

---

# A Data-Balanced Geohash for Distributed Spatiotemporal Database Indexing

---

**Taylor B. Arnold**  
AT&T Labs Research  
33 Thomas Street  
New York, NY 10007  
`taylor@research.att.com`

## Abstract

We present a modification of the standard geohash algorithm for which the data volume, rather than spatial area, is constant for a given hash prefix length. This property is particularly useful for indexing large distributed databases, where load distribution of large range scans is an important aspect of query performance. Distributed spatiotemporal databases, which typically require interleaving spatial and temporal elements into a single key, reap additional benefits from a balanced geohash by creating a consistent balance between spatial and temporal precision even across areas of varying data density. We apply our algorithm to data generated proportional to population as given by census block population counts provided from the US Census Bureau. An efficient implementation for calculating an arbitrary balanced geohash is also provided.

## 1 Introduction

Queries over large distributed databases often take the form of a series of large range scans; balancing

## 2 A Balanced Geohash

### 2.1 Formal definition of standard geohash

### 2.2 Weighting scheme

### 2.3 Error analysis

$$K(x) \equiv -\log_2 (|h(x) - f(x)|) \tag{1}$$

## 3 Application to Census Data

### 3.1 Population based

### 3.2 Robustness analysis

## 4 Implementation

## 5 Conclusions