Sveučilište u Rijeci
POMORSKI FAKULTET
FACULTY OF MARITIME STUDIES
University of Rijeka

University of Zagreb
Faculty of Transport
and Traffic Sciences

Royal Institute of Navigation
Science Technology Practice

ISSN 1849-7306 (Print)
ISSN 2670-8981 (Online)

13th Annual Baška GNSS Conference

PROCEEDINGS

Under the High Auspicies of The European Academy of Sciences and Arts

Baška, Krk Island, Croatia

12 – 15 May 2019

Sveučilište u Rijeci
POMORSKI FAKULTET
FACULTY OF MARITIME STUDIES
University of Rijeka

University of Zagreb
Faculty of Transport
and Traffic Sciences

Royal Institute of Navigation
Science Technology Practice

13th Annual Baška GNSS Conference

PROCEEDINGS

Baška, Krk Island, Croatia

12 – 15 May 2019

13th Annual Baška GNSS Conference – PROCEEDINGS

ISSN 1849-7306 (Print) | ISSN 2670-8981 (Online)

Published by:
University of Rijeka, Faculty of Maritime Studies, Rijeka, Croatia &
The Royal Institute of Navigation, London, UK

For the Publisher:
Prof Alen Jugović, PhD, Dean,
University of Rijeka, Faculty of Maritime Studies, Rijeka, Croatia

Publishing Associates:
Terry Moore, PhD FRIN FION, President,
The Royal Institute of Navigation, London, UK

John Pottle, FRIN, Director,
The Royal Institute of Navigation, London, UK

Editors:
Assist Prof David Brčić, PhD,
University of Rijeka, Faculty of Maritime Studies, Rijeka, Croatia

Assist Prof Marko Valčić, PhD,
University of Rijeka, Faculty of Engineering, Rijeka, Croatia

Academician Serdjo Kos, PhD, FRIN,
University of Rijeka, Faculty of Maritime Studies, Rijeka, Croatia

Front-page photo credits:
David Brčić: *Dance*

Text design:
Tempora, Rijeka

Print:
AKD d.o.o. Zagreb

Publisher's address:
University of Rijeka, Faculty of Maritime Studies
Studentska 2, 51000 Rijeka, Croatia
Phone: +385 (0)51 338 411
Fax: +385 (0)51 336 755
URL: http://www.pfri.uniri.hr/
E-mail: dekanat@pfri.hr

Contents

13th Annual Baška GNSS Conference – International Programme and Organising Committee

John R Pottle, Director, The Royal Institute of Navigation, London, UK, Conference Chair

Jasna Prpić-Oršić, University of Rijeka, Faculty of Engineering, Croatia, Conference Co-chair

Serdjo Kos, FRIN, University of Rijeka, Faculty of Maritime Studies, Croatia, Conference Co-chair

Tomislav Kos, FRIN, University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia, Conference Co-chair

Hrvoje Gold, University of Zagreb, Faculty of Transportation and Traffic Sciences, Croatia,

Conference Co-chair

Mia Filić, MRIN, independent satellite navigation, statistical learning and space weather specialist, Sesvete, Zagreb, Croatia, Conference Programme Chair

Terry Moore, FRIN, President, The Royal Institute of Navigation, London, UK

Jeffrey M Auerbach, US Department of State, Washington, DC, USA

David Brčić, University of Rijeka, Faculty of Maritime Studies, Rijeka, Croatia

Ljiljana R Cander, Rutherford Appleton Laboratory, Chilton, UK

Olja Čokorilo, Faculty of Transport and Traffic Engineering, University of Belgrade, Belgrade, Serbia

Jordi Corbera, Institut Cartografic i Geologic de Catalunya, Barcelona, Spain

Giovanni E Corazza, University of Bologna, Bologna, Italy

Robert Crane, National Coordination Office for Space-Based PNT, Washington, DC, USA

Franc Dimc, Faculty of Maritime Studies and Transport, University of Ljubljana, Portorož, Slovenia

Sharafat Gadimova, UN Office of Outer Space Affairs, Vienna, Austria

Alan Grant, FRIN, The GLA of the United Kingdom and Ireland, London, UK

Luka Grubišić, Dept for Matematics, Faculty of Science, University of Zagreb, Zagreb, Croatia

Josip Josipović, CROCONTROL, Zagreb, Croatia

Oliver Jukić, College for Management in Tourism and Informatics, Virovitica, Croatia

Ines Kolanović, University of Rijeka, Faculty of Maritime Studies, Rijeka, Croatia

Siniša Krajnović, LM Ericsson, Stockholm, Sweden

Bal Krishna, Coordinates Journal, Delhi, India

Marta Krywanis-Brzostowska, European GNSS Agency, Prague, Czech Republic

David Last, FRIN, Consultant Engineer and Expert Witness, Conwy, UK

Kristijan Lenac, MRIN, University of Rijeka, Faculty of Engineering, Rijeka, Croatia

Igor Larin, ISS Reshetnev Corporation, Moscow, Russia

Anastasia Lyubimova, GLONASS/GNSS Forum Association, Moscow, Russia

Roger A McKinley, Consultant, Leatherhad, UK

Krešimir Meštrović, Zagreb University of Applied Sciences, Zagreb, Croatia

Tomislav Josip Mlinarić, University of Zagreb, Faculty of Transport and Traffic Sciences, Zagreb, Croatia

Washington Y Ochieng, FRIN, Imperial College, London, UK

Sergey Revivnykh, ISS Reshetnev Corporation, Moscow/Zheleznogorsk, Russia

Enik Shytermeja, Ecole Nationale de l'Aviation Civile, Toulouse, France

Grigory Stupak, Russian Space Systems, Moscow, Russia

Marko Ševrović, FRIN, University of Zagreb, Faculty of Transport and Traffic Sciences, Zagreb, Croatia

Darko Špoljar, PhD student, University of Rijeka, Faculty of Engineering, Rijeka, Croatia

James Taylor, FRIN, The Royal Institute of Navigation, London, UK

Marko Valčić, University of Rijeka, Faculty of Engineering, Rijeka, Croatia

Josip Vuković, University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia

Adam Weintrit, FRIN, Faculty of Navigation, Gdynia, Poland

Jingnong Weng, Beihang University, Beijing, China

List of reviewers

Andrej Androjna (*Faculty of Maritime Studies and Transport, University of Ljubljana, Portorož, Slovenia*)

Mate Barić (*University of Zadar, Nautical Department, Zadar, Croatia*)

David Brčić (*University of Rijeka, Faculty of Maritime Studies, Rijeka, Croatia*)

Jasmin Ćelić (*University of Rijeka, Faculty of Maritime Studies, Rijeka, Croatia*)

Franc Dimc (*Faculty of Maritime Studies and Transport, University of Ljubljana, Portorož, Slovenia*)

Renato Filjar (*Zagreb University of Applied Sciences, Zagreb, Croatia*)

Hrvoje Gold (*University of Zagreb, Faculty of Transportation and Traffic Sciences, Zagreb, Croatia*)

Serđo Kos (*University of Rijeka, Faculty of Maritime Studies, Rijeka, Croatia*)

Tomislav Kos (*University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia*)

Kristijan Lenac (*University of Rijeka, Faculty of Engineering, Rijeka, Croatia*)

Terry Moore (*University of Nottingham, The Notthingham Geospatial Institute, Nottingham, United Kingdom*)

Boris Sviličić (*University of Rijeka, Faculty of Maritime Studies, Rijeka, Croatia*)

Marko Valčić (*University of Rijeka, Faculty of Engineering, Rijeka, Croatia*)

Josip Vuković (*University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia*)

Falin Wu (*Beihang University, School of Instrumentation and Optoelectronic Engineering, Beijing, China*)

Srđan Žuškin (*University of Rijeka, Faculty of Maritime Studies, Rijeka, Croatia*)

We dedicate this book to Professor David Last.

Professor, it was an honour to know You. Thank You.

Editors

Sveučilište u Rijeci
POMORSKI FAKULTET
FACULTY OF MARITIME STUDIES
University of Rijeka

University of Zagreb
Faculty of Transport
and Traffic Sciences

Royal Institute of Navigation
Science Technology Practice

13th Annual Baška GNSS Conference

PROCEEDINGS

Sveučilište u Rijeci
POMORSKI FAKULTET
FACULTY OF MARITIME STUDIES
University of Rijeka

University of Zagreb
Faculty of Transport
and Traffic Sciences

Royal Institute of Navigation
Science Technology Practice

13th
Annual
Baška GNSS
Conference

THE RESULTS OF GNSS TIMING PARAMETERS MONITORING

Denis Aganov, Petr Bogdanov, Andrei Druzhin, Olga Nechaeva, Tatiana Primakina

Russian Institute of Radionavigation and Time, Saint-Petersburg, Russia
e-mail: bogdanov_pp@rirt.ru*)*

ABSTRACT

The monitoring of timing parameters of Global Navigation Satellite Systems (GNSS) is important as the parameters influence the accuracy of positioning and timing as well as the accuracy of calculating and predicting GNSS-GNSS Time Offsets for GNSS interoperability. As GNSS time scales are produced at GNSS control centers they are inaccessible "from outside" but the need for their estimation exists. The authors produced the estimates of GNSS timing parameters based on available indirect data: broadcast corrections to convert from GNSS Time to Reference Time and the values of GNSS Time - Reference Time offsets based on measurements at Reference Time Generating Facility.

The monitoring results obtained at Russian Institute of Radionavigation and Time by independent means on the basis of available data showed that the values of GNSS time-Reference Time offsets for different GNSS are well within the limitations specified by GNSS providers in their Interface Control Documents and time scale templates.

Key words: *Global Navigation Satellite System, time scale, time offset, correction*

1 INTRODUCTION

To provide users with high-accuracy determination of position, velocity and time, in all GNSS time scales of space vehicles (SV) are referenced to GNSS Time which is synchronized to a local laboratory k realization of Coordinated Universal Time, namely UTC(k). Corrections to convert from SV time to GNSS Time and from GNSS Time to Reference Time are broadcast in the navigation messages of all GNSS. Besides, to provide time interoperability with other GNSS, corrections for GNSS-GNSS time offset (GGTO corrections) are broadcast or specified to be broadcast in the navigation messages.

The monitoring of GNSS Time accuracy parameters is important as they influence the accuracy of positioning and timing as well as the accuracy of calculating and predicting GNSS-GNSS Time Offsets for GNSS interoperability.

The paper presents the main GNSS timing parameters and particularities of their monitoring. The results were obtained at Russian Institute of Radionavigation and Time on the basis of publicly open information provided by international monitoring centers.

2 GNSS TIMING PARAMETERS

2.1 GPS Time. GPS Time is a composite time scale based on clock ensemble of monitor station frequency standards and satellite clocks (UNOOSA, 2012). GPS Time is computed as part of the overall clock and orbit estimation process. GPS time is referenced to UTC(USNO) produced by the U.S. Naval Observatory. GPS Time- UTC(USNO) offset shall be within 1 μs. GPS Time is not corrected by ±1 s with UTC leap second corrections and, as a result, there is whole second GPS Time–UTC (USNO) offset and since January 2017 GPS Time has been 18 s ahead of UTC. The error of corrections to convert from SV time to GPS Time is a part of User Range Error (URE), which is specified to be within 6 m. The specified accuracy of the corrections to convert from system to Reference Time is 40 ns (95 % probability). Broadcasting corrections for GPS Time offsets relative to other GNSS (Galileo, GLONASS and others) is declared without specifying their accuracy.

2.2 GLONASS Time. GLONASS Time is a mathematical time scale produced based on GLONASS Central Synchronizers (CS) (RISDE, 2008). There are two CSs: The Main CS and Reserved one. GLONASS Time is produced based on the

Main Central Synchronizer's time scale. GLONASS Reference Time is the national time scale of Russia UTC(SU) that is generated by State Time/Frequency Reference (STFR). GLONASS Time is maintained within 1 ms of UTC(SU). GLONASS Time is corrected by ±1 s simultaneously with UTC corrections and, as a result, there is no whole second time offset between GLONASS Time and UTC(SU). However, there is a three-hour constant offset between GLONASS Time and UTC due to GLONASS Ground Control Segment operational principles. The specified accuracy of the corrections to convert from SV time to GLONASS Time is 5.6 ns (rms), and corrections to convert from GLONASS Time to Reference Time – 1 μs. The specified accuracy of broadcast GLONASS-GPS time offset corrections is 30 ns.

2.3 Galileo Time. Galileo Time is a continuous time scale produced at two Galileo Control Centers and synchronized to UTC based on contributions from European UTC Timing Laboratories (UNOOSA, 2016b). The offset between Galileo Time and UTC (modulo 1 s) shall be less than 50 ns (95 %). Galileo Time is not corrected by ±1 s with UTC leap second corrections and, as a result, since January 2017 Galileo Time has been 18 seconds ahead of UTC. The error of corrections to convert from SV time scales to Galileo Time is a part of URE that is specified to be 65 cm (rms). The specified accuracy of the corrections to convert from Galileo Time to Reference Time is 28 ns (95 %) and corrections for Galileo-GPS Time Offset – 5 ns (95 %).

2.4 BeiDou Time. BeiDou Time (BDT) is a continuous time scale that is produced and maintained by the Master Control Station based on the clocks at Master Control Station and monitor stations (UNOOSA, 2016a). BeiDou Time is linked to UTC through UTC(NTSC) provided by China National Time Service Center. The offset of BDT from Reference Time is specified to be within 100 ns (modulo 1 s). Since January 2006 BeiDou Time is not corrected by ±1 s with UTC leap second corrections and, as a result, since January 2017 BeiDou Time has been 4 s ahead of UTC. The specified accuracy of corrections to convert from SV time to BeiDou Time is 2 ns, from BeiDou Time to UTC(NTSC) – 5 ns (95 %). Broadcasting corrections for BeiDou Time offsets relative to GPS Time, Galileo Time and GLONASS Time is specified without specifying their accuracy.

2.5 QZSS Time. Time is produced similarly to GPS Time as part of the overall clock and orbit estimation process (UNOOSA, 2016c). QZSS Reference Time is UTC(NICT) produced by the National Institute of Information and Communications

Technology. QZSS Time is maintained within 1 μs of UTC(NICT). QZSS Time is not corrected by ±1 s with UTC leap second corrections and, as a result, since January 2017 QZSS Time has been 18 s ahead of UTC. The specified accuracy of corrections to convert from SV time to QZSS Time is 1.6 m (95 %), from QZSS Time to Reference Time – is not specified. The specified accuracy of QZSS Time offset relative to GPS Time is 6.67 ns.

2.6 The Key GNSS Timing Parameters. According to the analysis mentioned above the key timing performance parameters are the following:

- GNSS Time – Reference Time offset;
- the accuracy of corrections to convert from SV time to GNSS Time;
- the accuracy of corrections to convert from GNSS Time to Reference Time;
- the accuracy of GGTO corrections.

The specified values of these parameters are introduced in Table 1.

Table 1. GNSS performance parameters specified by GNSS Providers

Parameter	GPS	GLONASS	Galileo	Beidou	QZSS
GNSS Time-Reference Time offset (mod 1 s)	1 μs	1 ms	50 ns (95 %)	100 ns	1 μs
The accuracy of SV -GNSS Time offset corrections	6 m*	5.6 ns (rms)	65 cm* (rms)	2 ns	1.6 m (95 %)
The accuracy of GNSS Time-Reference Time offset corrections	40 ns (95 %)	1 μs	28 ns (95 %)	5 ns (95 %)	-
The accuracy of GGTO corrections	-	30 ns (rms)	5 ns (95 %)	-	6.67 ns

* User Range Error (URE) that includes the error of SV – GNSS Time offset corrections

As it was mentioned above, GNSS time scales are produced and maintained by GNSS control centers and inaccessible "from outside". Therefore, even to estimate all the GNSS timing parameters based on accessible data to the user is impossible, but the need for their estimated predictions exists.

The trustworthy estimates "from outside" for GNSS Time-Reference Time offsets can be calculated based on the following indirect data:

- broadcast corrections to convert from GNSS Time to Reference Time;
- the values of GNSS Time-Reference Time offsets based on measurements at Reference Time Generating Facility.

In the first case, the accuracy of the estimates depends on the accuracy of broadcast corrections to convert from GNSS Time to Reference Time. In the second case, it depends on the error of the measured SV-GNSS Time offsets of satellite clocks from the Reference Time and the accuracy of broadcast corrections to convert from SV time to GNSS Time.

Since the specified accuracy of broadcast corrections is limited the trustworthy results can be obtained only for the values of GNSS Time offsets of Reference Time.

3 MONITORING RESULTS OF GNSS TIMING PARAMETERS

The offsets dT of GPS Time, GLONASS Time, Galileo Time, Beidou Time and QZSS Time relative to their Reference Time based on broadcast corrections are presented in Figure 1. GPS Time-UTC(USNO), GLONASS Time-UTC(SU) and BeiDou Time–UTC(NTSC) offsets dT based on broadcast corrections and measurement processing at USNO, STFR and NTSC are presented in Figures 2, 3 and 4, respectively.

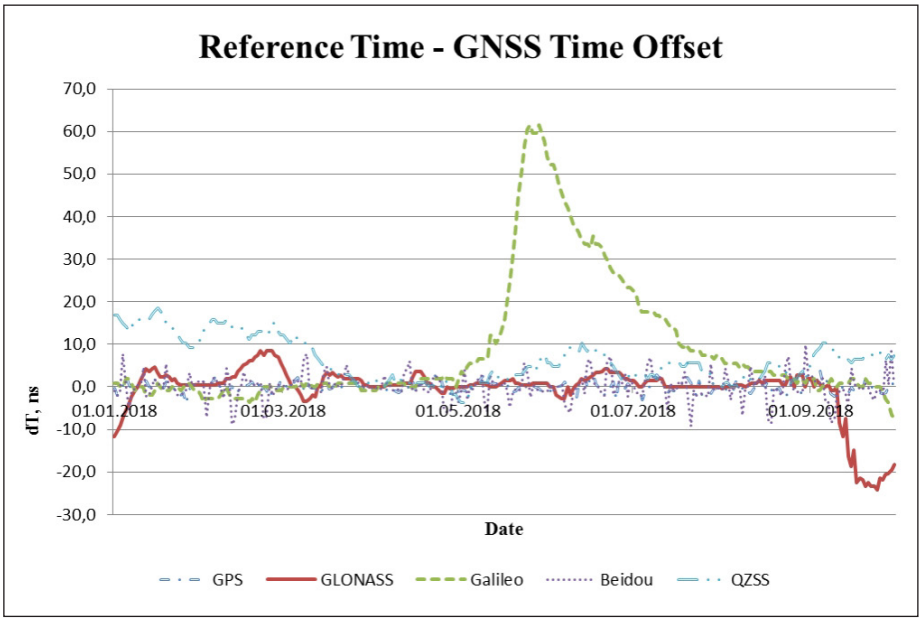


Figure 1. GNSS Time-Reference Time offsets based on broadcast corrections

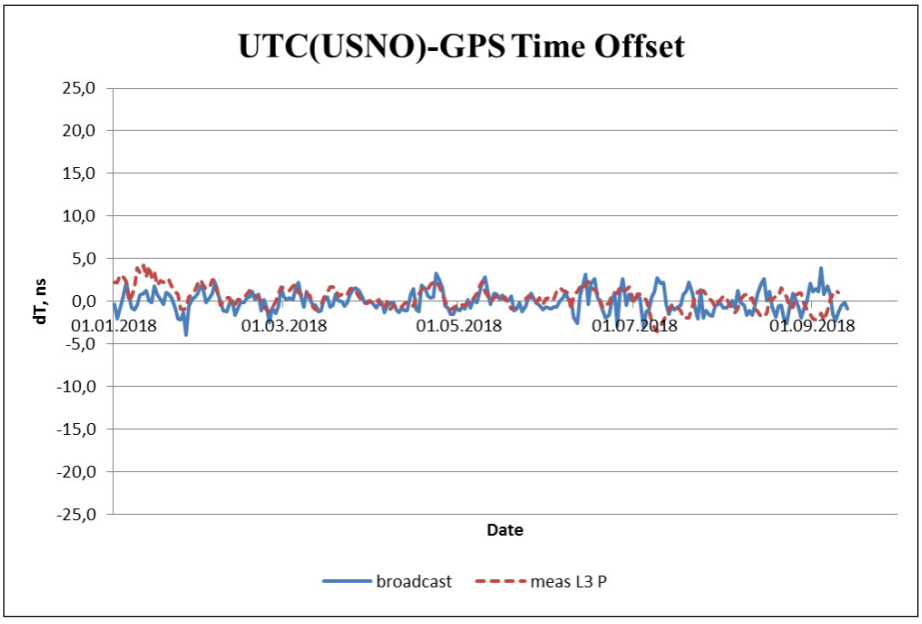


Figure 2. GPS Time-UTC(USNO) offset based on broadcast corrections and measurement processing at USNO

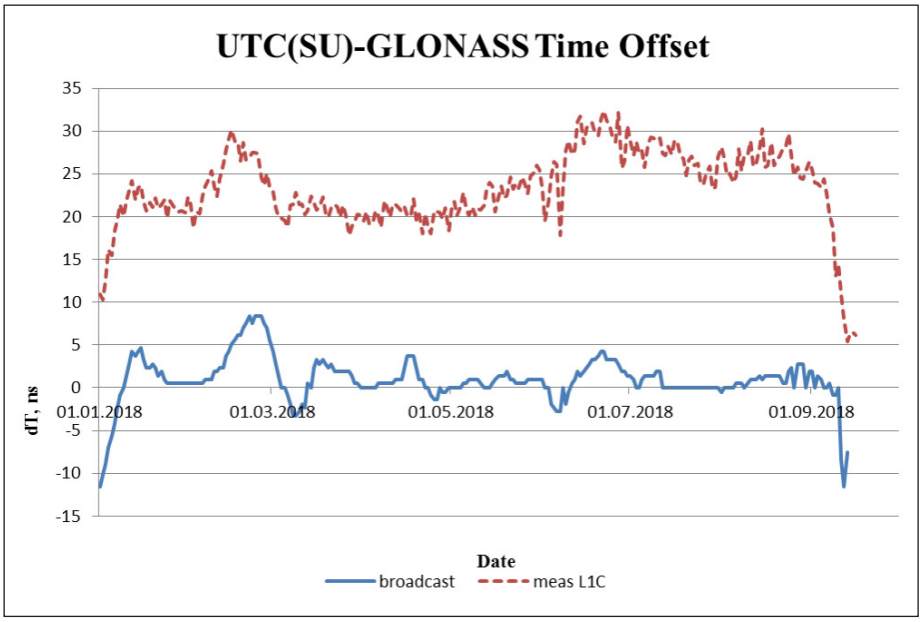


Figure 3. GLONASS Time-UTC(USNO) offset based on broadcast corrections and measurement processing at STFR

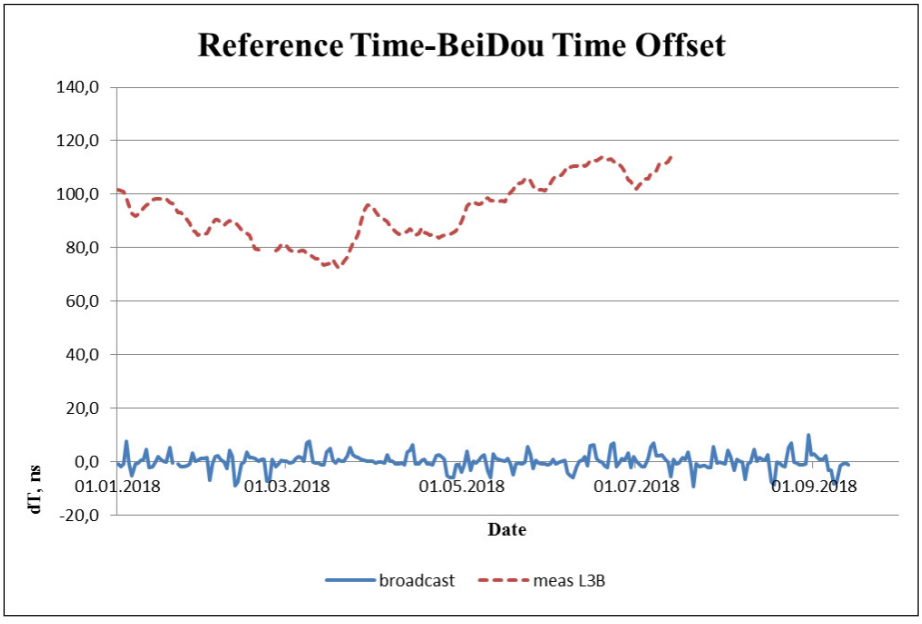


Figure 4. BeiDou Time-UTC(NTSC) offset based on broadcast corrections and measurement processing at NTSC

The analysis of the GNSS Time-Reference Time offsets for the period of January-September 2018 shows the following:

- GPS Time-UTC(USNO) offset is maintained within ±5 ns;
- GLONASS Time-UTC(SU) offset is maintained within ±10 ns, however, there is a systematic error component of broadcast GLONASS Time of about (15-20) ns;
- Galileo Time-UTC offset is mostly maintained within ±5 ns;
- Beidou Time-Reference Time offsets based on broadcast corrections and measurement processing at NTSC differ significantly. Additional analysis of open sources of information showed that broadcast corrections are actually calculated for BeiDou Time offset relative to the time scale produced by Beijing Satellite Navigation Center BSNC. BeiDou Time–BSNC time offset is maintained within ±10 ns and BeiDou Time–UTC(NTSC) offset is within ±120 ns;
- QZSS Time-UTC(NICT) offset is maintained within ±20 ns.

4 CONCLUSIONS

GNSS time scales are either mathematical or physical time scales generated at GNSS Control Centers or Master Stations. As a result, they can be analyzed "from outside" only indirectly by using either broadcast corrections to convert from GNSS Time to Reference Time or the values of GNSS Time–Reference Time offsets calculated by measurement processing at the Reference Time Generating Facility.

The monitoring results obtained at Russian Institute of Radionavigation and Time by independent means based on available data showed that the values of GNSS time-Reference Time offsets for different GNSS are well within the limitations specified by GNSS providers in their Interface Control Documents and time scale templates.

REFERENCES

Russian Institute of Space Device Engineering (RISDE). (2008). *GLONASS Interface Control Document (ICD)*, Edition 5.1. Moscow: RISDE.

United Nations Office for Outer Space Affairs (UNOOSA). (2012). *GNSS Timescale Description: GPS*. Available at: www.unoosa/pdf/icg/2012/Timescale-GPS.pdf, accessed 15 April 2019.

United Nations Office for Outer Space Affairs (UNOOSA). (2016a). *Beidou Timescale Description*. Available at: www.unoosa/pdf/icg/2016/Beidou-Timescale2016.pdf, accessed 15 April 2019.

United Nations Office for Outer Space Affairs (UNOOSA). (2016b). *GNSS Timescale Description: Galileo*. Available at: www.unoosa/pdf/icg/2016/gst2016.pdf, accessed 15 April 2019.

United Nations Office for Outer Space Affairs (UNOOSA). (2016c). *GNSS Timescale Description: QZSS*. Available at: www.unoosa/pdf/icg/2016/QZSS-Timescale2016.pdf, accessed 15 April 2019.

Sveučilište u Rijeci
POMORSKI FAKULTET
FACULTY OF MARITIME STUDIES
University of Rijeka

University of Zagreb
Faculty of Transport
and Traffic Sciences

Royal Institute of Navigation
Science Technology Practice

13th
Annual
Baška GNSS
Conference

METHOD OF ESTIMATING THE CARRIER PHASE GENERATION ERROR BY GNSS SIMULATORS

Svyatoslav Yurievich Burtsev,
Dmitry Stanislavovich Pecheritsa,
Anatoly Aleksandrovich Frolov

Federal Agency on Technical Regulating and Metrology
Russian Metrological Institute of Technical Physics and Radio Engineering
Moscow region, Mendeleevo, Solnechnogorsk, Russia
burtsevsy@vniiftri.ru

ABSTRACT

When calibrating the receiver using a signal simulator, special attention is paid to the quality of the generated navigation signal. For high-precision carrier phase measurements by receiver is necessary to calibrate it using the navigation signals simulator with stable time parameters. The time characteristics of the generated navigation signal can be expressed by the its phase jitter, provided that the navigation signal is generated with a constant frequency. This article presents the method of estimating carrier phase generation simulators error by simulator. The method consists in measuring the phase difference absolute value between the carrier signal and pure reference sinusoidal signal of the same frequency as the carrier. The advantages and disadvantages of the proposed method are noted.

Key words: *simulator, calibration, carrier-phase, phase measurements*

1 INTRODUCTION

Global navigation satellite system (GNSS) simulators is intended for navigation radio frequency signals generation and uses for development, debugging and receiver's calibration. GNSS simulators have different metrological characteristics: dynamic range of power level, power level setting error, frequency error of the internal reference generator etc. Pseudorange generation and pseudorange change rate of most importance to receiver calibration. The pseudorange error may be estimate by code pseudorange and carrier-phase. The carrier-phase generation GNSS simulators error is of particular interest to simulators developers because of receiver calibration results are applied for more accurate coordinates determination by the receiver than the code phase.

At present, the carrier-phase generation GNSS simulators error is estimated by means of either direct calculation or the reference receiver. The first method is strictly theoretical and is available only to manufacturers, as it requires the knowledge of the simulator circuitry. Analytical methods have issues associated with the mathematical model description, the choice of initial conditions and others. The latter using simulators therefore the error value is uncertain.

The phase measurements error may be expressed by sum of bias and random components.

The systematic component of phase measurements difficult to identify because of circuit Phase-Frequency Characteristics (PFC) unevenness from digital-analog conversion output of a simulator to analog-digital conversion output of a phase measuring instrument. A PFC unevenness leads to group-delay time unevenness, which determinate a bias component of the carrier-phase measurement error. A PFC unevenness to one degree or another also possess amplifiers, filters and other passive devices which form the navigation signal carrier (Perov and Harisov, 2010).

The absolute measurement error of the carrier-phase generation simulator devoid of physical meaning and cannot be determined for GNSS simulators in which the carrier frequency is not tied to a time scale.

The article describes theoretical and practical parts of the method of estimating carrier phase generation GNSS simulators error. Theory describes the mathematical justification, and practical implementation of the method shows its limitations in the application. Finally, the main conclusions present about the practical value of the applied method.

2 METHOD DESCRIPTION

The authors propose an experimental method for determining the random error of carrier-phase generation using oscilloscope for measuring navigation signal parameters.

The method essence consists in determine the standard deviation of the phase difference absolute value between the carrier signal and pure reference sinusoidal signal of the same frequency as the carrier. The navigation signal may be represented by in-phase and quadrature components of the carrier (QPSK-modulation):

$$s_{0n} = \ln \cdot \cos(\omega_0 \cdot t_n + \varphi_0) + Q_n \cdot \sin(\omega_0 \cdot t_n + \varphi_0) \quad (1)$$

where $I(t)$ and $Q(t)$ is a real and imaginary parts of the QPSK signal complex envelope.

The pure sinusoidal signal is in the same time scale as navigation signal and it has identical amplitude and zero initial phase, and can be described in complex form:

$$s_{ref_n} = A_{ref} \cdot e^{i \cdot \omega_0 \cdot t_n} \quad (2)$$

where A_{ref} is the reference signal amplitude that can be found by the expression:

$$A_{ref} = \left| \frac{1}{N} \cdot \sum_{n=1}^{N} s_{0n} \right| \cdot \frac{\pi}{2} \quad (3)$$

where N is the number of measurements.

The phase difference between the model and digital navigation signal is calculated by the following equations (Voronov, 2007):

$$\Delta\varphi_{ref-0} = \frac{2}{A_{ref}^2} \left(\frac{1}{N} \cdot \sum_{n=1}^{N} \left(s_{ref_n} \cdot s_{0n} \right) \right) \quad (4)$$

The method accuracy increases with increasing ratio of the sample length to the signal period.

The found phase difference is a QPSK signal complex envelope parameter and takes four values.

In short, the phase difference after deduction phase manipulations can be represented by the expression:

$$\Delta\varphi_{ref-0} = \varphi_0 + n \cdot {}^{\pi}/_{2}, \text{ where } n \in Z \tag{5}$$

Having N measurements, the standard deviation of the carrier-phase generation GNSS simulators instrumental error can be calculated with the next formula:

$$\sigma_{\varphi} = \sqrt{\frac{1}{N} \cdot \sum_{n=1}^{N} \left(\varphi_{0_n} - \bar{\varphi}_0\right)^2} \tag{6}$$

where $\bar{\varphi}_0$ is mean initial phase.

3 METHOD VALIDATION 1

Validation purpose is to confirm the characteristics declared by the manufacturer.

The experiment was carried out using next measuring means:

- Oscilloscope LeCroy WaveMaster 820-Zi with vertical resolution 8 bits and 40 GS/s sample rate (sampling interval 25 ps), jitter between channels 250 fsrms;
- GNSS simulator Spirent GSS7000, phase noise (max) 0.02 rad RMS;
- Frequency rubidium standard FS 725, accuracy $\pm 5 \times 10^{-11}$.

The oscilloscope gets the samples on the positive edge of the simulator output 1PPS. Jitter can reach 100 ps, which makes phase measurements very noisy. The simulator and the oscilloscope are synchronized by the frequency standard for reducing the phase measurements noise. The measurement time scale is tied to the reference signal simulator output.

Method validation was performed using the following connections shown in Figure 1.

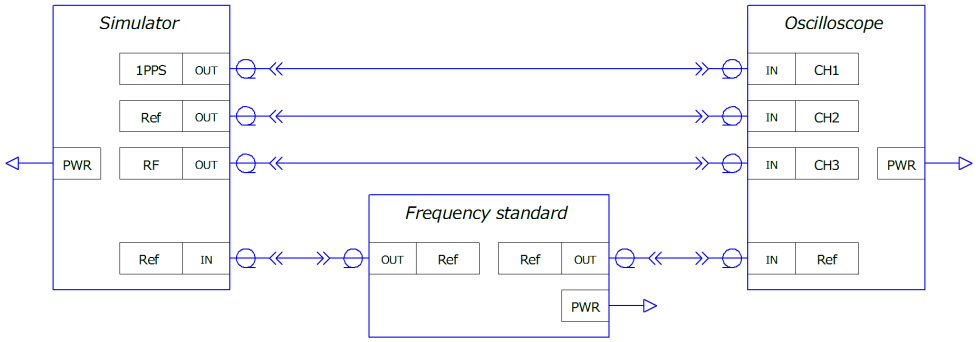


Figure 1. Connection for measurements of the phase difference

In practical implementation, the proposed method has the following limitations:

- the navigation signal generation should be from one type of system, in one frequency band and from one satellite;
- the navigation signal generation should be constant frequency and power level for oscilloscope vertical resolution;
- sensitivity of measuring instruments to environmental parameters, primarily to temperature.

Figure 2 shows the navigation signal phase modulation with the pure sinusoidal signal.

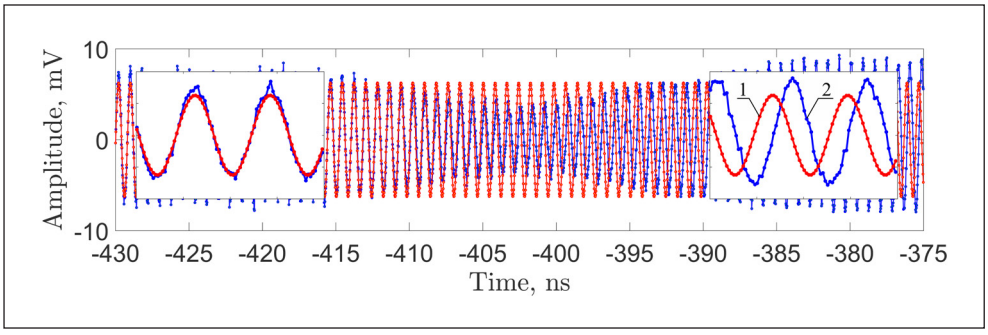


Figure 2. The navigation signal phase modulation compared to the pure sinusoidal signal (1 – pure sinusoidal signal, 2 – navigation signal)

As a next step, we get the phase difference between the model and navigation signal according to equation (4), as shown in Figure 3.

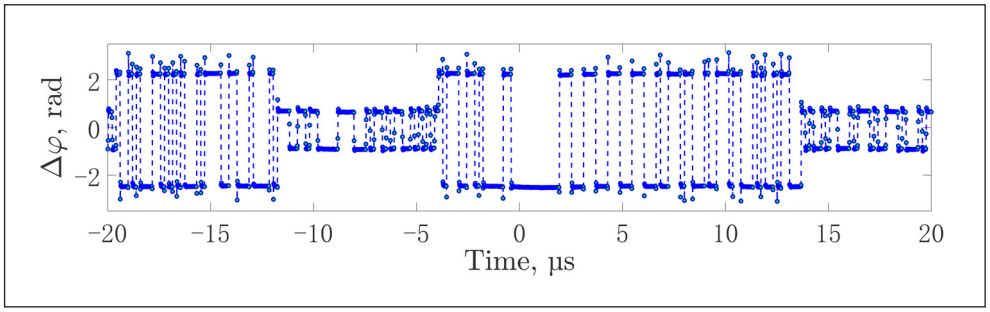


Figure 3. Phase difference

For sections selected with a constant phase difference, the navigation signal must be demodulated. Decoding of QPSK signals allows to obtain demodulated orthogonal signal components containing Pseudo-Random Sequences (PRS) with a specific clock frequency and period duration. The frequency of carrier phase shift keying is determined by the frequency of the orthogonal component PRS chips (Hofmann-Wellenhof, Lichtenegger and Wasle, 2008).

Figure 4 shows accurate determine the phase difference on sections without phase modulation, which can be represented by the expression (3).

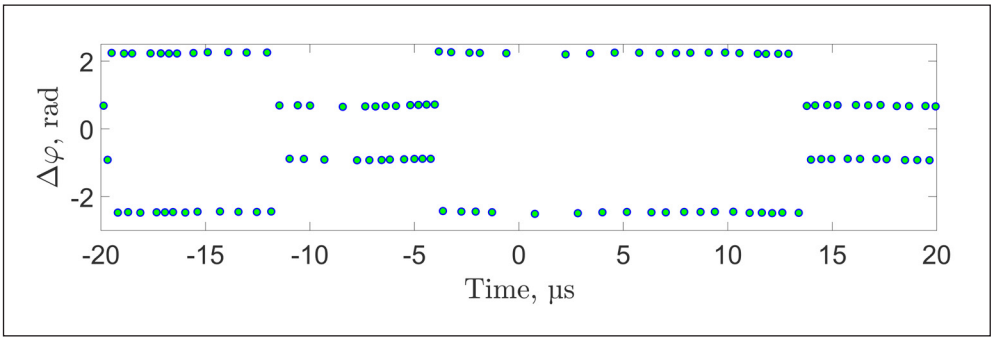


Figure 4. Accurate phase difference

Measurements have conducted on the GLONASS and GPS systems (RISDE, 2008; GPSD, 2018). Results are presented in Table 1.

Table 1. Standard deviation of the phase difference measurements of GLONASS and GPS, rad

GLO			GPS		
Freq Slot	**L1**	**L2**	**PRN**	**L1**	**2C**
-7	0.015	0.014	10	0.013	0.011
-6	0.016	0.012	20	0.013	0.010
-5	0.021	0.016	30	0.012	0.011
-4	0.014	0.011			
-3	0.022	0.012			
-2	0.009	0.012			
-1	0.014	0.015			
0	0.013	0.008			
1	0.017	0.009			
2	0.016	0.013			
3	0.015	0.013			
4	0.014	0.011			
5	0.015	0.012			
6	0.015	0.010			

Results have confirmed the technical parameters stated by the manufacturer.

4 METHOD VALIDATION 2

Validation purpose is to confirm the phase offset set by the manufacturer. The second experiment was carried out using GNSS simulator designed by NAVIS Inc. It has the ability to change the carrier phase. Results are shown in Table 2.

Table 2. The measurement result of the offset carrier-phase

Offset, ps	Measured value, ps
100	99.5
150	149.8
200	199.7
500	499.7

Results have confirmed the phase offset set by the manufacturer.

5 CONCLUSIONS

The carrier-phase generation GNSS simulators error is a most importance to receiver calibration. The considered method allows us to estimate the simulator carrier phase quality. Low phase noise features stability of time parameters and high signal quality.

The method allows estimating the in-phase and quadrature QPSK-signal components orthogonality. The proposed method had validation and showed the simplicity of its practical implementation despite the limited.

REFERENCES

Global Positioning Systems Directorate (GPSD). (2018). *Systems engineering & integration interface specification IS-GPS-200J*. Los Angeles: GPSD.

Hofmann-Wellenhof, B., Lichtenegger, H. and Wasle, E. (2008). *GNSS – Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and more*. Wien: Springer-Verlag.

Perov, A. I. and Harisov, V. N. (eds). (2010). *GLONASS: Principles of Construction and Operation: Collective Monograph [Rus]*. Moscow: Radiotekhnika.

Russian Institute of Space Device Engineering (RISDE). (2008). *GLONASS Interface Control Document (ICD)*, Edition 5.1. Moscow: RISDE.

Voronov, A. S. (2007). *Signal phase difference measurement [Rus]*, Doctoral dissertation. Barnaul: Polzunov Altai State Technical University.

Sveučilište u Rijeci
POMORSKI FAKULTET
FACULTY OF MARITIME STUDIES
University of Rijeka

University of Zagreb
Faculty of Transport
and Traffic Sciences

Royal Institute of Navigation
Science Technology Practice

13th
Annual
Baška GNSS
Conference

GNSS NAVIGATION MESSAGE AUTHENTICATION USING TESLA PROTOCOL

Mia Filić

Faculty of Computer and Information Science, University of Ljubljana
Ljubljana, Slovenia, e-mail: mia.filic@fri.uni-lj.si

ABSTRACT

Satellite navigation is an enabling technology for rising number of technology and socio-economic systems and services, and a component of the national infrastructure. Information security threats add to the list of known Global Navigation Satellite System (GNSS) vulnerabilities, calling for the GNSS resilience development. Here we contribute to the efforts and results through introduction of the improved GNSS Navigation Message Authentication scheme based on the Timed Efficient Stream Loss-tolerant Authentication (TESLA) protocol. We fine-tuned the TESLA protocol utilisation by bootstrapping the GNSS receiver, and the utilisation of the Elliptic Curve Digital Signature Algorithm (ECDSA). The authentication performance accomplished is discussed in the context of GNSS applications, and the subjects of further research.

Key words: *GNSS, navigation message authentication, TESLA algorithm, anti-spoofing*

1 INTRODUCTION

The usage of the Global Navigation Satellite System (GNSS) is widespread. The GNSS term refers to the system and the constellation of satellites broadcasting ranging signals and Navigation Messages (US DoD, 2008). Many applications use GNSS to provide or maintain their services and operations (Dukare et al., 2015; Subirana, Zornoza, and Hernandez-Pajares, 2013, including:

(i) Goods/Assets tracking,

(ii) Road tolling,

(iii) Tracking (money) transaction.

In an attempt to connect an object with its position at a certain time, the information about position and time must be reliable. In explaining the relationship, here we refer to the process of geolocation. The geolocation of an object is usually determined either by using one of the GNSS systems (for instance, Global Positioning System, GPS) or using the information from telecommunication base stations (Network-based positioning) to triangulate the approximate position (Filić, 2017). The GNSS approach is more accurate and widely used. The telecommunication base station approach is used if the GNSS system fails, or as an augmentation (Subirana, Zornoza and Hernandez-Pajares, 2013).

The increasing commercial usage of GNSS raises concerns about the GNSS information authentication, usually called Navigation Message Authentication (NMA) process (see Section 3). While the military use GNSS signals are strongly authenticated, commercial use signals are exposed to spoofing or retransmission (meaconing). GNSS spoofing involves transmission of signals of greater strength and mimicking the attributes of another GNSS signal, thus taking over a GNSS receiver.

There are two main types of spoofing defence, cryptographic and non-cryptographic. Cryptographic defence has significant protection against spoofing attacks relative to the additional cost and bulk required for implementation. Non-cryptographic defence involves inertial measurement units or other hardware, which exceeds the cost, mass, or size constraints of a broad range of applications (Wesson, Rothlisberger and Humphreys, 2012).

We concentrate on the cryptographic spoofing mitigation (anti-spoofing) approach, and propose utilisation of the Timed Efficient Stream Loss-tolerant Authentication (TESLA) protocol (Perrig et al., 2001). Here, TESLA protocol was assessed for its

authentication capacity and potentials to mitigate the spoofing effects on GNSS. TESLA protocol relies on secret keys that encrypt and digitally sign components of the broadcast signals. It modulates asymmetric properties using only symmetric cryptography.

This manuscript is structured as follows. Section 2 introduces previous work exploring navigation message authentication opportunities. In Section 3, motivation and alternatives for GNSS Navigation Message authentication are discussed. It reveals the importance of cryptography in the navigation message authentication. Section 4 provides the TESLA protocol definition with emphasis on the utilisation in GNSS NMA. In addition, it provides essentials for the TESLA protocol implementation (Message Authentication Code, MAC, the one-way chain, loose time synchronisation). Finally, the final security notes and reasons for combining the Elliptic Curve Digital Signature Algorithm (ECDSA) with the TESLA protocol are to obtain an improved GNSS authentication protocol are given in Section 5.

This manuscript aims at provision of an overview of the usage of the cryptography in GNSS, and at justified suggestion for adaptation of the TESLA protocol for GNSS navigation channels, e.g. usage of commitment chain technology or modification to some stages of the TESLA protocol. Furthermore, it brings cryptographic defence against spoofing to general GNSS public, thus rendering satellite navigation more resilient and robust against the artificial sources of vulnerabilities.

2 RELATED WORK

In 2001, the U.S. Department of Transportation published a report estimating the vulnerability of the U.S. transportation infrastructure due to disruption of civil GPS (NTSC, 2001). The report emphasised the threats of spoofing and meaconing attacks, which motivated greater research of the mentioned attacks. That has led to a greater progress in anti-spoofing development. Pozzobon (2011) proposed the Navigation Message authentication concept based on signal authentication sequences. Furthermore, Scott (2003) discussed anti-spoofing methods available to civil users within the common GNSS architecture. In recent years, several NMA processes for the Gallileo GNSS have been proposed. The one described in (Fernandez-Hernandez, 2016) is based on the ECDSA and the TESLA protocol.

3 NAVIGATION MESSAGE AUTHENTICATION

To provide a position accuracy, there is a need to protect not only the content of the navigation message, but also the process of travelling time measurement. If the signal arrives delayed, pseudo-ranges are increased, thus causing the incorrect position estimate. The authentication of the navigation message should include the authentication of the navigation message content and the time-of-arrival. TESLA protocol provides authentication of the navigation message content on the packet-by-packet basis. The packet is considered to be a one frame, subframe, or the whole navigation message. Navigation message contains 25 frames where each of frames has 5 sub-frames (US DoD, 2008).

A simple deployment of a standard point-to-point authentication protocol, e.g. appending the MAC, does not provide secure broadcast authentication. The problem is that any receiver with the secret key can impersonate the sender. To prevent such an attack, we look at asymmetric cryptography schemes, digital signatures. Such a scheme provides secure broadcast authentication, but has a considerable large time and bandwidth overhead.

Another approach is to modulate asymmetric properties using only symmetric cryptography, more specifically the MACs and delayed disclosure of keys by the sender. This scheme was proposed by Cheung (Cheung, 1997) in the context of authenticating link state routing updates. Similar approach was used in the Fawkes protocol for interactive unicast communication (Anderson et al., 1998).

4 TESLA PROTOCOL DEFINITION

The Time-Efficient-State-Less-Authentication (TESLA) protocol enables all receivers to check the integrity of data in the signal and authenticate the source of the signal in the multicast or broadcast data stream environment. TESLA protocol is an efficient protocol with low communication and computation overhead, which tolerates packet loss (Perrig et al., 2002). It is based on loose time synchronisation between the sender and the receiver. Despite using symmetric cryptographic functions (MAC), the TESLA protocol achieves asymmetric properties due to a delayed disclosure of keys by the sender. TESLA protocol is widely applicable, from broadcast authentication in sensor networks (Perrig et al., 2000) to authentication of messages in ad hoc network routing protocols (Hu, Perrig and Johnson, 2005).

4.1 Essentials for TESLA protocol. In order to better understand the TESLA protocol, this subsection outlines a simple loosely time synchronisation protocol, and introduces the basic concepts of the Message Authentication Code (MAC).

The Message Authentication Code (MAC) is generally used for message authentication. In cryptography, the plain message to be sent is called plaintext, denoted by M (Figure 1). The cryptographically changed (encrypted)/appended message that will be actually sent by the sender is called ciphertext, denoted by C. As the main goal of authentication is not to hide data, but to authenticate. The ciphertext C sent by the sender is simply the original message M appended with the tag T.

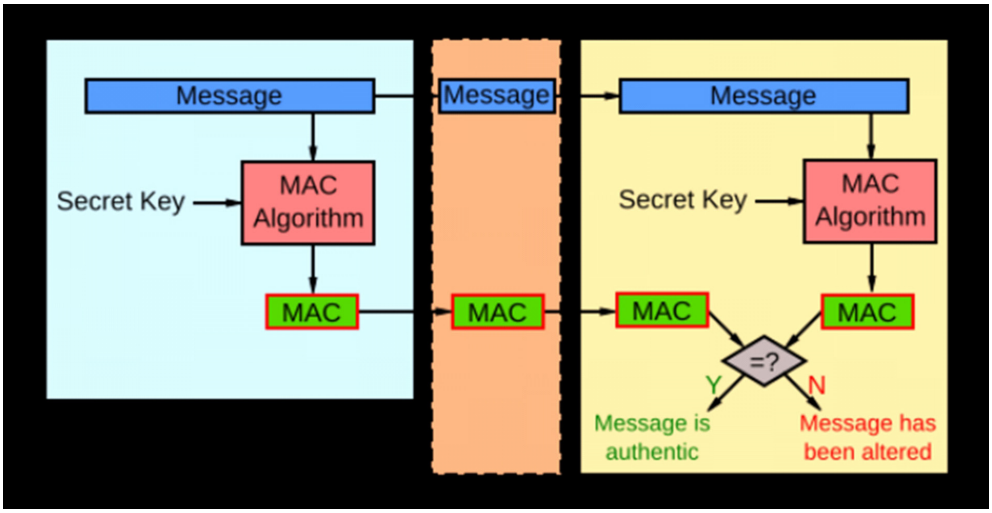


Figure 1. Message Authentication Code

Thus, the C equals to (M,T). The tag T provides an authentication of the message M. When the ciphertext has this form, we call the corresponding communication mechanism a message-authentication scheme. Such a scheme is specified by (i) the tag-generation (TG) algorithm and (ii) tag-verification (TV) algorithm. The tag-generation algorithm TG produces a tag T from a key K and the message M. The tag-verification algorithm produces a bit from a key K, a message M, and corresponding tag T. When the bit B equals to 1, it indicates to accept the message M, and to reject otherwise. If the tag generation algorithm is stateless and deterministic, it is called Message Authentication Code (MAC). The scheme whose TG and TV algorithms are MACs is simply called a MAC scheme.

MAC provides the data integrity and the authentication of a message. One example of MAC is the *hash-MAC,* or *HMAC,* (Krawczyk, Bellare and Canetti, 1997), which includes the cryptographic hash function together with the cryptographic key. Computation of the HMAC over the message M is performed as shown in equation (1).

$$H(K \; XOR \; opad; H(K \; XOR \; ipad; M)) \tag{1}$$

where:

ipad … denotes the byte 0x36, repeated b times,

opad ... denotes the byte 0x5C, repeated b times.

TESLA protocol uses MAC authentication scheme with *TG* algorithm as a keyed hash function with certain property relying on the secret key *K*.

One-way chains method is a representative of the methods class that commits a sequence of random values (Figure 2). Generation of a one-way chain involve utilisation of a hash function. The sender generates chain of the size l by randomly selecting s_l and repeatedly applying the one-way function, denoted as F, to s_l. The sender obtain the sequence (chain): $s_l; s_{l-1}; s_{l-2}; \ldots ; s_0$ where $s_i = F^{l-i}(s_i)$.

Furthermore, $s_i = F^{i-j}(s_j)$, for $j \geq i$. In addition, every element of the chain can be verified having s_0 (self-verification). The value s_0 is called the commitment to the chain. The chain can either be created at once and stored, or each element can be calculated on demand having only s_l stored in advance. Usually, the hybrid type is used to balance the storage and computation overhead.

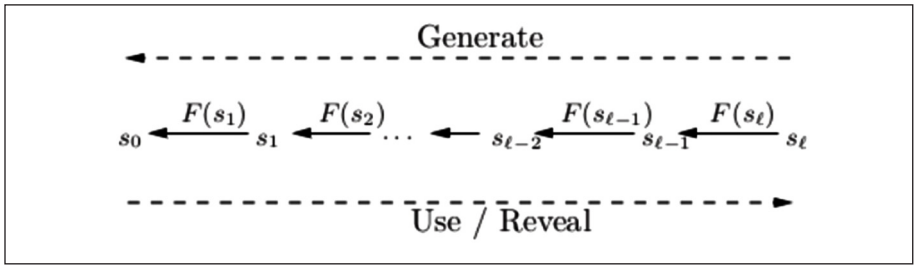


Figure 2. One way chain example (Perrig et al., 2002)

The TESLA protocol uses a one-way chain as an effective mechanism to authenticate MAC keys. The *loose time synchronisation* between the receiver and sender requires the receiver to know only the upper bound on the sender's (satellite) local time. For the needs of TESLA, a two-round time synchronisation is sufficient.

Let δ be the real time difference between the sender's and the receiver's time. In the loose time synchronisation, the receiver does not need to know the exact d but only an upper bound on it, denoted as Δ, with the reference to nomenclature presented in Figure 3. The protocol runs as follows:

(i) Receiver saves time *tr*, and sends the synchronisation request containing *Nonce* to sender.

(ii) Following the reception of the synchronisation request, sender records time *ts*, signs it together with the *Nonce*, and sends it back to receiver.

(iii) Receiver verifies the digital signature, and checks that the *Nonce* in the packet is identical to the *Nonce* it randomly generated. If the message is authentic, the receiver stores $t_R := tr$, and $t_S := ts$.

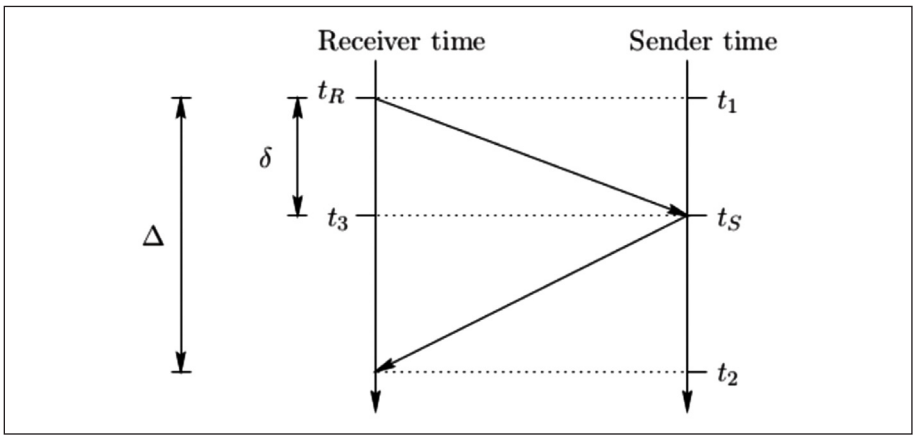


Figure 3. Direct time synchronisation between the sender and the receiver (Perrig et al., 2001)

After the initial synchronisation, at any given time t, the bounds on the sender time t_s can be calculated as

$$tr - vt_R + t_S \leq ts \leq tr - t_R + t_S + \Delta. \quad (2)$$

The risk of denial-of- service attacks, where an attacker floods the sender with time synchronisation requests, can be reduced with the aggregation of multiple requests, and signing it with the Merkle hash tree that is generated from all requesters *Nonces* (Merkle, 1980), which is sent to all the receivers (Perrig et al., 2001).

In a GNSS application, Δ is obtained from one of the positioning determination processes for which is considered to rely on authentic data. The satellite-receiver communication is a one-way communication, with $D:= 2 * \delta$. The travelling time is the same elapsed time used in positioning determination process.

4.2 Protocol definition. The cryptographic protocol TESLA separates communication between parties into sessions/communications. It sets the agreement between a receiver and a sender at the beginning of each conversation. Each communication between satellite and receiver can contain one or more conversations. Each conversation is attached to one successful Acquisition-Tracking (Subirana, Zornoza and Hernandez-Pajares, 2013) period. At the beginning of a conversation, satellite provides a receiver with data about the starting time of the conversation and the time interval between the sequence of messages correlated to the conversation. The time interval between messages can be static or change dynamically. In later case, each message needs to be supplemented with the time difference to the subsequent message. In this set up, the term message refers to the data content/useful data sent as one unit (packet or similar) through the communication channel.

Elapsed time between two messages is set to be constant. If one wants to forge the arrival time, the arrival time of all messages in the conversation need to be forged (delayed), which is a complex task to achieve. Production of a valid message sequence is hard because, at the beginning of the conversation, the first message sent by the satellite is signed by its private key, which is usually hard to forge. The definition of the TESLA protocol makes one chain-message even harder to forge.

The main idea of TESLA is to expand each message with its MAC computed with specific key K (Figure 4). The key K is known only to the sender. After the reception of the message, the receiver buffers the message, not being able to authenticate. After a short period, a previously agreed time interval, the sender reveals the key K. This way the broadcast authentication with only one MAC per message is enabled.

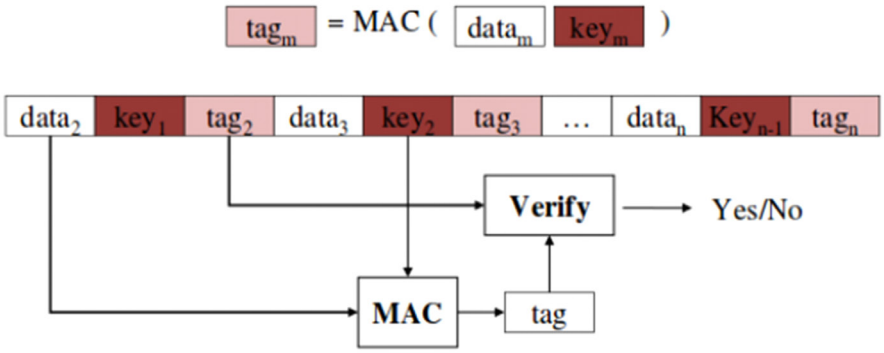


Figure 4. Overview of TESLA protocol

Under the assumption that the receiver has been already bootstrapped, the protocol runs as follows.

1. The sender determines n, natural number, length of the key chain and limits the number of messages before the new one-way key chain need to be generated. It limits the number of messages in one conversation. The sender generates random value for R_N. Using the one-way hash function, the sender generates the one-way chain of values R_N ; . . . ; R_1 ; R_0.

2. The sender splits time into N time intervals of equal duration $t_i nt$: t_0 ; . . . ; t_{N-1}; t_N. The key R_i is attached to the i-th time interval. Observe that one-way chain is used in the reverse order, so any key can be used to derive keys attached to previous intervals. The sender publishes the key R_i after the disclosure time $i \cdot L \cdot t_i nt$. The disclosure time is often given in a number of time intervals between usage and disclose of the key, denoted by L.

3. The sender attaches the MAC to each message M (i-th message sent from the time t_0). The MAC of the message M to be sent in interval $[t_i, t_{i+1}]$ is calculated using the key R_{i+1}. Along with the $MAC(M) = MAC(R_{i+1}, M)$, sender attaches the most recent key that can be disclosed, R_j, $j = i + 1 - L$.

4. Each receiver that receives the extended message (MAC(M); M; R_j) does the following:

- Calculates the travelling time (within or without the positioning determination process),

- Checks if the key used to compute the MAC is still secret by determining if the sender could not have yet reached the time interval for disclosure. If the key is still secret, buffers the extended message with its travelling time.
- Checks if the disclosed key is correct using the self-verification ($R_0 = F_j(R_j)$) and previous keys. If the key is not correct, the message is removed from the buffer.
- Checks the MAC of buffered extended messages which were sent in the interval $[t_i - L,\ t_i + 1 - L]$. If the MAC is valid, the receiver accepts only messages with the travelling time smaller than Δ.

In step 1, a pseudo random function usually generates R_N for each key chain generation. Furthermore, step 2 does not have to include the splitting of the predefined time interval into N, but defining the time duration of each. That way N subsequent intervals are again defined.

The TESLA protocol steps are divided into 4 stages (Perrig et al., 2002): 1. Sender Setup (step 1 and 2), 2. Bootstrapping receiver (prior running the protocol), 3. Broadcasting Authenticated Messages (step 3), 4. Authentication at Receiver (step 4).

In the 'Bootstraping receiver' stage, the receiver needs to be loosely time-synchronised with the sender, to find out the disclosure schedule of the keys, and to receive an authenticated key (commitment to the chain) of the one-way key chain. Observe that a number of receivers need to be synchronized with the same sender and it is really opportunistic to aspect that all of them will start the synchronisation process at the same time. The TESLA enables that the synchronisation stage can occur at any given time, not only at t_0. As all synchronized receivers receive the same data from the sender, we propose that the Bootstraping receiver stage can occur at time interval i, not only in $[t_0, t_1]$.

Bootstrapping receiver (t_0) and Broadcasting Authenticated Messages with delay factor/disclose factor $L = 1$ are defined as

$$t_0: \quad \text{Sign}(t_1, t_i - t_{i-1}, R_0); \quad R_0 = F(R_0), \tag{3}$$

$$t_1: \quad (\text{MAC}(M_1), M_1, R_1); \quad R_1 = F(R_2), \tag{4}$$

$$t_2: \quad (\text{MAC}(M_2), M_2, R_2); \quad R_2 = F(R_3), \tag{5}$$

$$t_3: \quad (\text{MAC}(M_3), M_3, R_2); \quad R_3 = F(R_4). \tag{6}$$

The process starts with the receiver synchronisation request. A response to the request, in general, contains the following information:

1. A time interval schedule, e.g. interval duration t_i *nt*, together with the start time of current period t_i, index of interval i, length of one-way key chain.
2. A key commitment to the one-way key chain R_j, where $\max_j(j) \leq i - d$, the most recent key that can be disclosed (d is the disclose offset in number of intervals),
3. A disclosed factor L.

The information is signed by the sender's private key before sending.

Observe that a key commitment to the key chain R_0 is replaced by R_j, where $\max_j(j) \leq i - d$, the most recent key that can be disclosed. Each R_i is the commitment to the chain R_i; R_{i+1}; R_{i+2}; . . .; R_N.

Finally, it should be noted that the one-way chain has a property that if one of the intermediate keys (messages) are lost, the message can be authenticated by recomputing the lost key using the later values. This renders TESLA a packet loss-resistant method. In the explained set up, the protocol protects not only the content of the navigation message, but also the process of travelling time measurement.

4.3 TESLA in Navigation Channels. In the Navigation Channels, senders are satellites, and receivers are GNSS receivers. The application of TESLA protocol in the Navigation Channels does not require considerable investments and system modifications, but the redefinition of the format of the Navigation Message NM). The NM should be extended with the MAC and corresponding hash key. Furthermore, each satellite needs to be attached with appropriate hash function and the private-public key pair. The hash function is used to generate the one-way chain of values R_N; ...;R_1; R_0. The private-public key pair used for verifying and generating signatures is usually set in Bootstraping receiver stage. The hash function can be the same for all satellites (Fernandez-Hernandez, 2016) or each can use its own (Kerns, Wesson and Humphreys, 2014). In addition, TESLA was found to perform as a good oracle for geoencryption on Loran (Qiu et al., 2007; Clifford, 2017).

In the Navigation Channels where the sender transmits messages continuously (as satellites does) and each sender's message is extended with the corresponding R_j and the disclosure factor, there is no need for an receiver to send the synchronisation request as it can synchronize with the sender using its continuously sent messages.

The only thing he must pay attention to is to be sure that the message he uses for synchronization is authentic. This shows how the TESLA protocol can be adopted for current GNSS systems. Majority of the existing and emerging GNSS systems allows only one-way communication between satellites and receivers, with Beidou being the exception.

4.4 TESLA security considerations. The security of TESLA relies on the following presumptions:

1. The receiver's clock is time synchronized up to a maximum error of D.
2. The function F is the Pseudo Random Function (PRF), and weak collision-resistant.

As long the above holds, it is computationally intractable for an attacker to forge a TESLA packet that the receiver authenticate (Perrig et al., 2000). The same PRN codes can be used by several satellites at different points in time and a satellite may have used different PRN codes at different points in time. PRN codes are created by XORing 2 bit streams generated by the linear feedback shift registers (LFSR) with maximal period 10. Different codes are obtained by delaying one of the bit streams. They repeat themselves over time. The PRN generator cannot be used as pseudo random number generator in the key-chain construction without modification.

This brings us to the need of having a pseudo random function for generation RN values for each key chain generation. If the receiver timing accuracy δ_r is smaller than $T_{int}d$˙, or larger than the time disclosed offset TESLA does not prevent an attacker from replaying an old data, or creating an arbitrary message extension for a valid packet. This can be overcome using a hybrid approach (Kerns, Wesson and Humphreys, 2014), combining TESLA with the signature scheme ECDSA. Due to (Kerns, Wesson and Humphreys, 2014), the TESLA-ECDSA protocol achieves the best Navigation Message Authentication. The scheme drastically reduces overhead of verifying a digital signature in ECDSA while preserving cryptographic authentication of navigation data for all users, with arbitrary timing accuracy. Furthermore, the cryptographic strength of the chain generation mechanism can be significantly increased by breaking the symmetry of iterations. This is done by adding additional information to the hashing process that is known to the receiver, e.g., a counter or the time Tag. This is a recommended approach. In the proposed set up, if we allow the multiuse of the same hash-chain, if an attacker can somehow extract the sender hash-chain s_0; ...; s_l from previous receiver-satellite communication, it cannot reuse it as it is valid only for the specified period of time.

5 DISCUSSION

The security of GNSS communication channels becomes increasingly important. The increasing number of GNSS applications raise the global concern about its security that can be enforced trough authentication processes. This paper discussed cryptographically based spoofing defence using TESLA protocol. It gives definition of the protocol and general utilisation in GNSS systems. The TESLA protocol uses symmetric cryptography and achieves asymmetric property, which is crucial for reliable broadcast message authentication. TESLA can avoid denial-of-service attacks, is robust, packet lost resistant with low communication and computation overhead. Used wisely, its utilisation can provide the high level of authentication security without significant changing the physical model (architecture) of the system. A variety of applications (Hu, Perrig and Johnson, 2005; Perrig et al., 2002; Suwannarath, 2016) makes the protocol evolve further. With each application, a new security analysis is provided which can form a basis for protocol evolution and maintainability. Additionally, the protocol evolution most of the time aims to increase the level of provided security.

GNSS is a part of national infrastructure, ensuring and enabling development and operation of a rising number of technology and socio-economic systems and services. It suffers from a number of vulnerabilities, including potentials of information security failures including GNSS spoofing (Filić and Dimc, 2019; Filić, 2018), thus deserving advanced protection (Caparra, 2017). Introduction of the new pseudorange measurement PRN codes (Filić and Dimc 2019) may minimise the probability of GNSS spoofing-related disruptions of GNSS Positioning, Navigation, and Timing (PNT) services, while the suitable GNSS positioning environment monitoring for GNSS spoofing (Filić, 2018) may contribute to the resilient GNSS development. GNSS modernisation processes have already anticipated the utilisation of information security measures in GNSS: The most recent example is Galileo Public Regulated Service (PRS) with encryption technique utilised. Not being disclosed for the obvious reasons, it may involve TESLA protocol, presented in this manuscript as a suitable candidate for rendering GNSS information attacks-resilient global positioning, navigation, and timing system.

6 CONCLUSION

Satellite navigation has become a grown-up technology that supports the national infrastructure, enables development and operation of technology and socio-economic systems and services, and serves as public goods. However, those

emphasise the vulnerabilities of Global Navigation Satellite Systems (GNSS) and call for the resilient GNSS development. Rise of the general computing capacity reveal the information security as the new critical GNSS vulnerability. GNSS was assumed to be protected from cyber-attacks by its way of operation and utilisation. However, system integration and technology and business development made GNSS as an underlying and enabling technology prone to cyber-attacks.

Research has been under way for some time aimed at fortification the GNSS against the information attacks. Modernisation of the GNSS systems render them more resilient against the information threats. Here we presented a contribution to the efforts by proposing deployment of the TESLA protocol for GNSS Navigation Message authentication. Developing our contribution on the TESLA protocol performance and the existing GNSS infrastructure, we fine-tuned the TESLA protocol utilisation with the introduction of the TESLA bootstrapping procedure into GNSS receiver thus enhancing the information security of the GNSS navigation message transfer process. Finally, we examined the effects of the enhancement through assessment of the GNSS application performance improvement accomplished by the TESLA-based GNSS navigation message authentication. Further research will concentrate on the prospects for the GNSS information security advancement through further utilisation of the Elliptic Curve Digital Signature Algorithms (ECDSA) throughout the GNSS architecture and the GNSS position estimation process.

Acknowledgements

Author acknowledges partial support of the research from the *Research of environmental impact on the operation of satellite navigation systems in maritime navigation* project (Project Code: uniri-tehnic-18-66), funded by University of Rijeka, Rijeka, Croatia.

REFERENCES

Anderson, R et al. (1998). A new family of authentication protocols. *ACM Operating Systems Review*, 32(4), pp. 9–20.

Caparra, G. (2017). *Authentication and Integrity Protection at Data and Physical Layer for Critical Infrastructures*. Doctoral dissertation. Padua: University of Padua.

Cheung, S. (1997). An efficient message authentication scheme for link state routing. *Proceedings of 13th Annual Computer Security Applications Conference*, pp. 90–98. San Diego: IEEE.

Clifford, G. J. (ed). (2017). *American Practical Navigator (Bowditch)*. Springfield: NGA. Available at: https://msi.nga.mil/MSISiteContent/StaticFiles/NAV_PUBS/APN/Chapt-12.pdf, accessed 12 February 2019.

Dukare, S. et al. (2015). Vehicle Tracking, Monitoring and Alerting System: A Review. *International Journal of Computer Applications*, 119(10), pp. 39-44. Available at: http://search.proquest.com/openview/767aabdc8b343ff9b5c72c1194232e4f/1?pq-origsite=gscholarcbl=136216, accessed 4 February 2019.

Fernandez-Hernandez, I. (2016). A navigation message authentication proposal for the Galileo open service. *Navigation - Journal of The Institute of Navigation*, 63(1), pp. 85-102. Available at: http://spcomnav.uab.es/docs/journals/ Navigation-FERNANDEZ 2016.pdf, accessed 2 February 2019.

Filić, M. (2018). Foundations of GNSS spoofing detection and mitigation with distributed GNSS SDR receiver. *TransNav*, 12(4), pp. 649-656.

Filić, M. (2017). *Analiza postupka procjene položaja temeljem zadanih pseudoudaljenosti u programski određenom prijamniku za satelitsku navigaciju [Cro]*, Master thesis. Zagreb: University of Zagreb, Faculty of Science, Department of Mathematics. Available at: https://zir.nsk.hr/islandora/object/pmf%3A3230, accessed 12 February 2019.

Filić, M. and Dimc, F. (2019). Logistic map-encrypted PRN code as a proposed alternative to GNSS PRN pseudo-range code. *TransNav*, 13(3), pp. 587-590.

Hu, Y.C, Perrig, A. and Johnson, D. B. (2005). Ariadne: A secure on-demand routing protocol for ad hoc networks. *Wireless Networks*, 11(1-2), pp. 21-38.

Krawczyk, H., Bellare, M. and Canetti, R. (1997). *Hmac: Keyed-hashing for message authentication*. p.11. Freemont: IETF. Available at: https://tools.ietf. org/html/rfc2104, accessed 2 February 2019.

John A. Volpe National Transportation Systems Center (NTSC). (2001). *Vulnerability assessment of the transportation infrastructure relying on the global positioning system*. Cambridge: NTSC.

Kerns, A. J., Wesson, K. D. and Humphreys, T. E. (2014). A blueprint for civil GPS navigation message authentication. *Proceedings of IEEE/ION PLANS 2014*, Monterey, California, pp. 262-269. doi: 10.1109/PLANS.2014.6851385. Manassas: ION.

Merkle, R. (1980). Protocols for public key cryptosystems. *Proceedings of 1980 IEEE Symposium on Security and Privacy* Vol. I, Oakland, California, pp. 122-134. Wahsington, D. C: IEEE Computer Society.

Qiu, D., Lo, S., Enge, P. and Boneh, D. (2007). Geoencryption using Loran. *Proceedings of the 2007 National Technical Meeting of The Institute of Navigation*, San Diego, California, pp. 104-115. Manassas: ION. Available at: https://web.stanford.edu/group/scpnt/gpslab/pubs/papers/Qiu_IONNTM_2007.pdf, accessed 4 February 2019.

Pany, T. (2010). *Navigation Signal Processing for GNSS Software Defined Receivers*. London: Artech House.

Perrig, A., Canetti, R., Song, D., and Tygar, J. (2001). Efficient and secure source authentication for multicast. *Proceedings of the Network and Distributed System Security Symposium NDSS 2001*, San Diego, California, p.12. Reston: The Internet Society. Available at: https://users.ece.cmu.edu/~adrian/projects/tesla-ndss/ndss.pdf, accessed 4 February 2019.

Perrig, A., Canetti, R., Song, D., and Tygar, J. (2000). Efficient authentication and signing of multicast streams over lossy channels. *Proceedings 2000 IEEE Symposium on Security and Privacy S&P 2000,* Berkeley, California, pp. 56-74. Piscataway: IEEE. Available at: https://people.eecs.berkeley.edu/~dawnsong/papers/tesla.pdf, accessed 4 February 2019.

Perrig, A., Canetti, R., Tygar, J. D., and Song, D. (2002). The TESLA broadcast authentication protocol. *CryptoBytes*, 5(2), pp. 2-13. Available at: https://people.eecs.berkeley.edu/~tygar/papers/TESLA_broadcast_authentication_protocol.pdf, accessed 2 February 2019.

Perrig, A., Szewczyk, R., Wen, V., Culler, D. and Tygar, J. (2002). SPINS: Security protocols for sensor networks. *Wireless Networks*, 8(5), pp. 521-534. Available at: https://netsec.ethz.ch/publications/papers/spins-wine-journal.pdf, accessed 2 February 2019.

Pozzobon, O. (2011). Keeping the spoofs out: Signal authentication services for future GNSS. *Inside GNSS,* 6, pp. 48–55. Available at: http://www.insidegnss.com/node/2570, accessed 15 February 2019.

Scott, L. (2003). Anti-spoofing and authenticated signal architectures for civil navigation systems. *Proceedings of the 2003 ION GPS/GNSS Meeting,* Portland, Oregon, pp.1542–1552. Manassas: ION.

Subirana, J. S., Zornoza, J. J., and Hernandez-Pajares, M. (2013). *GNSS Data Processing: Fundamentals and Algorithms*, Vol. 1. Paris: ESA. Available at: https://gssc.esa.int/navipedia/GNSS_Book/ESA_GNSS-Book_TM-23_Vol_I.pdf, accessed 9 February 2019.

Suwannarath, S. (2016). *The TESLA-alpha broadcast authentication protocol for building automation system,* Master thesis. Long Beach: California State University. Available at: http://media.proquest.com/media/pq/classic/doc/4092454841/fmt/ai/rep/NPDF?_s= aVHQogcJ3K6hpQ3AJdvhVG5Kr48, accessed 21 February 2019.

Thomas, M et al. (2011). *Global Navigation Space Systems: reliance and vulnerabilities*. London: RAENG. Available at: http: //www.raeng.org.uk/publications/reports/ global-navigationspace-systems, accessed 4 February 2019.

US Department of Defense (US DoD). (2008). *Global positioning system standard positioning service performance standard*. Washington D.C: US DoD. Available at: https://www.gps.gov/technical/ps/2008-SPS-performance-standard.pdf, accessed 30 January 2019.

Wesson, K., Rothlisberger, M. and Humphreys, T. (2012). Practical Cryptographic Civil GPS Signal Authentication. *Navigation – Journal of the Institute of Navigation*, 59(3), pp. 177-193. Available at: https://radionavlab.ae.utexas.edu/images/stories/files/papers/nma.pdf, accessed 6 February 2019.

Sveučilište u Rijeci
POMORSKI FAKULTET
FACULTY OF MARITIME STUDIES
University of Rijeka

University of Zagreb
Faculty of Transport
and Traffic Sciences

Royal Institute of Navigation
Science Technology Practice

13th
Annual
Baška GNSS
Conference

SATELLITE-BASED POSITIONING MODEL AS AN OPTIMISATION PROBLEM SOLUTION

Darko Špoljar[1], Stefan Ivić[2], Kristijan Lenac[2]

[1] PhD candidate, University of Rijeka, Faculty of Engineering, Rijeka, Croatia, e-mail: ds.spoljar@gmail.com

[2] University of Rijeka, Faculty of Engineering, Rijeka, Croatia

ABSTRACT

Satellite positioning can be understood as an optimisation problem. After a brief outline of the optimisation theory, this paper formulates the solution for satellite position estimation in manner of the optimisation theory application. Minimisation function is defined, along with a set of constraints, both related to satellite positioning variables. Paper concludes with a brief discussion of different optimisation variants to yield satellite position estimates, and concludes with the outline of further research in utilisation of the mathematical formulation of satellite positioning problem in assessment of satellite positioning performance.

Key words: *satellite positioning, estimation, positioning performance assessment*

1 INTRODUCTION AND MOTIVATION

Engineering problems are usually of optimisation nature, thus offering a room for fine-tuning the methodology to suit a particular engineering application. A series of classical optimisation methods, such as: single and multivariable optimisation techniques without constraints, or with inequality or with equality constraints are used traditionally in a range of engineering disciplines: constructions, mechanical, electrical and electronics engineering (Rao, 1996; Sarker and Newton, 2008; Weise, 2009). Gustafsson (2010) assembled an overview of optimisation techniques deployed in various problems of position estimation. Here we address the problem of satellite-based positioning and present it as an optimisation problem. We argue that the approach taken allows for improvement in provision of: (i) optimised Global Navigation Satellite System (GNSS) positioning performance for targeted GNSS applications, (ii) optimised utilisation of GNSS positioning resources, the computing and battery capacity in mobile user GNSS receiver in particular, and (iii) optimised computing resources utilisation in distributed communication and computing environment, such as Internet-of-Things (IoT). The aim of research is to demonstrate and raise awareness of research and commercialisation opportunities through innovative, advanced and educated utilisation of proper optimisation-based mathematical methods.

This report is structured as follows. Section 1 introduces reader into the research subject and outlines motivation for research. Section 2 defines satellite positioning as an optimisation problem, providing foundations for optimisation problem solution understanding. Section 3 discusses solution for satellite-based positioning problem from the perspective of optimisation, and provides details on prospects for GNSS positioning error mitigation with the optimised positioning procedure. Section 4 discusses research and development opportunities in satellite positioning quality improvements through utilisation of educated and innovative optimisation approach, outlines advantages and shortcomings and concludes with the prospects of further research.

2 SATELLITE POSITIONING AS AN OPTIMISATION PROBLEM

Satellite-based positioning is an experimental-based technique for global position estimation that utilises simultaneous measurements of the satellite signal propagation times from satellite to receiver aerial for at least four satellites at the

time (Filić, 2017; Filić, Grubišić and Filjar, 2018; Oxley, 2017). Measurements are prone to a number of effects caused by natural and artificial sources, including: space weather and ionospheric conditions, tropospheric conditions, malicious suppressing (jamming) of GNSS signals, and provision of faked (engineered) GNSS signals aimed at fooling the GNSS receiver to calculate pre-determined wrong position (Filić and Filjar, 2018; Filić, 2018). With a growing number of GNSS applications and rising reliance on satellite navigation, it is of utmost importance to minimise the risk of degradation of quality of GNSS-based positioning (HM GOS, 2018). Optimisation approach is suggested as a candidate solution for the growing problem and concerns (Filić, 2017; Filić, Grubišić and Filjar, 2018; Filić, 2018; Gustafsson, 2010).

The general GNSS positioning model is defined as in Equation (1) (Filić, 2017; Oxley, 2017)

$$\rho_i = \sqrt{(x_{si} - x_u)^2 + (y_{si} - y_u)^2 + (z_{si} - z_u)^2} + ct_{rec} + \epsilon_{\rho i} \tag{1}$$

where: (x_{si}, y_{si}, z_{si}) ... (known) coordinates of position of the *i*-th satellite at the time of signal transmission, (x_u, y_u, z_u) ... (un-known) co-ordinates of the user's position, c ... speed of satellite signal propagation (assumed to be equal to speed of light in vacuum), t_{rec} ... (un-known) user receiver clock error, $\epsilon_{\rho i}$ ... summarised un-corrected pseudorange measurement error due to error sources effects on satellite signal propagation (ionospheric and tropospheric delays, respectively, multipath error, etc).

Solution of the GNSS positioning problem (1) should be given as the user state vector (2), comprising the unknown variables of (1).

$$\vec{\boldsymbol{x}} = [x_u, y_u, z_u, t_{rec}]^T \tag{2}$$

Optimisation problem can be defined as minimisation of the objective function of optimisation variable under given constraints, expressed by constraints function, as given with Equation (3) (Boyd, Vandenberghe, 2004; Sarker, Newton, 2008)

$$\begin{aligned} &\min f_0(\boldsymbol{x}) \\ &\text{s.t. } f_i(\boldsymbol{x}) \leq b_i, \quad i = 1, \ldots, m \end{aligned} \tag{3}$$

where: $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ ... denotes optimisation variables of the problem, $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$ ... denotes objective function, $f_i : \mathbb{R}_n \rightarrow \mathbb{R}$, $i = 1, ..., m$ ... constraint, or constraint function, $b_1, b_2, ..., b_m$ ... denote limits of constraints.

Solution of optimisation problem is a vector $\boldsymbol{x}^*$, called optimal, providing it extends the smallest objective value of all the other candidate vectors, that satisfy the constraints (Boyd, Vandenberghe, 2004).

GNSS pseudorange measurement errors are statistically described by their mean error and the covariance matrix (Gustafsson, 2010; Filić, 2017). The mean error m of pseudorange measurement errors $\varepsilon_{\rho i}$ is defined as expectation E of $\varepsilon_{\rho i}$ (4)

$$m(\epsilon_{\rho i}) = E[\epsilon_{\rho i}]. \tag{4}$$

The covariance matrix $\boldsymbol{R}$ of pseudorange measurement errors $\boldsymbol{\epsilon}_{\rho}$ is defined by Equation (5) (Gustafsson, 2010; Filić, 2017), with the operator T denoting matrix transpose

$$\boldsymbol{R} = E[\boldsymbol{\epsilon}_{\rho} \cdot \boldsymbol{\epsilon}_{\rho}^{T}]. \tag{5}$$

Filić and Filjar (2018) defined the GNSS position estimation model in the form of (6)

$$\boldsymbol{x} = \boldsymbol{G} \cdot \boldsymbol{y} + \boldsymbol{\epsilon} \tag{6}$$

where: $\boldsymbol{G}$ ... GNSS Geometric matrix, $\boldsymbol{x}$ ... state vector (GNSS position estimation), $\boldsymbol{y}$ … observations (GNSS pseudorange measurements), $\boldsymbol{\epsilon}$ … vector of positioning errors.

The GNSS estimation process (Filić, 2017; Gustafsson, 2010; Filić and Filjar, 2018) is defined as in (7), with *hat marks* (^) denoting estimates

$$\hat{\boldsymbol{x}} = \boldsymbol{G} \cdot \hat{\boldsymbol{y}}. \tag{7}$$

Considering the presence of errors, the solution of the estimation problem is defined through the optimisation problem (Gustafsson, 2010; Filić, 2017), as in (8)

$$min\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| = min \sum_{i=1}^{n} (x_i - \hat{x})^2. \tag{8}$$

Assuming the least-square approach, according to (Filić, 2017; Filić, Grubišić, Filjar, 2018), Equation (8) can be re-written as in (9)

$$\boldsymbol{x}^{LS} = \underset{\hat{x}}{argmin} \sum_{i=1}^{n} (x_i - \hat{x})^2. \tag{9}$$

Introducing (7), and with differentiation and equalling with zero, (9) yields (10) (Filić, 2017)

$$\boldsymbol{x}^{LS} = \left(\sum_{i=1}^{n} \boldsymbol{G}_k^T \boldsymbol{G}_k \right)^{-1} \sum_{i=1}^{n} \boldsymbol{G}_k^T \boldsymbol{y}_k. \tag{10}$$

This yields the following solution of the optimisation problem (Gustafsson, 2010; Filić, 2017), as in (11)

$$\hat{\boldsymbol{x}} = (\boldsymbol{G}^T \boldsymbol{G})^{-1} \boldsymbol{G}^{-1} \boldsymbol{y}. \tag{11}$$

The residual model (Gustafsson, 2010; Filić, 2017) of (11) is given with (12)

$$\boldsymbol{r} = [\boldsymbol{I} - \boldsymbol{G}(\boldsymbol{G}^T \boldsymbol{G})^{-1} \boldsymbol{G}^T] \boldsymbol{y}. \tag{12}$$

Optimisation as defined in (8) can be resolved using the least-square weighted pseudo-inverse matrix, as presented by (Filić, 2017) in a form of (13)

$$\hat{\boldsymbol{x}}_w = [\boldsymbol{G}^T \boldsymbol{W} \boldsymbol{G}]^{-1} \boldsymbol{G}^T \boldsymbol{W} \boldsymbol{y}. \tag{13}$$

Optimised GNSS positioning defined by (13) allows for mitigation of error sources with known statistical description through bespoke weighted coefficients $\boldsymbol{W}$ (Gustafsson, 2010; Filić, 2017). Filić (2017) addressed the ionospheric effects on GNSS positioning performance with suitably defined weights $\boldsymbol{W}$ related to ionospheric conditions. Our research will address the multipath effects in accordance to their statistical description (Hannah, 2001) in the manner analogous to (Filić, 2017).

3 SOLUTION TO SATELLITE POSITION ESTIMATION PROBLEM FROM OPTIMISATION PERSPECTIVE

Satellite position estimation model (1) is of a non-linear nature, and comprises the over-all pseudorange measurement error $\boldsymbol{\epsilon}_{\rho i}$ from the *i*-th satellite, of generally stochastic nature.

A 4 × 4 independent equations (13) system can be formed, with pseudoranges ρi, i = 1, ..., 4 measured simultaneously. The system's solution will be derived using the optimisation approach.

Let us define the optimisation function as given in (14)

$$p_i(\vec{\boldsymbol{x}}) = -\boldsymbol{\epsilon}_i. \tag{14}$$

Using (1), Equation (14) may be expressed as in (15), with notation $d_{trec} \coloneqq ct_{rec}$

$$p_i(\vec{\boldsymbol{x}}) = \sqrt{(x - x_{si})^2 + (y - y_{si})^2 + (z - z_{si})^2} + d_{trec} - \rho_i. \tag{15}$$

Using the optimisation function $p_i(\vec{\boldsymbol{x}})$, satellite position estimation problem may be defined as the optimisation problem using (16) for points $\vec{\boldsymbol{x}} + \Delta\vec{\boldsymbol{x}}$

$$\min p_i(\vec{\boldsymbol{x}}) \coloneqq p_i(\vec{\boldsymbol{x}} + \Delta\vec{\boldsymbol{x}}) \to 0. \tag{16}$$

With the introduction of incremental steps, the satellite position estimation solution may be assumed as an iterative process in which every iteration yields new estimate of the vector of state variables incremental steps, as given in (17)

$$\vec{\boldsymbol{x}}_{k+1} = \vec{\boldsymbol{x}}_k + \Delta\vec{\boldsymbol{x}}, \quad k = 0, \ldots, N. \tag{17}$$

Method (17) requires definition of initial conditions in the form of initial estimate of position and receiver clock error, and a set of targeted precision constraints $(\epsilon_{con-northing}, \epsilon_{con-easting}, \epsilon_{con-vertical})$ (18) to terminate iteration (17)

$$N = f(\epsilon_{con-northing}, \epsilon_{con-easting}, \epsilon_{con-vertical}). \tag{18}$$

A GNSS receiver embedded in a smartphone may utilise the alternative means for initial position estimate, such as an estimate resulting from utilisation of a telecommunications position estimation method (Cell ID, E-OTD, or any other).

A system of equations of the type (15) may be linearized using Taylor series in the vicinity of position estimate defined by iteration process (17). Taylor series of (15) is given in (19) as

$$p_i(\vec{x} + \Delta\vec{x}) - p_i(\vec{x}) = \frac{\partial p_i(\vec{x})}{\partial x} \cdot \Delta x + \frac{\partial p_i(\vec{x})}{\partial y} \cdot \Delta y + \frac{\partial p_i(\vec{x})}{\partial z} \cdot \Delta z + \frac{\partial p_i(\vec{x})}{\partial d_{trec}} \cdot \Delta d_{trec}. \quad (19)$$

Using the optimisation function (16), Equation (19) may be re-written as (20), with Δx, Δy, Δz and Δd_{trec} as the unknowns, to yield Equation (19)

$$-p_i(\vec{x}) = \frac{\partial p_i(\vec{x})}{\partial x} \cdot \Delta x + \frac{\partial p_i(\vec{x})}{\partial y} \cdot \Delta y + \frac{\partial p_i(\vec{x})}{\partial z} \cdot \Delta z + \frac{\partial p_i(\vec{x})}{\partial d_{trec}} \cdot \Delta d_{trec}. \quad (20)$$

Equation (20) expresses the estimate of the over-all pseudorange measurement error $\boldsymbol{\epsilon}_i$, as evident from (14).

Let us define $R_{i,k}$ as a k-step estimate of true range between the i-th satellite and user receiver, as given in (21)

$$R_{i,k} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2 + (z_i - z_k)^2}. \quad (21)$$

After Taylor expansion (19) and (20), a system of four Equation (15) becomes (22), written in matrix form for simplicity. The large 4 x 4 matrix is known as the GNSS *geometric matrix* (6), due to its definition related to position estimates of both satellite and receiver

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ t_{clock} \end{bmatrix} = \begin{bmatrix} \frac{x_u - x_{s1}}{R_{1,k}} & \frac{y_u - y_{s1}}{R_{1,k}} & \frac{z_u - z_{s1}}{R_{1,k}} & c \\ \frac{x_u - x_{s2}}{R_{2,k}} & \frac{y_u - y_{s2}}{R_{2,k}} & \frac{z_u - z_{s2}}{R_{2,k}} & c \\ \frac{x_u - x_{s3}}{R_{3,k}} & \frac{y_u - y_{s3}}{R_{3,k}} & \frac{z_u - z_{s3}}{R_{3,k}} & c \\ \frac{x_u - x_{s4}}{R_{4,k}} & \frac{y_u - y_{s4}}{R_{4,k}} & \frac{z_u - z_{s4}}{R_{4,k}} & c \end{bmatrix} \cdot \begin{bmatrix} \rho_1 - R_{1,k} \\ \rho_2 - R_{2,k} \\ \rho_3 - R_{3,k} \\ \rho_4 - R_{4,k} \end{bmatrix}. \quad (22)$$

Equation (22) may be written as expressed with (23), with $\boldsymbol{G}$ as the GNSS geometry matrix (6), and $\Delta\vec{\boldsymbol{\rho}}$ denoting the vector of pseudorange measurement errors on the right side of (22)

$$\Delta\vec{x} = \boldsymbol{G} \cdot \Delta\vec{\rho}. \tag{23}$$

The iteration process (24) is repeated until all the unknown variable meet the iteration closure criteria, set by the minimum acceptable pseudorange measurement errors (18) from related satellites.

$$\begin{aligned} x_{k+1} &= x_k + \Delta x \\ y_{k+1} &= y_k + \Delta y \\ z_{k+1} &= z_k + \Delta z. \end{aligned} \tag{24}$$

Satellite position estimation process defined by (23) and (24), and resulting from the utilisation of an optimisation approach, usually converges quickly, mostly after several iterations (Filić, 2017; Filić, Grubišić and Filjar, 2018).

We examined potential implementation of optimised GNSS positioning methods in various environments, including supporting libraries in C/C++, Matlab, and the R framework for statistical computing. We developed code implementation of the (24) & (25) system, with the particular emphasis on utilisation of the R framework for statistical computing.

During assessment of various computing environments, we explored the R library CVXR (Fu, Balasubramanian and Boyd, 2017), for its flexibility in formal definition of various classes of optimisation problems, and deployment of numerous methods for optimisation problems solutions. In a sense of an exercise, we programmed the (17) problem with the CVXR directly, and compared the GNSS single point solution with the one obtained with the (24) & (25) system. Comparison yields just a minor difference in the two concurrent approaches examined.

4 CONCLUSION

This paper aims at theoretical definition of the satellite positioning as an optimisation problem solution. The optimisation approach taken allowed for introduction of inherent methodology for partial satellite positioning errors mitigation. Our research is to address the potentials for multipath effects mitigation, thus providing the framework for improved GNSS utilisation quality in positioning environment encountered by growing number of GNSS applications users (urban areas, forests and parks).

We observed just a minor difference in two competitive results. CVXR provides a systematic way for direct implementation of optimisation approach in computer

engineering. We intend to pursue the examination of performance of different variants of optimised GNSS positioning methods in CVXR to assess the potentials for GNSS multipath mitigation using optimisation approach, thus contributing to GNSS resilience development through utilisation of mathematics.

Acknowledgments

Authors acknowledge partial support of his research by the research project *Embedded system for 3D perception*, funded by University of Rijeka, Rijeka, Croatia.

REFERENCES

Boyd, S, Vandenberghe, L. (2004). *Convex Optimization*. New York: Cambridge University Press. Available at: https://stanford.io/1FPvwqC, accessed 13 April 2019.

Filić, M. (2018). Foundations of GNSS spoofing detection and mitigation with distributed GNSS SDR receiver. *TransNav*, 12(4), pp. 649-656. DOI: 10.12716/1001.12.04.01

Filić, M. (2017) *Analiza postupka procjene položaja temeljem zadanih pseudo-udaljenosti u programski određenom prijamniku za satelitsku navigaciju [Cro],* Master thesis. Zagreb: University of Zagreb, Faculty of Science, Department of Mathematics. Available at: https://bit.ly/2vGmyN8, accesed 12 March 2019.

Filić, M. and Filjar, R. (2018). *Forecasting model of space weather-driven GNSS positioning performance*. Riga: Lambert Academic Publishing.

Filić, M, Grubišić, L. and Filjar, R. (2018). Improvement of standard GPS position estimation algorithm through utilization of Weighted Least-Square approach. *Proceedings of 11th Annual Baška GNSS Conference*, Baška, Croatia, pp. 7-19. Rijeka: University of Rijeka, Faculty of Maritime StudiesAvailable at: https://bit.ly/2sLuR82, accessed 1 May 2019.

Fu, A., Balasubramanian N. and Boyd, S. (2017). *CVXR: An R Package for Disciplined Convex Optimization,* Working Paper. Available at: https://stanford.io/2QhSFip, accessed 1 May 2019.

Gustafsson, F. (2010). *Statistical Sensor Fusion*. Linkoepping: Studentlitteratur.

Hannah, B. M. (2001). *Modelling and Simulation of GPS Multipath Propagation,* Doctoral dissertation. Brisbane: Queensland Institute of Technology.

HM Government Office for Science (HM GOS). (2018). Satellite-Derived Time and Position: A Study of Critical Dependencies. London: HM GOS. Available at: https://bit.ly/2E2STnd, accessed 25 March 2019.

Oxley, A. (2017). *Uncertainties in GPS Positioning: A Mathematical Discourse*. London: Academic Press/Elsevier.

Rao, S. S. (1996). *Engineering Optimization: Theory and Practice (3rd edition)*. Toronto: John Wiley & Sons.

Sarker, R. A. and Newton, C. S. (2008). *Optimization Modelling: A Practical Approach*. Boca Raton: CRC Press.

Weise, T. (2009). *Global Optimization – Theory and Application [e-book]*. Available at: http://www.it-weise.de/projects/book.pdf, accessed 12 March 2019.

Sveučilište u Rijeci
POMORSKI FAKULTET
FACULTY OF MARITIME STUDIES
University of Rijeka

University of Zagreb
Faculty of Transport
and Traffic Sciences

Royal Institute of Navigation
Science Technology Practice

13th
Annual
Baška GNSS
Conference

GNSS-BASED MARITIME NAVIGATION SYSTEMS: CYBER THREATS SOURCING

Boris Sviličić

University of Rijeka, Faculty of Maritime Studies, Studentska ulica 2, 51000 Rijeka, Croatia, *e-mail: svilicic@pfri.hr*

ABSTRACT

The Global Navigation Satellite System (GNSS) has strongly improved the safety of navigation by providing highly accurate position data in the real time. The GNSS data integration with electronic navigation charts and radar data was the basis for development of complex and cyber technology based shipboard navigation systems. This paper presents a comparative cyber security analysis of risks threatening two GNSS-based systems, a shipboard ECDIS and chart-radar. The analysis is based on the combination of the cyber security testing results and the ships' crew interview. The results suggest that the cyber threats are mainly in vulnerabilities of the GNSS-based systems' software underlying operating system.

Key words: *GNSS, navigation safety, ECDIS, chart-radar, maritime cyber security, cyber security testing*

1 INTRODUCTION

The Global Navigation Satellite System (GNSS) has strongly influenced the development of the shipboard navigation equipment by providing highly accurate position data in the real time, and thus significantly improved the safety of navigation (Brčić and Žuškin, 2018). The integration of the GNSS data with the Electronic Navigation Chart (ENC) and with the radar data has resulted in development of critical shipboard navigation systems, such as the Electronic Chart Display and Information System (ECDIS) and the chart-radar. The both GNSS-based systems rely on computing and communication (cyber) technologies, meaning that the ECDIS and chart-radar are actually software platforms that provide the data integration. The International Maritime Organization (IMO) has regulated the functionality of the ECDIS and radar software with the performance standards (IMO, 2017; IMO, 2007).

The maritime community has recognized a need for protecting the ship navigation systems from cyber threats (Svilicic et al., 2019a; Tam and Jones, 2019; Svilicic et al., 2019b; Svilicic et al., 2019c; Safet4sea, 2019; Hareide et al., 2018; Kessler, Craiger and Haass, 2018; Lewis et al., 2018; Goudossis and Katsikas, 2018; Lee et al., 2017). Therefore, IMO has issued the general guidelines to manage the maritime cyber risks (IMO, 2017a), where the cyber risk is defined as a measure of the extent to which a technology asset is threatened by a potential circumstance or event, which may result in shipping-related operational, safety or security failures as a consequence of information or systems being corrupted, lost or compromised. In addition, IMO encouraged administrations to include cyber risks assessment in the safety management systems no later than the first annual verification of the document of compliance after the 1st January 2021 (IMO, 2017b). Furthermore, jointly with the International Electrotechnical Commission, IMO is preparing a new maritime standard on the cyber security of maritime navigation and radiocommunication equipment and systems, IEC 63154 “Cybersecurity - General requirements, methods of testing and required test results” (IEC 2019).

Recently, cyber security testing of two GNSS-based shipboard systems, an ECDIS and a chart-radar are presented (Svilicic et al., 2019a; Svilicic et al., 2019b). The both ECDIS and chart-radar are type approved systems from different manufacturers. The systems are installed on two ships that are of the different types, a training ship and a ro-ro passenger ferry (Figure 1). The cyber security testing was performed using an industry vulnerability scanner. This paper presents a

comparative analysis of the cyber risk threatening the GNSS-based systems. The analysis is based on the combination of the cyber security test results and the ships' crew interview.

(a)

(b)

Figure 1. The training ship (a) and ro-ro passenger ferry (b).

2 GNSS-BASED SHIPBOARD SYSTEMS

The analysed GNSS-based systems are an ECDIS and a chart-radar that are installed on a training ship and ro-ro passenger ferry, respectively. The ships are involved in international voyages. The systems are from different manufactures, the Japan Radio Company (ECDIS model JAN-901B) and Wärtsilä SAM Electronics (chart radar model NACOS RADARPILOT Platinum). Technical specifications are given in Table 1.

Table 1. The GNSS-based systems' specifications.

		ECDIS	Chart-radar
GNNS-based system	Manufacturer	Japan Radio Co. Ltd.	Wärtsilä SAM Electronics GmbH
	Model	JAN-901B	NACOS RADARPILOT Platinum
	Software version	KG 01130	2.1.02.10
	IMO compliant	Yes	Yes
Charts	IHO ENC	IHO S-57	IHO S-57
	IHO RNC	IHO S-61	IHO S-61
	IHO Chart Content	IHO S-52	IHO S-52
	IHO Data Ptotection	IHO S-63	IHO S-63
Interfaces	Serial NMEA	IEC1162-1	IEC61162-1
	Serial high speed	IEC61162-2	IEC61162-2
	Network	Ethernet (LAN)	Ethernet (LAN)
	Chart Update	USB	USB
	Remote maintenance	Possible	Possible

The both systems are type approved. The ECDIS integrates mandatory GPS data with electronic navigational charts and mandatory position information from GPS, mandatory sensors (gyrocompass and Doppler log) and additional sensors (AIS, Navtex and echo sounder). The chart-radar integrates mandatory GPS data with electronic navigational charts and data of the x-band radar scanner. In addition, sensor data from gyrocompass, speed log, AIS, EFPS, Navtex, echo sounder and anemometer are also integrated.

3 CYBER SECURITY TESTING

The cyber security testing of the GNSS-based systems was preformed using a vulnerability scanner, the Nessus Professional (Nessus, 2019). The main goal of the vulnerability scanning is detection of all known vulnerabilities of the shipboard systems (Svilicic et al., 2018). The testing setup is shown in Figure 2.

(a)

(b)

Figure 2. The testing of the ECDIS (a) and chart-radar (b).

The testing results are shown in Figure 3 (Svilicic et al., 2019a; Svilicic et al., 2019b). In the case of the ECDIS, the detected critical cyber vulnerabilities alert that the ECDIS software is running on the Microsoft XP Embedded operating system. This version of the operating system has not been supported by the manufacturer for more than four years. In the case of the chart-radar, the version of the operating system is the Microsoft Windows 7 Professional (Service Pack 1). The manufacturer will discontinue the support of this operating system by the end of the current 2019 year (Microsoft, 2019). The lack of the support implies that manufacturer does not investigate or publish reports on new vulnerabilities, which allows an attacker to exploit known vulnerabilities using widely available guidelines. The manufacturer recommends migration to the actual version of the operating systems, which, however, could significantly affect the performance of the systems, and it is to be implemented only by the equipment manufacturer.

Figure 3. The testing results for the ECDIS (a) and chart-radar (b).

For the both ECDIS and chart-radar, the test results alert detection of the critically vulnerable version of the Server Message Block (SMB) service that is running on the systems. The vulnerable service is as an integral part of the both operating systems. The service provides file and printer sharing, and due to lack of security features (Microsoft, 2017), represents a threat vector for distribution of the NotPetya malicious software (CERT US, 2017). The NotPetya is ransomware that caused one of the most known maritime cyber incidents, the NotPetya attack on the Maersk shipping company (CERT CH, 2017). The recommended secure setup of the underlying operating systems by discontinuation of the use of the vulnerable service and update of the operating systems with a set of security patches, could also affect the systems, and it is to be done by the equipment manufacturer. It is worth mentioning that the same vulnerability is detected on the ECDIS and the chart-radar that are from different manufactures (see Table 1), installed on two different ships, with different underlying operating systems, but from the same manufacturer of the underlying operating system.

4 CYBER THREATS OF GNSS-BASED SYSTEMS

Even the cyber security testing allows for detection of all known vulnerabilities existing on the GNSS-based systems, the outcomes could incorrectly represent the real level of risk (Svilicic et al., 2018). Therefore, the outcomes are analysed regarding to the GNSS-based system operating environment and implemented protection measures. Identification of implemented protection measures was conducted by interviewing the ships' crew, in particular the ships' master and first officer. The interview was focused on two segments of the protections regarding the GNSS-based systems' operating environment: the security management system implemented on the ships and the systems' network integration. For the both ships, the same security level of implemented protections is identified. The GNSS-based systems operate in the stand-alone configuration with no Internet

connection, strong access controls are implemented, the crews' training and awareness is at high level, the security policies and procedures are adhered by the crews, and the systems are continuously evaluated, which all is traditionally ingrained in the shipping industry.

On the basis on the collected results with the cyber security testing and the interviews, the cyber threats identified have been qualitatively analysed in order to determine the risk level. Table 2 shows the identified cyber threats, together with estimated impact magnitude level and likelihood rate. The impact magnitude represents a damage resulting from a threat execution (with a value from 0 to 100), while likelihood represents a probability that a threat is executed (with a value from 0 to 1). The given values of the impact magnitude and likelihood are discussed with the risk level analysis in the following part of the chapter.

Table 2. GNSS-based system cyber threats.

No.	Treat	Description	Impact magnitude	Likeli-hood
1.	Underlying operating system out of date	Allows exploitation of well known vulnerabilities of the INS underlying operating system	100	0.4
2.	Underlying operating system insecure setup	Backdoors are open for possible intrusions and performance are reduced	100	0.4
3.	Crew training	Ship crew is not able to adequately perform their duties and responsibilities	50	0.2
4.	Crew awarnwess	Ship crew is not able to adequately adhere policies and procedures	50	0.2
5.	Internet connection establishment	Remote attacker is provided with access to the INS's navigational tools	100	0.1
6.	Unauthorized access	Attacker is provided with physical or logical access to the INS's navigational tools	100	0.1
7.	Cyber security polices and procedures	Shep crew is not aware of their roles and responsibilities	20	0.5
8.	Continuous assessment and improvement	Lack of ability to respond to rapid technological development	20	0.2

Risk level of the cyber threats is calculated by multiplying the impact magnitude and likelihood values, and are shown on the radar graph in Figure 4. The product of multiplication represented the qualitative cyber risk level: (i) acceptable low risk (the product is lower than 25), (ii) medium risk that is acceptable for a short time (the product is between 25 and 50), (iii) high risk demanding a risk mitigation plan (the product is between 50 and 75), and (iv) critical risk demanding instant action (the product is higher than 75). As it can be seen from the graph in Figure 4, our analysis has shown that the risk level of cyber threats determined coincide well for each of the GNSS-based systems, and that why only one curve is shown in Figure 4. From in total eight identified threats, six of them are assigned with low risk level. The acceptable low risk level is attributed to the protection measures that are traditionally implemented in the shipping industry, the continues training and awareness of the crew, strong access controls, security policies and procedures, and continuous assessment. In addition, the GNSS-based systems are not connected to Internet.

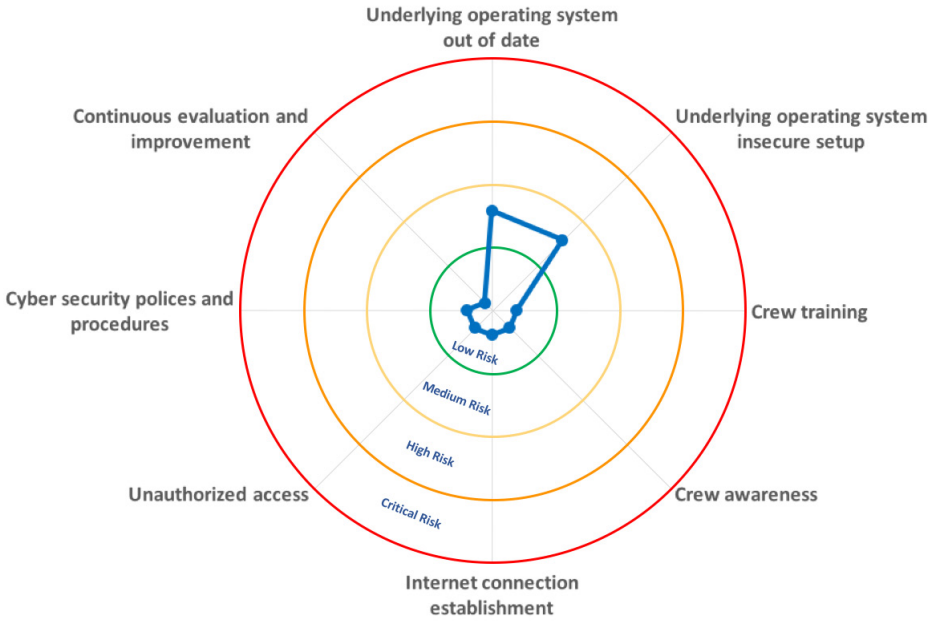


Figure 4. Radar graph of the cyber risks threating GNSS-based systems.

While most of the threats are classified with acceptable risk level, the two distinguish threats (medium risk level) are related to the GNSS-based systems underlying operating system update and secure setup. The distinguish threats identified imply that an attacker can exploit a known vulnerability using publicly available instructions without significant expertise in GNSS navigation and computing technologies. The secure setup of the underlying operating system by disabling services and features that are not needed for the GNSS-based systems operation, allows not only better performance of the systems, but also provides proactive protection from unknown threats.

5 CONCLUSIONS

The comparative cyber security analysis of two shipboard GNSS-based systems, the ECDIS and chart-radar, is presented. The GNSS-based systems are the type approved and installed on-board of two ships of different types that are involved in international voyage. The analysis is based on the combination of the cyber security testing with a vulnerability scanner and the ships' crew interview. The cyber threats identified were analysed qualitatively. The obtained results suggest that the cyber threats sources are mainly in weaknesses of the shipboard GNSS-based systems software underlying operating system. The results contribute to understanding of the cyber security of the shipboard GNSS-based navigation systems. In addition, the study indicates importance of the cyber security testing and contributes to the development of the new standard IEC 63154.

Acknowledgments

The research was financially supported by the University of Rijeka under the research project Cyber Security of Maritime ICT-Based Systems (grant number: uniri-tehnic-18-68).

REFERENCES

Brčić D. and Žuškin S. (2018). Towards Paperless Vessels: A Master's Perspective. *Journal of Maritime and Transportation Sciences – Pomorski zbornik,* 55, pp. 183-199.

Goudossis, A. and Katsikas, S. K. (2018). Towards a secure automatic identification system (AIS). *Journal of Marine Science and Technology*, 24, pp. 410-423.

Hareide, O. S., Jøsok, Ø., Lund, M. S., Ostnes, R. and Helkala, K. (2018). Enhancing Navigator Competence by Demonstrating Maritime Cyber Security. *Journal of Navigation*, 71, pp. 1025-1039.

International Electrotechnical Commission (IEC). (2019). *IEC 63154 ED1: Maritime navigation and radiocommunication equipment and systems - Cybersecurity - General requirements, methods of testing and required test results*. Geneva: IEC.

International Maritime Organization (IMO). (2017a). *MSC-FAL.1/Circ.3: Guidelines on maritime cyber risk management*. London: IMO.

International Maritime Organization (IMO). (2017b). *MSC 98/23/Add.1: Maritime Cyber Risk Management in Safety Management Systems*. London: IMO.

International Maritime Organization (IMO). (2017c). *MSC.1/Circ.1503/Rev.1: ECDIS – Guidance for Good Practice*. London: IMO.

Kessler, G. C., Craiger, J. P. and Haass, J. C. (2018). A Taxonomy Framework for Maritime Cybersecurity: A Demonstration Using the Automatic Identification System. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, 12, pp. 429-437.

Lee, Y. C., Park, S. K., Lee, W. K. and Kang, J. (2017). Improving cyber security awareness in maritime transport: A way forward. *Journal of the Korean Society of Marine Engineering*, 41, pp. 738-745.

Lewis, S., Maynard, L., Chow, C. E. and Akos, D. (2018). Secure GPS Data for Critical Infrastructure and Key Resources: Cross-Layered Integrity Processing and Alerting Service. *Navigation, Journal of The Institute of Navigation*, 65, pp. 389-403.

Microsoft. (2019). Microsoft: Search product lifecycle. Available at: https://support.microsoft.com/en-us/lifecycle, accessed 31 April 2019.

Microsoft. (2017). Microsoft Security Bulletin MS17-010 - Critical. Available at: https://technet.microsoft.com/library/security/MS17-010, accessed 25 April 2019

Nessus. (2019). Tenable Products: Nessus Professional. Available at: https://www.tenable.com/products/nessus/nessus-professional, accessed 1 March 2019.

Safey4sea. (2019). 2018 Highlights: Major cyber attacks reported in maritime industry. Available at: https://safety4sea.com/cm-2018-highlights-major-cyber-attacks-reported-in-maritime-industry, accessed 10 July 2019.

Svilicic, B., Kamahara, J., Rooks, M. and Yano, Y. (2019a). Maritime Cyber Risk Management: An Experimental Ship Assessment. *Journal of Navigation*, 72, pp. 1108-1120.

Svilicic, B., Rudan, I., Jugović, A. and Zec, D. (2019b). Cyber Security Testing of Shipboard Chart Radar. *Proceedings of 20th IAMU AGA Conference*, Tokyo, Japan, pp. 129-134. Tokyo: IAMU.

Svilicic, B., Rudan, I., Frančić, V., Mohović, Đ. (2019c). Towards a Cyber Secure Shipboard Radar. *Journal of Navigation,* p. 12. DOI: 10.1017/S0373463319000808.

Svilicic, B., Celic, J., Kamahara, J. and Bolmsten, J. (2018). A Framework for Cyber Security Risk Assessment of Ships. *Proceedings of 19th IAMU AGA Conference,* Barcelona, Spain, pp 21-28. Barcelona: UPC/CIMNE.

Swiss Government Computer Emergency Response Team (CERT CH). (2017). Notes About The NotPetya Ransomware. Available at https://www.govcert.admin.ch/blog/32/notes-about-the-notpetya-ransomware#, accessed 10 May 2019.

Tam, K. and Jones, K. (2019). MaCRA: a model-based framework for maritime cyber-risk assessment. *WMU Journal of Maritime Affairs*, 18, pp. 129-163.

United States Computer Emergency Readiness Team (CERT US). (2017). Alert (TA17-181A) Petya Ransomware. Available at: https://www.us-cert.gov/ncas/alerts/TA17-181A, accessed 1 September 2019.

Sveučilište u Rijeci
POMORSKI FAKULTET
FACULTY OF MARITIME STUDIES
University of Rijeka

University of Zagreb
Faculty of Transport
and Traffic Sciences

Royal Institute of Navigation
Science Technology Practice

13th
Annual
Baška GNSS
Conference

A BLOCKCHAIN APPLICATION FOR VALIDATING A PATH TRAVELED BY A DRONE

Silvio Šimunić, Kristijan Lenac

University of Rijeka, Faculty of Engineering, Rijeka, Croatia,
e-mail: klenac@riteh.hr

ABSTRACT

This paper examines how blockchain technology can be applied in validation of a path traveled by a drone. A system that consists of smart contracts, drones, gateway and web interface was developed. Drones are registered within smart contract with their control information and set up with configuration containing private key. After each flight, collected and signed data containing the path is dispatched to the gateway and saved on the blockchain. End users are able to reliably verify the information by comparing the cryptographic hash of the asserted path to the one saved on the blockchain. Beacons that are able to observe nearby drones and record those events on the blockchain can further help strengthen validation claims.

Key words: *path validation, drones, blockchain*

1 INTRODUCTION

As usage of drones and their application in various industries continues to rise, there is a need to be able to testify their traveled path. Common method of verifying traversed locations relies on centralized client-server model, where plain Global Navigation Satellite System (GNSS) coordinates are stored locally or in the cloud, and later distributed to end users (Shakhatreh et al., 2019). Data from such sources might be subject to tampering and not satisfactory in situations where the importance is put on trust and validity of information (Gaetani et al., 2017).

This paper examines how blockchain technology can be applied in validation of traveled paths. Blockchain is a decentralized and immutable ledger that is used to store records without the need of central authority (Zheng et al., 2017). It provides trust where involved parties don't trust each other. Secured by cryptography, it can only be updated via consensus between all parties involved, but its history can't be changed. The reason for using blockchain over a regular database is because it provides proof of data existence at a certain time and guarantees that data wasn't tampered with (Parker, 2015). Proof-of-Location concept (Amoretti et al., 2018) can be achieved by storing data originating from both the drone and also from external beacons encountered during the flight. Flight data stored on centralized services can now be decentralized, simply by storing a hash (i.e. fingerprint) of information on blockchain. Today, modern blockchain technology is equipped with smart contracts, which furthermore enrich blockchain's functionality (Christidis and Devetsikiotis, 2016). Smart contract is a program that runs on the blockchain whose methods, when remotely invoked, perform actions that execute logic and either store or return the data. By leveraging those properties, end users can perform validation and be sure that the drone really traveled the path that is claimed to be traveled.

In this paper, focus is put on how can drones and blockchain technology work together, and more specifically, how can blockchain be used to validate the path traveled by a drone. First, the system consisting of smart contracts, drones, gateway and web interface is described. Next, drone flight session is broken down and different ways of path validations are described. Finally, conclusion is made and future work is discussed.

2 SYSTEM COMPONENTS

2.1 Smart contracts. In total, three smart contracts that work together and run on a public Ethereum (Wood, 2014) blockchain platform were developed. Functionalities

for generic Internet of Things (IoT) devices, drones and beacons were put in separate smart contracts to achieve extensibility. As a first step, drone must be registered on the blockchain as an IoT device through one smart contract, and then take a role of a drone through another smart contract. To become a beacon within the system, device must register itself as such through separate smart contract. It is important to note that writing to public blockchains is expensive and cost depends on the current network state, while reading from them is free of charge. Communication between all smart contracts is shown in Figure 1. Solid lines represent data stores and dashed lines represent invocable methods.

DeviceManager is a main smart contract that is primarily responsible for registration of IoT devices on the blockchain. Each device is registered with identifier, hash of metadata information and hash of firmware. Identifier is a public key or Ethereum address, which is controlled by the corresponding private key. Metadata information, such as device manufacturer and model can be saved in form of a Merkle tree root hash -- type of binary hash tree that allows efficient and secure verification of content (Mykletun, Narasimha and Tsudik, 2015). Additionally, hash of currently installed firmware can be saved for integrity verification purposes. After successful registration with described properties through smart contract, each device is assigned a unique identifier *device_ID* for further use. At a later date, when the data from device needs to be saved on the blockchain, appropriate methods for validating signed messages sent by devices will be available to consumers and are free of charge. This smart contract provides authentication, integrity and non-repudiation in communication between devices.

DroneManager extends DeviceManager smart contract and stores drone flights. Any previously registered device can take a role of a drone, and as such can store flights on the blockchain. Each flight record consists of *device_ID*, source and target location of the flight, departure and arrival time, and most importantly, flight data hash. Flight data is what drone collects during its travel, including path coordinates. If drone is equipped with sensors, their readings can be saved as well during flight session, and then later hashed and finally stored on the blockchain. To be able to access data at further date, flight_ID is assigned to each newly created record.

BeaconManager is a final smart contract that adds additional functionality of validating flights with beacons. Same as DroneManager, it also connects to DeviceManager smart contract and enables any device to register itself as a beacon on a specific location with an identifier. Role of a beacon is to observe nearby devices

and save information about those events on the blockchain. In a case of observing a drone, beacon will save drone identifier, location and time of observation. Additionally, a drone can also save the corresponding information about observed beacons on its flight through DroneManager smart contract.

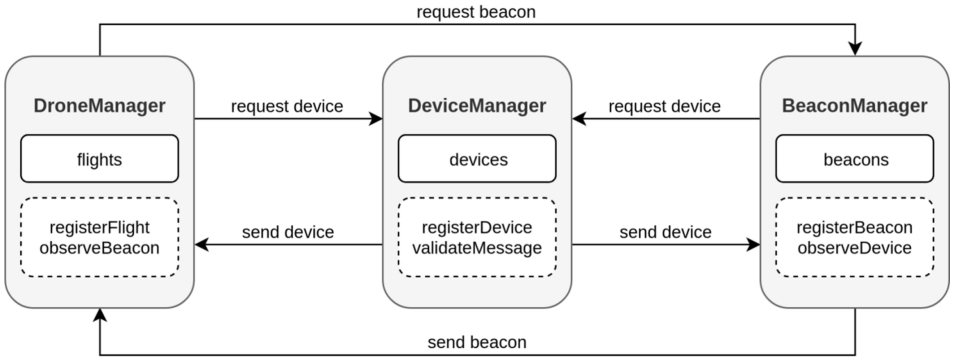


Figure 1. Communication between smart contracts

2.2 Web interface. To ease registration of devices, configuration of drones, and validation of the path, simple web interface was built. First, owner registers the device with necessary properties and retrieves configuration in the form of a file, and sets up a drone, which is then ready to fly. When the flight is stored on the blockchain, a unique identifier of the same flight, *flight_ID* is returned, which can then be entered on web interface to display a report of the flight, including the timeline, the validation form and the flight path on a map.

2.3 Drones. For a drone to operate properly and be compatible with the system, it needs to be configured and have the ability to dispatch messages after flights. Once the device is registered via DeviceManager smart contract, configuration in the form of a file that includes private key and other properties is downloaded and uploaded to the drone. Once set up, drone is ready to go on series of flights and collect data. Any drone that can run appropriate software and has working connection is able to utilize configuration and send signed messages to the receiver.

2.4 Gateway. Gateway is a server connected to the blockchain that acts as a receiver for incoming messages. Once the message is sent by a drone and arrives to the gateway, its content is first extracted into *device_ID*, payload P and signature S and then validated via appropriate DeviceManager smart contract method. In a case of a

valid signature, message payload gets processed and flight information contained in the payload is hashed and saved on the blockchain. In case of an invalid signature, whole message is discarded.

3 DRONE FLIGHT

Relation between system components and drone flight life cycle is shown in Figure 2 and is described in subsections below.

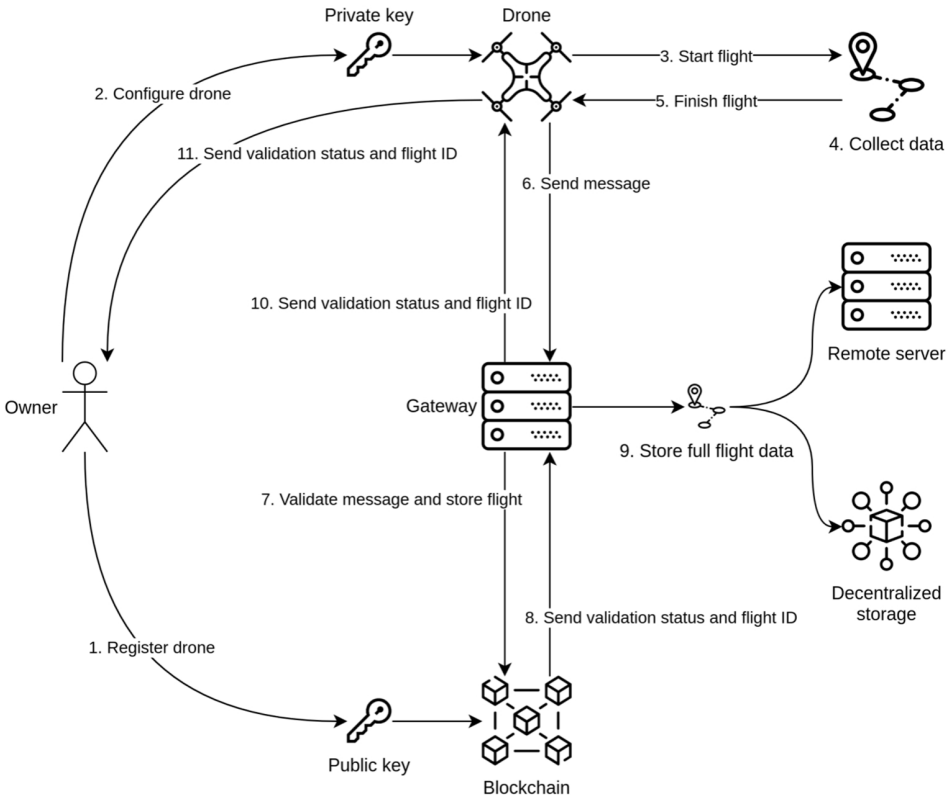


Figure 2. Drone flight lifecycle

3.1 Collecting data. After a drone is configured and ready to fly, it departs from its source location. During its flight from one location to another, it is primarily receiving coordinates from the GNSS receiver. At the end of the flight, these

coordinates collectively denote the traveled path. Each coordinate data point can contain a signature to be sure of its validity. Moreover, if a drone is equipped with additional sensors, their readings can be saved in addition to the GNSS coordinates. Once a drone arrives at the target location, its payload P containing flight data is signed using private key K_{Pr}, stored on the drone, as shown in Equation (1). The message M is then constructed from the *device_ID*, payload P and signature S, and as such sent to the gateway:

$$S = Sign(P, K_{Pr}). \quad (1)$$

3.2 Storing data. Once the message M arrives at the gateway, it is deconstructed into *device_ID*, payload P and signature S. Public key K_{Pu} is recovered from payload P and signature S, as shown in Equation (2)

$$K_{Pu} = Recover(P, S). \quad (2)$$

Blockchain is queried with the received *device_ID* to retrieve the corresponding device identifier, which is then compared with recovered public key K_{Pu}. If those two matches, message is successfully validated and can be processed further. Payload consists of flight header, flight data and beacons data. Flight header contains source and target location of the flight and departure and arrival time. Flight data is a path consisting of GNSS coordinates and optionally, sensor readings. Beacons data contains information about beacons that were encountered during the flight, such as their identifier, location and time of encounter. Due to writing to blockchain being expensive and flight data having potential to grow very big, it is hashed using a cryptographic hash function and saved as such on the blockchain together with the flight header and data on encountered beacons. Full flight data can be saved directly on the gateway, remote server or decentralized storage like InterPlanetary File System (IPFS) (Benet, 2014.) and then distributed to end users that are going to use it to validate the flight.

4 PATH VALIDATION

Step by step path validation process is shown in Figure 3 and is described in subsections below.

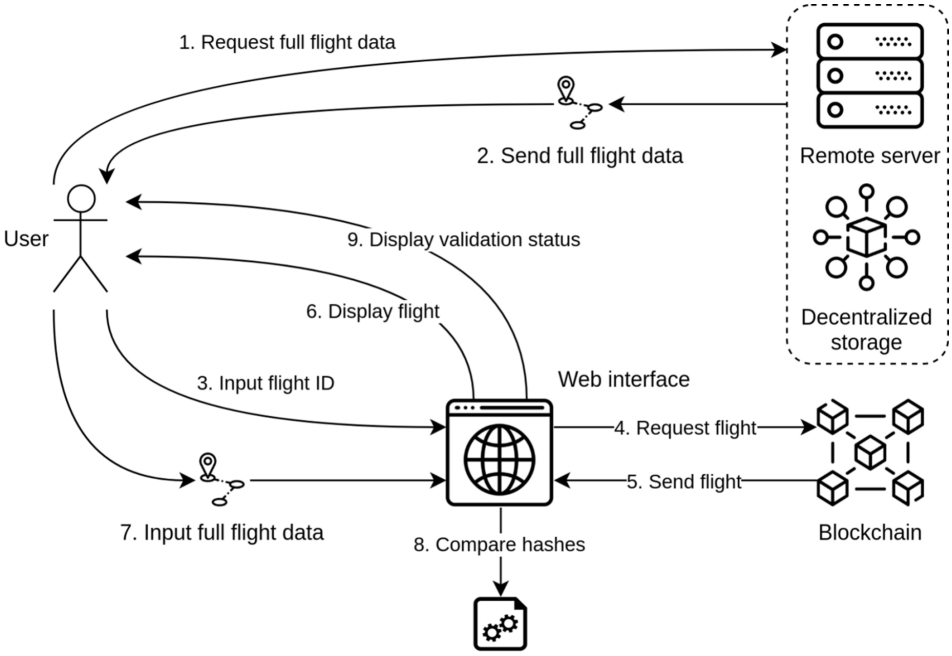


Figure 3. Validation process

4.1 Path hash. In order to properly validate the path, end user needs to have full flight data that contains coordinates through which drone traveled. Using an ID of a flight, end user can request and download full flight data from remote server or decentralized storage, depending on where it was saved at the time of flight creation.

To ease the process, web interface that connects to the blockchain is used to display flight information back to the end user. On input of a valid flight ID, information such as flight header, path hash and associated beacons is retrieved from the blockchain and displayed in form of a timeline and a map. Furthermore, once full flight data is entered, validation process is triggered and utilizes key element, path hash. Cryptographic hash function is applied on full flight data and the result is then compared with the path hash that was retrieved from the blockchain. If mentioned two hashes match, it means that drone traveled through coordinates contained in

full flight data and validation is successful. In case of even a slight tweak of coordinates, resulting hash will be different and not match the one on the blockchain, making validation unsuccessful. Figure 4 shows map with estimated path and no validation, map with a path in green after successful validation and map with a path in red after unsuccessful validation, respectively.

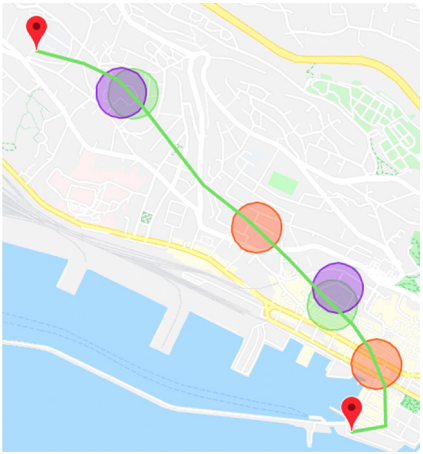
a) No path validation (b) Successful path validation

(c) Unsuccessful path validation

Figure 4. Path validation with path hash

4.2 Beacons. To further strengthen the validation, beacon data associated with a flight can be used. The existing coordinates, if any, where a drone observed the beacons as well as those where the beacons observed a drone are both displayed on the map with a disk of appropriate radius. Figure 5 shows different results of path validation. The locations where a drone observed registered beacons are colored in green, while the locations where a drone observed unregistered beacons are colored in red. Finally, the locations where beacons observed the drone are colored in purple. Overlapping green and purple circles mark those areas where both a drone and a beacon observed each other. As a result, more overlapping green and purple circles there are, the stronger the claim is that the traveled path is valid. A mechanism for allowing both storage of the path with arbitrary resolution and successful validation of a set of possible paths confirmed by beacons can be based on geohashed coordinates (Niemeyer, 2008).

a) Path not validated by beacons



(b) Path successfully validated by beacons

(c) Another path successfully validated by beacons

Figure 5. Path validation with beacons

5 CONCLUSION AND FUTURE WORK

A system for path validation of drones was proposed and described consisting of drones, smart contracts running on a blockchain and a gateway. It can be used with or without beacons to provide reliable validation of traveled path. The same system and path validation principles described here can be applied not only to drones, but also to other vehicles such as cars. In addition to the path validation function, the

information on the beacons detected during the flight (position and time) and saved on the blockchain can also be used to achieve Proof-of-Location concept. Any device can query the blockchain with identifiers of nearby beacons whose signals it received and retrieve their latest locations, which can then be used to determine current location by performing triangulation.

REFERENCES

Amoretti, M., Brambilla, G., Medioli, F. and Zanichelli, F. (2018). Blockchain-Based Proof of Location. *Proceedings 2018 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, Lisbon, Portugal, pp. 146-153. Washington, D. C: IEEE CS CPS.

Benet, J. (2014). IPFS – Content Addressed, Versioned, P2P File System. *eprint arXiv:1407.3561*. p. 11.

Christidis, K. and Devetsikiotis, M. (2016). Blockchains and Smart Contracts for the Internet of Things. *IEEE Access*, 4, pp. 2292-2303.

Gaetani, E., Aniell, L., Baldoni, B., Lombardi, F., Margher, A. and Sassone V. (2017). Blockchain-based Database to Ensure Data Integrity in Cloud Computing Environments. *Proceedings of the First Italian Conference on Cybersecurity (ITASEC17),* Venice, Italy, pp. 8-18. Aachen: CEUR.

Mykletun, E., Narasimha, M. and Tsudik, G. (2015). Providing Authentication and Integrity in Outsourced Databases using Merkle Hash Tree. *CS261: Security in Computer Systems*, p. 7. Berkeley: UC Berkeley.

Niemeyer, G. (2008). Labix Blog: Geohash. Available at: https://web.archive.org/web/20080305223755/http://blog.labix.org/, accessed 13 October 2019.

Parker, L. (2015). Brave New Coin: Timestamping On The Blockchain. Available at: https://bravenewcoin.com/insights/timestamping-on-the-blockchain, accessed 13 August 2019.

Shakhatreh, H., Sawalmeh, A., Al-Fuqaha, A., Dou, Z., Almaita, E., Khalil, I., Othman, N., Khreishah, A. and Guizani, M. (2019). Unmanned Aerial Vehicles: A Survey on Civil Applications and Key Research Challenges. *IEEE Access*. DOI: 10.1109/ACCESS.2019.2909530

Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. EIP-150 Revision. *Ethereum Project Yellow Paper*, 151, pp. 1-32.

Zheng, Z., Xie, S., Dai, S., Chen, X. and Wang, H. (2017). An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. *Proceedings of The 6th IEEE International Conference on Big Data (Big Data 2018),* Seattle, USA, pp. 557-564.

Sveučilište u Rijeci
POMORSKI FAKULTET
FACULTY OF MARITIME STUDIES
University of Rijeka

University of Zagreb
Faculty of Transport
and Traffic Sciences

Royal Institute of Navigation
Science Technology Practice

13th
Annual
Baška GNSS
Conference

DECODING AIS MESSAGES WITH THE USE OF LOW PERFORMANCE SOFTWARE DEFINED RADIO

Matej Bažec, Franc Dimc

Faculty of Maritime Studies and Transport, Pot pomorščakov 4,
6320 Portorož, Slovenia, e-mail: matej.bazec@fpp.uni-lj.si

ABSTRACT

AIS signal was captured with an SDR device using direct DSP techniques only. A computationally effective way to receive AIS signals was developed that could be implemented in embedded devices. For this purpose, the signal processing was divided in three stages. The first is computationally as simple as possible and serves only as a discriminator passing through only the samples that have a high probability of carrying an AIS message. This is done in order to reduce the load of the computationally more intensive latter stages. In stage 2 the signal is downconverted, filtered and FM demodulated. In stage 3 the proper message is acquired and checked for the CRC match. Some theoretical description as well as a concrete example will also be given.

Key words: *Automated Identification System, Software Defined Radio*

1 INTRODUCTION

Automatic Identification System (AIS) (IALA, 2011; ETSI, 2019) is a compulsory part of equipment for ships in international voyages of 300 gross tons and upwards, 500 tons and upwards for carrying cargoes not in international waters and on all types of passenger vessels. Since 2014, also all European Union fishing vessels of the length over 15 meters (LOA) have been required to be equipped with AIS.

AIS serves as a regularly selfreporting aid for conducting the telemetric ship data predominately the internal GNSS receiver's derived position, speed, course, rate of turn, but also as a source of static data as the vessel identity and its dimensions. The reported voyage data include the type of cargo on board, destination, draught, and estimated time of arrival. Beside the system's data are reliable to a certain degree, its information are also vulnerable to manipulations, which both degrades the maritime situation awareness and thus the safety at sea. The reporting interval depends upon the message type and the sea activity of the vessel.

As a ship to ship and ship to shore reporting system, AIS supports the safety of life at sea and enhances the control and monitoring of the traffic. Range, normally referred as a line of sight (IALA, 2011), of implemented VHF transmissions is typically 40-50 NM with coastal AIS base stations' receivers, but with AIS satellites in Low Earth Orbit it exceeds 600 km (Skauen and Olsen, 2016).

Today there is a large choice of commercial AIS transponders on the market available at moderate prices. An obvious question arises, what could be a possible motivation to build another AIS receiver (or even transmitter, although the latter would be more of a regulatory problem rather than technical). There were three main reasons for this decision. The first is the price. Although the commercial AIS transponders are mass produced and their prices can reach as low as €400, this is still too expensive for some applications (for instance to set up a network of low-budget AIS receivers). The second one is the commercial receivers hide the signal details (which is pretty obvious from the end-user point of view). However, the knowledge of an eventual signal anomaly could provide an indication to malfunctioning of the equipment. And finally, a potential AIS as a cloud service could be considered with such an approach.

Keeping those factors in mind the obvious choice was to build an AIS receiver based on a Software Defined Radio (SDR) device. The SDR technique is based on the acquisition of the in-phase (i) and the out-of-phase or quadrature (q) signal in a frequency band of interest. All the demodulation and signal post-processing is then

done in software using various mathematical algorithms. Since we wanted to keep the budget as low as possible, an obvious choice was to use a RTL2832U based USB dongle (Palosaari, Fry and Markraf, 2010) with TCXO (temperature compensated crystal oscillator). In order to keep the price even lower, the SDR should be possibly connected to an embedded device. This excluded the use of any sort of Gnuradio (Blossom, 2004) or Matlab based blocks such as AIS Tools for Gnuradio (Moratto, 2019), since their use would impact the computational effectiveness. On the other hand, there already exist applications such as AISmon (MT, 2014), GnuAIS (GNU AIS, 2012) and Ships (Videgro, 2019) that acquire the AIS signal from the audio demodulation in the VHF region. We found such an approach at least non-elegant and its use would probably imply a significant processing penalty. For those reasons and the lack of any C library for AIS acquisition, a decision was made to build up an SDR based AIS receiver written completely in C.

2 TECHNICAL DETAILS

An AIS message is sent over one of the two channels in the VHF band (161.975 MHz and 162.025 MHz). A transmitter can use a reserved time slot of 26.7 ms (although longer messages can use up to 5 slots). The CRC (cyclic redundancy check) of the raw data is first appended to the message. Then the data is bit-stuffed – a 0-bit is inserted after 5 consecutive 1-bits in order to help the receiver to stay synchronized. After that, a start-flag and stop-flag (a 7E hexadecimal byte) are prefixed and postponed respectively. At the end, the message is prefixed with the preamble (a 24-bit long sequence of alternating 0 and 1). The armored data is then GMSK (Gaussian minimal shift keying) modulated with a bitrate of 9600 bits per second.

In order to receive both the channels simultaneously and to get rid of the DC offset, the SDR device was tuned to 161.875 MHz with a sample rate of 2 MSps. It should be mentioned that in northern Adriatic sea (where the acquisition was taking place) the AIS slots are very sparsely populated. This means that the majority of the time the receiver acquired RF noise. So the first step was to build a rough yet computationally effective way to sieve slots with potentially present AIS messages. For this purpose, the first component of the Fourier transform of 16 consecutive samples was calculated. Keeping in mind the effects of using no window at all, a single FT component roughly corresponds to a 125 kHz wide frequency band. This means that the DC component holds the frequencies from 161.8125 MHz to 161.9375 MHz and the first component from 161.9375 MHz to 162.0625 MHz.

This covers equidistantly both AIS channels and it was the reason the device was tuned to 161.875 MHz.

It should be stressed out that a single component and not the entire Fourier transform (FT) is to be computed. This increases the computational speed significantly especially if the FT coefficients are precomputed. A further improvement can be achieved taking into account that a multiple of 2π/16 corresponds to an angle of a multiple of 22.5° so the summation order can be permuted to reflect this fact with vanishing factors omitted. The square of the amplitude (the total energy of the frequency band of interest) is then calculated. This can be abbreviated in a mathematical form:

$$
\begin{aligned}
I_1 &= \sum_{n=0}^{15} i_n e^{i\frac{2\pi}{16}n} = \\
&= x_0 + c(x_1 - x_7) + d(x_2 - x_6) + s(x_3 - x_5) + \\
&+ i(x_4 + c(x_3 + x_5) + d(x_2 + x_6) + s(x_1 + x_7)) \\
Q_1 &= \sum_{n=0}^{15} q_n e^{i\frac{2\pi}{16}n} = \\
&= y_0 + c(y_1 - y_7) + d(y_2 - y_6) + s(y_3 - y_5) + \\
&+ i(y_4 + c(y_3 + y_5) + d(y_2 + y_6) + s(y_1 + y_7)) \\
A^2 &= |I_1 + iQ_1|^2,
\end{aligned}
\tag{1}
$$

where $x_n = i_n - i_{n+8}$, $y_n = q_n - q_{n+8}$, $c = \cos(\pi/8)$, $d = 1/\sqrt{2}$ and $s = \cos(\pi/8)$. The calculation for A^2 takes totally 43 additions or subtractions and 14 multiplications for the sequence of 16 consecutive samples. That yields 2.7 additions and 0.9 multiplications per sampled pair (i and q).

If A^2 exceeds a particular threshold value, then the signal gets further processed, otherwise the samples are dropped. The threshold value should be set empirically such that it includes all the messages in the test run. It does not need to be particularly precise, since this is only an early discriminator. It is better to set it a little bit lower than the weakest accepted slot, since it is better to get some false-positive data that can be dropped later rather than missing a valid message. On the other hand, setting the value too low would unnecessarily spend the processing power in stage 2 that is computationally much more intensive. It should be also mentioned that a false-positive event could also be triggered by the spectral leakage of a signal in the vicinity of the frequency band of interest. It should be stressed that both the AIS channels are unoccupied most of the time (in our experience at least 80 %, even 90 %). This means that most of the received signal will be dropped at this point, consuming almost no computational power. This should be compared to conventional AIS SDR receivers

that perform FM demodulation (in reality they are doing it twice – once for each channel) continuously which is very power consuming. Although a quantitative measurement of CPU consumption has not been done yet, it is evident that such an approach presents a significant improvement.

3 LATTER STAGE POSTPROCESSING

Once the acquired signal passes stage 1, it gets downconverted twice (for each channel once). The downconverted signal then gets filtered in order to get rid of all the signals out of the particular AIS channel frequency band. A more detailed mathematical description of the above procedure can be reduced to these steps. First, a sequence of 2048 consecutive samples (both in i and q) is joined in a single complex sequence

$$a_n = i_n + iq_n \tag{2}$$

and processed with a Hann window to prevent spectral leakage

$$h_n = a_n \frac{1-\cos\left(\frac{2\pi n}{2048}\right)}{2}. \tag{3}$$

Then, the fast Fourier transform is applied as

$$H_k = \sum_{n=0}^{2047} h_n e^{i\frac{2\pi}{2048}nk} \tag{4}$$

In frequency space, the downconversion is trivial: the signal for channel A is displaced by 102 indices (frequency components – representing 161.975 MHz) and for channel B by 154 (162.025 MHz). In order to apply the filter, all the values outside the band of interest are ignored (set to 0). To cover all the 25 kHz range, beside the central frequency, also a side band of twice the 12 indices (in each direction) is included. In order to further speed up the procedure, the decimation to a sequence of 128 samples (higher frequency components are cut off) is performed before the inverse Fourier transform takes place. Higher decimation could result in failure to use the differential formula for the FM demodulation.

As such, the signal enters stage 2 discriminator. Once again, the amplitude square (of the modified signal) is calculated and only if the value exceeds some threshold it is considered for further processing, otherwise it is dropped. Once again the reasoning above applies. Although a value close to the threshold might indicate the

presence of a weak AIS signal, it is pointless to try to decode it, since the noise would probably prevent a successful read.

If the downconverted and filtered signal passes the second stage discriminator, it enters the FM demodulator. If the changes in the signal are small (this can be safely assumed, since the signal is filtered to hold only low frequencies), then the infinitesimal formula can be used:

$$f_n = \frac{q_n i_{n-1} - q_{n-1} i_n}{i_n^2 + q_n^2}. \tag{5}$$

The demodulated signal then gets NRZI (no return to zero) decoded. It should be stressed out that even with a TCXO the tuned frequency may significantly vary from the nominal one (few kHz). This means that the central frequency of the received signal does not need to be exactly at 0. For this reason, the mean in the AIS message preamble is used to set the offset.

After the message is NRZI decoded the flags are sought. If the flags are at the right place, then the message gets cleaned from stuffed 0-bits and finally, a CRC check is calculated. If the two match, it is considered a valid AIS message.

4 RESULTS

AIS channels were monitored on 03/28/19 around 1 pm. The raw signal was stored on the disk and postprocessed later. Due to the large amount of data produced, the acquisition was limited to 1 minute. In the mean time, 43 AIS messages were captured. This number is not very informative due to the lack of an additional reference receiver. However, we have access to a monitoring station that is situated approximately 9 km from the acquisition place. In the same time it spotted 148 events, although those numbers should not be compared directly.

There is a difference in line of sight, antenna, amplifiers, etc. Figure 1 shows an example of a successful stage 1 event detection, where a good candidate for AIS message can be seen at the beginning of the chunk (a chunk is equivalent to 0.13 ms).

The discriminating function shows a significant increase so the acquired signal is a good candidate for further processing.

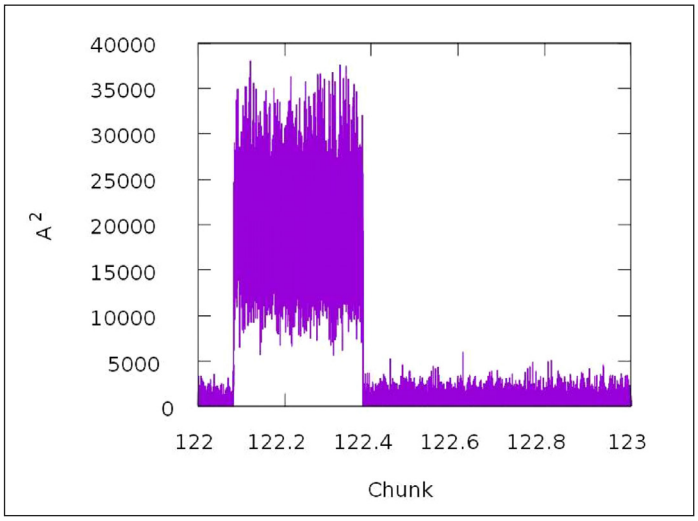


Figure 1. First stage discriminator

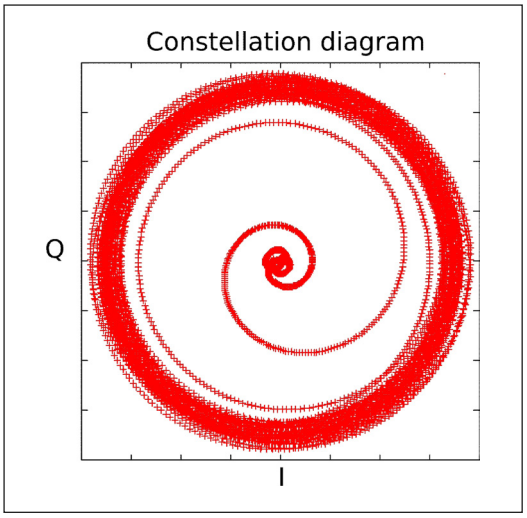


Figure 2. Constellation diagram of the AIS message during ramp-up

After entering the stage 2 processing block the signal is downconverted and filtered and its amplitude calculated. Figure 2 shows the constellation diagram of the modified signal at the current stage during the ramp-up stage of AIS transmission. When the transmission is absent the point fluctuates around the coordinate system origin in a

random fashion. After the transmission enters the ramp up phase, the point gradually moves to an orbit with almost constant radius. Once the distance from the origin exceeds the threshold value, the signal can enter the FM demodulation stage.

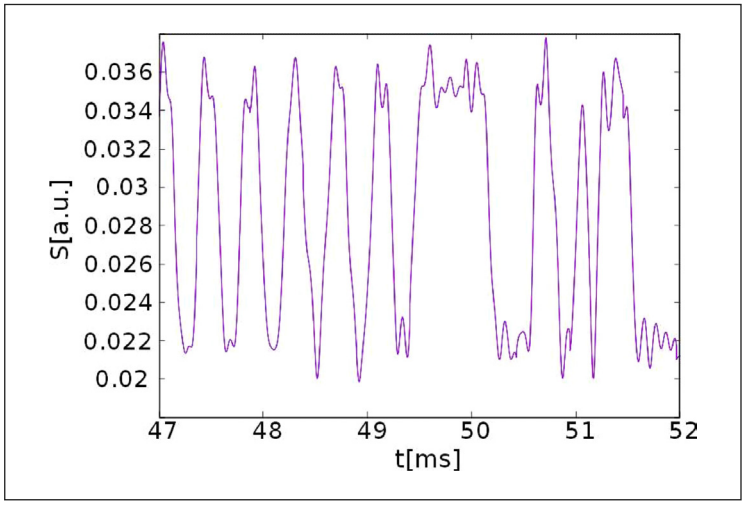


Figure 3. FM demodulated AIS signal

After the signal is FM demodulated, the AIS message can be read. Figure 3 shows the beginning of the AIS message: the preamble, start flag and some bits of the actual data. The message starts with the preamble followed by the start flag. Actual data begins around 50.5 ms.

The preamble is also used to calibrate the offset. After the flags are spotted, stuffed bits removed and CRC matches the transmitted value the message can be considered a valid AIS message reception.

As an optional step, the message was encoded in the way it could be used directly into the NMEA sentence body. In this particular case the output was:

```
{START_FLAG@50.2ms}
EvjO‘E2RqJrt@30a7h22V60h;4b0OO;o=4F8010888N0053lR@@
{STOP_FLAG@84.9ms}
```

This particular message is of type 21 (ATON – aids to navigation report).

5 CONCLUSION

It has been shown that a completely SDR based AIS receiver can be in principle constructed. However, at this moment the process is not fully automated so this is considered an obvious continuation step. After that, the receiver should be able to acquire AIS messages in real time. Since the algorithms used are optimized for a maximal computational efficiency, it can be safely assumed that the postprocessing can be done even on some embedded devices with lesser computational power (Raspberry PI or even some of the high-end MCUs, for instance). Although this is yet to be tested, according to our experience so far, we firmly believe that this is possible. In this case, the price for the complete receiver would be well below € 100 and it would allow to set up an affordable network of dispersed AIS receivers.

REFERENCES

Blossom, E. (2004). GNU Radio: Tools for Exploring the Radio Frequency Spectrum. *Linux Journal*, 122, pp. 4. Available at: http://dl.acm.org/citation.cfm?id=993247.993251, accessed 7 November 2019.

European Telecommunications Stanards Institute (ETSI) (2019). *EN 303 098, Maritime low power personal locating devices employing AIS; Harmonised Standard for access to radio spectrum*, Version 2.2.1, p. 50. Valbonne: ETSI.

GNU AIS. (2012). Automatic Identification System for Linux *[open source software]*. Available at: http://gnuais.sourceforge.net/, accessed 7 November 2019.

International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA). (2011). *IALA Guideline 1082: An overwiev of AIS*, Edition 1.0, p. 44. St Germain en Laye: IALA.

Marine Traffic (MT). (2014). AISMon *[freeware]*. Available at: https://help.marinetraffic.com/hc/en-us/articles/205339707-AISMon, accessed 7 November 2019.

Moratto Z. (2012). GNU Radios tools for viewing AIS *[open source software]*. Available at: https://github.com/zmoratto/ais-tools, accessed 7 November 2019.

Palosaari, A., Fry, E. and Markraf, S. (2010). RTL-SDR with Realtek 2832U tuners *[open source software]*. Available at: https://www.rtl-sdr.com/, accessed 7 November 2019.

Skauen A. N. and Olsen, Ø. (2016). Signal environment mapping of the Automatic Identification System frequencies from space. *Advances in Space Research*, 5, pp. 725-741.

Videgro Ships. (2019). Receive AIS data from air *[open source software]*. Available at: https://github.com/videgro/Ships, accessed 7 November 2019.

Sveučilište u Rijeci
POMORSKI FAKULTET
FACULTY OF MARITIME STUDIES
University of Rijeka

University of Zagreb
Faculty of Transport
and Traffic Sciences

Royal Institute of Navigation
Science Technology Practice

13th
Annual
Baška GNSS
Conference

CHALLENGES OF ADAPTIVE COASTAL VOYAGE PLANNING

Davor Šakan[1], Srđan Žuškin[1], Marko Valčić[2], Duško Pavletić[2]

[1] University of Rijeka, Faculty of Maritime Studies, Rijeka, Croatia

[2] University of Rijeka, Faculty of Engineering, Rijeka, Croatia,
e-mail: sakan@pfri.hr

ABSTRACT

Creating an efficient and safe voyage plan is a complex and challenging task. Compliance with safe mandatory voyage planning procedures must be considered alongside shortest distance, sailing time and efficiency. Mentioned factors reflect on decisions in voyage appraisal, planning, execution and monitoring. Decisions in all voyage stages are further influenced by navigational knowledge and experience of those involved in the actual planning. Subjective interpretation, quality of used sources and uncertainties can result in different voyage plan outcomes. Research in weather routing and quality and assessment of hydrographic data have enabled multi-criteria and multi-objective voyage planning approaches. Usage of Electronic Chart Display and Information System (ECDIS) voyage planning features and advanced third-party solutions (including improved GNSS solutions) have improved the voyage planning process on-board. For a safe voyage planning, determination of safety distances to non-navigable areas or dangers is essential. Furthermore, uncertainty must be assessed for all elements of voyage planning. Moreover, the route has to be adaptable to forecasted or present environmental conditions.

The aim of this paper is to present the concept of safety distance determination as a function of the adaptive planning process in coastal navigation. The concept integrates all relevant parameters influencing the decisions of the navigational planning task in an adaptive way, with the final output subject to pre-defined limit values. Ship particulars, hydrographic data accuracy and reliability, hydrographical and meteorological conditions, etc. set the basis

for the safety distance determination. The concept is further enhanced with information on current and forecasted environmental conditions prevailing in the area of interest.

Key words: *maritime navigation, adaptive voyage planning, safety distance determination*

1 INTRODUCTION

Creating a safe and efficient voyage plan is a complex and challenging process. Procedures describing elements of planning are based on official resolutions and documents. These documents describe items contributing to safe and effective voyage planning. Adherence to the procedures should create a similar voyage plan, regardless of the person who plans. However, voyage planning is subjective and different voyage plan outcomes arise. There are slight variations in personal interpretation of voyage planning procedures and elements. More, Officers' and Masters' voyage planning skills depend on education, training and experience. Technological and organizational advancements have simplified and changed parts of voyage planning phases. Global Navigation Satellite Systems (GNSS) have simplified position fixing with unprecedented positional accuracies (Šakan et al., 2019), both for navigation and surveying. Electronic Chart Display and Information System (ECDIS), which uses GNSS as primary position source, simplified voyage planning. Conversely, other challenges emerged such as single position source over reliance, discrepancy between positional accuracy and quality of hydrographic data sources. Furthermore, focus on voyage optimization objectives is increasing due to regulations compliance and availability of vessel monitoring data.

Procedures should simplify the planning and ensure that is carried out accordingly. Detailed voyage planning procedures are usually described in company's Safety Management System (SMS) manual. Such procedures include specifics of the vessel, trade and other relevant considerations. Official IMO resolution: Guidelines for voyage planning A.893 (21), describes voyage planning and important objectives (IMO, 2000). Essential objectives are safety of life, safe and efficient navigation and environmental protection.

There are several voyage planning stages: appraisal, planning, execution and monitoring. For each stage details about vessel, equipment, crew or environment must be considered. In appraisal stage, dedicated Officer usually gathers and prepares

relevant information. The sources are navigational charts, publications and even personal experiences, if any. Master supervises appraisal, changes and approves the voyage. In planning phase, a designated Officer evaluates variations of the proposed route. They are created in ECDIS and/or on paper charts. After departure, the execution stage begins. Factors such as arrival times, weather forecast, or expected maritime traffic must be considered. Voyage continues in monitoring phase until arrival. The Officer will change the plan if necessary, adapting to the objectives of a safe and effective voyage.

Vessel navigates in several distinctive phases: ocean, coastal and restricted waters phase (IALA, 2018). The phase depends on the navigational area in which ship sails. Ocean phase is usually beyond the continental shelf. Depths are over 200 m and distance from shore is 50 or more nautical miles (NM). Coastal phase refers to areas less than 50 NM from shore or are in the limit of the continental shelf with depths less than 200 m. Restricted water phase can occur in coastal phase and in straits.

For every voyage plan, there are several general objectives. Prime objective is the safety of the vessel and all on board. A safe voyage plan will reduce risks and improve safety of navigation. Additionally, the navigator considers shortest distance, efficiency and voyage optimization objectives. The importance of voyage optimization is increasing. An effective voyage plan will result in reduced costs and damages on both vessel and cargo thus increasing savings. Furthermore, reductions of harmful emissions from ships and climate changes are all part of present and future transport policies.

To conform to general objectives is not an easy task even for the experienced Officers and Masters. Weather, maritime traffic and activities, hydrographic constraints and uncertainties influence the voyage plan objectives. How much they influence the voyage plan, depends on the navigational phase. To adapt and optimize frequently, in the execution and monitoring stage is even more challenging. Considering navigation phases, it is simpler to adapt the voyage plan in less constrained ocean phase. Consequently, weather routing optimization is used for decades on ocean passages. However, such solutions have not been extensively researched and developed for other more constrained stages. This was because of perceived limited or non-existent benefit, complexity and low-resolution weather models. Shorter distances between waypoints, manoeuvring limitations or coastal areas restrictions decrease the number of alternative routes. Traffic intensity is greater and collision avoidance occurs more in coastal voyage phase, adding small

route deviations. However, accuracy and quality of hydrographic data, weather forecasting methods, models and information sharing are increasing. This will put more emphasis on more efficient solutions for adaptive voyage planning in all voyage navigation stages.

Research and application of coastal voyage planning solutions is rather limited (Takashima et al., 2009; Grifoll et al., 2018; Vettor et al., 2016). Moreover, the research was focused on certain elements of voyage planning, routing or optimization, as it will be presented in following sections. The aim of this paper is to evaluate challenges for adaptive, multi-objective and multi-criteria route planning, with focus on coastal navigational phase. Major research areas are assessed related to voyage planning solutions. Finally, safety distance usage and possible application is analysed.

The structure of the paper is as follows. In Section 2, we evaluate challenges of coastal voyage planning. In Section 3, safety distance is presented. In Section 4, we discuss the findings and results. Finally, we conclude the paper with suggestions for future research in Section 5.

2 COASTAL VOYAGE PLANNING CHALLENGES

2.1 Route and path planning. Many path planning methods have been researched and developed for marine surface vehicles (Singh et al., 2018), collision avoidance (Lazarowska, 2015) and ship routing (Simonsen et al., 2015). Methods and algorithms used can solve a single objective such as shortest distance (Dramski, 2011) or multiple objectives such as path and speed (Lee et al., 2018). There are several common algorithms used: Dijkstra's, A* (A-star), Genetic algorithm, Theta algorithm and Route Binary Tree algorithm (Jia et al., 2019). In the following section, we present Dijkstra's and A* algorithms.

One of the most common algorithms used is Dijkstra's algorithm (Dijkstra, 1959) published in 1959 (Dramski and Mąka, 2013) presented in Figure 1. It is a non-heuristic algorithm. It calculates the shortest path from initial (A) to goal vertex (G). First, two sets of vertices are created: empty visited set and unvisited set with all other vertices. Starting vertex (A) value is set to 0, while values for the other vertices are set to infinity. The edge weight is the cost which can represent distance, required passage time or fuel. It has non-negative value.

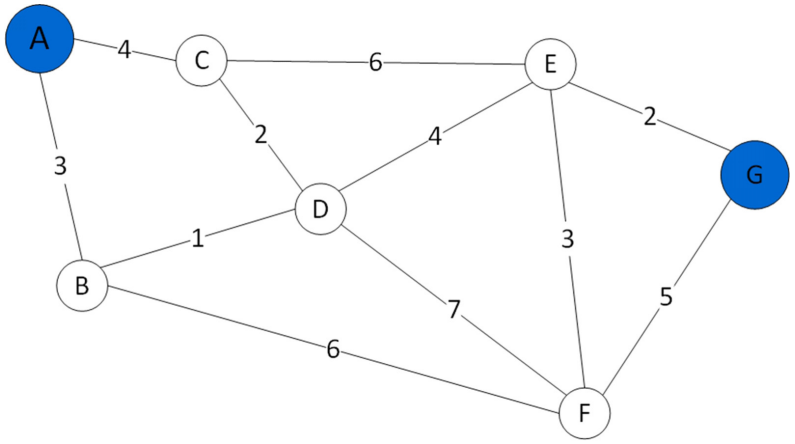


Figure 1. Representation of Dijkstra's algorithm (Dijkstra, 1959)**.**

Made by authors

When started, algorithm visits adjacent vertices from the starting vertex (A) beginning from vertex with minimum weight. Then it visits the remaining unvisited vertices calculating the cost from starting vertex for each adjoining vertex.

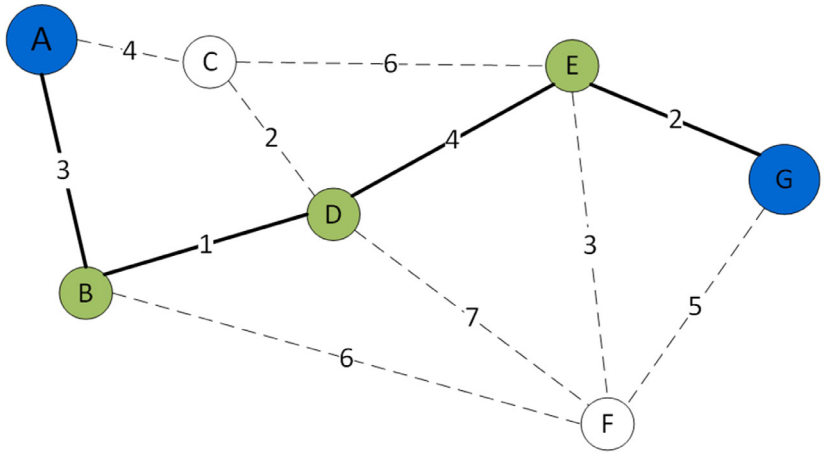


Figure 2. Representation of determined Dijkstra's shortest path (Dijkstra, 1959)**.**

Made by authors

If the cost value is less than the previous known value, the cost is updated. Visited starting vertex (A) is then added to the visited set. The algorithm proceeds to the next vertex with minimum cost (B), from the start vertex. The process repeats for the adjacent vertices (D and F). The cost from the start vertex (A) is determined and updated if less than the previous cost. Vertex (B) is added to the visited set. Process continues until there are no more unvisited vertices, and minimum cost is calculated as it is presented in Figure 2.

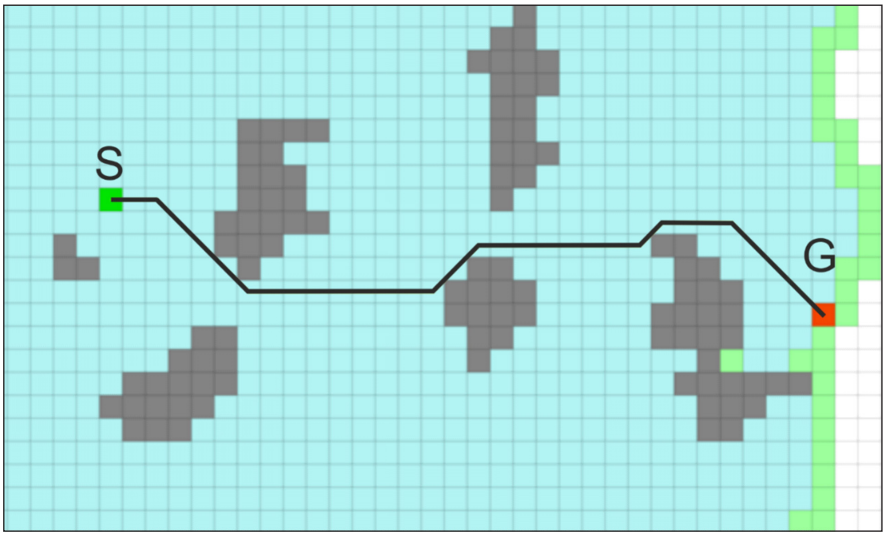


Figure 3. Representation of determined Dijkstra's shortest path between obstacles.

Made with PathFinding.js (Xu, 2012). Adapted by authors.

A* algorithm (Hart et al., 1968) is a best-first search heuristic algorithm. We can consider it as an extended Dijkstra's algorithm (Kim et al., 2014). By minimizing $f(n) = g(n) + h(n)$, the shortest path is selected. Element $g(n)$ is the cost from the source vertex. Heuristic function $h(n)$ estimates the cost to the goal vertex using values such as Euclidean distance or transit time (Dramski, 2011).

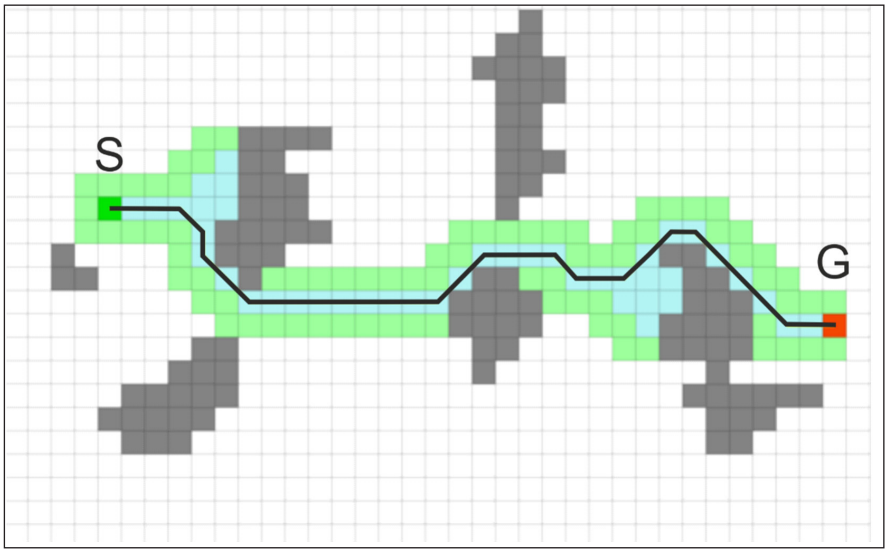
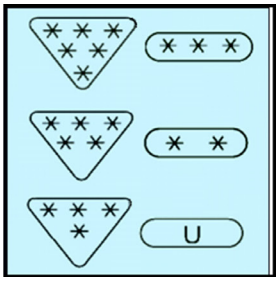


Figure 4. Representation of determined A* shortest path between obstacles.

Made with PathFinding.js (Xu, 2012). Adapted by authors.

As it can be seen in Figures 3 and 4, the number of visited vertices is reduced when using A*, thus resulting in faster computation time (Grifoll and Castells, 2016).

2.2 Quality and* accuracy *of hydrographic data. One of the challenges of voyage planning and particularly of coastal and restricted phases is assessing the quality of data. Electronic Chart Display and Information System (ECDIS) uses Official Electronic Navigational Chart (ENC). ENC is a database created from official hydrographic data. Relevant regulations and standards regulate specifications of the ENC (Žuškin et al., 2017). Hydrographic data can be from various sources, historical periods and made with different survey methods. ENC attribute "Category of Zone of Confidence in Data" (CATZOC) represents the quality of data. When selected for display, star symbols (*) represent CATZOC categories, ranging from the highest six to the lowest two stars. Categories range from highest A1, A2, B, C and lowest D. Unassessed data are presented and labelled with "U". IHO publication S-67 represents current status of data quality for worldwide coverage areas (IHO, 2017) presented in Figure 5. Assessment criteria include survey characteristics, seafloor coverage, position accuracy and depth accuracy. IHO S-67 states seafloor coverage as most important feature regarding minimum ship clearance from the keel to the seabed. Position accuracy of chart features also results in uncertainness, requiring larger margin of ship passing distance is as in (IHO, 2017, p. 15).



Category	% area of world's coastal ENC	Confidence
A1 (6 stars)	0.7%	Very Good
A2 (5 stars)	1.0%	Very Good
B (4 stars)	30.5%	Good
C (3 stars)	21.8%	Fair
D (2 stars)	20.5%	Low
Unassessed (U)	25.4%	Low

Figure 5. CATZOC categories and accuracy categories of world's coastal areas (IHO, 2017)

Although surveys in coastal areas are more frequent, still the highest accuracy categories are not featured extensively.

When planning, the navigator assesses the quality of data, survey methods used and related accuracies. These values are used as one of the safety settings in ECDIS (Žuškin et al., 2016). Assessment, transfer and representation of surveyed data can be found in (Kang et al., 2015). Methods for ENC compilation such as Delaunay Algorithm (Kang et al., 2014) of ENC data have been proposed for route planning (Kang et al., 2015). Further evaluation and suggested risk-based hydrographic uncertainty model can be found in (Calder, 2015).

2.3 ECDIS safety settings setup. We presented challenges arising from shortest path objective and interpretation of ENC data. When planning, the navigator must appropriately enter safety settings in ECDIS. The settings should be adapted throughout the whole voyage, thus corresponding to environmental and other voyage related changes. Basic settings available for representation of safe and unsafe waters are: safety contour, safety depth and deep and shallow area indication. Other functions and tools improving navigational awareness exist. Anti-grounding or look-ahead, safety frame functions are available; however, we present only basic settings.

Safety depth and safety contour can be set either as separated values or as a single value depending on the implementation of the ECDIS manufactures. Safety contour represents a visual delineation between safe and unsafe water. Its calculated value is the basis for crossing of safety contour alarm. Methods for safety contour and safety depth calculation are not established internationally and are set by ship company recommendations or theoretical navigational background (Žuškin et al., 2016). As authors suggest in (Žuškin et al., 2016), values for safety depth can be c alculated as:

$$Safety\ depth = T + d_{density} + d_{heel} + d_{squat} + d_{SM} + d_{TIDE} + d_{ZOC} \quad (1)$$

where T is vessel's static draft, $d_{density}$ is draft correction for change of water density, d_{heel} is draft correction for potential heeling angle, d_{squat} is draft correction due to ship squat calculation, d_{SM} is correction for safety margin, d_{TIDE} is correction for tidal heights, and d_{ZOC} is correction for Zone of Confidence. However, different calculation for safety depth and contour can be found in (Rutkowski, 2018):

$$SD = T_{max} + R_{UKC} + R_{squat} + R_d - H_{tide} \tag{2}$$

$$SC = SD + CATZOC \tag{3}$$

where SD is safety depth [m], SC is safety contour [m], T_{max} is ship's draught [m], R_{UKC} is required Under Keel Clearance [m], R_{squat} is estimated squat [m], R_d is dynamic reserve caused by ship's seakeeping characteristics related to roll and pitch [m], H_{tide} is tide height above chart datum [m], and $CATZOC$ is Category of Zone of Confidence [m].

The shortcoming of safety contour is that calculated value does not have to respond to available depth contours on ENC. Then, the next higher available contour becomes and is depicted as a safety contour. The problem is known and recognized by the International Hydrographic Organization (IHO). Solution is the usage of high-resolution bathymetric data. Standards and solutions for high-resolution bathymetry data ENCs are considered in ("ENCWG," 2018). In addition, future hydrographic standards such as IHO S-100 (IHO, 2018) will be adapted to current and fore-coming navigational needs.

The navigator sets up safety values for checking of voyage plan. If necessary and before departure, he updates values for execution and monitoring phase. The values can be set up either for the whole voyage or for each leg of the route separately. If we consider distances between legs, there is a possibility that values should change more often, following the quality and resolution of data changes. For future adaptive planning, execution and monitoring, safety margins should be more precise reflecting hydrographic and vessel constraints. Contrary, today's worst-case minimum safety settings are overestimated but safe. In such case, the values for future systems could be too restrictive for certain navigational areas.

Contemporary ECDIS usage and misunderstanding of safety settings setup has resulted in several accidents and groundings in coastal waters. To find the causes of misinterpretation of ECDIS data and settings, a study started by United Kingdom and Danish national maritime accident investigation boards (MAIB, 2017).

2.4 Coastal weather routing systems. Weather routing is the creation of an optimal route based on weather forecasting and sea state considering vessel's characteristics. Although designed for ocean transits, solutions for coastal passages are also emerging. The most popular general weather routing approach is based on isochrones, single-criterion method first proposed by James in 1957. Since inception, it has been extensively developed. Isochrones represent time-fronts that are lines or envelopes. They are made of positions reachable by the vessel after departure under certain weather conditions. First isochrone is created within a defined time limit. In the same way, next isochrones are created until destination. Then, by backtracking, the optimal route with a minimal passage time is created. Other common methods include isopones (energy fronts), calculus of variation or 3D programming. Weather routing methods were primary single-objective considering passage time optimization (Szlapczynska, 2015). Recently, multi-objective and multi-criteria approaches based on evolutionary algorithms have been developed and used (Walther et al., 2016). An overview of available voyage optimization and weather routing services can be found in (Lu et al., 2015), while structure of a weather routing systems and algorithms used can be found in (Simonsen et al., 2015).

Usage of coastal weather routing so far has not been considered extensively. In 2009, Minimum Fuel Route consumption (MFR) method based on Dijkstra's algorithm (Takashima et al., 2009) for fuel saving operation was considered for coastal vessels. The MFR simulation was carried out for a RO/RO and a cement-carrier vessel with routes between Japanese coastal ports. Data used for the simulation included forecasted data of surface winds, waves, ocean and tidal currents. The result confirmed fuel savings with usage of MFR method.

Recently, coastal weather routing was also considered in (Vettor et al., 2016) for a 29 m trawler from Portugal to Norway based on different selection criteria. A similar VISIR-I model and system were considered for coastal navigation of smaller vessels in the Mediterranean Sea. Shortcomings of current weather routing literature and implementation for coastal areas were stated in (Mannarini, 2016). Correspondingly, the benefits of voyage optimization for short sea shipping in the European Union (EU) have also been considered (Grifoll Colls et al., 2016). In the presented paper ship, routing algorithm was implemented for the route between Barcelona and Palma de Mallorca with a length of 132 NM. The authors used Dijkstra's and A* algorithm and high-resolution wind and wave models. They established the horizontal edge resolution in 16 edges per node. Both algorithms gave the same optimal path; however, A* had significantly lower computational time.

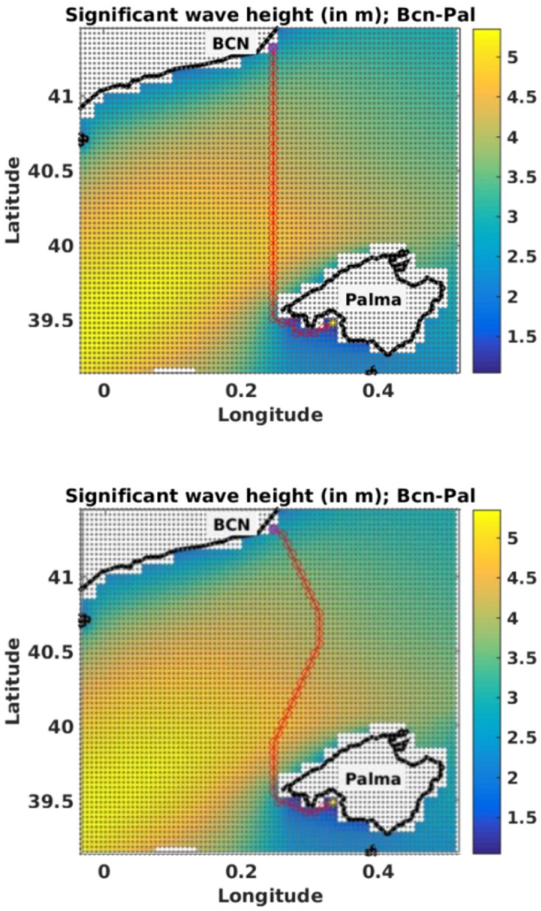


Figure 6. Optimum path without and with wave resistance (Grifoll Colls et al., 2016)

Wave height and direction relative to the vessel's course influenced the optimal path solution. The difference between passage times for the shortest and optimal path is larger or smaller depending on wave direction. However, the optimal path solution resulted in the shortest passage times for both calculated cases. One of the observed shortcomings for the coastal weather routing was spatial resolution of the meteorological and oceanographic predictions. An extended analysis and feasibility of the concept and algorithm termed SIMROUTEv2 with several ports and routes, was presented in (Basiana Ribera et al., 2017). The savings and values of the proposed routing were greater when the sea state was moderate, rough and high. Furthermore, it also depended on the width of the wave field.

2.5 Ship safety domain. We have presented challenges to adaptive coastal planning related to hydrographic conditions, quality and interpretation of data and weather routing during planning and while underway. Finally, we must consider avoidance of other vessels or objects along the vessel's planned route. Maritime traffic and risk of collision is greater in coastal phase of the voyage. Every collision avoidance will cause route deviation thus increasing total distance travelled. Although the deviations are relatively small, they influence the efficiency objective, both on vessels engaged in trans-oceanic passages or in short sea shipping.

To evaluate navigational safety of the vessel and extending water area, the concept of ship domain was created. Introduced by (Fujii and Tanaka, 1971), it has been developed extensively. Although used for waterway capacity analysis, collision risk, near-miss detection, its primary purpose is collision avoidance (Szlapczynski et al., 2018). Definitions, domain dimensions, criteria vary, thus increasing the complexity of the concept. Domain determination methods used can be empirical, theoretical analyses or based on experts' knowledge (Szlapczynski and Szlapczynska, 2017). Broadly, they can be described as a circle, ellipse and polygon ship domains (Wang et al., 2009). Besides domain parameters and coefficients used, the size of the domain will depend on the available manoeuvring sea area (Wielgosz, 2017). The ship domain or similar safety concepts must be included in future adaptive routing systems. (Pietrzykowski and Uriasz, 2010; Tsou et al., 2010).

3 SAFETY DISTANCE

We presented ship domain in previous section. Domain is a term encompassing various area representations, determined safe for the vessel. We can base domain size and shape on different criteria. Moreover, it can be considered as a generalization of safety distance, since the safety distance is not same in all directions. The term and measurement of safety distance is used in waterway capacity analysis despite a recent increase of safety domain usage. Safety distance and correlated values such as Closest Point of Approach (CPA) and Time to Closest Point of Approach (TCPA) are also used. They are preferred in real-time collision avoidance. This is because of their usage on shipboard systems such as Automatic Radar Plotting Aid (ARPA) radar.

Compared to the safety domain, one of the major advantages of safety distance concept is simplicity. It is simpler for interpretation and decision making, implementation and computational time (Szlapczynski and Szlapczynska, 2017). Safety distance can

represent distance considered safe from the vessel to nearest navigational danger (Weintrit and Kopacz, 2004) or depth contour. On the other hand, it can be a safe passing distance from coastal objects such as rocks and reefs (Gao et al., 2017). Moreover, safety distance has been used correlated with determination of realistic ship turns (Ari et al., 2013). Furthermore, it has been used in adaptation of constrained A* algorithm used for an optimal path of an unmanned surface vehicle. The distance defines a circular boundary, which is a constraint for generation of waypoints (Singh et al., 2018).

We presented dissimilarities of calculations and approaches for safety values, safety depth and contour. The anti-grounding alarms are based on safety contour and distance from navigational dangers. In addition, the base values for activation of the alarms are fixed for the whole voyage or for each route leg. They are static values, and during the voyage, the values and settings should be changed manually.

So, based on identified findings in previous sections, we formulate the safety distance for coastal and other navigational phases applicable to adaptive route planning. The proposed safety distance value changes with quality and accuracy of hydrographic data, vessel characteristics and movement. We assume that vessel's draft safety settings are set appropriately for the whole voyage.

For route planning and shortest path determination, proposed safety distance would add additional constraints. In the appraisal and planning process the shortest path would be adapted if necessary. While underway, the value would change depending on the vessel motion. When the vessel is making a turn or heads to navigational danger safety distance would change. It would increase for the value of vessel's characteristics, such as turning circle or stopping distance. Safety distance would also increase or decrease with the probability of crossing or grounding. It is based on realistic manoeuvring characteristics of the vessel–not only on draft, beam, length and approximated safe passing distance values from hydrographic objects. Therefore, we do not base respective safety distance value for alarm activation only on simplified passing or crossing distance from navigational danger, area or contour. By incorporating CATZOC values and vessel characteristics, minimal horizontal safety values created are not determined subjectively.

In the following example presented in Figure 7, vessel is underway in a coastal area with low CATZOC rating D (two stars **). This value is worse than ± 500 m of positional accuracy and worse than 2 m of depth accuracy (IHO, 2017). The shortest route is adapted for the required value. If the vessel's manoeuvring characteristics require an even larger safety distance, the waypoint position or cross-track distance is adjusted.

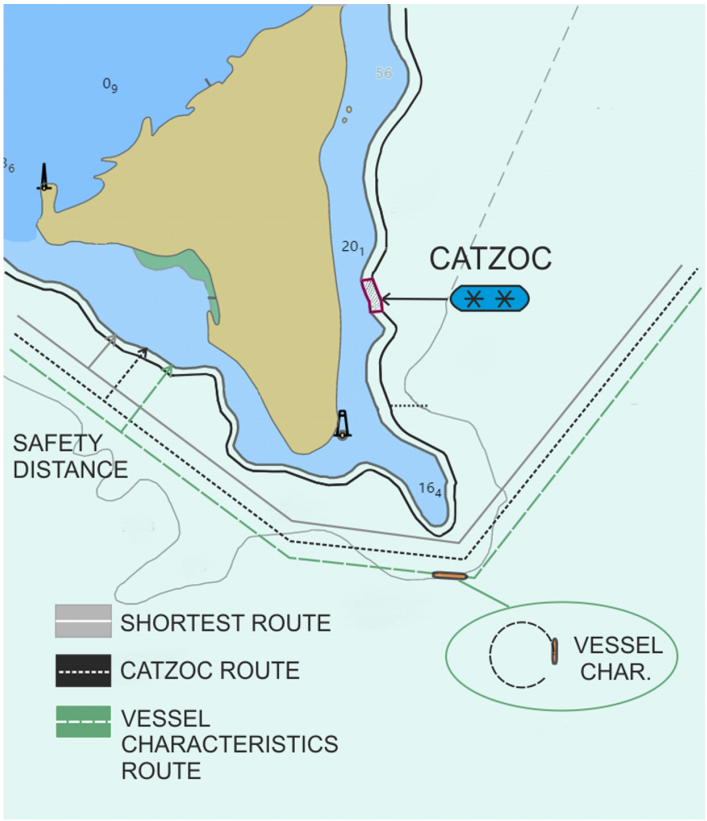


Figure 7. Safety distance.

Made by authors

The same principle can be extended to forecasted weather data. If necessary, the required value of corresponding safety distance would increase or decrease depending on weather conditions. Accordingly, additional waypoints could be created to avoid higher waves or winds. With every weather or navigational conditions change, the route would be adapted and possibly optimized.

The precedence of objectives would change depending on the prevailing conditions and navigational phase. In coastal and restricted phases, shortest path would be adapted based on hydrographic obstacle avoidance and quality. It would be further adapted in restricted navigational phase if the vessel manoeuvring constraints do not allow required course alterations, passing distances and manoeuvres. In unrestricted coastal areas, the route would be adapted based on forecasted data.

4 DISCUSSION

The voyage planning process is a challenging task, as it was presented in previous sections. However, present and near future challenges require a next step in voyage planning solutions. It is a great task to quantify the knowledge developed throughout centuries of practical navigation and skills of the past and present-day navigators. To model the environment, vessel and human factors and their interactions in a complete solution is even greater. Each navigational phase has its own set of factors influencing the possibility of adaptation. In the most unconstrained ocean phase, routing solutions have been developed and used for decades. They started from single-objective and evolved towards multi-criteria or multi-objective approaches. This will extend to the coastal phase as we presented in previous sections.

When we discuss coastal multi-criteria approach, besides previously stated objectives and criteria, we did not elaborate extensively navigational risk. However, it was elaborated in recent research on vessel (Jeong et al., 2018) and Autonomous Surface Vehicle (ASV) (Jeong et al., 2019) route planning. The authors represented risk as a gradient value. Base for risk gradient is a probabilistic analysis of maritime traffic accidents, with the use of near-miss data along a route. Beside standard hydrographic data, risk contour is created representing the risk gradient graphically. Static factors include hydrographic data, thus excluding other vessels. Other objectives include safety, efficiency and convenience (Jeong et al., 2019). In (Jeong et al., 2018), multi-criteria route planning is considered for voyage appraisal and planning phases only. The authors stated that method and data presented are for a limited coastal area. They recommended the need for validation in other regions with various conditions. Furthermore, they observed that most research and commercial routing solutions focused on efficiency. Solutions for coastal weather routing are available; however, the data resolution and scope require further evaluation.

To develop adaptive routing optimization solutions for the coastal phase, high-resolution of hydrographic and weather data is required. High-Resolution Rapid Refresh (HRRR) (Benjamin et al., 2015) forecasting and improved global and regional models running every hour will increase adaptability options for coastal routing (Masters, 2019).

It is important to state that the we did not evaluate extensively collision avoidance in this research. Any complete adaptive voyage planning system should include decision-support collision avoidance. Several solutions such as NAVDEC (Borkowski, 2017) and Totem Plus (Totem Plus, 2019) are available; however, they

are installed on a limited number of vessels. We presented safety distance usage from other research areas and our own proposal. Understanding benefits or shortcomings of this approach requires an extended research, including more detailed and realistic scenarios. Scenarios should include route checking and adaptation in pre-departure and while underway stages. They should be evaluated to domain models in terms of simplicity and application.

5 CONCLUSION

We have presented elements of a contemporary voyage planning process. It is subjective and dependent on skills and experience of the Masters and Officers. Similarly, we have presented several major areas of voyage planning research with focus on coastal voyage phase. Research on usage of Automated Identification System (AIS) data, navigational risk assessment, detailed collision and obstacle avoidance, position methods and sources, contributes to effective and adaptive voyage planning. These and other topics will be a part of future and extensive research.

We observed dissimilarities in ECDIS safety settings setup and determination. The parameters can be set for each leg or the whole voyage. However, it is still challenging to interpret data and to adapt a voyage plan frequently to navigational changes in-between. Effort to optimize voyage plan, especially in coastal phase, increases with each voyage objective change. Fulfilment of present and future safety, environmental and economic goals is also important. Therefore, we require better solutions for decision-support or automated route adaptation in coastal areas. Improved solutions need high-resolution and high-frequency of data updates. Despite challenges, multi-criteria and multi-objective approaches will improve voyage planning and optimization. This will reflect in new adaptive solutions for coastal phase of voyage planning and execution.

We will further evaluate and develop safety distance concept presented, including details on vessel characteristics, movement and hydrographic constraints.

Acknowledgments

This work has been supported by the Croatian Science Foundation under the project IP-2018-01-3739.

REFERENCES

Ari, I., Aksakalli, V., Aydoğdu, V. and Kum, S. (2013). Optimal ship navigation with safety distance and realistic turn constraints. *European Journal of Operational Research,* 229, pp. 707–717.

Basiana Ribera, L., Castells Sanabra, M., Grifoll Colls, M., Martínez de Osés, F.X. and Borén Altés, C. (2017). Ship-weather routing applied to short sea distances: study of the feasibility of SIMROUTEv2 algorithm. *Proceedings of 18th IAMU AGA Conference,* Varna, Bulgaria, pp. 330–340. Varna: Nikola Vaptsarov Naval Academy.

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., Olson, J. B., James, E. P., Dowell, D. C., Grell, G. A., Lin, H., Peckham, S. E., Smith, T. L., Moninger, W. R., Kenyon, J. S. and Manikin, G. S. (2015). A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh. *Monthly Weather Review*, 144, pp. 1669–1694.

Borkowski, P. (2017). The ship movement trajectory prediction algorithm using navigational data fusion. *Sensors,* 17(6): 1432, pp.1-12. DOI: 10.3390/s17061432.

Calder, B. R. (2015). On Risk-Based Expression of Hydrographic Uncertainty. *Marine Geodesy,* 38, pp. 99–127.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1, pp. 269–271.

Kang, D., Jung, J., Oh, S. and Kim, S. (2015). Automatic route checking method using post-processing for the measured water depth. *Proceedings of 2015 International Association of Institutes of Navigation World Congress (IAIN),* Prague, Czech Republic, pp. 132-136. London: IAIN.

Dramski, M. (2011). Shortest path problem in static navigation situations. *Metody Informatyki Stosowanej,* 5, pp. 173–180.

Dramski, M. and Mąka, M. (2013). Algorithm of Solving Collision Problem of Two Objects in Restricted Area. In: Mikulski, J. (ed.) *Activities of Transport Telematics, Communications in Computer and Information Science*, pp. 251–257. Berlin/ Heidelberg: Springer.

Fujii, Y. and Tanaka, K. (1971). Traffic Capacity. *The Journal of Navigation,* 24, pp. 543–552.

Gao, M., Shi, G., Li, W., Wang, Y., Liu, D. (2017). An Improved Genetic Algorithm for Island Route Planning. *Proceeding of 13th Global Congress on Manufacturing and Management (GCMM 2016) (Procedia Engineering Vol 174),* Zhengzhou, China, pp. 433–441. Amsterdam: Elsevier B. V.

Grifoll Colls, M., Castells Sanabra, M., Martínez de Osés, F. X. (2016). Enhancement of Maritime Safety and Economic Benefits of Short Sea Shipping Ship Routing. *Proceedings of the International Conference on Maritime Safety and Human Factors (SEAHORSE) 2016,* Glasgow, United Kingdom, p. 5.

Grifoll, M., Martínez de Osés, F. X., Castells, M. (2018). Potential economic benefits of using a weather ship routing system at Short Sea Shipping. *WMU Journal of Maritime Affairs,* 17, pp. 195–211.

Hart, P. E., Nilsson, N. J. and Raphael, B. (1968). A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics*, 4, pp. 100–107.

International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA). (2018). *IALA NAVGUIDE 2018: Marine Aids to Navigation Manual,* Edition 8. St Germain en Laye: IALA.

International Hydrographic Organization (IHO). (2018). ENC Standards Maintenance Working Group (ENCWG3-5.2): Paper for Consideration by ENCWG3 High Density Bathymetric ENCs. Available at: https://www.iho.int/mtg_docs/com_wg/ENCWG/ENCWG3/ENCWG3-5.2_HighDensityBathymetricENCs.pdf, accessed 11 October 2019. Monaco: IHO.

International Hydrographic Organization (IHO). (2018). *S-100: Universal Hydrographic Data Model,* Edition 4.0.0. Monaco: IHO.

International Hydrographic Organization (IHO). (2017). *S-67: Mariners' guide to accuracy of electronic navigational charts (ENC),* Edition 0.5. Monaco: IHO.

International Maritime Organization (IMO). (2000). *Resolution A.893(21): Guidelines for voyage planning.* London: IMO.

Jeong, M., Lee, E. and Lee, M. (2018). An Adaptive Route Plan Technique with Risk Contour for Autonomous Navigation of Surface Vehicles. *OCEANS 2018 MTS/IEEE Charleston,* Charleston, USA, pp. 1–4. Piscataway: IEEE.

Jeong, M. G., Lee, E. B., Lee, M. and Jung, J. Y. (2019). Multi-criteria route planning with risk contour map for smart navigation. *Ocean Engineering,* 172, pp. 72–85.

Jia, S., Dai, Z. and Zhang, L. (2019). Automatic Ship Routing with High Reliability and Efficiency between Two Arbitrary Points at Sea. *The Journal of Navigation,* 72, pp. 430–446.

Kang, D., Shim, W., Oh, S. and Kim, S. (2014). Assessment of ENC sounding by Delaunay Triangulation method in aspect of fine compilation for safe navigation. *Proceedings of 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS),* Kitakyushu, Japan, pp. 226–230. Piscataway: IEEE.

Kim, H, Kim, D., Shin, J. U., Kim, H. and Myung, H. (2014). Angular rate-constrained path planning algorithm for unmanned surface vehicles. *Ocean Engineering,* 84, pp. 37–44.

Lazarowska, A. (2015). Ship's Trajectory Planning for Collision Avoidance at Sea Based on Ant Colony Optimisation. *The Journal of Navigation,* 68, pp. 291–307.

Lee, S. M., Roh, M. I., Kim, K. S., Jung, H., Park, J. J. (2018). Method for a simultaneous determination of the path and the speed for ship route planning problems. *Ocean Engineering,* 157, pp. 301–312.

Lu, R., Turan, O., Boulougouris, E., Banks, C. and Incecik, A. (2015). A semi-empirical ship operational performance prediction model for voyage optimization towards energy efficient shipping. *Ocean Engineering,* 110, pp. 18–28.

Mannarini, G. (2016). VISIR-I: Small vessels - Least-time nautical routes using wave forecasts. *Geoscientific Model Development,* 9, pp. 1597–1625.

Marine Accident Investigation Branch (MAIB). (2017). *Grounding of bulk carrier Muros*. Available at: https://www.gov.uk/maib-reports/grounding-of-bulk-carrier-muros, accessed 27 August 19.

Masters J. (2019). *IBM Introducing the World's Highest-Resolution Global Weather Forecasting Model*. Available at: https://www.wunderground.com/cat6/IBM-Introducing-Worlds-Highest-Resolution-Global-Weather-Forecasting-Model, accesed 3 September 2019.

Pietrzykowski, Z. and Uriasz, J. (2010). Knowledge Representation in a Ship's Navigational Decision Support System. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation,* 4, pp. 265–270.

Xu, X. (2012). PathFinding.js: Path-finding library *[online algorithm]*. Available at http://qiao.github.io/PathFinding.js/visual/, accessed 10 August 2019.

Rutkowski, G. (2018). ECDIS Limitations, Data Reliability, Alarm Management and Safety Settings Recommended for Passage Planning and Route Monitoring on VLCC Tankers. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation,* 12, pp. 483–490.

Šakan, D., Žuškin, S., Brčić, D. and Valčić, S. (2019). Analysis of Primary Position Validation in ECDIS System. *Proceedings of the 13th International Conference on Marine Navigation and Safety of Sea Transportation,* Gdynia, Poland, pp. 5-15. Boca Raton: CRC Press.

Simonsen, M. H., Larsson, E., Mao, W. and Ringsberg, J. W. (2015). State-of-the-Art Within Ship Weather Routing. *Proceedings of the ASME 2015 34th International Conference on Ocean, Offshore and Arctic Engineering, American Society of Mechanical Engineers Digital Collection*, St. John's, Canada, p. 1-11. New York: ASME.

Singh, Y., Sharma, S., Sutton, R., Hatton, D. and Khan, A. (2018). A constrained A* approach towards optimal path planning for an unmanned surface vehicle in a maritime environment containing dynamic obstacles and ocean currents. *Ocean Engineering,* 169, pp. 187–201.

Szlapczynska, J. (2015). Multi-objective Weather Routing with Customised Criteria and Constraints. *The Journal of Navigation*, 68, pp. 338–354.

Szlapczynski, R., Krata, P. and Szlapczynska, J. (2018). Ship domain applied to determining distances for collision avoidance manoeuvres in give-way situations. *Ocean Engineering,* 165, pp. 43–54.

Szlapczynski, R. and Szlapczynska, J. (2017). Review of ship safety domains: Models and applications. *Ocean Engineering,* 145, pp. 277–289.

Takashima, K., Mezaoui, B. and Shoji, R. (2009). On the Fuel Saving Operation for Coastal Merchant Ships using Weather Routing. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation,* 3(4), pp. 401–406.

Totem Plus. (2019). Totem Plus–DST according to COLREGS. Available at: http://www.totemplus.com/DST.php, accessed 4 September 2019.

Tsou, M. C., Kao, S. L., and Su, C. M. (2010). Decision Support from Genetic Algorithms for Ship Collision Avoidance Route Planning and Alerts. *The Journal of Navigation,* 63, pp. 167–182.

Vettor, R., Tadros, M., Ventura, M. and Guedes Soares, C. (2016). Route planning of a fishing vessel in coastal waters with fuel consumption restraint. In: Guedes Soares, C. and Santos, T. A. (eds.) *Maritime Technology and Engineering III: Proceedings of the 3rd International Conference on Maritime Technology and Engineering (MARTECH 2016*), Lisbon, Portugal, pp. 167–173. Boca Raton: CRC Press.

Walther, L., Rizvanolli, A., Wendebourg, M. and Jahn, C. (2016). Modeling and Optimization Algorithms in Ship Weather Routing. *International Journal of e-Navigation and Maritime Economy,* 4, pp. 31–45.

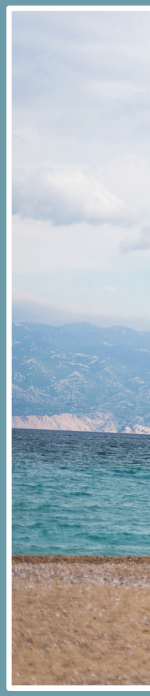
Wang, N., Meng, X., Xu, Q. and Wang, Z. (2009). A Unified Analytical Framework for Ship Domains. *The Journal of Navigation,* 62, pp. 643–655.

Weintrit, A. and Kopacz, P. (2004). The Use of Distance Measurement to the Nearest Navigational Danger in Route Planning, Route Monitoring and Voyage Recording in ECDIS. *Proceedings of the XIV-th International scientific and technical conference The part of navigation in support of human activity on the sea,* Gdynia, Poland, pp. 398-409.

Wielgosz, M. (2017). Ship Domain in Open Sea Areas and Restricted Waters: an Analysis of Influence of the Available Maneuvering Area. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation,* 11, pp. 99–104.

Žuškin, S., Brčić, D. and Kos, S. (2016). Partial structural analysis of the ECDIS EHO research: The safety contour. *Proceedings of the 7th International Conference on Maritime Transport,* Barcelona, Spain, pp. 246-262. Barcelona: UPC.

Žuškin, S., Brčić, D. and Valčić, S. (2017). ECDIS Possibilities for BWE Adoption. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation,* 11, pp. 477–482.



Technical co-sponsors

Media coverage