An Overview of the Tactical Secure Voice Cryptographic Interoperability Specification
Thomas Moran
David Heide
Swati Shah
U.S. Naval Research Laboratory

Abstract
The Tactical Secure Voice Working Group was formed by the National Security Agency in 2008 to ensure that modernized tactical secure voice devices will be interoperable across the Department of Defense. The Naval Research Laboratory was tasked by the working group to develop the Tactical Secure Voice Cryptographic Interoperability Specification (TSVCIS), which became an official National Security Agency (NSA) document in July of 2009.

The TSVCIS consists of two documents, one classified and the other unclassified. These documents define the voice encoding, encryption, framing, synchronization, key management, and other functions for tactical secure voice and data radio communications.

This paper describes the unclassified portion of the TSVCIS and explains some of the improvements over the legacy tactical secure voice systems. It also describes how the modernized secure voice modes can provide even wider benefits in the future.

Main Objective and Overall Design Goals
The primary objective in developing the TSVCIS was to ensure that the cryptographically modernized modes of two families of tactical secure voice devices now under development be interoperable. These two families of devices are the Air Force led VINSON/ANDVT Cryptographic Modernization (VACM) program and the Navy led ARC-210 Generation 5 radio program.

The main goal when specifying the modernized modes was to improve operational performance wherever possible and at least meet legacy performance when improvements were not possible. While interoperability across devices is an obvious concern, simultaneously modernizing both narrowband and wideband tactical secure voice was a rare opportunity to do much more than that. It provided the chance to address the longstanding issue of legacy stovepipes in tactical secure voice.

The ANDVT and VINSON family of devices being replaced are incompatible; both with each other and with themselves across the different voice modes. The differences between these links include: data rates, methods for voice compression, encryption, synchronization, and framing. These differences not only hinder existing tactical communications but also prevent the modernization of the overall communications networks. This total lack of commonality prevents these links from someday being converged into a larger network that could provide more effective and efficient tactical communications.

The overall design goal with the TSVCIS was to implement all the voice modes using a common, layered approach. One that provides the basis for secure interoperability across the specified voice modes and allows for wider network convergence in the future. It is important to note that the TSVCIS only provides the basis for this wider interoperation. Actually achieving narrowband/wideband voice interoperability will be dependent on the implementation of the specific devices and systems.

Due to the urgent need of the program offices the draft TSVCIS was produced in a little over six months. In spite of the timeframe, significant improvements were made in many areas. A number of these improvements stem from taking a common approach to implementing all of the modes of operation of the combined VINSON and ANDVT families of devices being modernized. Consequently, the goal of commonality across all modes drove many of the design decisions.

Scope
The TSVCIS defines the required voice encoding, encryption, framing, synchronization, key management, and other functions for tactical secure voice and data radio communications capabilities employed in emerging tactical secure voice devices.

It will first apply to the VACM and ARC-210 Generation 5 radio programs. Both of these programs will produce a family of tactical secure voice devices for a variety of aircraft, ship, and land-mobile platforms. All future programs using
modernized cryptographic devices for tactical secure voice will need to adhere to the TSVCIS.

The TSVCIS was developed to ensure that modernized devices implementing new cryptographic algorithms can interoperate. Simply having one common mode of operation that is shared by the different devices would satisfy this requirement. But the TSVCIS accomplishes more than that by providing the following areas of improvement over current tactical secure voice technology: enhanced interoperability, crypto and framing synchronization, over-the-air key distribution, higher data rates, more robust transmissions.

The TSVCIS also calls for a single, multirate voice compression algorithm for all of the standard-rate modes of voice operation. So all TSVCIS-compliant devices will interoperate in all modes, and can potentially interoperate across all the voice and data modes; even when encrypted.

Direct Benefits

The following section describes some of the direct benefits that arise from the TSVCIS.

1) Interoperable Modes: To provide interoperability between the narrowband and wideband voice modes the narrowband vocoder is embedded in the bitstream of all the voice modes. In addition, the wideband and narrowband modes have common synchronization methods, common Mode Control Words, and use common cryptographic techniques.

The vocoder core is the DoD standard for 2400 bits/second voice; Mixed Excitation Linear Predictor; enhanced version (MELPe).

<table>
<thead>
<tr>
<th>54 bits from MELPe 2400 bit/s standard</th>
<th>FEC to protect MELPe bits</th>
<th>Additional spectral coefficients to improve speech quality</th>
<th>Additional FEC to protect spectral coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>54 bits from MELPe 2400 bit/s standard</td>
<td>FEC to protect MELPe bits</td>
<td>Additional spectral coefficients to improve speech quality</td>
<td>Additional FEC to protect spectral coefficients</td>
</tr>
</tbody>
</table>

Figure 1 - Example of Embedded Bitstream approach (16000 bit/s with FEC mode shown)

All four wideband voice modes include the 2400 bits/sec MELPe as the bitstream kernel. An example is illustrated in figure one. The first 54 bits of each wideband mode bitstream frame are the exact 54 bits of the narrowband vocoder speech frame.

Converting from a wideband mode to the narrowband mode involves just stripping the first 54 bits and passing them directly to the narrowband coder. Converting from narrowband to a wideband mode involves just adding FEC to the 54-bit kernel. This simple stripping or appending bitstreams makes interoperability clearly achievable and does not involve any voice parameter decoding/re-encoding that can typically hurt speech quality. See Figure Two.

The common synchronization methods include frame synchronization, the Initialization Vector (IV), Mode Control Word (MCW), ID Fields, and End Of Message fields for narrowband and wideband modes.

The MCW automates switching between voice and data modes, thereby eliminating the need for the receive operator to manually switch between modes. Although legacy narrowband equipment already uses MCWs, legacy wideband equipment does not. Now MCWs are specified for both narrowband and wideband communications.

Also, certain MCWs are now specified to be the same for narrowband, wideband, and the voice modem modes. This should simplify the design of an interoperable gateway by, in part, allowing the transmission preamble to remain intact within the gateway. Otherwise, the preamble would need to be decoded, translated, and re-coded at the gateway.

Importantly, common cryptographic techniques (algorithm/framing/mode) are also specified for the narrowband and wideband modes that allow direct interoperability across these modes. That is, no process of decrypting and re-encrypting is needed to interoperate across the voice modes.

To summarize, all these design features, taken together, enable direct digital interoperability between narrowband and wideband voice modes (5-kHz and 25-kHz devices) in the black (encrypted) via a simple, secure, gateway, and without speech degradation. See Figure Three

2) Improved Voice Quality: One of the goals in writing the TSVCIS was to improve voice quality as compared to existing equipment. A key aspect of overall vocoder performance is
how well a vocoder performs in two common military environments: a) harsh acoustic platform noise and b) harsh RF channel noise (bit errors).

To judge the performance of candidate voice modes, the prospective modes were analyzed with formal speech intelligibility and acceptability tests in various noise conditions. These test scenarios included channel conditions with up to a 10% Bit Error Rate (BER) and acoustic environment such as the HMMWV and the Blackhawk helicopter.

For the 2400 bits/s narrowband voice mode, MELPe is the DoD and NATO standard. It is designed to provide good performance overall in tactical environments and has been tested extensively as part of the standardization process.

The goals for the wideband voice modes, besides improving voice quality, were to provide multiple voice rates that can be matched to varying channel conditions while still maintaining direct interoperability with the other voice modes, both wideband and narrowband.

These goals were met with four wideband voice modes that are all supersets of the MELPe 2400 bit/s mode. There is an 8000 bit/s mode, a 12,000 bit/s mode, and two modes at 16,000 bit/s. The rates of these modes were selected to conform to the voice data rates of existing radio equipment. The modes themselves were defined to accommodate the expected range of acoustic and channel noise condition. The 8000 bit/s and 12,000 bit/s modes are both super-protected MELPe frames. They both have Forward Error Correction (FEC) layered on the narrowband bitstream to strongly protect it from channel noise and so provide higher voice quality over longer transmission distances.

The first 16000 bit/s mode contains both layered FEC to protect from channel noise and a layer of additional voice information (spectral coefficient parameters) to improve voice quality in harsh acoustic environments.

Finally, the second 16000 bit/s mode contains no FEC but even more voice information. This mode is designed to provide high voice quality when the channel has low bit errors or FEC is provided out of band.

The users can select the voice mode that best matches their needs. Ideally, a radio system would be flexible enough to switch between modes to adjust to changing noise conditions. These four modes were defined to provide this flexibility while still maintaining direct interoperability through the MELPe vocoder.

3) Channel Bit Error Resistance: As stated above, ensuring good voice quality even in harsh RF channel noise is very important. To address this issue, three of the four wideband modes have inband FEC to correct channel bit errors. In addition, two optional narrowband modes were designed with inband FEC. The main FEC technique used was BCH block coding of various lengths. These block codes give good performance, can be implemented in hardware or software, and are memoryless (which is important because interoperability involves stripping portions of the bitstream). The performance of the FEC protected 8000, 12000, & 16000 bit/s modes is given in the table below. Note that 8000 & 16000 FEC protected modes yield less than 1% BER even up to 8% uncoded random bit error rate. And the 12000 bit/s protected mode yields less than 1% BER at 10% uncoded BER. While the narrowband specification includes optional inband FEC, the TSVCIS also included the specification of modern FEC for HF radios.

<table>
<thead>
<tr>
<th>Uncoded bit error rate (%)</th>
<th>Coded bit error rate (%) BCH (15,5) 8000 &amp; 16000 bit/s non-DAMA channel modes</th>
<th>Coded bit error rate (%) 1st layer BCH (15,5) 2nd layer BCH (125,55) 12000 bit/s non-DAMA channel mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1.0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2.0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>3.0</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>4.0</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>5.0</td>
<td>0.19</td>
<td>0.00</td>
</tr>
<tr>
<td>6.0</td>
<td>0.36</td>
<td>0.02</td>
</tr>
<tr>
<td>7.0</td>
<td>0.58</td>
<td>0.05</td>
</tr>
<tr>
<td>8.0</td>
<td>0.94</td>
<td>0.08</td>
</tr>
<tr>
<td>9.0</td>
<td>1.40</td>
<td>0.29</td>
</tr>
<tr>
<td>10.0</td>
<td>2.13</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Table 1 - Performance of in-band FEC in 8000/12000/16000 bit/s wideband modes.

4) Cryptographic Synchronization:
Many improvements with cryptographic synchronization were achieved through the use of NSA-approved modern Suite A and B algorithms for traffic encryption and over-the-air key distribution. Use of these algorithms also allows emergent tactical secure voice devices to comply with the NSA's Cryptographic Modernization requirements.

The TSVCIS cryptographic synchronization and framing is based on the Secure Communications Interoperability Protocol (SCIP) as much as possible. SCIP is an existing protocol designed mainly for secure telephony. The approach in using elements of SCIP was to support future SCIP interoperability via a gateway but provide good tactical secure voice device performance.

Having minimal delay is an important requirement for tactical communications. Several layers of synchronization between the transmitter and receiver need to properly occur before the received speech can be heard. Since this synchronization must occur at the start of every half-duplex transmission it is a critical factor in communications response time.

For narrowband modes only a few changes were made as the legacy methods for tactical secure voice synchronization still perform well. These changes were to allow for more modes, some security improvements, and a slightly faster synchronization time.

For the wideband modes, more changes were made. Bit synchronization was made user selectable down to 2ms. This is a great improvement over legacy wideband modes. Frame synchronization was much improved by using a more robust Pseudo Random Noise (PN) sequence, versus the legacy Phi pattern. The PN sequence is the same as in the narrowband modes, but repeated three times. This provides robust synchronization and supports interoperability by simplifying the design of a wideband/narrowband gateway.

5) Key Management: Current tactical secure voice devices support over the air distribution (OTAD) of key material from a broadcasting device to receiving devices. The TSVCIS improves on legacy OTAD capabilities by specifying: (a) Suite A and Suite B modern cryptographic algorithms for key wrapping, thus, promoting releasability and the Cryptographic Interoperability Strategy; (b) Wrapping the traffic key being delivered with a separate pre-placed encryption key prior to transmission, thus adding better protection of the delivered key and offering layered security over legacy OTAD techniques; (c) Delivering the traffic key with metadata that include at a minimum key identification attributes, which offers improved key management and allows easier identification of the key upon receipt (current TSV devices do not identify the transmitted key which requires more manually intensive pre-coordination on the broadcast); (d) Allowing the receiving unit to extract the received encrypted key into a connected key fill device; (e) Improved robustness by using FEC coupled with Majority Voting Logic (legacy only uses FEC); (f) Transmission times are kept low (less than 2s at 2400 bit/s) while still providing the above improvements.

Potential Benefits

The following section describes some of the potential benefits that could be achieved by taking advantage of the inherent possibilities of the flexible architecture, established in the TSVCIS v.1.0. These are potential benefits because they depend upon how the TSVCIS may be used, not on how it must be used.

1) All voice mode black interoperability: Because interoperability between modes is achieved by simply stripping bits or appending bits to a bitstream, secure black interoperability can be achieved. This secure interoperability is possible because all wideband and narrowband bitstream formats are layered and are defined by mode. The boundaries for all the voice parameters, FEC bits, etc. are known without the need to decrypt the bitstream. This makes end-to-end security possible even at network junctions and across dissimilar data transport networks.

2) Automatic/dynamic voice mode selection at receiver: One of the main advantages of the 16000 bit/s Wideband Mode (non-DAMA channels) is that there are two options for synthesizing speech at the receiver. The complete bitstream contains both the heavily error protected MELPe 8000 bit/s mode followed by 8000 bits/s of lightly error protected additional spectral components to enhance the speech quality. The bitstream format is given in figure one above. Because FEC gives detected
errors along with error correction, the receiver can use this information to determine the severity of the channel noise. In addition, the receiver can detect the signal-to-noise ratio of the speech in the presence of acoustic noise. By combining these two measurements (channel noise vs. acoustic noise), the receiver can decide to use the heavily FEC protected synthesis option in high bit error rate conditions or in the case of especially severe acoustic noise, the synthesis option with additional spectral components would be best. These synthesis options can be periodically switched during a conversation based on changing channel or acoustic noise conditions. Because this choice of synthesis does not affect interoperability between radios, it is not strictly part of the TSVCIS, but is certainly a capability that could be utilized for improved voice quality.

This concept could be extended to all the narrowband and wideband modes if a protocol for feeding the channel information to the transmitter were to be developed. Similar protocols already exist for IP links.

3) Interoperable Secure Voice Core for Strategic Voice and VoIP: One goal for tactical voice communications is to find ways to extend the range of tactical voice across the globe to strategic networks. Currently, most tactical vocoders are designed for a fixed data rate, while future strategic voice networks may use Voice over Internet Protocol (VoIP) as the mode for voice communication. For these VoIP systems automatically variable data rate (VDR) vocoders provide efficient use of bandwidth by constantly changing the data rate based on voice content and the current network congestion. NRL has designed a variable data rate (VDR) vocoder for VoIP applications that has these features. In designing the tactical modes for the TSVCIS, NRL used a fixed rate subset of this technique for the 16000 bit/s tactical modes. Both vocoders use the same approach to spectral coefficient enhancement and both use the common MELPe kernel.

The main issue in making the VDR approach directly interoperable with the fixed rate tactical versions is that the different spectral coefficient constellations need to be mapped to one another. Specifically, the fixed rate vocoders only use a single 8-bit table to represent the spectral coefficients while VDR uses a series of 3 to 9 bit tables, which makes the design variable. By mapping these spectral coefficients tables to each other, a direct, straightforward method for interoperability with VoIP networks could be provided for both 16000 bit/s wideband modes. In addition, the narrowband modes are always directly interoperable with the MELPe kernel portion of the bitstream. Because all of the vocoders were designed with a common approach, interoperability across narrowband tactical networks, to wideband tactical networks, all the way to high bandwidth strategic voice communication is possible, securely and with very little voice quality degradation.

4) Simple Black Gateway:
Future interoperability between the 5-kHz and 25-kHz radio channels can be achieved within a gateway. The TSVCIS allows for a simple and efficient 5-kHz to 25-kHz gateway design through the use of the common vocoder, common synchronization methods, common Mode Control Words, and common cryptographic techniques present in both wideband and narrowband modes. Furthermore, the need to use RF bandwidth more efficiently may someday lead to the 25-kHz channels being reduced to 12.5-kHz. The 8000 bit/s voice mode will likely fit that channel well. Any transition from 25-kHz to 12.5-kHz channels will be greatly simplified by the fact that the 8000 bit/s voice mode is interoperable with the 16000 bit/s and 2400 bit/s voice modes.

Possible Future Work

Enhanced voice quality

1) Wideband Audio: One way to improve voice quality is to extend the audio frequency being transmitted. Vocoders typically cut off the portion of the audio spectrum above 4 kHz. Extending this range up to 7 kHz would produce much clearer consonants and maintain the upper resonant frequencies of certain vowels. Fortunately, this upper band (4-7 kHz) can be encoded independently from the lower band (0-4 kHz). This allows for the extended audio band to be added as an enhanced option when there is available capacity in the channel. Since it is independent of the lower band vocoding it could be appended to any of the other voice modes; from the 2400 bit/s MELPe mode to either one of the 16,000 bit/s modes. This could be done automatically when the channel has the capacity to support the added bandwidth.
2) Additional Voice Modes: The need for the TSVCIS to conform to existing radio equipment limited the number of voice modes that could be designed. With future radios it may be possible to further optimize voice quality by designing many more voice modes with varying proportions of vocoding and error control. These new modes could provide the optimal mix based on network congestion, channel conditions and acoustic noise conditions. The communications system could automatically switch between these different modes, constantly adjusting to changing external conditions, and providing the highest voice quality at all times.

Conclusion

The Tactical Secure Voice Cryptographic Interoperability Specification will ensure that future tactical secure voice devices will be interoperable. It provides improvements for tactical secure voice communications in the areas of: voice quality, security, time to synchronize, robustness, and ease of implementation. Many of these benefits result from the use a common, layered approach to implementing all the voice modes. This approach defines, in effect, a secure voice core that allows all the voice modes to directly interoperate; even when encrypted.

It also lays the foundation for benefits beyond the specification itself. These include providing the basis for automatic channel selection based upon channel conditions and the design of a relatively simple radio gateway between tactical secure voice devices operating in different modes.

A future version of the TSVCIS could expand on this work to provide even wider interoperability. This could include SCIP and secure VoIP interoperability.

---

**Fig. 2 Legacy Interoperation between Wideband (WB) & Narrowband (NB) Devices**

Analog Tandeming between devices:
- adds latency and complexity,
- reduces security,
- degrades speech.

**Fig. 3 TSVCIS Approach for Interconnecting Wideband (WB) & Narrowband (NB) Devices**

Direct digital interoperability between devices in the black:
- simple,
- quick,
- secure,
- no speech degradation.