



Supporting secure dynamic alert zones using searchable encryption and graph embedding

Sina Shaham¹ · Gabriel Ghinita² · Cyrus Shahabi¹

Received: 2 January 2023 / Revised: 12 June 2023 / Accepted: 23 June 2023 / Published online: 18 July 2023
© The Author(s) 2023

Abstract

Location-based alerts have gained increasing popularity in recent years, whether in the context of healthcare (e.g., COVID-19 contact tracing), marketing (e.g., location-based advertising), or public safety. However, serious privacy concerns arise when location data are used in clear in the process. Several solutions employ searchable encryption (SE) to achieve *secure* alerts directly on encrypted locations. While doing so preserves privacy, the performance overhead incurred is high. We focus on a prominent SE technique in the public-key setting—hidden vector encryption, and propose a graph embedding technique to encode location data in a way that significantly boosts the performance of processing on ciphertexts. We show that the optimal encoding is NP-hard, and we provide three heuristics that obtain significant performance gains: gray optimizer, multi-seed gray optimizer and scaled gray optimizer. Furthermore, we investigate the more challenging case of dynamic alert zones, where the area of interest changes over time. Our extensive experimental evaluation shows that our solutions can significantly improve computational overhead compared to existing baselines.

Keywords Hidden vector encryption · Secure alert zones · Graph embedding

1 Introduction

Location data play an important part in offering customized services to mobile users. Whether they are used to find nearby points of interest, to offer location-based recommendations, or to locate friends situated in proximity to each other, location data significantly enrich the type of interactions between users and their favorite services. However, current service providers collect location data in clear, and often share it with third parties, compromising users' privacy. Movement data can disclose sensitive details about an individual's health status, political orientation, alternative lifestyles, etc. Hence,

it is important to support such location-based interactions while protecting privacy.

Our focus is on *secure alert zones*, a type of location-based service where users report their locations in encrypted form to a service provider, and then they receive alerts when an event of interest occurs in their proximity. This operation is very relevant to contact tracing, which is proving to be essential in controlling pandemics, e.g., COVID-19. It is important to determine if a mobile user came in close proximity to an infected person, or to a surface that has been exposed to the virus, but at the same time one must prevent against intrusive surveillance of the population. More applications of alert zones include public safety notifications (e.g., active shooter), and commercial applications (e.g., notifying mobile users of nearby sales events).

Searchable encryption (SE) [5, 16, 26] is very suitable for implementing secure alert zones. Users encrypt their location before sending it to the service provider using a special kind of encryption, which allows the evaluation of predicates directly on ciphertexts. However, the underlying encryption functions are not specifically designed for geospatial queries, but for arbitrary keyword or range queries. As a result, a data mapping step is typically performed to transform spatial queries to the primitive operations supported on

✉ Gabriel Ghinita
gghinita@hbku.edu.qa

Sina Shaham
sshaham@usc.edu

Cyrus Shahabi
shahabi@usc.edu

¹ Department of Computer Science, University of Southern California, Los Angeles, CA, USA

² College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

ciphertexts. Due to this translation, the performance overhead can be significant. Some solutions use *Symmetric Searchable Encryption (SSE)* [9, 16, 26], where a trusted entity knows the secret key of the transformation, and collects the location of all users before encrypting them and sending the ciphertext to the service provider. While the performance of SSE can be quite good, the system model that requires mobile users to share their clear text locations with a trusted service is not adequate from a privacy perspective, since it still incurs a significant amount of disclosure.

To address the shortcomings of SSE models, the work in [5] introduced the novel concept of hidden vector encryption (HVE), which is an *asymmetric* type of encryption that allows direct evaluation of predicates on top of ciphertext. Each user encrypts her own location using the *public* key of the transformation, and no trusted component that accesses locations in clear is required. This approach has been considered in the location context in [13, 18], with encouraging results. However, the performance overhead of HVE in the spatial domain remains high. Motivated by this fact, we study techniques to reduce the computational overhead of HVE. Specifically, we derive special types of spatial data mapping using graph embeddings, which allow us to express spatial queries with predicates that are less computationally-intensive to evaluate.

In existing HVE work for geospatial data [13, 18], the data domain is partitioned into a hierarchical data structure, and each node in this structure is assigned a binary string identifier. The binary representation of each node plays an important part in the query encoding, and it influences the amount of computation that needs to be executed when evaluating predicates on ciphertexts. However, the impact of the specific encoding is not evaluated in-depth. Our approach embeds the geospatial data domain to a high-dimensional hypercube, and then it applies graph embedding [7] techniques that directly target the reduction of computation overhead in the predicate evaluation step. Finally, no existing work considers the case of alert zones that change over time. Support for dynamic alert zones is very important, given that in most use case scenarios, phenomena of interest evolve over time (e.g., places visited by COVID carriers, area affected by a gas leak, etc). Our work tackles this important challenge¹.

Our specific contributions are:

- We introduce a novel transformation of the spatial data domain based on graph embedding and Gray codes that is able to model accurately the performance overhead

incurred when running HVE queries for spatial predicates;

- We transform the problem of minimizing HVE computation to a graph problem, and show that the optimal solution is NP-hard;
- We devise three heuristics: Gray Optimizer (GO), Multi-Seed Gray Optimizer (MSGO) and Scaled Gray Optimizer (SGO), that can solve the problem efficiently in the embedded space, while reducing significantly the computational overhead. The heuristics produce distinct trade-offs between the time required to compute the cell mappings and the runtime overhead when matching ciphertexts;
- We propose models that take into account the spatial and temporal evolution of alert zones, and choose encodings that improve performance under dynamic conditions;
- We perform an extensive experimental evaluation which shows that the proposed approaches are able to halve the performance overhead incurred by HVE when processing spatial queries.

The rest of the paper is organized as follows: Sect. 2 introduces necessary background on the system model, including an HVE primer. Section 3 provides the details of the proposed graph embedding transformation. Section 4 introduces several heuristic algorithms that solve the problem efficiently. Section 5 focuses on modeling of dynamic alert zones, and on advanced encodings under changing conditions. Section 6 evaluates thoroughly the proposed approach on real-life datasets. We survey related work in Sect. 7 and conclude in Sect. 8.

2 Background

2.1 System model

Consider a $[0,1] \times [0,1]$ spatial data domain divided into n non-overlapping partitions, denoted as

$$\mathcal{V} = \{v_1, v_2, \dots, v_n\}. \quad (1)$$

We use the term *cell* to refer to partitions, which can have an arbitrary size and shape. In practice, a well-established indexing technique can first be applied on top of the data domain to partition the space according to some application-specific criterion, such as expected density. For example, a k-d-tree can be used in conjunction with a public dataset of points of interest to split popular areas into more fine-grained cells, whereas coarser cells can be used for sparser areas.

An example of such a partitioning is provided in Fig. 3a. The system architecture of location-based alert system is represented in Fig. 1, and consists of three types of entities:

¹ This submission is an extended version of the work in [25]. Additional contributions include: a complexity analysis of Gray-based approaches (Sect. 4.3); novel techniques for supporting dynamic alert zones (Sect. 5); and an updated experimental section, including empirical evaluation of dynamic alert zones algorithms.

Table 1 Summary of notations

Symbol	Description
n	Number of cells
k	Length of HVE index
$\mathcal{V} = \{\cup v_i\}$	Set of all cells
C_j	Encrypted location of user j
TK	Token j
M_j	message of user j
$G_1(\mathcal{C}, \mathcal{E}_1)$	k -cube; vertices $\mathcal{C} = \{\cup c_i\}$; edges \mathcal{E}_1
$G_2(\mathcal{V}, \mathcal{E}_2)$	Complete graph; vertices $\mathcal{V} = \{\cup v_i\}$; edges \mathcal{E}_2
$(\mathcal{H}_1, \mathcal{H}_2, \mathcal{E}_3)$	Bipartite graph; node sets \mathcal{H}_1 and \mathcal{H}_2 ; edges \mathcal{E}_3
$\mathcal{D}_{i c_j}$	Set of nodes in \mathcal{C} with Hamming distance i from c_j
\mathcal{L}_x	Set of all complete x -bit gray cycles, $\mathcal{L}_x = \{\cup l_i\}$
S_n	state space with n cells
Q_n	state transition matrix with n cells
$\text{Pois}(\lambda)$	Poisson distribution; occurrence rate λ
s	Stationary distribution vector
\mathbf{i}	state i

1. **Mobile Users** subscribe to the alert system and periodically submit encrypted location updates.
2. The **Trusted Authority (TA)** is a trusted entity that decides which are the alert zones, and creates for each zone a search *token* that allows to check privately if a user location falls within the alert zone or not.
3. The **Server (S)** is the provider of the alert service. It receives encrypted updates from users and search tokens from TA, and performs the predicate evaluation to decide whether encrypted location C_i of user i falls within alert zone j represented by token TK_j . If the predicate holds, the server learns message M_i encrypted by the user, otherwise it learns nothing.

Table 1 summarizes the notations used throughout the manuscript.

The system supports location-based *alerts*, with the following semantics: a *Trusted Authority (TA)* designates a subset of cells as an *alert zone*, and all the users enclosed by those cells must be notified. The TA can be, for instance, the Center for Disease Control (CDC), who is monitoring cases of a pandemic, and wishes to notify users who may have been affected; or, the TA can be some commercial entity that the users subscribe to, and who notifies users when a sales event occurs at selected locations.

The *privacy requirement* of the system dictates that the server must not learn any information about the user locations, other than what can be derived from the match outcome, i.e., whether the user is in a particular alert zone or not. In case of a successful match, the server S learns that user u is enclosed by zone z . In case of a non-match, the server S learns only that the user is outside the zone z ,

but no additional location information. Note that, this model is applicable to many real-life scenarios. For instance, users wish to keep their location private most of the time, but they want to be immediately notified if they enter a zone where their personal safety may be threatened. Furthermore, the extent of alert zones is typically small compared to the entire data domain, so the fact that S learns that u is *not* within the set of alert zones does not disclose significant information about u 's location. The TA can be an organization such as CDC, or a city's public emergency department, which is trusted not to compromise user privacy, but at the same time does not have the infrastructure to monitor a large user population, and outsources the service to a cloud provider.

2.2 HVE encryption primer

Hidden vector encryption (HVE) [5] is a searchable encryption system that supports predicates in the form of conjunctive equality, range and subset queries. Search on ciphertexts can be performed with respect to a number of *index attributes*. HVE represents an attribute as a bit vector (each element has value 0 or 1), and the search predicate as a *pattern* vector where each element can be 0, 1 or '*' (star) that signifies a wildcard (or "don't care") value. Let l denote the HVE *width*, which is the bit length of the attribute, and consequently that of the search predicate. A predicate evaluates to *True* for a ciphertext C if the attribute vector I used to encrypt C has the same values as the pattern vector of the predicate in all positions that are non-star in the latter. Figure 2 illustrates the two cases of *Match* and *Non-Match* for HVE.

HVE is built on top of a symmetrical bilinear map of composite order [5], which is a function $e : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}_T$ such

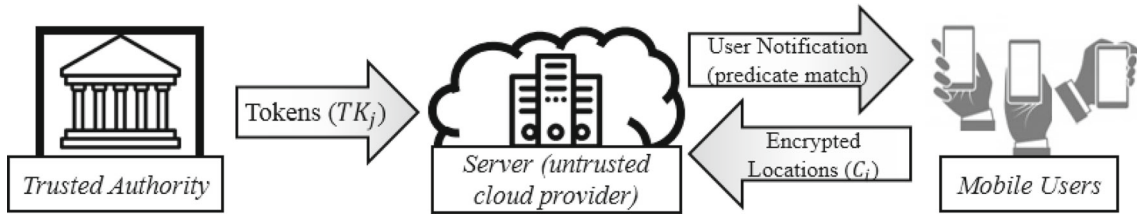


Fig. 1 Location-based alert system

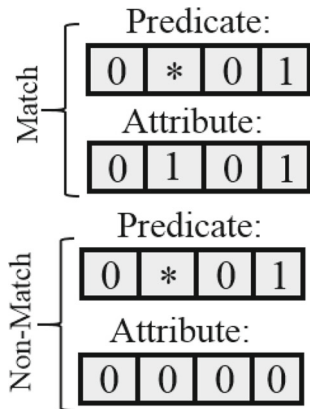


Fig. 2 HVE evaluation

that $\forall a, b \in G$ and $\forall u, v \in \mathbb{Z}$ it holds that $e(a^u, b^v) = e(a, b)^{uv}$. \mathbb{G} and \mathbb{G}_T are cyclic multiplicative groups of composite order $N = P \cdot Q$ where P and Q are large primes of equal bit length. We denote by $\mathbb{G}_p, \mathbb{G}_q$ the subgroups of \mathbb{G} of orders P and Q , respectively. Let l denote the HVE width, which is the bit length of the attribute, and consequently that of the search predicate. HVE consists of the following phases:

Setup. The *TA* generates the public/secret (PK/SK) key pair and shares PK with the users. SK has the form:

$$SK = (g_q \in \mathbb{G}_q, a \in \mathbb{Z}_p, \forall i \in [1..l] : u_i, h_i, w_i, g, v \in \mathbb{G}_p)$$

To generate PK , the *TA* first chooses at random elements $R_{u,i}, R_{h,i}, R_{w,i} \in \mathbb{G}_q, \forall i \in [1..l]$ and $R_v \in \mathbb{G}_q$. Next, PK is determined as:

$$PK = (g_q, V = vR_v, A = e(g, v)^a, \forall i \in [1..l] : U_i = u_i R_{u,i}, H_i = h_i R_{h,i}, W_i = w_i R_{w,i})$$

Encryption uses PK and takes as parameters index attribute I and message $M \in \mathbb{G}_T$. The following random elements are generated: $Z, Z_{i,1}, Z_{i,2} \in \mathbb{G}_q$ and $s \in \mathbb{Z}_n$. Then, the ciphertext is:

$$C = (C' = MA^s, C_0 = V^s Z, \forall i \in [1..l] : C_{i,1} = (U_i^{I_i} H_i)^s Z_{i,1}, C_{i,2} = W_i^s Z_{i,2})$$

Token Generation. Using SK , and given a search predicate encoded as pattern vector I_* , the *TA* generates a search token TK as follows: let J be the set of all indices i where $I_*[i] \neq *$. *TA* randomly generates $r_{i,1}$ and $r_{i,2} \in \mathbb{Z}_p, \forall i \in J$. Then

$$TK = (I_*, K_0 = g^a \prod_{i \in J} (u_i^{I_*[i]} h_i)^{r_{i,1}} w_i^{r_{i,2}}, \forall i \in [1..l] : K_{i,1} = v^{r_{i,1}}, K_{i,2} = v^{r_{i,2}})$$

Query is executed at the server, and evaluates if the predicate represented by TK holds for ciphertext C . The server attempts to determine the value of M as

$$M = C' / (e(C_0, K_0) / \prod_{i \in J} e(C_{i,1}, K_{i,1}) e(C_{i,2}, K_{i,2})) \quad (2)$$

If the index I based on which C was computed satisfies TK , then the actual value of M is returned, otherwise a special number which is not in the valid message domain (denoted by \perp) is obtained.

2.3 Problem statement

Prior work [13, 18] assumed that all cells are equally likely be in an alert zone. However, that is not the case in practice. Some parts of the data domain (e.g., denser areas of a city) are more likely to become alert zones. The cost of encrypted alert zone enclosure evaluation is given by the number of operations required to apply HVE matching at the service provider. As we discussed in our HVE primer in Sect. 2.3, the evaluation cost is directly proportional to the number of non-star bits in the tokens. Armed with knowledge about the likelihood of cells to be part of an alert zone, one can create superior encodings that reduce processing overhead.

Our goal is to find an enhanced encoding that reduces non-star bits for a given set of alert zone tokens. Denote by $p(v_i)$ the probability of cell v_i being part of an alert zone. The mutual probability of multiple cells indicates how likely they are to be part of the same alert zone. Given individual cell probabilities, the mutual probability of a set of i cells

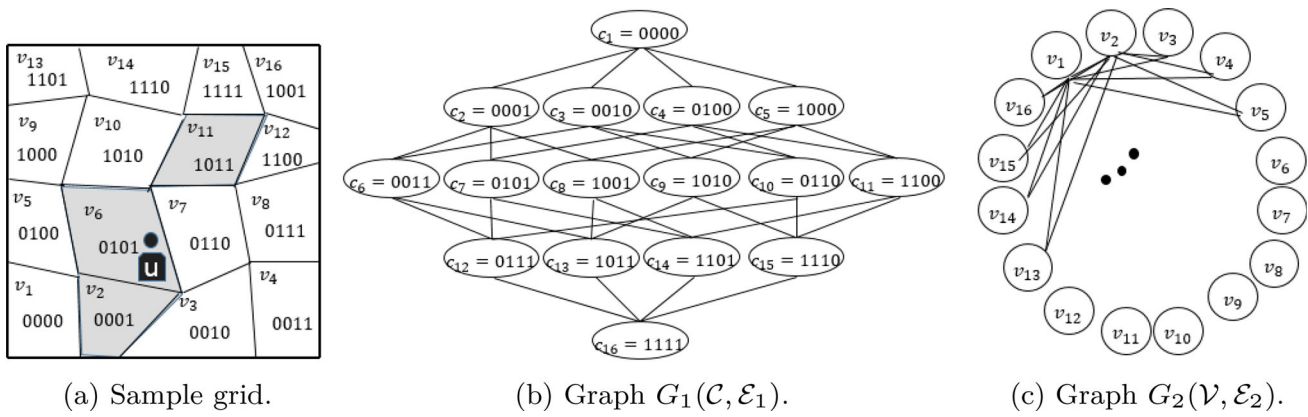


Fig. 3 An example of embedding graphs generated based on a sample grid

$\mathcal{L} = \{v'_1, v'_2, \dots, v'_i\}$ is calculated as:

$$p(\mathcal{L}) = \prod_{j=1}^i p(v'_j). \tag{3}$$

The problem we study is formally presented as follows:

Problem 1 Find an encoding of the grid that on average reduces the number of non-star bits in the tokens generated from alert zone cells.

In the above formulation, the correlation between cells becoming part of an alert zone is assumed to be negligible. In essence, the assumption is that cells are independent in time and space (in Sect. 5, we provide an advanced modeling of the correlation of alert zones over space and time).

3 Location domain mapping through graph embedding

Our approach minimizes the number of non-star bits in alert zone tokens by modeling the data domain partitioning as an embedding problem of a k -cube onto a complete graph. We denote a k -cube as $G_1(\mathcal{C}, \mathcal{E}_1)$, where $\mathcal{C} = \{c_1, c_2, \dots, c_n\}$ and $c_i = \{0, 1\}^k$. Figure 3b illustrates a k -cube generated based on the sample partitioning in Fig 3a. In G_1 , two nodes c_i and c_j are connected if their Hamming distance is equal to one. We refer to such a bit as Hamming bit.

Definition 1 (Hamming Distance and Bits). The Hamming distance between two indices c_i and c_j in $G_1(\mathcal{C}, \mathcal{E}_1)$ is the minimum number of substitutions required to transform c_i to c_j , denoted by the function $d_h(\cdot)$. We refer to the bits that need to be substituted as the Hamming bits of the indices.

Example 1 The Hamming distance between indices $c_1 = 0100$ and $c_2 = 0010$ is two ($d_h(c_i, c_j) = 2$), and the Hamming bits are the second and third most significant bits of the indices.

The second graph required to formulate the problem of minimizing the number of non-stars is a complete graph generated by all cells in the partitioning, denoted by $G_2(\mathcal{V}, \mathcal{E}_2)$. The set \mathcal{V} represents the nodes corresponding to cells, and an undirected edge connects every two nodes in G_2 .

Note that, every token (including those containing stars), can be related to several cycles on the k -cube. For example, token 00^{**} represents four indices $0000, 0001, 0010, 0011$, which correspond to cycles (c_1, c_2, c_6, c_3) and (c_1, c_3, c_6, c_2) on the k -cube in Fig. 3b. Unfortunately, there is no one-to-one correspondence between the tokens and the cycles. In particular, for a larger number of stars, there exist several cycles representing the same token. To generate a one-to-one correspondence, we incorporate Binary-Reflected Gray (BRG) encoding on the k -cube to create unique cycles corresponding to tokens.

Definition 2 (BRG path on k -cube). A BRG path between two nodes with non-zero Hamming distance is defined as the path on the k -cube going from one node to another based on BRG coding on Hamming bits.

As an example, the Hamming bits between 0001 and 1000 are the least and most significant bits, and the BRG path connecting them on the k -cube in Fig. 3b includes indices $0001, 1001, \text{ and } 1000$ in the given order. One can see that as the BRG codes are unique, the BRG path between two indices on the k -cube is also unique. This characteristic of BRG paths is formulated in Lemma 1.

Lemma 1 A BRG path between two nodes on a k -cube is unique.

Proof The uniqueness of the path between two nodes on the k -cube follows from the uniqueness of BRG code, as only one such path can be constructed. \square

Definition 3 (Complete x -bit BRG cycle). Given a k -cube, a complete x -bit BRG cycle is a cyclic BRG path with the

length of 2^x , in which only x bits are affected. We denote the set of all possible complete x -bit BRG cycles by $\mathcal{L}_x = \{\cup l_i\}$.

Example 2 In Fig. 3b, token *0** entails eight indices 0000, 0001, 0011, 0010, 1010, 1011, 1001, 1000. This token maps uniquely to the complete 3-bit BRG cycle on the 4-cube with nodes $(c_1, c_2, c_6, c_3, c_9, c_{13}, c_8, c_5)$ and start point c_1 .

We can uniquely associate a token to a cycle on the k -cube. Consider a token with k bits and x stars. This token is mapped to a complete x -bit BRG cycle on the k -cube, starting from a node in which all the star bits are set to zero. Such a cycle is unique and has a length of 2^x . Based on this mapping, every token is associated with a unique cycle on the k -cube, and every complete x -bit BRG cycle is mapped to a unique token with x -stars. Therefore, there is a one-to-one correspondence between tokens and complete BRG cycles. The formulation of Problem 1 based on graph embedding can be written as follows:

Problem 2 Given two graphs $G_1(\mathcal{C}, \mathcal{E}_1)$ and $G_2(\mathcal{V}, \mathcal{E}_2)$, find a mapping function $\mathcal{F} : G_1 \rightarrow G_2$ with the objective to

$$\text{Maximize } \left\{ \sum_{i=1}^k p(\mathcal{L}_i) \right\}. \tag{4}$$

3.1 Gray optimizer (GO)

The problem of embedding a complete graph within a minimized size k -cube has been shown to be NP-hard [7]. We develop an heuristic algorithm called *Gray Optimizer* that solves Problem 2. Consider an initial node of the complete graph $v_r \in \mathcal{V}$, and without loss of generality assume that it is assigned to index c_1 . We refer to nodes in \mathcal{G}_1 interchangeably using their vertex id or binary index. The optimization problem can be formulated as follows.

Problem 3 Given two graphs $G_1(\mathcal{C}, \mathcal{E}_1)$ and $G_2(\mathcal{V}, \mathcal{E}_2)$, and the node $v_r \in \mathcal{V}$ assigned to index c_1 , find a mapping function $\mathcal{F} : G_1 \rightarrow G_2$ that

$$\text{Maximize } \left\{ \sum_{i=1}^k p(\mathcal{L}_i|v_r) \right\}. \tag{5}$$

Problem 2 requires an assignment of vertices in G_2 to the nodes of G_1 such that the probability of complete BRG cycles is maximized; whereas Problem 3 seeks to maximize the probability of cycles with respect to a particular node, in this case v_r , which is assigned to the index c_1 . A reasonable candidate for assignment to c_1 is the cell with the highest probability, as it is most likely to be part of an alert zone. To solve this problem, we propose the heuristic in Algorithm 1. The input of the algorithm is the root index $c_1 \in G_1$, the root node $v_r \in G_2$ (also called seed) and the graphs G_1 and G_2 .

Algorithm 1: Gray Optimizer.

Input : $G_1; G_2; c_1; v_r$
 1 Sort nodes in G_2 based on probabilities
 2 Assign v_r to c_1
 3 **for** i in $[1 : k]$ **do**
 4 Initialize $\mathcal{H}_1, \mathcal{H}_2 = \emptyset$
 5 $\mathcal{H}_1 \leftarrow \left\{ \binom{k}{i} \right.$ non-assigned nodes in G_2 with the highest probability
 6 **for** $c_j \in \mathcal{D}_{i|c_1}$ **do**
 7 Calculate $p(l_j/c_j) = \prod_{v \in l_j/c_j} p(v)$
 8 $\mathcal{H}_2 \leftarrow p(l_j/c_j)$
 9 **end**
 10 Sort nodes in \mathcal{H}_2
 11 Match vertices in \mathcal{H}_1 to \mathcal{H}_2
 12 **end**

Denote by $\mathcal{D}_{i|c_1}$ the set of nodes on \mathcal{C} that have a Hamming distance of i from c_1 . Note that $\mathcal{D}_{i|c_1}$ includes $\binom{k}{i}$ nodes, each one having a Hamming distance of i from c_1 . The overall assignment structure is as follows: first, Algorithm 1 assigns the remaining nodes of \mathcal{V} of the graph \mathcal{G}_2 to nodes in $\mathcal{D}_{1|c_1}$. After assignment of all nodes in $\mathcal{D}_{1|c_1}$, the algorithm assigns the nodes in $\mathcal{D}_{2|c_1}$ and follows the same process until all nodes are assigned ($\mathcal{D}_{1|c_1}$ to $\mathcal{D}_{k|c_1}$). An initial sorting of nodes in \mathcal{V} is conducted at the start of the algorithm, and is used throughout the assignment process to reduce the computation complexity.

The assignment objective in stage i of the process is to maximize $p(\mathcal{L}_i|v_r)$.

Note that (5) can be written as:

$$\sum_{i=1}^k \text{Maximize } \{p(\mathcal{L}_i|v_r)\}. \tag{6}$$

where $p(\mathcal{L}_i|v_r)$ represents the probability of all complete i -bit BRG cycles that include c_1 ($v_r \rightarrow c_1$). Denote such a cycle by l . Based on the following lemma, there exists one and only one node c_j in l that has a Hamming distance of i from c_1 , which means that $c_j \in \mathcal{D}_{i|c_1}$. Therefore, every complete i -bit BRG cycle given index c_1 includes one node in $\mathcal{D}_{i|c_1}$. On the other hand, every node in $\mathcal{D}_{i|c_1}$ corresponds to a unique complete i -bit BRG cycle passing through c_1 , as it results from Lemma 1. Therefore, all complete i -bit BRG cycles are considered in stage i and we maximize their probabilities in this stage of the assignment.

Lemma 2 For each node c_i in a complete x -bit BRG cycle, there exists one and only one node with the Hamming distance of x from c_i .

Proof A complete x -bit BRG cycle includes 2^x nodes and only x bits are affected. Therefore, the only index that can exist with the Hamming distance of x from c_i is the one in which all x Hamming bits are flipped. \square

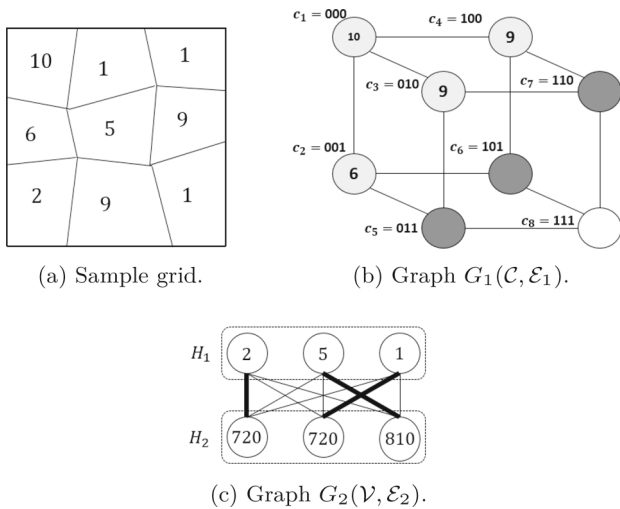


Fig. 4 An example of embedding graphs generated based on a sample grid

The assignment process in the stage i of GO creates a bipartite graph, i.e., $(\mathcal{H}_1, \mathcal{H}_2, \mathcal{E}_3)$, where \mathcal{H}_1 and \mathcal{H}_2 are two set of nodes, and \mathcal{E}_3 represents the set of edges. In this stage, the nodes in sets $\mathcal{D}_{1|c_1}, \mathcal{D}_{2|c_1}, \dots, \mathcal{D}_{i-1|c_1}$ are already assigned and we aim to find the best assignment for the nodes in $\mathcal{D}_{i|c_1}$ such that $p(\mathcal{L}_i|v_r)$ is maximized. Among the remaining nodes in \mathcal{V} , we choose $\binom{k}{i}$ of them that have the highest probabilities, as $|\mathcal{D}_{i|c_1}| = \binom{k}{i}$, and allocate them to \mathcal{H}_1 .

On the other hand, for each node c_j in $\mathcal{D}_{i|c_1}$, we construct the unique complete i -bit BRG cycle including c_j and c_1 . Let us represent this cycle by l_j . Note that all nodes included in l_j are assigned except c_j . The algorithm calculates the probability of the set of nodes in l_j excluding c_j and allocates it to a node in \mathcal{H}_2 . Based on (3), this probability can be calculated as:

$$p(l_j \setminus \{c_j\}) = \prod_{v \in l_j \setminus \{c_j\}} p(v), \tag{7}$$

The algorithm repeats the process for all nodes in $\mathcal{D}_{i|c_1}$. Next, the nodes in \mathcal{H}_2 are sorted, and the best matching is conducted between these two sets of nodes by assigning the i^{th} node of \mathcal{H}_1 to the i^{th} node of \mathcal{H}_2 . The optimality of the matching process is proven in Lemma 3, and the achievement of maximal assignment in each stage is proven in Lemma 4 (Fig. 4).

Lemma 3 Suppose in the i^{th} step of the algorithm h_1 to $h_{\binom{k}{i}}$ are the members of \mathcal{H}_1 and h'_1 to $h'_{\binom{k}{i}}$ are the members of \mathcal{H}_2 such that $h_1 \leq h_2 \leq \dots \leq h_{\binom{k}{i}}$ and $h'_1 \leq h'_2 \leq \dots \leq h'_{\binom{k}{i}}$. The optimal value of matching is achieved when h_i is matched with h'_i .

Proof Suppose that the converse is true. Hence, there exist two nodes h_i and h_k which are paired with h'_j and h'_t , respectively, such that $h_i \leq h_k$ and $h'_j \geq h'_t$. Since the current matching is maximal by swapping h'_j and h'_t , we have

$$h_i h'_j + h_k h'_t + R > h_i h'_t + h_k h'_j + R, \tag{8}$$

where R indicates the remaining pairing summation. Rewriting equation (9) results in

$$(h_i - h_k) \times (h'_j - h'_t) > 0. \tag{9}$$

However, $h_i \leq h_k$ and $h'_j \geq h'_t$, therefore, the left hand side of the equation is always less than or equal to zero, which is a contradiction. The case for equality of equation (9) is removed as swapping does not change the summation and the lemma holds. \square

Lemma 4 In stage i , GO maximizes $p(\mathcal{L}_i|v_r)$ given the currently assigned nodes $(\mathcal{D}_{1|c_1}, \mathcal{D}_{2|c_1}, \dots, \mathcal{D}_{i-1|c_1})$.

Proof We prove the lemma based on mathematical induction. *Base case:* For $i = 1$, given that the node v_r is assigned to c_1 , we aim to prove that GO maximizes $p(\mathcal{L}_1|v_r)$. To start with, GO chooses $\binom{k}{1}$ remaining nodes of \mathcal{V} for the purpose of assignment. The optimal assignment of nodes in $\mathcal{D}_{1|c_1}$ is a permutation of the chosen nodes; otherwise, they could be replaced with a node with a higher probability that would result in a higher value for $p(\mathcal{L}_1|v_r)$. Next, the algorithm generates a bipartite graph $(\mathcal{H}_1, \mathcal{H}_2, \mathcal{E}_3)$. The probability of chosen nodes are allocated to \mathcal{H}_1 , and the nodes in \mathcal{H}_2 represent the probability of complete 1-bit gray cycles constructed from $c_j \in \mathcal{D}_{1|c_1}$ and the node c_1 , excluding the probability of c_j itself. Next, the optimal matching is done by assigning the j^{th} maximum node in \mathcal{H}_2 to the j^{th} maximum node in \mathcal{H}_1 , achieving maximal $p(\mathcal{L}_1|v_r)$ given the node c_1 .

Induction step: Let us assume that GO has maximized the probabilities of complete x -bit BRG cycles for $x = 1$ to $i - 1$ in stages one to $i - 1$. We prove that in stage i , the algorithm maximizes complete i -bit gray cycles, given the previously assigned nodes.

Based on Lemma 2, all complete i -bit BRG cycles are considered in stage i , as each such cycle includes exactly one node in $\mathcal{D}_{i|c_1}$, which has the highest Hamming distance from c_1 . GO starts by choosing the cells with the highest probabilities and assigning them to \mathcal{H}_1 . Same as in the base case, we know that the optimal assignment in this stage includes the chosen set of nodes. Next, the nodes in \mathcal{H}_2 are assigned based on finding the probability of complete i -bit BRG cycles for nodes in $\mathcal{D}_{i|c_1}$, excluding the nodes themselves from the probability. As the matching process is optimal match, the best permutation of nodes in \mathcal{H}_1 is matched to complete i -bit BRG cycles. \square

4 Scaling up gray optimizer

The GO algorithm can lead to significant improvements in the processing of HVE operations; however, there are two major drawbacks once the algorithm is applied to grids with high granularities. (i) The complexity of the algorithm creates a processing time bottleneck for its application in HVE; (ii) The calculation of probabilities for large complete BRG cycles may result in numerical inaccuracies. To make GO applicable to grids with higher levels of granularity, we propose two variations.

The first proposed algorithm, called *Multiple Seed Gray Optimizer (MSGO)* (Sect. 4.1), generates non-overlapping clusters and applies GO within each one of them. The second algorithm, called Scaled Gray Optimizer (SGO) (Sect. 4.2) takes a *Breadth-First Search (BFS)* [17] approach. The performance of BFS is preferred to its counterpart *Depth-First Search (DFS)* as the nodes closer to the seed have higher probabilities. Thus, it is reasonable to consider those nodes earlier in the process.

4.1 Multiple seed gray optimizer (MSGO)

The starting point of the GO algorithm, which we refer to as *seed*, was chosen as the node in G_2 with the maximum probability. However, the algorithm can work starting with any initial seed, then follow the assignment process for other nodes in ascending order of their Hamming distance from the seed. Furthermore, as BRG cycles become larger, their associated probability becomes smaller. Thus, one way to reduce the complexity of GO is to run the algorithm up to a particular *depth*. Essentially, the algorithm aims at optimizing BRG cycles up to a certain length. We enhance GO by running Algorithm 1 with multiple seeds, and also by limiting the depth of the assignment.

Definition 4 *Depth*: For a given seed c_j , the GO algorithm is said to run with a depth of i if it only considers the assignment of nodes in $\mathcal{D}_{1|c_j}, \mathcal{D}_{2|c_j}, \dots, \mathcal{D}_{i|c_j}$.

The pseudocode of the proposed approach is presented in Algorithm 2. The algorithm starts by assigning the node with the highest probability in G_2 to the origin of G_1 or a random index. However, instead of running GO with respect to this index for all depths from one to k , MSGO runs GO with the specified depth as input. The algorithm completes the process of assignment for a cluster of indices in G_1 . MSGO then chooses a random index of G_1 among the remaining indices and assigns it to the node in G_2 with maximum probability among remaining nodes. Similarly, this index is used as a seed for GO with the specified depth and generates a new cluster. The cluster-based approach continues until all nodes are assigned to an index. The algorithm supports variable cluster sizes based on the underlying application.

Algorithm 2: Multiple Seed Gray Optimizer (MSGO).

Input : $G_1; G_2; depth$

- 1 Sort nodes in G_2 based on probabilities
- 2 Select a random index on G_1 which is not currently assigned
- 3 Assign the index with the node that has the maximum probability in G_2
- 4 Apply Algorithm 1 on the selected index with the specified depth
- 5 **Repeat** lines 2-4 **until** all indices are assigned

The MSGO algorithm provides a robust solution for grids with higher granularity. The algorithm no longer suffers the drawbacks of GO when the grid size grows, such as numerical inaccuracies in the calculation of the probability of large cycles. The complexity of the algorithm depends on the depth chosen as input, and in low depths, it can be implemented in $\mathcal{O}(n(\log_2 n))$. MSGO can significantly reduce the number of operations required for the implementation of HVE in location-based alert systems, and therefore, making it a practical solution for preserving the privacy of users in location-based alert systems.

4.2 Scaled gray optimizer (SGO)

SGO considers overlapping clusters and necessitates that all nodes act as seed during the assignment process. The pseudocode of the proposed approach is presented in Algorithm 3. SGO starts by assigning the node with the highest probability to an index on G_1 . However, instead of assigning indices with all depths from one to k with respect to index c_1 , the SGO algorithm runs GO with the depth of one. Next, SGO sorts the indices in $\mathcal{D}_{1|c_1}$ based on their assigned probabilities in descending order and runs GO with the depth of one on each index. Once the algorithm is applied on all the indices in $\mathcal{D}_{1|c_1}$, the process repeats for indices in $\mathcal{D}_{2|c_1}, \mathcal{D}_{3|c_1}, \dots$, etc. The algorithm continues until all indices are assigned to a node.

Algorithm 3: Scaled Gray Optimizer (SGO).

Input : $G_1; G_2$

- 1 Sort nodes in G_2 based on probabilities
- 2 Assign $v_r \in G_2$ with the highest probability to the origin of G_1 , i.e., c_1
- 3 Apply Algorithm 1 on c_1 with the depth of one
- 4 **for** i in $[1 : k]$ **do**
- 5 Sort $\mathcal{D}_{i|c_1}$ in descending order of probabilities assigned to its indices
- 6 **for** c_j in $\mathcal{D}_{i|c_1}$ **do**
- 7 Apply Algorithm 1 on c_j with the depth of one
- 8 **end**
- 9 **end**

4.3 Complexity analysis

The key computation overhead of the GO algorithm is in the calculation of probability of BRG cycles. Let the function $T(\cdot)$ return the computational complexity. In the i^{th} step of the algorithm, the nodes with the hamming distance of i from c_1 are assigned to an index on the k -cube, i.e., $\mathcal{D}_{i|c_1}$. The number of nodes in $\mathcal{D}_{i|c_1}$ is $\binom{\log_2(n)}{i}$. For each one of such nodes the complete BRG path is calculated which requires the multiplication of $2^i - 1$ probabilities. Therefore, the assignment process for the nodes in $\mathcal{D}_{i|c_1}$ requires

$$T(\mathcal{D}_{i|c_1}) = (2^i - 2) \times \binom{\log_2(n)}{i} \quad (10)$$

operations. Hence, the total number of operations required for the algorithm is

$$\bigcup_i T(\mathcal{D}_{i|c_1}) = \sum_{i=1}^{\log_2(n)} (2^i - 2) \times \binom{\log_2(n)}{i} \quad (11)$$

$$= \sum_{i=1}^{\log_2(n)} 2^i \binom{\log_2(n)}{i} - 2 \sum_{i=1}^{\log_2(n)} \binom{\log_2(n)}{i}. \quad (12)$$

From the binomial theorem,

$$\sum_{i=1}^{\log_2(n)} 2^i \binom{\log_2(n)}{i} = (1 + 2)^{\log_2(n)} - 1 = 3^{\log_2(n)} - 1, \quad (13)$$

and

$$\sum_{i=1}^{\log_2(n)} \binom{\log_2(n)}{i} = 2^{\log_2(n)} - 1. \quad (14)$$

Therefore, Eq. (12) can be written as

$$\bigcup_i T(\mathcal{D}_{i|c_1}) = n^{\log_2 3} - 2n + 1. \quad (15)$$

In addition to the above operations, there exists an initial sorting of the probabilities that can be implemented based on merge sort with the complexity of $\mathcal{O}(n \log_2 n)$, and a sorting process in each stage for the nodes in \mathcal{H}_2 . For the latter, the complexity can be written as

$$\sum_{i=1}^{\log_2(n)} \mathcal{O}\left(\binom{\log_2(n)}{i} \times \log_2\left(\binom{\log_2(n)}{i}\right)\right) = \quad (16)$$

$$\sum_{i=1}^{\log_2(n)} \mathcal{O}\left(\log_2\left(\binom{\log_2(n)}{i}\right)^{\binom{\log_2(n)}{i}}\right) \leq \quad (17)$$

$$\sum_{i=1}^{\log_2(n)} \mathcal{O}\left(\log_2 n^{\binom{\log_2(n)}{i}}\right) \leq \quad (18)$$

$$\mathcal{O}\left(\log_2 n^{\sum_{i=1}^{\log_2(n)} \binom{\log_2(n)}{i}}\right) = \mathcal{O}(n \log_2 n) \quad (19)$$

Therefore, accounting for sorting, the closed-form expression for the total complexity is $\mathcal{O}(2n \log_2 n) + n^{\log_2 3} - 2n + 1$.

The MSGO algorithm is based on executing the GO algorithm with shorter depths in a cluster based approach. Suppose that the depth is set to r where $r \leq \log_2 n$. Running the algorithm in each cluster with similar logic as the GO requires the following number of operations.

$$\sum_{i=1}^r (2^i - 2) \binom{\log_2(n)}{i}. \quad (20)$$

On the other hand, the total number of clusters is approximately

$$\#clusters \approx n / \sum_{i=1}^r \binom{\log_2(n)}{i}. \quad (21)$$

Therefore, the total complexity considering the initial sorting algorithm is calculated as

$$s \mathcal{O}(n \log_2 n) + \left(\sum_{i=1}^r (2^i - 2) \binom{\log_2(n)}{i} \right) \times \left(n / \sum_{i=1}^r \binom{\log_2(n)}{i} \right). \quad (22)$$

Defining the binary entropy function as

$$H_2(x) = x \log_2\left(\frac{1}{x}\right) + (1 - x) \log_2\left(\frac{1}{1 - x}\right), \quad (23)$$

the following approximation can be used for deriving closed-form expression for various cluster sizes in Eq. (22)

$$\binom{n}{i} \simeq 2^{n H_2(r/n)}. \quad (24)$$

Lastly, the SGO algorithm executes the GO algorithm with the depth of one and has the computational complexity of $\mathcal{O}(n \log_2 n)$. The low computational complexity of SGO makes it a suitable option for the encoding of grids with higher levels of granularity.

5 Supporting dynamic alert zones

So far, we considered the case of static alert zones, and we optimized the data encoding and token generation under this

scenario. However, in practice, alert zones vary over time. Whether an alert corresponds to a natural phenomenon (e.g., gas leak) or a human activity (e.g., COVID carrier movement), alert zones exhibit spatio-temporal patterns that must be accounted for in order to obtain fast performance.

We maintain the grid-based partition of the spatial domain used for the static case, and we denote by *state* of the grid the set of all alert cells at a given time. The occurrence probability of a state can be modeled analytically and used as a basis for grid encoding. The higher the statistical model accuracy, the more precise the encoding becomes, reducing HVE operations overhead. Next, we build a comprehensive statistical model to characterize alert zone evolution in space and time.

Definition 5 (State Space). For a given grid

$$\mathcal{V} = \{v_1, v_2, \dots, v_n\},$$

let X be a random variable defined on all possible subsets of the cells. The state space of X is defined as the power set $\mathcal{S}_n = \{\mathbf{1}, \mathbf{2}, \dots, \mathbf{2}^n\}$.

The *cardinality* of a state \mathbf{i} represents the number of cells included in the state and is denoted by $|\mathbf{i}|$. The set of all states with the cardinality of j are denoted by $\mathcal{S}_n^{|\mathbf{j}|}$. Note that, the notation is not concerned with a precise order of states. For example, a grid with two cells $\{v_1, v_2\}$ leads to the state space of

$\mathcal{S}_2 = \{\{\emptyset\}, \{v_1\}, \{v_2\}, \{v_1, v_2\}\}$, which is depicted by $\mathcal{S}_2 = \{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}\}$; however, the order of states is not captured by the notation. Two examples of such an assignment can be $\{\mathbf{1} = \{\emptyset\}, \mathbf{2} = \{v_1\}, \mathbf{3} = \{v_2\}, \mathbf{4} = \{v_1, v_2\}\}$, and $\{\mathbf{1} = \{\{v_1, v_2\}\}, \mathbf{2} = \{v_2\}, \mathbf{3} = \{v_1\}, \mathbf{4} = \emptyset\}$. We provide more details on the construction of the state space and ordering in Sect. 5.4.

Let $X_0, X_1, \dots, X_i, \dots$ denote the sequence of random variables modeling the occurrence of alert zones. The set of possible values for X_i is the state space of the grid, and the index i denotes the evolution of the process in time. The probability of X_i being in a particular state \mathbf{j} is denoted as $p(X_i = \mathbf{j})$. The probability of a cell becoming part of an alert zone depends on underlying phenomena properties, existing correlations among cells, and the history of alert zones on the map. Moreover, probabilities do not remain constant over time. We identify several distinct scenarios, and we create a statistical model for each: (i) the states are independent in both space and time; (ii) the states are independent in space, but dependent in time (i.e., temporal causality); (iii) the states are independent in time but exhibit space correlation (i.e., spatial causality); and (iv) the states are dependent in both time and space. The first case corresponds to the static case introduced in the previous sections; the last case is the most general one, whereas cases (ii) and (iii) are special cases of (iv). Each case may be relevant under different

types of applications and data domains. Next, we investigate in details each of the cases, and propose a data encoding and token generation technique for each. Our goal is to obtain an accurate representation of how the probabilities X_i are distributed over the state space.

5.1 Independence in time and space

Having the independence assumption in space and time greatly simplifies the problem formulation as the sequence of random variables $X_0, X_1, \dots, X_i, \dots$ become a sequence of independent and identically distributed (iid) random variables defined over the state space. Such modeling indicates that the random variables X_1 to X_i provide no information about the random variable X_{i+1} . Therefore, the *probability mass function (PMF)* of X_i depends on the probabilities of individual cells. For a given X_i , the probability of cell $v_i \in \mathcal{V}$ becoming part of the alert zone is denoted by $p(v_i)$, and corresponds to a value between zero and one. The mutual probability of a subset of cells $\mathcal{L} = \{v_1, v_2, \dots, v_i\}$ being in an alert zone can be calculated as

$$p(\mathcal{L}) = \prod_{j=1}^i p(v_j). \quad (25)$$

The calculation of mutual probabilities is the direct result of the independence assumption, which indicates that there are no correlations between cells.

5.2 Independence in space, dependence in time

In this case, the grid state no longer consists of iid random variables following the same PMFs. The probability of state \mathbf{i} at time j is no longer assumed to be equal to the probability of being in state \mathbf{i} at a different time k , i.e., $p(X_j = \mathbf{i}) \neq p(X_k = \mathbf{i})$. Our objective is to determine whether the system reaches a steady state in which the probabilities no longer change significantly over time. We model the evolution of alert cells over time using Markov chains. We assume that alert zones evolve incrementally by addition or removal of a single cell at a time (this can always be achieved by properly choosing the time granularity). Our choice of modeling approach for alert zone evolution is guided by the fact that several types of natural phenomena have been successfully captured using Markov chains, such as gas leaks [27], infectious disease spread [28] and even earthquakes [6]. Markov models have also been used extensively to model human mobility, which would be relevant to alerts that are triggered by human behavior (e.g., traffic jams, active shooter situations).

The proposed model is represented in Fig. 5. States \mathbf{i} and \mathbf{j} are connected if and only if the difference between their

cardinality is one, $|\mathbf{i} - \mathbf{j}| = 1$. The only exception is the state including all cells (if all cells are within the alert zone, then all have the same status). The model assumes that each state depends only on the previous state, and therefore, it follows Markov chain properties, i.e., for all $k \geq 0$,

$$\begin{aligned}
 p(X_{k+1} = \mathbf{j} | X_k = \mathbf{i}, X_{k-1} = \mathbf{i}_{k-1}, \dots, X_0 = \mathbf{i}_0) \\
 = p(X_{k+1} = \mathbf{j} | X_k = \mathbf{i}).
 \end{aligned}
 \tag{26}$$

The forward propagation to a state with a higher cardinality indicates the addition of an alert cell, whereas forward propagation to a state with a lower cardinality indicates the removal of an alert cell.

The value of $p(X_{k+1} = \mathbf{j} | X_k = \mathbf{i})$ is called the *transition probability* from state \mathbf{i} to state \mathbf{j} and we implicitly make the assumption that the transition probabilities are homogeneous over time. We are interested in understanding what the likelihood of being in a state is starting from any other state, and whether the chain reaches a *stationary distribution* in which the probabilities of individual states do not change over time.

First, we review three properties of the proposed Markov chain:

Property 1 *All states in the proposed model are recurrent. Therefore, starting from any state of the chain, it is possible to reach any other state, eventually.*

Property 2 *The proposed Markov chain is irreducible, as for any two states \mathbf{i} and \mathbf{j} , it is possible to reach one from the other in a finite number of steps.*

Property 3 *The proposed Markov chain for modeling alert zones is aperiodic, as the period of states is equal to one.*

The above properties help to characterize the long-term behaviour of the Markov chain. If after a certain period of time the *transition matrix* of the chain reaches a stationary distribution, it enables us to know the probability of each state in the state space. The state *transition matrix* is defined as follows:

Definition 6 (Transition matrix). For a Markov chain $X_0, X_1, \dots, X_i, \dots$ with a state space $\mathcal{S}_n = \{\mathbf{1}, \mathbf{2}, \dots, \mathbf{2}^n\}$, let $q_{ij} = p(X_{k+1} = \mathbf{j} | X_k = \mathbf{i})$ be the transition probability from state \mathbf{i} to state \mathbf{j} . The $2^n \times 2^n$ matrix $Q_n = (q_{ij})$ is called the transition matrix of the chain. The value of q_{ij} for $i < 2^n$ is defined as $p(v)$, where v is the alert cell which exists in state \mathbf{i} (row) and does not exist in state \mathbf{j} (column).

Recall that two states are connected if and only if their cardinality differs by one. The last row of the matrix represents the only outgoing directed edge from the state with the cardinality of n to the state with cardinality of zero. Thus,

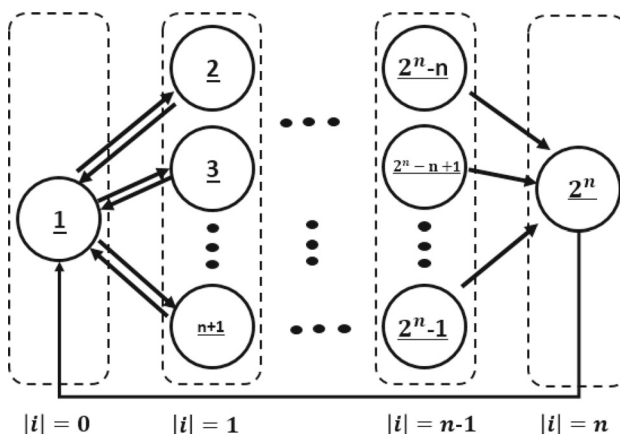


Fig. 5 Proposed Markov model for alert zone evolution

the first element of the last row is one ($q_{2^n 1} = 1$) and all its other elements are zero. Such a row ensures the aperiodicity of the chain.

It can be inferred that the i^{th} row of the transition matrix corresponds to outgoing edges from the state \mathbf{i} of the Markov chain. Therefore, in order for the matrix to maintain the Markovian properties, the values in each row should sum up to one, which is indeed the case for the proposed transition matrix. This property is termed as *Markovian matrix property*. Let a row vector $\mathbf{t} = [t_1, t_2, \dots, t_{2^n}]$ be the PMF of X_0 , where $t_i = p(X_0 = \mathbf{i})$. Then, based on the properties of Markovian chains, the marginal distribution of X_m is given by the j^{th} component of $\mathbf{t} Q_n^m$, i.e., $p(X_n = \mathbf{j})$. The marginal distribution indicates that given a initial state \mathbf{i} , the probability of being in state \mathbf{j} after m transitions is the j^{th} component of the vector $\mathbf{t} Q_n^m$. We are interested in the long run behaviour of the system and to understand if the proposed model will reach a *stationary distribution*.

Definition 7 (Stationary distribution). Given a Markov chain with the transition matrix Q_n , a row vector $\mathbf{s} = [s_1, \dots, s_{2^n}]$, such that $s_i \geq 0$ and $\sum_i s_i = 1$, is a stationary distribution if

$$\mathbf{s} Q_n = \mathbf{s}
 \tag{27}$$

We elaborate further on the meaning of the vector \mathbf{s} . Suppose that the i^{th} element of the vector corresponds to the state \mathbf{i} . If the proposed Markov chain reaches a stationary distribution, this value represents probability $p(X_n = \mathbf{i})$ for any n after reaching the stationary distribution. Thus, the importance of each state is revealed by its corresponding value in \mathbf{s} .

Example 3 Consider a map with two cells v_1 and v_2 , where $p(v_1) = 0.2$ and $p(v_2) = 0.8$. The state space includes four states $\{\{\emptyset\}, \{v_1\}, \{v_2\}, \{v_1, v_2\}\}$, and the transition matrix

is calculated as

$$Q_2 = \begin{bmatrix} 0 & 0.2 & 0.8 & 0 \\ 0.2 & 0 & 0 & 0.8 \\ 0.8 & 0 & 0 & 0.2 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

Solving Eq. (27) for the matrix Q_2 results in the eigen vector $s = [0.4310, 0.0862, 0.3448, 0.1379]$. Hence, the probability of states are $p(\{\emptyset\}) = 0.4310, p(\{v_1\}) = 0.0862, p(\{v_2\}) = 0.3448, p(\{v_1, v_2\}) = 0.1379$.

There are three important questions to be answered about the stationary distribution: (a) *does it exist?* (b) *is it unique?* and (c) *does the Markov chain converge to the stationary distribution?* The stationary distribution is the left eigenvector of the transition matrix corresponding to the eigenvector of one as shown by Eq. (27). The existence and uniqueness of a stationary distribution for the proposed Markov model is proven in the following theorem.

Theorem 1 *There exists a unique stationary distribution for the proposed Markov chain to model alert zones.*

Proof According to [2], a stationary distribution exists for any finite-state Markov chain, and if the chain is irreducible, the solution is unique. Based on property 2, there exists a unique stationary distribution for the model. Later in Sect. 5.4, we present the recursive construction of matrix Q_n and show that the cardinality of the null space of the matrix $s(Q_n - I)$ is one. \square

The above theorem shows that there exists a unique stationary distribution for the proposed Markov model regardless of the initial probabilities of the cells; however, to reach the stationary distribution, the chain needs to be aperiodic as well as irreducible. Based on Property 3, the proposed model is aperiodic. However, particular initial probabilities, including zero values, can result in periodic chains. To address this problem, we adopt a similar approach as the PageRank algorithm [23], used to rank the relevance of webpages. Suppose that before moving to a new state on the chain, a coin is tossed with probability α of heads. If the result of the coin toss is heads, the state evolves using the transition matrix Q ; otherwise, the system jumps to a state in a uniformly random distribution. The resulting transition matrix is represented as:

$$O_n = \alpha Q + (1 - \alpha) \frac{J_n}{2^n}, \tag{28}$$

where J_n is a $2^n \times 2^n$ matrix of all ones. The recommended value [23] of α is 0.85. It can be observed that all elements of O_n are positive, and therefore, the aperiodicity of the chain is guaranteed. Hence, solving Eq. (27) for O_n has a solution leading to a stationary distribution (s) as well as converging

to the stationary distribution. Similarly, the i^{th} element of the vector s for the new transition matrix O_n indicates the significance of state \mathbf{i} , as it represents $p(X_m = \mathbf{i})$ for any large value of m . In the following, we consider that the transition matrix is aperiodic, and we use the matrix Q_n as our reference.

Example 4 Going back to Example 3, the transition matrix is derived as

$$O_2 = \begin{bmatrix} 0.0375 & 0.2075 & 0.7175 & 0.0375 \\ 0.2075 & 0.0375 & 0.0375 & 0.7175 \\ 0.7175 & 0.0375 & 0.0375 & 0.2075 \\ 0.8875 & 0.0375 & 0.0375 & 0.0375 \end{bmatrix}.$$

Solving Eq. (27) for the matrix O results in the eigenvector of $s = [0.4111, 0.1074, 0.3171, 0.1644]$. Hence, the new probability of states are $p(\{\emptyset\}) = 0.4111, p(\{v_1\}) = 0.1074, p(\{v_2\}) = 0.3171, p(\{v_1, v_2\}) = 0.1644$. One can check the convergence by choosing a large enough value of m and calculating tO_2^m starting by an arbitrary PMF on t . As an example, if the vector $t = [0.25, 0.25, 0.25, 0.25]$, then tO_2^{50} will result in

$$[0.4111, 0.1074, 0.3171, 0.1644]$$

which is the stationary distribution vector s .

5.3 Dependence in both space and time

In this section, we study how to capture correlation among alert cells over time by incorporating spatial distance between cells within the Markov model. We embed spatial correlations in the transition matrix while maintaining Markovian properties, and thus the long-term behaviour of the model can be better defined. We use as starting point the proposed model from Fig. 5.

Consider a grid with two cells $\{v_1, v_2\}$ and the state space of $S_2 = \{\mathbf{1} = \{\emptyset\}, \mathbf{2} = \{v_1\}, \mathbf{3} = \{v_2\}, \mathbf{4} = \{v_1, v_2\}\}$. The matrix Q_2 is derived as

$$Q_2 = \begin{bmatrix} 0 & p(\{v_1\}) & p(\{v_2\}) & 0 \\ p(\{v_1\}) & 0 & 0 & p(\{v_2\}) \\ p(\{v_2\}) & 0 & 0 & p(\{v_1\}) \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Investigating the transition matrix closely, one can see the impact of independence between cells in the matrix. Consider the entry $Q_2(2, 4)$ as an example. This entry indicates the probability of going from state $\mathbf{2} = \{v_1\}$ to state $\mathbf{4} = \{v_1, v_2\}$ is $p(\{v_2\})$. In other words, the transition captures the fact that the existence of another alert zone cell v_1 did not impact

the cell v_2 (i.e., spatial independence between cells). More formally, from the *Bayes rule*:

$$p(\{v_1, v_2\}) = p(v_2|v_1)p(v_1) \rightarrow p(\{v_1, v_2\}) = p(v_2)p(v_1), \quad (29)$$

given that

$$p(v_2|v_1) = p(v_2). \quad (30)$$

In Sect. 5.2, we assumed independence between states. To address this issue, we propose the following method to capture the correlations between states without eliminating the Markov property of the matrix Q_n . The main idea behind the approach is that cells that are in close proximity to the alert zone are more likely to become part of the zone in the future.

Let $X_0, X_1, \dots, X_i, \dots$ be an order one Markov sequence of random variables modeling the occurrence of the alert zones, where X_i 's are defined over the state space of the grid. Without loss of generality assume that the j^{th} row of the matrix Q corresponds to the state $\{v_1, v_2, \dots, v_j\}$. Based on the proposed Markov model in Fig. 5, it is known that this state can evolve by the addition or removal of a single alert cell. Therefore, there exist n non-zero elements in each row of the matrix. For all $v_k \in \mathcal{V}$, we calculate the probability of its removal or addition as:

If $v_k \notin \{v_1, v_2, \dots, v_j\}$ then;

$$p(\{v_k\} \cup \{v_1, v_2, \dots, v_j\}) = p(v_k)/(d(v_k, c)) \times \beta$$

If $v_k \in \{v_1, v_2, \dots, v_j\}$ then;

$$p(\{v_1, v_2, \dots, v_j\} - \{v_k\}) = p(v_k)/(d(v_k, c)) \times \beta,$$

where the function $d(\cdot)$ returns the Euclidean distance between two points, β is a normalization factor over the entire row, and the point c is the centre point of $\{v_1, v_2, \dots, v_j\}$, calculated as

$$c = \left(\sum_{i=1}^j v_j \right) / j. \quad (31)$$

Note that, in all above calculations, each cell's center point is used as its representative. The intuition behind the approach is that the correlation between cells becomes smaller as we go further away from the alert zone. The only special case is when there exists a single-cell alert zone, and we seek the probability of its removal. In this case $d(v_k, c)$ becomes close to zero and $p(v_k)/(d(v_k, c))$ tends to go to infinity. As there exist no other alert zone cell for this case, we consider this probability as $p(v_k)$ instead of $p(v_k)/(d(v_k, c))$ to avoid inaccuracies. As an example, consider a grid with three cells $\{v_1, v_2, v_3\}$ and the average point c . Suppose that

the j^{th} row of the matrix Q_3 corresponds to the state $\{v_1, v_2\}$. In this row, there exist three nonzero elements:

$$\begin{aligned} p(\{v_1\}) &= p(v_2)/(d(v_2, c)) \times \beta \\ p(\{v_2\}) &= p(v_1)/(d(v_1, c)) \times \beta \\ p(\{v_1, v_2, v_3\}) &= p(v_3)/(d(v_3, c)) \times \beta, \text{ where} \\ \beta &= 1/(p(v_2)/(d(v_2, c)) + p(v_1)/d(v_1, c) + p(v_3)/d(v_3, c)) \end{aligned} \quad (32)$$

The proposed method satisfies the Markovian matrix property. Hence, it can be used as part of the Markov model in Sect. 5.2 to capture the long-term behavior of the system.

5.4 Recursive construction and monte carlo sampling

Finding the eigenvector of matrix Q_n corresponding to eigenvalue one is necessary to determine the probability of being in a particular state at a given time $p(X_n = \mathbf{i})$. The eigenvector provides valuable information that enables us to prioritize more likely states in the grid encoding process. However, there are two important issues with its calculation: (i) The matrix Q_n has dimensions of $2^n \times 2^n$. Even considering a small grid with 100 cells, it requires an extremely large storage capacity. (ii) The calculation of the eigenvector for such a large matrix is expensive, with $\mathcal{O}(n^3)$ [21] complexity. For example, based on Householder transformations, eigenvalues and eigenvectors can be calculated with complexity $\mathcal{O}(n^2) + 4n^3/3$. To address the high computational overhead, we approximate the stationary distribution based on random walks on the Markov model.

We start by explaining the recursive construction of the matrix Q_n . The rows and columns of the matrix depend on the order in which states are chosen. We propose to construct the states of the $n + 1$ cells, v_1 to v_{n+1} , from the grid with n cells, v_1 to v_n as follows:

$$\mathcal{S}_2 = \{\{\emptyset\}, \{v_1\}, \{v_2\}, \{v_1, v_2\}\}, \quad (33)$$

$$\mathcal{S}_{n+1} = \{\mathcal{S}_n, \mathcal{S}_n \cup v_{n+1}\}. \quad (34)$$

For instance, \mathcal{S}_3 is constructed as

$$\begin{aligned} \mathcal{S}_3 = \{ & \{\emptyset\}, \{v_1\}, \{v_2\}, \{v_1, v_2\}, \\ & \{v_3\}, \{v_1, v_3\}, \{v_2, v_3\}, \{v_1, v_2, v_3\} \}, \end{aligned} \quad (35)$$

The matrix Q_{n+1} can be constructed recursively as

$$Q_n = \begin{bmatrix} W_{n-1} & p(v_n)I_{2^{n-1}} \\ p(v_n)I_{2^{n-1}} & W_{n-1} \end{bmatrix} - W_n(2^n, :) + K_{2^n}, \quad (36)$$

where I_{2^n} is the identity matrix and K_{2^n} is an all-zero $2^n \times 2^n$ matrix except for element $K_{2^n}(2^n, 0) = 1$, and

$$W_n = \begin{bmatrix} W_{n-1} & p(v_n)I_{2^{n-1}} \\ p(v_n)I_{2^{n-1}} & W_{n-1} \end{bmatrix} \quad (37)$$

given that

$$W_2 = \begin{bmatrix} 0 & p(v_1) & p(v_2) & 0 \\ p(v_1) & 0 & 0 & p(v_2) \\ p(v_2) & 0 & 0 & p(v_1) \\ 0 & p(v_2) & p(v_1) & 0 \end{bmatrix}.$$

The above representation of Q_n works under the spatial independence assumption, but the construction of states holds regardless of that assumption.

To tackle the high computational complexity of determining eigenvectors, we use a probabilistic approach. PageRank's approach [23] to this problem is incorporating the power iteration method to calculate the eigenvectors, but still incurs a high computational complexity. An alternative approach is the Monte Carlo approximation, which is widely used in literature and results in an enhanced estimation of the stationary distribution. The Monte Carlo method provides several advantages over deterministic power iteration methods such as significantly lower computation complexity, opportunities for parallel implementation, and it facilitates updating of probabilities.

The main idea behind the Monte Carlo approximation is to start R random walks on the Markov model's primary node, i.e., state **1**. Each random walk terminates with the probability of $1 - c$ and makes a transition to the next outgoing node with the PMF specified in the transition matrix Q_n . The fraction of walks ending at a state over all the random walks indicates the probability or significance of that state. The vector of calculated probabilities for all states is the approximation of stationary distribution. The number of samples required to estimate the stationary distribution is shown to pessimistically be in the order $\mathcal{O}(n^2)$, where n indicates the number of states; however, it is shown that n random walks are enough to provide a reasonable approximation of a stationary distribution [1].

6 Experimental evaluation

6.1 Experimental setup

We conduct our experiments on a 3.40GHz core-i7 Intel processor with 8GB RAM running 64-bit Windows 7 OS. The code is implemented in Python, and we used the LogicMin Library [10] for binary minimization of token expressions.

We compare the proposed approaches (GO, MGSO and SGO) against the hierarchical Gray encoding technique from [13] (labeled *HGE*), the state-of-the-art in location alerts on HVE-encrypted data.

To model the probability of partition cells becoming alert zones, we use the sigmoid function $\mathcal{S}(x) = 1/(1 + \exp^{-b(x-a)})$, where a and b are parameters controlling the function shape. The output value is between zero and one. The sigmoid function is a frequent model used in machine learning, and we chose it because in practice, the probability of individual cells becoming part of an alert zone can be computed using such a model built on a regions' map of features (e.g., type of terrain, building designation, point-of-sale information, etc). Parameter a of the sigmoid controls the *inflection* point of the curve, whereas b controls the gradient.

For each data point in the graphs, we average results over 500 queries. We use ten distinct seeds for all the random components of our solution, and we display error bars (standard deviation of results) for performance improvement figures. (We excluded the error bars from the HVE operation count results, since their variability may be high due to the difference in queries. Our purpose is to measure the improvement in overhead per query among methods, and not differences in overhead for distinct queries, which are expected to differ significantly even for the same method).

Section 6.2 evaluates GO in comparison with the HGE benchmark from [12]. Section 6.3 investigates the performance of the proposed MSGO and SGO heuristics relative to GO. Section 6.4 provides an analysis of the effect that imperfect knowledge about cell probabilities has on performance for GO, MSGO and SGO. Finally, Sect. 6.5 focuses on techniques that support dynamic alert zones.

6.2 Gray optimizer evaluation

GO is our core proposed algorithm to reduce the number of HVE operations required to support alert zones. Specifically, by *HVE operations* we refer to the computation executed by the server to determine matches between tokens and encrypted user locations. Recall that, for each non-star item in a token, a number of expensive bilinear map operations are required. GO aims to minimize the number such non-star items in tokens by choosing an appropriate encoding of the domain. Our comparison benchmark is the approach from [13] which uses a hierarchical quadtree structure to partition the data domain. We refer to this approach as *HGE*, and we present our result as an improvement in terms of computation overhead compared with [13].

6.2.1 Improvement in HVE operations

Figure 6 summarizes the evaluation results of GO for three logistic function parameter settings. The grid size is set to

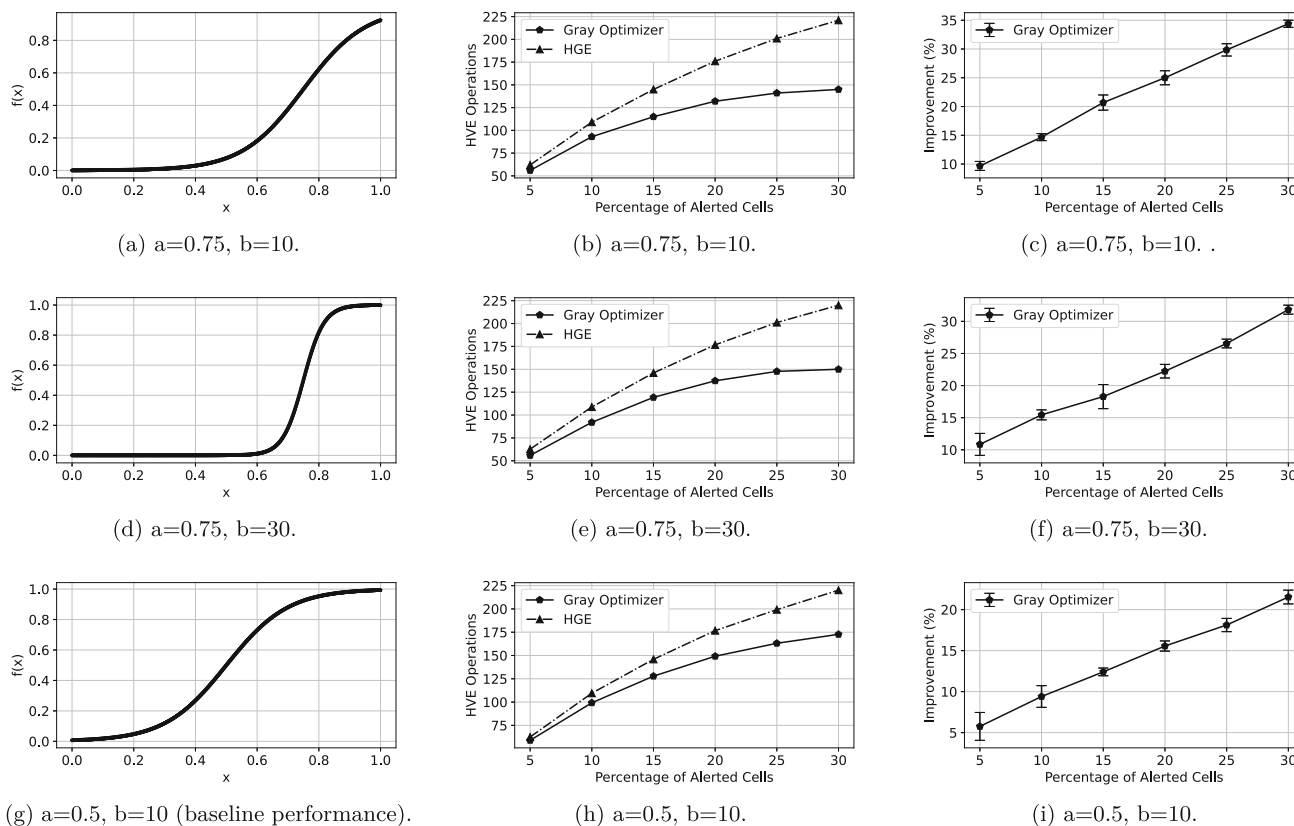


Fig. 6 Evaluation of GO, grid size = 100 cells

100 cells (recall from our earlier discussion that GO can only support relatively low granularities). Figure 6 shows the total number of bilinear pairings performed for a ciphertext-token pair. GO clearly outperforms the approach from [13]. The relative gain in performance of GO increases when the size of the alert zone increases (i.e., when there are more grid cells covered by the alert zone). This can be explained by the fact that a larger input set gives GO more flexibility to optimize the encoding and decrease the number of non-star entries in a token. In terms of percentage gains, GO can improve performance by up to 40%, which is quite significant. Also, note that the gains are significant for all parameters of the sigmoid function used. In general, we identified that a higher a value leads to more pronounced gains. This is an encouraging factor, because a higher a corresponds to a more skewed probability case, where a relatively small number of cells are more likely to be included in an alert zone than others. In practice, one would expect that to be the case, since events that trigger alerts also tend to be concentrated over a relatively small area (e.g., very popular hotspots, certain facilities that present higher risks, like a chemical plant, etc.).

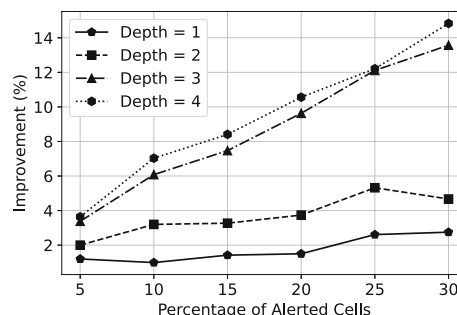


Fig. 7 Performance evaluation of GO for varying depth (100 cells)

6.2.2 Impact of depth

Recall that the reduction in computation achieved by GO depends on the depth at which the algorithm is run (GO works similar to a depth-first search graph algorithm). In general, running the algorithm with a higher depth will produce better results in terms of performance gain at runtime (i.e., when matching is performed at the server), but it also requires a lot more computational time to compute a good encoding (which is a one-time cost). Figure 7 captures the impact of depth on improvement. In this experiment, GO is executed on a single cell with different depths, and the remaining cells are

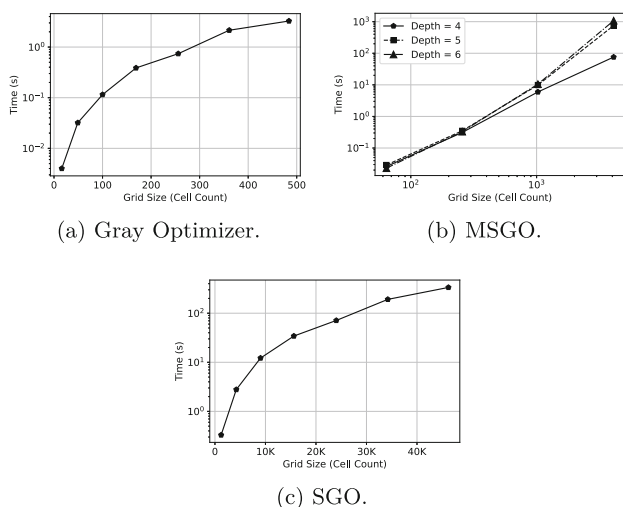


Fig. 8 Execution time

assigned randomly (the experiment is specifically designed to show the effect of using lower depths on GO). As expected, there is a clear increasing trend, with higher depths resulting in better improvement factors. However, after a sharp initial gain (illustrated by the large distance between the chart graphs corresponding to depths 2 and 3), the improvement stabilizes, and it may no longer be worth increasing the depth of the computation considerably (the gains are stabilizing between depths 3 and 4).

6.2.3 Execution time

Figure 8a illustrates the execution time of GO. Recall that, the execution time of GO is influenced by the granularity of the grid (finer granularities increase execution time). The results show that GO can complete within a short execution time for smaller grid sizes; however, as the grid granularity increases, there is a sharp increase in execution time. Therefore, GO may not be practical to apply for high granularity grids, and that is the main motivation behind our two variations, MSGO and SGO (which are evaluated next). Moreover, as the grid granularity increases, the length of cycles becomes larger, which will also result in numerical inaccuracies when executing GO. The execution time required by GO for values up to 600 cells is around 10s. We observed that this value is the maximum number of cells for which GO performs reasonably; beyond this level, the algorithm is not suitable due to increased execution time and numerical inaccuracies associated with the calculation of probabilities for large cycles.

6.3 Evaluation of GO variations on higher granularity grids

As discussed previously, GO does not perform well when directly applied to high granularity grids. To improve the computational complexity of GO, we proposed two extensions of the algorithm, namely, MSGO and SGO. Next, we evaluate experimentally both these variations.

6.3.1 MSGO

Figure 9 illustrates the performance of MSGO for various algorithms depths. Unlike the single seed GO, we are able to evaluate the performance of MSGO for grids with much higher granularity (i.e., 1024 cells in this case). There is a similar trend in terms of gain as we have observed with GO, where larger alert zones provide more opportunities for advantageous encodings, and thus overall performance is improved (the percentage of HVE operations eliminated is higher). The relative gain obtained is very close to 40% compared to the benchmark. Also, the absolute amount of improvement is better than for GO in all cases. This occurs due to the fact that MSGO can support higher-granularity grids, and in this setting there is more flexibility in choosing a good encoding (due to the larger number of cells, there are significantly more choices for our algorithm). As expected, increasing the depth of MSGO leads to higher improvement percentage, but the trade-off is a larger computation complexity.

Comparing Figs. 6 and 9, we remark that the MSGO algorithm obtains similar performance gains as the core algorithm GO for low granularity grids, but with a much lower computational overhead. For high granularity grids, GO cannot keep up in terms of computational overhead, whereas MSGO scales reasonably well, and it is able to still obtain significant improvements. One main reason is that MSGO no longer requires the calculation of probabilities of large cycles, avoiding numerical inaccuracies and reducing overall computational overhead. The complexity of the algorithm can be as low as $\mathcal{O}(n(\log_2 n))$ depending on the chosen depth value, which provides a robust and efficient solution for reducing the number of HVE operations.

The execution time of MSGO is illustrated in Fig. 8b. The graph indicates that even for a high level of granularity, such as 4,000, the algorithm requires less than 15 minutes to encode the grid, depending on the specified depth at the input. As expected, by increasing the depth of the algorithm, better performance can be achieved in terms of HVE operations, at the cost of higher computational overhead. The MSGO algorithm can be extended for an arbitrary number of cells on the grid, and also it may have various cluster sizes depending on the application.

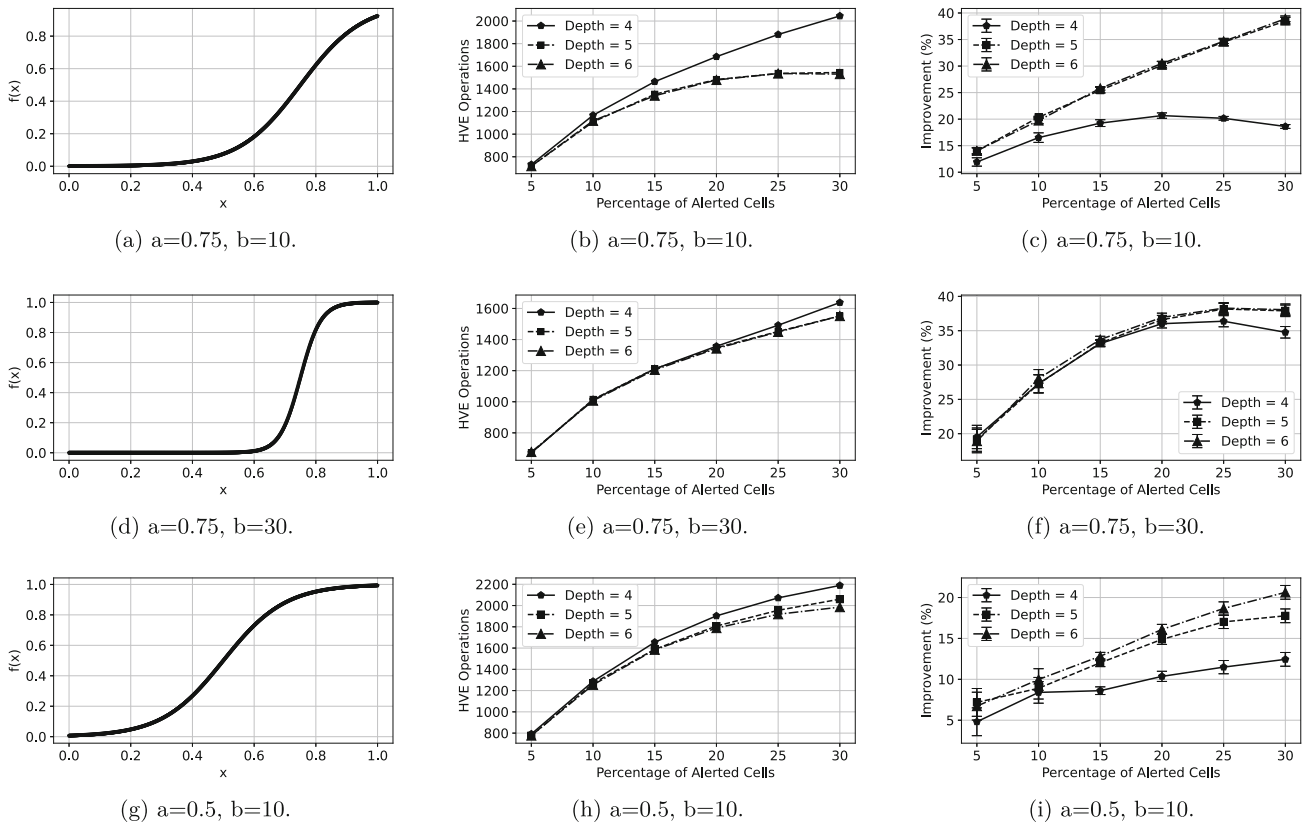


Fig. 9 Performance evaluation of MSGO (grid size = 1024 cells)

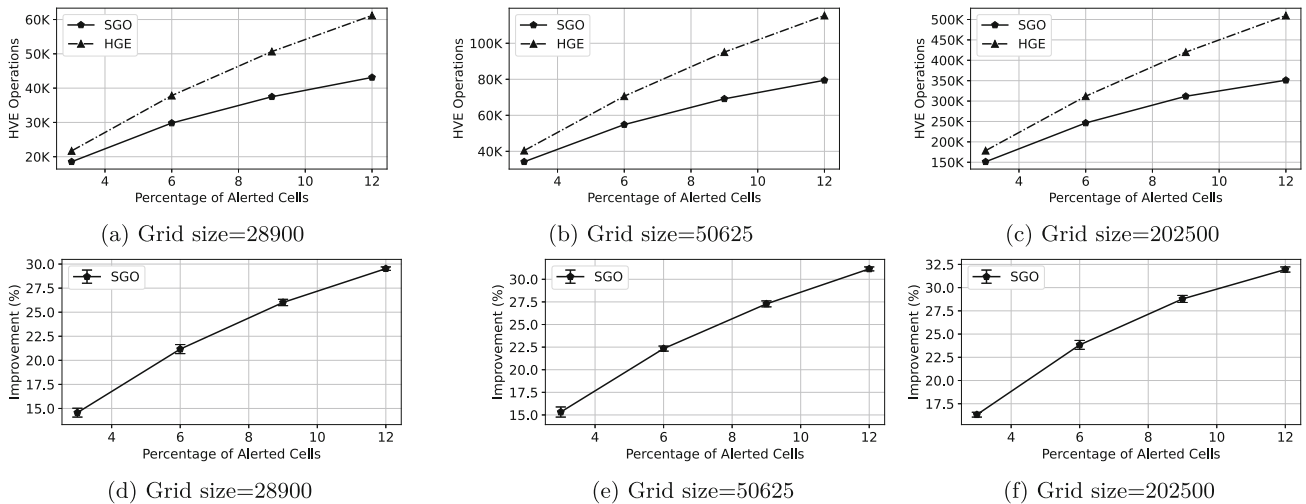


Fig. 10 SGO performance evaluation, varying grid size

6.3.2 SGO

Figure 10 illustrates the performance gain obtained by SGO. In this experiment, we focused on applying the algorithm to much larger number of cells, up to 202, 500 (which is equivalent to a 450 × 450 square grid). Similar to the MSGO algorithm, the improvement achieved by SGO occurs even

when the alert zones are small. Since the overall number of cells is larger, the SGO algorithm has even more flexibility in choosing an advantageous encoding, resulting in strong performance gains. For example, at 9% ratio of alert cells, the SGO algorithm results in 25.8, 26, and 27.3% improvements for grid sizes of 10, 000, 28, 900, and 50, 625, respectively.

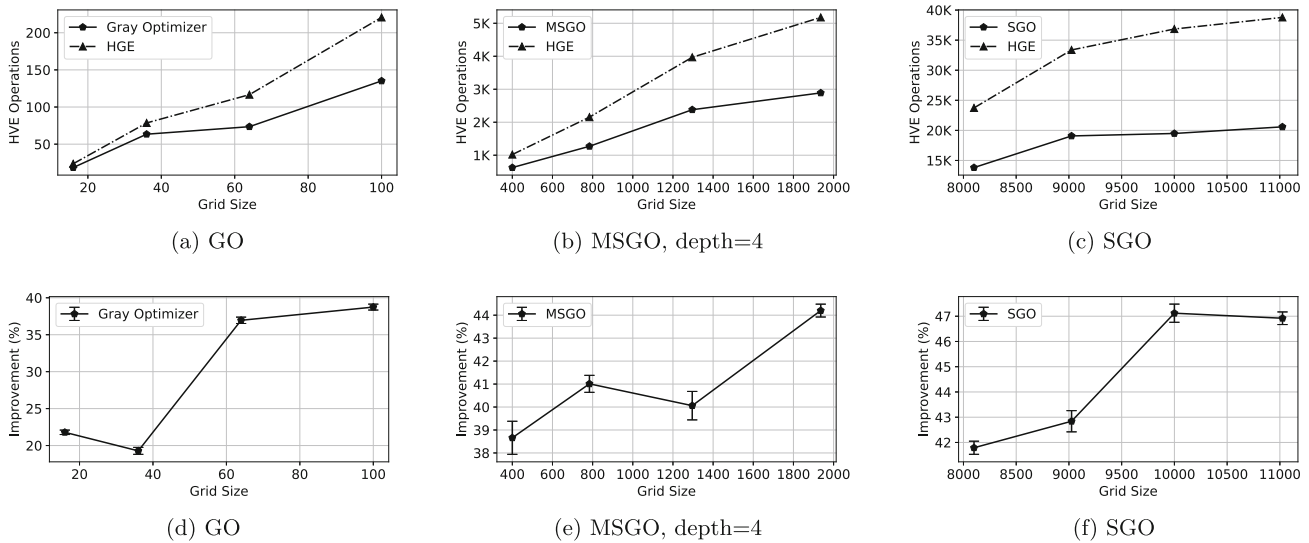


Fig. 11 Performance of algorithms, varying grid size, 30% of cells on alert

The execution time of SGO is shown in Fig. 8c. Even for very large grid sizes, such as 50, 625, the algorithm requires less than six minutes to encode the grid. Therefore, the system can be set to regularly update the probabilities and run the algorithm at six-minute intervals, if needed. To compare this time performance with GO, consider the maximum grid size for which the encoding can be computed within 60s in each case. As shown in Fig. 8a, this number corresponds to a grid size of 1200 for GO, whereas in a similar time, SGO can be applied on the grid size of 22,000 cells. Therefore, the SGO algorithm requires significantly lower computation overhead to execute compared with GO and even MSGO algorithms, while the performance gain in terms of HVE operations reductions is still solid.

Figure 11 presents the result of algorithms by fixing the percentage of alerted cells to 30% and varying the grid size. It can be seen that the performance improvement of algorithms stays in a comparable margin for varying grid sizes. The slight fluctuation in graphs is due to two primary reasons (i) as all codewords have the same length, increasing the quantization level result in an addition of a bit to all codewords, and (ii) the initial probabilities are assigned to the cells in a random process based on the sigmoid activation function.

6.4 Imperfect probabilities information

The knowledge of cell probabilities plays an important role in the reduction of HVE operations. These probabilities are input to GO and its extensions SGO and MSGO, used to find an enhanced encoding of space. Having imperfect initial cell probabilities can negatively impact the performance of algorithms by deviating the optimization result. Therefore, we aim at investigating the effect of imperfect initial prob-

abilities on the improvement achieved compared with the previous work (HGE). This is done by the addition of noise to cell probabilities at the input of algorithms modeling the inaccuracies that might exist. Let us briefly illustrate how the addition of noise is conducted. Given the vector of cell probabilities:

$$\text{Probability of alert} = [p(v_1), p(v_2), \dots, p(v_n)],$$

each probability is added with an iid uniformly distributed random noise n between $[0, u]$, where u indicates the maximum noise value. For example, if the percentage of noise is 20%, this value is set to 0.2, and the random noise is generated uniformly in the interval of $[0, 0.2]$. Doing so, the transformed probabilities are acquired as

$$[p(v_1)', p(v_2)', \dots, p(v_n)'], \text{ where } p(v_i)' = p(v_i) + n_i.$$

Note that the values are considered to be cyclic between zero and one, i.e., if the noise of 0.5 is added with a cell probability of 0.8, the resulting value would be recorded as 0.3. Hence, with 100% of noise, it is expected that the numbers would be uniformly distributed.

Figure 12 indicates the sensitivity of GO, MSGO, and SGO to imperfect probability values used as input. For each algorithm, the number of HVE operations required is shown as well as the improvement gained in the performance compared with the previous work. The x -axis represents the percentage of noise added to the perfect probability information varied between 0 to 100, and the y -axis indicates HVE operations required side by side to the improvement achieved. The percentage of alerted cells is set to 40% in all graphs.

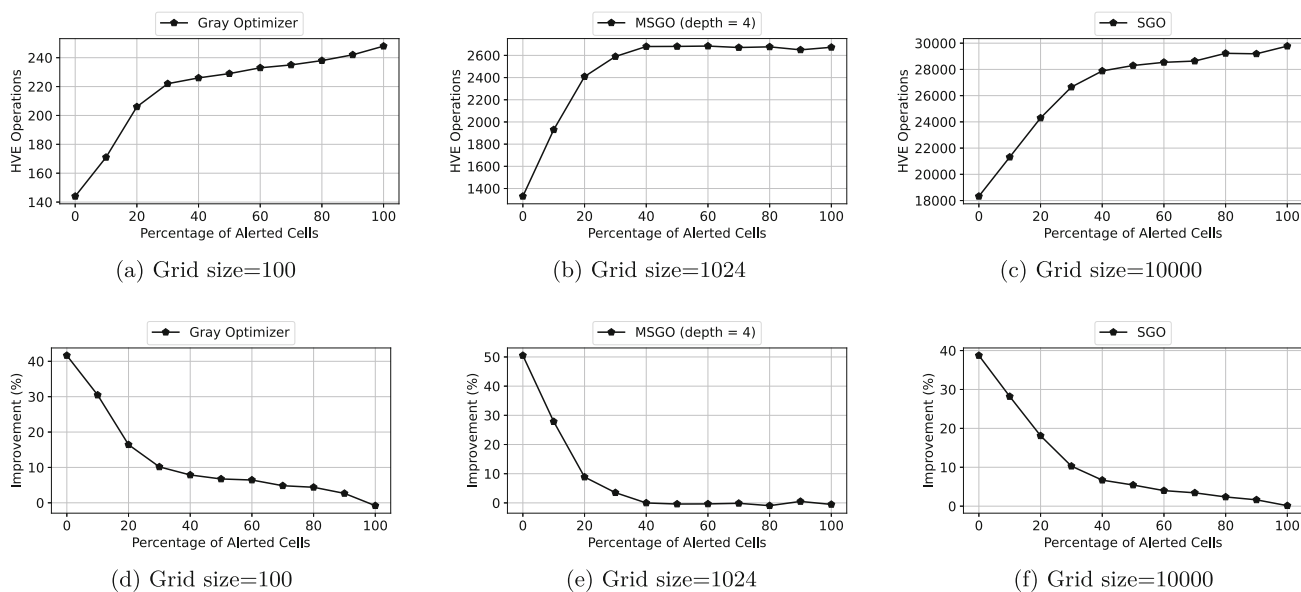


Fig. 12 Sensitivity analysis of algorithms, 40% alerted cells

The overall trend of reduction in the performance improvement by the addition of noise is consistent across all three algorithms. The improvement gained from algorithms stays at its highest when there exists no amount of noise at the input. The figure gradually drops as more noise is applied between zero to 50%, after which the performance improvement becomes almost negligible.

As expected, in the case of maximum noise, no information is available regarding probabilities, and therefore, no further gain could be made with respect to HGE. Hence, at 100% of noise, the number of HVE operations required from all algorithms converges to the HVE operations of HGE. The rate of sensitivity to imperfect information varies among algorithms. Looking at 10% of noise, it can be seen that the drop in MSGO performance occurs at a higher rate than the other two algorithms with GO and SGO indicating 25% loss in the performance against the loss of 40% for the MSGO algorithm. Overall, MSGO shows a higher sensitivity compared with the GO and SGO algorithms.

6.5 Dynamic alert zones

So far, we evaluated techniques for static alert zones. Next, we measure the performance of our proposed technique for dynamic alert zones introduced in Sect. 5.

Figure 13 investigates the performance gain acquired by applying the proposed Markov model. The random path approach (Monte Carlo sampling) is used as the underlying method to compute the transition matrix’s stationary distribution, minimizing the induced computation complexity on the system. The *x*-axis of the graphs shows the percentage of alert cells, and the *y*-axis represents the percentage of improve-

ment as well as the number of HVE operations required. To distinguish between the two modeling approaches, performance improvement achieved by incorporating the Markov model is labeled as *dynamic*, and the scenario in which the time dependence is not considered is referred to as *static*. The experiment is designed by initializing both static and dynamic approaches with the same set of initial probabilities; however, the system would continue evolving in a uniformly distributed manner. Therefore, if there are *m* outgoing edges from a state of the model, the corresponding probability is set to $1/m$. The aim is to see if the Markov model is able to capture the evolution of the system and how much improvement can be achieved with the gained information. As before, the value of the *a* and *b* are set to 0.75 and 10 with the termination probability of 0.4.

Figure 13 shows that the dynamic method can predict well the evolution of alert zones, as the resulting encoding requires far fewer HVE operations. The performance gain achieved for all three of the algorithms is significant. The percentage of improvement is approximately 35% to 50%, indicating more impact on GO compared to MSGO and SGO.

7 Related work

7.1 Location privacy

Preserving the privacy of users in communication networks and online platforms has been one of the most challenging research problems in the past two decades. In the widely accepted scenario, users provide their location to service providers in exchange for location-based services they offer.

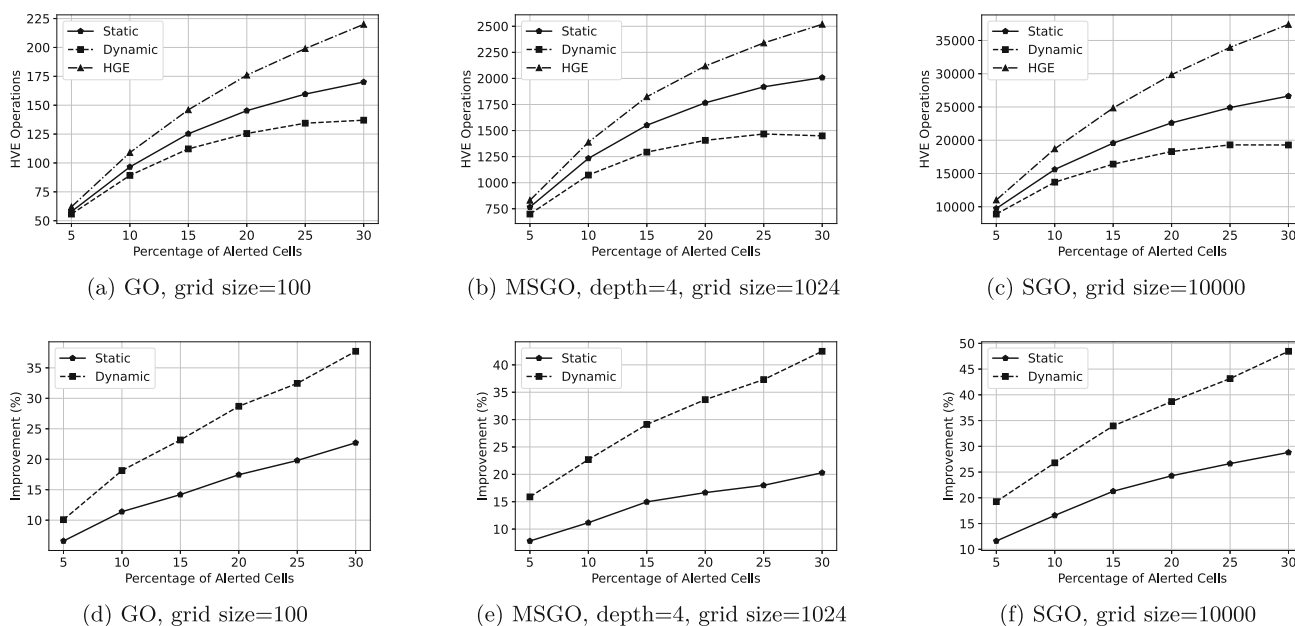


Fig. 13 Markov model versus static approach

The goal is to provide the service without user privacy being compromised by any of the parties involved. Early works to tackle this problem were focused on hiding or obfuscating user locations to achieve a privacy metric termed as k -anonymity. The location of a user is said to be k -anonymous if it is not distinguishable from at least $k - 1$ other queried locations [22].

In [15], the authors aim to provide k -anonymity by hiding the location of user among $k - 1$ fake locations and requesting for desired services for all k locations at the same time. The generation of such dummy locations based on a virtual grid or circle was considered in [20]. The authors in [19] conducted the selection of dummy locations predicated on the number of queries made on the map and aimed at increasing the entropy of k locations in each set. In [8], random regions that enclose the user locations were introduced to bring uncertainty in the authentication of user locations. Unfortunately, fake locations can be revealed particularly in trajectories and with the existence of prior knowledge about the map and users.

Later on, approaches based on *Cloaking Regions (CRs)* proposed by [14] gained momentum in the literature. The principal idea behind this method is to use a trusted anonymizer that clusters k real user locations and query the area they are enclosed by to retrieve points of interest. Doing so, CRs aim to achieve k -anonymity for users and preserve their privacy. This approach is partially effective when snapshots of trajectories are considered, but once users are seen in trajectories, their location privacy would be severely at risk [24]. Even for individual snapshots, it must be noted that a coarse area of real locations is released to the ser-

vice provider, which could threaten the location privacy of users. Moreover, the CR-based approaches are susceptible to inference attacks predicated on the background knowledge or so-called side information. One such side information is the knowledge about the number of queries made on different locations of the map [19].

More recently, a model for privacy preservation in statistical databases named *differential privacy (DP)* was developed in [11]. DP only supports aggregate queries, thus it is not suitable for alert systems, such as ours, since it forbids any disclosure that leaks the presence of an individual record in the data. In the case of location alerts, we need to assess spatial predicates and notify individual users, which goes against the founding principle of DP. Therefore, protecting location attributes with searchable encryption is the preferred solution.

Closer to HVE approach, a private information protocol was proposed in [12]. The PIR technique is based on cryptography and shown to be secure for private retrieval of information. Despite the promising results, there exists an assumption behind PIR approach that the user already knows about the points of interest. Therefore, PIR is not suitable for location-based alert systems as users are not aware of alert zone whereabouts.

7.2 Searchable encryption

Originated from works such as [26], the concept of search encryption was proposed to provide a secure cryptographic search of keywords. Initially, only the exact matches of keywords were supported and later on the approach was extended

Table 2 Summary of the proposed algorithms

Algorithms	Possible granularity	Complexity	Recommended depth	Number of seeds	Application
Gray Optimizer	Low	$\mathcal{O}(2n(\log_2 n) + n^{\log_2 3} - 2n + 1)$	$\lceil \log_2 n \rceil$	one	The core algorithm used in MSGO and SGO, which is suitable for low granularity grids due to high complexity
MSGO	Medium to high	The minimum of $\mathcal{O}(n(\log_2 n))$ depending on the depth variable	Depth=4 when the clusters are of equal length, otherwise it depends on the application	Equal to number of clusters	Advantageous in scenarios where the optimization is required for particular subsets of the grid. For example, the service is needed for two major non-overlapping organizations
SGO	High	$\mathcal{O}(n(\log_2 n))$	One	n	SGO requires the lowest complexity, and it is advantageous in scenarios where the aim is to optimize the number of HVE operations for all the cells with equal significance

for comparison queries in [4], and to subset queries and conjunctions of equality in [5]. The authors in [5] also proposed the concept of HVE, used as the underlying tool to provide a secure location-based alert system. This approach and its extension in [3] preserves the privacy of encrypted messages and tokens with the overhead of high computational complexity. The authors in [13] introduced and adopted the HVE for location-based alert systems, conducting the predicate match at a trusted provider, preserving the privacy of encrypted messages as well as tokens. Despite the promising results of the approach for privacy preservation in location-based alert systems, further reduction of computational overhead is necessary to increase the practicality.

8 Conclusion

We proposed a family of techniques to reduce the computational overhead of HVE predicate evaluation in location-based alert systems. Specifically, we used graph embeddings to find advantageous domain space encodings that help reduce the required number of expensive HVE operations. Our heuristic solutions provide a significant improvement in

computation compared to existing work, and they can scale to domain partitionings of fine granularity. In addition, we studied how to extend these techniques to work for the challenging setting of dynamic alert zones. Table 2 summarizes the properties of the proposed approaches.

In future work, we will focus on deriving cost models and strategies to reduce the HVE overhead based on workload-specific requirements. Certain families of tasks may exhibit specific patterns of operations, which can be taken into account to optimize HVE matching performance, as well as to re-use computation. We will also investigate extending the graph embedding approach to other types of searchable encryption, beyond HVE (e.g., Inner Product Evaluation), which exhibit different types of internal algebraic operations.

Acknowledgements This research has been funded in part by NSF grants IIS-1,910,950, IIS-1,909,806 and CNS-2,125,530, the USC Integrated Media Systems Center (IMSC), and unrestricted cash gifts from Microsoft and Google. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of any of the sponsors such as the National Science Foundation.

Funding Open Access funding provided by the Qatar National Library.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Avrachenkov, K., Litvak, N., Nemirovsky, D., Osipova, N.: Monte carlo methods in pagerank computation: When one iteration is sufficient. *SIAM J. Numer. Anal.* **45**(2), 890–904 (2007)
- Blitzstein, J.K., Hwang, J.: *Introduction to Probability*. Crc Press, Florida (2019)
- Blundo, C., Iovino, V., Persiano, G.: Private-key hidden vector encryption with key confidentiality. In: *International Conference on Cryptology and Network Security*, pp. 259–277. Springer (2009)
- Boneh, D., Sahai, A., Waters, B.: Fully collusion resistant traitor tracing with short ciphertexts and private keys. In: *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 573–592. Springer (2006)
- Boneh, D., Waters, B.: Conjunctive, subset, and range queries on encrypted data. In: *Theory of Cryptography Conference*, pp. 535–554. Springer (2007)
- Cavers, M., Vasudevan, K.: Spatio-temporal complex markov chain (scmc) model using directed graphs: earthquake sequencing. *Pure Appl. Geophys.* **172**, 225–241 (2015)
- Chandrasekharam, R., Vinod, V., Subramanian, S.: Genetic algorithm for embedding a complete graph in a hypercube with a vlsi application. *Microprocess. Microprogr.* **40**(8), 537–552 (1994)
- Cheng, R., Zhang, Y., Bertino, E., Prabhakar, S.: Preserving user location privacy in mobile data management infrastructures. In: *International Workshop on Privacy Enhancing Technologies*, pp. 393–412. Springer (2006)
- Curtmola, R., Garay, J., Kamara, S., Ostrovsky, R.: Searchable symmetric encryption: improved definitions and efficient constructions. *J. Comput. Secur.* **19**(5), 895–934 (2011)
- Drake, C.: Two-level logic minimization (2012). <https://pyeda.readthedocs.io/en/latest/2llm.html>
- Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: *Theory of cryptography conference*, pp. 265–284. Springer (2006)
- Ghinita, G., Kalnis, P., Khoshgozaran, A., Shahabi, C., Tan, K.L.: Private queries in location based services: anonymizers are not necessary. In: *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pp. 121–132 (2008)
- Ghinita, G., Rughinis, R.: An efficient privacy-preserving system for monitoring mobile users: making searchable encryption practical. In: *Proceedings of the 4th ACM conference on Data and application security and privacy*, pp. 321–332. ACM (2014)
- Gruteser, M., Grunwald, D.: Anonymous usage of location-based services through spatial and temporal cloaking. In: *Proceedings of the 1st international conference on Mobile systems, applications and services*, pp. 31–42 (2003)
- Kido, H., Yanagisawa, Y., Satoh, T.: An anonymous communication technique using dummies for location-based services. In: *ICPS'05. Proceedings. International Conference on Pervasive Services, 2005.*, pp. 88–97. IEEE (2005)
- Lai, S., Patranabis, S., Sakzad, A., Liu, J.K., Mukhopadhyay, D., Steinfeld, R., Sun, S.F., Liu, D., Zuo, C.: Result pattern hiding searchable encryption for conjunctive queries. In: *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pp. 745–762 (2018)
- Leiserson, C.E., Rivest, R.L., Cormen, T.H., Stein, C.: *Introduction to Algorithms*, vol. 6. MIT press, Cambridge, MA (2001)
- Nguyen, K., Ghinita, G., Naveed, M., Shahabi, C.: A privacy-preserving, accountable and spam-resilient geo-marketplace. In: *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 299–308. ACM (2019)
- Niu, B., Li, Q., Zhu, X., Cao, G., Li, H.: Achieving k-anonymity in privacy-aware location-based services. In: *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pp. 754–762. IEEE (2014)
- Niu, B., Zhang, Z., Li, X., Li, H.: Privacy-area aware dummy generation algorithms for location-based services. In: *2014 IEEE International Conference on Communications (ICC)*, pp. 957–962. IEEE (2014)
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: *Numerical recipes in c* (1988)
- Samarati, P.: Protecting respondents identities in microdata release. *IEEE Trans. Knowl. Data Eng.* **13**(6), 1010–1027 (2001)
- Sarma, A.D., Molla, A.R., Pandurangan, G., Upfal, E.: Fast distributed pagerank computation. In: *International Conference on Distributed Computing and Networking*, pp. 11–26. Springer (2013)
- Shaham, S., Ding, M., Liu, B., Dang, S., Lin, Z., Li, J.: Privacy preserving location data publishing: a machine learning approach. *IEEE Trans. Knowl. Data Eng.* **33**(9), 3270–83 (2020)
- Shaham, S., Ghinita, G., Shahabi, C.: Enhancing the performance of spatial queries on encrypted data through graph embedding. In: *Data and Applications Security and Privacy DBSec*, pp. 289–309 (2020)
- Song, D.X., Wagner, D., Perrig, A.: Practical techniques for searches on encrypted data. In: *2000 IEEE Symposium on Security and Privacy*, pp. 44–55. IEEE (2000)
- Wang, J., Zhang, R., Yan, Y., Dong, X., Li, J.M.: Locating hazardous gas leaks in the atmosphere via modified genetic, mcmc and particle swarm optimization algorithms. *Atmos. Environ.* **157**, 27–37 (2017)
- Yaesoubi, R., Cohen, T.: Generalized markov models of infectious disease spread: a novel framework for developing dynamic health policies. *Eur. J. Op. Res.* **215**(3), 679–687 (2011)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.