Compression of PIREPs for Throughput-Limited Transmission

Joseph Rios[*]

NASA Ames Research Center, Moffett Field, CA 94035

Miguel de la Cruz[†]

Universidad Autónoma de Nuevo León, San Nicolás de los Garza, N.L. 66450 México

Pilot weather reports (commonly referred to as PIREPs) are an essential part of flight planning and situational awareness for pilots. These reports are provided by pilots to flight services stations for dissemination to other pilots. PIREPs are used during pre-flight planning as well as en route by pilots to make more informed decisions about their flight. In some areas, PIREPs are not available (before or after departure) due to lack of traditional aviation communication infrastructure. With the concept of PIREP compression presented here, multiple PIREPs can be bundled together for a single transmission over throughput-limited channels not typically used for aviation. For example, 10 to 20 compressed PIREPs might be transmitted via a cellular text message and even more might be sent via satellite using this concept, which is an increase of up to 10 times the amount of data. In this paper, the compression algorithm is presented along with some analysis of the loss of information incurred through the compression-decompression cycle.

I. Introduction

PILOTS rely heavily upon reports from other pilots to aid in flight planning and situational awareness. The formal mechanism for reporting weather information by pilots is the Pilot Weather Report or PIREP.[1] These reports contain information about turbulence, wind, visibility, icing, and temperature as well as information about the aircraft and location involved with the report. Typically, these PIREPs are solicited by Flight Service Stations (FSS) or pilots offer them freely to FSS. These PIREPs are then entered into a central database for access by other FSS or users via various web interfaces. The PIREPs may then become part of the briefing given during pre-flight planning or perhaps en route when requested by a pilot via radio. PIREPs are not typically available to en route pilots in an automated fashion. They were originally designed to be easily transmitted via voice over the radio.

As new technologies for communication and for pilot situational awareness have become more readily available, an opportunity has presented itself in terms of how PIREP data is provided to pilots. With the possibility of providing PIREPs over a broadcast (or multi-cast) channel, or via a specialized data subscription service, there may be a need to compress the data contained within the PIREP message into a form more amenable to those potential channels. For instance, a typical text message (SMS) can contain only 140 bytes (1,120 bits) of data. A service such as Iridium's Short Burst Data (SBD) provides 270 bytes (2,160 bits) of data per message. Given constraints on this order, at most 2 or 3 PIREPs could be transmitted per message. With proper compression, as presented in this paper, that number can be increased by 10 times. Providing up to 20 PIREPs per message offers the opportunity to dramatically increase the situational awareness of pilots, especially those in general aviation. If a common compression tool, such as gzip, were used to compress 20 PIREPs, the resulting file would still be over 5,000 bits in size, which is too large for either of the transmission technologies noted above.

The remainder of the paper is organized as follows. In the next section (II) some background information

---

American Institute of Aeronautics and Astronautics

is provided to frame the data compression discussion that follows. Next, the data used for the compression analysis is fully described in Section III. Then Section IV details the compression algorithm. Analysis of the compression and decompression of the input data is then provided in Section V along with a complete example of the compression concept. Finally, concluding remarks and future directions are discussed in Section VI.

II.   Background

This section is divided into three parts. First, an overview of the data reported within a PIREP is provided. Second, the system for collecting and distributing PIREPs is described. Finally, the field of data compression as it relates to this project is briefly overviewed.

II.A.   PIREP Data Description

The data included in PIREPs together with their appropriate formatting are completely described in the Federal Aviation Administration's (FAA) documentation on Flight Services.[1] In this section, an overview of that data is provided in enough detail to ground the discussion of the compression approach to follow in Section IV.

Every PIREP includes the message type, time (TM), location (OV), altitude (FL), and aircraft type (TP) for the reporting aircraft. For this work, these elements are considered to describe the "header" for the PIREP. The message type is either routine (UA) or urgent (UUA). A PIREP is labeled urgent if it reports tornadoes, funnel clouds, waterspouts, severe or extreme turbulence, severe icing, hail, low level wind shear, volcanic ash, or any other phenomenon considered hazardous by the flight services specialist. Time is reported in UTC and represents when the reported phenomenon was encountered. The location is described in reference to a navigational aid or airport (or two such points in the case of route segments). The flight level is reported in hundreds of feet. Aircraft type is reported using the appropriate designator.[2] If the PIREP is reported by a pilot who has been certified as a "SKYSPOTTER" (one who has received specialized training), '/AWC' is appended to the PIREP.

Each element except the leading message type element is preceded by a solidus (/), thus cleanly dividing the various elements into distinct fields. Each element is officially referred to as a Text Element Identifier (TEI).

The rest of the PIREP contains information about the weather phenomena being reported. Those phenomena are divided into one of six categories. A seventh "remarks" category is allowed for describing additional details or phenomena not covered by the six categories. These categories are summarized in Table 1 and then further detailed in the sections below.

Table 1. Summary of PIREP weather phenomenon elements

Phenomenon	Identifier (TEI)	Examples
Turbulence	TB	/TB LGT 040, /TB MOD CHOP 220
Air Temperature	TA	/TA 08, /TA M08
Icing	IC	/IC LGT-MOD MX 085, /IC SEV CLR 035-062
Sky Condition	SK	/SK OVC-TOP085, /SK OVC015-TOP035
Wind Direction and Speed	WV	/WV 28080KT, /WV 010105KT
Visibility and Weather	WX	/WX FV20SM, /WX FV10SM HZ

II.A.1.   Turbulence Element

The turbulence (TB) TEI has up to three components. The only required field is the intensity described as light, moderate, severe, or extreme. The abbreviations LGT, MOD, SEV, and EXTRM, respectively, are used to denote the intensities. These intensities may also be combined to show a range of intensity, such as LGT-MOD. If the pilot reports it, the PIREP will indicate if the turbulence is clear air turbulence (CAT),

American Institute of Aeronautics and Astronautics

choppy (CHOP), or both. If the encounter with turbulence occurs at a different altitude than indicated in the /FL TEI, the /TB element may also include altitude information.

II.A.2. Air Temperature Element

Air temperature is reported in degrees Celsius. The value is always described with two digits. For negative values, an "M" is prepended.

II.A.3. Icing Element

Icing (IC) is described with an intensity, type, and optional flight level. The intensity is indicated as TRACE, LGT, MOD, or SEV. The type is rime, clear, or mixed, which are indicated by RIME, CLEAR, and MX, respectively. When icing is reported, the pilot is also supposed to supply a TA report.

II.A.4. Sky Condition Element

The Sky Condition (SK) TEI reports on cloud coverage. Descriptions of the cloud cover as broken, few, overcast, scattered, or sky clear are abbreviated as BKN, FEW, OVC, SCT, and SKC, respectively. In addition to this qualitative description, an range of flight levels (or just a top) is also provided.

II.A.5. Wind Vector Element

The Wind Vector (WV) TEI describes both the direction and speed of the wind at the reported time and location. Three digits are used for the magnetic direction (000-359) and two to three digits indicate the speed in knots. A zero is prepended to a wind speed less than 10 Kts. Finally, the characters "KT" are appended to the element.

II.A.6. Weather Element

The WX element describes flight visibility, flight weather, or both. Visibility, if reported, is provided in statute miles and is indicated with a preceding "FV" followed by a two-digit value followed by "SM". If visibility is unrestricted, a value of 99 is indicated. For weather types, the FAA provides a set of abbreviations to describe 40 phenomena such as rain, drifting sand, shallow fog, and others. For precipitation phenomena, an intensity value may be indicated with a '-' or '+', describing light and heavy precipitation, respectively. For extreme weather phenomena, additional information may be required within the WX TEI or the remarks section. An altitude range may also be provided.

II.A.7. Example PIREPs

To provide a complete idea of what all of these elements look like in an actual PIREP, we now provide 4 examples each with a brief description.

1. UA /OV ILI047045/TM 0013/FL210/TP SF34/TA M23/IC MOD RIME

2. UUA /OV OMN115095/TM 0027/FL360/TP B737/TB SEV

3. UA /OV FSD/TM 0236/FL100/TP PAT4/SK UNKN050-TOP067/TA M08/IC LGT MX

4. UA /OV PIR/TM 0107/FL110/TP C208/WX -SN/TA M10/IC NEG

The first PIREP is a routine PIREP (UA) reported 47° and 45nmi from ILI (ILI047045) at 0013Z. The pilot was at flight level 210 in a Saab 340 turboprop (SF34). The pilot reported an air temperature of -23°C (M23) and moderate rime icing (MOD RIME).

American Institute of Aeronautics and Astronautics

The second PIREP was urgent (UUA) reported 115° and 95nmi from OMN (OMN115095) at 0027Z. The pilot was at flight level 360 in a Boeing 737. The report contains only one element: severe turbulence (TB SEV). Severe turbulence reports necessitate the tagging as urgent.

The third PIREP was reported over FSD at 0236Z while flying at flight level 100. From a Piper T-1040 (PAT4), the pilot reported a sky condition of "unknown" from flight levels 050 to 067 (UNKN050-TOP067). The air temperature was -8°C (M08) and there was light mixed (rime and clear) icing (LGT MX).

The fourth and final PIREP shown here was flying over PIR at 0107Z. The pilot's flight level was 110 in a Cessna 208 Grand Caravan (C208) when there was a report of light snow (-SN). There was no icing (IC NEG) and the air temperature was -10°C (M10).

II.B.   PIREP Reporting and Distribution

PIREPs are filed by pilots to Flight Service Stations (FSS). FSS actively solicit PIREPs from pilots at various times and especially when certain weather conditions are present. These conditions include ceilings at or below 5,000 feet, visibility of 5 miles or less, thunderstorms, significant turbulence, icing conditions, wind shear, and volcanic ash. Requests for PIREPs may also be sent out via certain navigation aids (NAVAIDs) radio frequencies. Also, during any standard communication between the pilot and FSS (e.g. preflight weather briefing, post-flight contact, etc.) PIREPs may be solicited.

After receiving the report from a pilot, the FSS will be formatted and entered by the FSS such that other users of the system may view the data. PIREPs are then disseminated as part of the regular pre-flight weather briefing supplied by the FSS to pilots. PIREPs can also be obtained via other services such as the Aviation Digital Data Service (ADDS) run by the National Weather Service.[3] ADDS provides data in text or XML-based formats.

II.C.   Data Compression

Data compression is the process of converting an initial data source into another data source of smaller size. In the context of this paper, the initial data sources are text files, though in other contexts, streams of data may be the target of compression. For ease of discussion, we will consider the initial data source to always be a file. A key design choice for creating a compression method is whether the compression algorithm will be lossless or lossy. A lossless compression approach ensures that a file that is compressed can be decompressed such that original file and the decompressed file are identical. A lossy approach does not ensure that decompression of the compressed data file will provide a file identical to the original. Lossy approaches are not typically applied to text data since loss of information could cause loss of meaning or understanding in the text file. However, in the approach presented in Section IV, the approach is lossy and tailored for PIREP encoding.

There are several lossless file compression tools available on most desktop computer systems. These include zip, gzip, bzip2, and Z. In general, these tools use some combination of creating a dictionary of common strings within the input data and reducing the space needed to store commonly seen sequences. For the dictionary portion, an example might be if the phrase "the dog is large" shows up many times in a file, only the first occurrence is kept and future occurrences are stored as a pointer to the first. For the reduction of bits for common symbols, arithmetic coding[4] is often used. Huffman coding[5] is a special case of arithmetic coding. The basic idea behind Huffman coding is that common symbols (e.g. 'e' in English documents) will be encoded with fewer bits than those symbols that occur with lower frequency (e.g 'z'). For further information on the compression topics, a modern textbook on data compression[6] or the original articles by Ziv and Lempel;[7] Witten, Neal, and Cleary;[4] and Huffman[5] are informative.

III.   PIREP Data

To develop the PIREP compression concept, 25,265 PIREPs from January 13, 2013 through February 6, 2013 were used. PIREP data is publicly available from the National Oceanic and Atmosphere Administration via

American Institute of Aeronautics and Astronautics

their Aviation Digital Data Service.[3] However, only the PIREPs for the previous 36 hours are accessible. To perform this work, the data service was queried at regular time periods and the resulting data stored. Figure 1 shows the frequency of the various TEIs within our test set. It also shows some of the relationships between all of the pairs of TEIs. For example, it shows that of the 12,836 elements with a TB element, 1,186 also had a TA element. It is worth noting that IC elements often have an accompanying TA element and that TB elements are very likely to be the only information in a given PIREP. In addition, it can be seen that WV and WX elements are seldom reported relative to the other elements. It is important to recognize that these reports are from across the entire United States and that the content could vary from season to season.
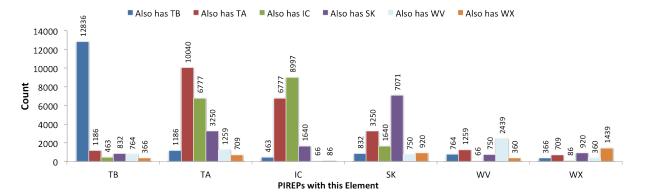

Figure 1. Counts of various elements within the 25,265 PIREPs from Jan 13 to Feb 6, 2013. The chart illustrates how often various pairs of elements occurred.

For this work we have also performed some initial analysis on the aircraft types sending the PIREPs. In the compression algorithm presented in Section IV, the specific aircraft type will be lost and replaced with an aircraft class to save transmitted data. This design choice is justified by the understanding that particular aircraft types are not as important as knowing the size of the aircraft affected by the phenomenon. For instance, a Cessna 152 has a max takeoff weight of 1,670lb and a Piper Cub PA-32 has a max takeoff weight of 3,400lb, so they could be good candidates for inclusion into a class of aircraft with a max takeoff weight less than 5,000lb. The existing categorization of aircraft by ICAO into Light, Medium, and Heavy as well as the FAA classification of aircraft into Small, Small Plus, Medium, Large, and Heavy are perhaps too gross for PIREP classification. In fact, using the distribution of PIREPs and some initial feedback from pilots, a more applicable classification may be determined. With this in mind, the proposed classification for this work is described in Table 2.

Table 2. Proposed Classification of Aircraft Types

Class	Label	Max Takeoff Weight (lb)	% of PIREPs
1	Light	$\leq 5,000$	4.7%
2	Small	$\leq 15,000$	18.9%
3	Small Plus	$\leq 25,000$	10.5%
4	Medium	$\leq 100,000$	12.3%
5	Large	$\leq 250,000$	33.4%
6	Heavy	$> 250,000$	14.7%
7	Unknown	Unknown	4.6%

IV. Compression

In order to minimize the size of each compressed PIREP transmission as much as possible, a lossy compression approach is adopted. As will be shown below, if a standard lossless approach (like, say, gzip) is used, the achieved compression rate would not be suitable for sending a reasonable number of PIREPs in a single data message via SMS or SBD. Hence the decision for a lossy approach.

The size of each PIREP can be decreased when some redundant time data is extracted to a header for a

American Institute of Aeronautics and Astronautics

group of related PIREPs. We begin by describing the high-level message format, then the message header itself, next the structure of the individual PIREPs, and finally, the approach to compressing each of the TEIs (as described in Section II.A) contained within each PIREP. It is worth noting at this point that many of the values presented in this work are malleable. If, in the future, any of the specific fields would require more or fewer bits (binary digits) to describe it, those changes could be easily made. For proof-of-concept, the values chosen here provide a basis for initial analysis and should not be taken as a final recommendation of parameters to be used in an operational system.

IV.A. Message and Main Header Format

The overall message is composed of a header common to all PIREPs together with some number of PIREPs (see Figure 2). This main header will contain exactly three fields. The first is the time, the second is the day, and the third is a number representing the count of PIREPs contained in the message. See Figure 3 for a visual description.

Figure 2. Overall format for a set of n compressed PIREPs.

The time field will be a count of 10-minute time steps (or ticks) since 0000Z. There are 144 10-minute time bins in a 24-hour period, so eight bits is the minimum number of bits to describe time in this way. Note the loss in precision that occurs by keeping time in such a manner. We will discuss in Section V the effect of such precision loss. This time field represents the earliest time that any PIREP within the message occurs. For example if the earliest PIREP in a message occurs at 0121Z, then this time field will contain the value '8' representing the number of 10-minute ticks that occur from 0000Z to 0121Z. All PIREPs in this message will have its own time field that will represent an offset from this main header time. If all PIREPs occur within, say, the last few hours, then the number of bits to represent the time for each individual PIREP is much less than the 8 needed for the main header's time field.

The day field is used to ensure the correct interpretation of the time given that the message may span more than one day. For instance the message could be received on Tuesday morning, but the PIREPs may have been reported on Monday night. Since there are 7 possible values for the day, 3 bits are necessary for encoding it. If the delivery system for these messages could guarantee timely delivery of the message, then this field could likely be eliminated. Timely delivery is *not* guaranteed with SMS, for example, so it is possible for a message to be delivered much later (potentially hours) than when it was actually sent.

The final field is the count of PIREPs included in the message. For the experimentation in this paper, the main header was targeted to be 16 bits long, thus the count field is 5 bits (time bits + day bits + count bits = 16). That provides for a maximum of 31 PIREPs within any one message. Recall that these values could be changed in the future based on feedback from users and developers of an operational system. For example, if a larger number of messages were expected to be included in a message, the number of bits in the count field (and the overall size of the header) could be increased.

Figure 3. Format of the main header for a compressed PIREP message.

IV.B. Compressed PIREP Format

Each compressed PIREP will contain a PIREP header and a PIREP payload. The PIREP header contains 6 fields: the size of the payload (in bits), a field of 2 bits to indicate if the PIREP is UUA and/or from a Skyspotter-certified reporter (AWC), the flight level of the reporting aircraft using 9 bits (flight levels 0-511), the type of aircraft (by weight) in 3 bits, the time of the reported phenomenon in a count of 10-minute ticks

American Institute of Aeronautics and Astronautics

from the main header's time using 5 bits, and the latitude-longitude position of the reported phenomenon using 35 bits (as described below). These fields are summarized in Table 3 and Figure 4.

Table 3. Summary of Compressed PIREP Header Elements

Field	Size (bits)	Description
Payload Size	8	An integer indicating the number of bits in the PIREP payload.
Flags	2	Bits to indicate if PIREP is UUA and/or Skywatch-reported.
Flight Level	9	Integer representing flight level (100's of feet).
Type	3	Code representing the weight class of the aircraft.
Time	5	Number of 10-minute ticks from the Main Header time.
Position	35	A geohash value indicating the horizontal position of the aircraft.

Payload Size: 8 bits	Flags: 2 bits	Flight Level: 9 bits	Type: 3 bits	Time: 5 bits	Position: 35 bits

Figure 4. Compressed PIREP Header Fields.

The position field deserves further description. For this concept, we have chosen to encode the position as a geohash.[8] The geohash algorithm uses a nested bisection approach to indicate 2D position on the globe. Each bisection is encoded as a bit. For example, the first bit describing the longitudinal bisection indicates whether the position is in the eastern or western hemisphere. Depending on the resolution required, further bisections provide greater precision in the position. After the series of longitudinal and latitudinal bisections are complete, one is left with an ordered set of bits for each direction. Those sets are then interleaved for final encoding. By interleaving the sets of bits, one is able to truncate the resulting interleaved set to sacrifice resolution, but not accuracy. For this work, we chose 35 total position bits: 17 for latitude and 18 for longitude. This allows for less than 0.1 km error in either direction. By decreasing to 30 bits, we could still have sub-km precision, which still seems reasonable for the reporting of aviation weather phenomena.

The organization of the compressed PIREP payload mirrors the data in a standard, text-based PIREP. There are one or more elements with a designator for the type of element followed by the data associated with that element. Since there are 6 weather-related TEIs, a field of 3 bits will allow for designating which type of element is being described. For each TEI, we will describe the approach to compressing them, but first we discuss how flight level ranges are compressed. This is an important topic since these ranges can be part of the IC, TB, WX, and SK elements.

Flight levels and flight level ranges appear in several elements of a PIREP. As such, encoding altitudes for correctness and efficiency takes on significant importance in our compression scheme. Note that flight levels are altitudes rounded to the nearest 100 feet. This means that turbulence encountered between altitudes 3120 feet and 4000 feet would be described in a PIREP as occurring between flight levels 031 and 040. Our initial concept proposal was to use 1000 foot steps instead of 100 to decrease the number of bits needed to describe the altitudes. Informal feedback from pilots indicated that this would be too inaccurate to use in PIREPs.

Note that using 8 bits only allows for describing numbers between 0 and 256, but flight levels for PIREPs may range up to 400+. To describe all potential flight levels that might occur in a PIREP, 9 bits are required, allowing for the description of flight levels from 0 through 511. This implies that for flight level ranges, we could encode them completely with two 9-bit fields. However, analysis of the 25,265 PIREPs described in Section III shows that the ranges are seldom large enough to require 9 bits to describe. More specifically, for IC, TB, and SK elements (all of which may have flight level ranges), over 98% of the instances of IC, TB, and SK elements in our data set with flight level ranges had a difference in altitudes of less than 12000 ft or 120 flight levels. See Figure 5 for illustration.

Knowing that the range in flight level values can be described more efficiently than the actual values, we encode the base flight level and the spread instead of both flight levels. The spread is described with 7 bits. This allows for the encoding of values from 0-120, while saving values 121-127 for certain flags useful in decoding special altitude cases. These might include altitude ranges where only the "TOP" value is given or when the altitude is listed as "UNKN". Thus, for all PIREP elements needing to describe altitude ranges,
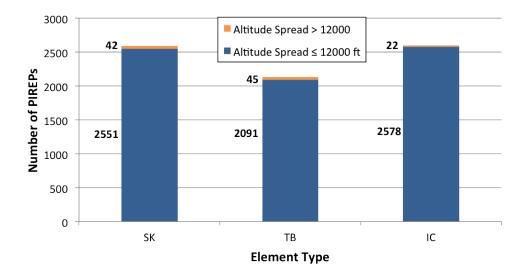
American Institute of Aeronautics and Astronautics

Figure 5. Analysis of PIREP elements with altitude ranges showing over 98% have a spread of less than 12000 ft.

a total of 16 bits is always used. In the cases where the spread between the lower and upper flight level is greater than 120, we encode it as 120. In an operational system, we would be able to deal with large altitude ranges by including multiple instances of a given element.

We now describe the compression of the individual weather elements. It may be useful to reference Figure 7, which summarizes the fields and their sizes, as the following sections are read.

IV.B.1. Turbulence Element

Turbulence (TB) is described with 27 or 11 bits depending on whether a flight level range is specified or not, respectively. We use three bits to describe the intensity. This allows for up to 8 different intensities to be specified. For this work, the following values were specified: LGT, LGT-MOD, MOD, MOD-SEV, SEV, SEV-EXTRM, EXTRM, and NEG. Note that this does not allow for specifying a range across non-adjacent values such as LGT-SEV. If a range is provided, then the two values are almost always adjacent, so this simplifying assumption seems safe. We use one bit each for CAT and CHOP. Then the duration is described with two bits to encode the values CONT (continuous), OCNL (occasional), INTMT (intermittent), or 'none' if no duration was supplied. Finally, one bit denotes whether flight level information is provided. If it is given, an additional 16 bits describe that flight level information.

IV.B.2. Air Temperature Element

The air temperature (TA) element is another relatively easy element to encode. We use 7 bits to represent the values of -85 to 42, which encompasses all practical aviation air temperatures. Specifically, in the data described in Section III, the maximum temperature reported is 28°C, while the minimum is -76°C. Some of the lower extreme temperatures that would not be typically reported (like -85 or -84) could be used as flags for values such as "UNKNOWN", which is sometimes reported. There is no loss of precision with this encoding unless extreme temperatures needed to be "rounded" to the limits of the encoding range. For example, a reported temperature of -89 would be forced to -85 and a temperature of 45 would be roused to 42. To encode values from the range -85 to 42, we simply add an offset value of 85 to the temperature and store a value from 0-127 (the non-negative integers described by 7 bits). To decode, we would then reverse the operation and subtract 85 from the encoded number to put the value back to the range of -85 to 42. Together with the 3 bits designating this a TA element, a total of 10 bits is sufficient to describe any TA element. Further feedback could be used to reduce the number of bits describing the temperature. For

American Institute of Aeronautics and Astronautics

example, there may not be much operational difference between 10°C and 13°C or between -20°C and -17°C. In that case, instead of encoding the value in degrees, it may make more sense to encode the temperature into some set of classes of temperature to reduce the number of bits needed. For 8 temperature classes, only 3 bits would be required, thus reducing the number of bits needed for a TA element from 10 to 6.

IV.B.3.   Icing Element

The icing element (IC) element is 25 or 9 bits depending on whether flight levels are provided or not, respectively. After the 3 bits for the IC identifier, 3 more bits describe the intensity of the icing as TRACE, TRACE-LGT, LGT, LGT-MOD, MOD, MOD-SEV, SEV, or NEG. Two bits then describe the type of icing: 1 bit for clear, 1 bit for rime. If both bits are active, then the interpretation is mixed icing. The next bit indicates whether there flight level information is provided. If it is provided, then the next 16 bits describe the flight level information as described above.

IV.B.4.   Sky Condition Element

The sky condition element (SK) is described with a constant value of 32 bits. Along with the three bits for designating the element as SK, we use a one bit flag to denote whether the "SKC" (sky clear above) is indicated. Then we have two 3-bit fields for describing the cloud cover range. The cloud cover takes one of 7 values as described in Section II.A.4. The pilot may report these as a range, thus necessitating the use of two 3-bit fields. Thus, the pilot can report OVC-SCT cloud cover or just BKN. In the latter case, both fields are still present in the compressed data, but the second is encoded as empty. Finally, 16 bits are used to describe the supplied flight level range.

IV.B.5.   Wind Vector Element

The wind vector (WV) element is one of the easier to describe. Its data are typically well-formatted, i.e. there is little to no anomalous data reported in a WV TEI. Since the general direction of the wind versus its precise direction, we generalize the direction of the wind to be one of 16 values conforming to the 16 major compass rose directions. Since there are 22.5° between each of the 16 directions, there is a maximum error of 11.25° using this encoding. For the wind speed, we can use 9 bits to represent anywhere from 0 to 511 knots which covers all practical wind speeds. Specifically, in the data described in Section III, besides an outlier of 411 knots, the maximum wind speed is 255 knots. Thus, for encoding a WV element, we use 4 bits for the direction and 9 for the speed. Together with the 3 bits for designating this a WV element, a total of 16 bits is sufficient for describing any WV at a cost of 11.25° in precision. We note that future refinements to this compression approach may indicate the need for better wind direction precision. In that case, each additional bit used to encode the direction would halve the error in precision. Likewise, if less precision were deemed to be acceptable, each fewer bit would double the error in precision.

IV.B.6.   Weather Element

The weather element (WX) is slightly different from the other TEIs since it technically has two optional data items. Those are visibility and weather. To start, we specify one bit to indicate if there are visibility data and one bit to indicate if there are weather data. If there is visibility data, it will be described with 7 bits allowing for description of values 0-99 without loss of information. Then, if there are weather data they is described with 8 bits. The first two bits provide the intensity ('-', '+', or no modifier) and the remaining 6 bits are an index to a look-up table of weather phenomena. Those phenomena are shown in Figure 6 as duplicated from the FAA source document.[1] Since 6 bits is sufficient for describing up to 63 different phenomena, we also encode "IMC", "VMC", and "CLR/CLEAR" since these are often reported in the WX TEI.

American Institute of Aeronautics and Astronautics

Type	METAR Code
Drifting / Blowing Snow	DRSN/BLSN
Drifting Dust	DRDU
Drifting Sand	DRSA
Drizzle/Freezing Drizzle	DZ/FZDZ
Dust / Blowing Dust	DU/BLDU
Duststorm	DS
Fog (vis < 5/8SM)	FG
Freezing Fog	FZFG
Freezing Rain	FZRA
Funnel Cloud	FC
Hail (aprx 1/4" dia or more)	GR
Hail Shower	SHGR
Haze	HZ
Ice Crystals	IC
Ice Pellets/ Showers	PL/SHPL
Mist (vis 5/8SM or more)	BR
Patchy Fog	BCFG
Patchy Fog on part of Arpt	PRFG
Rain / Showers	RA/SHRA
Sand / Blowing Sand	SA/BLSA
Sandstorms	SS
Shallow Fog	MIFG
Sml Hail/Snow Pellet Showers	SHGS
Sml Hail/Snow Pellets	GS
Smoke	FU
Snow Grains	SG
Snow / Showers	SN/SHSN
Spray	PY
Squalls	SQ
Thunderstorm	TS
Tornado/Waterspout	+FC
Unknown Precipitation	UP
Volcanic Ash	VA
Well developed Dust/Sand Whirls	PO

Figure 6. FAA-defined weather phenomena for use in the WX TEI.

IV.C. Summary of Compression Concept

Figure 7 summarizes all of the compression fields and sizes described above for each TEI. We emphasize once again that the encoding choices described in this paper are reasonable, but not necessarily the best for implementation. To perfect the concept, subject matter experts including FAA officials, pilots, communication engineers, and others would provide feedback on each of the choices described herein. Changes to individual field encodings would not detract from the overall concept in any significant way. It is also important to note that we lose all information in the RM element if it is provided in the PIREP. This decision should be re-examined in future work. Much of the information in the RM element might be deemed irrelevant or could be encoded more appropriately in one of the main weather elements. Some of it, however, is important, like indications of low-level wind shear and descriptions of funnel clouds. It might be that these items deserve flags in the PIREP header like UUA and skyspotters, even though they happen infrequently.

V. Analysis

This section is provided in two distinct parts. First, we provide some analysis of the parsing of text PIREPs, which is an important step in the compression process. Second, we analyze the performance of the compression concept in relation to other compression techniques.
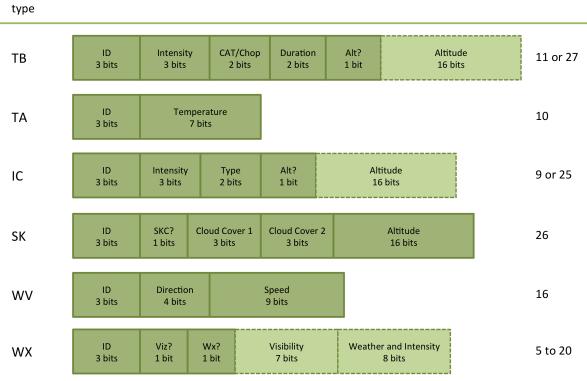
American Institute of Aeronautics and Astronautics

Element type	Formatting (lighter shading indicates an optional field)	Size (bits)

TB	ID 3 bits	Intensity 3 bits	CAT/Chop 2 bits	Duration 2 bits	Alt? 1 bit	Altitude 16 bits	11 or 27

(Table rendered as figure)

Figure 7. Visual summary of compression for each TEI.

V.A. Parsing PIREPs

While the specification for each of the TEIs is clear, there is a significant amount of data included in PIREPs that do not conform to the specification. The reason for this can be traced to the point of data entry by the FSS. The tools used by FSSs to enter PIREP data do not uniformly enforce correct formatting of the data. As such, "messy" data resides in the PIREP database. In practice, this is not seen as a major problem due to the fact that PIREPs are intended for interpretation by humans for voice communication to other humans. However, correctly formatted data has an important impact on any potential machine-to-machine communication, as is proposed in this work. The best solution, in our view, would be to improve the data entry tools to enforce "legal" entries into the database. Until that occurs, it is incumbent upon users of the data (like us during this work) to parse the input data as well as possible.

We implemented a parser to handle the various PIREP elements, including some of the more common data entry anomalies. We will outline some of those anomalies below. A PIREP element is considered "correctly parsed" using the following criteria. First read in the original element. Extract the data from the element. From the extracted data, create a new element. Compare the two elements for matching within the precision bounds of the compression algorithm. Sometimes the original element and the processed element differ, but they differ in an understandable way. For instance, the original element may have used "MDT" to specify a phenomenon of moderate intensity, but "MOD" is the correct specifier. If our parser correctly identified the intent of the element, we consider it correctly parsed. However, if the element contains some text that is not understood at all, then that element would be considered incorrectly parsed. Some examples of correctly parsed elements that contain non-conforming data are provided in Table 4, and some examples of elements that are not recognized or incorrectly parsed are provided in Table 5. Overall performance of our parser is shown in Figure 8.
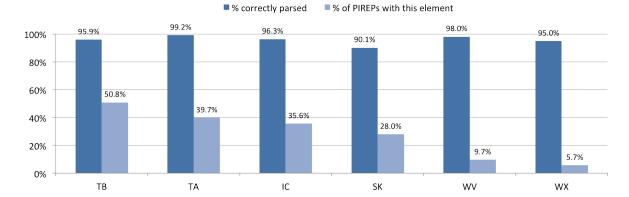
American Institute of Aeronautics and Astronautics

Figure 8. Counts of the various elements in the experimental data set and the percentage correctly parsed.

Table 4. Examples of Incorrectly Formatted PIREP Elements that are Correctly Parsed

Actual Element	Interpreted Element	Reason for Difference
/TB SMTH	/TB NEG	"SMTH" refers to "smooth" and is not in specification.
/TB CONS MOD CHOP	/TB CONT MOD CHOP	"CONT" is correct description of continuous turbulence.
/TB LGT-MOD TURB	/TB LGT-MOD	Trailing "TURB" is superfluous and not in specification.
/TB MDT	/TB MOD	"MOD" is correct description of moderate turbulence.
/IC LGT RIME ICE	/IC LGT RIME	Trailing "ICE" is superfluous and not in specification.
/IC NIL	/IC NEG	"NEG" is correct description of lack of icing.
/IC MDT MXD 020-050	/IC MOD MX 020-050	"MOD" and "MX" are correct forms of moderate and mixed.
/TA -6	/TA M06	Negative temperatures specified with "M", temps use 2 digits.
/TA UNKWN	/TA UNKN	"UNKN" is correct description of unknown temperature.
/TA 8	/TA 08	Temperatures use two digits.
/SK OVC027-TOP065/SKC	/SK OVC027-TOP065/ SKC	Space required before trailing "SKC".

Table 5. Examples of Incorrectly Formatted PIREP Elements that are Inorrectly Parsed

Actual Element	Reason for Parsing Error
/TB LGT SFC - 020	Incorrect altitude range formatting.
/TB LGT- MOD CHOP	Extra spacing in intensity description.
/TB MOD IN SECTOR 4/ZLC	Completely unspecified remarks.
/IC SEV 060-030	Icing type is a required field and is missing.
/IC LGT C LR	Extra spacing in icing type.
/IC LGT TO MOD RIME FL240-FL270	Incorrect intensity range formatting.
/IC MOD RIME BTN FL180-FL190	Superfluous "BTN".
/TA MSG	"MSG" completely unspecified.
/TA M54 TO M64	Ranges not allowed in TA elements.
/SK 120BKN	Transposed cloud cover and altitude.
/SK 020 OVC TOP 032	Incorrect altitude range formatting.

V.B. Compression Performance

Examination of the compression performance versus standard compression tools is shown in Figure 9. Recall that all tools to which we are comparing our compression technique are *lossless* and ours is *lossy*. It is atypical to use lossy compression on text files (usually lossy compression is reserved for audio and visual files), but in this domain, it does seem reasonable. Note that lower compression ratio values are better,

American Institute of Aeronautics and Astronautics

and that the custom compression approach maintains a relatively constant value. This is due to the overall statistics of PIREPs as a whole. For example, over a large number of PIREPs, their respective sizes will be relatively similar, thus each PIREP takes up about the same number of bits. Also note that the lossless approaches improve in their ability to compress the PIREPs as more and more are compressed. This is due to those algorithms being better able to take advantage of statistical redundancy within the PIREPs at a lower relative cost to the overhead of bookkeeping that redundancy. If this concept was being developed to deliver a large number of PIREPs (1000's) at any given time, it may make more sense to use a traditional, lossless approach. However, pilots seldom need 1000's of PIREPs at any given time. Rather, based on their current location, 10-20 would seem a reasonable number to provide the best PIREP information available to help the pilot make decisions while en route.
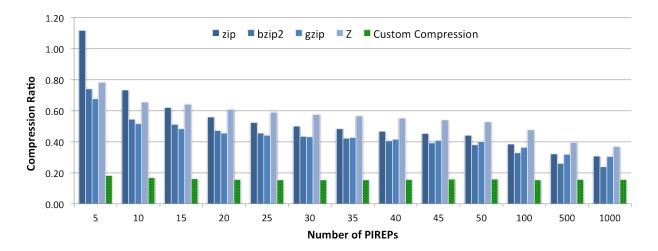

Figure 9. Comparison of compression approaches on varying numbers of PIREPs. Lower values are better.

V.C.   Complete Compression Example

Using the PIREPs described in Section II.A.7, we now provide an example of the compression concept. For those 4 PIREPs, we will show the generation of the main header as described in Section IV.A and the complete generation of bits for the third PIREP.

To construct the main header, a time must be chosen that is no later than the earliest PIREP time stamp and occurs at a 10-minute mark. We must also verify that all PIREPs were reported within 310 minutes of the time in the main header. In this case the PIREPs were reported at 0013Z, 0027Z, 0236Z, and 0107Z, so a main header time of 0010Z would be appropriate. Since time is reported as the number of 10-minute steps since 0000Z, the time in this main header would be stored in binary as 00000001. Let us assume the day of the reports is Wednesday, which could be encoded as 011. Finally, the main header would contain the count of PIREPs in the complete message, which would be 4, or 00100. Thus, the main header would be 0000000101100100.

For the PIREP

UA /OV FSD/TM 0236/FL100/TP PAT4/SK UNKN050-TOP067/TA M08/IC LGT MX

we provide a complete example of how it would be encoded using the presented scheme. For reference, the encoding is summarized below in Table 6.

To begin, we can determine the bits for the header. The payload for this PIREP contains SK, TA, and IC elements. SK will take 26 bits, TA will take 10 bits, and IC (without an altitude range) will take 9 bits for a total payload size of 45 bits. The 8-bit payload field will then contain 00101101. Since this is not an urgent PIREP and was not reported by a certified skywatcher, the flags for those fields will be set to 0's (00). The flight level is reported as 100, so the 9-bit field for flight level will be 001100100. This is a small aircraft, so

American Institute of Aeronautics and Astronautics

we encode the type as 000. Recall that the time field for a PIREP is given relative to the time in the main header. Since the time was reported as 0236Z, and the main header has a time of 0010Z, we can calculate 14 10-minute time steps to this report time. So the time 5-bit time field for this PIREP will contain the binary description of 14: 01110. Finally, the position of the aircraft at the time of this report is provided to the geohash algorithm to determine the appropriate 35 bits to describe it within 0.1km. in this case, FSD is Sioux Falls Regional Airport, which is located at 43° 34' 55" N, 96° 44' 31" W. This translates into a geohash value of 01001111110110110101111101100101110.

As for the actual weather elements, we start with the SK element. To flag the cloud cover as unknown, we use a value of 101 in the first cloud cover field and a value of 111 in the second cloud cover field to indicate there is only one cloud cover value provided. The altitude range of 050-067 is stored with 9 bits for the base of 050 (000110010) and 7 bits to store the difference in altitudes of 17 (0010001). With a prepended value of 001 to indicate that this is a SK element, and a 0 to indicate the lack of a "sky clear" from the pilot, we describe the complete SK element as 00101011110001100100010001.

The TA element is simply stored as the binary value for -8 plus the offset of 85 (as described in Section IV.B.2), which is 77. Together with the three bits to indicate this is a TA element (101), we have 1011001101.

The IC element is indicated with the preceding three bits of 100. The intensity of LGT is described by 010 and the type (MX) is given by 11. Since there is not a flight level range provided, the altitude flag is given as 0. The ultimate bit string for this IC element is then 100010110.

Table 6. Encoding "UA /OV FSD/TM 0236/FL100/TP PAT4/SK UNKN050-TOP067/TA M08/IC LGT MX"

Field	Text	Bits	Encoding	Comments
UA	-	1	0	Not an urgent (UUA) PIREP.
AWC	-	1	0	Not reported by a certified Skyspotter.
OV	FSD	35	01001111...100101110	Geohash encoding of FSD, decoded as lat-lon location.
TM	0236	5	01110	14 time steps from main header time, decodes to 0230.
FL	100	9	001100100	Binary encoding of 100, decoded value matches original.
TP	PAT4	3	000	Size class for a Piper T-1040, decoded as "light" aircraft.
SK	UNKN050-TOP067	26	00101011110001100100010001	Decoded value matches original.
TA	M08	10	1011001101	-8 (temp) + 85 (offset) = 77, decoded value matches original.
IC	LGT MX	9	100010110	Decoded value matches original.

Putting all of the bits for the main header and the 4 PIREPs would then constitute a single message ready for transmitting to pilots. The total size of this message would be 373 bits (16 for main header, 81 for first PIREP, 75 for second, 107 for third, and 94 for the fourth). This is well under the limit for a standard SMS message, but the original, uncompressed 4 PIREPs would require two SMS messages to send in their entirety. A mobile-based application on a phone, tablet, laptop, or other commercial hardware for example, could then interpret and display this data in a useful way to the pilot after the transmission was received. Note that the data loss through the compression-decompression cycle for the example PIREP includes only the exact aircraft type (PAT4 becomes "Light") and a 6-minute discrepancy in reporting time (0236 becomes 0230), which seems reasonable for the increased transmission efficiency.

VI.   Conclusions and Future Directions

In this paper we presented a method for compressing pilot weather reports (PIREPs) so that those reports could be efficiently transmitted over throughput-limited channels. Example channels might include standard Short Message Service (text) messages or satellite data services for short data messages. The approach achieves a compression ratio much greater than standard, lossless text compression tools at the cost of some loss of information. This loss seems operationally reasonable and is open for modification in the future by using slightly more bits to encode the data if deemed reasonable by stakeholders.

There is potential for use of this approach in areas where there may be difficulty contacting flight services before departure or en route. For instance, a satellite-based system may periodically broadcast or multi-cast

American Institute of Aeronautics and Astronautics

PIREP data to a particular geographic region giving pilots who are listening access to better information for decision making.

Acknowledgements

References

[1]U.S. Department of Transportation, *Flight Services – JO 7110.10V*, Federal Aviation Administration, Washington, DC, July 2012.

[2]U.S. Department of Transportation, *Contractions – JO 7340.2C*, Federal Aviation Administration, Washington, DC, February 2012.

[3]NOAA's National Weather Service, "Aviation Digital Data Service," http://www.aviationweather.gov/adds/pireps, February 2013.

[4]Witten, I. H., Neal, R. M., and Cleary, J. G., "Arithmetic Coding for Data Compression," *Communications of the ACM*, Vol. 30, No. 6, 1987, pp. 520–540.

[5]Huffman, D. A., "A Method for the Construction of Mimimum-Redundancy Codes," *Proceedings of the Institute of Radio Engineers*, September 1952.

[6]Salomon, D., *Data Compression: The Complete Reference*, Springer, 4th ed., 2006.

[7]Ziv, J. and Lempel, A., "Compression of Individual Sequences via Variable-Rate Coding," *IEEE Transactions on Information Theory*, Vol. 24, No. 5, 1978, pp. 530–536.

[8]Niemeyer, G., "Geohash," http://en.wikipedia.org/wiki/Geohash, Feb 2013.