CLOUD COMPUTING AND HYPERBOLIC VORONOI DIAGRAMS ON THE SPHERE

PAVEL BLEHER AND CAROLINE SHOURABOURA

Abstract. In this work we study the minimization problem for the total distance in a cloud computing network on the sphere. We prove that a solution to this problem is given in terms of a hyperbolic Voronoi diagram on the sphere. We present results of computer simulations illustrating the solution.

1. Introduction

We consider a network

\[ S = \{ S_1, \ldots, S_n \} \]

of computer data centers on the earth surface. We assume that the earth surface is represented by a sphere \( \Lambda_ R \) of a radius \( R > 0 \), and each data center \( S_m \) is described by its geographical coordinates, the latitude \( \phi_m \) and the longitude \( \theta_m \). The data centers provide a variety of cloud computing services to a big set of users

\[ U = \{ U_1, \ldots, U_N \}. \]

Our main goal will be to solve the minimization problem for the total distance \( \mathcal{D} \) between the data centers and the users. We assume that each user \( U_k \) is assigned to a data center \( S_{j(k)} \), and the total distance is the sum of the distances between the users and the corresponding data centers,

\[ \mathcal{D}(j) = \sum_{k=1}^{N} d(S_{j(k)}, U_k), \quad (1.1) \]

where \( j = j(k) \) is the function assigning the user \( U_k \) to the data center \( S_{j(k)} \).

For any two points \( x, y \) on the sphere we denote \( d(x, y) \) the distance between \( x \) and \( y \) on the sphere, that is the length of the arc of the great circle connecting \( x \) to \( y \). We will consider the problem of the minimization of the total distance under certain constraints. We will assume that each data center \( S_j \) has a capacity \( C_j \) which characterizes how many users it can service. The assumption of the limited capacity for the data centers leads to the minimization problem with constraints: Find an assignment function \( j_0(k) \) such that

\[ \min_{j \in \mathcal{J}(C)} \mathcal{D}(j) = \mathcal{D}(j_0), \quad (1.2) \]

where

\[ \mathcal{J}(C) = \{ j = j(k) \mid N_m = \#\{ k : j(k) = m \} \leq C_m \quad \text{for} \quad m = 1, \ldots, n \}. \quad (1.3) \]
Observe that $N_m = \#\{k : j(k) = m\}$ is just the number of users assigned to the data center $S_m$.

Since the number of users is usually large, it can be useful to consider a (nonnegative) measure $\mu$ of users on the sphere. In this case, the total distance functional looks like

$$D(j) = \int_{\Lambda_R} d(S_j(U), U) d\mu(U), \quad (1.4)$$

where $j(U)$ is the assignment function of the users to the data centers. Respectively, constraint (1.3) takes the form

$$J(C) = \{ j = j(U) \mid \mu\{U : j(U) = m\} \leq C_m \text{ for } m = 1, \ldots, n \}. \quad (1.5)$$

The main result of this work is Theorem 2.1, in which we prove that the solution to minimization problem (1.2) is given by a hyperbolic Voronoi diagram on the sphere. The classical Voronoi diagrams have numerous applications to science and technology (for many of these applications see the review paper [1] by Aurenhammer, the lectures [6] by Möller, the collection of papers [4], the monograph [8] of Okabe, Boots, Sugihara, and Chiu, and references therein). For applications of Voronoi diagrams to cloud computing see the recent paper [9] of Shouraboura and Bleher and references therein. In our Theorem 2.1 we show that the presence of the constraints leads to an extension of the classical Voronoi diagrams to hyperbolic Voronoi diagrams on the sphere. It is interesting to notice that in a completely different setting similar diagrams on the plane or in the 3D space appear in crystallography as the Johnson-Mehl model of the crystal growth [5]. We formulate and prove Theorem 2.1 in Section 2. In the subsequent section we illustrate Theorem 2.1 by results of computer simulations.

2. Solution of the constrained minimization problem and hyperbolic Voronoi diagrams on the sphere

In this section we will solve minimization problem (1.2). We will consider a general measure $\mu$ of users. For a given measurable assignment function $j(U)$ on $\Lambda_R$, we introduce the sets

$$\sigma_m(j) = \{ U \in \Lambda_R : j(U) = m \}, \quad m = 1, \ldots, n, \quad (2.1)$$

so that $\sigma_m(j)$ is the set of users assigned to the data center $S_m$. The sets $\{ \sigma_m(j) \}$ are disjoint and they form a measurable partition $\pi(j)$ of the sphere $\Lambda_R$, so that

$$\pi(j) : \Lambda_R = \bigsqcup_{m=1}^n \sigma_m(j). \quad (2.2)$$

The functional $D(j)$ reduces to

$$D(j) = \sum_{m=1}^n \int_{\sigma_m(j)} d(S_m, U) d\mu(U), \quad (2.3)$$

and constraint (1.5) to

$$J(C) = \{ j = j(U) \mid \mu(\sigma_m(j)) \leq C_m \text{ for } m = 1, \ldots, n \}. \quad (2.4)$$

Let us begin with the unconstrained minimization problem. In the absence of the capacity constraint, minimization problem (1.2) can be solved by assigning each user $U$ to its closest
data center $S_j$, so that $S_{j_0(U)}$ is the closest data center to $U$. Geometrically, we partition the sphere $\Lambda_R$ into the cells $\{\sigma_m\}$ of the Voronoi diagram $V$ on the sphere with the points $\{S_m\}$. The cell $\sigma_m$ is defined as the set of points on $\Lambda_R$ which are closer (or at the same distance) to $S_m$ than to other $S_j$’s, that is
\[ \sigma_m = \{ x \in \Lambda_R : d(x, S_m) \leq d(x, S_l) \text{ for all } l \neq m \}. \]

Observe that the cells $\{\sigma_m\}$ are convex spherical polygons on the sphere $\Lambda_R$. With the Voronoi diagram we associate a graph $\Gamma_V$. The vertices of the graph $\Gamma_V$ are the vertices of the polygons $\{\sigma_m\}$, and the edges of the graph $\Gamma_V$ are the sides of the polygons $\{\sigma_m\}$. The Delaunay triangulation, associated with the Voronoi diagram, is the dual graph $\Gamma_D$ to $\Gamma_V$, with vertices $\{S_m\}$ and edges connecting vertices $S_m, S_l$ if and only if the cells $\sigma_m, \sigma_l$ have a common side.

Thus, in the absence of the constraint, we assign to each data center $S_m$ all users in the cell $\sigma_m$ of the Voronoi diagram. To describe the assignment in the presence of the constraint we introduce hyperbolic Voronoi diagrams on the sphere.

Suppose that to each point $S_m$ a number $d_m$ is assigned. Then the hyperbolic Voronoi diagram
\[ V(d), \quad d = (d_1, \ldots, d_n), \]
with the parameters $\{d_m\}$ and the points $\{S_m\}$, is defined as follows. The cell $\sigma_m$ of the point $S_m$ is defined as
\[ \sigma_m = \{ x \in \Lambda_R : d(x, S_m) + d_m \leq d(x, S_l) + d_l \text{ for all } l \neq m \}. \]

Observe that on the curve $\gamma_{ml}$ separating two neighboring cells $\sigma_m, \sigma_l$ we have the equation,
\[ d(x, S_m) + d_m = d(x, S_l) + d_l, \]

hence $\gamma_{ml}$ is a part of the spherical hyperbola on the sphere $\Lambda_R$. A vertex $v$ of the hyperbolic Voronoi diagram is a point which belong to three or more cells. The graph $\Gamma_V(d)$ of the hyperbolic Voronoi diagram $V(d)$ consists of the vertices $\{v\}$ and the edges $\{e\}$, which are the curves $\{\gamma_{ml}\}$ separating neighboring cells.

A solution to the constrained minimization problem can be obtained as follows. We will take a natural assumption that the total capacity of all data centers is not less than the number of the users,
\[ \sum_{m=1}^{n} C_m \geq \mu \Lambda_R. \]

Otherwise, it would be impossible to service all the users.

**Theorem 2.1.** For any measure $\mu$, a minimizer $j_0(U)$ to constrained minimizing problem (2.2) exists, which can be obtained as follows. There exist numbers $(d_1, \ldots, d_n)$ such that the minimizer $j_0$ is obtained by assigning all users in the cell $\sigma_m$ of the hyperbolic Voronoi diagram $V(d)$ to the data center $S_m$, so that
\[ j_0(U) = m \quad \text{if and only if} \quad U \in \sigma_m. \]

**Proof.** For any $\varepsilon > 0$, we will introduce a regularization $\mu_\varepsilon(U)$ of the measure $\mu(U)$, and we will first prove Theorem 2.1 for the measure $\mu_\varepsilon(U)$ and then we will take the limit $\varepsilon \to 0$. The measure $\mu_\varepsilon(U)$ is defined as follows. Consider the function
\[ g_\varepsilon(x) = e^{-x^2/\varepsilon}. \]
on the real line, and for a given point \( U_0 \in \Lambda_R \), introduce the probability density \( f_\varepsilon(U; U_0) \) on the sphere \( \Lambda_R \) as
\[
f_\varepsilon(U; U_0) = \frac{1}{Z(\varepsilon)} g_\varepsilon (d(U, U_0)) > 0
\] (2.11)
where
\[
Z(\varepsilon) = \int_{\Lambda_R} g_\varepsilon (d(U, U_0)) \, dU,
\] (2.12)
Then, as \( \varepsilon \to 0 \), the function \( f_\varepsilon(U; U_0) \) converges to a \( \delta \)-function at the point \( U_0 \). Consider the function
\[
p_\varepsilon(U) = \int_{\Lambda_R} f_\varepsilon(U; U_0) d\mu(U_0) > 0,
\] (2.13)
and define the measure \( \mu_\varepsilon(U) \) as
\[
d\mu_\varepsilon(U) = p_\varepsilon(U) dU.
\] (2.14)
Then in the weak sense,
\[
\lim_{\varepsilon \to 0} \mu_\varepsilon = \mu,
\] (2.15)
so that for any continuous function \( f(U) \),
\[
\lim_{\varepsilon \to 0} \int_{\Lambda_R} f(U) d\mu_\varepsilon(U) = \int_{\Lambda_R} f(U) d\mu.
\] (2.16)

Consider the minimization problem
\[
\min_{j \in \mathcal{J}(C)} D_\varepsilon(j) = D_\varepsilon(j_\varepsilon),
\] (2.17)
for the functional
\[
D_\varepsilon(j) = \int_{\Lambda_R} d(S_j(U), U) d\mu_\varepsilon(U).
\] (2.18)

**Lemma 2.2.** A minimizing function \( j_\varepsilon \) for minimization problem (2.17) exists.

**Proof.** Let
\[
\inf_{j \in \mathcal{J}(C)} D_\varepsilon(j) = \lim_{k \to \infty} D_\varepsilon(j_\varepsilon^k).
\] (2.19)
Consider the partitions \( \pi(j_\varepsilon^k) \). By the compactness argument, there exists a weakly converging subsequence
\[
\lim_{r \to \infty} \pi(j_\varepsilon^{k_r}) = \pi(j_\varepsilon).
\] (2.20)
Since \( d(U, U_0) \) is a continuous function of \( (U, U_0) \) and \( p_\varepsilon(U) \) is a continuous function of \( U \), this implies that \( j_\varepsilon \) solves (2.17). \( \square \)

**Lemma 2.3.** Let \( j_\varepsilon \) be a solution to (2.17) and
\[
\pi(j_\varepsilon) : \Lambda_R = \bigsqcup_{m=1}^{n} \sigma_m(j_\varepsilon)
\] (2.21)
the corresponding partition of \( \Lambda_R \). Then for any \( m \neq l \),
\[
\essinf_{U \in \sigma_m(j_\varepsilon)} [d(U, S_l) - d(U, S_m)] \geq \esssup_{U \in \sigma_l(j_\varepsilon)} [d(U, S_l) - d(U, S_m)].
\] (2.22)
Proof. Suppose, for the sake of contradiction, that (2.22) is false. Then there exist sets \( A \subset \sigma(j_\varepsilon), B \subset \sigma_m(j_\varepsilon) \) and numbers \( a > b \) such that

1. \( d(U, S_l) - d(U, S_m) > a \) for all \( U \in A \),
2. \( d(U, S_l) - d(U, S_m) < b \) for all \( U \in B \),
3. \( \mu_\varepsilon(A) = \mu_\varepsilon(B) > 0 \).

Let us reassign all users \( U \in A \) from \( S_l \) to \( S_m \) and all users \( U \in B \) from \( S_m \) to \( S_l \). This reassignment, \( j_\varepsilon^R \), will not violate the constraints and it will decrease the total distance functional \( D(j) \). Indeed,

\[
D_\varepsilon(j_\varepsilon) - D_\varepsilon(j_\varepsilon^R) = \int_A [d(U, S_l) - d(U, S_m)] \, d\mu_\varepsilon(U) - \int_B [d(U, S_l) - d(U, S_m)] \, d\mu_\varepsilon(U) \\
\geq a \mu_\varepsilon(A) - b \mu_\varepsilon(B) = (a - b) \mu_\varepsilon(A) > 0,
\]

hence \( D(j_\varepsilon) \) is not the minimum of \( D(j) \). This contradiction proves (2.22). \( \square \)

It follows from (2.22) that there exist sets \( \{E_m, m = 1, \ldots, n\} \) such that

1. \( E_m \subset \sigma_m(j_\varepsilon) \) for \( m = 1, \ldots, n \),
2. \( \mu_\varepsilon(\sigma_m(j_\varepsilon) \setminus E_m) = 0 \) for \( m = 1, \ldots, n \),
3. for any \( m \neq l \),

\[
\inf_{U \in E_m} [d(U, S_l) - d(U, S_m)] \geq \sup_{U \in E_l} [d(U, S_l) - d(U, S_m)].
\]

Consider the closure \( F_m \) of each set \( E_m \),

\[
F_m = \overline{E_m}, \quad m = 1, \ldots, n.
\]

Then, since \( \{\sigma_m(j_\varepsilon)\} \) is a partition of the sphere \( \Lambda_R \), we have that

\[
\Lambda_R = \bigcup_{m=1}^n F_m.
\]

By continuity, (2.24) holds if we replace \( E_m, E_l \) for \( F_m, F_l \), respectively, that is for any \( m \neq l \),

\[
\inf_{U \in F_m} [d(U, S_l) - d(U, S_m)] \geq \sup_{U \in F_l} [d(U, S_l) - d(U, S_m)].
\]

Moreover, if the intersection of \( F_m \) and \( F_l \) is nonempty,

\[
\gamma_{ml} \equiv F_m \cup F_l \neq \emptyset,
\]

then

\[
\inf_{U \in F_m} [d(U, S_l) - d(U, S_m)] = \sup_{U \in F_l} [d(U, S_l) - d(U, S_m) \equiv d_{ml}].
\]

This implies that for \( U \in \gamma_{ml} \),

\[
d(U, S_l) - d(U, S_m) = d_{ml},
\]

Thus, \( \gamma_{ml} \) is a part of hyperbola (2.30) on the sphere \( \Lambda_R \). Since the function \( d(U, S_m) \) is an analytic function of \( U \) outside of the point \( S_m \), it follows that \( \gamma_{ml} \) consists of a finite number of arcs of hyperbola (2.30). Moreover, the boundary of \( F_m \) is equal to

\[
\partial F_m = \bigcup_{l \neq m} (F_m \cap F_l),
\]

hence \( \partial F_m \) consists of a finite number of arcs of hyperbolas on the sphere \( \Lambda_R \).
Assume that for some $m$ the sets $\gamma_{ml}$ and $\gamma_{mp}$ intersect at some point $v$. Then $v$ belongs to the boundaries of $F_l$ and $F_p$, hence $\gamma_{lp}$ passes through $v$, that is $v$ is a triple point (or higher). By taking $U = v$ in (2.30), we obtain that

$$d_{ml} + d_{lp} + d_{pm} = 0. \quad (2.32)$$

Define the one-chain

$$C_1 = \sum_{m,l} d_{ml} \gamma_{ml}. \quad (2.33)$$

Then by (2.32),

$$\partial C_1 = 0. \quad (2.34)$$

Since the first homology group of the sphere $\Lambda_R$ is trivial, this implies that there exists a zero-chain $C_0$ such that

$$\partial C_0 = C_1. \quad (2.35)$$

In other words, there exist numbers $\{d_m\}$ such that

$$d_{ml} = d_m - d_l. \quad (2.36)$$

By (2.30), this proves (2.7) and hence Theorem 2.1 for the distribution $\mu_\varepsilon$ of users. Let us consider the distribution $\mu$.

The numbers $\{d_m\}$ are defined uniquely up to a common additive constant. To fix the constant, let us put $d_1 = 0$. Then the other numbers, $d_m = d_m(\varepsilon)$, $m > 1$, are bounded by a constant independent of $\varepsilon$. Let us take a converging subsequence

$$d_m = \lim_{\varepsilon \to 0} d_m(\varepsilon_p), \quad p = 1, 2, \ldots; \quad m = 1, \ldots, n. \quad (2.37)$$

This defines the numbers $\{d_m\}$. Consider the corresponding hyperbolic Voronoi diagram $V(d)$. Since $V(d(\varepsilon))$ gives a solution to the minimization problem for the users distribution $\mu_\varepsilon$, we obtain that $V(d)$ gives a solution for the distribution $\mu$. This proves Theorem 2.1. \qed

**Remark.** Observe that, in general, the hyperbolic Voronoi diagram solving the minimization problem is not unique. Consider, for instance, a discrete distribution of users

$$d\mu(U) = \sum_{k=1}^{N} \delta(U - U_k) \, dU. \quad (2.38)$$

Then, if there are no users on the edges of the Voronoi diagram, then we can shift the numbers $d_1, \ldots, d_n$ a little, without changing the number of users in the Voronoi cells. On the other hand, the solution to the minimization problem is unique for any users distribution which has an absolutely continuous component with a positive density, like $\mu_\varepsilon$. This explains why we used the distribution $\mu_\varepsilon$ in the above proof, establishing the theorem first for $\mu_\varepsilon$ and then passing to the limit $\varepsilon \to 0$.

### 3. Computer simulations of the minimizer

Consider a minimizer $j_0$ of (1.2). We say that constraint (1.5) is active on a server $S_m$ if

$$\mu\{U : j_0(U) = m\} = C_m, \quad (3.1)$$

and it is inactive on $S_m$ if

$$\mu\{U : j_0(U) = m\} < C_m. \quad (3.2)$$
If all constraints are inactive then the minimizer \( j_0 \) is the usual Voronoi diagram on the sphere.

In the computer simulations we begin with the usual Voronoi diagram. To that end, we construct the convex hull \( \mathcal{H} \) of the set \( S \). This can be done in \( \mathcal{O}(n \ln n) \) operations. The convex hull gives the Delaunay triangulation, and the spherical bisectors of the edges of the Delaunay triangulation give the edges of the Voronoi diagram (see, e.g., [2], [3], [7]). We check the condition

\[
\mu(\sigma_m) \leq C_m, \quad (3.3)
\]

for the Voronoi cells \( \sigma_m \), \( m = 1, \ldots, n \). If the condition is satisfied for all \( m \)'s, then the minimizer is the usual Voronoi diagram.

**Figure 1.** The solution to the constrained minimization problem with 10 data centers, Seattle, Atlanta, New York, Phoenix, San Francisco, Denver, Houston, Chicago, Boston, and Miami. The Voronoi Diagram on the sphere with these 10 points is shown by the cyan line. We impose two constraints shown in Red: in San Francisco the capacity is equal to 15 million users and in Atlanta it is 20 million.

In application to cloud computing, it is plausible to assume that most of the servers have a big capacity, and condition (3.3) violates only on a set \( S_0 \) of isolated \( S_m \)'s. To set a notation, let

\[
S_0 = \{ S_m, \ m \in I_0 \}, \quad (3.4)
\]

where \( I_0 \subset \{1, \ldots, n\} \). In this case we are looking for parameters \( \{ d_m, \ m \in I_0 \} \) such that

\[
\mu(\sigma_m(d)) = C_m, \quad m \in I_0, \quad (3.5)
\]
where \( \sigma_m(d) \) is the cell of the hyperbolic Voronoi diagram \( V(d) \), with given \( d_m \) when \( m \in I_0 \) and \( d_m = 0 \) when \( m \not\in I_0 \). We solve equation (3.5) numerically for each \( m \in I_0 \). We assume that the capacity of the neighboring vertices to \( S_m \in S_0 \) is big so that the resulting hyperbolic Voronoi diagram \( V(d) \) satisfies condition (3.3) for all \( m \).

In Fig. 1 we illustrate the numerical solution by an example. In this example, we consider 10 data centers located in Seattle, San Francisco, Phoenix, Denver, Chicago, Houston, Atlanta, Boston, New York, and Miami. We assume that the distribution of users is proportional to the population distribution. The blue lines show the spherical Voronoi diagram. We further assume that the data centers in San Francisco and Atlanta have a limited capacity equivalent to the population of 15 and 20 million, respectively. We calculate the hyperbolic Voronoi diagram under these constraints. The change from the usual Voronoi diagram to the hyperbolic one is shown in Fig. 1 in red.

4. Conclusion

In this work we studied the minimization problem for the total communication distance in a computer cloud under the condition of restricted capacity of the data centers. Our main result is Theorem 2.1 which shows that a solution to the minimization problem is given by a hyperbolic Voronoi diagram constructed on the data centers \( S_1, \ldots, S_n \). The parameters \( d_1, \ldots, d_n \) of the hyperbolic Voronoi diagram can be found from the condition that the number of users in each cell \( \sigma_m \) of the diagram does not exceed the capacity of the corresponding data center \( S_m \).

Although we discuss the application to the computer cloud only, the hyperbolic Voronoi diagrams on the sphere can find other important applications. We can mention the problem of location of air-bases [8], the distribution of facilities in global Internet companies like Amazon.com, the distribution of the telecommunication centers for mobile telephones in global telephone companies, data collection centers, and others. Also, theorem 2.1 can be extended, under appropriate conditions, to Riemannian manifolds in dimension 2 and higher.
REFERENCES

[1] Aurenhammer, F., Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure. 1991, ACM Computing Surveys, pp. 23(3):345–405.

[2] Brown, K.Q., Geometric Transforms for Fast Geometric Algorithms. Ph.D. Dissertation. 1979. Carnegie Mellon Univ., Pittsburgh, PA, USA. AAI8012772.

[3] Brown, K.Q., Voronoi diagrams from convex hulls, Inform. Process. Lett. 9 (1979) 223–228.

[4] Generalized Voronoi Diagram: A Geometry-Based Approach to Computational Intelligence (Studies in Computational Intelligence 158). Editor: M. Gavrilova, Springer, 2008, 334 pages.

[5] Johnson, W. A. and Mehl, R. F. Reaction kinetics in processes of nucleation and growth. Trans. Amer. Inst. Min. Engrs. 135, (1939) 416–458.

[6] Möller, J. Lectures on Random Voronoi Tessellations. Lecture Notes in Statistics. Springer-Verlag, New York, 1994.

[7] Na, H.-S., Lee, C.-N., and Cheong, O., Voronoi diagrams on the sphere. Computational Geometry: Theory and Applications. September 2002, 183–194.

[8] Okabe, A., Boots, B., Sugihara, K., and Chiu, S.N. Spatial Tessellations – Concepts and Applications of Voronoi Diagrams, 2nd edition. John Wiley, 2000, 671 pages.

[9] Shouraboura, C. and Bleher, P., Placement of applications in computing clouds using Voronoi Diagrams. Journal of Internet Services and Applications, v.2, no.3, 2011 Dec, p.229(13) (ISSN: 1867-4828)

Department of Mathematical Sciences, Indiana University-Purdue University Indianapolis, 402 N. Blackford St., Indianapolis, IN 46202, U.S.A.

E-mail address: bieber@math.iupui.edu

Forest Ridge 9927 171st Ave SE, Newcastle, WA 98059

E-mail address: carolinesh@forestridge.org