the trends is to research the use of other algebraic structures than the traditional, such as a quasigroup. Quasigroups are equivalent to the more familiar Latin squares. There are many characteristics that must quasigroups have from the cryptography point of view. They have to be non-commutative, non-associative, non-idempotent and so on. If one want to work with quasigroups of a large order, effective methods of testing their properties are necessary. In this paper we present several experiments on various types of guasigroups and their results.**

*Keywords*-**Quasigroup; Cryptography; Hamming distance; Slot distiribution;**

## I. Introduction

The goal of this paper is to introduce a method for properties of huge quasigroups testing. We want to show first series of experiments and their results. There were used two types of quasigroups, the table quasigroup and analytic one; their properties were tested by analyzing values of a simple hash function based on them.

Quasigroups, or more famous Latin squares, are well known combinatorial designs with a lot of theoretical results concerning them. Researchers have focused in quasigroups' usage in the information and communication technologies, mainly when cryptographic tools are necessary, more seriously from the beginning of century. Cryptographic techniques are part of many protocols. There were published many cryptographic algorithms based on quasigroups primitives, from simple till more ambitious ones. We can mention e.g. [1], [2]; Hassinen et al. propose to secure SMS messages using quasigroup encryption [3]; authors of [4] have implemented a class of public key algorithms MQQ, that are based on quasigroup string transformations, in FPGA. Quasigroups may also be used for error detection applications [5].

For a cryptographic purpose quasigroups have to be of a good quality. Examination of required properties may not be easy if we want to work with the quasigroups of a large order, e.g. $2^{16}$ and larger. Number of distinct quasigroups (e.g. Latin squares) of a given order grows exceedingly quickly with the growing order of quasigroup, there is no known easily-computable formula for the number of distinct quasigroups.

Various approaches are used for the good quality quasigroup generation. It is possible to use quasigroups of a small order represented as a look-up table only, whose properties may by verified by the exhaustive search. There is proposed stream cipher Edon80 [6] as an eSTREAM[1] candidate. The cipher Edon80 uses quasigroup of order $4$. Principle of the selection appropriate quasigroups of order $4$ is described in [7].

Another approach is to construct quasigroup of large order e.g. by applying some basic permutations. Permutations may be constructed in many ways. Authors of NIST's SHA-3 competition[2] candidate, hash function Edon$\mathcal{R}$, have used high-quality Latin squares of order $8$ for constructing the permutations on 256-bit words and these permutations are then used

[1] http://www.ecrypt.eu.org/stream/

[2] http://csrc.nist.gov/groups/ST/hash/sha-3/index.html

9781-4244-3941-6/09/$25.00 ©2009 IEEE

for the quasigroup of order $2^{256}$ construction [8].

The organization of the paper is following: In section 2 we give brief introduction to the math background. Experiments, tested data and quasigroups are presented in section 3. In section 4 we describe experiments results. Finally in last section we give conclusions and future research directions.

## II. Mathematical Background

**Definition 1.** A grupoid $(Q, *)$ is said to be a quasigroup (i.e. algebra with one binary operation $(*)$ on the set $Q$) satisfying the law:

$$\forall(u, v \in Q)(\exists! x, y \in Q)(u * x = v \wedge y * u = v).$$

This implies:

1) $x * y = x * z \vee y * x = z * x \Rightarrow y = z$
2) The equations $a * x = b, y * a = b$ have an unique solutions $x, y$ for each $a, b \in Q$.

Quasigroups are equivalent to the well known Latin squares. A Latin square of order $n$is an $n \times n$ matrix filled with $n$ different symbols in such a way that each symbol occurs exactly once in each row and exactly once in each column. The multiplication table of a quasigroup of order $n$ is a Latin square of order $n$, and conversely every Latin square of order $n$ is the multiplication table of a quasigroup of order $n$ [9]. We say that quasigroup $(Q, *)$ is of order $n$ if $|Q| = n$.

However, in general, the operation $(*)$ is neither a commutative nor an associative operation. As is known, every quasigroup satisfying the associative law has an identity element and is, hence, a group, e.g. a group is, by definition, an associative loop.

**Definition 2.** A loop is a quasigroup $(Q, *)$ with an identity element $e \in Q$ such that:

$$(\exists e \in Q)(\forall x \in Q)(x * e = x = e * x).$$

While some Latin squares do represent associative operations and can form a group, most Latin squares do not. For example, at order 4 there are 576 Latin squares, but only 16 are associative (about 2.8 percent). So non-associative (and thus non-group) squares dominate heavily, and are desirable for cryptography anyway.

By a quasigroup of a high cryptographic quality we mean a non-commutative, non-associative, non-idempotent quasigroup without right or left units and without proper sub-quasigroups.

**Definition 3.** A groupoid $(Q, *)$ is said to be a left (a right) quasigroup if the equation

$$x * a = b \ (a * y = b)$$

have a unique solution $x$ $(y)$ in $Q$ for every $a, b \in Q$.

**Definition 4.** A quasigroup is called idempotent if the identity $x * x = x$ is satisfied for all $x \in Q$.

That quasigroup $(Q, *)$ should not be linear, in the sense that no output bit of $a * b$ is a linear combination of the input bits of $a$ and $b$, for each $a, b \in Q$. Also quasigroups with various identities seem to be a problematic structures. Every quasigroup satisfying e.g. the Moufang identity $M1 : (x * (y * z)) * x = (x * y) * (z * x)$, and other, is a loop (let this equation with variables is understood to be universally quantified) [10].

Testing properties of the large ausigroups is done through a simple hash function defined below. Experiments are described in the following section.

**Construction 1.** Let $(Q, *)$ be a quasigroup and $Q^+$ be a set of all nonempty words formed by the elements $q_i \in Q$, $1 \leq i \leq n$. For a fixed $a \in Q$ let the hash function $h_Q : Q \times Q^+ \rightarrow Q^+$ be defined as

$$h_Q(q_1 q_2 \dots q_n) = ((\dots(a * q_1) * q_2 * \dots) * q_n.$$

**Definition 5.** Let $(G, \cdot)$, $(H, *)$ be two quasigroups. An ordered triple $(\pi, \rho, \omega)$ of bijections $\pi, \rho, \omega$ of the set $G$ onto set $H$ is called an isotopism of $(G, \cdot)$ upon $(H, *)$ if

$$\forall u, v \in G, \pi(u) * \rho(v) = \omega(u \cdot v).$$

Quasigroups $(G, \cdot)$, $(H, *)$ are said to be isotopic.

We can imagine an isotopism of quasigroups as a permutation of rows and columns of quasigroup's multiplication table.

## III. Experiment Description

### A. Quasigroups generation

A various methods of generating a practically unlimited number of quasigroups of a (theoretically) arbitrary order (problem that is crucial to cryptography) are used in various papers. The usage of a general quasigroup in computation requires to store its Cayley table, i.e. $n^2$ elements. Various tricks based on some "easy to evaluate" expression can be used to overcome the problem of storage requirements. We proposed e.g. the quasigroup of modular subtraction to be used [11], where the operation $*$ defined on $Q$ is given as

$$a * b = (a + n - b) \ mod \ n, \ n = |Q|, \ a, b \in Q.$$

The isotopism of quasigroups gives us then the power to use a large number of isotopic quasigroups, quasigroup isotopic to quasigroup of modular subtraction can be used. The multiplication in such an isotopic quasigroup is defined as follows:

$$a * b = \pi((\omega(a) + n - \rho(b)) \bmod n).$$

This allows us to use quasigroups with a very large number of elements without necessity of their storage.

- We've used also so called "table quasigroup" that comes from the quasigroup isotopic of quasigroup of modular subtraction. Three random permutations were generated and obtained table was used to modify the original table. Disadvantage of this method is a huge space complexity ($n^2$ elements must be stored).
- Second method of the quasigroup generation [12] is that follows, such a quasigroup we call "analytic quasigroup". This quasigroup is based on Bruck-Toyoda theorem, see [13].For the proof of this concept see [14].

**Construction 2.** Let $m > 1$ and $h, k, l$ are integers such that $GCD(h, m) = 1 = GCD(k, m)$, where $GCD$ denotes a greatest common divisor. If operation $(*)$ is defined as follows: $a*b = (h\cdot a+k\cdot b+l) \ mod \ m$, then $Q_m = (Z_m, *)$ is quasigroup over $Z_m$.

### B. Experiments

The aim of our experiments was to check empirically the characteristics of quasigroups by valuing the hash function based on group $(\mathbb{Z}_n, +)$, on the table quasigroups and on quasigroups defined analytically. The experiments can be divided into two groups:

1) distribution of hash values, for given quasigroup and for given testing data,
2) distribution of hash value's changes, for given quasigroup and for given testing data, with respect to bit change in them.

### C. Quasigroups used in tests

Quasigroups of following orders have been used in our experiments:

- quasigroup of order $5003$ defined by table and analytically,
- quasigroup of order $2^{16}$ defined analytically,
- quasigroup of order $2^{32}$ defined analytically[3],
- group $(\mathbb{Z}_{5003}, +)$, we have conducted some experiments with additive group $(\mathbb{Z}_{5003}, +)$ for comparison. The additive group of integers $\mathrm{mod}\ n$ satisfies Latin square law. This group was used as a representative of cryptographically bad quasigroups (associativity and commutativity holds). Generation and work with the group isn't problematic.

To compare properties of the table group, quasigroups and analytic quasigroups, quasigroups of order $5003$ were used. Simultaneously these quasigroups are used for the testing quasigroups of prime order. Quasigroups of this order can be used as a base of the hash functions for the hash table data structure. Quasigroups of order $64K$ and $4G$ were used to test large quasigroup with order equal to power of 2. These quasigroups are intended rather for the cryptography.

### D. Testing data

We've used the words extracted from text as testing data. Words were extracted from a .GOV collection of web pages WebTREC [15]. The original text size was approximately $18$ gigabytes. The dictionary was drawn up from this text. The dictionary contains a total of $4319200$ unique words. In following text, this file is referred as `webtrec`. This file is used to simulate hashing of the text messages.

[3] Orders of these qusigroups will be referred as $64K$ and $4G$ in following text.

## IV. Experimental Results

### A. Distribution of hash values

The first part of our experiment was determined to check the distribution of probability of occurrence of different hash values for the given hash function – quasigroup[4]. The `webtrec` was used as an input data in this case. The experiment was as follows:

1) Initialize a quasigroup $Q$ of order $k$.
2) Create hash table with $k$ slots. Collisions were solved using separate chaining i.e. words with the identical hash values are stored in the same slot, in one list[5].
3) For each input word $w$ compute hash value $h_Q(w)$.
4) Insert word $w$ into slot with index $h_Q(w)$.

*1) Ideal results:* The ideal result of this experiment is the uniform distribution of input words across the whole table. Mean value of the number of words in one slot should match the total number of input words divided by the number of the slots. Histogram of the slot lengths should copy the normal distribution in this case.

*2) Experimental results:* Charts 1(a), 1(c), 1(e) show histograms of the slot lengths for table quasigroup of order $5003$, analytic quasigroup of order $5003$ and analytic quasigroup of order $64K$. X-axis represents the number of words in the slot, i.e. length of the slot; y-axis represents the number of slots with a given length. Histograms of both quasigroup, table and analytic, of order $5003$ are very similar, the histogram of analytic quasigroup of order $64K$ is already deformed.

Basic statistical results of this experiment are shown in Table I[6]. Table II shows that distribution of the length of slots corresponds 68-95-99.7 rule and can therefore are considered as a normal distribution. Table I clearly shows that the hash functions based on table and analytic quasigroups of order $5,003$ provide almost identical results. The parameters $\mu$ and $\sigma$ were used to construct normal distribution curve. We can graphically compare experimentally measured distribution curve and normal distribution curve $N(\mu, \sigma^2)$ in the graph 1(b) for table quasigroup and in the graph 1(d) for analytic quasigroup.

Table I

Slot Distribution of `webtrec` — Statistical Evaluation

| Quasigroup | Order | $\mu$ | $\sigma$ |
|---|---|---|---|
| Table quasigroup | 5003 | 863.322 | 29.438 |
| Analytic quasigroup | 5003 | 863.322 | 28.978 |
| Analytic quasigroup | 64K | 65.906 | 13.791 |
| Group $(\mathbb{Z}_{5003}, +)$ | 5003 | 863.322 | 1577.503 |

The situation is different for analytic quasigroup of order $64K$. The histogram is deformed due to normal distribution curve $N(\mu, \sigma^2)$ computed according to the parameters from Table I. This deviation is due to higher order of used quasigroup, this corresponds to the greater number of slot in the hash table. Thanks to the higher number of slots, one slot in table will contain less number of input words in general, about $66$ words per slot. In this case the difference of a few words in comparison with the theoretical number of words in the slot represents a significant deviation and causes deformation of distribution curve. We can compare this situation with quasigroups of order $5003$. There is approximately $863$ words per one slot, and the difference of a few words in relation to the theoretical number of words in the slot plays no role.

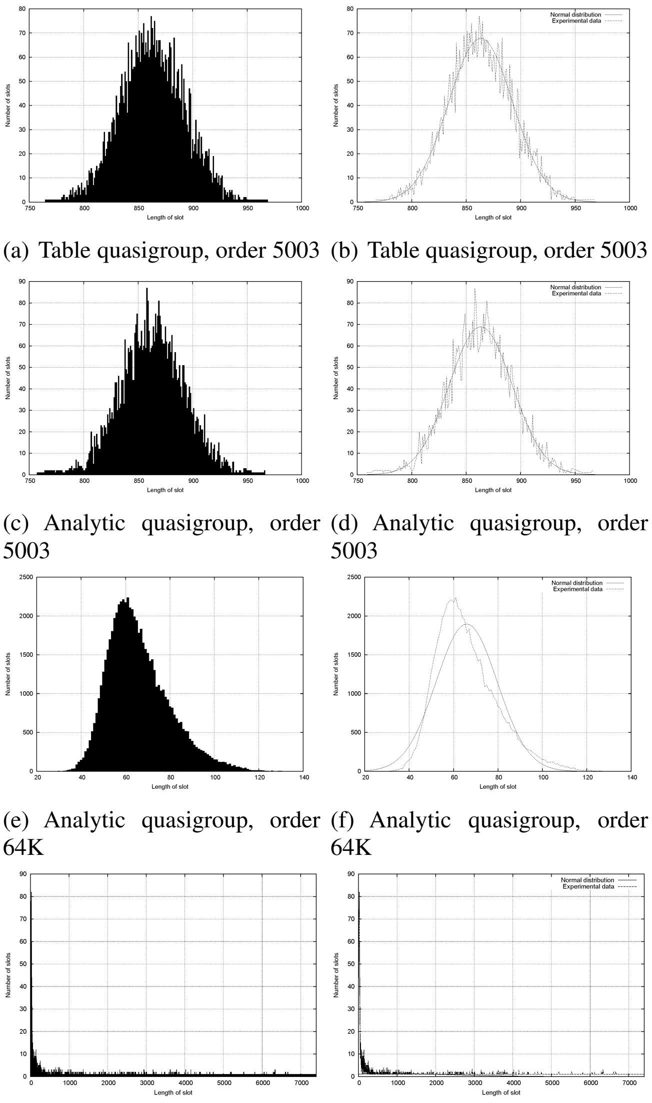
Results for additive group $(\mathbb{Z}_{5003}, +)$ are shown in Figure 1(g), comparison with the normal distribution curve can be seen in Figure 1(h). Distribution of slots' length differs completely from normal distribution. These results are due to other characteristics of the group – associativity and commutativity. In this case, the words $aab$, $aba$ a $baa$ will have same hash value.

In this experiment, we does not use the quasigroup of order $4G$. Corresponding hash table should have four billion slots, while there are approximately only 4 million input words. The result of such experiment will be meaningless.

[4] Hash function is same in all experiments.

[5] The table in this experiment was fictitious, instead of storing of data in slots, only counters in slots were incremented.

[6] In following tables $\mu$ represents mean value of experimental data and $\sigma$ standard deviation of the data.



(a) Table quasigroup, order 5003 (b) Table quasigroup, order 5003

(c) Analytic quasigroup, order 5003 (d) Analytic quasigroup, order 5003

(e) Analytic quasigroup, order 64K (f) Analytic quasigroup, order 64K

(g) Group $(\mathbb{Z}_{5003}, +)$ (h) Group$(\mathbb{Z}_{5003}, +)$

Figure 1. Slot Distribution of `webtrec`

### *B. Changes of hash values – The number of changed bits*

The second part of our experiment was to monitor changes in the hash values for the given hash function – quasigroup. The `webtrec` was used as an input data in this case. The experiment was as follows:

1) Initialize quasigroup $Q$ of order $k$.
2) For each input word $w$ compute hash value $h_Q(w)$.
3) Assume that binary representation of word $w$ has $l$ bits. Create $l$ words $w_i$ from word $w$ such that $i$-th bit in word $w$ is changed.

Table II
SLOT DISTRIBUTION OF `WEBTREC` — CONFIDENCE INTERVALS

(a) Table quasigroup, 5003

| Confidence interval | Percentage |
|---|---|
| $[-\sigma, \sigma]$ | 68.479 |
| $[-2\sigma, 2\sigma]$ | 95.623 |
| $[-3\sigma, 3\sigma]$ | 99.820 |

(b) Analytic quasigroup, 5003

| Confidence interval | Percentage |
|---|---|
| $[-\sigma, +\sigma]$ | 68.559 |
| $[-2\sigma, +2\sigma]$ | 95.583 |
| $[-3\sigma, +3\sigma]$ | 99.620 |

(c) Analytic quasigroup, 64K

| Confidence interval | Percentage |
|---|---|
| $[-\sigma, +\sigma]$ | 69.081 |
| $[-2\sigma, +2\sigma]$ | 95.355 |
| $[-3\sigma, +3\sigma]$ | 99.194 |

(d) Group $(\mathbb{Z}_{5003}, +)$

| Confidence interval | Percentage |
|---|---|
| $[-\sigma, +\sigma]$ | 85.509 |
| $[-2\sigma, +2\sigma]$ | 92.544 |
| $[-3\sigma, +3\sigma]$ | 96.882 |

4) For each word $w_i$ compute hash value $h_Q(w_i)$ and measure Hamming distance between $h_Q(w)$ and $h_Q(w_i)$.

*1) Ideal results:* The ideal result of this experiment is to change half of bits in hash values when one bit is changed in the input. Histogram of Hamming distances in this case should copy the normal distribution.

*2) Experimental results:* Charts in Figure 2 show histograms of Hamming distances between original word hash value and hash value of word with one bit changed. X-axis represents the number of changed bits; y-axis represents the probability (relative frequency) of change of given number of bits.

Basic statistical results of this experiment are shown in Table III. Table IV show, that distribution of bit changes corresponds to 68-95-99.7 rule, with some exception in $[-\sigma, +\sigma]$ interval. Distribution of Hamming distances can be considered as normal distribution. From these tables, it is clear that table and analytic quasigroups of order 5003 provides almost identical results. The parameters $\mu$ and $\sigma$ were used for normal distribution curve $N(\mu, \sigma^2)$ construction.

Table III
HAMMING DISTANCES, webtrec — STATISTICAL EVALUATION

| Quasigroup | Order | $\mu$ | $\sigma$ |
|---|---|---|---|
| Table quasigroup | 5003 | 6.263 | 1.800 |
| Analytic quasigroup | 5003 | 6.414 | 1.662 |
| Analytic quasigroup | 64K | 8.043 | 1.938 |
| Analytic quasigroup | 4G | 16.163 | 2.684 |
| Group $(\mathbb{Z}_{5003}, +)$ | 5003 | 1.996 | 1.374 |

We can graphically compare experimentally measured distribution curve and normal distribution curve $N(\mu, \sigma^2)$ in the graphs 2.

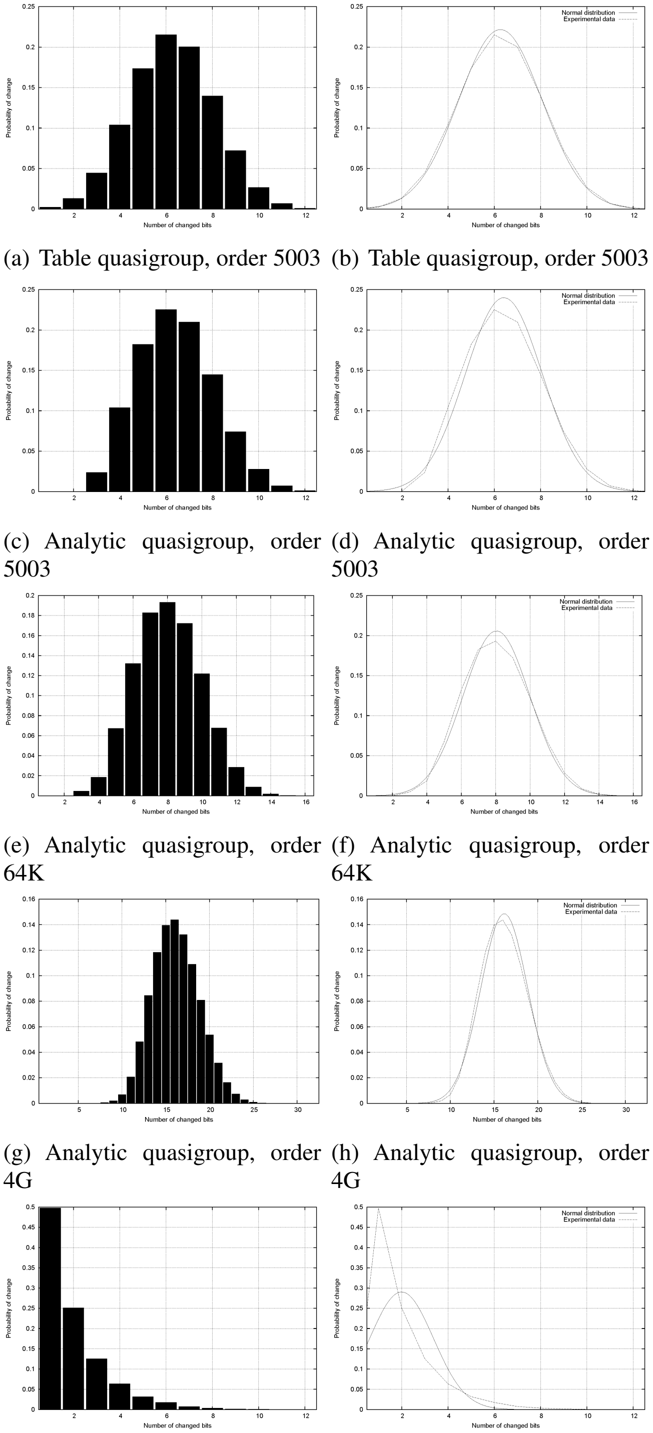
In all cases, we can see very good agreement between experiment and the ideal results, the theoretical results.

Completely different situation arises for the additive group $(\mathbb{Z}_{5003}, +)$. From the histogram 2(i) is quite clear that in half of the cases changed only one bit, in the quarter of the cases, only two bits. This illustrates the mean number of changed bits in Table III. Comparison with the normal distribution curve is shown in Figure 2(j). A significant deformation of distribution of changed bits can be seen there.

## V. CONCLUSION

Our goal is to find a way of testing the properties of large quasigroups and to explore the interpretation of experimental results. We've reported two types of experiments conducted and their results in this paper - distribution of hash values and distribution of hash values with respect to bit change in testing data, for given quasigroup and for given testing data. Two types of quasigroup generation were used, we call them table and analytical quasigroup. It seems that if quasigroup is randomly selected from the space of all possible quasigroups of the given order, it is relatively difficult to choose bad quasigroup. For comparison we used group $(\mathbb{Z}_{5003}, +)$, which represents quasigroup with bad cryptographic quality. Influence of associativity and commutativity is clearly evident on the charts.

Next time we are going to use different data collections, e.g. whole web pages from .GOV collection and Enron e-mail dataset. Different experiments will



(a) Table quasigroup, order 5003 (b) Table quasigroup, order 5003

(c) Analytic quasigroup, order 5003 (d) Analytic quasigroup, order 5003

(e) Analytic quasigroup, order 64K (f) Analytic quasigroup, order 64K

(g) Analytic quasigroup, order 4G (h) Analytic quasigroup, order 4G

(i) Group $(\mathbb{Z}_{5003}, +)$ (j) Group $(\mathbb{Z}_{5003}, +)$

Figure 2. Histogram of Hamming Distances – webtrec

Table IV
CONFIDENCE INTERVALS, WEBTREC

(a) Table quasigroup, 5003

| Confidence interval | Percentage |
|---|---|
| $[-\sigma, \sigma]$ | 72.895 |
| $[-2\sigma, 2\sigma]$ | 94.986 |
| $[-3\sigma, 3\sigma]$ | 99.888 |

(b) Analytic quasigroup, 5003

| Confidence interval | Percentage |
|---|---|
| $[-\sigma, +\sigma]$ | 76.205 |
| $[-2\sigma, +2\sigma]$ | 94.006 |
| $[-3\sigma, +3\sigma]$ | 99.882 |

(c) Analytic quasigroup, 64K

| Confidence interval | Percentage |
|---|---|
| $[-\sigma, +\sigma]$ | 54.818 |
| $[-2\sigma, +2\sigma]$ | 93.738 |
| $[-3\sigma, +3\sigma]$ | 99.781 |

(d) Analytic quasigroup, 4G

| Confidence interval | Percentage |
|---|---|
| $[-\sigma, +\sigma]$ | 64.292 |
| $[-2\sigma, +2\sigma]$ | 96.286 |
| $[-3\sigma, +3\sigma]$ | 99.824 |

(e) Group $(\mathbb{Z}_{5003}, +)$

| Confidence interval | Percentage |
|---|---|
| $[-\sigma, +\sigma]$ | 87.386 |
| $[-2\sigma, +2\sigma]$ | 93.752 |
| $[-3\sigma, +3\sigma]$ | 98.676 |

be presented and we will explore ways of visualization quasigroups' characteristics, too.

## References

[1] D. Gligoroski and S. Markovski, "Cryptographic potentials of quasigroup transformations," Talk at EIDMA Cryptography Working Group, Utrecht, 2003.

[2] J. Kong, "The role of latin square in cipher systems: Matrix approach to model encryption modes of operation," UCLA Computer Science Department Technical Report CSTR030038, 2008.

[3] M. Hassinen and S. Markovski, "Secure sms messaging using quasigroup encryption and java sms api," in *SPLST*, 2003, pp. 187–.

[4] M. El-Hadedy, D. Gligoroski, and S. J. Knapskog, "High performance implementation of a public key block cipher - mqq, for fpga platforms," in *RECONFIG '08: Proceedings of the 2008 International Conference on Reconfigurable Computing and FPGAs*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 427–432.

[5] G. L. Mullen and V. Shcherbacov, "n-T-quasigroup codes with one check symbol and their error detection capabilities," *Comment. Math. Univ. Carolinae*, vol. 45, no. 2, pp. 321–340, 2004.

[6] D. Gligoroski, S. Markovski, L. Kocarev, and M. Gusev, "Stream cipher edon80," *eSTREAM candidate, http://www.ecrypt.eu.org/stream/*, 2005.

[7] S. Markovski, D. Gligoroski, and J. Markovski, "Classification of quasigroups by random walk on torus," *Journal of Applied Mathematics and Computing*, vol. 19, no. 1-2, pp. 57–75, March 2005.

[8] D. Gligoroski and et al., "EdonR cryptographic hash function," *Submition to NIST's SHA-3 hash function competition, http://csrc.nist.gov/groups/ST/hash/sha-3/index.html*, 2008.

[9] J. Dénes and A. Keedwell, *Latin Squares and their Applications*. New York: Akadémiai Kiadó, Budapest, Academic Press, 1974.

[10] R. H. Bruck, *A survey of binary systems*. Springer - Verlag, 1958.

[11] E. Ochodková, J. Dvorský, and V. Snášel, "Hash functions based on large quasigroups," in *Proceedings of Workshop Velikonoční kryptologie (Easter Cryptology)*, 2002.

[12] R. H. Schulz, *Codierungstheorie*. Vieweg Verlag, 1991.

[13] K. Toyoda, "On axioms of linear functions," *Proc. Imp. Acad. Tokyo*, vol. 17, p. 221227, 1941.

[14] K. Kunen, "Quasigroups, loops, and associative laws," *Journal of Algebra*, vol. 185, 1995.

[15] D. Harman, Ed., *The Forth REtrieval Conference (TREC-4)*. NIST, 1997.