# Attack-Resilient Multitree Data Distribution Topologies

Sascha Grau*

sascha.grau@tu-ilmenau.de
Technische Universität Ilmenau, Germany

**Abstract.** We consider a scenario of information broadcast where a source node distributes data in parallel over a fixed number of trees spanning over a large audience of nodes. The trees used for data dissemination are called distribution topology. Particular implementations of this scenario are peer-to-peer live streaming systems. Encoding data partially redundant, nodes are satisfied as long as they receive packets in at least a certain portion of trees. Otherwise, they are called *isolated*. We study distribution topologies limiting the worst-case consequences of attacks suddenly removing nodes from the trees. In particular, we aim to minimize the maximum possible number of isolated nodes for each number of removed nodes. We show necessary conditions on distribution topologies closely approximating this goal. Then, we demonstrate that the attack-resilience of topologies adhering to these conditions is characterized by specific matrices that have to be Orthogonal Arrays of maximum strength. The computational complexity of finding such matrices for arbitrary dimensions is a long-standing research problem. Our results show that finding representatives of the studied distribution topologies is at least as hard as this problem.

**Keywords:** network topologies, dependability, P2P, orthogonal arrays

## 1 Introduction

In many applications data shall be reliably broadcast from a resource-restricted source to a large audience of nodes. Applying multiple description coding [1] or error-correcting codes [2], it is possible to split each block of data into $k$ subblocks, such that the reception of a certain portion of these subblocks already satisfies the participating nodes (i.e. they can restore the original data to satisfactory degree).

Distributing each of the $k$ subblocks from node to node over a distinct tree rooted at the source, a data distribution system is obtained which

* This work was supported by the *Deutsche Forschungsgemeinschaft* under grant number KU 658/10-2.

Published at OPODIS 2012.
The original publication is available at www.springerlink.com
(http://dx.doi.org/10.1007/978-3-642-35476-2_14).

is tolerant to failures. Furthermore, the number of participants in such a system can scale independently from resource restrictions of the source. Popular implementations of such approaches can be found in peer-to-peer live streaming systems like [3,4,5].

Due to their spreading application and growing importance, such data distribution systems are target of attacks. Abstracting from technical details, these attacks can often be modeled as a removal of nodes from the system. The consequences of such a removal can be measured as damage and depend on the layout of the distribution topology, i.e., the trees used for data dissemination. This motivated the study of distribution topologies minimizing the maximum damage that is achievable on them.

Here, different measures of damage can be of interest. In the past, distribution topologies minimizing notions of system-wide damage, like the global number of disturbed source-to-node paths, have been identified [3,6]. However, in many applications a damage measure based on the user-perceived quality of the data distribution service is more relevant. This corresponds to counting the number of nodes that are no longer satisfied since they lost too many paths from the source.

In the following, distribution topologies minimizing this kind of damage are called *attack-resilient*. Despite their relevance, the author is not aware of any analytical study of such topologies or of related network design problems based on a similarly generalized concept of connectivity. Current applications resort to following rules of thumb, as building 'diverse trees' [5]. Some insights for scenarios considering random node removal instead of worst-case attacks were obtained by simulation in [7].

*Contribution* In this document, we introduce *forward-stable* distribution topologies and show that they closely approximate attack-resilient topologies in situtations where the number of nodes considerably exceeds the number of source neighbors. This is a usual condition in applications of multitree data distribution topologies. We find necessary and sufficient requirements for forward-stable topologies and show that they can be characterized by matrices representing certain successor relations in the trees. By showing that these matrices have to be Orthogonal Arrays of maximum possible strength, we discover connections to design and coding theory. In particular, we show that the identification of an efficient construction scheme for forward-stable topologies would solve several long-standing open problems in these areas.

*Structure of This Document* In Section 2, we specify our system model and formalize the notion of attack-resilient distribution topologies. Section 3 introduces and motivates an alternative damage measure which

is then used in Section 4 to define forward-stable distribution topologies. Their properties are studied in-depth in the following Subsections. Finally, Section 5 concludes this document.

## 2 System Model and Attack-Resilient Topologies

In our system model, a source $s$ distributes data to a set $V = \{1, \ldots, n\}$ of nodes. Each block of data is encoded into $k$ subblocks and a node is satisfied as long as it receives *more* than $k - z$ such subblocks, for a fixed $z \in \{1, \ldots, k\}$ (see [1,2] for suitable encoding schemes). Otherwise, the node is called *isolated*. Each subblock is distributed over one of $k$ distribution trees (also called *stripes*). Those have node set $\{s\} \cup V$, are rooted at $s$, and are directed towards the leafs. A *distribution topology* is a $k$-tupel $\mathcal{T} = (T_1, \ldots, T_k)$ of stripes. The nodes that are adjacent to the source in stripe $T_i$ of $\mathcal{T}$ are the *heads* $H_i^{\mathcal{T}}$. The nodes $H^{\mathcal{T}} = \bigcup_{i \in \{1,\ldots,k\}} H_i^{\mathcal{T}}$ are the *heads of* $\mathcal{T}$.

We assume that the maximum degree of source node $s$ is limited to a value of $Ck$, for $C \in \mathbb{N}$ and $n \geq Ck$. The class of all distribution topologies with $k$ trees, node set $V = \{1, \ldots, n\}$ and source degree limit $Ck$ is denoted as $\mathbb{T}(n, C, k)$.

The data distribution over a topology $\mathcal{T} \in \mathbb{T}(n, C, k)$ can be disturbed in a number of ways. In this document, we study the consequences of sudden removals of nodes. Such events are common, especially in peer-to-peer systems with their unreliable and vulnerable participants. Considering the worst-case, we assume that the sets of removed nodes are the result of a malicious planning process. For this reason, they are termed as *attacks*. Note that we do not account for a removal of the source node, since it would always result in a non-functional distribution topology. Furthermore, in practical applications it is usual to take special measures to safeguard source functionality.

When a node $v$ is removed from topology $\mathcal{T}$, in each stripe $T_i$ with $i \in \{1, \ldots, k\}$, the paths between $s$ and all nodes in the subtree rooted at $v$ become disturbed. The set of nodes in this subtree is denoted as *successor set* $\mathrm{succ}_i^{\mathcal{T}}(v)$ and contains $v$. For node sets $X$, we correspondingly define $\mathrm{succ}_i^{\mathcal{T}}(X) = \bigcup_{v \in X} \mathrm{succ}_i^{\mathcal{T}}(v)$. Figure 2 gives an example.

Assuming that a node is isolated by the loss of at least $z$ paths from the source, the number of nodes isolated by attack $X$ is counted as *damage*

$$\mathrm{b}^{\mathcal{T}}(X, z) := \left| \bigcup_{I \subseteq \{1,\ldots,k\}, |I| = z} \bigcap_{i \in I} \mathrm{succ}_i^{\mathcal{T}}(X) \right|. \tag{1}$$

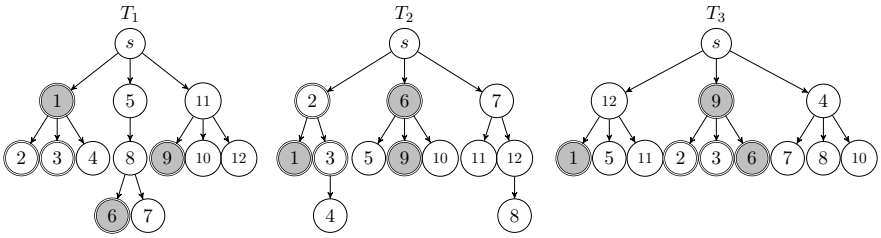



**Fig. 1.** Attack $X = \{1, 6, 9\}$ on this topology $\mathcal{T} \in \mathbb{T}(12, 3, 3)$ leads to $\mathrm{b}^{\mathcal{T}}(X, 2) = 5$ (attacked nodes gray, isolated nodes double-lined).

Figure 1 shows an example in which nodes are isolated by the loss of at least 2 paths from the source.

Given an arbitrary class $\mathbb{T}(n, C, k)$, we are generally interested in finding topologies $\mathcal{T} \in \mathbb{T}(n, C, k)$ minimizing the maximum damage that can occur for every possible number $x$ of removed nodes and every value of threshold $z$. Note that for $x \geq Cz$, it is possible to remove all heads of $z$ stripes (the ones with the least number of heads) and isolate all nodes. Hence, the maximum damage on topologies in $\mathbb{T}(n, C, k)$ can only differ for $x < Cz \leq Ck$. This leads to the following definition.

**Definition 1.** *A topology $\mathcal{T} \in \mathbb{T}(n, C, k)$ is* attack-resilient*, if for all $z \in \{1, \ldots, k\}$, all $x \in \{1, \ldots, Ck\}$, and all $\mathcal{C} \in \mathbb{T}(n, C, k)$, it holds that*

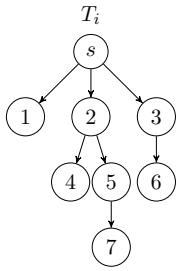
$$\max_{X \subseteq V, |X| = x} \mathrm{b}^{\mathcal{T}}(X, z) \leq \max_{X \subseteq V, |X| = x} \mathrm{b}^{\mathcal{C}}(X, z).$$

## 3 An Approximative Damage Measure

The function $\mathrm{b}^{\mathcal{T}}(X, z)$ used to characterize attack-resilient topologies counts nodes of two different kinds. On the one hand, it considers all removed nodes in the set $X$. On the other hand, it counts nodes positioned in subtrees below removed nodes in at least $z$ stripes. Furthermore, there are nodes falling into both categories. This superposition of different causes of damage complicates an analyis. For this reason, we choose to study a slightly altered notion of damage. At first, we define the *forward successor set* of a node $v$ in stripe $T_i$ of $\mathcal{T}$:

$$\mathrm{succ}_i^{\mathcal{T}\rightarrow}(v) := \begin{cases} \mathrm{succ}_i^{\mathcal{T}}(v) & \text{, if } |\mathrm{succ}_i^{\mathcal{T}}(v)| > 1 \text{ or } v \in H_i^{\mathcal{T}} \\ \emptyset & \text{, otherwise.} \end{cases} \tag{2}$$

It is equal to the successor set in most cases, but is empty if $v$ is neither




(a) A tree $T_i$ from a topology $\mathcal{T}$.

| $X$ | $\mathrm{succ}_i^{\mathcal{T}}(X)$ | $\mathrm{succ}_i^{\mathcal{T}\rightarrow}(X)$ |
|---|---|---|
| $\{1\}$ | $\{1\}$ | $\{1\}$ |
| $\{3\}$ | $\{3,6\}$ | $\{3,6\}$ |
| $\{4\}$ | $\{4\}$ | $\emptyset$ |
| $\{3,5,6\}$ | $\{3,5,6,7\}$ | $\{3,5,6,7\}$ |

(b) Successor and forward successor sets.

**Fig. 2.** Different concepts of successor sets.

head nor has children in $T_i$. Again, this definition extends to node sets: $\mathrm{succ}_i^{\mathcal{T}\rightarrow}(X) = \bigcup_{v\in X} \mathrm{succ}_i^{\mathcal{T}\rightarrow}(v)$. Figure 2 provides an example.

For $\mathcal{T} \in \mathbb{T}(n, C, k)$, $z \in \{1, \ldots, k\}$, and attacks $X \subseteq V$, we define the corresponding damage function as *forward damage*

$$\mathrm{bf}^{\mathcal{T}}(X, z) := \left| \bigcup_{I\subseteq\{1,\ldots,k\},|I|=z} \bigcap_{i\in I} \mathrm{succ}_i^{\mathcal{T}\rightarrow}(X) \right|. \tag{3}$$

Since it holds that $\mathrm{succ}_i^{\mathcal{T}}(X) = X \cup \mathrm{succ}_i^{\mathcal{T}\rightarrow}(X)$, we observe that

$$\mathrm{bf}^{\mathcal{T}}(X, z) \leq \mathrm{b}^{\mathcal{T}}(X, z) \leq \mathrm{bf}^{\mathcal{T}}(X, z) + |X|. \tag{4}$$

The maximum value of both, possible damage and forward damage, is $n$ on each topology $\mathcal{T} \in \mathbb{T}(n, C, k)$ if at least $Cz$ nodes may be removed (removing the heads of $z$ stripes). Together with Equation (4), we obtain the following theorem.

**Theorem 1.** *For every topology $\mathcal{T} \in \mathbb{T}(n, C, k)$, $z \in \{1, \ldots, k\}$, and $x \in \{1, \ldots, n\}$, it holds that*

$$\max_{\substack{X\subseteq V\\|X|=x}} \mathrm{bf}^{\mathcal{T}}(X, z) \leq \max_{\substack{X\subseteq V\\|X|=x}} \mathrm{b}^{\mathcal{T}}(X, z) \leq \max_{\substack{X\subseteq V\\|X|=x}} \mathrm{bf}^{\mathcal{T}}(X, z) + \min(Cz - 1, x).$$

In applications of multitree data distribution topologies, we usually have $n \gg Ck$. Furthermore, the maximum achievable forward damage for threshold $z$ on each topology $\mathcal{T} \in \mathbb{T}(n, C, k)$ is in $\Omega(\frac{n}{C^z})$ if at least $z$ nodes are removed. Consequently, with growing node numbers, the maximum possible forward damage *dominates* the value of the maximum possible damage.

## 4 Forward-Stable Distribution Topologies

Theorem 1 motivates the study of distribution topologies minimizing maximum forward damage for all numbers of removed nodes and thresholds $z$. In the following, we will distinguish between different levels of this resilience concept by restricting the possible sets of removed nodes. For this, we introduce the *$t$-restricted attacks* $\chi(\mathcal{T}, t)$ for each $\mathcal{T} \in \mathbb{T}(n, C, k)$ and $t \in \{1, \ldots, k\}$. An attack $X \subseteq V$ satisfies $X \in \chi(\mathcal{T}, t)$, if there is a set $I \subseteq \{1, \ldots, k\}$ of $t$ stripe indices such that each $v \in X$ either has forward successors in at least one of the stripes $I$, or it has no forward successors at all. Thus, if topology $\mathcal{T}$ has inner-node disjoint stripe trees, $\chi(\mathcal{T}, t)$ is the set of all attacks containing inner-nodes from *at most* $t$ stripes and an arbitrary number of nodes that are leaf in all stripes.
The definition ensures that $\chi(\mathcal{T}, t-1) \subseteq \chi(\mathcal{T}, t)$ is true and that $\chi(\mathcal{T}, k)$ equals the power set $\mathcal{P}(V)$ of $V$. Furthermore, for each $t \in \{1, \ldots, k\}$, the set $\chi(\mathcal{T}, t)$ contains *all* subsets of $V$ that have cardinality up to $t$.

Now, we can define $t$-forward-stable and forward-stable distribution topologies.

**Definition 2.** *A topology $\mathcal{T} \in \mathbb{T}(n, C, k)$ is called* $t$-forward-stable*, if for all $z \in \{1, \ldots, k\}$, $x \in \{1, \ldots, n\}$, and $\mathcal{C} \in \mathbb{T}(n, C, k)$, it holds that*

$$\max_{X \in \chi(\mathcal{T}, t), |X| = x} \mathrm{bf}^{\mathcal{T}}(X, z) \leq \max_{X \in \chi(\mathcal{C}, t), |X| = x} \mathrm{bf}^{\mathcal{C}}(X, z).$$

*If $\mathcal{T}$ is $t$-forward-stable for all $t \in \{1, \ldots, k\}$, it is called* forward-stable*.*

Consequently, a topology $\mathcal{T}$ is $t$-forward-stable, if it minimizes the maximum possible forward damage that is achievable by $t$-restricted attacks (for all attack cardinalities and thresholds $z$), while forward-stable topologies are $t$-forward-stable for all possible values of $t$. As we have seen in Section 3, the latter closely approximate attack-resilient topologies.

In the following, we show necessary and sufficient requirements for ($t$-)forward-stable topologies. Furthermore, we give a notion of the computational complexity of finding a forward-stable topology in a given class $\mathbb{T}(n, C, k)$. In particular, we show that a corresponding oracle could be used to efficiently determine so-called Orthogonal Arrays of given dimension and maximum strength. The latter is a notorious problem in both design and coding theory [8,2].

### 4.1 Basic Requirements

**Lemma 1.** *A $t$-forward-stable topology $\mathcal{T} \in \mathbb{T}(n, C, k)$ with $t \in \{1, \ldots, k\}$ has the following properties:*

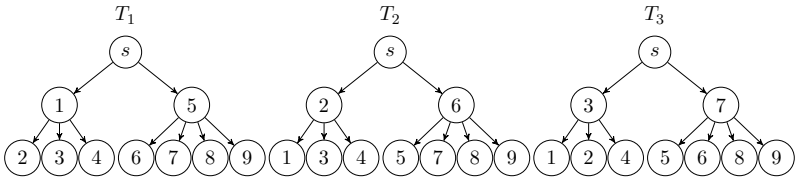



**Fig. 3.** A distribution topology $\mathcal{C} \in \mathbb{T}(9, 2, 3)$ as in the proof of Lemma 1.

1. $\forall v \in V : |\{i \in \{1, \ldots, k\} \mid \mathrm{succ}_i^{\mathcal{T}\rightarrow}(v) \neq \emptyset\}| \leq 1$
2. $\forall v \in V : |\bigcup_{i\in\{1,\ldots,k\}} \mathrm{succ}_i^{\mathcal{T}\rightarrow}(v)| \leq \lceil \frac{n}{C} \rceil$

*Proof.* We compare $\mathcal{T}$ with a topology $\mathcal{C} \in \mathbb{T}(n, C, k)$ that has $Ck$ distinct heads in total and $C$ heads per stripe. In each stripe $i \in \{1, \ldots, k\}$, all nodes $V \setminus H_i^{\mathcal{C}}$ are leafs below the heads $H_i^{\mathcal{C}}$. They are grouped such that each head $h \in H_i^{\mathcal{C}}$ satisfies $|\mathrm{succ}_i^{\mathcal{C}\rightarrow}(h)| \in \{\lceil n/C \rceil, \lfloor n/C \rfloor\}$. Figure 3 gives an example of such a topology.

Since it is $t$-forward-stable with $t \geq 1$, topology $\mathcal{T}$ should minimize the maximum possible forward-damage for attacks of cardinality 1 and all values of $z$. However, if $\mathcal{T}$ lacks one of the mentioned properties, we show that, for certain $z$, there are attacks of cardinality 1 on $\mathcal{T}$ that achieve more forward-damage than any such attack can achieve on $\mathcal{C}$:

1. Assume there is $v \in V$ and two distinct stripes $i, j \in \{1, \ldots, k\}$, such that $\mathrm{succ}_i^{\mathcal{T}\rightarrow}(v) \neq \emptyset$ and $\mathrm{succ}_j^{\mathcal{T}\rightarrow}(v) \neq \emptyset$. Then, it holds that $v \in \mathrm{succ}_i^{\mathcal{T}\rightarrow}(v) \cap \mathrm{succ}_j^{\mathcal{T}\rightarrow}(v)$. In contrast, for all $w \in V$ there is no pair $i, j$ of distinct stripes of $\mathcal{C}$ such that $\mathrm{succ}_i^{\mathcal{C}\rightarrow}(w) \cap \mathrm{succ}_j^{\mathcal{C}\rightarrow}(w) \neq \emptyset$. It follows that $\max_{u\in V} \mathrm{bf}^{\mathcal{T}}(\{u\}, 2) \geq 1$ and $\max_{u\in V} \mathrm{bf}^{\mathcal{C}}(\{u\}, 2) = 0$. Consequently, $\mathcal{T}$ is not $t$-forward-stable.
2. Assume that $\exists v \in V : |\bigcup_{i\in\{1,\ldots,k\}} \mathrm{succ}_i^{\mathcal{T}\rightarrow}(v)| > \lceil \frac{n}{C} \rceil$. For every topology $\mathcal{D} \in \mathbb{T}(n, C, k)$, the definition of forward damage guarantees that

$$\max_{X\subseteq V, |X|=1} \mathrm{bf}^{\mathcal{D}}(X, 1) = \max_{u\in V} \Big| \bigcup_{i\in\{1,\ldots,k\}} \mathrm{succ}_i^{\mathcal{D}\rightarrow}(u) \Big|. \tag{5}$$

   In $\mathcal{C}$, this value is $\lceil \frac{n}{C} \rceil$, whereas it is higher in $\mathcal{T}$. Again, $\mathcal{T}$ is not $t$-forward-stable. □

The first property ensures the construction of inner-node disjoint stripe trees. The second one corresponds to a balanced distribution of successors to the heads of each stripe. Both are frequent optimization goals in peer-to-peer live streaming systems such as [4] and [3]. Note that

topologies from $\mathbb{T}(n, C, k)$ with both properties will have $C$ unique heads per stripe.

Additionally, such topologies have another interesting property.

**Lemma 2.** *Let $\mathcal{T} \in \mathbb{T}(n, C, k)$ satisfy the requirements of Lemma 1. For all $z \in \{1, \ldots, k\}$ and each $X \subseteq V$, there exists an attack $Y \subseteq H^{\mathcal{T}}$ with $\mathrm{bf}^{\mathcal{T}}(Y, z) \geq \mathrm{bf}^{\mathcal{T}}(X, z)$ and $|Y| = \min(|X|, Cz)$.*

*Proof.* The stripe trees of topology $\mathcal{T}$ are inner-node disjoint. Therefore, the node sets $V_i := \{v \in V \mid \mathrm{succ}_i^{\mathcal{T}\rightarrow}(v) \neq 0\}$ for $i \in \{1, \ldots, k\}$ together with set $V_0 := V \setminus \bigcup_{i \in \{1,\ldots,k\}} V_i$ form a partition of $V$.

Since each $T_i$ is a tree, it holds that $\mathrm{succ}_i^{\mathcal{T}\rightarrow}(v) \subseteq \mathrm{succ}_i^{\mathcal{T}\rightarrow}(u)$ for each $v \in \mathrm{succ}_i^{\mathcal{T}\rightarrow}(u)$. Hence, for each stripe $T_i$ the set $\{\mathrm{succ}_i^{\mathcal{T}\rightarrow}(v) \mid v \in V_i\}$ is a *laminar family of sets*. In particular, the forward successor sets of the heads $H_i^{\mathcal{T}}$ are the only sets that are not subsets of others.

Now, let $X \subseteq V$ be an arbitrary attack on $\mathcal{T}$. If $|X| \geq Cz$, then *all* nodes can be isolated by attacking the (at most) $Cz$ heads of $z$ stripes with the smallest number of heads. Otherwise, set $X := X \setminus V_0$, and let

$$Y' := \{h \in H^{\mathcal{T}} \mid \exists i \in \{1, \ldots, k\}, v \in V_i \cap X \colon \mathrm{succ}_i^{\mathcal{T}\rightarrow}(v) \subseteq \mathrm{succ}_i^{\mathcal{T}\rightarrow}(h)\}.$$

Due to the node partition and set laminarity, it holds that $|Y'| \leq |X|$. Furthermore, we have $\forall i \in \{1, \ldots, k\} \colon \mathrm{succ}_i^{\mathcal{T}\rightarrow}(X) \subseteq \mathrm{succ}_i^{\mathcal{T}\rightarrow}(Y')$ and therefore $\mathrm{bf}^{\mathcal{T}}(X, z) \leq \mathrm{bf}^{\mathcal{T}}(Y', z)$ for all $z \in \{1, \ldots, k\}$. No superset $Y \subseteq H^{\mathcal{T}}$ with $Y' \subseteq Y$ and $|Y| = |X|$ can create less forward-damage. □

We see, that on every topology with the properties given in Lemma 1, a maximum value of forward damage can always be achieved by removing only heads. Consequently, the optimization of their forward successor sets is the key to finding forward-stable topologies.

### 4.2 A Matrix Representation and Orthogonal Arrays

For every distribution topology $\mathcal{T}$, there is a convenient matrix representation of its heads' forward successor sets.

**Definition 3.** *Let $\mathcal{T} \in \mathbb{T}(n, C, k)$ be given. Using per stripe $i \in \{1, \ldots, k\}$ a bijection $\sigma_i \colon H_i^{\mathcal{T}} \to \{1, \ldots, |H_i^{\mathcal{T}}|\}$, the* matrix $M^{\mathcal{T}}$ of forward successor sets of the heads $H^{\mathcal{T}}$ *is an $n \times k$ matrix $M^{\mathcal{T}} = (m_{vi})$, such that*

$$m_{vi} = \sigma_i(j) \Leftrightarrow v \in \mathrm{succ}_i^{\mathcal{T}\rightarrow}(j).$$

*For $v \in V$, $M^{\mathcal{T}}[v] = (m_{v1}, \ldots, m_{vk})$ denotes the $v$-th row of $M^{\mathcal{T}}$.*

Consequently, the $i$-th entry of the $v$-th row of $M^{\mathcal{T}}$ encodes the head supplying node $v$ in stripe $i$. Its numeric value is determined by bijection $\sigma_i$. As an example for this definition, Figure 4(b) shows a matrix corresponding to the topology in Figure 4(a).

Reusing the bijections $\sigma_i$ from $M^{\mathcal{T}}$, we can also transform attacks on the heads of $\mathcal{T}$ into sets of $k$-dimensional vectors. In their $i$-th position, these vectors contain entries from $\{0, \dots, |H_i^{\mathcal{T}}|\}$.

**Definition 4.** *Let topology $\mathcal{T} \in \mathbb{T}(n, C, k)$, matrix $M^{\mathcal{T}}$, and the the corresponding bijections $\sigma_i \colon H_i^{\mathcal{T}} \to \{1, \dots, |H_i^{\mathcal{T}}|\}$ for $i \in \{1, \dots, k\}$ be given.*

*The* vector attack $\sigma(X)$ *for an attack $X \subseteq H^{\mathcal{T}}$ contains each vector $\mathbf{y} \in \{0, \dots, Ck\}^k$ such that for all $i \in \{1, \dots, k\}$ either $\sigma_i^{-1}(\mathbf{y}_i) \in X$ or $(\mathbf{y}_i = 0) \wedge (X \cap H_i^{\mathcal{T}} = \emptyset)$ is true.*

Due to its definition, $\sigma(X)$ will contain $\prod_{i=1}^{k} \max(1, |X \cap H_i^{\mathcal{T}}|)$ vectors. In position $i$, such a vector either contains the value $\sigma_i(h)$ for some $h \in X \cap H_i^{\mathcal{T}}$ or the value 0 if $X \cap H_i^{\mathcal{T}} = \emptyset$.

Using vector attacks, the forward damage $\mathrm{bf}^{\mathcal{T}}(X, z)$ of an attack $X \subseteq H^{\mathcal{T}}$ on $\mathcal{T}$ can be determined by counting row vectors of $M^{\mathcal{T}}$ that are in *Hamming Distance* at most $k - z$ to an element of $\sigma(X)$. With $\mathrm{d}(\cdot, \cdot)$ as the Hamming Distance function, we can write

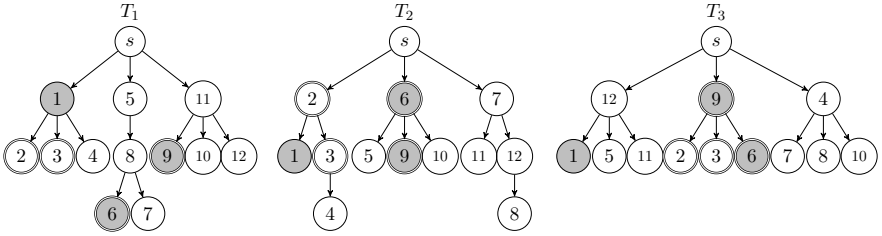
$$\begin{aligned}
\mathrm{bf}^{\mathcal{T}}(X, z) &= \left| \bigcup_{I \subseteq \{1,\dots,k\}, |I|=z} \bigcap_{i \in I} \mathrm{succ}_i^{\mathcal{T}\rightarrow}(X) \right| \\
&= \left| \left\{ v \in V \;\middle|\; \exists I \subseteq \{1, \dots, k\}, |I| = z, \forall i \in I \colon \sigma_i^{-1}(m_{vi}) \in X \right\} \right| \\
&= \left| \left\{ v \in V \;\middle|\; \exists \mathbf{x} \in \sigma(X) \colon \mathrm{d}(M[v], \mathbf{x}) \leq k - z \right\} \right| . \qquad (6)
\end{aligned}$$

Figure 4(c) gives a graphical example.

Next, we introduce a special class of matrices, the *Orthogonal Arrays* [8].

**Definition 5.** *For $n, k, C \in \mathbb{N}$ and $t \in \{0, \dots, k\}$, an $n \times k$ matrix $M$ with entries $m_{vi} \in \{1, \dots, C\}$ is called an* Orthogonal Array $\mathrm{OA}(n, k, C, t)$ *if in every $n \times t$ submatrix $M'$ consisting of $t$ complete columns of $M$, each $\mathbf{x} \in \{1, \dots, C\}^t$ appears exactly $\lambda := \frac{n}{C^t}$ times as a row.*

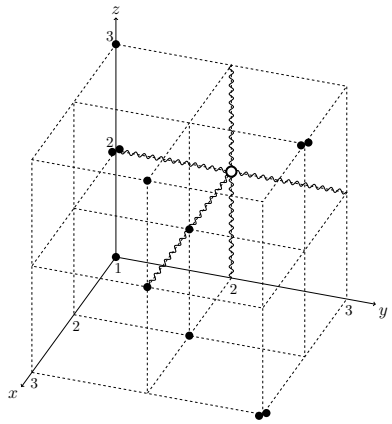
An $\mathrm{OA}(n, k, C, t)$ is said to have *strength* $t$. It minimizes the maximum frequency of a row vector in each of its $t$-column submatrices. Every Orthogonal Array of strength $t > 1$ also has strength $t - 1$. The strength of a given $n \times k$ matrix is computable in time $O(n^2 k)$ [8, Chapter 4.4]. Figure 5 shows an $\mathrm{OA}(18, 3, 3, 2)$.




(a) Attack $X = \{1, 6, 9\}$ on topology $\mathcal{T} \in \mathbb{T}(12, 3, 3)$ leads to $\mathrm{bf}^{\mathcal{T}}(X, 2) = 4$ (attacked nodes gray, nodes suffering forward-damage double-lined).

$$M^{\mathcal{T}} = \begin{pmatrix} 1\,1\,1 \\ 1\,1\,2 \\ 1\,1\,2 \\ 1\,1\,3 \\ 2\,2\,1 \\ 2\,2\,2 \\ 2\,3\,3 \\ 2\,3\,3 \\ 3\,2\,2 \\ 3\,2\,3 \\ 3\,3\,1 \\ 3\,3\,1 \end{pmatrix}$$

(b) $M^{\mathcal{T}}$ for
$\sigma_1(1) = \sigma_2(2) = \sigma_3(12) = 1$,
$\sigma_1(5) = \sigma_2(6) = \sigma_3(9) = 2$ and
$\sigma_1(11) = \sigma_2(7) = \sigma_3(4) = 3$.




(c) Rows of $M^{\mathcal{T}}$ (dots) in Hamming distance $\leq 1$ (snaked) from vector attack $\sigma(X) = \{(1, 2, 2)\}$ (circled) correspond to nodes $\{2, 3, 6, 9\}$.

**Fig. 4.** A distribution topology $\mathcal{T}$, a corresponding matrix $M^{\mathcal{T}}$, and forward damage due to the removal of node set $X = \{1, 6, 9\}$ from $\mathcal{T}$.

**Lemma 3.** *For every* $\mathrm{OA}(n, k, C, t)$ $M$ *with* $n \geq Ck$ *and strength* $t \geq 1$, *there is a topology* $\mathcal{T} \in \mathbb{T}(n, C, k)$ *with* $M^{\mathcal{T}} = M$ *that satisfies the requirements of Lemma 1.*

*Proof.* We construct a suitable topology $\mathcal{T}$ of depth 2. For the use as heads $H^{\mathcal{T}}$, we determine the indices of $Ck$ suitable rows of $M$. For this, construct a bipartite graph $G = (\{1, \ldots, n\} \dot\cup (\{1, \ldots, C\} \times \{1, \ldots, k\}), E)$. Its node set contains the nodes $V = \{1, \ldots, n\}$ of $\mathcal{T}$ and head positions $(i, j)$. A head position $(i, j)$ corresponds to the role as $i$-th head in stripe $j$ of $\mathcal{T}$. The edge set $E$ satisfies $\{v, (i, j)\} \in E \Leftrightarrow M[v]_j = i$.

$$\begin{pmatrix} 1\,1\,1\,2\,2\,2\,3\,3\,3\,1\,1\,1\,2\,2\,2\,3\,3\,3 \\ 1\,2\,3\,1\,2\,3\,1\,2\,3\,1\,2\,3\,1\,2\,3\,1\,2\,3 \\ 1\,2\,3\,2\,1\,3\,3\,1\,2\,1\,3\,2\,3\,2\,1\,2\,3\,1 \end{pmatrix}$$

**Fig. 5.** Transpose of an OA$(18, 3, 3, 2)$ with $\lambda = 2$.

For each node $u$, let $N(u)$ be the set of $u$'s neighbors in $G$. Since $M$ has $k$ columns, each node $v \in V$ satisfies $|N(v)| = k$. Since $M$ has strength at least 1, each head position $(i, j)$ has $|N((i, j))| = n/C$. Due to Hall's Theorem (cmp. [9]) there is a matching covering all head positions in $G$, if it holds that $\forall S \subseteq \{1, \ldots, C\} \times \{1, \ldots, k\} \colon |\bigcup_{u \in S} N(u)| \geq |S|$. This is the case in $G$. For each possible subset $S$ of head positions, there are $|S| \cdot n/C$ edges to nodes from $V$. Since these $|\bigcup_{u \in S} N(u)|$ nodes have $|\bigcup_{u \in S} N(u)| \cdot k$ edges in total and since $n/C \geq k$, we obtain

$$|S| \cdot \frac{n}{C} \leq |\bigcup_{u \in S} N(u)| \cdot k \quad \Rightarrow \quad |S| \leq |\bigcup_{u \in S} N(u)|. \tag{7}$$

Hence, a maximum matching $R$ in $G$ connects each head position with a unique node from $V$. For each $\{v, (i, j)\} \in R$, we use $v$ as head in stripe $T_j$ of $\mathcal{T}$, define $\sigma_j(v) := i$, and set $\mathrm{succ}_j^{\mathcal{T}}(v) := \{u \in V \mid M[u]_j = i\}$. In each stripe of the emerging topology $\mathcal{T}$, every node is either head or child of a head. The matching $R$ guarantees that we have $|H^{\mathcal{T}}| = Ck$ and that each head forwards in only one stripe. The defined bijections $\sigma_j$ with $j \in \{1, \ldots, k\}$ establish $M^{\mathcal{T}} = M$. Since $M$ is of strength at least 1, for all $j \in \{1, \ldots, k\}$ each head $v \in H_j^{\mathcal{T}}$ satisfies $|\mathrm{succ}_j^{\mathcal{T}\rightarrow}(v)| = n/C$. All other forward successor sets are empty. □

A matrix $M^{\mathcal{T}}$ of high strength is beneficial for the forward-stability of $\mathcal{T}$.

**Theorem 2.** *A topology $\mathcal{T} \in \mathbb{T}(n, C, k)$ is t-forward-stable, if it has the properties of Lemma 1 and $M^{\mathcal{T}}$ is an* OA$(n, k, C, t)$.

*Proof (sketch).* For reasons of space, we can only give a proof sketch. See [10, Theorem 5.3.14] for all details.

Given a topology $\mathcal{T} \in \mathbb{T}(n, C, k)$, we call each vector $\mathbf{x} \in \{0, \ldots, Ck\}^k$ satisfying $\forall i \in \{1, \ldots, k\} \colon \mathbf{x}_i \leq |H_i^{\mathcal{T}}|$ an *attack distribution for* $\mathcal{T}$ and say that an attack $X \subseteq H^{\mathcal{T}}$ *follows* $\mathbf{x}$ if $\forall i \in \{1, \ldots, k\} \colon |X \cap H_i^{\mathcal{T}}| = \mathbf{x}_i$ holds.

If $\mathcal{T}$ has the properties given in Lemma 1 and $M^{\mathcal{T}}$ has strength $t$, then for each threshold $z \in \{1, \ldots, k\}$ the forward damage of all attacks $X \in \chi(\mathcal{T}, t)$ on $\mathcal{T}$ following the same attack distribution $\mathbf{x}$ is equal. Furthermore, the value of this forward damage on $\mathcal{T}$ gives a lower bound on

the average (and maximum) forward damage of attacks following $\mathbf{x}$ on other topologies from $\mathbb{T}(n, C, k)$. Consequently, for each $z \in \{1, \ldots, k\}$ and each $\mathcal{C} \in \mathbb{T}(n, C, k)$ on which attacks with distribution $\mathbf{x}$ exist, there is $Y \in \chi(\mathcal{C}, t)$ following $\mathbf{x}$ with $\mathrm{bf}^{\mathcal{T}}(X, z) \leq \mathrm{bf}^{\mathcal{C}}(Y, z)$.

If there is no attack with distribution $\mathbf{x}$ on $\mathcal{C}$, then a suitable distribution $\mathbf{x}'$ can be found by adapting $\mathbf{x}$ with regard to the number of heads available in $\mathcal{C}$. Thus, for each $\mathcal{C} \in \mathbb{T}(n, C, k)$ and each attack $X \in \chi(\mathcal{T}, t)$, we can find an attack $Y \in \chi(\mathcal{C}, t)$ creating at least the same forward damage on $\mathcal{C}$ as $X$ does on $\mathcal{T}$. Consequently, $\mathcal{T}$ is $t$-forward-stable. □

Next, we show that the matrix $M^{\mathcal{T}}$ of a forward-stable topology $\mathcal{T} \in \mathbb{T}(n, C, k)$ must necessarily be an Orthogonal Array of maximum possible strength.

**Theorem 3.** *If an* $\mathrm{OA}(n, k, C, t)$ *exists, then for every* $t'$*-forward-stable distribution topology* $\mathcal{T} \in \mathbb{T}(n, C, k)$ *with* $t' \geq t$, $M^{\mathcal{T}}$ *is an* $\mathrm{OA}(n, k, C, t)$.

*Proof.* Topology $\mathcal{T}$ must have the properties listed in Lemma 1. Furthermore, assume that $\mathcal{T}$ is not an $\mathrm{OA}(n, k, C, t)$. If $\mathcal{T}$ were $t'$-forward-stable, it had to minimize maximum forward-damage for attacks of cardinality $t$ and threshold $z = t$. We show that under the above assumption, this is not the case. For this, let $\mathcal{C} \in \mathbb{T}(n, C, k)$ be a topology with the properties listed in Lemma 1 and $M^{\mathcal{C}}$ being an $\mathrm{OA}(n, k, C, t)$ (the existence of $\mathcal{C}$ is guaranteed by Lemma 3).

Set $z = t$ and study the possible forward-damage of attacks of cardinality $t$. Due to Lemma 2, it suffices to consider attacks removing only heads. Such attacks may target heads from less than $t$ different stripes. This would lead to forward-damage of 0 on both $\mathcal{T}$ and $\mathcal{C}$ since they have inner-node disjoint stripes. Alternatively, attacks can target one head from each stripe of a combination of $t$ stripes. In this case, the maximum possible forward-damage on $\mathcal{T}$ and $\mathcal{C}$ equals the maximum row frequency in $M^{\mathcal{T}}$ resp. $M^{\mathcal{C}}$ over all possible restrictions to $t$ columns (cmp. Equation (6)). An attack achieving this damage contains the heads corresponding to the entries in the respective columns of the most frequent row vector. Since $\mathcal{C}$ is an $\mathrm{OA}(n, k, C, t)$ but $\mathcal{T}$ is not, this frequency is smaller on $\mathcal{C}$ than on $\mathcal{T}$. Hence, $\mathcal{T}$ is not $t$-forward-stable and, thus, not $t'$-forward-stable. □

Summing up, this subsection has shown that – given the basic properties identified in Lemma 1 – the forward-stability of a distribution topology $\mathcal{T} \in \mathbb{T}(n, C, k)$ is characterized by its matrix $M^{\mathcal{T}}$. In particular, if an $\mathrm{OA}(n, k, C, t)$ exists, it is necessary and sufficient that $M^{\mathcal{T}}$ is such

an Orthogonal Array to obtain a $t$-forward-stable topology. To reach a maximum level of forward-stability, $M^{\mathcal{T}}$ must be an Orthogonal Array of maximum possible strength $t$. This observation is used in Subsection 4.3 to provide a notion of the computational complexity of finding forward-stable distribution topologies.

### 4.3 Hardness of Finding Forward-Stable Topologies

For given parameters $n, C, k \in \mathbb{N}$, let $\hat{\mathrm{t}}(n, C, k)$ be the maximum value $t$ such that an $\mathrm{OA}(n, k, C, t)$ exists. If $\hat{\mathrm{t}}(n, C, k)$ is efficiently computable, it is also possible to use binary search to efficiently determine extremal values for the parameter $k$ of Orthogonal Arrays.

However, resolving the computational complexity of finding such extremal parameters and finding Orthogonal Arrays featuring them are long-standing open problems in design theory (cmp. [8, p.32]). A special case in coding theory is the *MDS conjecture* [2,11] which claims to specify the maximum length of MDS codes. Its disputed part was first stated in 1955 [12].

We show that finding an efficient construction strategy for $t$-forward-stable distribution topologies would resolve many of the above questions.

**Theorem 4.** *Let $\mathcal{O}$ be an oracle returning a $t$-forward-stable topology $\mathcal{T} \in \mathbb{T}(n, C, k)$ on input $(n, k, C, t)$ if one exists.*

- *If one exists, an $\mathrm{OA}(n, k, C, t)$ can be constructed by one call to $\mathcal{O}$ plus $O(nk)$-time post-processing.*
- *The function $\hat{\mathrm{t}}(n, k, C)$ can be evaluated by $\lceil \log(k) \rceil$ calls to $\mathcal{O}$ plus $O(n^2 k)$-time post-processing.*

*Proof.* Due to the Theorems 2 and 3, there is a $t$-forward-stable topology $\mathcal{T} \in \mathbb{T}(n, C, k)$ if an $\mathrm{OA}(n, k, C, t)$ exists. In this case, $M^{\mathcal{T}}$ must be an $\mathrm{OA}(n, k, C, t)$. Using input $(n, k, C, t)$, such a $\mathcal{T}$ is obtained by one call to $\mathcal{O}$. The information necessary to return the $n \times k$ matrix $M^{\mathcal{T}}$ can be gathered by a traversal of all stripe trees. This needs time $O(nk)$.

Applying binary search, we need $\lceil \log(k) \rceil$ oracle calls to find the maximum $t' \in \{0, \ldots, k\}$ such that a $t'$-forward-stable topology $\mathcal{T} \in \mathbb{T}(n, C, k)$ exists. By Theorem 3, $M^{\mathcal{T}}$ must be an $\mathrm{OA}(n, k, C, \hat{\mathrm{t}}(n, k, C))$. The strength of $M^{\mathcal{T}}$ can be determined in time $O(n^2 k)$. □

In the light of these results, the goal of identifying efficient construction schemes for forward-stable distribution topologies turns out to be a challenging task. Advancements would lead to a breakthrough in multiple connected fields of research.

Until then, it is possible to make use of the large number of constructions and catalogues for Orthogonal Arrays that are already available [8]. However, most of them are specific for certain parameter combinations and not of provably maximum strength. All algorithmic approaches known to the author that try to find Orthogonal Arrays with given parameters rely on metaheuristics and local search schemes (e.g., [13,14]).

## 5 Conclusion

In this document, we studied multitree data distribution topologies aiming to minimize the maximum number of nodes that can be isolated by an attack. In particular, this minimization should hold for every possible number of removed nodes and every level of redundancy in data encoding. We introduced the notion of forward-stable multitree data distribution topologies and showed that they closely approximate this goal if the number of nodes considerably exceeds the number of possible source neighbors. This is a common condition in applications of the studied topologies.

We found basic requirements for forward-stable distribution topologies and pointed out that the resilience of topologies adhering to these requirements is captured by a matrix representation of their heads' forward successor sets. We showed that such a topology is $t$-forward-stable if its matrix is an Orthogonal Array of strength $t$. Furthermore, the use of Orthogonal Arrays of maximum strength is necessary for forward-stable topologies. This result allowed to connect the problem of finding forward-stable topologies to long-standing open problems in design and coding theory.

Since for higher numbers of nodes, attack-resilient and forward-stable topologies must be very similar, this also provides a notion of hardness of finding attack-resilient distribution topologies. The identified topologies and results are relevant for data distribution applications such as peer-to-peer live streaming systems. Furthermore, the studied model could also be applied to certain data aggregation tasks in wireless sensor networks.

## References


1. Goyal, V.: Multiple description coding: compression meets the network. IEEE Signal Proc. Mag. **18**(5) (September 2001) 74–93
2. MacWilliams, F.J., Sloane, N.J.A.: The Theory of Error-Correcting Codes. North-Holland Mathematical Library (1993)

3. Brinkmeier, M., Schaefer, G., Strufe, T.: Optimally DoS Resistant P2P Topologies for Live Multimedia Streaming. IEEE T. Parall. Distr. **20**(6) (2009) 831–844
4. Castro, M., Druschel, P., Kermarrec, A.M., Nandi, A., Rowstron, A., Singh, A.: Splitstream: high-bandwidth multicast in cooperative environments. SIGOPS Oper. Syst. Rev. **37** (October 2003) 298–313
5. Padmanabhan, V.N., Wang, H.J., Chou, P.A., Sripanidkulchai, K.: Distributing streaming media content using cooperative networking. In: NOSSDAV '02, New York, NY, USA, ACM (2002) 177–186
6. Grau, S., Fischer, M., Schäfer, G.: On the Dependencies between Source Neighbors in Optimally DoS-stable P2P Streaming Topologies. In: IEEE International Conference on Distributed Computing Systems 2011. ICDCS (2011) 121–130
7. Dán, G., Fodor, V.: Stability and performance of overlay multicast systems employing forward error correction. Perform. Eval. **67** (2010) 80–101
8. Hedayat, A.S., Sloane, N.J.A., Stufken, J.: Orthogonal Arrays: Theory and Applications. Springer-Verlag, New York (1999)
9. Diestel, R.: Graph Theory. Third edn. Volume 173 of Graduate Texts in Mathematics. Springer-Verlag, Heidelberg (2005)
10. Grau, S.: On the Stability of Distribution Topologies in Peer-to-Peer Live Streaming Systems. PhD thesis, Technische Universität Ilmenau, Germany (2012)
11. Roth, R.M.: Introduction to Coding Theory. Cambridge University Press (2006)
12. Segre, B.: Curve razionali normali e k-archi negli spazi finiti. Ann. Math. Pura Appl. (39) (1955) 357–359
13. Nguyen, N.K., Liu, M.Q.: An algorithmic approach to constructing mixed-level orthogonal and near-orthogonal arrays. Comput. Stat. Data An. **52** (2008) 5269–5276
14. Xu, H.: An Algorithm for Constructing Orthogonal and Nearly Orthogonal Arrays with Mixed Levels and Small Runs. Technometrics **44** (2002) 356–368