Using the Opus Codec for Simulated Radio Communications

Peter Ross, William Oliver and Peter Ryan
Defence Science & Technology Group
506 Lorimer St, Fishermans Bend VIC 3207, Australia
peter.ross@dst.defence.gov.au, william.oliver@dst.defence.gov.au, peter.ryan@dst.defence.gov.au

Keywords:
audio, communications, codec, DIS, HLA interoperability, network, protocol, radio, simulation,

ABSTRACT: *Since 1995 μ-law has served as the minimum required audio encoding type in DIS and RPR-FOM-based distributed simulation standards. Despite huge advances in audio compression technology, its status as the lowest common denominator means μ-law continues to be used today in distributed simulation exercises. The introduction of DIS version 8 presents an opportunity to update the minimum required encoding type to a more efficient and capable system.*

We propose Opus Codec as the new minimum required encoding type for DIS version 8. Opus is a high performance, royalty free, open standards audio codec that supports both interactive voice and music applications. At 12 kbps it can reproduce better-than-telephone quality audio, in comparison to the 64 kbps bitrate required by μ-law. Opus is used in many internet services, including Skype and YouTube. This paper describes how to use the Opus Codec in DIS and RPR based simulations. We provide theoretical and synthesised bandwidth comparisons, and estimate the real-world benefits by applying Opus to historical exercise data.

1. Introduction

Simulated voice communications make up a large portion of overall simulation network data [1,2]. Recent simulation exercises in Australia have approached the capacity limits of existing network configurations, resulting in symptoms such as distorted audio, blinking entities, and reduced responsiveness of other non-DIS network services. While more network capacity can and is being provisioned, there is potential to reduce network utilisation through other means. This paper explores adding a new audio encoding type to DIS, one that offers better compression ratios and achieves better audio quality.

2. Background

Voice-based radio communication, first embraced in the Second World War, remains a critical warfighting tool [3]. Its importance is reflected by the integration of simulated voice communications in military simulation exercises.

Interoperable simulated radio communications were first introduced in DIS version 5, and remain almost unchanged in DIS version 7. A minimal implementation requires support for two PDU types: the Transmitter PDU and the Signal PDU. Each radio transmitter issues a Transmitter PDU describing its current radio frequency, effective radiated power, modulation parameters and whether it is transmitting or in a standby or off state. The Transmitter PDU is a state PDU, and is sent at a heartbeat interval. When a radio is transmitting, audio samples – usually taken from an analogue-to-digital converter – are segmented into frames. Each frame is fed into an encoder, and the resulting encoded frame is appended to a Signal PDU header. The header describes essential information needed to process the PDU, namely the encoding type, audio sample rate and frame size. Radio receiver applications perform these operations in reverse. Audio encodings are described in the next section.

The High Level Architecture (HLA) Real-time Platform Reference Federation Object Model (RPR-FOM) uses the same approach, but with 'Transmitter' objects and 'Encoded Audio' interactions. The approach taken by DIS is also similar to that of other real-time audio streaming technologies. For example, Standard Voice over Internet Protocol (VoIP) services use the Session Initialisation Protocol (SIP) to describe the state of a telephone handset, and Real-time Transport Protocol (RTP) to communicate encoded frames.

2.1. Audio Encoding Types

Audio encoders transform analogue audio samples into a digital code. Throughout this paper, we define the rate of bits the encoder outputs as codec rate, measured in bits per second (bps).

Many different audio encoding types are supported by DIS and RPR-FOM. These are defined in SISO-REF-010 [4], and a list of current values is reproduced in Table 1. The DIS standard requires all radio implementations to support the μ-law[1]encoding at 8 and 16 kHz sampling rates [5]. Support for other encoding types is optional. A short description of each encoding system will now follow.

Table 1: Current audio encoding types supported by DIS and RPR-FOM

Enumeration Value	Description	Earliest Reference
1	8-bit μ-law (ITU-T G.711)	1972 [6]
2	CVSD (MIL-STD-188-113)	1966 [7]
3	ADPCM (ITU-T G.726)	1973 [8]
4	16-bit Linear PCM 2's complement, Big Endian	1937
5	8-bit Linear PCM unsigned	1937
8	GSM Full-Rate (ETSI 06.10)	1992
9	GSM Half-Rate (ETSI 06.20)	1992
10	Speex Narrow Band	2002
100	16-bit Linear PCM 2's complement, Little Endian	1937

Linear Pulse-Coded Modulation (LPCM) is frequently regarded as the ’uncompressed’ encoding scheme. Amplitude samples are quantized to linearly spaced levels. There are 256 levels in the case of 8-bit LPCM, and 65,536 levels in the case of 16-bit LPCM. This technology was invented in the early 20th century and is the foundation of digital audio. An 8 kHz 8-bit LPCM stream produces 64 kbps of encoded audio, while a 16 kHz 16-bit LPCM stream produces 256 kbps.

μ-law is a non-linear LPCM scheme, where amplitude samples are quantized to 256 non-linearly spaced levels. Because the human ear is less sensitive to volume level changes as volume increases, μ-law has greater perceived audio quality than 8-bit LPCM at the same sample rate. An 8 kHz μ-law stream produces 64 kbps of encoded audio.

Adaptive Differential Pulse-Code Modulation (ADPCM) is a differential encoding scheme, whereby only the delta values between adjacent amplitude levels are encoded. The number of quantization levels used to encode the delta values is varied automatically by the encoder, resulting in greater efficiency. An 8 kHz ADPCM stream produces 32 kbps of encoded audio.

Continuously Variable Slope Delta Modulation (CVSD) encodes the difference between adjacent amplitude samples as a single bit, indicating ’up’ or ’down’. An 8 kHz CVSD stream produces 8 kbps of encoded audio, but because so much information is lost during the encoding process, audio quality is poor. CVSD is typically used with higher sample rates, such as 12 and 16 kHz.

Based on the authors’ experience, spanning 20 years supporting distributed simulation, we observe that μ-law is almost always used in multi-site simulation exercises. We attribute this to three factors: *interoperability* – μ-law is a mandatory requirement and therefore guaranteed to work; *implementation complexity* – μ-law is by far the simplest algorithm to implement; and *audio quality* – in order to reduce codec rate ADPCM and CVSD further sacrifice audio quality. A further observation is that the codecs described above are all old and predate 1975; they do not feature in modern communication systems. We believe that this was done intentionally to avoid inclusion of then-patented technology in the 1995 DIS standard.

[1] SISO-REF-010 describes this encoding type as 8-bit μ-law, however as there are no other kinds of μ-law in existence, we omit the number of encoded bits-per-sample from the name for brevity.

2.2. Candidate Encoding Types

The internet and mobile telephony, and increases in computer power have resulted in the development of many newer audio codecs.

The first group of codecs, created by the 3rd Generation Partnership Project (3GPP), addresses the needs of mobile telecommunication users. These codecs include Global System for Mobile communications Full-Rate (GSM-FR), Adaptive Multi-Rate (AMR) and Enhanced Voice Services (EVS), which are deployed in the 2G, 3G/4G and 5G mobile networks respectively. All these codecs use Linear Predictive Coding (LPC) and provide exceptional performance, both in terms of audio quality, error resilience and delay. Open source implementations of the codecs are available. However, both AMR and EVS require paid patent licenses.

A separate group of codecs has primarily addressed the needs of internet users. These codecs include Internet Low Bitrate Codec (iLBC), Vorbis, Speex, CELT and SILK. Opus, now an Internet Engineering Task Force (IETF) standard, combines the latter two to create a hybrid codec that simultaneously supports speech and music applications [9]. SILK, the speech compression technology, was developed by makers of Skype and is based on the LPC approach. CELT, the wide-band audio compression technology, was developed by the Xiph.Org Foundation, and is successor to its earlier Vorbis codec. Open-source implementations of the Opus codec are available. The patented technology found within Opus is made available under royalty-free, open-source compatible licenses. These properties have made Opus popular amongst application developers, and it can be now found in computer games, internet telephony, and streaming websites.

Both groups of codecs support a range of input sample rates and different output codec rates. This paper does not attempt to provide a comparison of codec audio quality. Studies suggest Opus audio quality exceeds that of AMR, but is not as competitive as EVS [10].

2.3. Related Work

The authors are not aware of any other studies exploring encoding type changes to DIS or RPR-FOM. Other, more general techniques exist to reduce simulation network utilisation. Bundling is a tool for reducing network overhead by combining multiple PDUs into a single datagram prior to transmission. However, general purpose bundling is not suited to Signal PDUs as it introduces latency. For specific applications, such as Live Virtual Constructive integration, PDU fields can be modified to use more compact data types or be removed altogether [11]. Lossless compression algorithms can also be applied to PDUs [12]. Signal PDUs that have already been compressed will exhibit high entropy, and are unlikely to benefit from further compression. These techniques may also be combined.

3. Method

Three experiments were run to study the effects audio codecs changes have on the DIS protocol. The first experiment builds a theoretical model to estimate the network rate across different encoding settings. Network rate is a measure of bandwidth utilisation, and includes IPv4 and UDP overheads. The second experiment measures the network rate of a single radio simulation, configured to use different encoder settings. In the third experiment, we gauge the real-world benefit offered by re-encoding historical exercise data with different encoding settings.

3.1. Encoding Settings

Four encoder settings are used throughout the experiments. The first setting is 8 kHz μ-law and forms the baseline from which comparisons are drawn. The remaining three settings use 16 kHz Opus at 12, 16 and 24 kbps codec rates. Opus was selected on the basis that it represents a balance between quality and patent licensing. These settings were based on suggestions from the codec authors [13].

Opus supports constant and variable bitrate modes. The Constant Bitrate (CBR) mode forces the encoder to never exceed the target codec rate, and was used in all experiments. The Variable Bitrate (VBR) mode gives the encoder flexibility to go above or below the target codec rate in bursts, whilst achieving the target over a long-term average. VBR is more suited to archival applications.

Opus cannot encode arbitrary frame sizes, and supports only 2.5, 5, 10, 20, 40 and 60 milliseconds (ms). A 20 ms frame size was used in all experiments. Again, this was based on suggestions from the codec authors. The frame size constraint presents a problem when encoding the very last Signal PDU of an audio stream, as the available number of samples may be less than the frame size. Radio implementations have two options: drop the frame altogether, omitting up to 20 ms of audio, or pad the frame with zeros to the nearest supported frame size. We used the padding option, but chose to report the unpadded frame size in the Signal PDU.

libopus version 1.3, available to download on the Xiph.org Foundation website, was used in the practical tests. For the first two experiments, the Transmitter PDU heartbeat rate was configured to 2 seconds, representing a non-stationary platform.

4. Results

4.1. Experiment 1: Theoretical Model

In this experiment, we define the relationship between codec setting and network rate. The issuance rules for the DIS Transmitter PDU and Signal PDU provide a starting point for calculating the network rate.

Each PDU sent incurs transport overheads. The IPv4 header without option fields is 160 bits, and the UDP header is 64 bits.

The Signal PDU header is 256 bits, and is immediately followed by the encoded audio frame. Signal PDUs are sent as many times as needed to transfer the audio stream. Frame size affects latency and network rate. The larger the frame size, the longer the encoder must wait to receive samples from the analogue-to-digital converter, thus increasing latency. Reducing frame size increases the rate at which Signal PDUs are sent, resulting in more overhead.

The Transmitter PDU is 832 bits without any antenna pattern, modulation, or variable parameter records. The default DIS parameter settings require the Transmitter PDU to be sent every 2 seconds for moving entities, or every 30 seconds for stationary entities. These values may be changed by exercise agreement.

Using this information, a theoretical model of the network rate can be constructed. The relationship between codec rate and network rate is shown in Equations (1) to (3), where $CodecRate$ is in bits per second, $FrameSize$ is in milliseconds, $DataLength$ is in bits, and $TXHeartBeat$ is in seconds.

$$Overhead = 160 + 64 \tag{1}$$

$$DataLength = CodecRate \cdot \frac{FrameSize}{1000} \tag{2}$$

$$NetworkRate = CodecRate + (Overhead + 256) \cdot \frac{CodecRate}{DataLength} + \frac{Overhead + 832}{TXHeartBeat} \tag{3}$$

Table 2 presents the output of the model for different codec settings, and clearly demonstrates the overheads DIS adds to the codec rate. If we compare μ-law to 16 kbps Opus, despite reducing the codec rate by 75%, the network rate only decreases by 54.22%.

Table 2: Output of the theoretical network rate model

Encoding System	Codec rate (bps)	Network rate (bps)	Network rate decrease from μ-law
μ-law	64,000	88,528	
Opus	24,000	48,528	45.18%
Opus	16,000	40,528	54.22%
Opus	12,000	36,528	58.74%

It is important to note this model has no concept of the encoding system other than codec rate. The model presented is valid for DIS versions 5, 6 and 7. A similar model could be constructed for DIS version 8, or even HLA, assuming knowledge of how the Run-Time Infrastructure (RTI) encodes objects and interactions.

4.2. Experiment 2: Single Channel Radio

To permit validation of the theoretical model, Opus was added to an existing DIS radio implementation and an empirical test was undertaken. Audio Encoding Type 11 was selected for Opus encoded audio.

A 60 second digital recording of Melbourne Airport (YMML) air traffic control radio was fed into a simulated radio, and the resulting network packets captured in a controlled setting. The packets were analysed, starting from the initial 'transmitting' Transmitter PDU and ending with the 'not transmitting' Transmitter PDU, to determine average and peak network rate. No other simulation applications were present on the network, and the capture excluded other non-DIS traffic.

The data, summarised in Table 3, is consistent with the theoretical model results. There is a small increase in average network rate over the theoretical model, attributable to inclusion of the final Transmitter PDU.

Table 3: Single channel radio

Encoding System	Codec rate (bps)	Peak network rate (bps)	Average network rate (bps)	Network rate decrease from μ-law
μ-law	64,000	89,056	88,958	
Opus	24,000	49,056	48,960	44.96%
Opus	16,000	41,056	40,981	53.93%
Opus	12,000	37,056	36,991	58.42%

4.3. Experiment 3: Large Force Engagement Exercises

The theoretical model is useful for understanding an always transmitting, single channel, radio transmitter, but this is not representative of normal voice-based radio operations. In large force engagement exercises, voice-based radio communication occurs across multiple channels, varies in intensity level, and is accompanied by other data, such as that for entities and tactical data links. To appreciate the true impact of Opus encoding, we use the DIS recordings of six large force engagement exercises, and re-encode the Signal PDUs using different settings. A summary of the exercises is provided in Table 4. Each exercise was recorded from a central node in Australia, and involves different services conducting simulated operations of varying complexity. Bandwidth reduction techniques, described in Section 2.3, were not used in any of the exercises.

A communications analysis tool, ordinarily used to listen to DIS recording files, was modified to re-encode Signal PDUs whilst preserving all other PDU types. Transmitter PDU heartbeats were also preserved. The original recordings all used 8 kHz μ-law encoding, but with frame sizes ranging from 20 to 80 ms. Therefore, to ensure a fair comparison, the tool was also used to re-encode the existing μ-law recording to a 20 ms frame size. Spot checks also were made on all the re-encoded files to confirm the content matched that of the original recordings.

An obvious limitation of this approach is that using 8 kHz μ-law source data, and not that from an analogue-to-digital converter, as input to a 16 kHz Opus encoder is going to affect encoding performance. We acknowledge this limitation, and stress that evaluation of audio quality is not the intent of this evaluation.

The re-encoded files were analysed to determine total network data size and peak network rate. Total network data size is the cumulative number of bytes sent over the network. Peak network rate is the highest data rate observed in any one-second window over the exercise duration. The peak occurs when many packets are sent simultaneously, such as when multiple radios are transmitting, or when a large number of entity interactions occur at the same time. Peak network rate is an important consideration because when it exceeds the amount of available bandwidth, the packets are buffered by network equipment. When those buffers are completely filled, the network equipment has no option but to drop packets, producing the kind of symptoms motivating this research.

Table 4: Historical simulation exercise recordings

Exercise	Year	Duration (seconds)	Audio (seconds)	Entities	Radios
E1	2001	5,536	13,179	82	52
E2	2006	8,914	8,298	254	193
E3	2012	10,800	102,560	12,213	1,247
E4	2018	7,200	4,546	211	84
E5	2018	10,800	37,515	3,374	557
E6	2018	10,800	34,557	3,388	537

4.4. Total Network Data Size

Figures 1 and 2 below show the total network data transferred for each exercise under different codec settings.

Figure 1 and Table 5 show total network data size for all PDU types. The effect of changing encoder settings was found to vary widely across the six exercises (between 6.59% and 48.47% reduction). Specifically, in exercises E4, E5 and E6 the decrease from μ-law is much less than others.

Figure 2 and Table 6 show total network data size for only Transmitter and Signal PDUs. The decrease from μ-law is much more consistent in these graphs (ranging from 37.87% to 55.84%), and is close to the theoretical model.

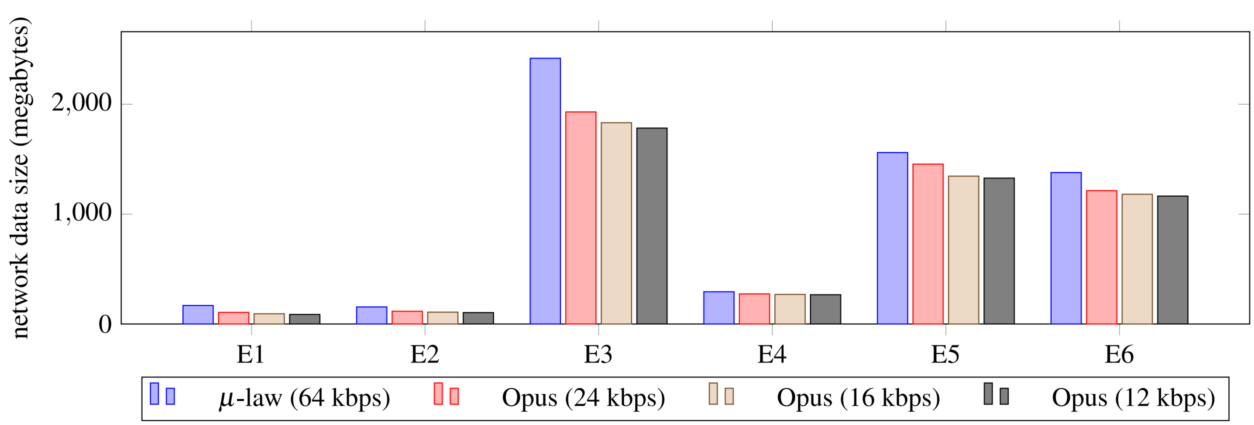

Figure 1: Total network data size for all PDU types

Table 5: Decrease in total network data size, for all PDU types

Encoding System	E1	E2	E3	E4	E5	E6
Opus (24 kbps)	37.26%	25.48%	20.19%	6.59%	6.74%	11.94%
Opus (16 kbps)	44.74%	30.59%	24.25%	8.34%	13.75%	14.34%
Opus (12 kbps)	48.47%	33.14%	26.27%	9.21%	14.90%	15.54%

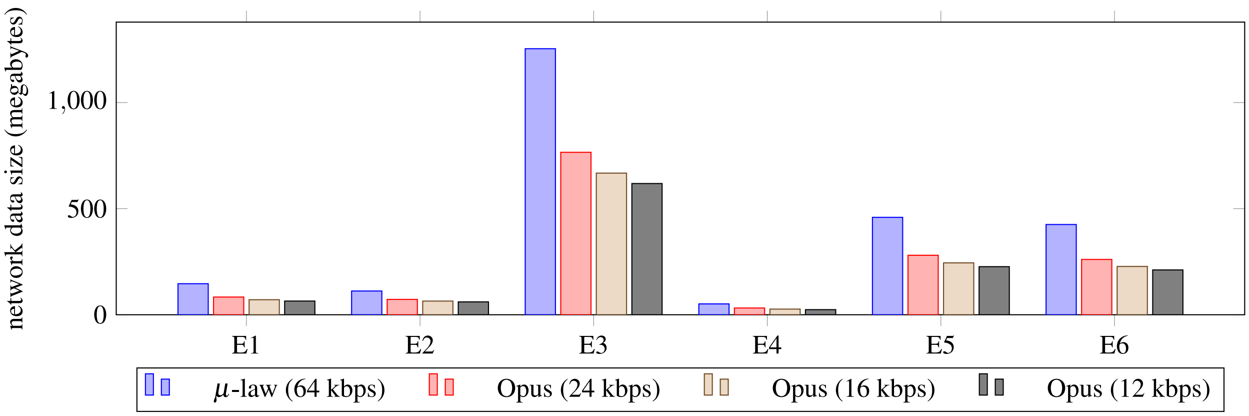

Figure 2: Total network data size for Transmitter and Signal PDUs only

Table 6: Decrease in total network data size, for Transmitter and Signal PDUs only

Encoding System	E1	E2	E3	E4	E5	E6
Opus (24 kbps)	42.93%	35.35%	38.94%	37.87%	38.93%	38.71%
Opus (16 kbps)	51.60%	42.43%	46.76%	47.90%	46.74%	46.47%
Opus (12 kbps)	55.84%	45.89%	50.68%	52.91%	50.64%	50.35%

4.5. Peak Network Rate

Figures 3 and 4 below show the peak network rate for each exercise under different codec settings.

Figure 3 and Table 7 show peak network rate for all PDU types. The peak network rate for exercise E1 does not change at all, because there was a significant entity interaction during the exercise when no radio communication was present.

Figure 4 and Table 8 show peak network rate for Transmitter and Signal PDUs only.

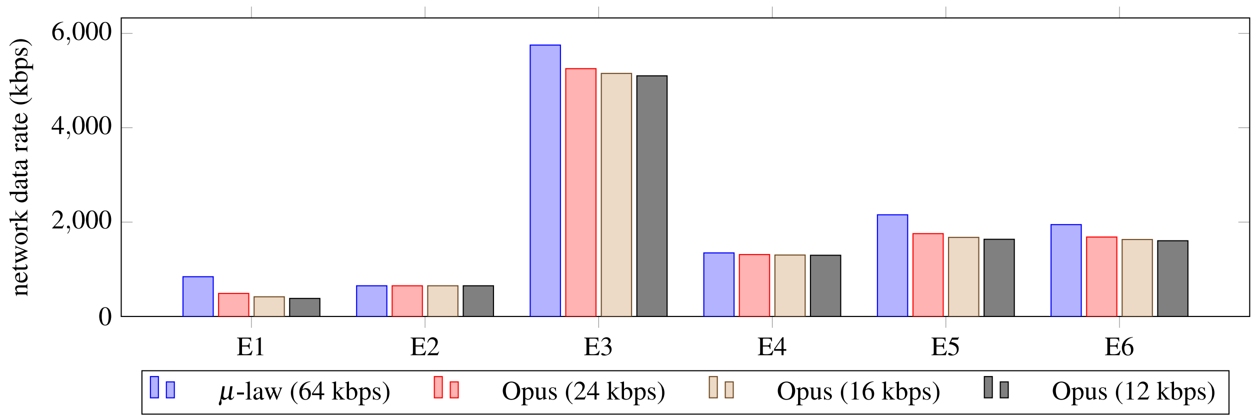

Figure 3: Peak network rate for all PDU types

Table 7: Decrease in peak network rate, for all PDU types

Encoding System	E1	E2	E3	E4	E5	E6
Opus (24 kbps)	42.07%	0.00%	8.68%	2.58%	18.53%	13.52%
Opus (16 kbps)	50.48%	0.00%	10.45%	3.34%	22.24%	16.22%
Opus (12 kbps)	54.63%	0.00%	11.33%	3.72%	24.10%	17.57%

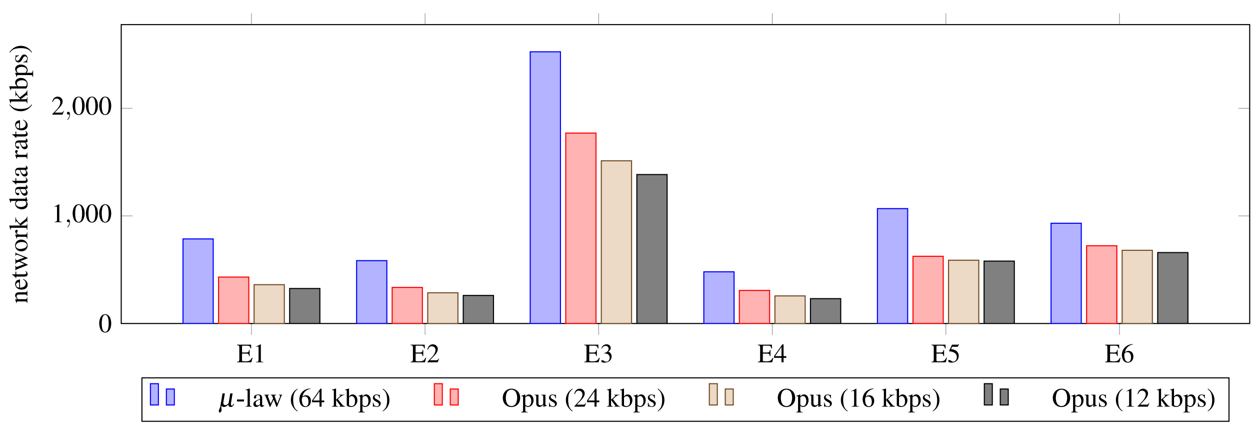

Figure 4: Peak network rate for Transmitter and Signal PDUs only

Table 8: Decrease in peak network rate, for Transmitter and Signal PDUs only

Encoding System	E1	E2	E3	E4	E5	E6
Opus (24 kbps)	45.09%	42.59%	29.92%	36.01%	41.49%	22.47%
Opus (16 kbps)	54.11%	51.11%	40.12%	46.60%	44.94%	26.99%
Opus (12 kbps)	58.61%	55.37%	45.18%	51.90%	45.69%	29.24%

5. Other Capabilities

Opus provides two additional capabilities not examined by this study.

Opus supports packet loss concealment. When the Opus decoder is informed of packet loss, it will attempt to fill in the missing information and produce comfort noise. DIS versions 5, 6 and 7 provide no mechanism for the receiving application to detect packet loss, so it was not possible to evaluate this feature. The current proposal for DIS v8 includes a sequence number in the new Voice Communications PDU, enabling the detection of packet loss.

The Opus standard supports stereophonic encoding. This will permit simulation of wide-band stereo broadcast radio, not presently achievable using existing encoding types.

6. Conclusion

Using the theoretical model, empirical tests and historical exercise data, the benefits of switching from μ-law to Opus codec have been presented. We consider peak network rate and 16 kbps Opus in our conclusion. In an unrealistic scenario of an always transmitting radio and standard heartbeats, Opus decreases peak network rate by 54%. This is the greatest reduction possible in accordance with the model. For a more realistic large force engagement scenario, the peak network rate for Transmitter and Signal PDUs was decreased by between 22.47% and 58.61%. The results are less impressive when all other PDU types are considered, and dependent on the length and type of exercise. Decreasing the codec rate further, to 12 kbps, appears to have diminishing effect on the network rate, because the network transport and Signal PDU header sizes quickly dominate the equation.

We believe there is a reasonable case to upgrade the mandatory audio encoding system in the next revision of the DIS standard. Making the new codec mandatory guarantees adoption and encourages interoperability. There is nothing technically stopping the use of Opus today in DIS versions 5, 6 and 7, however achieving sufficient adoption is going to be a challenge. For the interim, Opus could be applied selectively to network segments that are currently under stress, such as those between sites or major nodes. This would require the installation of gateways to translate between the existing encoding system and Opus in real-time.

We not have examined the computational cost of Opus versus other codecs. Undoubtably it requires more compute cycles than μ-law, however, even on embedded platforms Opus is not regarded as computationally expensive [14].

7. References

[1] A. Berry, "Lessons learned from Virtual Flag integration," *The Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*, pp. 118–126, 2004. Paper 1648.

[2] P. Ryan, P. Ross, P. Clark, and L. Zalcman, "Interoperability issues encountered with Australian/US Navy simulation exercises," *European Simulation Interoperability Workshop*, 2004. Paper 04E-SIW-030.

[3] K. Allard, *Command, control, and the common defense*. National Defense University, Institute for National Strategic Studies, 1996.

[4] SISO-REF-010-2018, "Reference for enumerations for simulation interoperability – Version 25." Simulation Interoperability Standards Organization.

[5] IEEE 1278.1-2012, "Distributed Interactive Simulation – Application protocols." Institute of Electrical and Electronics Engineers.

[6] ITU-T G.7111, "G711: Pulse Code Modulation (PCM) of voice frequencies," 1972. International Telecommunication Union.

[7] G. J. Anton, "Delta modulation signal transmission system," 1966. US Patent 3,249,870.

[8] P. Cummiskey, N. Jayant, and J. Flanagan, "Adaptive quantization in differential PCM coding of speech," *Bell System Technical Journal*, vol. 52, no. 7, pp. 1105–1118, 1973.

[9] J.-M. Valin, K. Vos, and T. Terriberry, *Definition of the Opus audio codec*. Internet Engineering Task Force, 2012. RFC 6716.

[10] A. Rämö and H. Toukomaa, "Subjective quality evaluation of the 3GPP EVS codec," *Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5157–5161, 2015.

[11] L. Call, S. Swigert, M. Zamba, and D. Noah, "Compressed DIS," *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, pp. 15–520, 2017. Paper 17074.

[12] D. J. Van Hook, J. O. Calvin, and D. C. Miller, "A Protocol Independent Compression Algorithm (PICA)," *Advanced Distributed Simulation Memorandum*, 1994.

[13] Xiph.Org Foundation contributors, "Opus recommended settings," 2018. https://wiki.xiph.org/index.php?title=Opus_Recommended_Settings; accessed 11-Dec-2018.

[14] M. Sundvall, "Opus audio codec in mobile networks," Master's thesis, Aalto University School of Electrical Engineering, 2014.

Author Biographies

WILLIAM OLIVER is a senior researcher in the Defence Science & Technology (DST) Group's Aerospace Division, specialising in interoperability issues and analysis techniques for advanced distributed simulation. Prior to joining DST, he developed software for flight simulators and simulated maintenance trainers, and studied Engineering and Mathematics.

PETER ROSS graduated from RMIT University in 2001 with a Bachelor of Applied Science, majoring in computer science. He joined DST's Aerospace Division in 2003, where he has undertaken a role in evaluating the use of advanced distributed simulation for collective training.

DR. PETER RYAN is an Honorary Research Fellow in DST Group's Aerospace Division. He has a 30 year background in the modelling and simulation of military operations. His main research interests include Advanced Distributed Simulation, real time simulation, synthetic environments, and their potential to provide enhanced training solutions for the Australian Defence Force.