DENIABLE GROUP TESTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Group Testing is a well-established area of searching and learning under uncertainty. One of its major applications is testing individuals for hidden defects, e.g., diseases or other sensitive characteristics. Surprisingly to these applications, very little has been done to assure privacy within the testing process and output. We introduce a new privacy concept, called δ-deniability, which guarantees the following with respect to the outputted set: (i) each of the defective individuals is in the outputted set; however, (ii) for each outputted individual v there are at least δ possible defective sets not containing v that give the same testing results. In other words, unless an observer could test negatively any δ pointed out individuals, it cannot be certain about defectiveness of any particular outputted individual. We show that any monotonic testing strategy can output a set at least $\log(\delta + 1) + 1$ times bigger than the number of defective elements, for some defective sets under binary testing feedback. For ℓ-capped quantitative feedback, $\ell \leq \delta$, the size of some output sets could be at least $\frac{\delta+1}{\ell}$ times bigger than the input set. On the positive side, we design a few testing strategies and prove that they guarantee δ-deniability and smooth trade-off between the number of tests and the overhead of the size of the output set over the number of defective individuals, called accuracy.

1 INTRODUCTION

Group testing (GT) Dorfman (1943) is a classic combinatorial problem commonly used in learning methods where, given a set of N items, the goal is to identify a subset $S \subset N$ of defective elements. This is achieved by performing a sequence of tests, which can be viewed as asking queries regarding specific subsets of N and obtaining feedback on their intersections with S. This simple concept, along with its numerous modifications, is surprisingly widely used in algorithm design and formal analysis within algorithmic learning theory. It also forms the foundation of many specific methods in the broader field of artificial intelligence and information theory.

However, in certain applications of group testing, knowing whether a specific element belongs to S may raise concerns. A primary example can be found in Dorfman's original 1943 work on disease screening. While learning an infected individual is crucial for medical treatment and epidemic control, the public disclosure of such information can be distressing for the infected individual.

Motivated by these observations, this paper introduces *Deniable Group Testing*. In this framework, an individual can plausibly claim that the obtained result is untrue; in other words, they can effectively deny it. While this approach introduces an element of uncertainty regarding individual results, the global outcomes remains useful in many aspects. This concept mirrors the well-known Randomized Response Technique by Warner (1965), where respondents provide untruthful answers with a certain probability to protect their privacy. Our proposed Deniable Group Testing differs in nature: it is based on a **deterministic** approach, and the perturbation of the result occurs at the test construction stage rather than through simple input randomisation.

The goal is to ensure δ-deniability while simultaneously obtaining a minimum number of *possibly defective* elements relative to the actual number of defective elements d.

1.1 OUR CONTRIBUTION

Model. We formalize deniable group testing framework (Section 3). A testing algorithm is δ-*deniable* if, for every element x and every feedback vector that some admissible set containing x

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

could produce, there are at least δ admissible sets not containing x that produce the same feedback. Its *accuracy* Δ bounds the size of the union of all sets consistent with the feedback relative to the number of defective elements, in the worst case or on average. The guarantee is deterministic and holds for the family of capped quantitative feedback functions $\mathcal{F}_\ell$, which includes binary feedback ($\ell = 1$). *In learning-theoretic terms, the sets consistent with the feedback form the version space of the observations: Δ measures how much of the domain it leaves undecided, and deniability requires that every possibly defective element is excluded by at least δ consistent hypotheses.*

Grouping algorithm. For binary feedback, we give a non-adaptive deterministic algorithm that applies a D-disjunct design to groups of $\delta + 1$ elements (Theorem 1). It is δ-deniable, has worst-case accuracy $\Delta = \delta + 1$, which is attained, and uses $t \leq \min\{\frac{n}{\delta+1}, 3(D+1)^2 \ln \frac{n}{\delta+1}\}$ tests; thus a larger δ trades accuracy for fewer tests. We also give an adaptive variant using at most $D(\lceil \log_2 \frac{n}{\delta+1} \rceil + 1) + 1$ tests, and a δ-deniable extension to $\mathcal{F}_\ell$ with $\ell \geq 2$ and $\Delta \leq D + \delta$.

SUBMARTINGALE algorithm. For $\mathcal{F}_\ell$, we propose an adaptive deterministic algorithm that refines a partition of the domain and tests a new part only if it is guaranteed to contain at least δ non-defective elements. We prove that it is δ-deniable on every input (Lemma 2) and evaluate its accuracy experimentally over a range of d, D, δ, and ℓ (Section 5.2, Appendix C).

We also prove a **lower bound on accuracy** of any monotonic (δ, Δ)-DGT in Theorem 3 in Section 6. Specifically, we show that any such algorithm satisfies $\Delta \geq \log(\delta + 1) + 1$ for $\ell = 1$, and accuracy $\Delta \geq \frac{\delta+1}{\ell}$ for $2 \leq \ell \leq \delta$. Monotonic algorithms is a wide class of solutions that produce monotonic feedback vectors with respect to the inclusion order of defective sets (which are inputs to the algorithm) and the position-wise order of feedback vectors (which are outputs of the algorithm); i.e., the feedback vector of a superset is bigger or equal than the feedback vector of a subset.

2 RELATED WORK

Group testing is generally traced back to the seminal paper by Dorfman (1943) and has since been studied extensively in both adaptive and non-adaptive settings; overviews can be found in Du & Hwang (2000) and, from an information-theoretic perspective, in Aldridge et al. (2019). In the adaptive setting, classical results include the generalized binary-splitting method of Hwang (1972) and the order-optimal two-stage scheme of De Bonis et al. (2005), while in zero-error non-adaptive group testing, superimposed codes and disjunct matrices provide a central combinatorial framework, with foundational constructions and bounds due to Kautz & Singleton (1964) and D'yachkov & Rykov (1982), and explicit constructions due to Porat & Rothschild (2011). Recent work studied the impact of different threshold functions on the number of tests Klonowski et al. (2022); Kowalski & Pajak (2022). Group testing (GT) has also served as a subroutine or inspiration in machine learning and related algorithmic problems. Zhou et al. (2014) use it to design a parallel feature selection algorithm, Engels et al. (2021) reduce approximate near-neighbor search to a GT problem, and Erginbas et al. (2026) use adaptive group testing to learn sparse real-valued polynomials over Boolean inputs. Ciampiconi et al. (2020) develop a MaxSAT-based framework for decoding in non-adaptive GT that also handles noisy test outcomes, and Silva & Singh (2021) study how to form testing groups using a transmission network, show that the resulting group-assembly problem is computationally hard, and propose efficient heuristics.

Although the proposed concept of group testing with deniability is, to the best of our knowledge, entirely new, there is a broad body of related literature. Our results can be viewed as a form of group testing in which there is an induced inaccuracy of the results or a deliberate limitation of released knowledge. This general framework also encompasses several recent works that address group testing designed to ensure privacy. In the paper Gonen et al. (2025), GT models are addressed strictly from the perspective of protecting the privacy of users, formally defined as *differential privacy* Dwork (2006). The algorithms presented there are fundamentally different in nature from the solutions presented in our work; they rely on simple randomization at various levels (although the analysis itself is highly non-trivial).

Our approach is, in some sense, analogous to the concept of *Positive-Unlabeled* (PU) Data, which has been intensively studied recently in statistics and machine learning theory. In this problem, we consider positive and negative elements where only a subset of positive elements is labeled, while all negative elements remain unlabeled (see e.g., Liu et al. (2002); Mielniczuk & Wawrzeńczyk (2025)).

In the motivation for our work, one can find certain analogies to the **Randomized Response Technique** Warner (1965), where survey respondents flip their Yes/No answers with a fixed probability. This guarantees the protection of **individual privacy**, while the aggregated data can still provide significant statistical insights. In terms of approach, our approach is similar to **k-anonymity** Sweeney (2002), one of the methods preceding the now-standard *differential privacy*. In this model, individual privacy protection relies on the fact that each record is indistinguishable from at least $k-1$ other individuals with the same attributes. Despite its practical weaknesses, this informal approach has an advantage over differential privacy in that it can be implemented **deterministically**; as a result, it is still used in numerous applications. This scenario is similar to our situation, where an element identified by a test as *possibly* defective may, in fact, be non-defective. In other words, classification errors may exist, but they are asymmetric—occurring only in one direction. A review of recent results in this field can be found in Bekker & Davis (2020).

Potential applications of our method may relate, in spirit, to scenarios such as *deniable digital signatures*Chaum & Van Antwerpen (1989); Klonowski et al. (2008), where the signer can, under certain circumstances, disavow the authorship of a signature to a third party.

3 Formal Model

Let N be the set of n elements, also called a ***domain***. Let $\mathcal{S}$ be a family of ***admissible*** defective subsets of N. That is, $\mathcal{S}$ is a family containing all sets of ***defective*** elements $S \subset N$, which may occur in the considered testing environment. In this work we focus on $\mathcal{S}$ being a family of all subsets of N containing at most D elements, where D is a given model/problem parameter. We denote this family as $\mathcal{S}_D$.

Let $\mathcal{F} : 2^N \to \mathbb{N}$ be the ***test feedback function***, where 2^N denotes a family of all subsets of N and $\mathbb{N}$ is the set of all natural numbers. In group testing, the test feedback function is a symmetric function, i.e., depending on the size of the argument (an integer between 0 and n). In this work, we consider the entire family of such test functions parameterized by a positive integer $\ell \leq D$: for any $X \subseteq N$, $\mathcal{F}_\ell(X) = \min\{|X|, \ell\}$. They are also called ***ℓ-capped quantitative feedback*** functions. Naturally, this family includes all of the most popular feedback functions used in the group testing literature, including quantitative feedback functions ($\ell = n$) and the standard binary feedback, also called a beeping model ($\ell = 1$).

A ***test***, also called a ***query***, is modeled as a subset of the domain N. Assume we are given a sequence of queries $\mathcal{Q} = \langle Q_1, Q_2, \ldots, Q_t \rangle$. Let us define a ***feedback vector*** for an admissible set $S \in \mathcal{S}$:

$$F(S) = \langle \mathcal{F}(Q_1 \cap S), \mathcal{F}(Q_2 \cap S), \ldots, \mathcal{F}(Q_t \cap S) \rangle \ .$$

In our work, we consider both non-adaptive and adaptive algorithms; in the former, the sequence of queries is fixed in advance (i.e., it does not depend on the defective set S on which it is executed), while in the latter, query Q_i may depend on the feedback $\mathcal{F}(Q_1 \cap S), \mathcal{F}(Q_2 \cap S), \ldots, \mathcal{F}(Q_{i-1} \cap S)$ obtained for the earlier queries $Q_1, \ldots, Q_{i-1}$ applied for the defective set S. In particular, the length of the feedback vector may vary for different admissible sets. In this work, we restrict our attention to deterministic algorithms, which means that the feedback vector for the same admissible set S is always the same. Let $A(S) = \{S' \in \mathcal{S} : F(S') = F(S)\}$, which denotes the family of all admissible sets that yield the same feedback as S. A family $\{A(S)\}_{S \in \mathcal{S}}$ contains classes of abstraction in $\mathcal{S}$ defined by the relation of having the same feedback vector.

We call a vector f ***admissible (feedback vector)*** if $f = F(S)$ for some admissible set S; it follows that each admissible feedback vector corresponds 1-1 to some class of abstraction in $\{A(S)\}_{S \in \mathcal{S}}$. We use the classic position-wise (partial) order on vectors: for two vectors f, g, we denote $f \leq g$ if: the length $|f|$ of vector f is not bigger than the length $|g|$ of vector g, and for any position $i \leq |f|$, $f[i] \leq g[i]$. By extension, we write $f < g$ if $f \leq g$ and the alternative holds: $|f| < |g|$ or for some position i, $f[i] < g[i]$. We also say that a feedback vector f is ***x-compatible***, for a given element $x \in N$, if there exists $S' \in \mathcal{S}$ such that $x \in S'$ and $f = F(S')$. In other words, the feedback f can be the result of a certain admissible set containing x.

We say that $\mathcal{Q}_t = \{Q_1, \ldots Q_t\}$ is a solution to (possibly adaptive) **Deniable Group Testing** with deniability δ and accuracy Δ, or to (δ, Δ)-DGT for short, if:

Deniability For each $x \in N$ and for any x-compatible feedback vector f, there exist δ pairwise different admissible sets $S_1, \ldots, S_\delta$ such that $F(S_1) = F(S_2) = \ldots = F(S_\delta) = f$ and $x \notin \bigcup_{i=1}^{\delta} S_i$.

Accuracy: we consider two forms of measuring accuracy:

Worst-case accuracy: for each $S \in \mathcal{S}$, we have $\left|\bigcup_{S' \in A(S)} S'\right| \leq \Delta \cdot |S|$.

Average-case accuracy: $\mathbb{E}\left[\frac{|\bigcup_{S' \in A(S)} S'|}{|S|}\right] \leq \Delta$, where randomness is taken uniformly over $\mathcal{S}$.

The motivation behind these definitions is as follows. We assume that an external observer, who knows the algorithm and feedback function, receives the test results (i.e., admissible feedback vector f) without knowing the exact set of defective elements S that generated f (i.e., $f = F(S)$ for some S unknown to the observer). Based on f and the knowledge of admissible sets, the observer considers whether a specific element x belongs to the tested set of defective elements S. If f is *not* x-compatible, then the observer clearly infers $x \notin S$. In this case, x is called *non-defective*. Conversely, if f is x-compatible, then there exists at least one admissible set S containing x. In this scenario, the *deniability* property means that there exist at least δ other admissible sets that yield the exact same feedback f but do **not** contain x. Consequently, one can plausibly deny that x is defective. From the observer's viewpoint, x holds the status *possibly defective*, only.

On the other hand, the *accuracy* property guarantees that the number of possibly defective elements is upper bounded by a factor of Δ relative to the true number of truly defective elements. In practice, depending on the scenario, this implies that the number of elements (e.g., patients/devices/processes) requiring treatment is at most Δ times larger than the actual number of defective elements.

The goal of algorithm design will therefore be to minimize the number of *possible defectives* while simultaneously ensuring δ-deniability for all $x \in N$.

4 Grouping Algorithm

We first present a simple non-adaptive algorithm, GROUPING, for the binary feedback function $\mathcal{F}_1$ and the admissible family $\mathcal{S}_D$. It reduces deniable group testing to classical zero-error group testing on groups of $\delta + 1$ elements.

Algorithm. Let $b = \delta + 1$ and assume that $b \mid n$. Partition N into $g = \frac{n}{b}$ groups $G_1, \ldots, G_g$ of size b each, fixed in advance and known to the observer. For $S \subseteq N$, let $K(S) = \{j : G_j \cap S \neq \emptyset\}$ be the set of groups hit by S; clearly $|K(S)| \leq |S|$. A $t \times g$ binary matrix M is *D-disjunct* if for every $K \subseteq \{1, \ldots, g\}$ with $|K| \leq D$ and every $j \notin K$ there is a row i with $M_{ij} = 1$ and $M_{ij'} = 0$ for all $j' \in K$ Du & Hwang (2000). Given a D-disjunct matrix M, row i defines query $Q_i = \bigcup_{j:\, M_{ij}=1} G_j$. In the blood-testing application of Dorfman (1943), a group can be represented by the pooled sample of its members. After observing the feedback vector f, the algorithm declares G_j *positive* if $f[i] = 1$ for every i with $M_{ij} = 1$ (the naive decoder of Cheraghchi & Nakos (2020)), marks all elements of positive groups as *possibly defective*, and marks all other elements as *non-defective*.

Number of tests. If $g \leq D$, the $g \times g$ identity matrix is D-disjunct, so $t = g$ tests suffice. If $g \geq D + 1$, there exists a D-disjunct matrix with

$$t \;\leq\; 3(D+1)\ln\left[(D+1)\binom{g}{D+1}\right]$$

rows (Du & Hwang, 2000, Theorem 8.1.3). Since $(D+1)\binom{g}{D+1} \leq g^{D+1}$, this implies $t \leq 3(D+1)^2 \ln g$, which is the bound stated for $g \gg D$ in Du & Hwang (2000, Corollary 8.1.4). These matrices are obtained by a probabilistic argument; explicit D-disjunct matrices with $O(D^2 \log g)$ rows are also known (Porat & Rothschild, 2011), and for $D \geq 2$, $O\big(D^2 \min\{\log g, (\log_D g)^2\}\big)$ rows suffice (Cheraghchi & Nakos, 2020, Theorem 2.9), building on Kautz & Singleton (1964) and Porat & Rothschild (2011). Hence, with $g = n/(\delta+1) \geq 2$,

$$t \;\leq\; \min\left\{\frac{n}{\delta+1},\; 3(D+1)^2 \ln\frac{n}{\delta+1}\right\}.$$

Thus, a larger δ reduces the effective number of items from n to $n/(\delta+1)$, and with it the bound on the number of tests, at the price of accuracy $\Delta = \delta + 1$ (Theorem 1).

Theorem 1. *Let $\ell = 1$, $D, \delta \geq 1$, and $(\delta+1) \mid n$. Algorithm* GROUPING *is a non-adaptive solution to $(\delta, \delta+1)$-DGT with worst-case accuracy for the admissible family $\mathcal{S}_D$. For every $S \in \mathcal{S}_D$, the set of elements marked* possibly defective *equals $\bigcup_{S' \in A(S)} S'$, and the accuracy bound $\delta+1$ is attained.*

Proof. Step 1: the feedback depends only on $K(S)$. For every i, $\mathcal{F}_1(Q_i \cap S) = 1$ if and only if $M_{ij} = 1$ for some $j \in K(S)$. Hence $F(S)$ is determined by $K(S)$.

Step 2: distinct $K(S)$ give distinct feedback. Let $S, S' \in \mathcal{S}_D$ with $K(S) \neq K(S')$, say $j \in K(S) \setminus K(S')$. Since $|K(S')| \leq |S'| \leq D$, D-disjunctness gives a row i with $M_{ij} = 1$ and $M_{ij'} = 0$ for all $j' \in K(S')$, so $\mathcal{F}_1(Q_i \cap S) = 1 \neq 0 = \mathcal{F}_1(Q_i \cap S')$. Together with Step 1, $A(S) = \{S' \in \mathcal{S}_D : K(S') = K(S)\}$. The same argument applied to $K = K(S)$ shows that the decoding rule declares positive exactly the groups G_j with $j \in K(S)$.

Step 3: deniability. Fix $x \in N$, let G_{j_x} be the group containing x, and let f be x-compatible, i.e., $f = F(S)$ for some $S \in \mathcal{S}_D$ with $x \in S$. Let $K = K(S)$; then $j_x \in K$ and $|K| \leq D$. For each $y \in G_{j_x} \setminus \{x\}$, let $S_y = \{y\} \cup \{r_j : j \in K \setminus \{j_x\}\}$. Then $|S_y| = |K| \leq D$, so $S_y \in \mathcal{S}_D$, and $K(S_y) = K$, so $F(S_y) = f$ by Step 1. Moreover, $x \notin S_y$, and the $b - 1 = \delta$ sets S_y are pairwise different.

Step 4: accuracy. By Step 2, every $S' \in A(S)$ satisfies $K(S') = K(S)$ and is thus contained in $U = \bigcup_{j \in K(S)} G_j$. Conversely, for $j \in K(S)$ and $z \in G_j$, the set $\{z\} \cup \{r_{j'} : j' \in K(S) \setminus \{j\}\}$ has $|K(S)| \leq D$ elements and hits exactly the groups in $K(S)$, so it belongs to $A(S)$ by Step 1. Hence $\bigcup_{S' \in A(S)} S' = U$, which by Step 2 is exactly the set marked *possibly defective*; in particular, $x \in U$ if and only if $x \in S'$ for some S' with $F(S') = F(S)$, i.e., if and only if $F(S)$ is x-compatible. Finally, $|U| = b\,|K(S)| \leq (\delta+1)|S|$, with equality whenever the elements of S lie in distinct groups, e.g., for $|S| = 1$. For $D = 1$ and $f = F(\{x\})$, the sets with feedback f are exactly the singletons of G_{j_x}, so exactly δ of them avoid x. □

Accuracy for a given number of defectives (worst case). Fix $S \in \mathcal{S}_D$ with $|S| = d \geq 1$, and let $k = |K(S)|$ be the number of positive groups. Since each group contains b elements, $\lceil d/b \rceil \leq k \leq \min\{g, d\}$. By Step 4 of the proof of Theorem 1, $\big|\bigcup_{S' \in A(S)} S'\big| = k\,b$, and hence

$$\frac{\big|\bigcup_{S' \in A(S)} S'\big|}{|S|} = \frac{k\,b}{d} \;\leq\; \frac{b}{d}\,\min\{g, d\},$$

with equality for every S of size d that hits $\min\{g, d\}$ groups. The right-hand side equals $\delta + 1$ if $d \leq g$, and equals n/d, which is less than $\delta + 1$, if $d > g$. The quantity $k\,b - d = |U \setminus S|$ is the number of non-defective elements marked *possibly defective*; it should not be confused with the deniability parameter, which counts admissible sets (Section 3) and is at least $\delta = b - 1$ for every x and every x-compatible f (Theorem 1).

Accuracy for a given number of defectives (average case). Let S be a uniformly random d-subset of N with $1 \leq d \leq D$. A fixed group contains no element of S with probability $p = \binom{n-b}{d} / \binom{n}{d}$ (with $\binom{a}{c} = 0$ for $c > a$). By linearity of expectation, $\mathbb{E}[k] = g(1-p)$, and hence

$$\mathbb{E}\left[\frac{\big|\bigcup_{S' \in A(S)} S'\big|}{|S|}\right] = \frac{b\,\mathbb{E}[k]}{d} = \frac{n}{d}\,(1-p) \;\leq\; \delta + 1.$$

Since $p = \prod_{i=0}^{d-1} \left(1 - \frac{b}{n-i}\right) \leq \left(1 - \frac{b}{n}\right)^d \leq e^{-bd/n}$ and $1 - e^{-u} \geq u - u^2/2$ for $u \geq 0$, this expectation is at least $(\delta+1)\big(1 - \frac{(\delta+1)d}{2n}\big)$; thus, for $(\delta+1)d \ll n$, the average-case accuracy is essentially $\delta + 1$. The expected number of non-defective elements marked *possibly defective* is $\mathbb{E}\big[|U \setminus S|\big] = n(1-p) - d$. This distribution fixes $|S| = d$ and thus differs from the uniform distribution over $\mathcal{S}_D$ used in Section 3; since the bound of Theorem 1 holds for every S, the average-case accuracy is at most $\delta + 1$ under any distribution over nonempty admissible sets.

Algorithm 1 SUBMARTINGALE Algorithm

Require: $D, \delta, l, N \in \{0,1\}^n$ with 1 on exactly d positions , $n > D + \delta$
Ensure: Ratio of elements marked as *possibly defective* to d

procedure SUBMARTINGALE(N, D, δ, l)
  $\mathcal{P} \leftarrow \{N\}$
  $\min_N \leftarrow 0$
  $\text{Max}_N \leftarrow D$
  while true **do**
    if $\exists S \in \mathcal{P}$ such that $k = \lfloor |S|/(\delta + \text{Max}_S) \rfloor > 1$ **then**
      for $j \leftarrow 1$ **to** $k-1$ **do**
        $S_j \leftarrow S[(\delta + \text{Max}_S) \cdot (j-1) + 1, (\delta + \text{Max}_S) \cdot j]$
      end for
      $S_k \leftarrow S[(\delta + \text{Max}_S) \cdot (k-1) + 1, |S|]$
      $\mathcal{P} \leftarrow \mathcal{P} \setminus \{S\}$
      for $j \leftarrow 1$ **to** k **do**
        PRUDENTTEST(S_j)
        $\mathcal{P} \leftarrow \mathcal{P} \cup \{S_j\}$
      end for
      UPDATEMAXES($\mathcal{P}$)
    else
      for each $S \in \mathcal{P}$ **do**
        Mark all $x \in S$ not marked as *non-defective* as *possibly defective*.
      end for
      break
    end if
  end while
end procedure

Remark. By Step 2, $A(S)$ is invariant under any permutation of elements within groups. Hence, given the feedback, all members of a positive group are indistinguishable: each of them belongs to some set in $A(S)$ and is excluded from another.

Extensions. *Adaptive version.* For the feedback function $\mathcal{F}_1$, the reduction works with any adaptive algorithm whose queries are unions of groups and which identifies $K(S)$ exactly for every $S \in \mathcal{S}_D$: by induction over the queries, the executed queries and the feedback depend only on $K(S)$, and exact identification makes this dependence injective, so Steps 2–4 carry over. For instance, repeatedly testing the union of the remaining groups and, if the answer is positive, locating one positive group by binary search uses at most $D(\lceil \log_2 g \rceil + 1) + 1$ tests and does not require knowing $|S|$. *Feedback $\mathcal{F}_\ell$ with $\ell \geq 2$.* Here groups of size $\delta + 1$ do not suffice: for $\ell = D = 2$, $\delta = 1$, the identity matrix, and $S = G_1$, the set S is the only admissible set with feedback $F(S)$. Instead, let $b = D + \delta$ and assume $b \mid n$. If f is x-compatible with witness $S \ni x$, then $|G_{j_x} \setminus S| \geq b - D = \delta$, and for each $y \in G_{j_x} \setminus S$ the set $(S \setminus \{x\}) \cup \{y\}$ has the same intersection size with every union of groups, hence the same feedback. Since $\mathcal{F}_\ell$ determines $\mathcal{F}_1$, every set in $A(S)$ is contained in $\bigcup_{j \in K(S)} G_j$, so the algorithm is δ-deniable with $\Delta \leq D + \delta$.

5 SUBMARTINGALE ALGORITHM

The adaptive SUBMARTINGALE algorithm ensures δ-deniability for any input in the bounded quantitative function model $\mathcal{F}_l$, assuming homogeneity of defective elements, while simultaneously achieving very favorable expected Δ values in settings with random input.

The main idea of the algorithm is as follows: a partition $\mathcal{P}$ is constructed from the set N, consisting of subsets tested with the function $\mathcal{F}_l$. Initially, $\mathcal{P}$ contains only the set N. In subsequent steps, the algorithm searches a set $S \in \mathcal{P}$ that is sufficiently large relative to the upper bound for maximum number of defective elements Max_S in S, such that after splitting S into $k > 1$ subsets $S_1, \ldots, S_k$, the revealed tests $\mathcal{F}_l(S_1), \ldots, \mathcal{F}_l(S_k)$ do not violate δ-deniability under any configuration of at

Algorithm 2 Sub-procedures

```
1: procedure PRUDENTTEST(S)
2:     (min_S, Max_S) ← F_l(S)
3:     if Max_S = 0 then
4:         Mark all x ∈ S as non-defective
5:     end if
6: end procedure

7: procedure UPDATEMAXES(P)
8:     for each S ∈ P do
9:         if Max_S ≥ l then
10:            Max_S ← D − Σ_{S'∈P, S'≠S} min_{S'}
11:        end if
12:    end for
13: end procedure
```

most Max_S defective elements. Concurrently, S is removed from the partition $\mathcal{P}$ and $S_1, \ldots, S_k$ are added. When no such set S can be found, the algorithm terminates.

During the execution of the algorithm, a variable $\min_{S'}$ is maintained for each $S' \in \mathcal{P}$, representing the minimum number of defective elements present in S'. Note that the maximal number of defective elements for a specific set S can be upper bounded by Max_S computed as $D - \sum_{S' \in \mathcal{P}, S' \neq S} \min_{S'}$, since the total sum of defective elements is fixed at D.

In the algorithm, the elements of a tested set S for which $\mathcal{F}_l(S) = 0$ are marked as *non-defective*. Those contained in untested sets $S' \in \mathcal{P}$ upon termination are assigned the status *possibly defective*.

Pseudocode. In the figure Algorithm 1 we present a pseudocode for Submartingale algorithm,[1] while the pseudocodes of the sub-procedures are given in the figure Algorithm 2. Let $S[i, j]$ denote the subvector of S containing elements from index i to j.

5.1 ANALYSIS

Lemma 2. *Let us assume the admissible family $\mathcal{S}_D$. Algorithm* SUBMARTINGALE *guarantees δ-deniability for any input.*

Proof. Assume that the observed feedback function $F(S) = (\mathcal{F}_l(T_1 \cap S), \ldots, \mathcal{F}_l(T_t \cap S))$ was generated for some non-empty set of defective elements $S \in \mathcal{S}_D$. If $S = \emptyset$ the Lemma is trivially true. It is equally easy to check that δ-deniability holds if $n < 2(D + \delta)$. In that case, the condition in line 6 of the algorithm is not satisfied, and no set is tested.

Assume therefore that $n \geq 2(D + \delta)$. Consider any $x \in N$. In the first case, suppose that $x \in S$. Let T be the last tested subset of N that contained x. From the construction of Algorithm (see line 6), it follows that $|T \setminus S| \geq \delta$. That is, every tested set containing at least one defective element must contain at least δ non-defective elements. Thus, let $y_1, \ldots, y_\delta \in T \setminus S$ be these non-defective elements. Consider the sets $S_i := (S \setminus \{x\}) \cup \{y_i\}$ for $1 \leq i \leq \delta$. Clearly $S_i \in \mathcal{S}_D$ for all i. Moreover $x \notin \bigcup_{1 \leq i \leq \delta} S_i$. Observe that each S_i yields the exact same feedback as S. This holds because the exact same sequence of subsets will be tested as in the case of S, and each tested subset will contain the same number of defective elements. Consequently, the feedback function produces the identical result.

In the case where $x \notin S$, choose an arbitrary element $x' \in S$. Let T be the smallest tested set such that $x' \in T$. It can be observed that T contains at least $\delta - 1$ elements $y_1, y_2, \ldots, y_{\delta-1} \notin S$ distinct from x. Analogously to the previous case, we construct $S_i := (S \setminus \{x'\}) \cup \{y_i\}$ for $1 \leq i \leq \delta - 1$. Together with the set S (which does not contain x), we obtain δ pairwise distinct sets, none of which contain x, that yield the exact same feedback. □

[1] The name stems from the fact that running the algorithm can be viewed as the trajectory of a random process in which the expected number of revealed zeros ("non-defective elements") will grow (more precisely, do not decrease) in subsequent loop iterations, since each step yields information that enables further testing.

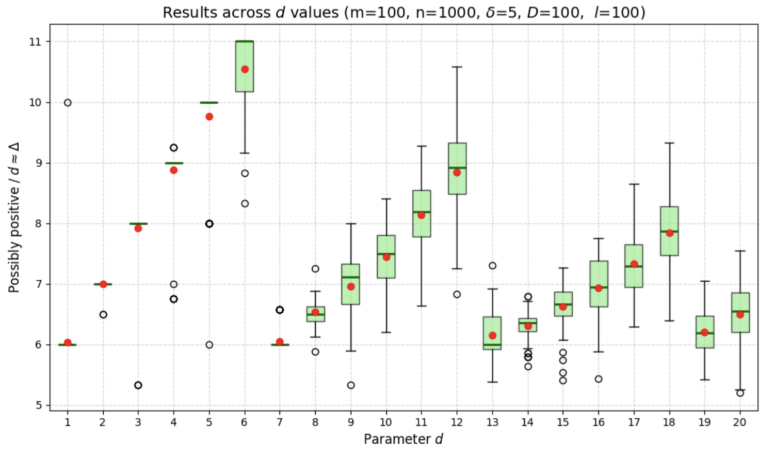


Figure 1: Distribution of Δ (number of elements marked *possibly defective* divided by d) for varying $d \leq D$, with $\delta = 5$, $D = 100$, $l = 100$, and $n = 1000$. Box plots over $m = 1000$ runs per value of d; red dots mark the means.

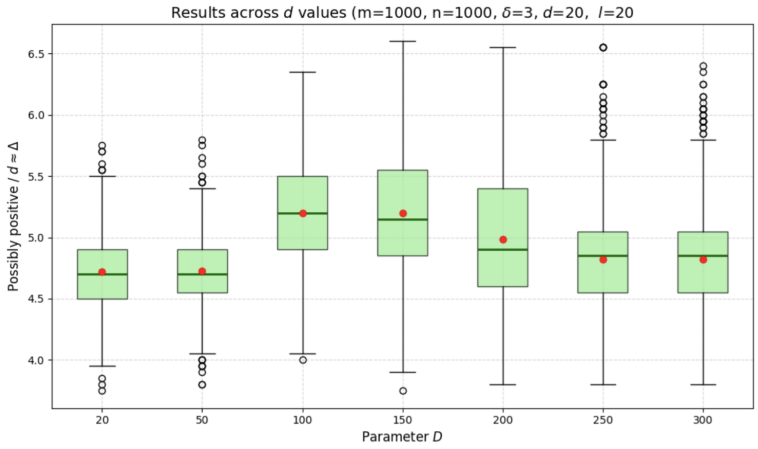


Figure 2: Distribution of Δ for varying $d \leq D$, with $\delta = 3$, $d = 20$, $l = 20$, and $n = 1000$. Box plots over $m = 1000$ runs per value of D; red dots mark the means.

5.2 Experimental results

Despite the relatively simple code, the analysis of the offered accuracy (parameter Δ, i.e, the ratio of the total number of possibly defective elements to defective elements) appears to be highly complex. The execution of the algorithm can be viewed as a random tree driven by the input and dependent on the parameters (d – number of defective elements, n – total number of elements, D – upper bound on the number of defective elements). Of course, the parameter d itself is not known to the SUBMARTINGALE algorithm, but it is used in experiments to perform experiments.

In this section, we calculate Δ by repeating the experiment $m = 1000$ (for fixed, various settings of parameters D, d, l) times and then present the distributions as box plots. We fix $n = 1000$.

Due to space limitations, in this section, we present only selected experiments, while comprehensive results are provided in Appendix C. The Python code used to conduct all experiments with the SUBMARTINGALE algorithm is available as supplementary material.

Conclusions from the Experiments. First, let us note that we need at least $\mathbb{E}(\Delta) \approx \delta + 1$ as long as $d \cdot \delta << n$, which corresponds to a situation where each defective element is "protected" by disjoint δ non-defective elements. However, these values may be significantly smaller or larger depending on various parameters.

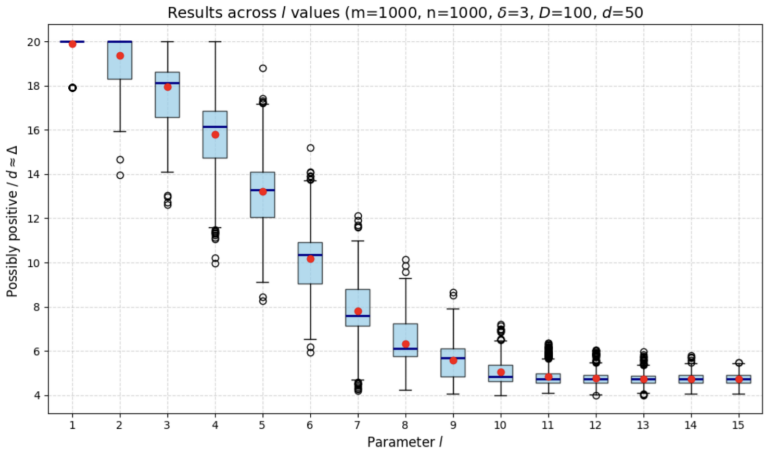


Figure 3: Distribution of Δ for varying l, with $\delta = 3$, $D = 100$, $d = 50$, and $n = 1000$. Box plots over $m = 1000$ runs per value of l; red dots mark the means.

It appears that the algorithm performs generally well – except for very specific settings, in the vast majority of cases, the accuracy is on the order of $\delta + 1$ with high probability (including outliers). Furthermore, in many cases, it remains well below this limit.

It is also worth noting that:

- The algorithm adapts very well to d, even when d differs significantly from the upper bound D (see Figure 1).
- The behavior of the algorithm is monotonic with respect to the parameters d and D (Fig. 1 and 2).
- Generally, increasing the accuracy of the feedback function (parameter ℓ) improves the overall accuracy (Fig. 3). Nevertheless, occasionally, increasing l may slightly reduce Δ (see App. C).

It appears that the main determinant of the SUBMARTINGALE algorithm's performance is the ratio n/d, i.e., the density of defective elements.

6 Lower Bound

We call a deniable group testing algorithm ***monotonic*** if the feedback vector is monotonic; more precisely, for any admissible sets $S' \subseteq S$, we get:

(a) the length t' of $F(S')$ is not bigger that the length t of $F(S)$, and

(b) for any $1 \leq i \leq t'$, we have $F(S')[i] \leq F(S)[i]$.

Observe that non-adaptive algorithms are monotonic for the considered feedback functions $\mathcal{F}_\ell$, because they use the same sequence of tests to all admissible sets of defective elements and the feedback function is monotonic.

Theorem 3. *Any monotonic (δ, Δ)-deniable GT algorithm under a feedback function $\mathcal{F}_\ell$, where $\delta \leq \ell$, has accuracy $\Delta \geq \log(\delta + 1) + 1$ for $\ell = 1$, and accuracy $\Delta \geq \frac{\delta+1}{\ell}$ for $2 \leq \ell \leq \delta$.*

Proof. The complete proof is deferred to Appendix B, due to space limitation. □

7 Conclusions, Limitations and Future Work

In this paper, we proposed a new paradigm of group testing that, at the cost of a controlled loss of accuracy, provides **all** individuals with plausible deniability regarding an unfavorable test result. In contrast to somewhat related concepts based on differential privacy, our approach does not require randomization and guaranteed deniability holds deterministically (although its accuracy/efficiency was tested on random inputs). Limitations and future directions are presented in Appendix A.

AI USE STATEMENT

We used AI tools solely for language editing and proofreading to improve the clarity and grammar of the manuscript. Additionally, AI was used to assist in reviewing and verifying the code for the numerical experiments, which is provided as supplementary material. We have not used generative AI tools for other tasks, and the rest of the required disclosure tasks are not applicable to this work. We take responsibility for the final content of this work, including text, claims, or artefacts produced with the aid of generative AI.

ETHICS STATEMENT

Our results present a concept of fundamental nature with no direct relation to issues that could give rise to any ethical or legal concerns.

REPRODUCIBILITY STATEMENT

Our work is primarily of theoretical in nature. We believe that the proofs presented in this paper are complete and verifiable. Furthermore, to ensure full reproducibility of the results, we have provided the code used for analyzing the experimental accuracy (parameter Δ) of the SUBMARTINGALE algorithm for various parameters δ, δ, D, n etc. in the supplementary material.

REFERENCES

Matthew Aldridge, Oliver Johnson, and Jonathan Scarlett. Group testing: An information theory perspective. *Foundations and Trends® in Communications and Information Theory*, 15(3-4): 196–392, 2019. doi: 10.1561/0100000099.

Jessa Bekker and Jesse Davis. Learning from positive and unlabeled data: A survey. *Machine Learning*, 109(4):719–760, 2020.

David Chaum and Hans Van Antwerpen. Undeniable signatures. In *Conference on the Theory and Application of Cryptology*, pp. 212–216. Springer, 1989.

Mahdi Cheraghchi and Vasileios Nakos. Combinatorial group testing and sparse recovery schemes with near-optimal decoding time. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 1203–1213. IEEE, 2020. doi: 10.1109/FOCS46700.2020.00115.

Lorenzo Ciampiconi, Bishwamittra Ghosh, Jonathan Scarlett, and Kuldeep S. Meel. A maxsat-based framework for group testing. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pp. 10144–10152. AAAI Press, 2020. doi: 10.1609/AAAI.V34I06.6574. URL https://doi.org/10.1609/aaai.v34i06.6574.

Annalisa De Bonis, Leszek Gąsieniec, and Ugo Vaccaro. Optimal two-stage algorithms for group testing problems. *SIAM Journal on Computing*, 34(5):1253–1270, 2005. doi: 10.1137/S0097539703428002.

Robert Dorfman. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, 14(4):436–440, 1943.

Ding-Zhu Du and Frank K. Hwang. *Combinatorial Group Testing and Its Applications*. World Scientific, 2nd edition, 2000.

Cynthia Dwork. Differential privacy. In *Automata, Languages and Programming*, pp. 1–12, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-35908-1. doi: 10.1007/11787006_1.

A. G. D'yachkov and V. V. Rykov. Bounds on the length of disjunctive codes. *Problems of Information Transmission*, 18(3):166–171, 1982. Translated from *Problemy Peredachi Informatsii*, 18(3):7–13.

Joshua Engels, Benjamin Coleman, and Anshumali Shrivastava. Practical near neighbor search via group testing. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 9950–9962, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/5248e5118c84beea359b6ea385393661-Abstract.html.

Yigit Efe Erginbas, Justin Singh Kang, Elizabeth Polito, and Kannan Ramchandran. Adaptive sparse möbius transforms for learning polynomials, 2026. URL https://arxiv.org/abs/2602.06246.

Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, Cambridge, 2009. ISBN 978-0521898065.

Mira Gonen, Michael Langberg, and Alex Sprintson. On subset retrieval and group testing problems with differential privacy constraints. In *2025 61st Allerton Conference on Communication, Control, and Computing Proceedings*. Allerton Conference on Communication, Control, and Computing, 2025. URL https://hdl.handle.net/2142/130305.

F. K. Hwang. A method for detecting all defective members in a population by group testing. *Journal of the American Statistical Association*, 67(339):605–608, 1972. doi: 10.1080/01621459.1972.10481257.

W. H. Kautz and R. C. Singleton. Nonrandom binary superimposed codes. *IEEE Transactions on Information Theory*, 10(4):363–377, 1964. doi: 10.1109/TIT.1964.1053689.

Marek Klonowski, Przemyslaw Kubiak, and Miroslaw Kutylowski. Practical deniable encryption. In Viliam Geffert, Juhani Karhumäki, Alberto Bertoni, Bart Preneel, Pavol Návrat, and Mária Bieliková (eds.), *SOFSEM 2008: Theory and Practice of Computer Science, 34th Conference on Current Trends in Theory and Practice of Computer Science, Nový Smokovec, Slovakia, January 19-25, 2008, Proceedings*, volume 4910 of *Lecture Notes in Computer Science*, pp. 599–609. Springer, 2008. doi: 10.1007/978-3-540-77566-9_52. URL https://doi.org/10.1007/978-3-540-77566-9_52.

Marek Klonowski, Dariusz R. Kowalski, and Dominik Pajak. Generalized framework for group testing: Queries, feedbacks and adversaries. *Theor. Comput. Sci.*, 919:18–35, 2022. doi: 10.1016/J.TCS.2022.03.026. URL https://doi.org/10.1016/j.tcs.2022.03.026.

Dariusz R. Kowalski and Dominik Pajak. Scalable and efficient non-adaptive deterministic group testing. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/61fc0928bc62ad9cf0cb5cab961fc178-Abstract-Conference.html.

Bing Liu, Wee Sun Lee, Philip S Yu, and Xiaoli Li. Partially supervised classification of text documents. In *ICML*, volume 2, pp. 387–394, 2002.

Jan Mielniczuk and Adam Wawrzeńczyk. Single-sample versus case-control sampling scheme for positive unlabeled data: the story of two scenarios. *Fundamenta Informaticae*, 193:29–45, 2025. doi: 10.46298/fi.12648.

Ely Porat and Amir Rothschild. Explicit nonadaptive combinatorial group testing schemes. *IEEE Transactions on Information Theory*, 57(12):7982–7989, 2011. doi: 10.1109/TIT.2011.2163296.

Arlei Silva and Ambuj Singh. Group testing on a network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 4348–4356, 2021. doi: 10.1609/aaai.v35i5.16560. URL https://ojs.aaai.org/index.php/AAAI/article/view/16560.

Latanya Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.*, 10(5):571–588, 2002. doi: 10.1142/S021848850200165X. URL https://doi.org/10.1142/S021848850200165X.

Stanley L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965. doi: 10.1080/01621459.1965.10480775.

Yingbo Zhou, Utkarsh Porwal, Ce Zhang, Hung Q. Ngo, XuanLong Nguyen, Christopher Ré, and Venu Govindaraju. Parallel feature selection inspired by group testing. In Zoubin Ghahramani, Max Welling, Corinna Cortes, Neil D. Lawrence, and Kilian Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pp. 3554–3562, 2014. URL https://proceedings.neurips.cc/paper/2014/hash/fb8feff253bb6c834deb61ec76baa893-Abstract.html.

A Appendix: Limitations and Future Work

The presented results have some limitations, which at the same time serve as starting points for future work. The follow-up questions we need to address include:

1. **Theoretical studies:**

- Tight lower bounds for a broader range of parameters.
- A formal accuracy analysis of the SUBMARTINGALE algorithm. It seems clear that the algorithm depends on multiple parameters, and empirical results suggest that the impact of individual parameters is highly complex. On the other hand, obtaining partial theoretical results regarding accuracy appears feasible using techniques from analytic combinatorics Flajolet & Sedgewick (2009) by analyzing the algorithm's execution tree (at least for selected, constant parameters ℓ and δ).
- Analysis of the required number of queries. Minimizing the query complexity is one of the core problems studied in the group testing literature. It remains unclear whether the deniability property forces an increase or allows a decrease in the total number of queries. On one hand, less information about the set of defective elements is required, which could potentially reduce the number of queries. On the other hand, maintaining accuracy sometimes necessitates asking additional queries.

2. **Extensions of the paradigm:**

- Although the underlying concept is general, the algorithms and analysis were presented for the highly regular case where the family of possible defective sets is $\mathcal{S}_D$. While this is the most frequently studied case in the literature, it may be far from realistic in many practical applications. A natural research question is whether an efficient algorithm analogous to SUBMARTINGALE exists that provides comparable properties for an arbitrary set family $\mathcal{S}$. It is clear that such an algorithm can be constructed if we allow a running time of $\Theta(|\mathcal{S}|)$, which, however, would be prohibitively inefficient.
- It appears possible to consider a randomized counterpart of the presented concept to achieve stronger deniability or even design a hybrid solution leveraging techniques from, e.g., Gonen et al. (2025), which could offer additional properties.

3. **Applications:**

- It is worth emphasizing that group testing is used as a subroutine in dozens of other protocols. It is crucial to determine to what extent the proposed deniable group testing scheme can serve as a building block for other protocols while preserving desirable privacy and performance properties.

B Appendix: Proof of the Lower Bounds in Theorem 3 from Section 6

Fix a monotonic (δ, Δ)-deniable GT algorithm under the feedback function $\mathcal{F}_\ell$. Let $F()$ be the feedback vector created by the algorithm on the parameter set. Let x be such that $F(\{x\})$ is minimal with respect to the vectors' order, in the sense that there is no element y such that $F(\{y\}) < F(\{x\})$.

Consider an admissible feedback vector $f = F(\{x\})$. By definition of f, we have that f is x-compatible, and therefore by δ-deniability, there are δ different sets $S_1, \ldots, S_\delta$ such that $F(S_1) = \ldots = F(S_\delta) = f$. Consider two cases.

Case $\ell = 1$**.** We argue that the union $U = \bigcup_{i \leq \delta} S_i$ of δ non-empty and pairwise-different sets must have at least $\log(\delta + 1)$ elements. Indeed, if it had less elements, the number of all possible different non-empty subsets of U would be smaller than $2^{\log(\delta+1)} - 1 = \delta$, and thus contradicting the fact that the number of such sets contributing to the union U is exactly δ. To finish the analysis of this case, observe that $F(U) = f$, because in this case we consider only a binary feedback $\mathcal{F}_1$ and thus $F(U)$ is a position-wise logic OR on the feedback vectors of its contributing sets, $F(S_1), \ldots, F(S_\delta)$, all being equal to f (we use monotonicity property of the algorithm here). Consequently, if we consider set $S = \{x\}$, we could establish the following bound:

$$\Big| \bigcup_{F(S')=F(S)} S' \Big| \geq |U \cup \{x\}| \geq \log(\delta + 1) + 1 = (\log(\delta + 1) + 1) \cdot |S| \,.$$

It implies that the equation in the worst-case accuracy definition may hold only if $\Delta \geq \log(\delta+1)+1$.

Case $2 \leq \ell \leq \delta$**.** Since $f = F(\{x\})$ and it is a minimal feedback vector with respect to the vector's order, sets $S_1, \ldots, S_\delta$ are (different) singletons. The proof of this fact is by contradiction. Assume otherwise, there was a set S_i, for some $i \leq \delta$, of size at least 2. Note that f contains only zeroes and ones, as the feedback is for a single defective element x. Let us choose arbitrarily two different elements $y_1, y_2 \in S_i$. By monotonicity of the algorithm, we have $F(\{y_1\}) \leq F(S_i) = f$ and $F(\{y_1\}) \leq F(S_i) = f$, and all these vectors are binary. Moreover, the positions with values 1 in both these vectors are *different* – otherwise, vector $F(S_i)$ would have a value bigger than 1 in some positions, which would violate storage of only values $0, 1$ in $F(S_i) = f$. Consequently, f subsumes the position-wise OR of $F(\{y_1\})$ and $F(\{y_2\})$ (recall, that they also have values 1 at disjoint positions). Since each feedback vector contains at least one non-zero value, both vectors $F(\{y_1\})$ and $F(\{y_2\})$ must be strictly smaller than their position-wise OR, and thus also than f. This contradicts the selection of x as an element with minimal feedback vector, and thus conclude the proof of the fact that $S_1, \ldots, S_\delta$ are disjoint singletons.

To finish the analysis of this case, consider set $S = S_1 \cup \ldots \cup S_\ell$. Since the union components are different singletons, each giving the same binary feedback vector f, we have $|S| = \ell$ and $F(S) = \ell \cdot f$, by monotonicity. Analogously to the previous case, let $U = \bigcup_{i \leq \delta} S_i$. Using analogous arguments as for the set S, we get $|U| = \delta$ and $F(U) = \ell \cdot f$. The latter is because the feedback function $\mathcal{F}_\ell$ has a cap ℓ on the maximum feedback value, while each single union component has the same binary feedback vector f and we could apply monotonicity property. Using these facts, we could derive the following bound:

$$\Big| \bigcup_{F(S')=F(S)} S' \Big| \geq |U \cup \{x\}| \geq \delta + 1 = \frac{\delta + 1}{\ell} \cdot \ell = \frac{\delta + 1}{\ell} \cdot |S| \,.$$

It implies that the equation in the definition of the worst-case accuracy could be satisfied only if $\Delta \geq \frac{\delta+1}{\ell}$.

C Appendix: Extended Experimental Results

In this Appendix, we present additional experimental results concerning the SUBMARTINGALE algorithm that could not be included in the main body of the paper due to space limitations.

Recall that the SUBMARTINGALE algorithm provably guarantees δ-deniability for all inputs. However, a formal accuracy analysis (i.e., determining the achieved value of Δ) was omitted in favor of an experimental evaluation across various choices of parameters l, δ, D, and d. In the experiments presented here, as in the main text, we set $n = 1000$ and executed the algorithm $m = 1000$ times for each configuration. The primary metric of interest is the ratio of the number of *possibly defective* items to the number of true defective items (represented by parameter d).

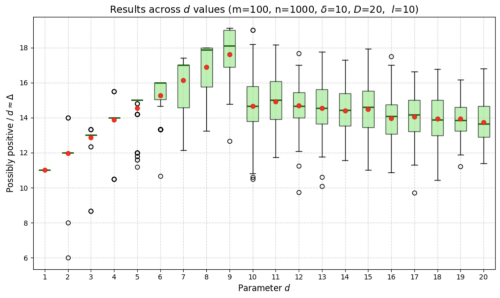
Below, we provide eight figures showing representative algorithm outputs for different fixed parameter settings. Each parameter set yields an empirical distribution of Δ, which is depicted using a box plot.

It can be observed that the algorithm yields a reasonable Δ across a wide range of parameters. Except for a few special cases, the obtained values of Δ are, with high probability, of the same order of magnitude as the median, which lies close to the optimal solution. Outliers are relatively scarce and occur only for specific parameter configurations.

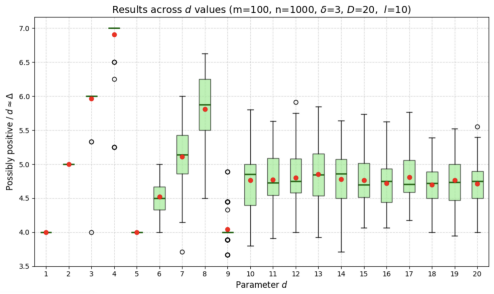
The behavior of the algorithm appears highly non-obvious. We can observe that the median value of Δ is **not** monotonic with respect to the parameter d (Figures 4a and 4b) or the parameter D (Figures 4c and 4d).

It is also worth noting that a larger value of l (representing the power of the feedback function) generally improves accuracy. In certain cases, the parameter l plays a decisive role. See Figure 4f, where the difference between capped quantitative feedback ($l \geq 2$) and binary feedback ($l = 1$) is dramatic. On the other hand, Figure 4g shows that accuracy can actually decrease as l increases, which is somewhat surprising.

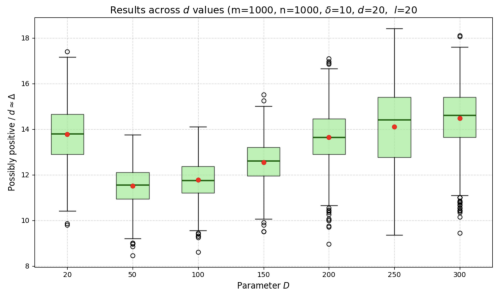
Finally, to ensure reproducibility, we have attached the experiment execution script as supplementary material. This script can also be utilized to run additional experiments for arbitrary parameter values.



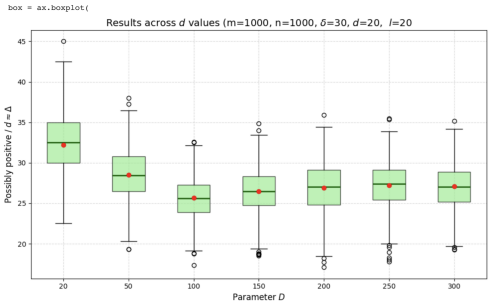
(a) Various d for fixed $\delta = 10,\ D = 20\ l = 10$



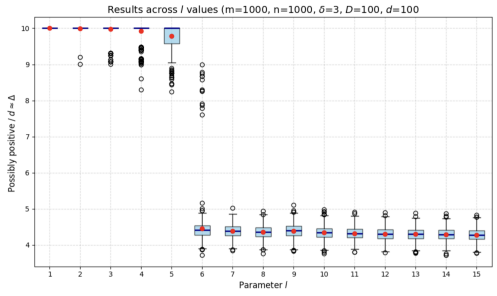
(b) Various d for fixed $\delta = 3,\ D = 20\ l = 10$



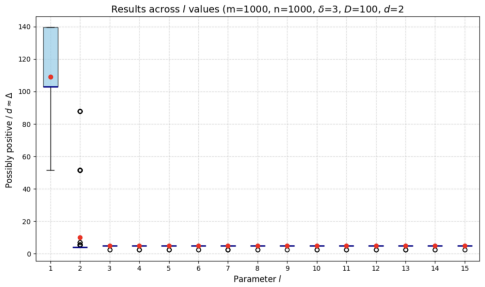
(c) Various D for fixed $\delta = 10,\ d = 20\ l = 20$



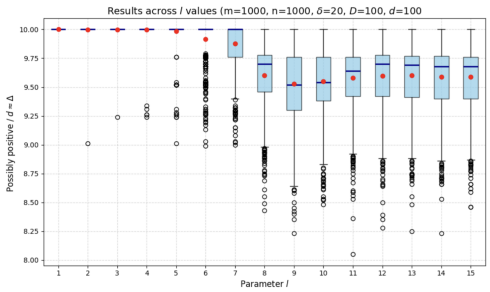
(d) Various D for fixed $\delta = 30,\ d = 20\ l = 20$



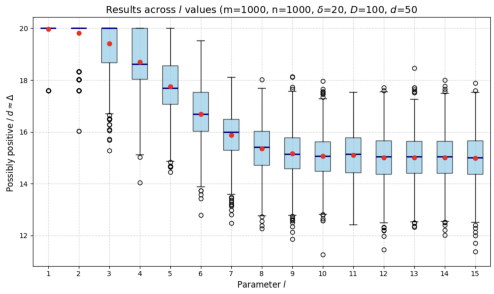
(e) Various l for fixed $\delta = 3,\ D = 100\ d = 100$



(f) Various l for fixed $\delta = 3,\ D = 100\ d = 2$



(g) Various l for fixed $\delta = 3,\ D = 100\ d = 100$



(h) Various l for fixed $\delta = 3,\ D = 100\ d = 50$

Figure 4: Examples of extended experiments.