Tree Tribes and Lower Bounds for Switching Lemmas

Jenish C. Mehta
California Institute of Technology, Pasadena, CA 91125, USA
jenishc@gmail.com

Abstract

Let f be a Boolean function on n variables, $\boldsymbol{\rho}$ a random p-restriction that independently keeps each variable unset (or free) with probability p and otherwise uniformly sets it to 0 or 1, and $\mathrm{DT}_{\mathrm{depth}}(f)$ denote the depth of the smallest depth decision tree for f. Let $\mathrm{R}_d(f|\boldsymbol{\rho})$ be the *resilience* of f to $\boldsymbol{\rho}$ for depth d, defined as

$$\mathrm{R}_d(f|\boldsymbol{\rho}) = \Pr_{\rho \leftarrow \boldsymbol{\rho}}[\mathrm{DT}_{\mathrm{depth}}(f|\rho) \geq d].$$

If $d \gg pn$, all functions have resilience close to 0 since less than d variables would remain unset with high probability. For $d \ll pn$, most functions f on n variables have resilience close to 1, and some functions, like AND and OR, have resilience close to 0. Håstad's Switching Lemma states that for t-DNFs, the resilience $\mathrm{R}_d(f|\boldsymbol{\rho})$ is upper bounded by $(5pt)^d$, and from known upper bounds on the size of constant depth circuits computing the parity function, it follows that there exist t-DNFs whose resilience is close to the bound obtained by Håstad. However, the exact bounds for such maximally resilient DNFs or their structure is unclear, and moreover, the argument is non-constructive.

In this work, we give an explicit construction of functions called Tree Tribes parameterized by an integer t and denoted Ξ_t (on n variables), such that

$$\mathrm{R}_d(\Xi_t|\boldsymbol{\rho}) \leq (4p2^t)^d,$$

and more importantly, the resilience is also *lower* bounded by the same quantity up to constants,

$$\mathrm{R}_d(\Xi_t|\boldsymbol{\rho}) \geq (c_0 p 2^t)^d,$$

for $0 \leq p \leq c_p 2^{-t}$ and $0 \leq d \leq c_d \frac{\log n}{2^t t \log t}$ (where c_0, c_p, c_d are universal constants). As a result, for sufficiently large n and small d, this gives a hierarchy of functions with strictly increasing resilience, and covers the entire region between the two extremes where functions have resilience (close to) 0 or 1.

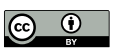
2012 ACM Subject Classification Theory of computation → Computational complexity and cryptography, Theory of computation → Oracles and decision trees

Keywords and phrases Tree Tribes, Resilience, Switching lemmas, lower bounds, decision tree

Digital Object Identifier 10.4230/LIPIcs.MFCS.2018.70

Related Version A full version of the paper is available at [5], `https://arxiv.org/abs/1703.00043`.

Funding Supported by NSF CAREER Grant CCF-1553477 and NSF Grant 1618795



43rd International Symposium on Mathematical Foundations of Computer Science (MFCS 2018).
Editors: Igor Potapov, Paul Spirakis, and James Worrell; Article No. 70; pp. 70:1–70:11

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

One useful and powerful idea to separate a Boolean function g from some set $\mathcal{F}$ of Boolean functions over $\{0,1\}^n$ is to use *restrictions*. By showing that restricted to some subset of $\{0,1\}^n$, the functions in $\mathcal{F}$ become *simple*, but the function g does not become simple, it can be concluded that $g \notin \mathcal{F}$. The extent to which functions in $\mathcal{F}$ become simple is captured by Restriction Lemmas, which informally try to answer the following question: Given a family $\mathcal{F}$ of Boolean functions characterized by some parameter t, and a family $\mathcal{S}$ of distributions over subsets of $\{0,1\}^n$ characterized by some parameter p, how complex does a function $f \in \mathcal{F}$ remain after it is restricted to a subset S chosen according to some distribution $\boldsymbol{\sigma} \in \mathcal{S}$? Defining the measure of complexity of a function and the sets $\mathcal{F}$ and $\mathcal{S}$ gives a Restriction Lemma of a particular type.

To make more concrete statements, let f be a Boolean function on n variables, $\boldsymbol{\sigma}$ a distribution over subsets of $\{0,1\}^n$, and let $f|\sigma$ denote the partial function obtained by restricting f to a subset σ chosen according to $\boldsymbol{\sigma}$. In general, we will use the boldface symbol $\boldsymbol{\sigma}$ to denote both the distribution over subsets and a subset chosen according to $\boldsymbol{\sigma}$, and the interpretation will be clear from context. Let $M : \{0,1\}^{2^n} \to [0,1]$ be some function such that $M(f)$ measures the complexity of a (total) Boolean function f. For a partial function g, $M(g)$ is defined naturally to be the minimum value of $M(f)$ for any total function f that is an extension of g. Given these definitions, we define the *resilience* of a Boolean function f to $\boldsymbol{\sigma}$ for the measure M and threshold $\gamma \in [0,1]$ as follows:

$$\mathrm{R}^M_\gamma(f|\boldsymbol{\sigma}) = \Pr_{\sigma \leftarrow \boldsymbol{\sigma}}[M(f|\sigma) \geq \gamma] = \Pr[M(f|\boldsymbol{\sigma}) \geq \gamma].$$

It is possible to study the variation in the resilience of functions restricted to various different distributions $\boldsymbol{\sigma}$ under measures of complexity M, and in this work, we study it for a particular choice of M and $\boldsymbol{\sigma}$. We will let $M(f)$ be the depth of the smallest depth decision tree for f, and since $M(f) \in [n] = \{0,1,...,n\}$, denoting the threshold specifically as d instead of γ, we will also modify the range of the threshold d and let $d \in [n]$. The restriction $\boldsymbol{\rho}$ is defined as follows: independently for every variable x_i, leave x_i unset with probability p and otherwise uniformly set it to 0 or 1. Note that $\rho \leftarrow \boldsymbol{\rho}$ chooses a p-biased subset of $\{0,1\}^n$. Under the measure M and $\boldsymbol{\rho}$ defined thus, the resilience can be written as:

$$\mathrm{R}_d(f|\boldsymbol{\rho}) = \Pr_{\rho \leftarrow \boldsymbol{\rho}}[\mathrm{DT}_{\mathrm{depth}}(f|\rho) \geq d] = \Pr[M(f|\boldsymbol{\rho}) \geq \gamma]$$

where the measure M is understood to be $\mathrm{DT}_{\mathrm{depth}}$ when omitted from the superscript.

Note that since every variable is independently left unset with probability p, there are $\approx pn$ variables that are left unset with high probability, and so if $d \gg pn$, then for any function f, $\mathrm{R}_d(f|\boldsymbol{\rho}) \approx 0$. Thus, to understand the resilience of different functions, the interesting range is $d \ll pn$, and in general, it will be worthwhile for us to intuitively understand d to be tiny, or about $O(\log \log n)$, since even the statements we make for $d = 1$ will be interesting in themselves. For one extreme, let f be the AND (or OR) function on n bits, and note that $\mathrm{DT}_{\mathrm{depth}}(f) = n$. However, since setting any bit or f to 0 (or 1) reduces it to a constant function, $\mathrm{DT}_{\mathrm{depth}}(f|\boldsymbol{\rho}) > 0$ with probability $(\frac{1+p}{2})^n \approx e^{-n} \approx 0$ (since $p \ll \frac{1}{2}$), and thus $\mathrm{R}_d(f|\boldsymbol{\rho}) \approx 0$ for all $d \geq 1$. On the other extreme, if f is the parity function, it is still the case that $\mathrm{DT}_{\mathrm{depth}}(f) = n$, but under the action of $\boldsymbol{\rho}$, there will be $\geq \frac{1}{2}pn$ variables that are unset with high probability, and $f|\boldsymbol{\rho}$ will be the parity function on the subset of variables, and if $d < \frac{1}{2}pn$, then $\mathrm{R}_d(f|\boldsymbol{\rho}) \approx 1$. In fact, since most functions require a circuit of

large size to decide them, it turns out that for most functions f on n variables, $\mathrm{R}_d(f|\boldsymbol{\rho}) \approx 1$. As a consequence of this, the property of not being reduced to a constant function when restricted to $\boldsymbol{\rho}$ satisfies the largeness condition of natural proofs [8].

This motivates the question about the type of Boolean functions f that have resilience between the two extremes of 0 and 1, and our main interest in this paper would be to understand the structure of such functions in the intermediate region. Upper bounds on $\mathrm{R}_d(f|\boldsymbol{\rho})$ are traditionally called Restriction or Switching Lemmas, which originated with the works of [2] and [1], and were proved in their strongest form by Håstad [3], who showed that if f is a DNF where each term has width t, then $\mathrm{R}_d(f|\boldsymbol{\rho})$ is upper bounded by $(5pt)^d$. Further, this bound can be used to prove tight lower bounds on the size of constant depth circuits deciding the parity function, and since there exist matching upper bounds, it implies that there exist t-DNFs whose resilience is close to $(5pt)^d$ (else it would be possible to get a better lower bound for the size of constant depth circuits deciding parity leading to a contradiction). However, the exact bounds achieved in terms of p, t and d are not known, and even the structure of such maximally resilient t-DNFs is unclear. Moreover, the argument is non-constructive.

Our result

Our main result is an *explicit* construction of functions for which we can prove tight lower and upper bounds on the resilience. The functions that we construct are called Tree Tribes, since they are similar to the Tribes function but on a tree-like structure, and are denoted by Ξ_t where t is an integer parameter. Our main theorem is the following.

▶ **Theorem 1.** *For every integer $t \geq 1$, for every n, there is an explicit function Ξ_t on n variables, such that for all p, t, and d,*

$$R_d(\Xi_t|\boldsymbol{\rho}) \leq (4p2^t)^d,$$

and for $0 \leq p \leq c_p 2^{-t}$ and $0 \leq d \leq c_d \left(\frac{\log n}{2^t t \log t}\right)$,

$$R_d(\Xi_t|\boldsymbol{\rho}) \geq (c_0 p 2^t)^d,$$

where c_p, c_d and c_0 are universal constants.

We would like to remark that the lower bound holds for d at most $\sim 2^{-t} \log n$, which is meaningful for $t \in O(\log \log n)$. However, the non-trivial part of the lower bound is that d is upper bounded by a function not only of t (which in general would be small), but of *both* t and n. Henceforth, when we consider the resilience of different functions, n will be assumed to be very large, and t and d will assumed to be small (about $O(\log \log n)$) compared to n.

An immediate corollary of Theorem 1 is the following:

▶ **Corollary 2.** *Let d and n be fixed. Let $i_1, \ldots, i_r$ be a sequence of integers such that $c_{t0} < i_1 < \ldots < i_r < c_{t1}(\log \log n - \log d)$ and $0 \leq p \leq c_p 2^{-i_r}$. Then the sequence of functions $\{\Xi_{i_j}\}_{j=1}^r$ (each on n variables) has strictly increasing resilience, i.e.*

$$R_d(\Xi_{i_j}|\boldsymbol{\rho}) < R_d(\Xi_{i_{j+1}}|\boldsymbol{\rho}),$$

where c_{t0}, c_{t1}, c_p are universal constants.

And similarly, we can get a resilience heirarchy in the following sense:

▶ **Corollary 3.** *Let d, n and r be fixed. Let t be an integer variable such that $c_{t0} \leq t \leq c_{t1}(\log\log n - \log d)/r$ and p a variable that is some function of t, i.e. $p = p(t)$ and $0 \leq p \leq c_p 2^{-rt}$. Then the sequence of functions $\{\Xi_{it}\}_{i=1}^{r}$ (each on n variables) has strictly increasing resilience, i.e.*

$$R_d(\Xi_{it}|\boldsymbol{\rho}) \in o(R_d(\Xi_{(i+1)t}|\boldsymbol{\rho}),$$

where c_{t0}, c_{t1}, c_p are universal constants.

The proof of the resilience upper bound in Theorem 1 uses a method of *conditioning on variables*, and thus avoids the complex conditioning in [3, 4], and also the combinatorial reasoning in [7, 6], and as a consequence, is *simpler* than both methods. The proof of the resilience lower bound is recursive, and proceeds by analyzing the coefficients of polynomials that arise in the analysis of Tree Tribes. Note that in Theorem 1, since we want to lower bound the probability of the event $\mathrm{DT}_{\mathrm{depth}}(\Xi_t|\boldsymbol{\rho}) \geq d$, we require that the decision tree with the least depth (and thus *every* decision tree) for $\Xi_t|\boldsymbol{\rho}$ must have depth greater than d with sufficient probability. To achieve this, our proof proceeds as follows: If T is the decision tree for Ξ_t, we lower bound the probability of finding *"paths with a split"* in $T|\boldsymbol{\rho}$. A *"path with a split"* is a subtree of $T|\boldsymbol{\rho}$, which consists of a path of distinct variables $y_1, \ldots, y_d$ (where y_1 is closest to the root in $T|\boldsymbol{\rho}$), such that y_d is connected to two leaves with *different* values (more specifically, y_d has a path to a leaf labelled 0 and a leaf labelled 1). Any decision tree for such a subtree of $T|\boldsymbol{\rho}$ must have depth at least d (at least if the variables $y_1, \ldots, y_d$ do not appear elsewhere in the tree), since the OR function is embedded in $T|\boldsymbol{\rho}$ *and* all the variables of T would be distinct in our construction. However, we need a sufficient number of vertices in the tree before we get sufficient probability mass for the event of finding such a path with a split, and this is the reason we get an upper bound on the depth d for which Theorem 1 holds.

We state the Preliminaries in Section 2, the construction of Tree Tribes in Section 3, prove the resilience upper bound in Theorem 1 in Section 4, and the resilience lower bound in Section 5. Due to space constraints, the proofs are relegated to the full version [5].

2 Preliminaries

The extended preliminaries, including the standard definitions of decision trees and random p-restrictions are given in the full version [5], and we state only the non-standard definitions here.

▶ **Definition 4** (Resilience)**.** Let f be a Boolean function on n variables, and ρ a random p-restriction. For $d \in [n]$, we let $\mathrm{R}_d(f|\boldsymbol{\rho})$ be the resilience of f to $\boldsymbol{\rho}$ at depth d, defined as

$$\mathrm{R}_d(f|\boldsymbol{\rho}) = \Pr[\mathrm{DT}_{\mathrm{depth}}(f|\rho) \geq d].$$

If T is some decision tree for some function f, we alternately write $\mathrm{R}_d(T|\boldsymbol{\rho})$ to mean $\mathrm{R}_d(f|\boldsymbol{\rho})$.

▶ **Definition 5** (Clipped decision trees, [6])**.** A decision tree T is t-clipped, if any vertex of T is at a distance of at most t from some leaf.

▶ **Definition 6** (Operators on polynomials)**.** For a univariate polynomial Q in the variable p, denoted as $Q = \sum_{i \geq 0} c_i p^i$ with a finite number of non-zero coefficients, define the operator $[p^i]$ as $[p^i]Q = c_i$ and the operator $[\uparrow p^i]$ as

$$[\uparrow p^i]Q = \sum_{j \geq i} \left([p^j]Q\right) p^{j-i}.$$

▶ **Definition 7** (Absolute maximizer)**.** Given a closed set $\mathcal{D} \subseteq [0,1]$ and some polynomial Q in $\mathbb{R}[p]$, we define the absolute maximizer of $[\uparrow p^i]Q$ by functions G_i as $G_i(Q) = \max_{p\in\mathcal{D}} \left|[\uparrow p^i]Q\right|$.

3 Tree Tribes

We now formally define Tree Tribes and their variants. A specific variant, t-clipped xor Tree Tribe, will be the function that will help us achieve the bounds in Theorem 1.

▶ **Definition 8** (Tree Tribe)**.** A Boolean function f is called a Tree Tribe, denoted by Ξ, if there is a decision tree T deciding f, such that all the variables at all the vertices of T are distinct, and from every vertex of T, there is a path to a leaf labeled 0 and a path to a leaf labeled 1.

An example for a Tree Tribe is the OR function. Given the above definition, it is possible to derive a variety of different Tree Tribes by imposing additional structure, and we define the specific structure that we'll need.

▶ **Definition 9** (Complete clipped decision trees)**.** Denote a complete t-clipped decision tree T on r levels by $W_t(r)$, and define it recursively as follows: $W_t(0)$ is a leaf. $W_t(r)$ consists of vertices $\{v_1, ..., v_t\}$, a leaf denoted by v_{t+1}, and edges $e_{i,0}$ and $e_{i,1}$ for $i \in \{1, ..., t\}$. The vertices v_1 to v_t will be said to belong to layer or level 1. Each edge $e_{i,0}$ is labelled 0 and it will be called a 0-edge, and it connects v_i and v_{i+1}. Each edge $e_{i,1}$ is labelled 1 and it will be called a 1-edge, and it connects v_i to the root of a copy of $W_t(r-1)$ on distinct variables.

▶ **Definition 10** (Clipped xor Tree Tribe)**.** A Boolean function f is called a t-clipped xor Tree Tribe, denoted by $\Xi_t(r)$ for an integer t, if $T(f)$ is a Tree Tribe, and can be expressed as a complete t-clipped decision tree, in which the leaves are labeled by the parity of the edges on the path from the root to the leaf.

Note that we define the level of some variable x in $\Xi_t(r)$ by the recursive step at which it was added to $\Xi_t(r)$. The variables at level 1 are x_1 to x_t, each of which is connected to a copy of $\neg\Xi_t(r-1)$ on distinct variables. The first t variables in each of the t copies of$\neg\Xi_t(r-1)$, a total of t^2 variables, are at level 2, and each of them is connected to a copy of $\Xi_t(r-2)$, and so on. An examples for $\Xi_2(3)$ is shown in the figure below.

Since our aim was to understand the structure of functions that are resilient towards random restrictions, we show that the functions $\Xi_t(r)$ exhibit many nice properties (given in the full version [5]). For instance, the Fourier coefficient of a subset $S \subseteq \{0,1\}^n$ depends *only* on the parameters (level in the tree and distance from root) of the variable in S that represents the vertex farthest from the root in S. These and other properties can be found in [5] along with some remarks on why they turn out to be resilient towards random restrictions.

We now proceed to show that these functions achieve the bounds stated in Theorem 1.

4 Resilience upper bound

We start by by showing the resilience upper bound in Theorem 1. Our bound will hold in general for any function f that has a t-clipped decision tree. It was shown in [6] that for any Boolean function f that has a t clipped decision tree, $\mathrm{R}_d(f|\boldsymbol{\rho}) \leq O(pt2^t)^d$, and we improve the bound to $\mathrm{R}_d(f|\boldsymbol{\rho}) \leq (4p2^t)^d$, an improvement that is necessary for us since it matches the resilience lower bound. As stated earlier, our proof uses a method of *conditioning on variables* that is different and simpler from the complex conditioning in [3, 4], and the combinatorial

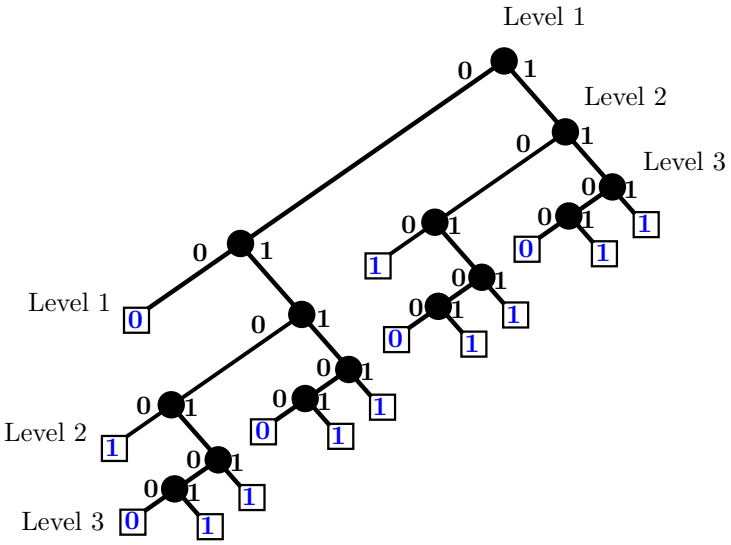


Figure 1 A 2-clipped xor tree tribe on 3 levels, or $\Xi_2(3)$. The variables at each of the vertices are distinct, and the leaves are labelled by the parity of the edges along the root to leaf path. Note that there are 2 vertices at level 1, 4 vertices at level 2, and 8 vertices at level 3.

reasoning in [7, 6], and is also amenable to extension to more general restrictions. The proofs of the lemmas in this section are given in the full version [5].

Let f be a Boolean function that has a t-clipped decision tree T.

▶ **Definition 11.** A decision tree T is (t_0, t)-clipped for $t_0 \leq t$, if the root has distance at most t_0 from some leaf, and every other vertex has distance at most t to some leaf.

▶ **Definition 12.** For positive integers t_0, t, n, d, define the probabilities $\gamma_{d,n}(t_0, t)$ as follows:

$$\gamma_{d,n}(t_0, t) = \max_{\substack{(t_0,t)\text{-clipped trees } T \text{ that} \\ \text{decide any function on } n \text{ variables}}} \mathrm{R}_d(T|\boldsymbol{\rho})$$

It is simple to see that γ is monotone in n.

▶ **Lemma 13.** *If $n' \leq n$, then $\gamma_{d,n'}(t_0, t) \leq \gamma_{d,n}(t_0, t)$.*

4.1 Recurrence for γ

We proceed by writing a recurrence for $\gamma_{d,n}(t_0, t)$.

Let T be the (t_0, t)-clipped decision tree for which $\gamma_{d,n}(t_0, t)$ has maximum value. Let x_1 be the root, and let T_0 and T_1 be subtrees out of the 0 and 1 edges of x_1. Under the action of a random p-restriction $\rho \leftarrow \boldsymbol{\rho}$, if x_1 is assigned 0 by ρ, we get a $(t_0 - 1, t)$ clipped decision tree T_0 on n_0 variables where $n_0 < n$. By the definition of decision trees, since x_1 appears as the root of T, it cannot appear again as a variable in T_0, and thus T_0 is indeed a $(t_0 - 1, t)$-clipped tree. If x_1 is assigned 1 by ρ, similarly, we get a (t, t) clipped decision tree T_1 on n_1 variables where $n_1 < n$.

Let the event $E_{T,d}$ be defined as follows:

$$E_{T,d} \equiv \mathrm{DT}_{\mathrm{depth}}(T|\boldsymbol{\rho}) \geq d.$$

▶ **Lemma 14.** *In case x_1 is assigned $*$ by a random p-restriction $\rho \leftarrow \boldsymbol{\rho}$,*

$$E_{T,d} = E_{T_0,d-1} \bigcup E_{T_1,d-1}.$$

Let $\boldsymbol{\rho} = \boldsymbol{\rho_{x_1}\rho'}$ where $\boldsymbol{\rho'}$ is a random restriction on variables different from x_1, and $q = (1-p)/2$. Thus, we can write the following,

$$\begin{aligned}
\Pr_{\rho\leftarrow\boldsymbol{\rho}}[E_{T,d}] &= \Pr_{\rho\leftarrow\boldsymbol{\rho}}[E_{T,d}|\rho(x_1)=0]\Pr_{\rho\leftarrow\boldsymbol{\rho}}[\rho(x_1)=0] + \Pr_{\rho\leftarrow\boldsymbol{\rho}}[E_{T,d}|\rho(x_1)=1]\Pr_{\rho\leftarrow\boldsymbol{\rho}}[\rho(x_1)=1] \\
&\quad + \Pr_{\rho\leftarrow\boldsymbol{\rho}}[E_{T,d}|\rho(x_1)=*]\Pr_{\rho\leftarrow\boldsymbol{\rho}}[\rho(x_1)=*] \\
&= \Pr_{\rho'\leftarrow\boldsymbol{\rho'}}[E_{T_0,d}]q + \Pr_{\rho'\leftarrow\boldsymbol{\rho'}}[E_{T_1,d}]q + \Pr_{\rho'\leftarrow\boldsymbol{\rho'}}[E_{T_0,d-1}\bigcup E_{T_1,d-1}]p \\
&\le \Pr_{\rho'\leftarrow\boldsymbol{\rho'}}[E_{T_0,d}]q + \Pr_{\rho'\leftarrow\boldsymbol{\rho'}}[E_{T_1,d}]q + \Pr_{\rho'\leftarrow\boldsymbol{\rho'}}[E_{T_0,d-1}]p + \Pr_{\rho'\leftarrow\boldsymbol{\rho'}}[E_{T_1,d-1}]p \qquad (1)
\end{aligned}$$

where the second line follows from Lemma 14 and the fact that the subtrees at x_1 do not contain x_1, and the last line follows from the union bound. Further, rewriting the above inequality in terms of $\gamma_{d,n}(t_0,t)$ and using Lemma 13, we can write,

$$\begin{aligned}
\gamma_{d,n}(t_0,t) &\le q\gamma_{d,n_0}(t_0-1,t) + q\gamma_{d,n_1}(t,t) + p\gamma_{d-1,n_0}(t_0-1,t) + p\gamma_{d-1,n_1}(t,t) \\
&\le q\gamma_{d,n}(t_0-1,t) + q\gamma_{d,n}(t,t) + p\gamma_{d-1,n}(t_0-1,t) + p\gamma_{d-1,n}(t,t).
\end{aligned}$$

The parameters n and t above are implicit, and we can rewrite the recurrence succinctly as

$$\gamma_d(t_0) \le q\gamma_d(t_0-1) + q\gamma_d(t) + p\gamma_{d-1}(t_0-1) + p\gamma_{d-1}(t) \qquad (2)$$

and for every integer d, we set

$$\gamma_d(0) = 0. \qquad (3)$$

We get the following recurrence for γ.

▶ **Lemma 15.** *After m iterations, the recurrence is,*

$$\gamma_d(t_0) \le \sum_{i=0}^{m}\binom{m}{i}q^{m-i}p^i\gamma_{d-i}(t_0-m) + \sum_{i=1}^{m}q^i\gamma_d(t) + \sum_{j=1}^{m}p^j\gamma_{d-j}(t)\left(\sum_{i=0}^{m-j}\binom{j+i}{i}q^i\right) \qquad (4)$$

4.2 Upper bound on γ

Setting $t_0 = t$ and $m = t$ in equation 4 and using 3, we get,

$$\frac{1-2q+q^{t+1}}{1-q}\gamma_d(t) \le \sum_{j=1}^{t}p^j\gamma_{d-j}(t)\left(\sum_{i=0}^{t-j}\binom{j+i}{i}q^i\right).$$

Using the induction hypothesis, for $\mu = \kappa 2^t$, we have that $\gamma_{d-c}(t) \le (\mu p)^{d-c}$ for all p, c, d, t where $p \le \frac{1}{\mu}$. Note that γ is a probability and since $p \le \frac{1}{\mu}$, it is also valid when $c > d$. [1] Then we have,

$$\frac{1-2q+q^{t+1}}{1-q}\gamma_d(t) \le (\mu p)^d\left(\sum_{j=1}^{t}\sum_{i=0}^{t-j}\left(\frac{1}{\mu}\right)^j\binom{j+i}{i}q^i\right). \qquad (5)$$

The next lemma is important, but the proof is relegated to the full version [5].

[1] In fact, whenever $d < t$, we can indeed get much fewer terms in the summation, and get a better bound, although asymptotically it does not make a difference.

▶ **Lemma 16.** *For* $\mu = 4.2^t$,

$$\left(\sum_{j=1}^{t}\sum_{i=0}^{t-j}\left(\frac{1}{\mu}\right)^j\binom{j+i}{i}q^i\right)\left(\frac{1-q}{1-2q+q^{t+1}}\right) \leq 1.$$

Using Lemma 16 in equation 5, we get that $\gamma_d(t) \leq (4p2^t)^d$, which concludes the proof of the resilience upper bound in Theorem 1.

5 Resilience lower bound

We now prove the lower bound on resilience in Theorem 1 which is the main contribution of this work. We prove the bound for t-clipped xor Tree Tribes or $\Xi_t(r)$, by induction on d. However, we cannot use $d = 0$ as the base case, and we discuss this when we do the inductive step. The base case will be $d = 1$, which we solve next. The proofs from this section are given in the full version [5].

5.1 The case for $d = 1$

This base case will turn out the most interesting. Here, we want to show the following.

$$\mathrm{R}_1(\Xi_t(r)|\boldsymbol{\rho}) \geq c_0 p 2^t.$$

If r is small, i.e., if the function has fewer than $\sim 2^t$ variables, then the number of variables assigned $*$ by $\rho \leftarrow \boldsymbol{\rho}$ will be low on average, and the event $\mathrm{DT}_{\mathrm{depth}}(\Xi_t(r)|\boldsymbol{\rho}) \geq 1$ would be extremely unlikely. Thus, we would like to show the following.

▶ **Lemma 17.** *For some universal constants* c_0 *and* c_p, *for* $r \in \Omega(t2^t)$ *and* $0 \leq p \leq c_p 2^{-t}$, *if* $\boldsymbol{\rho}$ *is a random* p*-restriction, then*

$$R_1(\Xi_t(r)|\boldsymbol{\rho}) \geq c_0 p 2^t.$$

Note that in Lemma 17, we require that the decision tree with the *smallest* depth that can compute $\Xi_t(r)|\boldsymbol{\rho}$ has depth greater than 1 with good probability, which means that we require that *any* decision tree representing $\Xi_t(r)|\boldsymbol{\rho}$ must query at least one variable. This will be made possible by a simple observation.

▶ **Lemma 18.** $DT_{depth}(\Xi_t(r)|\boldsymbol{\rho}) \geq 1$ *if and only if there is a path in* $\Xi_t(r)|\boldsymbol{\rho}$ *from the root to a leaf that evaluates to 0 and to a leaf that evaluates to 1.*

▶ **Definition 19.** Let $\Xi_t(r)$ be a t-clipped xor tribe on r levels and $\boldsymbol{\rho}$ a random p-restriction. Letting the parameters p and t be implicit, let $P_0(r)$, $P_1(r)$ and $P_*(r)$ be defined as follows.

$$\begin{aligned} P_0(r) &= \Pr_{\rho}[\Xi_t(r)|\boldsymbol{\rho} \equiv 0], \\ P_1(r) &= \Pr_{\rho}[\Xi(r)|\boldsymbol{\rho} \equiv 1], \\ P_*(r) &= \mathrm{R}_1(\Xi_t(r)|\boldsymbol{\rho}) \\ &= 1 - P_0(r) - P_1(r). \end{aligned} \tag{6}$$

Given Lemma 18, the proof for Lemma 17 proceeds as follows:

Step 1. We write the exact expressions for $P_0(r)$ and $P_1(r)$ as a polynomial in p by using counting arguments and recursion.

Step 2. In the next step, we reason about the constant coefficients of $P_0(r)$ and $P_1(r)$. We derive the expressions for $[1]P_0(r)$ and $[1]P_1(r)$, using which we show:

$$[1]P_0(r) + [1]P(r) = 1.$$

Step 3. In the third step, we first compute the recursive expressions for $[p]P_0(r)$ and $[p]P_1(r)$ and show the the following two claims.

▶ **Lemma 20.** *For every* $r \geq 1$, $-2.2^t \leq [p]P_0(r) \leq 0$ *and* $-2.2^t \leq [p]P_1(r) \leq 0$.

▶ **Lemma 21.** *For* $r \in \Omega(t2^t)$, $[p](P_0(r) + P_1(r)) \leq -c_1 2^t$ *where* $c_1 = \frac{1}{3}$.

The proofs of all the above lemmas are given in the full version [5]. Lemma 21 stated above is an important one, and it states there there is sufficient mass, $\sim 2^t$ in the coefficient of p in $P_0(r) + P_1(r)$. Combining lemma 20 and 21, we get that

$$-4.2^t \leq [p](P_0(r) + P_1(r)) \leq -c_1 2^t.$$

Step 4. In the equation at the end of step 3 above, we almost have sufficient information to conclude Lemma 17, however, we need to reason that higher powers of p in $P_0(r) + P_1(r)$ cannot substantially affect the coefficient of p. We do so by showing that the absolute value of $[\uparrow p^2](P_0(r) + P_1(r))$ is bounded by $O(2^{2t})$. Here we will use the fact that $p \leq c_p 2^{-t}$ where c_p is some universal constant. This lemma is also important, and its proof is given in the full version [5].

▶ **Lemma 22.** *For every* $r \geq 1$, *for* $0 \leq p \leq p_{\max} = \frac{1}{200.2^t}$,

$$G_2(P_0(r) + P_1(r)) \leq 30.2^{2t}.$$

Step 5. Finally we prove Lemma 17.

Proof of Lemma 17. The probability that the decision tree with r levels has depth more than or equal to 1, i.e. it does not become constantly 0 or 1 after being hit by a random restriction is given by

$$P_*(r) = 1 - P_0(r) - P_1(r).$$

For $r \in \Omega(t2^t)$, we can write

$$\begin{aligned} P_*(r) &= 1 - [1](P_0(r) + P_1(r)) - ([p]P_0(r) + [p]P_1(r))\,p - \left([\uparrow p^2](P_0(r) + P_1(r))\right)p^2 \\ &= -([p]P_0(r) + [p]P_1(r))\,p - \left([\uparrow p^2](P_0(r) + P_1(r))\right)p^2 \\ &\geq c_1 2^t p - p_{\max} 2.30.2^{2t} p \\ &= p2^t\left(\frac{1}{3} - \frac{60}{200}\right) \\ &\geq \frac{1}{30}p2^t \end{aligned}$$

where in the second line we used the fact that $[1]P_0(r) + [1]P(r) = 1$, and in the third line we used Lemmas 21 and 22. Note that we get $c_0 = \frac{1}{30}$. ◀

5.2 Inductive step

We remark that $d = 0$ cannot be taken as the base case for induction. In a decision tree T with x_1 as the root and having subtrees T_0 and T_1 out of the 0 and 1 edges respectively, if $\rho \leftarrow \boldsymbol{\rho}$ assigns $*$ to x_1, then $T_0|\rho$ or $T_1|\rho$ having depth greater than 0 does not imply that T has depth greater than 1. If both $T_0|\rho$ and $T_1|\rho$ are constant functions, we would additionally require that they evaluate to *different* values. Thus we cannot use $d = 0$ as the base case.

Doing the inductive step recursively is tricky. This is because we already need about $\sim t2^t$ levels in the tree for the base case to work. Thus, any induction must take care of the fact that once the number of levels are too few, the base case would not hold.

Let us write the recurrence first. Let $\gamma_d(r)$ be the probability that $\Xi_t(r)$ has depth greater than or equal to d under the action of a random restriction.

▶ **Lemma 23.** *Let $\mu = 1 - \gamma_{d-1}(r-1)$. Then,*

$$\gamma_d(r) \geq \sum_{k=1}^{t-1} q \sum_{i=1}^{k} \binom{k}{i} q^{k-i} p^i (1-\mu^{i+1}) + \sum_{i=0}^{t} \binom{t}{i} q^{t-i} p^i (1-\mu^i) + \sum_{k=0}^{t-1} q^{k+1} \gamma_d(r-1). \quad (7)$$

Note that we would use the above recurrence for m steps, since we need to pick up sufficient probability mass from each step. As such, we would want the inductive hypothesis to hold for *all* the m steps. The inductive steps will be different for the cases $t \geq 2$ and $t = 1$, shown in the following lemma.

▶ **Lemma 24.** *For $r \in \Omega(dt2^t)$, $\gamma_d(r) \geq (c_0 p 2^t)^d$.*

In the proof of Lemma 24, we use $O(2^t)$ steps to take the induction from d to $d-1$ levels, and the final $d = 1$ case can be carried out with $O(t2^t)$ levels. Thus, the maximum depth for which our conclusions hold is

$$dt2^t \leq O(r) \leq c_d \frac{\log n}{\log t}$$

or

$$d \leq c_d \left(\frac{\log n}{2^t t \log t} \right).$$

This concludes the proof of Theorem 1.

▶ Remark. If we unroll the induction and see how it worked, for depth d, we find some vertex that is unset by $\boldsymbol{\rho}$ and is connected to a tree that has depth $d-1$ under the action of ρ. For depth $d-1$, we again find an unset variable connected to a tree of depth $d-2$. This carries on until the last step, where we want to find a vertex that has depth greater than or equal to 1, i.e., has a path to both a 0-leaf and a 1-leaf. Thus, as described in the introduction, the whole proof essentially finds a *"path with a split"*.

References

1 Miklos Ajtai. Sigma 1-formulae on finite structures. *Annals of pure and applied logic*, 24(1):1–48, 1983.

2 Merrick Furst, James B Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Theory of Computing Systems*, 17(1):13–27, 1984.

3 Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the eighteenth annual ACM symposium on Theory of computing*, pages 6–20. ACM, 1986.

4 Johan Håstad. On the correlation of parity and small-depth circuits. *SIAM Journal on Computing*, 43(5):1699–1708, 2014.

5 Jenish C. Mehta. Tree tribes and lower bounds for switching lemmas. *CoRR*, 2017. URL: `https://arxiv.org/abs/1703.00043`.

6 Toniann Pitassi, Benjamin Rossman, Rocco A Servedio, and Li-Yang Tan. Poly-logarithmic frege depth lower bounds via an expander switching lemma. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 644–657. ACM, 2016.

7 Alexander A Razborov. An equivalence between second order bounded domain bounded arithmetic and first order bounded arithmetic, 1993.

8 Alexander A Razborov and Steven Rudich. Natural proofs. *Journal of Computer and System Sciences*, 55(1):24–35, 1997.