Galileo NMA Signal Unpredictability and Anti-Replay Protection

Ignacio Fernández-Hernández
European Commission DG GROW
Brussels, Belgium

Gonzalo Seco-Granados
Universitat Autònoma de Barcelona (UAB)
Barcelona, Spain

***Abstract*—Galileo is studying the addition of Navigation Message Authentication (NMA) to its Open Service. NMA cryptographic data is mostly unpredictable, and this unpredictability can provide certain protection against replay attacks, depending on the user environment and how the signal is processed in the receiver. This paper first characterises Galileo NMA symbol unpredictability, which depends on how the NMA data is packaged within the I/NAV message, and how the bits are coded into symbols and subsequenly interleaved. Next, a replay protection method is proposed and preliminarily characterised. By storing the first chips of every unpredictable symbol, a receiver can create a synthetic sequence whose correlation gain will be low if the tracked signal is being spoofed.**

I. Introduction

For the last years, Galileo is studying the addition of Navigation Message Authentication (NMA) to its Open Service. NMA is essentially aimed at the authentication of the satellite navigation data. However, the cryptographic information added to authenticate the data is unpredictable, unlike the standard navigation data, which is updated at intervals in the order of an hour. This unpredictability depends on both how the NMA data is packaged within the message, and how the bits are coded and interleaved. While better protection against signal replay attacks is achieved with spreading code-level authentication (SCA), NMA unpredictability can help in the protection against replay attacks, depending on the user environment and receiver. Given the potential variety of users, receivers and NMA applications, and given that SCA cannot be offered for OS users in the Galileo first generation, maximizing unpredictability through NMA seems to be desirable, assuming this does not penalize navigation or NMA performance.

This paper characterizes the NMA-induced unpredictability of Galileo I/NAV signal, after FEC coding and interleaving. It also presents how symbol unpredictability can theoretically help a user receiver discriminate between an authentic and a replayed signal, as the attacker needs to accumulate some energy at the beginning of each unpredictable symbol before estimating its sign.

II. NMA Unpredictability in Galileo E1-B I/NAV

The Galileo I/NAV is transmitted in the signals E1 (1575.42 MHz) and E5b (1207.14 MHz). NMA is currently designed for the E1-B (data) component. Satellites transmit a navigation frame every 750 seconds, composed by 25 subframes of 30 seconds duration each. Every subframe is divided into fifteen

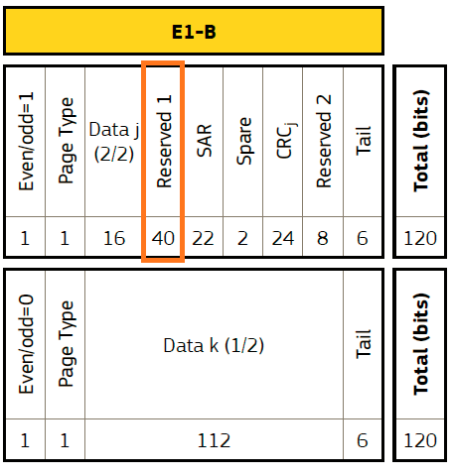

Fig. 1. "Reserved 1" field in I/NAV pages.

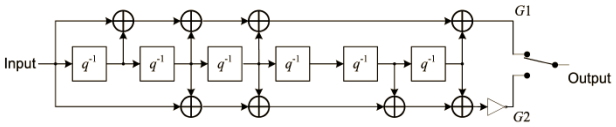

Fig. 2. Galileo I/NAV convolutional encoding.

2-second pages, each of which contains one word and some other fields [1]. The I/NAV effective bit rate is 120 bps. Every page has a 40-bit field, called "Reserved 1", which is the one proposed for NMA, as shown in Fig.1. As other Galileo signals, the I/NAV message is convolutionally encoded using the following coding parameters:

- Coding rate: 1/2
- Coding scheme: convolutional
- Constraint length: 7
- Generator polynomials: G1=171o ; G2 = 133o (e.g. 171o in octal is 1111001 in binary)
- Encoding sequence: G1, then G2.

Each bit is encoded into two symbols (i.e. coded bits), which depend on the present and past bits according to the scheme in Fig.2. Further details on convolutional encoding and interleaving can be found in [2], Sec. 8.8.2. In addition to the 120 bits leading to 240 symbols, a 10-symbol synchronization sequence is added, for a total of 250 symbols in 1 second.

978-1-5090-1757-7/16/$31.00 ©

1	9	17	25	33	41	49	57	65	73	81	89	97	105	113	121	129	233
2	10	18	26	34	42	50	58	66	74	82	90	98	106	114	122	130	234
3	11	19	27	35	43	51	59	67	75	83	91	99	107	115	123	131	235
4	12	20	28	36	44	52	60	68	76	84	92	100	108	116	124	132	236
5	13	21	29	37	45	53	61	69	77	85	93	101	109	117	125	133	237
6	14	22	30	38	46	54	62	70	78	86	94	102	110	118	126	134	238
7	15	23	31	39	47	55	63	71	79	87	95	103	111	119	127	135	239
8	16	24	32	40	48	56	64	72	80	88	96	104	112	120	128	...	240

Fig. 3. I/NAV 240-symbol unpredictability before interleaving. Green: unpredictable symbols. Yellow: predictable symbols based on unpredictable bits. White: predictable symbols based on predictable bits, assuming all 32 last "Reserved 1" bits are unpredictable.

1	9	17	25	33	41	49	57	65	73	81	89	97	105	113	121	129	233
2	10	18	26	34	42	50	58	66	74	82	90	98	106	114	122	130	234
3	11	19	27	35	43	51	59	67	75	83	91	99	107	115	123	131	235
4	12	20	28	36	44	52	60	68	76	84	92	100	108	116	124	132	236
5	13	21	29	37	45	53	61	69	77	85	93	101	109	117	125	133	237
6	14	22	30	38	46	54	62	70	78	86	94	102	110	118	126	134	238
7	15	23	31	39	47	55	63	71	79	87	95	103	111	119	127	135	239
8	16	24	32	40	48	56	64	72	80	88	96	104	112	120	128	...	240

Fig. 4. I/NAV 240 symbol unpredictability after encoding and interleaving

After the encoding, symbols are interleaved to add robustness against temporary fading effects in the channel. Galileo I/NAV interleaving consists of a block interleaver of size 240, i.e. all the symbols corresponding to the odd or even part of an I/NAV page except the synchronization sequence, and dimensions of 30 columns x 8 rows. That means that the 240 symbols are written column-wise in blocks of 8 symbols, and read row-wise. Most receivers use a Viterbi decoder to recover the bit from the received symbols [3].

One can anticipate that the entropy of the signal will not be increased by coding, i.e. that if there are 40 unpredictable bits, which are coded into 80 symbols, only 40 symbols will be unpredictable. However, in order to understand their place in the symbol stream, a deeper analysis is performed in this paper.

The 40 "Reserved 1" bits are those located between bit 19 and bit 58, both included. This means that symbols 37 to symbol 128, both included, are generated from "Reserved 1" bits. Each new unpredictable bit will lead to two new symbols. If the symbols were not interleaved, for each bit, the first one (from G1) would be unpredictable, and the second one (from G2) would not. The 240-bit stream would look as in Fig.3, including predictable and unpredictable symbols, before the interleaver.

According to the currently proposed NMA definition used as a baseline for this paper [4], the first 8 bits of the "Reserved 1" field are predictable, as they provide the signed root key of the TESLA chain. That means that the first unpredictable symbol is the 53th one, as shown in Fig.3.

After the interleaver, the sequential dependence between unpredictable symbols and bits is broken, so it is a priori not obvious which symbols based on initially unpredictable bits, are indeed unpredictable, and which not. However, a user may want to perform a test statistic based only the reception of unpredictable symbols to detect replay attacks. In this case, it must know a priori which symbols are unpredictable for the attacker. With this purpose, the positions of the post-interleaving unpredictable symbols in the message have been determined by defining a system of linear equations as follows:

- The unpredictable bits are considered as unknowns.
- Every time a new symbol is received, which depends on one or several unpredictable bits, a new linear equation is added to the system. This equation is based on the encoding polynomials from Fig.1 and adds the new unpredictable bits as new unknowns.
- When the number of equations equals the number of unknowns, i.e. when the number of received symbols based on unpredictable bits equals the number of unpredictable bits encoded by those symbols, the system can be solved.

The results of the algorithm are presented in Table I, for when the last 32 of the 40 "Reserved 1" bits are considered unpredictable. The first column is the index of the symbols as sent, i.e. after interleaving. The second column shows the symbol position before interleaving. The third column presents the number of equations of the system following reception of the symbol identified in the row. The fourth column presents the total number of unknowns, i.e. unpredictable bits on which the received symbols depend. The final column presents the number of possibilities that solve the underdetermined system of equations. Each unknown can take two values, and each new symbol received reduces the possible options to half, so the number of options is 2^{u-e}, where u is the number of unknowns and e is the number of equations, that is, u–e is the number of degrees of freedom of the system. This parameter is useful to assess the risk that a spoofer guesses the unpredictable symbols. We see that, for most of the unpredictable symbols, it is very high. Therefore, it is not considered worth to discard any unpredictable symbol from the test statistic.

The results of the analysis showing the positions of unpredictable symbols are presented in Fig.4. We can see that, after estimating the symbol in position 102, which corresponds to symbol in position 92 before interleaving, the system of equations can be solved, and the remaining symbols are predictable. Therefore, most of the unpredictable symbols are concentrated between symbol 57 and symbol 92, during a transmission time of 380 ms of every 2-second page. Note that the CRC of the page may use the unpredictable bits too and therefore be partly unpredictable but it is not considered here. Adding unpredictable symbols in each I/NAV word therefore guarantees signal unpredictability every 2 seconds, potentially constraining attacks in comparison with an NMA message structure where the signal is fully predictable for long time intervals.

We can therefore conclude that convolutional encoding and interleaving maintains the entropy and unpredictability of the signal in a way that can be determined a priori by the receiver, and depends on the message structure. The relevance of this aspect is that future NMA-enabled receivers tracking authentic signals, which can protect against spoofing attacks, can look at the samples of these particular symbols, and not others, in order to determine if they are spoofed. This analysis also

Symbol Post-Intl Position	Symbol Pre-Intl Position	Nb. Of Equations	Nb. Of Unknowns	Nb. Possibilities
8	57	1	3	4
9	65	2	7	32
10	73	3	11	256
11	81	4	15	2048
12	89	5	19	16384
13	97	6	23	131072
14	105	7	27	1048576
15	113	8	31	8388608
16	121	9	32	8388608
38	58	10	32	4194304
39	66	11	32	2097152
40	74	12	32	1048576
41	82	13	32	524288
42	90	14	32	262144
43	98	15	32	131072
44	106	16	32	65536
45	114	17	32	32768
46	122	18	32	16384
68	59	19	32	8192
69	67	20	32	4096
70	75	21	32	2048
71	83	22	32	1024
72	91	23	32	512
73	99	24	32	256
74	107	25	32	128
75	115	26	32	64
76	123	27	32	32
98	60	28	32	16
99	68	29	32	8
100	76	30	32	4
101	84	31	32	2
102	92	32	32	1

TABLE I
RESULTS OF UNPREDICTABILITY ANALYSIS AFTER ENCODING AND INTERLEAVING

confirms that, out of n unpredictable bits coded into $2n$ symbols, the first n symbols based on unpredictable bits after interleaving can be considered unpredictable.

III. SYMBOL UNPREDICTABILITY AND ANTI-REPLAY PROTECTION

This section presents an example of how Galileo NMA symbol unpredictability can protect against signal replay attacks. It proposes and preliminarily characterizes a method based on the accumulation of the first signal samples of each unpredictable symbol, and the correlation of that synthetic sequence with the known replica once the symbols are authenticated. Notice that a full characterization of security code estimation and replay (SCER) attack detectors using different test statistics is presented in [4] and improved in [5], but it is beyond the scope of this paper.

Before starting with the description of the method, we must note that a first protection method against non-sophisticated replay attacks where the attacker is not able to transmit the correct unpredictable symbols on time (e.g. because the estimation and replay method has a delay of several milliseconds), would be to look at the unpredictable symbols demodulated by the receiver before the Viterbi decoding.

The replay attack analyzed here consists of a zero-delay attack, as described in [4]. In this attack, the spoofer estimates and rebroadcasts the original signal with a zero or negligible delay, in order to initially take control of the tracking loops, and then it starts to gradually delay the signal, in order to spoof the pseudorange measurements and hence the position. Before the attack, we assume that the receiver is locked to the authentic signals.

A non-zero delay attack requires the attacker to force signal reacquisition by jamming the receiver for some seconds (e.g. 20 seconds for a two-microsecond delay with a static receiver using a 0.1-ppm Temperature-Controlled Cristal Oscillator, as per [4]). The receiver may have other means to detect this attack so it is left out of the scope of this paper.

In our zero-delay attack, we will assume that the attacker is continuously transmitting a signal with a zero delay, or a delay that is so close to zero that it does not represent any difference in the symbol detection process.

In order to detect if the signal is correct, the receiver stores the first samples of the N_C chips of the spreading code for each of the N_U unpredictable symbols over a given interval. The receiver will accumulate the samples of $N_C N_U$ chips in total. Once the N_U symbols are received and authenticated, the receiver can generate a replica of the samples corresponding to the $N_C N_U$ chips and correlate it with the signal samples. The correlation gain obtained by a receiver using the authentic signal can be approximated by the total number of chips correlated [7]:

$$G_r = N_C N_U \tag{1}$$

The previous section suggests that if the 32 symbols of every page are unpredictable, a receiver would gather 480 unpredictable symbols in one 30-second I/NAV subframe. However, this would only occur if all MAC and key bits of the TESLA protocol are unpredictable, which would only happen if all information is encrypted with the TESLA key later disclosed [9]. In addition, the last bits of a key may be predictable by an attacker through a brute force attack [6]. For all this reasons, we will assume in the rest of the analysis that the last 20 bits of the key are predictable, and only 10-bit MAC tags are unpredictable too. Based on these assumptions, $N_U = 276$ for a 30-second period. Note also that a receiver may follow a strategy whereby only the unpredictable key bits are used, and not the MACs. This would allow that, by retrieving the key from any satellite, one can compute the key unpredictable symbols from the keys transmitted by all other satellites, as they belong to the same chain. In any case, if N_U = 276 symbols and we use the first 5 chips of each unpredictable symbol ($N_C = 5$), we obtain a gain of $G_r = 1380$, or 31.39 dB.

One potential disadvantage of this method is that the correlation sequence loses the cross-correlation properties of PRN codes, which minimize the interference between satellites. One can expect a higher level of interference due to other satellites for both authentic and spoofed signals. In any case, for the rest of this section, we assume that the sequence is long enough to have a significant gain above the noise plus interference level. We also assume that a processing gain in the order of 30 dB will be sufficient, noting that the pre-correlation SNR of a typical receiver can be in the order of -20 dB, as shown in [8], Ch. 6. This assumption depends on the number of unpredictable symbols used, but assuming is high enough, the conclusions are still valid.

If the receiver is tracking a spoofed signal subject to a zero-delay attack, the gain will be lower and will depend on the probability for the spoofer of successfully estimating the symbols with a reduced number of samples in each symbol. For every set of m samples, we can calculate the probability of error in the estimation of an unpredictable symbol by the spoofer as follows:

$$p_{err}(m) = \frac{1}{2} erfc(\sqrt{mT_s C/(N_0 + I_0)}) \quad (2)$$

where m is the number of samples integrated by the spoofer, T_s is the sampling period and $C/(N_0 + I_0)$ is the carrier to noise plus interference density ratio as seen by the spoofer. Note that mT_s is the duration of the time interval used by the spoofer to estimate the symbol. Fig.5 shows the error probability in the estimation of a symbol as a function of mT_s, for four different spoofers, each one receiving the signal at different of 42 dBHz, 45 dBHz, 48 dBHz and 52 dBHz. As mentioned before, the I_0 term a priori cannot be neglected for a few samples. Due to this term, the equivalent carrier-to-noise-ration values can be lower than the standard C/N_0 of around 45 dBHz for GNSS signals tracked with standard receivers and antennas. However, as an advanced spoofer may incorporate additional hardware to raise the signal power, as directional antennas, the previous effect may be compensated and higher values need also to be considered. Notice that the term does not affect the results as we analyze in relative terms the gain difference between the spoofed and non-spoofed cases, which are both affected by the term in the same way.

Notice that $p_{err}(m)$ is the error probability after mT_s integration, for each m. If a spoofer integrates M samples, its average probability of error is higher, and will be between 0.5 and p_{err}:

$$p_{err,avg} = \frac{1}{M} \sum_{m=1}^{M} p_{err}(m) \quad (3)$$

where M is the total number of samples integrated per unpredictable symbol. Assuming that the receiver performs the abovementioned correlation over a spoofed signal, the power gain when tracking a spoofed signal would be:

$$G_s = N_C N_U (1 - 2p_{err,avg})^2 = G_a (1 - 2p_{err,avg})^2 \quad (4)$$

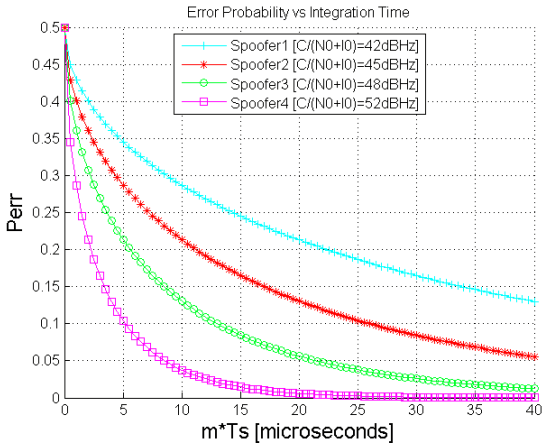

Fig. 5. Probability of error in the estimation of an unpredictable symbol wrt. Integration time ($m * Ts$), for different $C/(N_0 + I_0)$ values

The factor 2 multiplying $p_{err,avg}$ is due to the fact that, when the spoofer takes a wrong decision the symbols contribution is subtracted, instead of added, to the correlation. To compensate the fact that an intelligent spoofer will try to minimize this effect, we will approximate $p_{err,avg}$ by $p_{err}(M)$ in the rest of the analysis, which is a conservative assumption. Under this assumption, Fig.6 shows the ratio G_s and G_r for different integration times, for spoofers at a $C/(N_0 + I_0)$ of 42, 45, 48 and 52 dBHz.

The gain reduction is quite evident especially if a short interval at the beginning of the unpredictable symbols is used. Following the above example, if we look at the first 5 chips of each unpredictable symbol, and the spoofer received the signal with 45 dBHz, the gain would be reduced by more than 7 dB with respect to tracking an authentic signal.

The most obvious way to derive a test statistic would be to compare the gain based on an unpredictable sequence, and that based on a predictable sequence, i.e. a sequence based on samples from predictable symbols. A sustained difference between the two may indicate a replay attack.

A spoofer could use the SCER attack approach for both predictable and unpredictable symbols, in which case the receiver would not observe a difference. It could also interfere with the signal leading to low C/N_0 values. In these cases, the receiver may not be able to state if a replay attack is being performed, or the signal is just degraded due to other reasons. However, the level of protection obtained by this method can still be useful: if it is tuned to report a low probability of missed detection (i.e., the probability that the receiver considers the signal as not replayed when it actually is), a signal passing the threshold may be considered as trustable, even if it happens only in good visibility conditions, and with low or moderate availability. Notice that signal antireplay methods can be used in combination with other authentication receiver checks, as RAIM, INS, trusted timing, or antenna arrays.

In order to implement the proposed method, a receiver needs

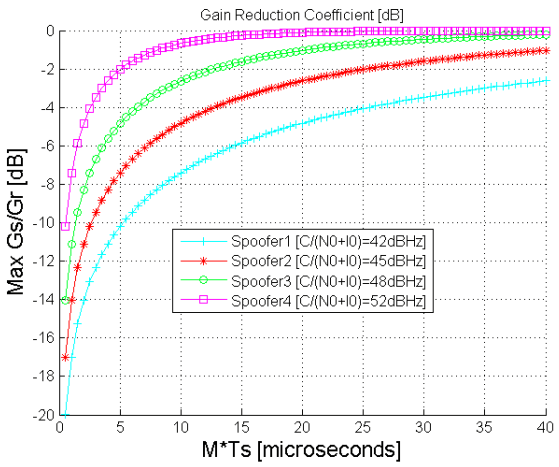

Fig. 6. Gain Reduction when spoofed (Maximum spoofed vs. non-spoofed gain), vs. Integration Time ($M * T_s$)

two inputs:

- The validation of the unpredictable symbols. This occurs every authentication event, that is, every 10 or 15 seconds for each satellite [9].
- A large-enough amount of chips from unpredictable symbols to generate a peak above the noise and interference. This depends on N_C and N_U.

As mentioned above, the receiver can obtain a processing gain of around 30 dB from 30 seconds of signal using 5 microseconds of each unpredictable symbol. If a higher gain is desired, the receiver can process more chips per symbol or take more unpredictable symbols. Increasing N_C improves the success rate of the spoofer. Increasing N_U implies waiting longer before computing the statistic, but computing it as frequently as possible makes the system more robust. While sensitivity analyses and trade-offs on this respect are left for further work beyond this section, here are some ideas for improvements and further study:

- N_U could be augmented by encrypting some predictable bits with the delayed TESLA keys.
- Even if the receiver waits for 30 seconds to get enough symbols, it could perform the check every 10-15 seconds (current Time Between Authentications as per [9]), though a 30-second sliding window. That is, the duration of the correlation interval and the period with which the check is performed need not coincide.

While more detailed analyses and metrics can be employed, including hypothesis testing versus missed detection and false alarm probabilities as those proposed in [4], [10] and [11], we can conclude that symbol unpredictability seems to be a relevant performance feature, enabled by NMA, against replay attacks, and unpredictability should be maximized, provided it does not degrade the overall performance.

IV. Conclusion

This paper has characterized Galileo NMA symbol unpredictability and the protection it can offer against replay attacks. As currently proposed, Galileo NMA in the "Reserved 1" field can add unpredictable symbols to the Galileo E1-B I/NAV message. The effect of interleaving on the position of the unpredictable symbols has been analyzed, resulting that the longest interval with only predictable symbols is 1620 milliseconds.

A replay protection method is proposed, whereby a receiver can store the first samples of every unpredictable symbol, creating thus a sequence whose correlation gain will be lower if the signal tracked starts to be replayed by a spoofer.

Areas of further work include the analysis of the hypothesis testing performance for the proposed method, including the missed detection and false alarm probabilities for different N_U and N_C values and under different spoofing conditions, and optimization strategies to maximize the difference in processing gain between authentic and replayed signals.

We can conclude that adding NMA unpredictability to the signals restricts the possibilities of a spoofer to perform replay attacks under certain conditions, and it can be a relevant building block for trusted PVT services.

Acknowledgment

The authors would like to thank the AALECS team and in particular Dr. J. Winkel from IFEN, Prof. K. Borre, and Prof. T. Larsen, from Aalborg University.

Disclaimer

The information and views set out in this paper do not necessarily represent any official view of the European Union.

References

[1] The European Union, *European GNSS (Galileo) Open Service Signal In Space Interface Control Document*, 2015.
[2] Andrea Goldsmith, *Wireless communications*, Cambridge university press, 2005.
[3] Andrew J Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *Information Theory, IEEE Transactions on*, vol. 13, no. 2, pp. 260–269, 1967.
[4] Todd E Humphreys, "Detection strategy for cryptographic gnss anti-spoofing," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 49, no. 2, pp. 1073–1090, 2013.
[5] Gianluca Caparra, Nicola Laurenti, Rigas T Ioannides, and Massimo Crisci, "Improving secure code estimate-replay attacks and their detection on gnss signals"," *Proceedings of NAVITEC 2014*, 2014.
[6] Ignacio Fernández-Hernández, *Snapshot and authentication techniques for satellite navigation*, Ph.D. thesis, Aalborg University, Faculty of Engineering and Science, jun 2015.
[7] Andrew J Viterbi, *CDMA: principles of spread spectrum communication*, Addison Wesley Longman Publishing Co., Inc., 1995.
[8] Frank van Diggelen, *A-GPS: Assisted GPS, GNSS, and SBAS*, Artech House, 2009.
[9] Ignacio Fernández-Hernández, Vincent Rijmen, Gonzalo Seco-Granados, Javier Simon, Irma Rodríguez, and J David Calle, "A navigation message authentication proposal for the galileo open service," *Navigation*, vol. 63, no. 1, pp. 85–102, 2016.
[10] Andrew J Kerns, Daniel P Shepard, Jahshan A Bhatti, and Todd E Humphreys, "Unmanned aircraft capture and control via gps spoofing," *Journal of Field Robotics*, vol. 31, no. 4, pp. 617–636, 2014.
[11] Christoph Günther, "A survey of spoofing and counter-measures," *Navigation*, vol. 61, no. 3, pp. 159–177, 2014.