On Overlappings of Digitized Straight Lines and Shared Steganographic File Systems

Maria Bras-Amorós* **and Josep Domingo-Ferrer***

* Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili,

Av. Països Catalans 26, 43007 Tarragona, Catalonia.

E-mail: {maria.bras,josep.domingo}@urv.cat

Abstract. We consider the unbounded integer grid and the digitized version of the straight line $y = \alpha x + \beta$, with $\alpha, \beta \in \mathbb{R}$ being the set of points $(i, [\alpha i + \beta]), i \in \mathbb{Z}$, where $[\cdot]$ is the integer rounding operator ($[x] - 0.5 \leq x < [x] + 0.5$). We address the problem of counting the number of points in the integer grid in which two digitized straight lines overlap each other in the particular case when the crossing point of the non-digitized version of the lines has integer coordinates and the slopes belong to the set $\{a/b : a \in \{-(N-1), .., (N-1)\}, b \in \{1, .., (N-1)\}\}$, that is, all the possible slopes of the segments between two different points in the $N \times N$ grid.

Applications of this problem are explained, with a special focus on a shared steganographic system with error correction.

Keywords. Digitized line, shared steganographic file system.

1 Introduction

When straight lines are mapped onto a grid, such as a computer screen or a bit matrix, they need to be digitized, that is, discretized. The properties of digitized lines have been studied since decades, mostly in connection with computer graphics [4]. For example, in [6] the problem of recognizing a sequence of pixels as a digitized straight line segment is considered; in [5], the number of different digitized lines in an $N \times N$ pixel array is investigated.

With a different motivation, namely shared steganographic file systems [1, 2], we pursue here the characterization of digitized lines and deal with the problem of counting the number of overlapping points when two digitized lines over a grid cross each other. Section 2 contains an analysis of the problem from a theoretical point of view. Section 3 goes further and gives an approximate result. Section 4 reports on computational results. Section 5 motivates the relevance of the problem for shared steganographic file systems and other areas in computer science. Section 6 summarizes conclusions and lists open research issues.

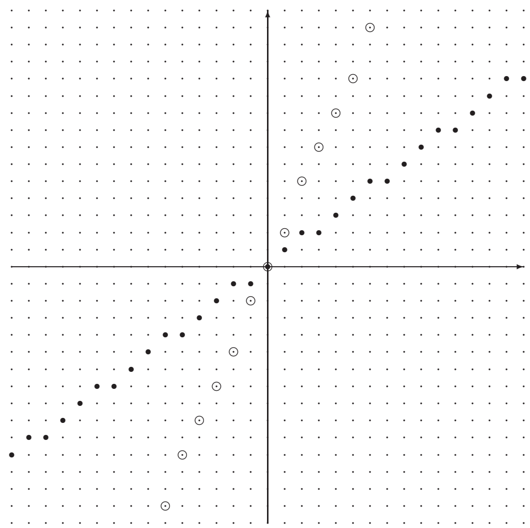

Figure 1: Digitized lines with slopes $3/4$ and $16/7$.

2 Analysis for an unbounded integer grid

We consider the unbounded integer grid and the digitized version of the straight line $y = \alpha x + \beta$, with $\alpha, \beta \in \mathbb{R}$ being the set of points $(i, [\alpha i + \beta]), i \in \mathbb{Z}$, where $[\cdot]$ is the integer rounding operator ($[x] - 0.5 \leq x < [x] + 0.5$).

Let two straight lines have rational slopes s_1 and s_2, $s_1 \neq s_2$ and suppose they intersect each other at a point with integer coordinates. Let λ be the number of positions in the integer grid in which the digitized versions of these straight lines overlap. Obviously λ does not depend on the integer point in which the two original lines cross each other. So without loss of generality we can think of the two lines as crossing at the origin and λ as a function depending only on s_1 and s_2. For instance, $\lambda(3/4, 16/7) = 1$ (see Figure 1), $\lambda(4/5, 5/4) = 4$ (Figure 2), $\lambda(1/6, 1/3) = 6$ (Figure 3).

It would seem natural if λ depended only on the angle between the two straight lines. However this is not true. In Figure 4 we depict the digitized versions of the lines with slopes 3 and 6, which are separated by the same angle as the lines with slopes $1/6$ and $1/3$. One can see that $\lambda(3, 6) = 1$ while as we noticed before, $\lambda(1/6, 1/3) = 6$. In fact λ is not determined by the difference $|s_1 - s_2|$ either. Indeed, $|11/12 - 13/12| = 1/6$ and $|1/6 - 1/3| = 1/6$ but $\lambda(1/6, 1/3) = 6$ while $\lambda(11/12, 13/12) = 11$ (see Figure 5).

In the next lemma we state some results related to $\lambda(s_1, s_2)$.

Lemma 1. 1. *$\lambda(s_1, s_2) = \lambda(s_1 - k, s_2 - k)$ for any integer k,*

2. *If $\lfloor s_1 \rfloor = \lfloor s_2 \rfloor$, then $\lambda(s_1, s_2) = \lambda(s_1 - \lfloor s_1 \rfloor, s_2 - \lfloor s_2 \rfloor)$,*

3. *If $|s| \geq 1$, then $\lambda(0, s) = 1$,*

4. *If $\lfloor s_2 \rfloor - \lfloor s_1 \rfloor \geq 2$ then $\lambda(s_1, s_2) = 1$,*

5. *$\lambda(0, s) = \lfloor \frac{1}{2|s|} \rfloor + \lceil \frac{1}{2|s|} \rceil$, whenever $s \neq 0$.*

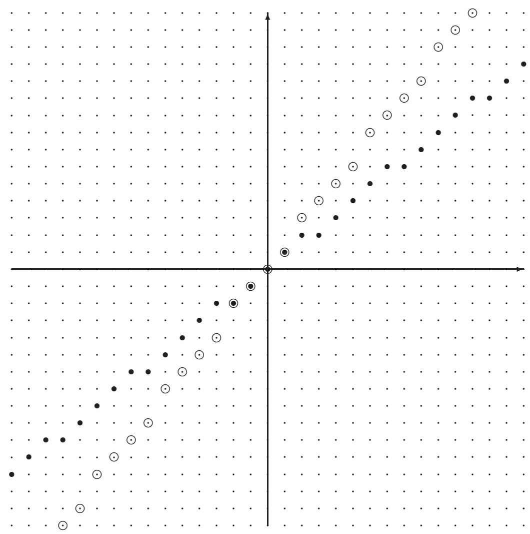

Figure 2: Digitized lines with slopes $4/5$ and $5/4$.

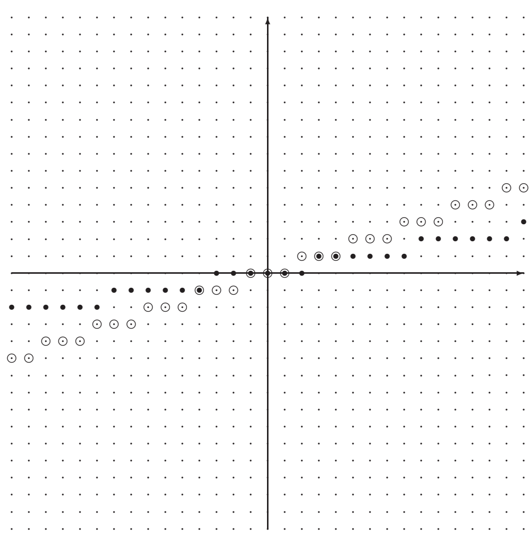

Figure 3: Digitized lines with slopes $1/6$ and $1/3$.

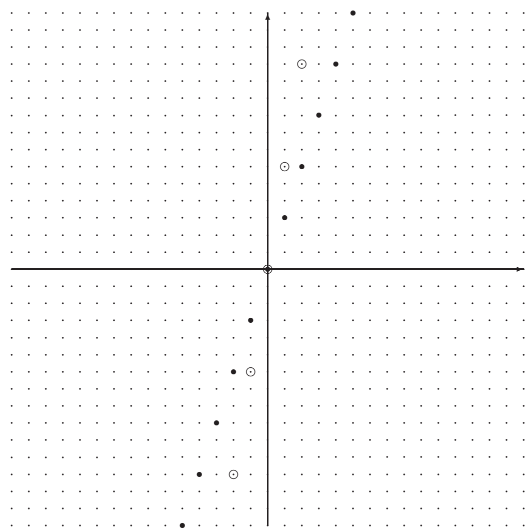

Figure 4: Digitized lines with slopes 3 and 6.

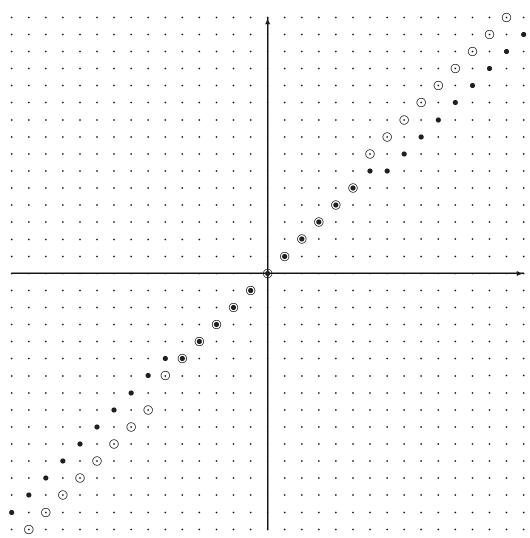

Figure 5: Digitized lines with slopes 11/12 and 13/12.

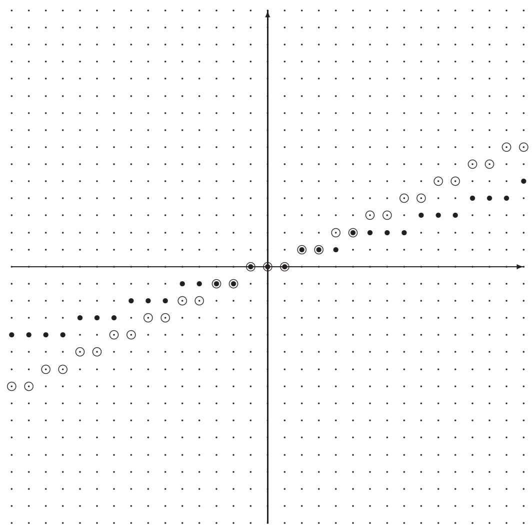

Figure 6: Digitized lines with slopes $3/10$ and $9/19$.

Proof. The first statement follows from the observation that $[x - j] = [x] - j$ for any integer j, so $(i, [(s_1 - k)i]) = (i, [(s_2 - k)i])$ if and only if $(i, [s_1 i]) = (i, [s_2 i])$. The second statement is a consequence of the first one. The third statement is obvious. To prove the fourth statement notice that $\lambda(s_1, s_2) = \lambda(s_1 - \lfloor s_1 \rfloor - 1, s_2 - \lfloor s_1 \rfloor - 1)$ but $s_1 - \lfloor s_1 \rfloor - 1 < 0$ while $s_2 - \lfloor s_1 \rfloor - 1 \geq \lfloor s_2 \rfloor - \lfloor s_1 \rfloor - 1 \geq 1$. Now the result follows easily from the third statement.

For the fifth statement notice that $(i, [s \cdot i]) = (i, 0)$ if and only if $-1/2 \leq s \cdot i < 1/2$ that is, $-1/2 \leq |s| \cdot i < 1/2$ whenever $s > 0$ and $-1/2 \leq |s| \cdot (-i) < 1/2$ whenever $s < 0$. Since $i \in \mathbb{Z}$, this is equivalent to

$$-\left\lfloor \frac{1}{2|s|} \right\rfloor \leq i < \left\lceil \frac{1}{2|s|} \right\rceil$$

if $s > 0$ and to

$$-\left\lfloor \frac{1}{2|s|} \right\rfloor \leq -i < \left\lceil \frac{1}{2|s|} \right\rceil$$

if $s < 0$. In both cases there are $\lfloor \frac{1}{2|s|} \rfloor + \lceil \frac{1}{2|s|} \rceil$ such integers and so $\lambda(0, s) = \lfloor \frac{1}{2|s|} \rfloor + \lceil \frac{1}{2|s|} \rceil$. □

3 Approximation

We empirically checked that if $0 \leq s_1, s_2 < 1$ then $\lambda(s_2 - s_1, 0)$ approximates quite well $\lambda(s_1, s_2)$. For instance, there are 28 rational numbers of the form a/b with $a \in \{0, \ldots, 9\}$, $b \in \{a + 1, \ldots, 9\}$ with the usual equivalence relationship, and so 378 combinations of them. There are only 12 pairs s_1, s_2 out of these 378 combinations such that $\lambda(s_1, s_2) > \lambda(s_2 - s_1, 0)$. Furthermore, in all these 12 cases, $\lambda(s_1, s_2) = \lambda(s_2 - s_1, 0) + 1$. Notice that however, we can find some further examples in which $\lambda(s_1, s_2) > \lambda(s_2 - s_1, 0) + 1$. For instance, for $s_1 = 3/10$, $s_2 = 9/19$ ($s_2 - s_1 = 33/190$), $\lambda(s_1, s_2) = 8$ while $\lambda(s_2 - s_1, 0) = 5$ (Figure 6). For more details on empirical results see the next section.

With the assumption that $\lambda(s_2 - s_1, 0) \approx \lambda(s_1, s_2)$ whenever $0 \leq s_1, s_2 < 1$ and Lemma 1 we can approximate $\lambda(s_1, s_2)$ quite well for any values of s_1, s_2:

- If $|\lfloor s_1 \rfloor - \lfloor s_2 \rfloor| > 1$, Statement 4 in the previous lemma gives us the exact value for $\lambda(s_1, s_2)$.
- If $\lfloor s_1 \rfloor - \lfloor s_2 \rfloor = 0$ then by Statement 2 we have $\lambda(s_1, s_2) = \lambda(s_1 - \lfloor s_1 \rfloor, s_2 - \lfloor s_2 \rfloor)$; since it holds that $0 \leq s_1 - \lfloor s_1 \rfloor, s_2 - \lfloor s_2 \rfloor < 1$, by the above assumption $\lambda(s_1, s_2)$ is quite well approximated by $\lambda(s_2 - \lfloor s_2 \rfloor - s_1 + \lfloor s_1 \rfloor, 0) = \lambda(s_2 - s_1, 0) = \left\lceil \frac{1}{2|s_2 - s_1|} \right\rceil + \left\lfloor \frac{1}{2|s_2 - s_1|} \right\rfloor$.
- Finally, suppose that $|\lfloor s_1 \rfloor - \lfloor s_2 \rfloor| = 1$ and $s_1 < s_2$. Then by Statement 1 $\lambda(s_1, s_2) = \lambda(s_1 - \lfloor s_2 \rfloor, s_2 - \lfloor s_2 \rfloor)$. But $s_1 - \lfloor s_2 \rfloor = s_1 - \lfloor s_1 \rfloor - 1 < 0 \leq s_2 - \lfloor s_2 \rfloor$. This means that the line with slope $s_1 - \lfloor s_1 \rfloor - 1$ is over the horizontal line before the origin and below it after the origin, while the line with slope $s_2 - \lfloor s_2 \rfloor$ is either the horizontal line or it is below the horizontal line before the origin and over it after the origin. So, the overlaps of both lines are the points at which both lines overlap the line with slope 0. The number of such overlaps is then bounded by the minimum between $\lambda(s_1 - \lfloor s_1 \rfloor - 1, 0)$ and $\lambda(s_2 - \lfloor s_2 \rfloor, 0)$. By Statement 5, this minimum can be computed as $\lceil \frac{1}{2m} \rceil + \lfloor \frac{1}{2m} \rfloor$, where $m = \max\{1 - s_1 + \lfloor s_1 \rfloor, s_2 - \lfloor s_2 \rfloor\}$.

Define $\varphi : \mathbb{Q} \times \mathbb{Q} \setminus \{(q, q) : q \in \mathbb{Q}\} \to \mathbb{N}$ as the symmetric mapping that, for $s_1 < s_2$ is defined by

$$\varphi(s_1, s_2) = \begin{cases} \left\lceil \frac{1}{2|s_2 - s_1|} \right\rceil + \left\lfloor \frac{1}{2|s_2 - s_1|} \right\rfloor & \text{if } \lfloor s_1 \rfloor = \lfloor s_2 \rfloor \\ \lceil \frac{1}{2m} \rceil + \lfloor \frac{1}{2m} \rfloor, \text{ where } m = \max\{1 - s_1 + \lfloor s_1 \rfloor, s_2 - \lfloor s_2 \rfloor\} & \text{if } \lfloor s_2 \rfloor - \lfloor s_1 \rfloor = 1 \\ 1 & \text{if } \lfloor s_2 \rfloor - \lfloor s_1 \rfloor > 1 \end{cases}$$

and for $s_1 < s_2$ is defined by $\varphi(s_1, s_2) = \varphi(s_2, s_1)$.

We have proved that the assumption $\lambda(s_2 - s_1, 0) \approx \lambda(s_1, s_2)$ whenever $0 \leq s_1, s_2 < 1$ implies the assumption $\lambda(s_1, s_2) \approx \varphi(s_1, s_2)$ for any pair of slopes s_1, s_2.

4 Computational experiments in finite grids

For each integer N we considered all the slopes a/b with $a \in \{-(N-1), \ldots, N-1\}$ and $b \in \{1, \ldots, N-1\}$, that is, the slopes of all lines passing through two different points of an $N \times N$ grid. We considered all pairs of slopes s_1, s_2 in this list and counted the number of cases in which $\lambda(s_1, s_2) \leq \varphi(s_1, s_2)$ and the number of cases in which $\lambda(s_1, s_2) > \varphi(s_1, s_2)$. Figure 7 shows for each N the count of pairs of slopes for which $\lambda(s_1, s_2) \leq \varphi(s_1, s_2)$ (black dot) and the count of pairs for which $\lambda(s_1, s_2) > \varphi(s_1, s_2)$ (white dot). For all tried N, it can be observed that the most common situation is $\lambda(s_1, s_2) \leq \varphi(s_1, s_2)$; in fact, asymptotically with N, the probability of this situation becomes overwhelming.

5 Practical relevance: shared steganographic file systems

While investigating the size of the overlap between two digitized straight line segments may have some interest in image processing (*e.g.* contour tracing, [3]), we motivate here the relevance of this problem in a less traditional context.

Steganographic file systems were introduced in [1] as file systems where the location and even the existence of files are unknown to the users not having stored them. This feature

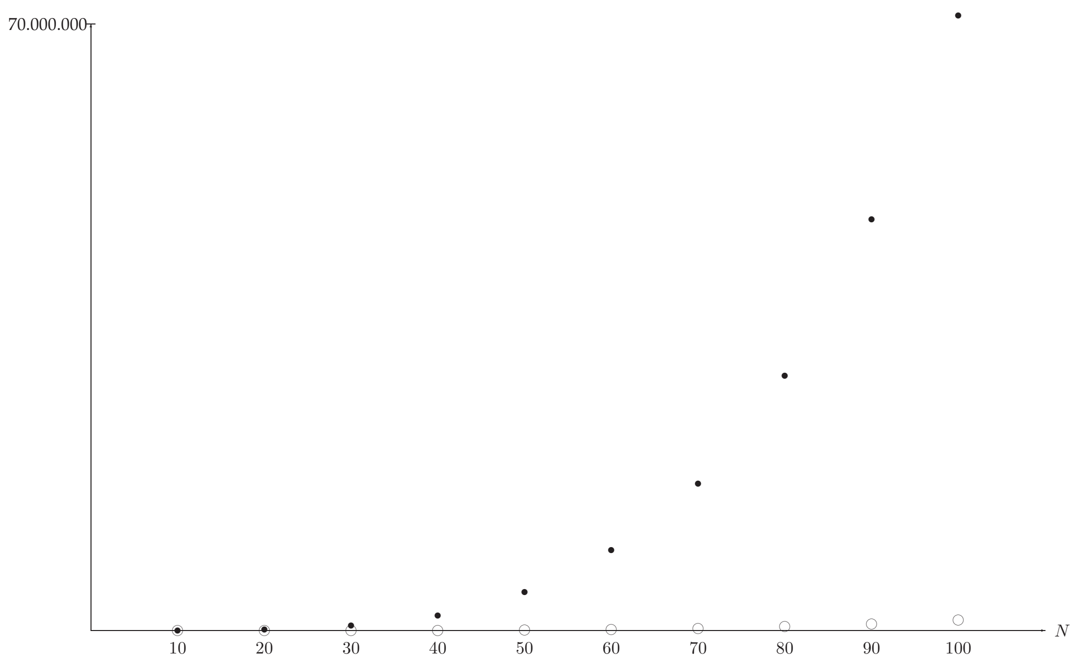

Figure 7: For several values of N, number of pairs of slopes (s_1, s_2) in an $N \times N$ grid for which $\lambda(s_1, s_2) \leq \varphi(s_1, s_2)$ (black dots) and the number of pairs for which $\lambda(s_1, s_2) > \varphi(s_1, s_2)$ (white dots)

becomes a problem when the file system is a shared one that can be written to by several users: as a result of an unknown collision, a user may inadvertently damage the files stored by other users [7].

There are two simplistic approaches to deal with the collision problem:

- *Privacy reduction*. A possible solution is to reduce the risk of collisions between users by reducing the freedom of each user for placing her file in the system. However, the smaller the freedom, the smaller is privacy: the location of the files of a user becomes easier to guess by other users or by external intruders.

- *Efficiency reduction*. If the total size of user files stored in the steganographic file system is less than the size of the file system by several orders of magnitude, collisions are unlikely to occur. However, this entails a very inefficient storage use.

In [2], intermediate solutions between the two approaches above are explored. The idea is to store user files as randomly chosen digitized straight line segments in a two-dimensional $N \times N$ bit matrix. Since the file system is shared, segments corresponding to files of different users may (inadvertently) cross each other. In a crossing between an older and a newer file, each overlapping bit in the older file is tweaked with probability $1/2$, thus resulting in one error. A natural countermeasure is to encode files with an error-correcting code before storing them. Clearly, the expected number of overlapping bits at the crossings between randomly chosen straight line segments is highly relevant in order to select an appropriate code with sufficient error-correcting capability.

6 Conclusions and open research issues

We have presented some results on the number of overlapping points at the crossing between two digitized straight lines whose continuous versions cross each other at a point in the integer grid (*i.e.* with integer coordinates).

Several issues remain to be investigated. One of them is to extend our analysis for the general case in which the crossing of the continuous versions has non-integer coordinates. A second line is to explore the effects of taking moduli in a finite grid: indeed, when representing an infinite digitized straight line on an $N \times N$ grid, both the abscissae and the ordinates must be reduced modulo N. Last but not least, it would be very useful for the application to shared steganographic file systems to be able to compute the expected number of overlapping points between two digitized straight line segments randomly chosen in an $N \times N$ grid.

Disclaimer and acknowledgments

The authors are with the UNESCO Chair in Data Privacy, but they are solely responsible for the views expressed in this paper, which do not necessarily reflect the position of UNESCO nor commit that organization. This work was partly supported by the Spanish Government through projects TSI2007-65406-C03-01 "E-AEGIS" and CONSOLIDER INGENIO 2010 CSD2007-00004 "ARES", and by the Government of Catalonia under grant 2005 SGR 00446.

References

[1] R. Anderson, R. Needham and A. Shamir. The steganographic file system. In D. Aucsmith (ed.), *Information Hiding, 2nd International Workshop*, Portland, Oregon, USA, 1998.

[2] J. Domingo-Ferrer and M. Bras-Amorós. Shared steganographic file systems with error correction. In V. Torra and Y. Narukawa (eds.), *Modeling Decisions for Artificial Intelligence-MDAI 2008*, Lecture Notes in Computer Science, 5285: 227–238, 2008.

[3] R. C. Gonzalez and R. E. Woods. *Digitized Image Processing*. Reading, MA: Prentice-Hall, 2007.

[4] R. Kletter and A. Rosenfeld. *Geometric Methods for Digitized Picture Analysis*. San Francisco, CA: Morgan Kaufmann, 2004.

[5] J. Koplowitz, M. Lindenbaum and A. Bruckstein. The number of digitized straight lines on an $N \times N$ grid. *IEEE Transactions on Information Theory*, 36(1): 192–197, 1990.

[6] A. Rosenfeld. Digitized straight line segments. *IEEE Transactions on Computers*, C-23: 1264–1269, 1974.

[7] X. Zhou. *Steganographic File System*. Ph. D. Thesis, School of Computing, National University of Singapore, 2005.