

GEOPRIV	M. Thomson
Internet-Draft	J. Winterbottom
Intended status: Informational	Andrew Corporation
Expires: September 11, 2011	March 10, 2011

Representation of Uncertainty and Confidence in PIDF-LO

[Abstract](#)

The key concepts of uncertainty and confidence as they pertain to location information are defined. Methods for the manipulation of location estimates that include uncertainty information are outlined.

[Status of this Memo](#)

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 11, 2011.

[Copyright Notice](#)

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

[Table of Contents](#)

- *1. [Introduction](#)
- *1.1. [Conventions and Terminology](#)
- *2. [A General Definition of Uncertainty](#)
- *2.1. [Uncertainty as a Probability Distribution](#)

- *2.2. [Deprecation of the Terms Precision and Resolution](#)
- *2.3. [Accuracy as a Qualitative Concept](#)
- *3. [Uncertainty in Location](#)
- *3.1. [Representation of Uncertainty and Confidence in PIDF-L0](#)
- *3.2. [Uncertainty and Confidence for Civic Addresses](#)
- *3.3. [DHCP Location Configuration Information and Uncertainty](#)
- *4. [Manipulation of Uncertainty](#)
- *4.1. [Reduction of a Location Estimate to a Point](#)
- *4.1.1. [Centroid Calculation](#)
- *4.1.1.1. [Arc-Band Centroid](#)
- *4.1.1.2. [Polygon Centroid](#)
- *4.2. [Conversion to Circle or Sphere](#)
- *4.3. [Three-Dimensional to Two-Dimensional Conversion](#)
- *4.4. [Increasing and Decreasing Uncertainty and Confidence](#)
- *4.4.1. [Rectangular Distributions](#)
- *4.4.2. [Normal Distributions](#)
- *4.5. [Determining Whether a Location is Within a Given Region](#)
- *4.5.1. [Determining the Area of Overlap for Two Circles](#)
- *4.5.2. [Determining the Area of Overlap for Two Polygons](#)
- *5. [Examples](#)
- *5.1. [Reduction to a Point or Circle](#)
- *5.2. [Increasing and Decreasing Confidence](#)
- *5.3. [Matching Location Estimates to Regions of Interest](#)
- *6. [Security Considerations](#)
- *7. [Acknowledgements](#)
- *8. [References](#)

*Appendix A. [Conversion Between Cartesian and Geodetic Coordinates in WGS84](#)

*Appendix B. [Calculating the Upward Normal of a Polygon](#)

*Appendix B.1. [Checking that a Polygon Upward Normal Points Up](#)

*[Authors' Addresses](#)

[1. Introduction](#)

Location information represents an estimation of the position of a Target. Under ideal circumstances, a location estimate precisely reflects the actual location of the Target. In reality, there are many factors that introduce errors into the measurements that are used to determine location estimates.

The process by which measurements are combined to generate a location estimate is outside of the scope of work within the IETF. However, the results of such a process are carried in IETF data formats and protocols. This document outlines how uncertainty, and its associated datum, confidence, are expressed and interpreted.

This document provides a common nomenclature for discussing uncertainty and confidence as they relate to location information.

This document also provides guidance on how to manage location information that includes uncertainty. Methods for expanding or reducing uncertainty to obtain a required level of confidence are described. Methods for determining the probability that a Target is within a specified region based on their location estimate are described. These methods are simplified by making certain assumptions about the location estimate and are designed to be applicable to location estimates in a relatively small area.

[1.1. Conventions and Terminology](#)

This document assumes a basic understanding of the principles of mathematics, particularly statistics and geometry.

Some terminology is borrowed from [\[RFC3693\]](#) and [\[I-D.ietf-geopriv-arch\]](#).

Mathematical formulae are presented using the following notation: add +, subtract -, multiply *, divide /, power ^ and absolute value |x|. Precedence is indicated using parentheses. Mathematical functions are represented by common abbreviations: square root sqrt(x), sine sin(x), cosine cos(x), inverse cosine acos(x), tangent tan(x), inverse tangent atan(x), error function erf(x), and inverse error function erfinv(x).

[2. A General Definition of Uncertainty](#)

Uncertainty results from the limitations of measurement. In measuring any observable quantity, errors from a range of sources affect the result. Uncertainty is a quantification of what is known about the

observed quantity, either through the limitations of measurement or through inherent variability of the quantity.

Uncertainty is most completely described by a probability distribution. A probability distribution assigns a probability to possible values for the quantity.

A probability distribution describing a measured quantity can be arbitrarily complex and so it is desirable to find a simplified model. One approach commonly taken is to reduce the probability distribution to a confidence interval. Many alternative models are used in other areas, but study of those is not the focus of this document.

In addition to the central estimate of the observed quantity, a confidence interval is succinctly described by two values: an error range and a confidence. The error range describes an interval and the confidence describes an estimated upper bound on the probability that a "true" value is found within the extents defined by the error.

In the following example, a measurement result for a length is shown as a nominal value with additional information on error range (0.0043 meters) and confidence (95%). x between 1.00312 and 1.01172 meters with 95% probability. No other assertion is made: in particular, this does not assert that x is 1.00742.

*e.g. $x = 1.00742 \pm 0.0043$ meters at 95% confidence

This result indicates that the measurement indicates that the value of x is 1.00742 meters with a 95% probability. This document uses the term *uncertainty* to refer in general to the concept as well as more specifically to refer to the error increment. Uncertainty and confidence for location estimates can be derived in a number of ways. This document does not attempt to enumerate the many methods for determining uncertainty. [\[ISO.GUM\]](#) and [\[NIST.TN1297\]](#) provide a set of general guidelines for determining and manipulating measurement uncertainty. This document applies that general guidance for consumers of location information.

2.1. Uncertainty as a Probability Distribution

The Probability Density Function (PDF) that is described by uncertainty indicates the probability that the "true" value lies at any one point. The shape of the probability distribution can vary depending on the method that is used to determine the result. The two probability density functions most generally applicable most applicable to location information are considered in this document:

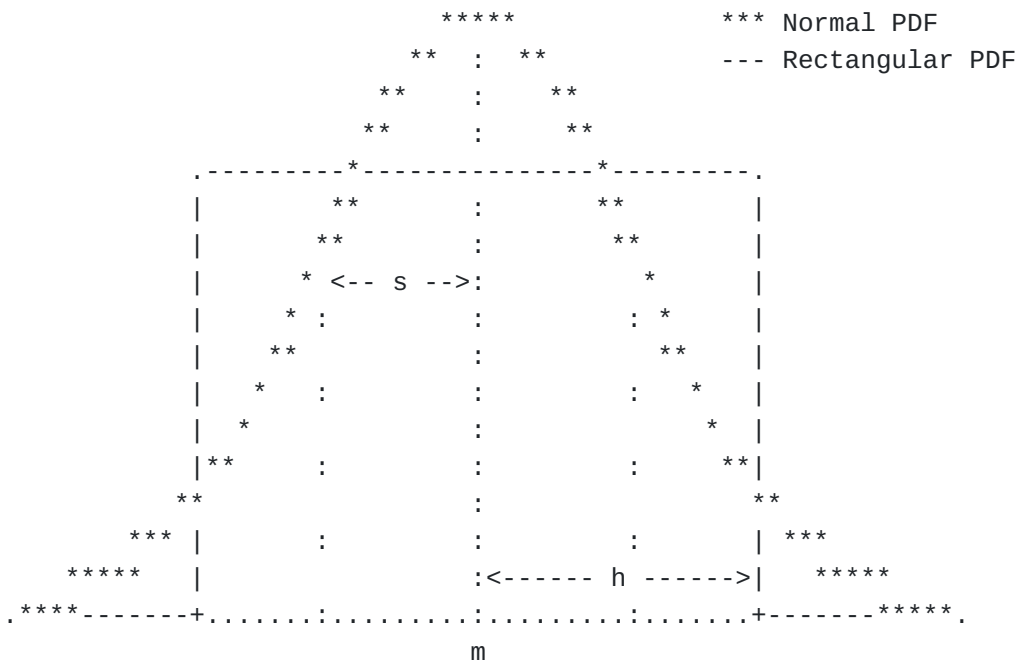
*The normal PDF (also referred to as a Gaussian PDF) is used where a large number of small random factors contribute to errors. The value used for the error range in a normal PDF is related to the standard deviation of the distribution.

*A rectangular PDF is used where the errors are known to be consistent across a limited range. A rectangular PDF can occur

where a single error source, such as a rounding error, is significantly larger than other errors. A rectangular PDF is often described by the half-width of the distribution; that is, half the width of the distribution.

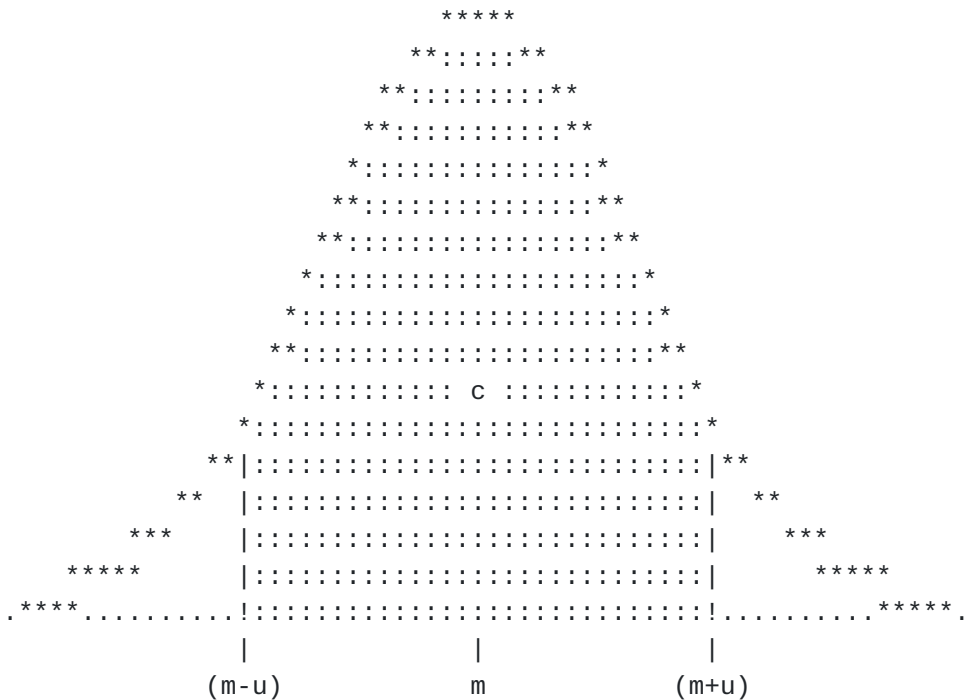
Each of these probability density functions can be characterized by its center point, or mean, and its width. For a normal distribution, uncertainty and confidence together are related to the standard deviation (see [Section 4.4](#)). For a rectangular distribution, half of the width of the distribution is used.

[Figure 1](#) shows a normal and rectangular probability density function with the mean (m) and standard deviation (s) labelled. The half-width (h) of the rectangular distribution is also indicated.



For a given PDF, the value of the PDF describes the probability that the "true" value is found at that point. Confidence for any given interval is the total probability of the "true" value being in that range, defined as the integral of the PDF over the interval. [Figure 2](#) shows how confidence is determined for a normal distribution. The area of the shaded region gives the confidence (c) for the interval between $m-u$ and $m+u$.

*The probability of the "true" value falling between two points is found by finding the area under the curve between the points (that is, the integral of the curve between the points). For any given PDF, the area under the curve for the entire range from negative infinity to positive infinity is 1 or (100%). Therefore, the confidence over any interval of uncertainty is always less than 100%.



In [Section 4.4](#), methods are described for manipulating uncertainty if the shape of the PDF is known.

2.2. Deprecation of the Terms Precision and Resolution

The terms *Precision* and *Resolution* are defined in [RFC 3693](#) [RFC3693]. These definitions were intended to provide a common nomenclature for discussing uncertainty; however, these particular terms have many different uses in other fields and their definitions are not sufficient to avoid confusion about their meaning. These terms are unsuitable for use in relation to quantitative concepts when discussing uncertainty and confidence in relation to location information.

2.3. Accuracy as a Qualitative Concept

Uncertainty is a quantitative concept. The term *accuracy* is useful in describing, qualitatively, the general concepts of location information. Accuracy is generally useful when describing qualitative aspects of location estimates. Accuracy is not a suitable term for use in a quantitative context.

For instance, it could be appropriate to say that a location estimate with uncertainty X is more accurate than a location estimate with uncertainty $2X$ at the same confidence. It is not appropriate to assign a number to "accuracy", nor is it appropriate to refer to any component of uncertainty or confidence as "accuracy". That is, to say that the "accuracy" for the first location estimate is X would be an erroneous use of this term.

3. Uncertainty in Location

A *location estimate* is the result of location determination. A location estimate is subject to uncertainty like any other observation. However, unlike a simple measure of a one dimensional property like length, a location estimate is specified in two or three dimensions.

Uncertainty in 2- or 3-dimensional locations can be described using confidence intervals. The confidence interval for a location estimate in two or three dimensional space is expressed as a subset of that space. This document uses the term *region of uncertainty* to refer to the area or volume that describes the confidence interval.

Areas or volumes that describe regions of uncertainty can be formed by the combination of two or three one-dimensional ranges, or more complex shapes could be described.

3.1. Representation of Uncertainty and Confidence in PIDF-LO

A set of shapes suitable for the expression of uncertainty in location estimates in the Presence Information Data Format - Location Object (PIDF-LO) are described in [\[GeoShape\]](#). These shapes are the recommended form for the representation of uncertainty in [PIDF-LO \[RFC4119\]](#) documents.

The PIDF-LO does not include an indication of confidence, but that confidence is 95%, by definition in [\[RFC5491\]](#). Similarly, the PIDF-LO format does not provide an indication of the shape of the PDF.

Absence of uncertainty information in a PIDF-LO document does not indicate that there is no uncertainty in the location estimate.

Uncertainty might not have been calculated for the estimate, or it may be withheld for privacy purposes.

If the Point shape is used, confidence and uncertainty are unknown; a receiver can either assume a confidence of 0% or infinite uncertainty. The same principle applies on the altitude axis for two-dimension shapes like the Circle.

3.2. Uncertainty and Confidence for Civic Addresses

[Civic addresses \[RFC5139\]](#) inherently include uncertainty, based on the area of the most precise element that is specified. Uncertainty is effectively defined by the presence or absence of elements -- elements that are not present are deemed to be uncertain.

To apply the concept of uncertainty to civic addresses, it is helpful to unify the conceptual models of civic address with geodetic location information.

Note: This view is one perspective on the process of geo-coding - the translation of a civic address to a geodetic location.

In the unified view, a civic address defines a series of (sometimes non-orthogonal) spatial partitions. The first is the implicit partition

that identifies the surface of the earth and the space near the surface. The second is the country. Each label that is included in a civic address provides information about a different set of spatial partitions. Some partitions require slight adjustments from a standard interpretation: for instance, a road includes all properties that adjoin the street. Each label might need to be interpreted with other values to provide context.

As a value at each level is interpreted, one or more spatial partitions at that level are selected, and all other partitions of that type are excluded. For non-orthogonal partitions, only the portion of the partition that fits within the existing space is selected. This is what distinguishes King Street in Sydney from King Street in Melbourne. Each defined element selects a partition of space. The resulting location is the intersection of all selected spaces.

The resulting spatial partition can be considered to represent a region of uncertainty. At no stage does this process select a point; although, as spaces get smaller this distinction might have no practical significance and an approximation if a point could be used.

Uncertainty in civic addresses can be increased by removing elements. This doesn't necessarily improve confidence in the same way that arbitrarily increasing uncertainty in a geodetic location doesn't increase confidence.

3.3. DHCP Location Configuration Information and Uncertainty

Location information is often measured in two or three dimensions; expressions of uncertainty in one dimension only are rare. The "resolution" parameters in [\[RFC3825\]](#) provide an indication of uncertainty in one dimension.

[\[RFC3825\]](#) defines a means for representing uncertainty, but a value for confidence is not specified. A default value of 95% confidence can be assumed for the combination of the uncertainty on each axis. That is, the confidence of the resultant rectangular polygon or prism is 95%.

4. Manipulation of Uncertainty

This section deals with manipulation of location information that contains uncertainty.

The following rules generally apply when manipulating location information:

- *Where calculations are performed on coordinate information, these should be performed in Cartesian space and the results converted back to latitude, longitude and altitude. A method for converting to and from Cartesian coordinates is included in [Appendix A](#).

- While some approximation methods are useful in simplifying calculations, treating latitude and longitude as Cartesian

axes is never advisable. The two axes are not orthogonal. Errors can arise from the curvature of the earth and from the convergence of longitude lines.

*Normal rounding rules do not apply when rounding uncertainty. When rounding, the region of uncertainty always increases (that is, errors are rounded up) and confidence is always rounded down (see [\[NIST.TN1297\]](#)). This means that any manipulation of uncertainty is a non-reversible operation; each manipulation can result in the loss of some information.

4.1. Reduction of a Location Estimate to a Point

Manipulating location estimates that include uncertainty information requires additional complexity in systems. In some cases, systems only operate on definitive values, that is, a single point. This section describes algorithms for reducing location estimates to a simple form without uncertainty information. Having a consistent means for reducing location estimates allows for interaction between applications that are able to use uncertainty information and those that cannot.

Note: Reduction of a location estimate to a point constitutes a reduction in information. Removing uncertainty information can degrade results in some applications. Also, there is a natural tendency to misinterpret a point location as representing a location without uncertainty. This could lead to more serious errors. Therefore, these algorithms should only be applied where necessary.

Several different approaches can be taken when reducing a location estimate to a point. Different methods each make a set of assumptions about the properties of the PDF and the selected point; no one method is more "correct" than any other. For any given region of uncertainty, selecting an arbitrary point within the area could be considered valid; however, given the aforementioned problems with point locations, a more rigorous approach is appropriate.

Given a result with a known distribution, selecting the point within the area that has the highest probability is a more rigorous method. Alternatively, a point could be selected that minimizes the overall error; that is, it minimises the expected value of the difference between the selected point and the "true" value.

If a rectangular distribution is assumed, the centroid of the area or volume minimizes the overall error. Minimizing the error for a normal distribution is mathematically complex. Therefore, this document opts to select the centroid of the region of uncertainty when selecting a point.

4.1.1. Centroid Calculation

For regular shapes, such as Circle, Sphere, Ellipse and Ellipsoid, this approach equates to the center point of the region. For regions of uncertainty that are expressed as regular Polygons and Prisms the center point is also the most appropriate selection.

For the Arc-Band shape and non-regular Polygons and Prisms, selecting the centroid of the area or volume minimizes the overall error. This assumes that the PDF is rectangular.

Note: The centroid of a concave Polygon or Arc-Band shape is not necessarily within the region of uncertainty.

4.1.1.1. Arc-Band Centroid

The centroid of the Arc-Band shape is found along a line that bisects the arc. The centroid can be found at the following distance from the starting point of the arc-band (assuming an arc-band with an inner radius of r , outer radius R , start angle a , and opening angle o): $a + (o/2)$. Negative values are possible if the angle of opening is greater than 180 degrees; negative values indicate that the centroid is found along the angle $a + (o/2) + 180$.

$$*d = 4 * \sin(o/2) * (R*R + R*r + r*r) / (3*o*(R + r))$$

This point can be found along the line that bisects the arc; that is, the line at an angle of

4.1.1.2. Polygon Centroid

Calculating a centroid for the Polygon and Prism shapes is more complex. Polygons that are specified using geodetic coordinates are not necessarily coplanar. For Polygons that are specified without an altitude, choose a value for altitude before attempting this process; an altitude of 0 is acceptable.

*The method described in this section is simplified by assuming that the surface of the earth is locally flat. This method degrades as polygons become larger; see [\[GeoShape\]](#) for recommendations on polygon size.

The polygon is translated to a new coordinate system that has an x-y plane roughly parallel to the polygon. This enables the elimination of z-axis values and calculating a centroid can be done using only x and y coordinates. This requires that the upward normal for the polygon is known.

To translate the polygon coordinates, apply the process described in [Appendix Appendix B](#) to find the normal vector $N = [N_x, N_y, N_z]$. This value should be made a unit vector to ensure that the transformation matrix is a special orthogonal matrix. From this vector, select two

vectors that are perpendicular to this vector and combine these into a transformation matrix.

If N_x and N_y are non-zero, the matrices in [Figure 3](#) can be used, given $p = \sqrt{N_x^2 + N_y^2}$. More transformations are provided later in this section for cases where N_x or N_y are zero.

$$T = \begin{bmatrix} -N_y/p & N_x/p & 0 \\ -N_x*N_z/p & -N_y*N_z/p & p \\ N_x & N_y & N_z \end{bmatrix} \quad T' = \begin{bmatrix} -N_y/p & -N_x*N_z/p & N_x \\ N_x/p & -N_y*N_z/p & N_y \\ 0 & p & N_z \end{bmatrix}$$

(Transform) (Reverse Transform)

To apply a transform to each point in the polygon, form a matrix from the ECEF coordinates and use matrix multiplication to determine the translated coordinates.

$$\begin{bmatrix} -N_y/p & N_x/p & 0 \\ -N_x*N_z/p & -N_y*N_z/p & p \\ N_x & N_y & N_z \end{bmatrix} * \begin{bmatrix} x[1] & x[2] & x[3] & \dots & x[n] \\ y[1] & y[2] & y[3] & \dots & y[n] \\ z[1] & z[2] & z[3] & \dots & z[n] \end{bmatrix}$$

$$= \begin{bmatrix} x'[1] & x'[2] & x'[3] & \dots & x'[n] \\ y'[1] & y'[2] & y'[3] & \dots & y'[n] \\ z'[1] & z'[2] & z'[3] & \dots & z'[n] \end{bmatrix}$$

Alternatively, direct multiplication can be used to achieve the same result:

$$\begin{aligned} *x'[i] &= -N_y * x[i] / p + N_x * y[i] / p \\ *y'[i] &= -N_x * N_z * x[i] / p - N_y * N_z * y[i] / p + p * z[i] \\ *z'[i] &= N_x * x[i] + N_y * y[i] + N_z * z[i] \end{aligned}$$

The first and second rows of this matrix (x' and y') contain the values that are used to calculate the centroid of the polygon. To find the centroid of this polygon, first find the area using: $x'[0] == x'[n]$ and $x'[n+1] == x'[1]$. Based on the area, the centroid along each axis can be determined by:

$$*A = \text{sum from } i=1..n \text{ of } (x'[i]*y'[i+1]-x'[i+1]*y'[i]) / 2$$

For these formulae, treat each set of coordinates as circular, that is

$$\begin{aligned} *Cx' &= \text{sum } (x'[i]+x'[i+1]) * (x'[i]*y'[i+1]-x'[i+1]*y'[i]) / (6*A) \\ *Cy' &= \text{sum } (y'[i]+y'[i+1]) * (x'[i]*y'[i+1]-x'[i+1]*y'[i]) / (6*A) \end{aligned}$$

Note: The formula for the area of a polygon will return a negative value if the polygon is specified in clockwise direction. This can be used to determine the orientation of the polygon.

The third row contains a distance from a plane parallel to the polygon. If the polygon is coplanar, then the values for z' are identical; however, the constraints recommended in [RFC5491](#) mean that this is rarely the case. To determine Cz' , average these values:

$$*Cz' = \text{sum } z'[i] / n$$

Once the centroid is known in the transformed coordinates, these can be transformed back to the original coordinate system. The reverse transformation is shown in [Figure 5](#).

$$\begin{bmatrix} -Ny/p & -Nx*Nz/p & Nx \\ Nx/p & -Ny*Nz/p & Ny \\ 0 & p & Nz \end{bmatrix} * \begin{bmatrix} Cx' \\ Cy' \\ \text{sum of } z'[i] / n \end{bmatrix} = \begin{bmatrix} Cx \\ Cy \\ Cz \end{bmatrix}$$

The reverse transformation can be applied directly as follows:

$$*Cx = -Ny * Cx' / p - Nx * Nz * Cy' / p + Nx * Cz'$$

$$*Cy = Nx * Cx' / p - Ny * Nz * Cy' / p + Ny * Cz'$$

$$*Cz = p * Cy' + Nz * Cz'$$

The ECEF value $[Cx,Cy,Cz]$ can then be converted back to geodetic coordinates. Given a polygon that is defined with no altitude or equal altitudes for each point, the altitude of the result can either be ignored or reset after converting back to a geodetic value.

The centroid of the Prism shape is found by finding the centroid of the base polygon and raising the point by half the height of the prism.

This can be added to altitude of the final result; alternatively, this can be added to Cz' , which ensures that negative height is correctly applied to polygons that are defined in a "clockwise" direction.

The recommended transforms only apply if Nx and Ny are non-zero. If the normal vector is $[0,0,1]$ (that is, along the z -axis), then no transform is necessary. Similarly, if the normal vector is $[0,1,0]$ or $[1,0,0]$, avoid the transformation and use the x and z coordinates or y and z coordinates (respectively) in the centroid calculation phase. If either Nx or Ny are zero, the alternative transform matrices in [Figure 6](#) can be used. The reverse transform is the transpose of this matrix.

$$\begin{array}{l|l} \text{if } Nx == 0: & \text{if } Ny == 0: \\ \begin{bmatrix} 0 & -Nz & Ny \\ 1 & 0 & 0 \\ 0 & Ny & Nz \end{bmatrix} & \begin{bmatrix} 0 & 1 & 0 \\ -Nz & 0 & Ny \\ Ny & 0 & Nz \end{bmatrix} \end{array} \quad \begin{array}{l} \\ T = T' = \end{array} \begin{bmatrix} -Nz & 0 & Nx \\ 0 & 1 & 0 \\ Nx & 0 & Nz \end{bmatrix}$$

4.2. Conversion to Circle or Sphere

The Circle or Sphere are simple shapes that suit a range of applications. A circle or sphere contains fewer units of data to manipulate, which simplifies operations on location estimates.

The simplest method for converting a location estimate to a Circle or Sphere shape is to determine the centroid and then find the longest distance to any point in the region of uncertainty to that point. This distance can be determined based on the shape type:

Circle/Sphere: No conversion necessary.

Ellipse/Ellipsoid: The greater of either semi-major axis or altitude uncertainty.

Polygon/Prism: The distance to the furthest vertex of the polygon (for a Prism, it is only necessary to check points on the base).

Arc-Band: The furthest length from the centroid to the points where the inner and outer arc end. This distance can be calculated by finding the larger of the two following formulae:

$$*X = \sqrt{d*d + R*R - 2*d*R*\cos(o/2)}$$

$$*x = \sqrt{d*d + r*r - 2*d*r*\cos(o/2)}$$

Once the Circle or Sphere shape is found, the associated confidence can be increased if the result is known to follow a normal distribution. However, this is a complicated process and provides limited benefit. In many cases it also violates the constraint that confidence in each dimension be the same. Confidence should be unchanged when performing this conversion.

Two dimensional shapes are converted to a Circle; three dimensional shapes are converted to a Sphere.

4.3. Three-Dimensional to Two-Dimensional Conversion

A three-dimensional shape can be easily converted to a two-dimensional shape by removing the altitude component. A sphere becomes a circle; a prism becomes a polygon; an ellipsoid becomes an ellipse. Each conversion is simple, requiring only the removal of those elements relating to altitude.

The altitude is unspecified for a two-dimensional shape and therefore has unlimited uncertainty along the vertical axis. The confidence for the two-dimensional shape is thus higher than the three-dimensional shape. Assuming equal confidence on each axis, the confidence of the circle can be increased using the following approximate formula: $C[2d]$ is the confidence of the two-dimensional shape and $C[3d]$ is the confidence of the three-dimensional shape. For example, a Sphere with a confidence of 95% can be simplified to a Circle of equal radius with confidence of 96.6%.

$$*C[2d] \geq C[3d] ^ (2/3)$$

[4.4. Increasing and Decreasing Uncertainty and Confidence](#)

The combination of uncertainty and confidence provide a great deal of information about the nature of the data that is being measured. If both uncertainty, confidence and PDF are known, certain information can be extrapolated. In particular, the uncertainty can be scaled to meet a desired confidence or the confidence for a particular region of uncertainty can be found.

In general, confidence decreases as the region of uncertainty decreases in size and confidence increases as the region of uncertainty increases in size. However, this depends on the PDF; expanding the region of uncertainty for a rectangular distribution has no effect on confidence without additional information. If the region of uncertainty is increased during the process of obfuscation (see [\[I-D.thomson-geopriv-location-obscuring\]](#)), then the confidence cannot be increased.

A region of uncertainty that is reduced in size always has a lower confidence.

A region of uncertainty that has an unknown PDF shape cannot be reduced in size reliably. The region of uncertainty can be expanded, but only if confidence is not increased.

This section makes the simplifying assumption that location information is symmetrically and evenly distributed in each dimension. This is not necessarily true in practice. If better information is available, alternative methods might produce better results.

[4.4.1. Rectangular Distributions](#)

Uncertainty that follows a rectangular distribution can only be decreased in size. Since the PDF is constant over the region of uncertainty, the resulting confidence is determined by the following formula: U_o and U_r are the sizes of the original and reduced regions of uncertainty (either the area or the volume of the region); C_o and C_b are the confidence values associated with each region.

$$*C_r = C_o * U_r / U_o$$

Where

Information is lost by decreasing the region of uncertainty for a rectangular distribution. Once reduced in size, the uncertainty region cannot subsequently be increased in size.

[4.4.2. Normal Distributions](#)

Uncertainty and confidence can be both increased and decreased for a normal distribution. However, the process is more complicated.

For a normal distribution, uncertainty and confidence are related to the standard deviation of the function. The following function defines the relationship between standard deviation, uncertainty and confidence along a single axis: $S[x]$ is the standard deviation, $U[x]$ is the

uncertainty and $C[x]$ is the confidence along a single axis. erfinv is the inverse error function.

$$*S[x] = U[x] / (\text{sqrt}(2) * \text{erfinv}(C[x]))$$

Where

Scaling a normal distribution in two dimensions requires several assumptions. Firstly, it is assumed that the distribution along each axis is independent. Secondly, the confidence for each axis is the same. Therefore, the confidence along each axis can be assumed to be: $C[x]$ is the confidence along a single axis and C_o is the overall confidence and n is the number of dimensions in the uncertainty.

$$*C[x] = C_o ^ (1/n)$$

Where

Therefore, to find the uncertainty for each axis at a desired confidence, C_d , apply the following formula:

$$*U_d[x] \leq U[x] * (\text{erfinv}(C_d ^ (1/n)) / \text{erfinv}(C_o ^ (1/n)))$$

For regular shapes, this formula can be applied as a scaling factor in each dimension to reach a required confidence.

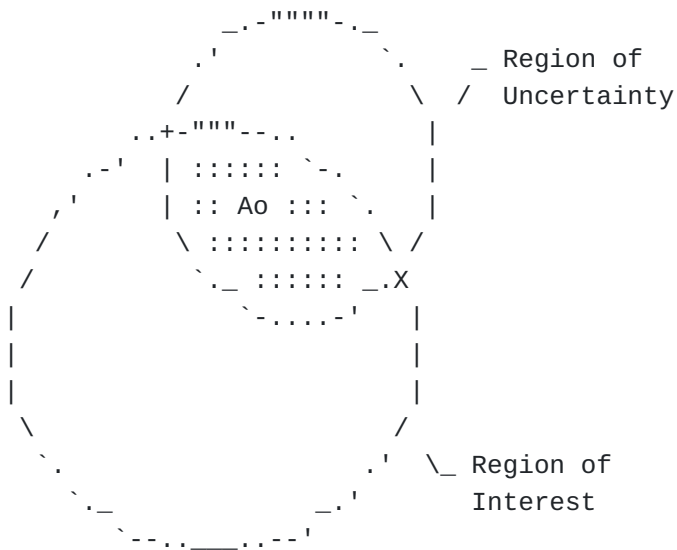
4.5. Determining Whether a Location is Within a Given Region

A number of applications require that a judgement be made about whether a Target is within a given region of interest. Given a location estimate with uncertainty, this judgement can be difficult. A location estimate represents a probability distribution, and the true location of the Target cannot be definitively known. Therefore, the judgement relies on determining the probability that the Target is within the region.

The probability that the Target is within a particular region is found by integrating the PDF over the region. For a normal distribution, there are no analytical methods that can be used to determine the integral of the two or three dimensional PDF over an arbitrary region. The complexity of numerical methods is also too great to be useful in many applications; for example, finding the integral of the PDF in two or three dimensions across the overlap between the uncertainty region and the target region. If the PDF is unknown, no determination can be made. When judging whether a location is within a given region, uncertainties using these PDFs can be assumed to be rectangular. If this assumption is made, the confidence should be scaled to 95%, if possible.

*Note: The selection of confidence has a significant impact on the final result. Only use a different confidence if an uncertainty value for 95% confidence cannot be found.

Given the assumption of a rectangular distribution, the probability that a Target is found within a given region is found by first finding the area (or volume) of overlap between the uncertainty region and the region of interest. This is multiplied by the confidence of the location estimate to determine the probability. [Figure 7](#) shows an example of finding the area of overlap between the region of uncertainty and the region of interest.



Once the area of overlap, A_o , is known, the probability that the Target is within the region of interest, P_i , is: A_o / A_u and the confidence is C_o .

$$*P_i = C_o * A_o / A_u$$

Given that the area of the region of uncertainty is A_u , this probability is often input to a decision process that has a limited set of outcomes; therefore, a threshold value needs to be selected. Depending on the application, different threshold probabilities might be selected. In the absence of specific recommendations, this document suggests that the probability be greater than 50% before a decision is made. If the decision process selects between two or more regions, as is required by [\[RFC5222\]](#), then the region with the highest probability can be selected.

[4.5.1. Determining the Area of Overlap for Two Circles](#)

Determining the area of overlap between two arbitrary shapes is a non-trivial process. Reducing areas to circles (see [Section 4.2](#)) enables the application of the following process.

Given the radius of the first circle r , the radius of the second circle R and the distance between their center points d , the following set of formulas provide the area of overlap A_o . d can be determined by

converting the center points to Cartesian coordinates and calculating the distance between the two center points:

*If the circles don't overlap, that is $d \geq r+R$, A_o is zero.

*If one of the two circles is entirely within the other, that is $d \leq |r-R|$, the area of overlap is the area of the smaller circle.

*Otherwise, if the circles partially overlap, that is $d < r+R$ and $d > |r-R|$, find A_o using:

$$-a = (r^2 - R^2 + d^2)/(2*d)$$

$$-A_o = r^2 \cdot \text{acos}(a/r) + R^2 \cdot \text{acos}((d - a)/R) - d \cdot \text{sqrt}(r^2 - a^2)$$

A value for

$$*d = \text{sqrt}((x_1-x_2)^2 + (y_1-y_2)^2 + (z_1-z_2)^2)$$

4.5.2. Determining the Area of Overlap for Two Polygons

A calculation of overlap based on polygons can give better results than the circle-based method. However, efficient calculation of overlapping area is non-trivial. Algorithms such as [Vatti's clipping algorithm](#) [Vatti92] can be used.

For large polygonal areas, it might be that geodesic interpolation is used. In these cases, altitude is also frequently omitted in describing the polygon. For such shapes, a planar projection can still give a good approximation of the area of overlap if the larger area polygon is projected onto the local tangent plane of the smaller. This is only possible if the only area of interest is that contained within the smaller polygon. Where the entire area of the larger polygon is of interest, geodesic interpolation is necessary.

5. Examples

This section presents some examples of how to apply the methods described in [Section 4](#).

5.1. Reduction to a Point or Circle

Alice receives a location estimate from her LIS that contains a ellipsoidal region of uncertainty. This information is provided at 19% confidence with a normal PDF. A PIDF-LO extract for this information is shown in [Figure 8](#).

```

<gp:geopriv>
  <gp:location-info>
    <gs:Ellipsoid srsName="urn:ogc:def:crs:EPSG::4979">
      <gml:pos>-34.407242 150.882518 34</gml:pos>
      <gs:semiMajorAxis uom="urn:ogc:def:uom:EPSG::9001">
        7.7156
      </gs:semiMajorAxis>
      <gs:semiMinorAxis uom="urn:ogc:def:uom:EPSG::9001">
        3.31
      </gs:semiMinorAxis>
      <gs:verticalAxis uom="urn:ogc:def:uom:EPSG::9001">
        28.7
      </gs:verticalAxis>
      <gs:orientation uom="urn:ogc:def:uom:EPSG::9102">
        43
      </gs:orientation>
    </gs:Ellipsoid>
  </gp:location-info>
  <gp:usage-rules/>
</gp:geopriv>

```

This information can be reduced to a point simply by extracting the center point, that is [-34.407242, 150.882518, 34].

If some limited uncertainty were required, the estimate could be converted into a circle or sphere. To convert to a sphere, the radius is the largest of the semi-major, semi-minor and vertical axes; in this case, 28.7 meters.

However, if only a circle is required, the altitude can be dropped as can the altitude uncertainty (the vertical axis of the ellipsoid), resulting in a circle at [-34.407242, 150.882518] of radius 7.7156 meters.

Bob receives a location estimate with a Polygon shape. This information is shown in [Figure 9](#).

```

<gml:Polygon srsName="urn:ogc:def:crs:EPSG::4326">
  <gml:exterior>
    <gml:LinearRing>
      <gml:posList>
        -33.856625 151.215906 -33.856299 151.215343
        -33.856326 151.214731 -33.857533 151.214495
        -33.857720 151.214613 -33.857369 151.215375
        -33.856625 151.215906
      </gml:posList>
    </gml:LinearRing>
  </gml:exterior>
</gml:Polygon>

```

To convert this to a polygon, each point is firstly assigned an altitude of zero and converted to ECEF coordinates (see [Appendix](#)

[Appendix A](#)). Then a normal vector for this polygon is found (see [Appendix Appendix B](#)). The results of each of these stages is shown in [Figure 10](#). Note that the numbers shown are all rounded; no rounding is possible during this process since rounding would contribute significant errors.

Polygon in ECEF coordinate space

(repeated point omitted and transposed to fit):

```

      [ -4.6470e+06  2.5530e+06 -3.5333e+06 ]
      [ -4.6470e+06  2.5531e+06 -3.5332e+06 ]
pecef = [ -4.6470e+06  2.5531e+06 -3.5332e+06 ]
      [ -4.6469e+06  2.5531e+06 -3.5333e+06 ]
      [ -4.6469e+06  2.5531e+06 -3.5334e+06 ]
      [ -4.6469e+06  2.5531e+06 -3.5333e+06 ]

```

Normal Vector: $n = [-0.72782 \quad 0.39987 \quad -0.55712]$

Transformation Matrix:

```

      [ -0.48152  -0.87643   0.00000 ]
t = [ -0.48828   0.26827   0.83043 ]
      [ -0.72782   0.39987  -0.55712 ]

```

Transformed Coordinates:

```

      [ 8.3206e+01  1.9809e+04  6.3715e+06 ]
      [ 3.1107e+01  1.9845e+04  6.3715e+06 ]
pecef' = [ -2.5528e+01  1.9842e+04  6.3715e+06 ]
      [ -4.7367e+01  1.9708e+04  6.3715e+06 ]
      [ -3.6447e+01  1.9687e+04  6.3715e+06 ]
      [ 3.4068e+01  1.9726e+04  6.3715e+06 ]

```

Two dimensional polygon area: $A = 12600 \text{ m}^2$

Two-dimensional polygon centroid: $C' = [8.8184e+00 \quad 1.9775e+04]$

Average of pecef' z coordinates: $6.3715e+06$

Reverse Transformation Matrix:

```

      [ -0.48152  -0.48828  -0.72782 ]
t' = [ -0.87643   0.26827   0.39987 ]
      [ 0.00000   0.83043  -0.55712 ]

```

Polygon centroid (ECEF): $C = [-4.6470e+06 \quad 2.5531e+06 \quad -3.5333e+06]$

Polygon centroid (Geo): $C_g = [-33.856926 \quad 151.215102 \quad -4.9537e-04]$

The point conversion for the polygon uses the final result, C_g , ignoring the altitude since the original shape did not include altitude.

To convert this to a circle, take the maximum distance in ECEF coordinates from the center point to each of the points. This results in a radius of 99.1 meters. Confidence is unchanged.

5.2. Increasing and Decreasing Confidence

Assuming that confidence is known to be 19% for Alice's location information. This is typical value for a three-dimensional ellipsoid uncertainty of normal distribution where the standard deviation is supplied in each dimension. The confidence associated with Alice's location estimate is quite low for many applications. Since the estimate is known to follow a normal distribution, the method in [Section 4.4.2](#) can be used. Each axis can be scaled by:

$$*scale = \text{erfinv}(0.95^{(1/3)}) / \text{erfinv}(0.19^{(1/3)}) = 2.9937$$

Ensuring that rounding always increases uncertainty, the location estimate at 95% includes a semi-major axis of 23.1, a semi-minor axis of 10 and a vertical axis of 86.

Bob's location estimate covers an area of approximately 12600 square meters. If the estimate follows a rectangular distribution, the region of uncertainty can be reduced in size. To find the confidence that he is within the smaller area of the concert hall, given by the polygon [-33.856473, 151.215257; -33.856322, 151.214973; -33.856424, 151.21471; -33.857248, 151.214753; -33.857413, 151.214941; -33.857311, 151.215128]. To use this new region of uncertainty, find its area using the same translation method described in [Section 4.1.1.2](#), which is 4566.2 square meters. The confidence associated with the smaller area is therefore $95\% * 4566.2 / 12600 = 34\%$.

5.3. Matching Location Estimates to Regions of Interest

Suppose than a circular area is defined centered at [-33.872754, 151.20683] with a radius of 1950 meters. To determine whether Bob is found within this area, we apply the method in [Section 4.5](#). Using the converted Circle shape for Bob's location, the distance between these points is found to be 1915.26 meters. The area of overlap between Bob's location estimate and the region of interest is therefore 2209 square meters and the area of Bob's location estimate is 30853 square meters. This gives the probability that Bob is less than 1950 meters from the selected point as 67.8%.

Note that if 1920 meters were chosen for the distance from the selected point, the area of overlap is only 16196 square meters and the confidence is 49.8%. Therefore, it is marginally more likely that Bob is outside the region of interest, despite the center point of his location estimate being within the region.

6. Security Considerations

This document describes methods for managing and manipulating uncertainty in location. No specific security concerns arise from most of the information provided.

7. Acknowledgements

Peter Rhodes provided assistance with some of the mathematical groundwork on this document. Dan Cornford provided a detailed review and many terminology corrections.

8. References

[RFC3693]	Cuellar, J., Morris, J., Mulligan, D., Peterson, J. and J. Polk, " Geopriv Requirements ", RFC 3693, February 2004.
[RFC3694]	Danley, M., Mulligan, D., Morris, J. and J. Peterson, " Threat Analysis of the Geopriv Protocol ", RFC 3694, February 2004.
[RFC3825]	Polk, J., Schnizlein, J. and M. Linsner, " Dynamic Host Configuration Protocol Option for Coordinate-based Location Configuration Information ", RFC 3825, July 2004.
[RFC4119]	Peterson, J., " A Presence-based GEOPRIV Location Object Format ", RFC 4119, December 2005.
[RFC5139]	Thomson, M. and J. Winterbottom, " Revised Civic Location Format for Presence Information Data Format Location Object (PIDF-LO) ", RFC 5139, February 2008.
[RFC5222]	Hardie, T., Newton, A., Schulzrinne, H. and H. Tschofenig, " LoST: A Location-to-Service Translation Protocol ", RFC 5222, August 2008.
[RFC5491]	Winterbottom, J., Thomson, M. and H. Tschofenig, " GEOPRIV Presence Information Data Format Location Object (PIDF-LO) Usage Clarification, Considerations, and Recommendations ", RFC 5491, March 2009.
[I-D.ietf-geopriv-policy]	Schulzrinne, H, Tschofenig, H, Cuellar, J, Polk, J, Morris, J and M Thomson, " Geolocation Policy: A Document Format for Expressing Privacy Preferences for Location Information ", Internet-Draft draft-ietf-geopriv-policy-25, October 2011.
[I-D.ietf-geopriv-arch]	Barnes, R, Lepinski, M, Cooper, A, Morris, J, Tschofenig, H and H Schulzrinne, " An Architecture for Location and Location Privacy in Internet Applications ", Internet-Draft draft-ietf-geopriv-arch-03, October 2010.
[I-D.thomson-geopriv-location-obscuring]	Thomson, M, " Obscuring Location ", Internet-Draft draft-thomson-geopriv-location-obscuring-03, June 2011.

[WGS84]	US National Imagery and Mapping Agency, "Department of Defense (DoD) World Geodetic System 1984 (WGS 84), Third Edition ", NIMA TR8350.2, January 2000.
[GeoShape]	Thomson, M. and C. Reed, "GML 3.1.1 PIDF-LO Shape Application Schema for use by the Internet Engineering Task Force (IETF)", Candidate OpenGIS Implementation Specification 06-142r1, Version: 1.0, April 2007.
[ISO.GUM]	ISO/IEC, "Guide to the expression of uncertainty in measurement (GUM) ", Guide 98:1995, 1995.
[NIST.TN1297]	Taylor, B. and C. Kuyatt, "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results ", Technical Note 1297, Sep 1994.
[Convert]	Burtch, R., "A Comparison of Methods Used in Rectangular to Geodetic Coordinate Transformations", April 2006.
[Sunday02]	Sunday, D, "Fast polygon area and Newell normal computation ", Journal of Graphics Tools JGT, 7(2):9-13,2002, 2002.
[Vatti92]	Vatti, B, "A generic solution to polygon clipping ", Communications of the ACM Vol35, Issue7, pp56-63, 1992.

[Appendix A. Conversion Between Cartesian and Geodetic Coordinates in WGS84](#)

The process of conversion from geodetic (latitude, longitude and altitude) to earth-centered, earth-fixed (ECEF) Cartesian coordinates is relatively simple.

In this section, the following constants and derived values are used from the definition of [WGS84](#) [WGS84]:

*{radius of ellipsoid} $R = 6378137$ meters

*{inverse flattening} $1/f = 298.257223563$

*{first eccentricity squared} $e^2 = f * (2 - f)$

*{second eccentricity squared} $e'^2 = e^2 * (1 - e^2)$

To convert geodetic coordinates (latitude, longitude, altitude) to ECEF coordinates (X, Y, Z), use the following relationships:

$*N = R / \sqrt{1 - e^2 * \sin(\text{latitude})^2}$

$*X = (N + \text{altitude}) * \cos(\text{latitude}) * \cos(\text{longitude})$

$$*Y = (N + \text{altitude}) * \cos(\text{latitude}) * \sin(\text{longitude})$$

$$*Z = (N * (1 - e^2) + \text{altitude}) * \sin(\text{latitude})$$

The reverse conversion requires more complex computation and most methods introduce some error in latitude and altitude. A range of techniques are described in [\[Convert\]](#). A variant on the method originally proposed by Bowring, which results in an acceptably small error, is described by the following: $p < 1$, the value for altitude that this method produces is unstable. A simpler method for determining the altitude of a point near the poles is:

$$*p = \sqrt{X^2 + Y^2}$$

$$*r = \sqrt{X^2 + Y^2 + Z^2}$$

$$*u = \text{atan}((1-f) * Z * (1 + e'^2 * (1-f) * R / r) / p)$$

$$*\text{latitude} = \text{atan}((Z + e'^2 * (1-f) * R * \sin(u)^3) / (p - e^2 * R * \cos(u)^3))$$

$$*\text{longitude} = \text{atan}(Y / X)$$

$$*\text{altitude} = \sqrt{(p - R * \cos(u))^2 + (Z - (1-f) * R * \sin(u))^2}$$

If the point is near the poles, that is

$$*\text{altitude} = |Z| - R * (1 - f)$$

[Appendix B. Calculating the Upward Normal of a Polygon](#)

For a polygon that is guaranteed to be convex and coplanar, the upward normal can be found by finding the vector cross product of adjacent edges.

For more general cases the Newell method of approximation described in [\[Sunday02\]](#) may be applied. In particular, this method can be used if the points are only approximately coplanar, and for non-convex polygons.

This process requires a Cartesian coordinate system. Therefore, [convert the geodetic coordinates of the polygon to Cartesian, ECEF coordinates \[ecef2geo\]](#). If no altitude is specified, assume an altitude of zero.

This method can be condensed to the following set of equations:

$(x[1], y[1], z[1])$ through $(x[n], y[n], z[n])$. Each array is treated as circular, that is, $x[0] == x[n]$ and $x[n+1] == x[1]$.

$$*N_x = \text{sum from } i=1..n \text{ of } (y[i] * (z[i+1] - z[i-1]))$$

$$*N_y = \text{sum from } i=1..n \text{ of } (z[i] * (x[i+1] - x[i-1]))$$

$$*N_z = \text{sum from } i=1..n \text{ of } (x[i] * (y[i+1] - y[i-1]))$$

For these formulae, the polygon is made of points
To translate this into a unit-vector; divide each component by the
length of the vector:

$$*N_x' = N_x / \sqrt{N_x^2 + N_y^2 + N_z^2}$$

$$*N_y' = N_y / \sqrt{N_x^2 + N_y^2 + N_z^2}$$

$$*N_z' = N_z / \sqrt{N_x^2 + N_y^2 + N_z^2}$$

Appendix B.1. Checking that a Polygon Upward Normal Points Up

[RFC 5491](#) [RFC5491] mandates that polygons be presented in anti-clockwise direction so that the upward normal is in an upward direction. Accidental reversal of points can invert this vector. This error can be hard to detect just by looking at the series of coordinates that form the polygon.

Calculate the dot product of the [upward normal of the polygon](#) [polygonupnormal] and any vector that points away from the center of the Earth from the location of polygon. If this product is positive, then the polygon upward normal also points away from the center of the Earth.

*The inverse cosine of this value indicates the angle between the horizontal plane and the approximate plane of the polygon.

A unit vector for the upward direction at any point can be found based on the latitude (lat) and longitude (lng) of the point, as follows:

$$*Up = [\cos(lat) * \cos(lng) ; \cos(lat) * \sin(lng) ; \sin(lat)]$$

For polygons that span less than half the globe, any point in the polygon - including the centroid - can be selected to generate an approximate up vector for comparison with the upward normal.

Authors' Addresses

Martin Thomson Thomson Andrew Corporation Andrew Building (39)
Wollongong University Campus Northfields Avenue
Wollongong, NSW 2522 AU EMail: martin.thomson@andrew.com

James Winterbottom Winterbottom Andrew Corporation Andrew Building
(39) Wollongong University Campus Northfields Avenue
Wollongong, NSW 2522 AU EMail: james.winterbottom@andrew.com