Conscious and controlling elements in combinatorial group testing problems

Dániel Gerbner*† Máté Vizer‡

MTA Rényi Institute
Hungary H-1053, Budapest, Reáltanoda utca 13-15.
gerbner@renyi.hu, vizermate@gmail.com

March 17, 2017

Abstract

In combinatorial group testing problems Questioner needs to find a special element $x \in [n]$ by testing subsets of $[n]$. Tapolcai et al. [24, 25] 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 special one. Using classical results of extremal set theory we prove that if $\mathcal{F}_n \subset 2^{[n]}$ solves the non-adaptive version of this problem and has minimal cardinality, then

$$\lim_{n\to\infty} \frac{|\mathcal{F}_n|}{\log_2 n} = \log_{(3/2)} 2.$$

This improves results in [24, 25].

We also consider related models inspired by secret sharing models, where the elements should share information among them to find out the special one. Finally the adaptive versions of the different models are investigated.

*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.

1 Introduction

In the most basic *model* of *combinatorial group testing* Questioner needs to find a special element x of $\{1, 2, ..., n\}(=: [n])$ 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*. 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 at the beginning.

Combinatorial group testing problems were first considered during the World War II by Dorfman [7] 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 [14], in quality control in product testing [22] or failure detection in wireless sensor networks [18]. 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 [8].

Here we consider a new type of combinatorial group testing model. The elements are conscious and they distrust the Questioner, thus they want to control the tests they are involved in. This gives the following extra condition: **each element knows the answer to those queries that contain it**. However, Questioner might not want to share all the information with them.

This is the generalization of the recently introduced node failure localization model by Tapolcai et al. [24, 25]. Failures in a network are checked by monitoring trails that turn into off state if interrupted by a failure event. The goal is to construct the monitoring trails such a way that any node can determine the network failure status solely by observing the on-off status of the monitoring trails traversing that node. The network is given by a graph and the monitoring trails are subgraphs satisfying certain properties.

In this paper we deal with the abstract version of the problem. Instead of a graph, we are given a set, and instead of the monitoring trails any subset can be used as a query. This way we get the same extra condition. The goal in this model is the following: each element should be able to identify the defective one.

Tapolcai et al. [24, 25] proved that $1.62088 \log_2 n$ queries are needed in the abstract setting, and gave examples of graphs where $2\log_2 n$ monitoring trails are needed. Here we improve their lower bound, and show that at least $\log_{(3/2)} 2\log_2 n \geq 1.70951 \log_2 n$ queries are needed, and this bound is asymptotically sharp in the abstract case.

We mention another motivation of our investigations: 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, 4, 9]. 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). Note that if each element knows there is exactly one defective, every query immediately shows several elements which are not defective - either the elements of the test, or the ones in the complement. As we do not use any encryption, the elements of that set gain significant information. This is why we can only require that elements cannot completely detect the defective one. They can narrow it down to two candidates, but cannot completely identify it.

Motivated by secret sharing schemes (see e.g. [2]), we also consider the following variant: 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 we do not deal with the way the data is transmitted. Information can not be distributed between different groups.

Structure. 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 introduce the investigated models. In Section 4 we provide the proofs of the statements we do not prove in Section 3, in Section 5 we investigate the adaptive scenario. We finish this article with some remarks and open problems in Section 6.

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 $\mathcal{F}^c := \{[n] \setminus F : F \in \mathcal{F}\}$, 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.

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

Definition 1. *We say that $\mathcal{F} \subset 2^{[n]}$ is:*

$\bullet_1$ ***intersection closed*** *if $F, G \in \mathcal{F}$ implies $F \cap G \in \mathcal{F}$.*

$\bullet_2$ ***Sperner*** *if there are no two different $F_1, F_2 \in \mathcal{F}$ with $F_1 \subset F_2$.*

$\bullet_3$ ***cancellative*** *if for any three $F_1, F_2, F_3 \in \mathcal{F}$ we have*

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

$\bullet_4$ ***intersection cancellative*** *if for any three $F_1, F_2, F_3 \in \mathcal{F}$ we have*

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

$\bullet_5$ ***separating*** *if for any two different $x, y \in [n]$ there is $F \in \mathcal{F}$ with*

$$x \in F \text{ and } y \notin F, \text{ or}$$

$$y \in F \text{ and } x \notin F.$$

$\bullet_6$ ***completely separating*** *if for any two different $x, y \in [n]$ there is $F \in \mathcal{F}$ with*

$$x \in F \text{ and } y \notin F.$$

$\bullet_7$ *a* ***pairwise balanced design*** *if for every two different elements $x, y \in [n]$ there is exactly one $F \in \mathcal{F}$ that contains both. If K is the set of cardinalities of the members of $\mathcal{F}$, we say $\mathcal{F}$ is a **PBD**(K). If $K = \{3\}$, we say $\mathcal{F}$ is a* ***Steiner triple system****.*

Some known results about these notions that we will use later

• The notion *cancellative* was introduced by Frankl and Füredi in [10], where they proved the following upper bound on the size of a cancellative family of subsets:

Theorem 2. *(Frankl, Füredi [10], Theorem 3)*
Suppose that $n \geq 14$ *and* $\mathcal{F} \subset 2^{[n]}$ *is cancellative. Then we have*

$$|\mathcal{F}| \leq n \cdot (\frac{3}{2})^n.$$

The following theorem was proved by Tolhuizen:

Theorem 3. *(Tolhuizen [26], Corollary 1)*
Suppose $\mathcal{F}_n \subset 2^{[n]}$ *is the largest cancellative family, then we have:*

$$\lim_{n\to\infty} \frac{1}{n} \log_2 |\mathcal{F}_n| = \log_2(\frac{3}{2}).$$

We will also use the following during the proof of our results:

Fact 4. $\mathcal{F} \subset 2^{[n]}$ *is intersection cancellative if and only if* $\mathcal{F}' := \{[n] \setminus F : F \in \mathcal{F}\}$ *is cancellative.*

• The notion of *separating family* in the context of combinatorial group testing was introduced and first studied by Rényi in [20]. We will use the following simple facts.

Fact 5. *Asking* $\mathcal{F} \subset 2^{[n]}$ *Questioner finds the defective element if and only if* $\mathcal{F}$ *is separating.*

Fact 6. *Suppose* $\mathcal{F}_n \subset 2^{[n]}$ *is a minimal separating family. Then we have:*

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

• The notion of *completely separating* family was introduced by Dickson in [6], where he determined the order of the smallest completely separating family. Later Spencer observed the following:

Theorem 7. *(Spencer, [23]) For* $\mathcal{F} \subset 2^{[n]}$ $(n \geq 1)$ *is completely separating if and only if its dual is Sperner. Thus for any* $n \geq 1$ *there exists a completely separating family* $\mathcal{F}_n \subset 2^{[n]}$ *with:*

$$|\mathcal{F}_n| \leq \lceil \log_2 n + \frac{1}{2} \log_2 \log_2 n \rceil.$$

• The notion of *Steiner triple systems* was introduced in the middle of the 19th century and has since developed into the huge area of combinatorial designs. Here we will use two of the most fundamental results. A subfamily of pairwise disjoint sets is a *partial matching*, and it is a *matching* if it covers all the elements. They are also called parallel classes in design theory.

Theorem 8. *(Kirkman [16], Bose [3], Skolem [21]) There exists a Steiner triple system on* $[n]$ *if and only if* $n = 6k+1$ *or* $n = 6k+3$ *for some integer* k*.*

Theorem 9. *(Ray-Chaudhuri, Wilson [19]) If* $n = 6k+3$*, then there exists a Steiner triple system that can be decomposed into* $3k+1$ *complete matchings.*

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 one of them ($d \in [n]$) is defective. We are dealing with non-adaptive models, so Questioner needs to construct a family $\mathcal{F} \subset 2^{[n]}$ at the beginning. A set $F \subset [n]$ correspond to a query of the following type: 'is the defective d an element of F?', and every element of F finds out the answer in addition to the Questioner. In each model we assume that knowing all the answers is enough information for Questioner to find the defective element, i.e. $\mathcal{F}$ is separating.

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 element, or oppositely, that they cannot find out something.

When we say that an element x *knows the defective element*, we mean that the query family satisfies the following property: no matter what the defective element is, after the answers x can find the defective one, i.e. the subfamily $\mathcal{F}_x$ is separating. In the opposite, when we say that x *does not know the defective*, we mean that the query family satisfies the following property: no matter what the defective element y is, after the answers x does not know that y is the defective, i.e. for any $y \in [n]$ there is a different $z \in [n]$ that is contained in exactly the same members of $\mathcal{F}_x$ as y.

Note that the above two cases do not cover all the possibilities. For example if x is contained only in the sets $\{x, y\}$ and $\{x, z\}$ and $n \geq 5$, then we can neither say x does not know the defective nor can say x knows the defective, since if the defective is x, y or z, then x knows, otherwise does not know the defective.

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 figure the defective element. 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 one of them is defective. 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 family that solves that specific model or state results about the cardinality of such query families. First we consider the models where we require the elements to find out something about the defective (the model by Tapolcai et al. [24, 25] that initiated this research is of this type). Then we consider the models where we require some information to remain hidden from the elements. Finally we mix these types of properties in Model 4.

3.1 Model 1

Perhaps the most natural model is the following:

Property: all elements find out (each about itself) if they are defective.

It is easy to see that this property is equivalent to the following: for every two different $x, y \in [n]$ there is a set $F \in \mathcal{F}$ such that $x \in F$, $y \notin F$, i.e. $\mathcal{F}$ is completely separating. By Theorem 7 we immediately have:

Proposition 10. *For any $n \geq 1$ there is $\mathcal{F}_n \subset 2^{[n]}$ that solves Model 1 with:*

$$|\mathcal{F}_n| \leq \lceil \log_2 n + \frac{1}{2} \log_2 \log_2 n \rceil.$$

3.2 Model 2

This model is the abstract version of the node failure localization model introduced by Tapolcai et al. [24, 25].

Property: all elements know the defective.

Lengert [17] proved that there is $\mathcal{F}_n \subset 2^{[n]}$ that solves Model 2 with $|\mathcal{F}_n| \leq 3\log_3 n$ (that is a better upper bound than the ones in [24, 25]. However we note again the latter results are about non-abstract cases.). In the following (Corollary 12) we prove an asymptotically sharp result on the minimal cardinality of the solutions of Model 3. Before that we characterize the query families that solve Model 3.

Theorem 11. *$\mathcal{F}_n \subset 2^{[n]}$ solves Model 2 if and only if its dual is Sperner and intersection-cancellative.*

With the help of the previous theorem we can prove the following:

Corollary 12. *Suppose $\mathcal{F}_n \subset 2^n$ solves Model 2 and has minimal cardinality. Then we have*

$$\lim_{n\to\infty} \frac{|\mathcal{F}_n|}{\log_2 n} = \log_{(3/2)} 2 \ (\approx 1.70951).$$

This result provides an improvement of the results of Theorem 1 of [24] and [25].

3.3 Model 3

In this model Questioner wants to find the defective such a way that its identity is hidden from the participants themselves.

Property: no element knows the defective.

Proposition 13. *No $\mathcal{F}$ can solve Model 3.*

Proof. Recall that we always assume that Questioner can find the defective, i.e. $\mathcal{F}$ is separating. Let us consider the families $\mathcal{F}_x$ ($x \in [n]$) and choose an element x such that $\mathcal{F}_x$ is inclusion-wise maximal among these families. We claim that if x is the defective, then it knows that. Indeed, x gets only YES answers. Suppose by contradiction that y could also be the defective according to x, then we would have $\mathcal{F}_y \supseteq \mathcal{F}_x$, which implies $\mathcal{F}_y = \mathcal{F}_x$. However it is impossible, as $\mathcal{F}$ is separating.

$\square$

3.4 Model 3'

As Model 3 is impossible to solve, in the next model the defective himself may find out he is the defective, but nobody else (note that we assume that knowing all the answers is enough to find the defective).

Property: no element knows the defective, except for the defective one.

Opposed to Model 3, this is easily achievable: we can ask all (or all but one) of the singletons. So a natural question that arises here is the cardinality of the smallest family that can solve Model 3'. In the next theorem we give an upper bound on this quantity.

Theorem 14. *For every $n \geq 1$ there is $\mathcal{F}_n \subset 2^n$ that solves Model 3' with*

$$|\mathcal{F}_n| \leq 3\lceil \log_3 n \rceil - t(n),$$

where $t(n)$ is the number of zeros in n written in ternary base.

3.5 Model 4

Now we start to investigate models where elements can share information among them. When we say that a group of elements together know the defective element, we mean that all of them in the group know the answers for the queries that contained one of them and using this information they can find the defective one. (Recall that information can not be distributed between different groups.) Let i and j be integers with $1 \leq i < j \leq n$.

Property: any j elements together know the defective, but i elements together do not know, unless one of them is the defective itself.

Note that $i = 0$ is another possibility. In that case the solution would be a family where any j elements together can find the defective. However, in this section we only deal with the existence of a solution, and a solution for Model 2 is obviously a solution for this model as well.

Let us continue with two simple observations. As long as we only consider the existence of a solution, we can assume the solution $\mathcal{F}$ is intersection-closed, as if $F, G \in \mathcal{F}$, then elements of $F \cap G$ know the answer to $F \cap G$ anyway. Another observation is that the family of singletons solves this model if $j \geq n-1$. Indeed, no set of elements has any information about the other ones, hence they know the defective if and only if he is one of them, or the only element not in the set. We show that if $i \geq 2$, then this is the only case, when Model 4 can be solved.

Theorem 15. *If $i \geq 2$ and $j \leq n-2$, then there is no solution for Model 4.*

The only remaining case is $i = 1$. Surprisingly, the solution here depends on divisibility conditions. First we deal with the $j = 2$ case. In the following two theorems we prove that a kind of minimal structure should be contained in any solution in this case.

Theorem 16. *If $n \geq 4,\ i = 1$ and $j = 2$, a Steiner triple system minus a partial matching solves Model 4.*

Theorem 17. *Let $i = 1$ and $j = 2$. If $\mathcal{F}$ is intersection-closed and solves Model 4, then it contains a Steiner triple system on n elements minus a partial matching.*

Note that if $i = 1$ and $j = 2$, then there is a solution for $n = 1$ and $n = 3$ and there is no solution for $n = 2$. So by the previous two theorems and Theorem 8 we have:

Corollary 18. *Let $i = 1$ and $j = 2$. There is a solution for Model 4 if and only if $n = 6k+1$ or $n = 6k+3$.*

Now we continue with larger j's.

Theorem 19. *Let $i = 1$. Then we have:*

a) if $j \geq 4$ and $n \neq 6$, then there is a solution for Model 4.

b) if $j = 3$, $n \neq 6$, $n \neq 6k+2$ and $n \neq 6k+5$ for some integer k, then there is a solution for Model 4.

The only remaining cases are $i = 1$, $j = 3$, $n = 6k+2$ or $6k+5$. In every other case we completely characterized the values of n where a solution for Model 4 exists. For our knowledge in the remaining cases see the Remark section.

4 Proofs

We use the expression x can *distinguish* y and z meaning that if one of y and z is the defective, x can tell which one it is, using only the information available to x, i.e. the answers to the queries containing x. Equivalently, there is a query that contains x and exactly one of y and z.

4.1 Proof of Theorem 11

We start the proof with the following easy lemma that gives some characterization of the query system that solves Model 2:

Lemma 20. *$\mathcal{F} \subset 2^{[n]}$ solves Model 2 if and only if the following two properties hold:*

- $\bullet_1$ *$\mathcal{F}$ is completely separating, and*
- $\bullet_2$ *for all pairwise different $a, b, c \in [n]$ there is $F \in \mathcal{F}$ with $a, b \in F$ and $c \notin F$ or with $a, c \in F$ and $b \notin F$.*

Proof of Lemma 20. We prove by contradiction.

First suppose that $\bullet_1$ is not true. So there are two different elements $a, b \in [n]$ such that for all $F \in \mathcal{F}$ if $a \in F$, then $b \in F$. In this case Adversary answers YES for all queries that contain a and a won't be able to distinguish a and b and decide whether a or b is the defective.

If $\bullet_2$ is not true, then there are three different $a, b, c \in [n]$ such that for all $F \in \mathcal{F}$ if $a, b \in F$, then $c \in F$ and if $a, c \in F$, then $b \in F$. If Adversary answers YES for all queries that contain a, b and c, then a won't be able to decide whether b or c is the defective.

To prove the other direction first observe that by $\bullet_1$ only the defective element gets YES answer for all the queries containing it. Thus any other element knows that he is not a defective (getting at least one NO answer (for a query containing it)). However by $\bullet_2$ they can decide who is the defective (as any of them should just consider the intersection of all the queries that were answered YES and contained it and the other element in the intersection is just the defective).

$\square$

Now we translate the properties of $\mathcal{F}$ given in Lemma 20 for the properties of the dual of $\mathcal{F}$.

Lemma 21. *$\mathcal{F} \subset 2^{[n]}$ satisfies properties $\bullet_1$ and $\bullet_2$ if and only if its dual is Sperner and intersection cancellative.*

Proof. The fact that the dual of a completely separating system (property $\bullet_1$ of Lemma 20) is Sperner was proved in [23] (as we mentioned it earlier in Theorem 7).

So it is enough to prove that the dual of a family with property $\bullet_2$ of Lemma 20 is cancellative. Property $\bullet_2$ means that for any three different sets A, B, C in the dual there is an element f (corresponding to F) such that either $f \in A$, $f \in B$ and $f \notin C$ or $f \in A$, $f \in C$ and $f \notin B$. This means either $f \in A \cap B \setminus C$ or $f \in A \cap C \setminus B$. The existence of f means either $A \cap B \not\subset C$ or $A \cap C \not\subset B$. Let us define three properties.

$\circ_1$ $A \cap B \not\subset C$.

$\circ_2$ $A \cap C \not\subset B$.

$\circ_3$ $C \cap B \not\subset A$.

Property $\bullet_2$ (for these three sets in this order) means that at least one of $\circ_1$ and $\circ_2$ holds. Considering the same three sets in different orders we get that also at least one of $\circ_1$ and $\circ_3$ and one of $\circ_3$ and $\circ_2$ holds. It is true if and only if at least two of these three properties hold.

To finish the proof of Lemma 20 we prove the following:

Claim 22. *$\mathcal{F}' \subset 2^{[n]}$ is intersection cancellative if and only if at least two out of $\circ_1, \circ_2$ and $\circ_3$ hold for any three members of it.*

Proof. Let us assume $\mathcal{F}'$ is intersection cancellative and let $A, B, C \in \mathcal{F}'$. Let us assume at most one, say $\circ_3$ of the three properties holds, thus $\circ_1$ and $\circ_2$ do not hold. The first one implies $A \cap B \subset C$, and obviously $A \cap B \subset A$. Thus we have $A \cap B \subset A \cap C$. Similarly the second one implies $A \cap C \subset A \cap B$, hence they together imply $A \cap C = A \cap B$, which contradicts the intersection cancellative property and our assumption that A, B, C are three different sets.

Let us assume now that $\mathcal{F}'$ is not intersection cancellative, thus we have $A \cap B = A \cap C$. This implies both $A \cap B \subset C$ and $A \cap C \subset B$, thus at most one of $\circ_1, \circ_2$ and $\circ_3$ can hold.

$\square$

We are done with the proof of Lemma 21.

$\square$

By Lemma 20 and Lemma 21 we are done with the proof of Theorem 11.

$\square$

4.2 Proof of Corollary 12

First note that by Theorem 11 and Fact 4 we have that $\mathcal{F}_n \subset 2^{[n]}$ solves Model 2 if and only if the complement of its dual is Sperner and cancellative. So now:

- upper bound:

The fact that

$$\limsup_{n\to\infty} \frac{|\mathcal{F}_n|}{\log_2 n} \leq \log_{(3/2)} 2$$

is a consequence of Theorem 2. Here we do not use that $\mathcal{F}_n$ is also Sperner.

- lower bound:

Theorem 3 gives a large (not necessarily Sperner) cancellative family. However, a more careful analysis of Tolhuizen's proof [26] shows that the family given there is Sperner. We just give a sketch here as it introduces a lot of new definitions.

A set $X \subset [n]$ is an *identifying set* for a family $\mathcal{G} \subset 2^{[n]}$ if for any members $G, G' \in \mathcal{G}$ there exists $x \in X$ such that either $x \in G \setminus G'$ or $x \in G' \setminus G$. Tolhuizen proves that for any family $\mathcal{G}$ the family of sets that are both members of $\mathcal{G}$ and identifying sets for $\mathcal{G}$ is intersection cancellative. To get a large intersection cancellative family he uses codes and constructs a family $\mathcal{G}$ that contains many sets that are also identifying sets for $\mathcal{G}$. Observe that if $A \subset B$ with $A, B \in \mathcal{G}$, then A cannot be identifying set, as elements of it cannot be in $A \setminus B$ nor in $B \setminus A$. This implies the resulting intersection cancellative family is also Sperner.

Thus we have

$$\liminf_{n\to\infty} \frac{|\mathcal{F}_n|}{\log_2 n} \geq \log_{(3/2)} 2.$$

We saw that Tolhuizen's construction is Sperner, however we note that even starting from a large cancellative family that is not Sperner, we could consider the largest subfamily of it that consists of sets of the same size. The resulting Sperner family would still be large enough to give the same asymptotic result.

$\square$

4.3 Proof of Theorem 14

We construct $\mathcal{F}_n$ recursively. If $n \leq 8$, then it is easy to check that there is $\mathcal{F}_n$ that solves Model 3' and $|\mathcal{F}_n| \leq 3\lceil \log_3 n \rceil - t(n)$.

Let us assume $n \geq 9$ and consider a family $\mathcal{F}$ that solves Model 3' on $\lfloor n/3 \rfloor$ elements. Let us replace each element x by a set A_x of three or four new elements to get n element altogether. For every set in $F \in \mathcal{F}$ let $A_F = \cup_{x\in F} A_x$. Let us also consider three disjoint sets B_1, B_2, B_3 such that $|A_x \cap B_i| = 1$ for every $x \in [\lfloor n/3 \rfloor]$ and $i = 1, 2, 3$. Let $\mathcal{A} = \{A_F : F \in \mathcal{F}\} \cup \{B_1, B_2, B_3\}$ and $\mathcal{A}_0 = \{A_x : x \in [\lfloor n/3 \rfloor]\} \cup \{B_1, B_2\}$.

Claim 23. *$\mathcal{A}$ solves Model 3' if $3 \nmid n$ and $\mathcal{A}_0$ solves Model 3' if $3 \mid n$.*

Proof of Claim 23. First we prove that both $\mathcal{A}$ and $\mathcal{A}_0$ satisfy the property of Model 3'. Indeed, let $y \in [n]$. Let us first forget about the queries B_1, B_2 (and B_3) and consider the remaining queries. By the construction of the remaining queries if (from that information): y can find out which one of the sets A_x contains the defective, then A_x contains the defective and $y \in A_x$. However in this case y (if y is not the defective) cannot distinguish the other elements of A_x, even using the answer for the B_i that contains it.

On the other hand if y can not decide (again, without the B_i's) which A_x contains the defective, then there are at least two sets A_x, A_z such that he cannot tell which one contains the defective element. Then without the sets B_i he cannot distinguish them at all, thus all the (at least) 6 elements of A_x and A_z should be considered as possible defective by y. However there is at most one B_i that y can use, and it intersects these (at least) two sets in (at least) two elements. Thus y cannot distinguish these (at least) two elements from each other, nor the other at least 4 elements from each other.

Finally we prove that both $\mathcal{A}$ and $\mathcal{A}_0$ are separating: if two elements are in different A_x, they are separated by the queries A_F. If they are in the same A_x, they are separated by B_1, B_2, B_3, or if $|A_x| = 3$, then by B_1, B_2. We are done with the proof of Claim 23 as if n is divisible by three, then every A_x has size 3.

□

By Claim 23 we are done with the proof of Theorem 14 as during this process we have a number divisible by three every time there is a 0 in the ternary form of n. □

Proofs about Model 4

Let us start with an easy observation. If $\mathcal{F}$ is a solution for some i and j, then it is a solution for i' and j' with $i' \leq i$ and $j' \geq j$.

We will give several constructions that share some basic properties. All the families are linear, meaning that any two query sets intersect in at most one element. There are no two-element query sets. Then an element x can find the defective element only if there are exactly $n-1$ elements contained in sets in $\mathcal{F}_x$. On the other hand, usually a straightforward case analysis shows that any two (or three, or four) elements together find the defective element, thus in some cases we omit the details.

4.4 Proof of Theorem 15

Let us assume indirectly that $\mathcal{F}$ is a solution. As we remarked earlier we can assume $\mathcal{F}$ is intersection-closed. Let us remove the singletons from $\mathcal{F}$ and let $\mathcal{F}'$ be the resulting family. We claim that $\mathcal{F}'$ is also intersection-closed. Indeed, if $F, G \in \mathcal{F}'$ and $|F \cap G| \geq 2$, then their intersection is in $\mathcal{F}$. On the other side if the intersection would be $\{x\}$, then let $y \in F \setminus \{x\}$, $z \in G \setminus \{x\}$. If x is the defective, y and z together finds that out, which is impossible since $i \geq 2$. Thus $|F \cap G| > 1$, hence it is in $\mathcal{F}'$.

For an element $x \in [n]$ let $F_x := \cap_{x \in F \in \mathcal{F}'} F$, the intersection of the sets in $\mathcal{F}'$ that contain x. We have $F_x \in \mathcal{F}'$. Let F_y be inclusion-wise minimal in $\{F_x : x \in [n]\}$. It has size larger than 1, thus it contains an element $z \neq y$, and we have $F_z \subseteq F_y$ by the definition of F_z. Thus we have $F_y = F_z$, which means that $\mathcal{F}'$ does not separate y and z, meaning that they are only separated by singletons (of $\mathcal{F}$). But then all the other elements ($= [n] \setminus \{y, z\}$) together cannot find which one of y or z is the defective, which is a contradiction as $n \geq 3$ and $j \leq n - 2$.

$\square$

4.5 Proof of Theorem 16

First we show that a Steiner triple system is a solution. Indeed, for any element a, if d is the defective with $a \neq d$, then a gets YES answer to the only query F containing both a and d. It contains a third element b, and a does not know if b or d is the defective as - using that the query family is a Steiner system - a has no more information about b.

On the other hand, let $a' \in [n] \setminus \{a, d\}$. There are two cases.

Case 1 : if $a' = b$.

By $n > 3$ there is another query containing a', the answer to that is NO, thus a' knows a' is not defective, similarly a knows about himself that he is not defective, but they both know the defective is in F and so they together can find out it is d.

Case 2: if $a' \neq b$.

Then there is a query F' containing both a' and d, thus a and a' together know the defective is in $F \cap F' = \{d\}$.

Let us finish the proof by noting that leaving out a partial matching does not change the information available to the elements. Theorem 8 implies $n = 6k + 1$ or $n = 6k + 3$ and we have assumed $n \geq 4$, thus we have $n \geq 7$, which means there are at least three queries containing a given element. It is easy to see that if $\{a, b, c\}$ is missing, a knows what the answer to that would be: If a gets exactly one YES answer to the other queries, then it is NO, otherwise it is YES. Indeed, a gets zero YES answer if b or c is the defective, only YES answers (thus at least two of those) if a is the defective, and one YES answer otherwise (for the query that contains a and the defective d).

$\square$

Now we prove that a Steiner triple system minus a partial matching is a minimal query family in this case, supposing that the query family is intersection-closed.

4.6 Proof of Theorem 17

For $a \in [n]$ let S_a be the set of elements that can be defective according to a after getting the answers, and let $S'_a := S_a \setminus \{d\}$, where d is the defective. Note that a knows S_a, but does not know S'_a. The property $i = 1$ implies $|S_a| \geq 2$ and the property $j = 2$ implies $S_a \cap S_b = \{d\}$ if $a \neq b$, $a, b \neq d$. Thus the sets S'_a ($a \in [n]$, $a \neq d$) are pairwise disjoint, non-empty sets on an underlying set of size $n-1$. Hence they are singletons as there are $n-1$ of them. This means that for any a there is exactly one element that he cannot distinguish from the defective.

Let us now consider $\mathcal{F}$. For any a, if $d \in [n] \setminus \{a\}$ is considered as the defective, then there is an element $c(a,d) \in [n] \setminus \{a, d\}$ such that a can not distinguish between d and $c(a,d)$. By the remarks above we know that there is exactly one such $c(a,d)$. If there are members of $\mathcal{F}_a$ that contain both d and $c(a.d)$, then using again the remarks in the previous paragraph, we have that the intersection of them is $\{a, d, c(a,d)\}$, thus it is in $\mathcal{F}$, as $\mathcal{F}$ is intersection closed. If there is no such member of $\mathcal{F}_a$, let us add $\{a, d, c(a,d)\}$ to $\mathcal{F}$. Let

$$\mathcal{F}' := \mathcal{F} \cup \{\{a, d, c(a,d)\} : a \in [n],\ d \in [n] \setminus \{a\},\ \{a, d, c(a,d)\} \notin \mathcal{F}\}.$$

First note that it is impossible that $\{a, d_1, c(a, d_1)\}, \{a, d_2, c(a, d_2)\} \notin \mathcal{F}$ with 4 different elements $d_1, d_2, c(a, d_1), c(a, d_2)$ as otherwise a could not distinguish between these elements, which would be a contradiction by the first paragraph of this proof.

Note also that if we add $\{a, b, c\}$ this way because a cannot distinguish b and c, then also b cannot distinguish a and c and c cannot distinguish a and b. Indeed, let us assume b can distinguish a and c, i.e. there is a set $F \in \mathcal{F}$ that contains b and c, but does not contain a. There is an element a' such that b cannot distinguish c and a', and thus $\{b, c, a'\} \subseteq F$. Moreover, $\{b, c, a'\} \in \mathcal{F}$ as it is the intersection of the sets in $\mathcal{F}$ containing both b and c. But this means a' cannot distinguish b and c, similarly to a, thus they together cannot either, a contradiction. This thought also shows that two sets from $\mathcal{F}' \setminus \mathcal{F}$ can not intersect in two elements. Altogether with the previuos paragraph we have that $\mathcal{F}' \setminus \mathcal{F}$ form a partial matching.

Let $\mathcal{F}_3 := \{F \in \mathcal{F}' : |F| = 3\}$. We claim that $\mathcal{F}_3$ is a Steiner triple system. For any two elements a, b there is a set in $\mathcal{F}_3$ that contains both as there is an element c such that

a cannot distinguish b and c; by the above either $\{a, b, c\} \in \mathcal{F}$ because $\mathcal{F}$ is closed under intersection, or $\{a, b, c\}$ was added to $\mathcal{F}$. Moreover, there is exactly one such element c, thus exactly one such set. $\square$

4.7 Proof of Theorem 19

First we note that a PBD-($\{3, 4\}$) solves Model 4 with $i = 1$, $j = 3$ and a PBD-($\{3, 4, 5\}$) solves Model 4 with $i = 1$ and $j = 4$. The proof of this statement goes similarly to the proof of Theorem 16, thus we provide only a sketch here. For any two elements there is a query containing them, and the other elements of that query cannot distinguish the first two. However, any other element can.

The sets of integers n such that there exists such pairwise balanced designs on n elements have been determined by Gronau, Mullin and Pietsch [13]. They showed that if $n = 3k$ or $n = 3k + 1$ with $n \neq 1, 6$, then there exists a PBD-($\{3, 4\}$). This proves b). They also showed that if $n \neq 1, 2, 6, 8$, then there exists a PBD-($\{3, 4, 5\}$). This proves a) except for the case $n = 8$. In that case consider the sets $\{1, 2, 3, 4\}, \{1, 5, 7\}, \{2, 5, 8\}, \{3, 6, 8\}, \{4, 6, 7\}$. One can easily check that these sets solve Model 4.

$\square$

5 Adaptive scenario

A natural idea is to consider the adaptive versions of these problems. However, the definition of these models are not straightforward. Earlier we assumed the existence of a Questioner only for notational convenience, the elements could come up with the query family in advance. However, in this case it is not clear which one of them should find out the next query in an adaptive algorithm, as they have different information available to them. Here we assume that there is a Questioner who knows all the answers and chooses the next query.

In this case he can ask queries to find the defective, and then share this information with the elements. In particular this gives an algorithm of length $\lceil \log_2 n \rceil + 2$ for Model 1 and an algorithm of length $\lceil \log_2 n \rceil + 2$ for Model 2. In Model 1, after a separating family is asked, then Questioner asks the defective $[n] \setminus \{d\}$ and $\{d\}$, if needed. To solve Model

2, it also is enough to ask the query $[n] \setminus \{d\}$ and $\{d\}$ after a separating family.

It is easy to see that Model 3 still cannot be solved. Indeed, let us assume that every answer is YES (unless it is impossible, but there is no point to ask such queries). If Questioner finds out that x is the defective, then it is separated from every other element y by a query. The answer to that query was YES, thus it contains x, and so x knows y is not defective for every $y \neq x$.

Model 3' can also be solved using $\lceil \log_2 n \rceil + 1$ queries. Questioner starts with the usual halving procedure: first asks a set F of size $\lceil n/2 \rceil$, and then depending on the answer continues recursively with F or $\overline{F}$ as the base set. Then stops when arrives to a set of size less than 6, and asks all but one of the singletons.

So far there was no difference between the adaptive and non-adaptive versions of the models when considering the existence of a solution. However the situation radically changes with Model 4.

Theorem 24. *Let $i = 1$. Model 4 can be solved adaptively if and only if $2 \leq j \leq n$ and n is odd, or $3 \leq j \leq n$ and n is even.*

Proof. Let Questioner start with asking the singletons to find the defective element d. If n is odd, he partitions the remaining elements into pairs and asks them together with d. Then every element $y \neq d$ knows that the defective is either its pair y' or d. On the other hand y and z together know it is d, as $y' = z'$ cannot happen unless $y = z$. If n is even, one of the parts should contain three of the remaining elements a, b, c. Then for example a knows the defective is d, b or c, and a and b together cannot find the defective, but any three elements can.

Let us now assume $j = 2$ and n is even. Let us assume every answer is NO, except if that would lead to a contradiction (note that it still makes sense for the Questioner to ask such queries, to help the elements find the defective, as we just saw in the algorithm described above). We claim that in this case there is no solution.

We repeat the beginning of the proof of Theorem 17. After the algorithm ends, let S_a be the set of elements that can be defective according to a, and let $S'_a = S_a \setminus \{d\}$, where d is the defective. We have $|S_a| \geq 2$ and $S_a \cap S_b = \{d\}$ if $a \neq b$, $a, b \neq d$. Thus the sets S'_a, $a \neq d$ are $n - 1$ pairwise disjoint, non-empty sets on an underlying set of size $n - 1$, thus they are singletons. This means that for any a there is exactly one element that he cannot distinguish from the defective.

Now let us define an auxiliary directed graph on the $n-1$ non-defective elements. Let $y \to z$ if y cannot distinguish d and z, i.e. among the sets that contain y, exactly the same sets contain d and z. By the above, every out-degree is one in this graph, thus it is the union of directed cycles. Let $y_1, \dots, y_k$ be the vertices of such a cycle C in the cyclic order. If a query contains d and y_1, it also contains y_2 by the definition of the edges. But then it also contains y_3, and so on. It means that the same queries from $\mathcal{F}_d$ contain the vertices of C. Then a vertex in C can distinguish d from other vertices of C only using queries that do not contain d. Let us assume $k \geq 3$. Then there is no query containing y_1 and y_2 and not containing d, as y_1 cannot distinguish y_2 and d. However, there must be such a query as y_2 can distinguish d and y_1 (as $y_1 \neq y_3$).

We claim that there is no cycle of length 1, showing that every cycle is of length 2, thus $n-1$ is even, finishing the proof. Indeed, a cycle of length 1 would mean that y_1 only received YES answers, thus it only appeared in queries containing d. There must be a query that separates d and y_1. Consider the first such query. By the above, it cannot contain y_1 and avoid d, hence it contains d and avoids y_1. Thus the answer to it was YES. However, it should have been NO (according to our assumption on the answers), as before that query it was a possibility that y_1 is the defective element, thus it would have lead to no contradiction.

□

Theorem 25. *If Model 4 can be solved adaptively, then* $(n-1)\binom{j-1}{i} \geq \binom{n-1}{i}$.

Proof. Let us consider again the sets S'_a (defined in the proof of Theorem 17) after the end of the algorithm. Let $\mathcal{G}$ be their family. Let $\mathcal{G}_k$ be the family of sets that can be written as the intersection of k sets in $\mathcal{G}$. Then we know that $\emptyset \notin \mathcal{G}_i$, but $\mathcal{G}_j = \{\emptyset\}$. Let us consider the family $\mathcal{G}'$ of the inclusion-wise minimal non-empty sets, that can be written as the intersection of sets in $\mathcal{G}$. The members of $\mathcal{G}'$ are pairwise disjoint, thus there are at most $n-1$ of them. On the other hand each of them can be written as the intersection of at most $j-1$ sets in $\mathcal{G}$. For every set $G \in \mathcal{G}'$ let $\mathcal{G}'_G$ be an inclusion-wise maximal subfamily of $\mathcal{G}$ such that every member of $\mathcal{G}'_G$ contains G. Then $|\mathcal{G}'_G| \leq j-1$.

Let us take i sets from $\mathcal{G}$. Their intersection is in $\mathcal{G}_i$, thus by definition it is a superset of a set $G \in \mathcal{G}'$. But this can only happen if those i sets are in $\mathcal{G}'_G$ (otherwise we could add one of those sets to $\mathcal{G}'_G$, contradicting its maximality). For any $G \in \mathcal{G}'$ there are at most $\binom{j-1}{i}$ i-element subfamilies of $\mathcal{G}'_G$, and there are $n-1$ $G \in \mathcal{G}'$. On the other hand there are $\binom{n-1}{i}$ ways to take i sets from $\mathcal{G}$.

$\square$

This theorem shows that if $i > 1$, then j should be large. On the other hand, unlike in the non-adaptive case, j can be smaller than $n-1$. Let us consider the following simple construction. Let Questioner ask the singletons first. He finds the defective and then partitions the other $n-1$ elements to $i+1$ sets in a balanced way, and asks all those sets. Any i elements not containing the defective get only NO answers, but there are at least $1 + \lfloor (n-1)/(i+1) \rfloor$ elements they do not know anything about. On the other hand if $j > n-1-\lceil (n-1)/(i+1) \rceil$, then j elements without the defective know all the answers to the non-singleton queries, thus they know the defective is the one not appearing in those queries.

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 Model 4 the only remaining case is $i = 1$, $j = 3$. In this case we only know that a solution exists if $n = 6k, 6k+1, 6k+3, 6k+4$. We do not know if it exists for the other values (it does not exist for some small values).

A simple way to construct a PBD-($\{3,4\}$) is the following. We take a Steiner triple system on a set X of $6k+3$ elements and its partition into $3k+1$ matchings. We take a set Y of $n-6k-3 \leq 3k+1$ additional elements and a PBD-($\{3,4\}$) on them. Finally, for every element $y \in Y$ we pick one of the matchings, and replace every set A in the matching by $A \cup \{y\}$.

Let us take a family $\mathcal{F}$ that is a solution for Model 4 with $i = 1$, $j = 3$ instead of a PBD-($\{3,4\}$) on Y. Then the resulting family is also a solution. Indeed, an element of

X and an element of Y can be distinguished by any element, two elements of X can be distinguished by any element except those two that are in a query with them, and two elements of Y can be distinguished by any elements of X (and all but two elements of Y by our assumption on $\mathcal{F}$). This argument would give a proof for Theorem 19 without using any characterization of PBDs.

Additionally, let us assume there is a solution $\mathcal{F}$ for Model 4 with $i = 1$, $j = 3$ on $6k_0 + 2$ elements. Let $k_1 \geq 2k_0 + 1$ and $n = 6k_1 + 3 + 6k_0 + 2 \geq 18k_0 + 5$ and take the above construction. Thus we get a solution for any large enough $n = 6k + 5$. Similarly if we start with a solution on $n = 6k_0 + 5$ elements (or continue with the solution found on $18k_0 + 5$ elements), we get a solution for large enough $n = 6k + 2$. Thus a solution for any of the remaining values of n would imply that for every n large enough there is a solution.

• All of the above mentioned models are also interesting in case of d defectives ($d \geq 2$). In a forthcoming paper ([11]) we started such investigations, however a lot of questions remained open.

• In this paper we considered the abstract version of the Model by Tapolcai et al. [24, 25]. It would be interesting to see if our other models or our methods work with their underlying graph structure.

• Recently there was some interest in the r round (or multi-stage) versions of combinatorial group testing problems (see e.g. [5, 12]). It would be interesting to investigate these models in this context. Note that the algorithm provided in Theorem 24 is in fact a 2-round algorithm: in the first round the singletons are asked. With those queries Questioner finds the defective, thus he knows the answer to every later queries (he uses them only to help the elements find the defective). This means whatever algorithm is used afterwards, that can be done in one round. As he gets no new information, there is no point in waiting for the answers.

Acknowledgement

We would like to thank Éva Hosszu [15], who asked us the first question of the type that was investigated in this article. We would also like to thank all participants of the Combinatorial Search Seminar at the Alfréd Rényi Institute of Mathematics for fruitful discussions.

References

[1] M. J. Atallah, K. B. Frikken, M. Blanton, Y. Cho, Private combinatorial group testing, In Proceedings of the 2008 ACM symposium on Information, computer and communications security, 312–320, (2008).

[2] A. Beimel, Secret-sharing schemes: a survey, In International Conference on Coding and Cryptology, Springer Berlin Heidelberg, pp. 11–46, (2011).

[3] R. C. Bose, On the construction of balanced incomplete block designs, Ann. Eugenics 9, 353-399, (1939).

[4] A. Cohen, A. Cohen, O. Gurewitz, Secure group testing, In IEEE International Symposium on Information Theory, IEEE, pp. 1391-1395, (2016).

[5] P. Damaschke, A. S. Muhammad, E. Triesch, Two new perspectives on multi-stage group testing, Algorithmica, 67(3), 324–354, (2013).

[6] T. J. Dickson, On a problem concerning separating systems of a finite set, Journal of Combinatorial Theory 7.3, 191–196, (1969).

[7] R. Dorfman, The detection of defective members of large populations, The Annals of Mathematical Statistics, 14(4), 436-440, (1943).

[8] D.-Z. Du and F. K. Hwang, Pooling designs and nonadaptive group testing: important tools for DNA sequencing. Vol. 18. World Scientific Publishing Company Incorporated, (2006).

[9] D. Eppstein, M. T. Goodrich, D. S. Hirschberg, Combinatorial pair testing: distinguishing workers from slackers, In Workshop on Algorithms and Data Structures, Springer Berlin Heidelberg, pp. 316-327, (2013).

[10] P. Frankl and Z. Füredi, Union-free hypergraphs and probability theory, European Journal of Combinatorics 5.2: 127–131, (1984).

[11] D. Gerbner and M. Vizer, Failure localization and information sharing in a combinatorial group testing problem with more defectives, *manuscript*

[12] D. Gerbner and M. Vizer, Rounds in a combinatorial search problem, arXiv preprint arXiv:1611.10133 (2016).

[13] H.O. Gronau, R.C Mullin, C. Pietsch, The closure of all subsets of $\{3, 4, ...10\}$ which include 3, Ars Comb., 41, (1995).

[14] 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 INFOCOM 2007. 26th IEEE International Conference on Computer Communications, 697–705, (2007)

[15] É. Hosszú, personal communication (2015).

[16] T. P. Kirkman, On a problem in combinations, Cambridge and Dublin Math. J. 2, 191–204, (1847).

[17] D. A. Lengert, Kombinatorikus keresési problémák, MSc thesis (*in Hungarian*), (2016).

[18] C. Lo, M. Liu, J. P. Lynch, and A. C. Gilbert, Efficient sensor fault detection using combinatorial group testing. In IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS) 199–206, (2013).

[19] D. K. Ray-Chaudhuri and R. M. Wilson, Solution of Kirkmans schoolgirl problem, In Proc. Symp. pure Math. Vol. 19, pp. 187–203, (1971).

[20] A. Rényi, On random generating elements of a finite Boolean algebra, Acta Sci. Math. Szeged 22.: 4, 75–81, (1961).

[21] T. Skolem. Some remarks on the triple systems of Steiner. Math. Scand. 6, 273-280, (1958).

[22] M. Sobel, and P. A. Groll, Group testing to eliminate efficiently all defectives in a binomial sample, Bell Labs Technical Journal, 38(5), 1179–1252, (1959).

[23] J. Spencer, Minimal completely separating systems, Journal of Combinatorial Theory 8.4: 446–447, (1970).

[24] J. Tapolcai, L. Rónyai, É. Hosszu, P. Ho, S. Subramaniam, Signaling Free Localization of Node Failures in All-Optical Networks, in Proc. IEEE INFOCOM, Toronto, Canada, (2014), 1860–1868.

[25] J. Tapolcai, L. Rónyai, É. Hosszu, L. Gyimóthi, P.-H. Ho, S. Subramaniam, Signaling Free Localization of Node Failures in All-Optical Networks IEEE Transactions on Communications 64(6), 2527-2538, (2016).

[26] 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(3), 1043-1046, (2000).