

Smart elements in combinatorial group testing problems with more defectives*

Dániel Gerbner[†] , Máté Vizer[‡] 

MTA Rényi Institute, Hungary H-1053, Budapest, Reáltanoda utca 13-15

Received 23 May 2017, accepted 18 February 2019, published online 14 August 2020

Abstract

In combinatorial group testing problems Questioner needs to find a defective element $x \in [n]$ by testing subsets of $[n]$. In [18] the authors introduced a new model, where each element knows the answer for those queries that contain it and each element should be able to identify the defective one.

In this article we continue to investigate this kind of models with more defective elements. We also consider related models inspired by secret sharing models, where the elements should share information among them to find out the defectives. Finally the adaptive versions of the different models are also investigated.

Keywords: Combinatorial group testing, defectives, cancellative.

Math. Subj. Class. (2020): 94A50

1 Introduction

In the most basic *model of combinatorial group testing* Questioner needs to find a special element x of $\{1, 2, \dots, n\}$ ($=: [n]$) by asking minimal number of *queries* (or *group tests* or *pools*) of type “does $x \in F \subset [n]$?”. Special elements are usually called *defective* (or *positive*). For every combinatorial group testing problem there are at least two main approaches: whether it is *adaptive* (or *sequential*) or *non-adaptive* (or *oblivious*). In the adaptive scenario Questioner asks queries depending on the answers for the previously asked queries, however in the non-adaptive version Questioner needs to pose all the queries

*We would also like to thank all participants of the Combinatorial Search Seminar at the Alfréd Rényi Institute of Mathematics for fruitful discussions.

[†]Research supported by the János Bolyai Research Fellowship of the Hungarian Academy of Sciences. Research supported by the National Research, Development and Innovation Office – NKFIH, grant K116769.

[‡]Research supported by the National Research, Development and Innovation Office – NKFIH, grant SNN 116095.

E-mail addresses: gerbner@renyi.hu (Dániel Gerbner), vizermate@gmail.com (Máté Vizer)

at the beginning. We call the *complexity* of a specific combinatorial group testing problem the number of the queries needed to ask by Questioner in the worst case during an optimal strategy.

Combinatorial group testing problems were first considered during the World War II by Dorfman [10] in the context of mass blood testing. Since then group testing techniques have had many different applications, for example in fault diagnosis in optical networks [20], in quality control in product testing [26] or failure detection in wireless sensor networks [23]. In this article we will mainly discuss non-adaptive models. The interested reader can find many variants and generalizations of the basic non-adaptive model and also many applications in the book [11].

1.1 Description of the new model

In [18] the authors introduced new combinatorial group testing models, inspired by the results of Tapolcai et al. [30, 29].

The main novel ingredient of these combinatorial group testing models is that the elements are smart and they distrust the Questioner, thus they want to control the tests they are involved in. So the following extra condition was introduced: **each element knows the answer for those queries that contain it**, and the goal: each element should be able to identify the defective one.

Motivated by secret sharing schemes (see e.g. [2]), the following variant was also considered: the elements can work together and share their knowledge. In this case we require certain sets of elements to be able to identify the defective, while we require other sets to be unable to identify the defective element. We emphasize that the way the data is transmitted does not play a role here. Information can not be distributed between different groups.

We mention here some other motivation to introduce these models: it is often mentioned in the group testing literature that an advantage of testing pools together is that it increases privacy. However, systematical research on this property has only started recently, see e.g. [1, 5, 16]. These papers focus on cryptographic versions of the problem. Here we deal with a simple combinatorial version, where privacy only means that an unauthorized participant cannot completely detect the defective element(s). In [18] the authors considered models with one defective element. The main aim of this article is to continue these investigations with more defectives.

1.2 Simple combinatorial models with d defectives

A well-studied generalization of the basic model is the following. There are exactly d defective elements, a query corresponds to a set F , and the answer shows if there is at least one defective elements in F or not.

About Questioner's strategy we remark that - as he should find all the defectives - the asked queries should form a *d-separating family* (see the next section for a definition) in the non-adaptive case, so for the minimum number of tests the known lower bound is $\Omega(\frac{d^2}{\log d} \log n)$, while the best upper bound construction yields $O(d^2 \log n)$ (see e.g. [14, 25]). It is one of the major open problems in the theory of combinatorial group testing models to close the gap between the previous upper and lower bound.

In the adaptive case there is a multiplicative constant factor between the information theoretic lower bound and the best existing algorithm. The known best lower bound is $d \log \frac{n}{d}$, while the upper bound is $O(d \log n)$.

1.3 Structure of the paper

We organize the paper as follows: in Section 2 we introduce some properties and related results about families of sets, that we will need later. In Section 3 we introduce the non-adaptive models that we investigate, while in Section 4 we prove the main results. In Section 5 we look at the adaptive scenario, and we finish the article with remarks and open questions in Section 6.

We also mention that in this article we use standard asymptotic notation.

2 Finite set theory background

Our topic is connected to several areas of finite set theory. In this section we introduce some notions on families of subsets and known results about them, that we will use during the proofs.

In this article we use the notation of $2^{[n]}$ for the power set of $[n]$ and for any $\mathcal{F} \subset 2^{[n]}$, $a \in [n]$ we use $\mathcal{F}_a := \{F \in \mathcal{F} : a \in F\}$. The *complement* of a family $\mathcal{F} \subset 2^n$ is $\overline{\mathcal{F}} := \{[n] \setminus F : F \in \mathcal{F}\}$, while the *dual* of a family $\mathcal{F} \subset 2^n$ is $\mathcal{F}' := \{\mathcal{F}_a : a \in [n]\}$. It is defined on the underlying set \mathcal{F} and has cardinality at most n . For a family $\mathcal{F} \subset 2^{[n]}$ and $d \geq 1$ let $\mathcal{F}^d := \{\cup_{i=1}^d F_i : F_i \in \mathcal{F}, F_i \neq F_j \text{ for } i \neq j\}$.

Now we introduce some notions about families of subsets of $[n]$.

Definition 2.1. We say that $\mathcal{F} \subset 2^{[n]}$ is:

- (1) **intersection closed** if $F, G \in \mathcal{F}$ implies $F \cap G \in \mathcal{F}$.
- (2) **Sperner** if there are no two different $F_1, F_2 \in \mathcal{F}$ with $F_1 \subset F_2$.
- (3) **cancellative** if for any three $F_1, F_2, F_3 \in \mathcal{F}$ we have

$$F_1 \cup F_2 = F_1 \cup F_3 \Rightarrow F_2 = F_3.$$

- (4) **intersection cancellative** if for any three $F_1, F_2, F_3 \in \mathcal{F}$ we have

$$F_1 \cap F_2 = F_1 \cap F_3 \Rightarrow F_2 = F_3.$$

- (5) **d-separating** for some $1 \leq d \leq n - 1$ positive integer, if for any two different $X_1, X_2 \subset [n]$ with $|X_1| = |X_2| = d$ there is $F \in \mathcal{F}$ with:

$$F \cap X_1 \neq \emptyset \text{ and } F \cap X_2 = \emptyset, \text{ or}$$

$$F \cap X_2 \neq \emptyset \text{ and } F \cap X_1 = \emptyset.$$

- (6) **d-union-free** for some $d \geq 1$ if for different $F_1, \dots, F_d \in \mathcal{F}$ and different $G_1, \dots, G_d \in \mathcal{F}$

$$\bigcup_{i=1}^d F_i = \bigcup_{i=1}^d G_i$$

$$\text{implies } \{F_1, \dots, F_d\} = \{G_1, \dots, G_d\}.$$

- (7) **d-cover-free** for some $d \geq 1$ positive integer if there are no $(d + 1)$ different $F_1, F_2, \dots, F_{d+1} \in \mathcal{F}$ with

$$F_{d+1} \subset \bigcup_{i=1}^d F_i.$$

- (8) **(r,d)-cover-free** for some $r, d \geq 1$ positive integers if there are no $(d + r)$ different $F_1, F_2, \dots, F_{d+r} \in \mathcal{F}$ with

$$\bigcap_{i=d+1}^{d+r} F_i \subset \bigcup_{i=1}^d F_i.$$

Before defining the last notion, we need some introduction. We will generalize a graph property, so it is more comfortable to use the word hypergraph instead of family of subsets of $[n]$ (where \mathcal{F} is the set of the hyperedges and $[n]$ is the set of vertices of the hypergraph). There are several ways to define cycles in hypergraphs. Here we use one due to Berge [4]. A *Berge-cycle* in a hypergraph of length k (a *Berge- C_k*) consists of k different hyperedges E_1, \dots, E_k and k different vertices x_1, \dots, x_k such that E_i contains x_i and x_{i+1} for $1 \leq i \leq k$ (modulo k , so E_k contains x_k and x_1). Note that for a 2-uniform hypergraph (that is a graph) this notion is the same as the 'usual' cycle in a graph. The *Berge-girth* (that we call just *girth* in this article) of a hypergraph \mathcal{H} is the smallest length of a cycle in \mathcal{H} (that is ∞ if there is no cycle in \mathcal{H}). A hypergraph is d -regular if every vertex is contained in exactly d hyperedges, r -uniform if every hyperedge has size r and linear if any two hyperedges intersect in at most one vertex.

2.1 Some known results about these notions that we will use later

- The notion cancellative was first introduced by Frankl and Füredi in [17].

Fact 2.2. $\mathcal{F} \subset 2^{[n]}$ is intersection cancellative if and only if $\overline{\mathcal{F}}$ is cancellative.

- The notion of separating family in the context of combinatorial search theory was introduced and first studied by Rényi in [24]. The following fact is rather trivial, so we omit its proof.

Fact 2.3. Suppose $\mathcal{F}_n \subset 2^{[n]}$ is a minimal separating family. Then we have:

$$|\mathcal{F}_n| \leq \lceil \log_2 n \rceil.$$

Fact 2.4. $\mathcal{F} \subset 2^{[n]}$ finds d defectives if and only if \mathcal{F} is d -separating. The dual of a d -separating family is d -union-free.

- The notion of d -union-free families was introduced by Hwang and T. Sós in [21] under the name of d -Sidon families. They proved the following:

Theorem 2.5 (Hwang, T. Sós, [21, Theorem 3]). *There exists a d -union-free family $\mathcal{F}_n \subset 2^{[n]}$ with:*

$$\frac{1}{2} \left(1 + \frac{1}{(4d)^2}\right)^n \leq |\mathcal{F}_n|.$$

- The notion of d -cover-free families was introduced by Kautz and Singleton in [22]. Note that a d -cover-free family is also d -union-free. They proved the following lower bound.

Theorem 2.6 (Kautz, Singleton, [22]). *There exists a d -cover-free family $\mathcal{F}_n \subset 2^{[n]}$ with:*

$$\Omega\left(\frac{1}{d^2}\right) = \frac{\log_2 |\mathcal{F}_n|}{n}.$$

D'yachkov and Rykov proved the following upper bound on the size of d -cover-free families:

Theorem 2.7 (D'yachkov, Rykov, [14]). *Suppose that $\mathcal{F}_n \subset 2^{[n]}$ is a d -cover-free family. Then we have:*

$$\frac{\log_2 |\mathcal{F}_n|}{n} \leq \frac{2 \log_2 d}{d^2} (1 + o(1)).$$

- The notion of (r, d) -cover-free families were introduced by D'yachkov, Macula, Torney and Vilenkin in [12]. They showed that a result by Stinson, Wei and Zhu [28] implies the following:

Theorem 2.8. *If the dual of $\mathcal{F}_n \subset 2^{[n]}$ is (r, d) -cover-free, then we have:*

$$|\mathcal{F}_n| = \Omega_{d,r}(\log_2 n).$$

- Ellis and Linial [15] studied regular uniform linear hypergraphs with large girth. They mention that a result of Cooper, Frieze, Molloy and Reed [6] implies that for any $d \geq 2$, $r, g \geq 3$ and sufficiently large n , if r divides n , then there is an r -uniform, d -regular, n -vertex linear hypergraph with girth at least g . Moreover the argument can be adapted to show the same statement in the case r divides dn .

Theorem 2.9. *Let $d \geq 2$, $r, g \geq 3$, n large enough and r divides dn . Then there exists a linear, d -regular, r -uniform hypergraph with girth at least g on n vertices.*

3 Models

In this section we start our investigations and give a systematic study of models with the extra property that each element knows the answers for those queries that contain it.

In all the models in this section an input set $[n]$ is given, and d of them are defectives ($d \leq n$). We are dealing with non-adaptive models, so Questioner needs to construct a family $\mathcal{F} \subset 2^{[n]}$. A set F correspond to a query of the following type: 'is there any defective element in $F \subset [n]$?'. In each model we assume that knowing all the answers is enough information for Questioner to find the defective elements, i.e. \mathcal{F} is d -separating. Note that this immediately implies a lower bound of $\Omega_d(\log_2 n)$ on the size of the query family in each model. We mention whenever the query family satisfies another property that could improve the factor depending only on d , but calculating the factors is outside the scope of this paper.

The main difference between the following models is what we want the elements to find out. Using only the information available to them, i.e. the answers to the queries containing them, we can require that they find out something about the defective elements, or oppositely, that they cannot find out something.

When we say that an element x *knows the defective elements*, we mean that the query family satisfies the following property: no matter what the defective element is, after the answers x can find out the defective ones, i.e. the subfamily \mathcal{F}_x is d -separating. In the opposite when we say that x *does not know any of the defective elements*, we mean that the

query family satisfies the following property: no matter what the defective elements are, after the answers x cannot identify any defective element. Equivalently, for any $D \subset [n]$, $y \in D$ with $|D| = d$ there is a $D' \subset [n]$ with $|D'| = d$, $y \notin D'$, such that the same members of \mathcal{F}_x intersect D and D' .

Another variant of this problem is when elements can share information among them. It is possible that in some model some element can not find out the defective, however if we pick two elements and they share their information among them, they can find the defective elements. We consider these kind of models.

We also assume that in each model the elements know the setup of the problem, i.e. that n elements are given and exactly d of them are defectives. We use the expression that a family *solves* a model if it satisfies the property that describes the model.

In each of the following models we first give a property describing what the elements should know, and then we examine if there is a query set that solves that specific model or state results about the cardinality of such query sets. Then we consider models where we require some information to remain hidden from the elements. Finally we mix these types of properties.

In this section we assume that there are exactly $d \geq 2$ defective elements (and every element knows that). We consider models analogous to the ones introduced in [18].

3.1 Model 1_d

Probably the most natural model is the following:

Property. All elements find out if they are defective.

We note that some cryptographic problems concerning this model were investigated in [1], where the authors observed that the dual of a d -cover-free family solves this model. Here we show that only such families solve this model.

Theorem 3.1. \mathcal{F} solves Model 1_d if and only if its dual is d -cover-free.

We prove this theorem in Section 4. By Theorem 3.1, Theorem 2.6 and Theorem 2.7 we have:

Corollary 3.2. If $\mathcal{F}_n \subset 2^{[n]}$ solves Model 1_d and has minimum cardinality, then we have:

$$\Omega\left(\frac{d^2}{\log_2 d} \log_2 n\right) = |\mathcal{F}_n| = O(d^2 \log_2 n).$$

3.2 Model 2_d

Another natural model is when the elements should find out everything.

Property. Every element finds all the defectives.

It is obvious that no \mathcal{F} can solve Model 2_d if $1 < d < n$: a defective element cannot gather any information about the other elements, as it gets only YES answers.

3.3 Model $2'_d$

As defective elements cannot gather any information about the other elements, in the next model we only require non-defective elements to find the defective ones.

Property. Every non-defective element finds all the defectives.

Theorem 3.3. *Suppose \mathcal{F}_n solves Model $2'_d$ and has minimum cardinality. Then we have*

$$|\mathcal{F}_n| = \Theta_d(\log_2 n).$$

Proof. We claim that the solution is the dual of a d -cover-free family, and the dual of a $(2, d)$ -cover-free family is always a solution. This together with Theorem 2.7 and Theorem 2.8 implies the statement.

Suppose that the dual is not d -cover-free. Then there are $F_1, F_2, \dots, F_{d+1} \in \mathcal{F}$ with $F_{d+1} \subset \bigcup_{i=1}^d F_i$. For the corresponding elements in the primal version we have x_{d+1} such that any set $F \in \mathcal{F}$ contains one of the other elements x_1, \dots, x_d . Thus if x_1, \dots, x_d are the defectives, x_{d+1} receives only YES answers, thus cannot distinguish this case from the case x_{d+1} and any $d - 1$ other elements are the defectives.

On the other hand let us assume \mathcal{F} is the dual of a $(2, d)$ -cover-free family. Then for every non-defective elements x, y there is a set $F \in \mathcal{F}$ such that F contains both x and y , but none of the defective elements, thus x finds out that y is not defective, \mathcal{F} solves Model $2'_d$. Indeed, there is an element in the intersection of the duals of x and y that is not contained by the duals of the defective elements by the $(2, d)$ -cover-free property. That element is the dual of a set $F \in \mathcal{F}$ that has the desired properties. \square

3.4 Model $2''_d$

The fact that defective elements cannot gather any information about the other elements shows that even $d - 1$ elements together cannot always find the defectives. However, if d elements share information, then either they are all the defectives and they do not need to gather information about the other elements, or at least one of them is not a defective, and then there is a solution by Model $2'_d$.

Property. d elements together know who the defective elements are.

Theorem 3.4. *\mathcal{F}_n solves Model $2''_d$ if and only if its dual \mathcal{G} is d -union-free and \mathcal{G}^d is Sperner and intersection-cancellative.*

We prove this theorem in Section 4. Note that we know the maximum possible size of a Sperner and intersection cancellative family (by results of Frankl and Füredi [17] and Tolhuizen [31]), but we do not know if that construction can be written as \mathcal{G}^d for a d -union-free family \mathcal{G} .

Theorem 3.5. *Suppose \mathcal{F}_n solves Model $2''_d$ and has minimum cardinality. Then we have*

$$|\mathcal{F}_n| = \Theta_d(\log_2 n).$$

Proof. It is easy to see that if a family solves both Model 1_d and Model $2'_d$, then it also solves Model $2''_d$. As we have seen in the proof of Theorem 3.3, a solution for Model $2'_d$ is the dual of a d -cover-free family, thus it also solves Model 1_d by Theorem 3.1. This implies the upper bound. \square

3.5 Model 3_d

Let us now examine the case when we require that elements do *not* find the defective. Note that as always, we assume that knowing all the answers is enough to find the defective element.

Property. No element knows any of the defective ones.

Note that for $d = 1$ there is a solution for Model 2_d and there is no solution for Model 3_d [18]. For $d \geq 2$ the situation is just the opposite: we will show that there is a solution for Model 3_d for n large enough. We will use arguments similar to the ones used in [3].

Theorem 3.6. *If $d \geq 2$, $r \geq 3$ and $n \geq dr + 2$, then an r -uniform, d -regular linear hypergraph with girth at least 5 solves Model 3_d .*

Proof. Let us consider an r -uniform, d -regular linear hypergraph \mathcal{F} of girth 5. For an arbitrary element x its neighborhood consists of d disjoint sets of size $r - 1$. Also, there are more than d elements not in its neighborhood. It is easy to see that by $r \geq 3$ x cannot identify any defective elements.

On the other hand, if we know all the answers, the YES answers form stars with the defective elements in the centers. The elements that get only YES answers are the candidates for being defective. Every candidate that is not defective has to be connected to all the defectives. Two such candidate would form a Berge- C_4 with any two of the defective elements, thus there is only one additional candidate. But then it is the only one among the $d + 1$ candidates that is connected to the other candidates, otherwise we could find a Berge- C_3 . \square

Corollary 3.7. *If n is large enough compared to $d > 1$, then there is a solution for Model 3_d .*

Proof. If $d \geq 3$, let us choose $r = d$, then Theorem 2.9 shows that we can find such a family. If $d = 2$, then Theorem 2.9 with $r = 4$ shows we can find such a family for n even. If n is odd, we find such a family for $n + 1$, and delete an element. The resulting family is not 4-uniform, but that property is not actually needed (in fact, we used only that every set in \mathcal{F} has size at least 3). \square

3.6 Model 4_d

Now we start to investigate models where elements can share information among them. Let i and j be integers with $1 \leq i < j \leq n$. When we say that a set of j elements together know the defective elements, we mean that knowing the answers to all the queries containing at least one element from the set is enough to find all the defectives. Similarly, when we say that a set of i elements do not know any of the defectives, we mean that knowing the answers to all the queries intersecting the set is not enough to identify any of the defective elements.

Property. Any j elements together know the defectives, but i elements together do not know any of the defectives, for some i and j with $1 \leq i < j \leq n$.

Note that Corollary 3.7 shows that there is a solution if $d > 1$, $i = 1$ and $j = n$, where n is large enough compared to d . In fact any $n - r + 1$ elements together know the answer to all the queries, thus it is enough to assume $j \geq n - r + 1$. A more precise version of Theorem 2.9 (see [15], Theorem 5) shows that about $\frac{n^{1/6}}{d}$ can be chosen as r , which shows that j can be as small as $n - \frac{n^{1/6}}{d}$.

Proposition 3.8. *If $i \geq d$ or $j < d$, then there is no solution.*

Proof. Let us assume first we are given j elements. If all of them are defectives, they only get YES answers, and do not gather any information about the other elements.

If $i \geq d$, for a set X let $\mathcal{F}_X := \cup_{x \in X} \mathcal{F}_x$. Let us consider among the d -element sets X such that \mathcal{F}_X is maximal. We claim that if the elements of X are the defectives, they can find it out by sharing information. Indeed, they get only YES answers. If they cannot be sure that they are the defective ones, then there is another d -set Y that could be the set of defectives. It means all the answers to the queries in \mathcal{F}_X would still be YES if Y was the set of defectives, i.e. $\mathcal{F}_X \subseteq \mathcal{F}_Y$. By the assumption on X we have $\mathcal{F}_X = \mathcal{F}_Y$, but then the family is not d -separating. \square

Proposition 3.9. *If $j = d$, then there is no solution.*

Proof. If x receives only YES answers, he cannot find out he is defective, thus there is a set $D = \{y_1, \dots, y_d\}$ no containing X that intersects every member of \mathcal{F}_x . On the other hand, if x and y_1, \dots, y_{d-1} are the defectives, they together can figure that out. In particular, they know that the set of defectives is not D , thus there is a set intersecting $\{x, y_1, \dots, y_{d-1}\}$ but not D . Such a set would be a member of \mathcal{F}_x that does not intersect D , a contradiction. \square

4 Proofs

4.1 Proof of Theorem 3.1

The dual of the d -cover-free property is that for every elements x_1, \dots, x_{d+1} we cannot have that the sets that contain x_{d+1} all contain at least one of the other x_i 's. Let

$$\mathcal{H}_x := \{F \setminus \{x\} : F \in \mathcal{F}_x\},$$

and $\tau(\mathcal{H}_x)$ be the size of the smallest set that intersects every member of \mathcal{H}_x . With these notation the following lemma finishes the proof of Theorem 3.1.

Lemma 4.1. *An element x always finds out if he is defective if and only if $\tau(\mathcal{H}_x) > d$.*

Proof. If x gets a NO answer, he learns he is not defective, thus we can assume he only gets YES answers. If \mathcal{H}_x cannot be covered by at most d elements different from x , then the only way to get YES answer to every element of \mathcal{F}_x is if x is defective (as defective elements cover the sets that get YES answers). On the other hand if \mathcal{H}_x can be covered by at most d elements different from x , then x cannot exclude the possibility that those are the defective elements, together with arbitrary additional elements to reach d defectives. \square

4.2 Proof of Theorem 3.4

Lemma 4.2. $\mathcal{F} \subset 2^{[n]}$ solves Model $2''_d$ if and only if the following two properties hold:

- (1) for any two different d -element sets $X, Y \subset [n]$ there is $F \in \mathcal{F}$ with $F \cap X \neq \emptyset$ and $F \cap Y = \emptyset$, and
- (2) for any three different d -element sets $X, Y, Z \subset [n]$ there is $F \in \mathcal{F}$ with $(F \cap X \neq \emptyset$ and $F \cap Y \neq \emptyset$ and $F \cap Z = \emptyset)$ or $(F \cap X \neq \emptyset$ and $F \cap Z \neq \emptyset$ and $F \cap Y = \emptyset)$.

Proof.

1. Note that the property that Questioner can find out the answer is: for any two different d -element sets $X, Y \subset [n]$ there is $F \in \mathcal{F}$ with $(F \cap X \neq \emptyset \text{ and } F \cap Y = \emptyset)$ or $(F \cap Y \neq \emptyset \text{ and } F \cap X = \emptyset)$. This property is contained in (1).

Let us assume now X is a set of size d .

2. If X is the set of defectives, they have to find this out. It means that for a different d -element set Y , there should be an $F \in \mathcal{F}$ with $X \cap F \neq \emptyset$ and $Y \cap F = \emptyset$.
3. If X is not the set of defectives, then another set Y is, and they have to identify Y . Thus for a third d -element set Z , there should be a set that intersects X (so they know the answer for it), and distinguishes Y and Z , i.e. it intersects exactly one of them. □

Lemma 4.3. $\mathcal{F} \subset 2^{[n]}$ satisfies properties (1) and (2) if and only if its dual \mathcal{G} is d -union-free and \mathcal{G}^d is Sperner and intersection cancellative.

Proof. The dual of (1) is the following statement:

- (3) for two different subfamilies each consisting of d sets $\{F_1, \dots, F_d\}, \{G_1, \dots, G_d\} \subset \mathcal{F}$ there is $f \in [n]$ with $f \in \cup_{i=1}^d F_i \setminus \cup_{i=1}^d G_i$.

The dual of (2) is the following statement:

- (4) for three different subfamilies each consisting of d sets

$$\{F_1, \dots, F_d\}, \{G_1, \dots, G_d\}, \{H_1, \dots, H_d\} \subset \mathcal{F}$$

there is $f \in [n]$ with either

$$f \in (\cup_{i=1}^d F_i \cap \cup_{i=1}^d G_i) \setminus \cup_{i=1}^d H_i, \text{ or}$$

$$f \in (\cup_{i=1}^d F_i \cap \cup_{i=1}^d H_i) \setminus \cup_{i=1}^d G_i.$$

It is easy to see that (3) is equivalent to the statement that \mathcal{G} is d -union-free and \mathcal{G}^d is Sperner. Now we claim that (4) means that \mathcal{G}^d is intersection cancellative. Let us use the following notation:

$$F := \cup_{i=1}^d F_i, \quad G := \cup_{i=1}^d G_i, \quad H := \cup_{i=1}^d H_i.$$

Using these, the existence of f means either $F \cap G \not\subset H$ or $F \cap H \not\subset G$. Let us define three properties.

- (i) $F \cap G \not\subset H$.
- (ii) $F \cap H \not\subset G$.
- (iii) $H \cap G \not\subset F$.

Property (2) (for these three sets in this order) means that at least one of (i) and (ii) holds. Considering the same three sets in different orders we get that also at least one of (i) and (iii) and one of (iii) and (ii) holds. It is true if and only if at least two of these three properties hold.

To finish the proof of Lemma 4.3 we prove the following:

Claim 4.4. A family $\mathcal{H} \subset 2^{[n]}$ is intersection cancellative if and only if at least two out of (i), (ii) and (iii) hold for any three members of it.

Proof. Let us assume \mathcal{F}' is intersection cancellative and let $F, G, H \in \mathcal{H}$. Let us assume at most one, say (iii) of the three properties holds, thus (i) and (ii) do not hold. The first one implies $F \cap G \subset H$, and obviously $F \cap G \subset F$. Thus we have $F \cap G \subset F \cap H$. Similarly the second one implies $F \cap H \subset F \cap G$, hence they together imply $F \cap H = F \cap G$, which contradicts the intersection cancellative property and our assumption that F, G, H are three different sets.

Let us assume now that \mathcal{H} is not intersection cancellative, thus we have $F \cap G = F \cap H$. This implies both $F \cap G \subset H$ and $F \cap H \subset G$, thus at most one of (i), (ii) and (iii) can hold. \square

We are done with the proof of Theorem 3.4. \square

5 Adaptive scenario

A natural idea is to consider the adaptive versions of these problems. Here we assume the Questioner knows all the earlier answers, and then he can choose the next query. He can find the defective, and then use further queries to share some information with the elements. However, there are two versions of this problem. The elements might know the algorithm, and use the order of the queries to gain information, or they only receive the answers to the queries at the end in no particular order.

For example in Model 2_d in the second version we require that for every element x the family \mathcal{F}_x with the answers is enough to find all the defectives, i.e. for two distinct sets D, D' of size d there is a query that contains x and intersects only one of D and D' . It is still obviously not solvable, as every defective element only gets YES answers and no information about the others. However, in the first version Questioner may start with a d -separating family, then ask the set of defectives and then the set of non-defectives. This way every element has to look only at the last query that contains it. If the answer to that is YES, then it is the set of defectives, if the answer is NO, it is the set of non-defectives. In both cases the defectives are identified.

From now on we consider only the second version, i.e. the elements receive the answer to the queries containing them at the end of the algorithm in no particular order, and they only know the underlying set and the number of defectives. It is still possible for the Questioner to find the defective, and then share some information using further queries.

Let $t^a(d, n)$ denote the number of queries in the fastest adaptive algorithm that finds the d defective (we mentioned some inequalities on $t^a(d, n)$ in the introduction), then $t^a(d, n)$ is a lower bound in every model. On the other hand $t^a(d, n) + d + 1$ queries are enough in Model 1_d , $t^a(d, n) + 1$ queries are enough in Model $2'_d$ and $t^a(d, n) + d + 1$ queries are enough in Model $2''_d$: first Questioner finds the d defectives, then ask them as singletons, and/or the set of non-defectives.

Let us consider now Model 3_d . By Corollary 3.7 there is a solution for n large enough, but that solution is linear in n . On the other hand it can be seen easily that for $n = d + 1$ there is no solution even adaptively. Here we give a faster algorithm.

Theorem 5.1. There is an adaptive algorithm that solves Model 3_d and uses at most $2d \log_2 n + 5d$ queries if n is large enough.

Proof. Questioner starts with asking a query Q of size $\lfloor n/2 \rfloor$ and its complement. Then in the next round he asks two complementing subsets of size differing by at most one in every query that was answered YES (say Q_1 and Q_2 with $Q_1 \cup Q_2 = Q$). He repeats this in every round except if the subset has size at most 5, he stops and does not ask that subset as a query. Since he asks disjoint sets in every round, he gets at most d YES answers, thus there are at most $2d$ queries in the next round. There are obviously at most $\log_2 n$ rounds. After that we have a family \mathcal{D} of at most d sets of size at least 3 and at most 5, each containing at least one defective element. Let $A := \{a_1, \dots, a_l\}$ be their union ($l \leq 5d$), we also know that every defective is in A . Let $D_i \in \mathcal{D}$ be the set that contains a_i . As n is large enough, we can assume that there were two disjoint queries B and C that were answered NO and have size at least $5d$. Let b_1, \dots, b_{5d} be distinct elements of B and c_1, \dots, c_{5d} be distinct elements of C . Then Questioner also asks the queries $\{a_i, b_i, c_i\}$ for $i \leq l$.

As we know b_i and c_i are not defective, Questioner finds out if a_i is defective for every i . On the other hand, if a_i is defective, every query Q that contains a_i also contains either other elements of D_i or contains b_i , thus a_i cannot be sure he is defective. If a_i is not defective, then all he knows is that another element of D_i is defective, but there are more than one such elements. Any other element x appears in a query Q_1 that got answer NO. At that point when this NO answer arrives, x might know that the answer was YES to a larger set Q that contains x . $Q \setminus Q_1$ has size at least 3, thus x does not know at this point which one is defective. If $x \notin B \cup C$, then he does not appear in any queries later, thus cannot find any defectives. If $x = b_i$ or $x = c_i$ for some i , he can get additional information about only one element of $Q \setminus Q_1$, thus there are two candidates remaining, again, x cannot find any defectives. Finally, if the answer to $\{a_i, b_i, c_i\}$ is YES, then b_i does not know if a_i or c_i is defective. \square

It is easy to see that Model 4_d still cannot be solved if $i \geq d$ or $j < d$. Indeed, the defectives still get only YES answers, thus less than d of them cannot have any idea about the remaining defectives. On the other hand we will show that there are possible answers such that the defectives together will find out they are the defectives, showing $i \geq d$ is impossible. Let us assume that every answer is YES, unless it is impossible. If Questioner finds out that D is the set of defectives, it means that for every other set D' of size d there was a query at some point that intersected exactly one of D and D' . At that point YES was a possible answer, thus the answer to that query was YES. Hence it intersected D and was disjoint from D' . Then an element of D knows D' is not the set of defectives, and this holds for every set $D' \neq D$ of size d .

6 Remarks

We finish this article with some possible directions that can be investigated:

- In some of the above models we proved that there is a family that solves the model, but did not say anything about its possible size.
- In case of Model 4_d our results can only be considered as the starting point of the investigations. In particular, it would be interesting to see if i can go above 1. It is tempting to try to extend the proof of Theorem 3.6 to this case, and use a linear hypergraph of large girth. However, it does not work even for $i = 2$. The property that the defectives can be identified forces the elements to be contained in many hyperedges, while the property that no 2 elements can identify any of the defectives forces the opposite.

If the query hypergraph is linear and d -separating, it is easy to see that for two elements contained by the same query there must be at least d other sets containing the two elements. This implies almost every element has to be contained in more than $(d + 1)/2$ queries.

On the other hand let us consider two elements x, y that are not contained in the same hyperedge. The large girth of the query hypergraph implies that there is at most one other element z contained in a hyperedge Q together with x and another hyperedge together with y (if there is no such z , then let Q be an arbitrary query containing x). Let us assume the answer to Q is NO, and the answer to every other query containing x or y is YES. Then x and y together know that x is not defective. If they cannot identify y as a defective, there cannot be more than d hyperedges containing x or y besides Q . This implies almost every element has to be contained in at most $(d + 1)/2$ queries.

- In [18] the authors considered the abstract version of the model introduced by Tapolcai et al. [29, 30]. Here we extended our models to the case of more defectives. It would be interesting to see if their model can be extended similarly.

- It is a phenomenon in combinatorial group testing that in most of the models the adaptive version actually means two round version of the problem (see e.g. [8]) Recently there was some interest in the r round (or multi-stage) versions of combinatorial group testing problems, where this phenomenon does not hold (see e.g. [7, 19]). It would be interesting to investigate these models in this context.

- One can consider a variant of these models, where instead of requiring that the elements find all (or none) of the defective elements, we require that they identify at least i and/or at most j of them.

ORCID iDs

Dániel Gerbner  <https://orcid.org/0000-0001-7080-2883>

Máté Vizer  <https://orcid.org/0000-0002-2360-3918>

References

- [1] M. J. Atallah, K. B. Frikken, M. Blanton and Y. Cho, Private combinatorial group testing, in: *Proceedings of the 2008 ACM symposium on Information, computer and communications security*, 2008 pp. 312–320, doi:10.1145/1368310.1368355.
- [2] A. Beimel, Secret-sharing schemes: a survey, in: *Coding and cryptography*, Springer, Heidelberg, volume 6639 of *Lecture Notes in Comput. Sci.*, pp. 11–46, 2011, doi:10.1007/978-3-642-20901-7_2.
- [3] F. S. Benevides, D. Gerbner, C. T. Palmer and D. K. Vu, Identifying defective sets using queries of small size, *Discrete Math.* **341** (2018), 143–150, doi:10.1016/j.disc.2017.08.023.
- [4] C. Berge, *Hypergraphs*, volume 45 of *North-Holland Mathematical Library*, North-Holland Publishing Co., Amsterdam, 1989, combinatorics of finite sets, Translated from the French, <https://books.google.com/books?id=jEyfse-EKf8C>.
- [5] A. Cohen, A. Cohen and O. Gurewitz, Secure group testing, in: *2016 IEEE International Symposium on Information Theory (ISIT)*, IEEE, 2016 pp. 1391–1395, <https://arxiv.org/abs/1607.04849>.

- [6] C. Cooper, A. Frieze, M. Molloy and B. Reed, Perfect matchings in random r -regular, s -uniform hypergraphs, *Combin. Probab. Comput.* **5** (1996), 1–14, doi:10.1017/s0963548300001796.
- [7] P. Damaschke, A. S. Muhammad and E. Triesch, Two new perspectives on multi-stage group testing, *Algorithmica* **67** (2013), 324–354, doi:10.1007/s00453-013-9781-4.
- [8] A. De Bonis, L. Gąsieniec and U. Vaccaro, Optimal two-stage algorithms for group testing problems, *SIAM J. Comput.* **34** (2005), 1253–1270, doi:10.1137/s0097539703428002.
- [9] T. J. Dickson, On a problem concerning separating systems of a finite set, *J. Combinatorial Theory* **7** (1969), 191–196, doi:10.1016/s0021-9800(69)80011-6.
- [10] R. Dorfman, The detection of defective members of large populations, *The Annals of Mathematical Statistics* **14** (1943), 436–440, doi:10.1214/aoms/1177731363.
- [11] D.-Z. Du and F. K. Hwang, *Pooling designs and nonadaptive group testing*, volume 18 of *Series on Applied Mathematics*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2006, doi:10.1142/9789812773463, important tools for DNA sequencing.
- [12] A. D’yachkov, P. Vilenkin, A. Macula and D. Torney, Families of finite sets in which no intersection of l sets is covered by the union of s others, *J. Combin. Theory Ser. A* **99** (2002), 195–218, doi:10.1006/jcta.2002.3257.
- [13] A. D’yachkov, P. A. Vilenkin and S. Yekhanin, Upper bounds on the rate of superimposed (s, l) -codes based on engel’s inequality, in: *Proceedings of the International Conf. on Algebraic and Combinatorial Coding Theory (ACCT)*, Citeseer, 2002 pp. 95–99.
- [14] A. G. D’yachkov and V. V. Rykov, Bounds on the length of disjunctive codes, *Problemy Peredachi Informatsii* **18** (1982), 7–13, <http://mi.mathnet.ru/eng/ppi1232>.
- [15] D. Ellis and N. Linial, On regular hypergraphs of high girth, *Electron. J. Combin.* **21** (2014), Paper 1.54, 17, doi:10.37236/3851.
- [16] D. Eppstein, M. T. Goodrich and D. S. Hirschberg, Combinatorial pair testing: distinguishing workers from slackers, in: *Algorithms and data structures*, Springer, Heidelberg, volume 8037 of *Lecture Notes in Comput. Sci.*, pp. 316–327, 2013, doi:10.1007/978-3-642-40104-6_28.
- [17] P. Frankl and Z. Füredi, Erratum: “Union-free hypergraphs and probability theory” [European J. Combin. **5** (1984), no. 2, 127–131; MR0753001 (85g:05110)], *European J. Combin.* **5** (1984), 395, doi:10.1016/s0195-6698(84)80025-6.
- [18] D. Gerbner and M. Vizer, Smart elements in combinatorial group testing problems, *J. Comb. Optim.* **35** (2018), 1042–1060, doi:10.1007/s10878-018-0248-z.
- [19] D. Gerbner and M. Vizer, Rounds in a combinatorial search problem, *Discrete Appl. Math.* **276** (2020), 60–68, doi:10.1016/j.dam.2019.11.016.
- [20] N. J. Harvey, M. Patrascu, Y. Wen, S. Yekhanin and V. W. Chan, Non-adaptive fault diagnosis for all-optical networks via combinatorial group testing on graphs, in: *IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications*, IEEE, 2007 pp. 697–705, doi:10.1109/infcom.2007.87.
- [21] F. Hwang and V. Sós, Non-adaptive hypergeometric group testing, *Studia Sci. Math. Hungar* **22** (1987), 257–263, <http://real.mtak.hu/id/eprint/107775>.
- [22] W. Kautz and R. Singleton, Nonrandom binary superimposed codes, *IEEE Transactions on Information Theory* **10** (1964), 363–377, doi:10.1109/tit.1964.1053689.
- [23] C. Lo, M. Liu, J. P. Lynch and A. C. Gilbert, Efficient sensor fault detection using combinatorial group testing, in: *2013 IEEE international conference on distributed computing in sensor systems*, IEEE, 2013 pp. 199–206, doi:10.1109/dcoss.2013.57.

- [24] A. Rényi, On random generating elements of a finite Boolean algebra, *Acta Sci. Math. (Szeged)* **22** (1961), 75–81, <http://acta.bibl.u-szeged.hu/id/eprint/13920>.
- [25] M. Ruszinkó, On the upper bound of the size of the r -cover-free families, *J. Combin. Theory Ser. A* **66** (1994), 302–310, doi:10.1016/0097-3165(94)90067-1.
- [26] M. Sobel and P. A. Groll, Group testing to eliminate efficiently all defectives in a binomial sample, *Bell System Tech. J.* **38** (1959), 1179–1252, doi:10.1002/j.1538-7305.1959.tb03914.x.
- [27] J. Spencer, Minimal completely separating systems, *J. Combinatorial Theory* **8** (1970), 446–447, doi:10.1016/s0021-9800(70)80038-2.
- [28] D. R. Stinson, R. Wei and L. Zhu, Some new bounds for cover-free families, *J. Combin. Theory Ser. A* **90** (2000), 224–234, doi:10.1006/jcta.1999.3036.
- [29] J. Tapolcai, L. Rónyai, É. Hosszu, L. Gyimóthi, P.-H. Ho and S. Subramaniam, Signaling free localization of node failures in all-optical networks, *IEEE Transactions on Communications* **64** (2016), 2527–2538, doi:10.1109/infocom.2014.6848125.
- [30] J. Tapolcai, L. Rónyai, É. Hosszu, P.-H. Ho and S. Subramaniam, Signaling free localization of node failures in all-optical networks, in: *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, IEEE, 2014 pp. 1860–1868, doi:10.1109/tcomm.2016.2545653.
- [31] L. M. Tolhuizen, New rate pairs in the zero-error capacity region of the binary multiplying channel without feedback, *IEEE Transactions on Information Theory* **46** (2000), 1043–1046, doi:10.1109/18.841182.