Loss of Votes in NIDC Applying Storage in Parallel Channels

Pablo García[1], Germán Montejano[1,2], Silvia Bast[1], and Estela Fritz[1]

[1] FCEyN - Universidad Nacional de La Pampa - Argentina
pablogarcia@exactas.unlpam.edu.ar,
WWW home page: http://www.exactas.unlpam.edu.ar
[2] FCFMyN - Universidad Nacional de San Luis - Argentina
gmonte@unsl.edu.ar,
WWW home page: http://www.unsl.edu.ar

Abstract. Birthday Paradox states that in a group of 23 people, the probability that there are at least two who share the same birthday is very close to $\frac{1}{2}$. This assertion is unacceptable for any scheme that proposes a vote storage method based on a vector of slots whose position is chosen at random. In this situation, it may produce collisions.
A collision occurs when two or more votes are stored in the same slot. It produces the loss of the coincident votes. This is the original model of the Non - Interactive Dining Cryptographers (NIDC) protocol.
The actual paper shows new achieved results obtained by analyzing the behaviour of a storage technique based on parallel channels. This scheme consists of replicating each vote in Q parallel channels, keeping the total number of slots (T) without variation.

Keywords: Parallel Channels, Storage Birthday Paradox, Non - Interactive Dining Cryptographers, Collisions.

1 Introduction

Within the scope of a research line that began at 2013 and which was formally presented in [1], the exact security level requested for anonymity in an electronic voting scheme was analyzed. Many of the proposed schemes (Mix Net based) give unconditional security to the votes' information and computational assurance to voter's privacy. However, it is easy to see that it is an erroneous proposal . In [2] it is concluded that it is necessary to give unconditional security for the privacy, because it must be protected indefinitely. Otherwise, votes must be kept for a finite period of time.

Consequently, those schemes, that include unconditional security as the main feature, acquire maximun interest. In this sense, one of the most interesting is Dinning Cryptographers (DC), which is described in detail in [3]. This protocol is resourceful and gives unconditional privacy.

The analysis is focused on a derivative of DC, called Non Interactive Dining Cryptographers (NIDC [4]), that relaxes the condition of concurrency online

for all participants. This protocol is suitable to be applied to electronic voting scheme.

The original version of NIDC, stores data in a vector of slots. This is observed in figure 1.

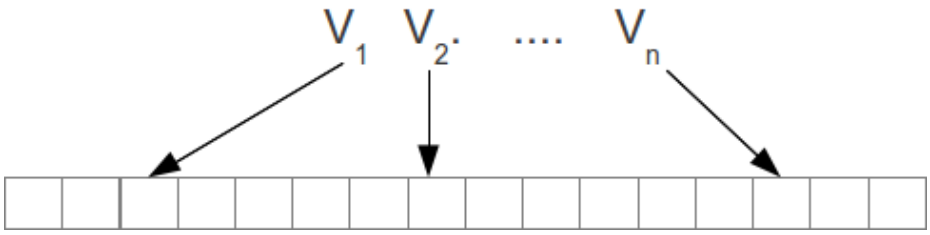

Fig. 1. Original NIDC Storage

If two or more votes are stored in the same slot, a collision occurs. That results in the loss of coincident votes. Simultaneously, the proposition of true randomness for the choice of position indicates that collisions may happen. It then seeks to ensure that the proportion of lost data is kept below a desired value with a certain probability. The proposed model in Figure 1 may be explained by Birthday Paradox ([5]). In those conditions, it is required a very significant number of slots to obtain suitable security levels.

Two interesting alternatives, aimed at improving the Birthday Paradox effect are presented in [6]. In that document an optimization for NIDC, applying multiple networks in serie and in parallel, is proposed. In this case, however, it seeks to generalize the approach to storage in parallel channels, a matter that may be generalized to multiple real-world problems, including NIDC. The alternative proposal consists on implementing N parallel channels, replicating each vote in all channels, in potentially different random positions in each case, as outlined in Figure 2.

It begins by describing the parameters involved:

T**:** # Total slots to implement. $T \in Z^+$.
S**:** # Parallel slots on each channel. $S \in Z^+ \wedge S \leq T$.
N**:** # Voters. $N \in Z^+$.
Q**:** # Parallel channels to implement. $Q \in Z^+$.
Q_{to}**:** # Parallel channels to implement (Theorically Optimal). $Q_{to} \in R^+$.
Q_{po}**:** # Parallel channels to implement (Practically Optimal). $Q_{po} \in Z^+$.
R**:** # Replicas of a vote on the same channel. $R \in Z^+$.
PLV**:** Percentage of Lost Votes.

Throughout previous papers ([7], [8] and [9]) the following relevant findings have been set forth (in adition, function CEIL will be used; it computes the nearest higher integer. That is necessary because Q_{to} could be non integer):

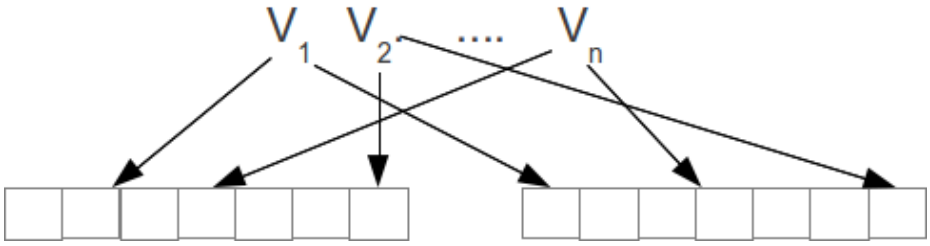

Fig. 2. Scheme based on Parallel Channels

– For a fixed number of voters N, the recommended number of slots for each parallel channel (S) is given by the formula:

$$S = CEIL(\frac{N}{ln2}) \tag{1}$$

– For given values of T and N, there exists an optimal number of parallel channels. Such value is expressed:

$$Q_{to} = ln2\frac{T}{N} \tag{2}$$

That formula should be taken to the next integer.

$$Q_{po} = CEIL(Q_{to}) \tag{3}$$

– An appropriate lower bound for the probability of X = "no vote is lost" is obtained by applying equation:

$$Pr(X) > (1 - (\frac{1}{s}(n-1))^q)^n \tag{4}$$

Besides, concrete methods were published to obtain optimal values for all parameters using a spreadsheet ([10]) and the pseudocode algorithm that must be applied for the same purpose was shown in [11].

In this document the variable PLV is analyzed. One equation is described to get an approximation of the expected value of that variable. The following section describes the deduction of such formula .

2 Expected Value of Percentage of Lost Votes (PLV)

By applying a parallel channels scheme, a question that quickly arises is: for a situation with N voters, and Q parallel channels of S slots (such that $T = SQ$), which is the expected Percentage of Lost Votes (PLV)?

At the beginnig, it is considered the original Birthday Paradox proposal. The first thing we see is that even in the best case (the 23 people Birthdays on different dates), the number of slots that will not be used is 342, then we have approximately 6,3 % of occupied slots and 93,7 % of empty slots. Consequently, for each slot containing a vote, more than 15 receive no ballots.

The proposal is to divide all the slots in $Q > 1$ parallel channels and to deposit an occurrence of each vote in each of the channels. In addition to what appeared in the simulations, the idea is related to the fact that a vote is lost on a given channel is independent of what happens in the other $Q - 1$ channels.

Independent events verify that:

$$Pr(A \cap B) = Pr(A)Pr(B) \tag{5}$$

Clearly a vote will be lost only if it collides on all the channels. The number of local collisions increases, since each channel will have a measure smaller than the single vector. However, an optimization based on replicas is obtained.

Let ε be:

$$\varepsilon = \frac{N}{S} \tag{6}$$

Initially the situation is analyzed if a single vector is implemented, therefore, $S = T$.

Several strategies based on analyzing the probability distribution are presented in [12]. Also, the approaches proposed by Feller [13], were mentioned. These approaches improve their behaviour when $N \to \infty$ y $T \to \infty$. A tool that may be useful is Stirling's approximation for calculating factorials:

$$N! = \sqrt{2\Pi N}(\frac{X}{N})^N \tag{7}$$

It is proposed another approach, which is simpler than the previous one because it only calculates expected values rather than probability distribution.

Considering the first vote, the probability that it falls into the slot 1 is:

$$p = \frac{1}{S} \tag{8}$$

Consequently, the probability that it does not fall into the slot 1 is:

$$q = 1 - p = (1 - \frac{1}{S}) \tag{9}$$

Generalizing to N votes, we get a binomial distribution with parameters N and p.

Let X_k be: "Exactly k votes are stored in slot 1" with $k \in Z^+$

$$Pr(X_k) = \binom{N}{k} p^k q^{N-k} \tag{10}$$

$$Pr(X_k) = \binom{N}{k} (\frac{1}{S})^k (1 - \frac{1}{S})^{N-k} \tag{11}$$

Given that:

$$\lim_{x \to \infty} (1 + \frac{1}{n})^n = e \tag{12}$$

We can assure:

$$Pr(X_0) = (1 - \tfrac{1}{S})^N \approx e^{-\varepsilon}$$
$$Pr(X_1) = N\tfrac{1}{S}(1 - \tfrac{1}{S})^{N-1} \approx \varepsilon e^{-\varepsilon}$$
$$Pr(X_2) = \tfrac{N(N-1)}{2}(\tfrac{1}{S})^2(1 - \tfrac{1}{S})^{N-2} \approx \tfrac{1}{2}\varepsilon^2 e^{-\varepsilon}$$

These probabilities also represent the expected number of votes in slot 1. It is obvious that the same reasoning can be applied to any slot. Therefore, it is possible to find the expected frequency.

Given that $\lim_{x\to\infty}(1 + \frac{1}{n})^n = e$, for $N = S = 1000$, $\varepsilon = 1$. Therefore:

$$Pr(X_0) = Pr(X_1) = e^{-1} \approx 0.3678 \tag{13}$$

Similarly, for $N = 500, S = 1000$, $\varepsilon = \frac{1}{2}$, in which case:

$$Pr(X_0) = e^{-\frac{1}{2}} \approx 0.6065$$
$$Pr(X_1) = \tfrac{1}{2}e^{-\frac{1}{2}} \approx 0.3032$$

Let $E(k)$ be: #expected slots containing k votes. Its value is obtained as follows:

$$E(k) = Sp(X_k) \tag{14}$$

This fits together with the Poisson approximation stated in [15] :

$$E(Poisson(\lambda)] = \lambda \tag{15}$$

$$\lambda[X_0] = Se^{-\frac{N}{S}} \tag{16}$$

For $k = 1$:

$$\lambda[X_1] = \frac{(ne^{\frac{-r}{n}})}{k!}(\frac{r}{n})^k = \frac{Se^{(-\varepsilon)}}{k!}\varepsilon = S\varepsilon e^{-\varepsilon} \tag{17}$$

For $k = 2$:

$$\lambda[X_2] = \frac{ne^{(\frac{-r}{n})}}{k!}(\frac{r}{n})^k = \frac{Se^{-\varepsilon}}{2}\varepsilon^2 = S\varepsilon^2 e^{-\varepsilon} \tag{18}$$

The Poisson approximation improves its quality when $S \to \infty$ and $N \to \infty$. The previous formula is related to S. It is more interesting yet, to obtain a connection with the number of successful votes, ie for $k = 1$, $E(1)$ is divided into N and it is obtained:

$$\frac{S\varepsilon e^{-\varepsilon}}{N} = e^{-\varepsilon} \tag{19}$$

It is possible to generalize the approach to Q channels , with $Q > 1$. For example, for $S = N = 1000$:

$$s[1] = 1000e^{-1} \approx 368 \tag{20}$$

Consequently, for $Q = 1$:

$$Pr(successfulvote) \approx 0.36 \tag{21}$$

$$Pr(lostvote) \approx 1 - 0.36 = 0.64 \tag{22}$$

For $Q = 2$, one vote is lost if collides in the two channels:

$$Pr(successfulvote) = 1 - 0.39 \approx 0.61 \tag{23}$$

$$Pr(lostvote) = 0.64^2 \approx 0.39 \tag{24}$$

The same scheme is generalized $\forall Q > 2$. Thus it is obtained a formula to calculate the expected value of the variable *Percentage of Lost Votes*.

$$\mid PLV \mid = (1 - e^{-\frac{n}{s}})^q \tag{25}$$

2.1 Practical Verification of the Proposed Formula

Given formulas above, a simulator has been implemented which two main aims:

1. To verify the correctness of formulas.
2. To bear out that the approach of storing in parallel channels optimizes the results in terms of several variables which may be considered.

The simulator is implemented allowing the following inputs:

1. Total number of slots to implement (T).
2. Number f voters (N).
3. Quantity of parallel channels to implement (Q).
4. Quantity of election acts that will be simulated by session (R).

The simulator verifies that the total number of slots (T) is a multiple of quantity of parallel channel, because the quantity of slots in each channel (S) must be an integer number.

When the simulation is complete, the following information may be obtained:

1. Total of successful votes (SV).
2. Total of lost votes (LV).
3. Quantity of runs where at least one vote is lost (R)
4. Quantity of runs (Votings) without lost votes ($RWLV$)
5. Quantity of runs (Votings) with lost votes (RLV)
6. Best case, that is to say, how many votes were lost in the most successful run (BC).
7. Worst case, that is to say, how many votes were lost in the less successful run (WC).

Therefore, we will observe the behaviour of the formula (25) based on the next ratio:

$$SPLV = \frac{LV}{SV + LV} \tag{26}$$

Table 1 shows the values that were obtained in different simulations and the difference between those and the analytical results obtained by application of equation (25). With this purpose, the following variables are introduced:

- FV: Values obtained by application of formula (25).
- SV: Values obtained by simulation.

Watching the values of Table 1, the difference between FV and SV remains at very low values. Specifically:

- The maximum one is 0,002452561.
- The minimum one is 2,80437E-07.

N	T	S	Q	FV	SV	$DIFFERENCE$
15	150	15	10	0,010185894	0,008466667	0,001719227
15	300	30	10	8,89424E-05	0	8,89424E-05
15	450	45	10	3,35005E-06	0	3,35005E-06
15	600	60	10	2,80437E-07	0	2,80437E-07
30	300	30	10	0,010185894	0,007733333	0,002452561
30	600	60	10	8,89424E-05	0	8,89424E-05
30	900	90	10	3,35005E-06	0	3,35005E-06
30	1200	120	10	2,80437E-07	0	2,80437E-07
60	600	60	10	0,010185894	0,0098	0,000385894
60	1200	120	10	8,89424E-05	0,00015	-6,10576E-05
60	1800	180	10	3,35005E-06	0	3,35005E-06
60	2400	240	10	2,80437E-07	0	2,80437E-07
120	1200	120	10	0,010185894	0,012025	-0,001839106
120	2400	240	10	8,89424E-05	0	8,89424E-05
120	3600	360	10	3,35005E-06	0	3,35005E-06
120	4800	480	10	2,80437E-07	0	2,80437E-07
240	2400	240	10	0,010185894	0,009129167	0,001056727
240	4800	480	10	8,89424E-05	0	8,89424E-05
240	7200	720	10	3,35005E-06	0	3,35005E-06
240	9600	960	10	2,80437E-07	0	2,80437E-07
360	3600	360	10	0,010185894	0,008391667	0,001794227
360	7200	720	10	8,89424E-05	1,11E-05	7,78313E-05
360	10800	1080	10	3,35005E-06	0	3,35005E-06
360	14400	1440	10	2,80437E-07	0	2,80437E-07
480	4800	480	10	0,010185894	0,00988125	0,000304644
480	9600	960	10	8,89424E-05	0	8,89424E-05
480	14400	1440	10	3,35005E-06	0	3,35005E-06
480	19200	1920	10	2,80437E-07	0	2,80437E-07

Table 1. Difference between FV and SV

– The average value is: 0,000227181

Another aspect which should be highlighted is that the formula (25) works better when $N < S$. As both values approach, the behaviour is worse. For example, Figure 3 shows the values of FV and SV with de following values for the parameters:

– $N = (7..15)$
– $T = 150$
– $Q = 10$
– $S = 15$

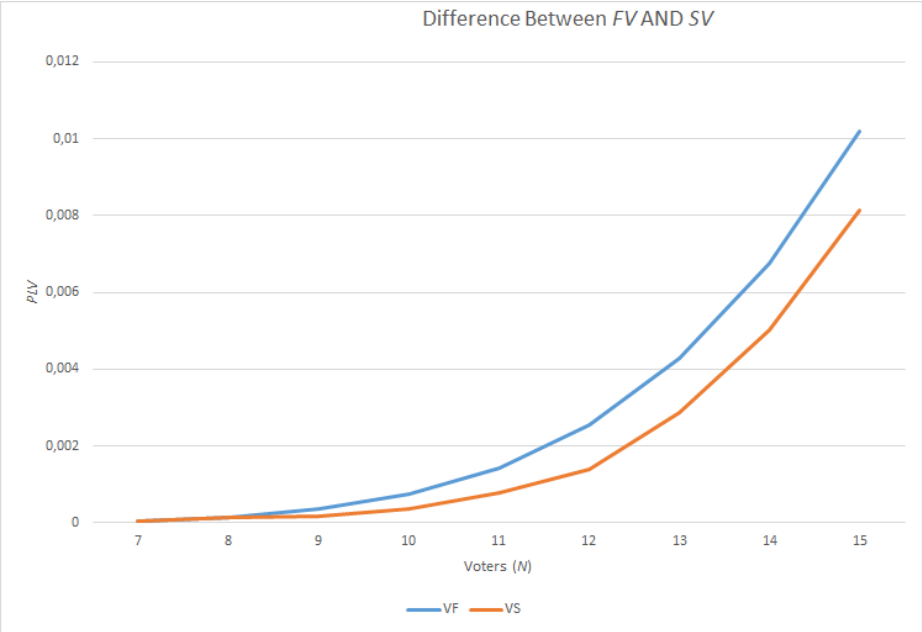

Fig. 3. PLV: Difference Between FV AND SV

3 Conclusions

The approach based on parallel channels optimizes the use of storage space intended to store data whose location is truly random. The formulas (1), (2), (3), (4) y (25) accurately describe the behaviour of the model.

Specifically, the results obtained by the formula (25) are very close to the values obtained in the simulations, though the difference increases when N is close to S. Even in that case, the behaviour of the formula is acceptable.

References

1. Uzal R., van de Graaf J., Montejano G., Riesco D., García P.: Inicio de la Línea de Investigación: Ingeniería de Software y Defensa Cibernética. WICC 2013. Ps. 769 - 773. ISBN: 9789872817961. (2013).
2. van de Graaf J., Montejano G., García P.: Manejo de Colisiones en un Protocolo Non Interactive Dining Cryptographers. JAIIO 2013. ISSN: 18502776. WSegI 2013. ISSN: 23139110. Ps. 29 a 43. (2013).
3. Chaum D.: The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability Journal of Cryptology. (1988).
4. van de Graaf J.: Anonymous One Time Broadcast Using Non Interactive Dining Cryptographer Nets with Applications to Voting Towards Trustworthy Elections. Ps 231-241. Springer-Verlag Berlin, Heidelberg. ISBN:9783642129797. (2010).
5. Flajolet P., Gardy D., Thimonier L.: Birthday paradox, coupon collectors, caching algorithms and self-organizing search. Discrete Applied Mathematics 39. Ps. 207-223. North-Holland. (1992).
6. García P., van de Graaf J., Hevia A., Viola A.: Beating the Birthday Paradox in Dining Cryptographer Networks. The third International Conference on Cryptology and Information Security in Latin America, Latincrypt 2014. September 17-19, 2014. Florianopolis, Brasil. Lecture Notes in Computer Science, Springer (2014).
7. García P.,van de Graaf J., Montejano G., Bast S., Testa O.: Implementacin de Canales Paralelos en un Protocolo Non - Interactive Dining Cryptographers. JAIIO 2014. ISSN 18502776. WSegI 2014. ISSN: 23139110. (2014)
8. García P., Montejano G., Bast S.: Aspectos optimizables en un Protocolo Non-Interactive Dining Cryptographers. CONAIISI 2014. ISSN: 23469927.(2014).
9. García P., Montejano G., Bast S, Fritz E.: Anonimato en Sistemas de Voto Electrónico: Últimos Avances. WICC 2016. ISBN: 9789506983772. (2016)
10. García P., van de Graaf J. Montejano G., Riesco D., Debnath N., Bast S.: Storage Optimization for Non - Interactive Dining Cryptographers (NIDC) Information Technology - New Generations (ITNG). Ps. 55 - 60. ISBN: 978-1-4799-8827-3. DOI: 10.1109/ITNG.2015.15. IEEE. (2015)
11. García P., Bast S., Fritz E., Montejano G., Riesco D., Debnath N.: A systematic method for choosing optimal parameters for storage in parallel channels of slots International Conference on Industrial Technology (ICIT)- Ps. 1700 - 1705. DOI:10.1109/ICIT.2016.7475019 IEEE - 2016
12. van de Graaf J., Montejano G., García P.: Optimización de un esquema Ocuppancy Problem orientado a E Voting. WICC 2013. Ps. 749 - 753. ISBN: 9789872817961. (2013).
13. Feller W.: An Introduction to Probability Theory and its Applications. Volmen I. Third Edition. John Wiley and Sons. New York. (1957).