Collided birthdays are welcome!

During inventory in a densely packed stock, multiple radio-frequency data transmitters often interfere with each other, leading to signal collisions. This reduces the efficiency of inventory. We present a method to resolve these collisions. The good news: despite these collisions, the items can still be identified, and their signals can be reconstructed. This advancement greatly enhances the performance of radio-frequency identification (RFID) systems. [1]

Keywords: RFID; I/Q data; aloha collision; signal separation; self-modeling regression

1 Collision-free inventory

The Aloha protocol addresses the issue of overlapping replies by dividing the inventory time segment into discrete time slots. When prompted, each tag waits for a random number of time slots before responding. Due to this randomness, some time slots remain unoccupied, some contain only a single tag ID, while others may still be occupied by multiple tags. The Aloha protocol resolves a mixture of replies: the inventory time-segment splits into time-slots. At request, each tag waits a random number of time slots and replies. Due to this randomness, some time slots are left unoccupied, some time slots keep a single tag ID, but some are still occupied by several tags. Can we avoid collisions in one inventory cycle? The problem of estimation of probability that two tags hit one slot is called *the birthday paradox* [1, 2]. What is probability of a two people have their birthdays in the same day? One tag hits any of D slots with the probability of $\frac{1}{D}$. Two tags do not hit the same slot with the probability $1 - \frac{1}{D}$. The third tag cannot hit both occupied slots, so the probability is

Figure 1: The goal is to inventory over 1,000 tags at once.

$$\frac{D-1}{D}\frac{D-2}{D} = \left(1 - \frac{1}{D}\right)\left(1 - \frac{2}{D}\right).$$

So for given D slots, the probability that none of N tags do not collide is

$$\frac{D!}{D^N(D-N)!}.$$

Figure 2 shows that the probability of successful inventory is small for any reasonable number of tags. So if the shopping cart has over 100 items with tags, most likely there is collision even for a long inventory cycle. See the green and red lines.

[1]Vadim Strizhov, vadim.vct@gmail.com, 18/03/2025

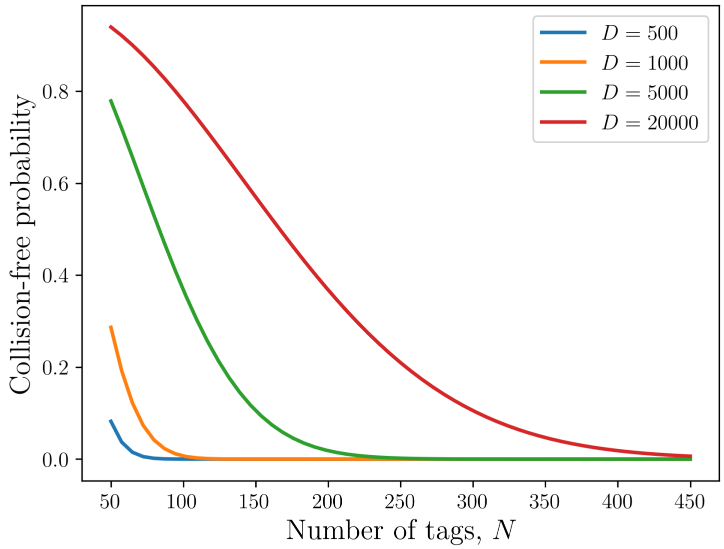

Figure 2: The probability of collision-free inventory of any of N given D time-slots.

If, with an insufficiently small number of slots, there is no initial period where the probability of getting two transmitters in one slot increases. That is, if there are enough transmitters to overlap at all, they will *immediately start crowding* into multiple transmissions per slot.

Briefly, *the collision is inaviodable in one inventory cycle.*

2 Signal self-modeling

The collision should be detected to avoid inventory errors. But there is no error in the signals, reconstructed after the collision. So we suppose two or more tags transmit at the same time.

The inventory reader decodes the high-frequency signal into the I/Q data signal (In-phase/Quadrature). This signal carries two time series, real and imaginary. Denote these time series by $\mathbf{x}$, a vector in the complex space.

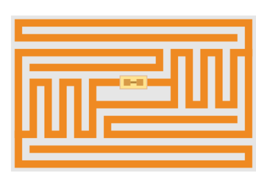

Figure 3: A tag emits an ultra-high-frequency signal through its antenna.

Since the tags are located in the different parts of shopping cart its signal is varied by phase and amplitude. Figure 4 shows the same signal with the phase and amplitude modifications. The self-regression model approximates these signals with only two parameters: scale and shift.

The self-modeling regression. It approximates the signal $\mathbf{x}$ with the standard signal $\mathbf{c}$ (call it the centroid) as

$$\hat{\mathbf{c}} = \text{scale} \cdot \big(\text{shift}(\mathbf{x})\big),$$

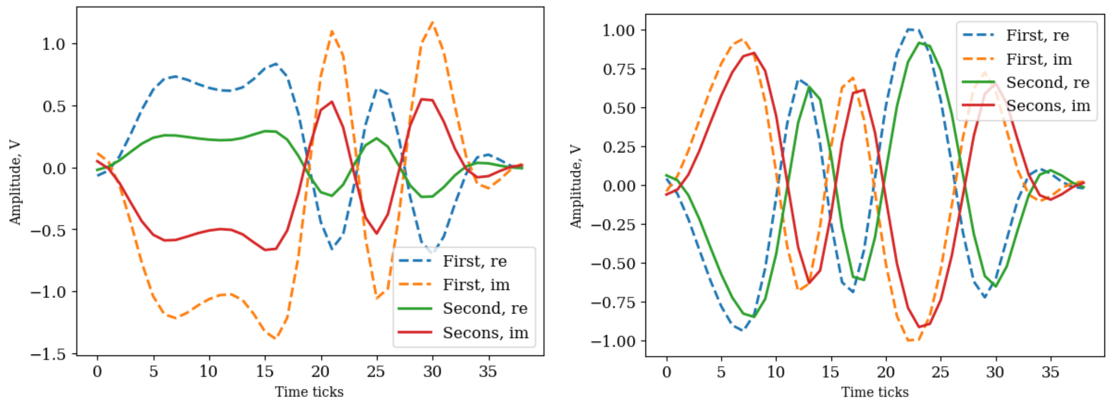

Figure 4: The self-modeling regression regresses the first signal to the second (left), while shifting the phase of the whole I/Q data signal to find the best fit (right). The legend shows the real and imaginary parts of the complex signal.

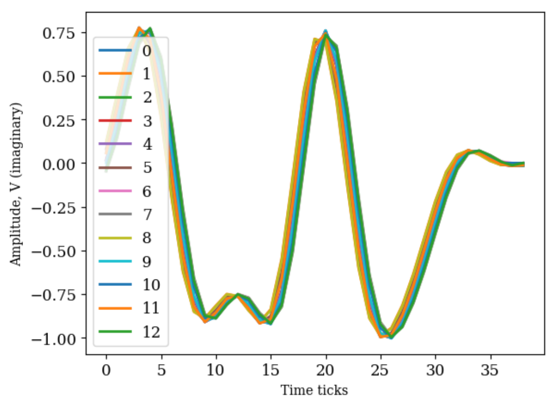

Figure 5: After the self-modeling regression, twelve different transmissions became similar (imaginary part is shown).

with two scalar parameters: scale and shift. The first parameter is calculated as the dot product ratio of the projection

$$\text{scale} \cdot \mathbf{x} = \frac{\mathbf{c}^\mathsf{T}\mathbf{x}}{\|\mathbf{c}\|^2}\mathbf{c}.$$

The second parameter calculated as an argument of minimum distance

$$\text{shift} = \min \|\hat{\mathbf{c}} - \mathbf{c}\|^2.$$

The result of self-modeling is shown in Figure 5

Briefly, *self-modeling unifies the signal shape* of I/Q data, and it makes it a tool to analyze the signal mixtures.

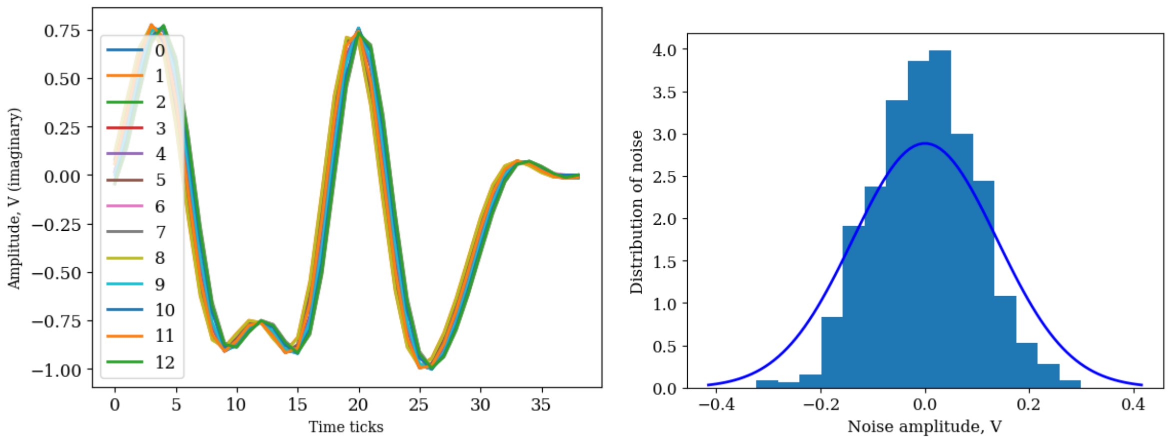

Figure 6: The left part of figure shows the cluster of transmitted signals with the same message without noise. They scaled to the same shift phase and amplitude. The ground truth noise (right part), samples from the dataset, has $\mu \leq 0.005$ (real or image) and standard deviation $\sigma = 0.1361$. The plot shows 3σ. This distribution function us used in the ≥ 3 signal reconstruction procedure.

3 Signal separation

When two or more tags hit the same time slot, their signals mix. Due to the various antenna orientations, they mix with different coefficients. For n tags denote the coefficients $[v_1, \dots, v_n]$. The received noisy mixture $\mathbf{y}$ approximated the weighted I/Q data signals in the best way:

$$\mathbf{y} \approx \mathbf{p} = v_1\mathbf{x}_1 + \dots + v_n\mathbf{x}_n = \mathbf{X}\mathbf{v}$$

The columns of the matrix $\mathbf{X}$ are the stacked I/Q data signals. The coefficients $\mathbf{v}$ and their number n are unknown. But for any mixture coefficients, the signals of collided tags are in the subspace of the space of the matrix $\mathbf{X}$. Using the self-regression model, find the source signals as the nearest linear combination to the received mixture, see Figure 7. There is no need to use methods like blind signal separation. The self-modeling regression works even for a single-antenna reader.

Figure 8 shows that the most expected types of collision: two and three tags hit the same slot. It delivers good TODO I/Q data identity reconstruction.

Briefly, *collision does not matter for the separated signals.*

4 Development

The code and report on the collision reconstruction: https://github.com/vadim-vic/Signal-separation

The report on the collision detection: https://github.com/vadim-vic/Signal-separation/blob/main/latex

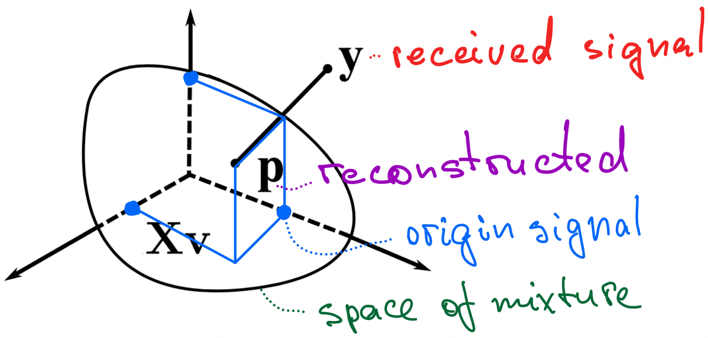

Figure 7: Two or more signals mix proportionally to their attenuation. It defines the vector span in the space of I/Q data signals. The vector $\mathbf{v}$ is the weights of the linear combinations of the signals. The vector $\mathbf{p}$ is the orthogonal projection to the span $\mathbf{Xv}$. The vector $\mathbf{y}$ is the mixture of signals and the added noise to be reconstructed. The basis of P independent (the transmitters can not send the same data) I/Q data signals $\mathbf{x}_1, \ldots, \mathbf{x}_P$ form the matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_P]$ as its columns.

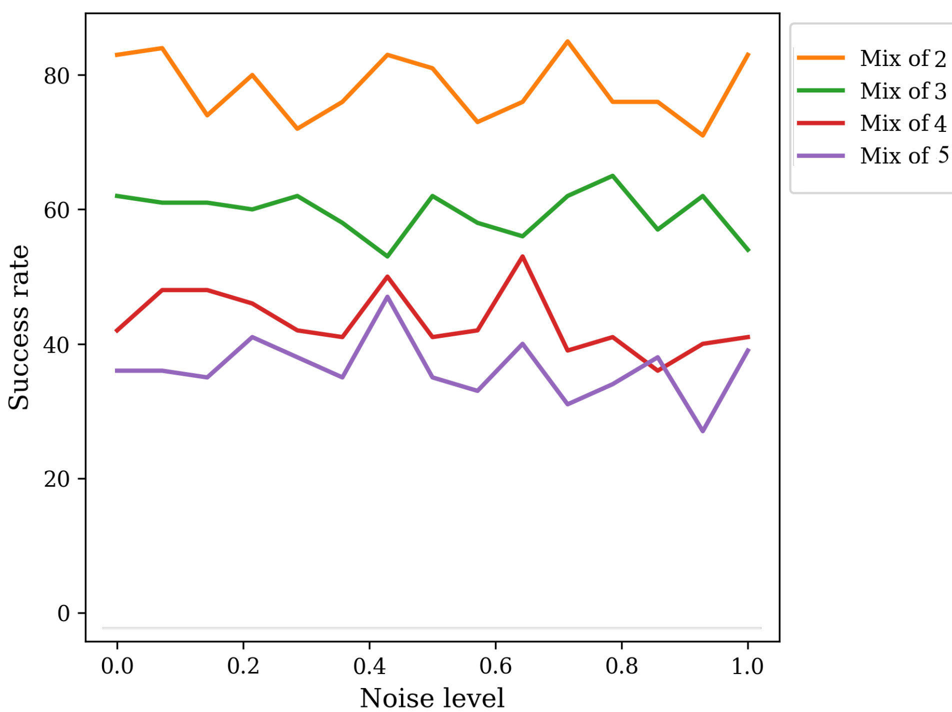

Figure 8: Each line is the number of successfully recognized IDs of the tags after the signal reconstruction. The x-axis shows level of noise from the expected standard deviation to zero. The y-axis shows the proportion (or number TODO) of randomly generated data.

Though the developed model is portable to an RFID reader, there are ways improve the signal separation. They origins are published by Aleksandr Katrutsa and Vadim Strijov in the papers [3, 4]

1. Comprehensive study of feature selection methods to solve multicollinearity problem // Expert Systems with Application, 2017

2. Stress test procedure for feature selection algorithms // Chemometrics and Intelligent Laboratory Systems, 2015

The further models include analysis of mixture in the high-frequency domain under signal interference conditions.

We welcome the company, involved to RFID software and hardware development: if you are interested to support this research and join a grant application, please contact the author.

References

[1] Carla Santos and Cristina Dias. A probabilistic approach on coincidences: the birthday paradox. *Pensamiento Matematico*, 2015.

[2] Frederick Mosteller. Understanding the birthday problem. *The Mathematics Teacher*, 55(5):322–325, May 1962.

[3] Alexandr Katrutsa and Vadim Strijov. Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria. *Expert Systems with Applications*, 76:1–11, June 2017.

[4] A.M. Katrutsa and V.V. Strijov. Stress test procedure for feature selection algorithms. *Chemometrics and Intelligent Laboratory Systems*, 142:172–183, March 2015.