


Brief Announcement: Let It TEE: Asynchronous Byzantine Atomic Broadcast with $n \geq 2f + 1$

Marc Leinweber  

Institute of Information Security and Dependability (KASTEL), Karlsruhe Institute of Technology (KIT), Germany

Hannes Hartenstein  

Institute of Information Security and Dependability (KASTEL), Karlsruhe Institute of Technology (KIT), Germany

Abstract

Asynchronous Byzantine Atomic Broadcast (ABAB) promises simplicity in implementation as well as increased performance and robustness in comparison to partially synchronous approaches. We adapt the recently proposed DAG-Rider approach to achieve ABAB with $n \geq 2f + 1$ processes, of which f are faulty, with only a constant increase in message size. We leverage a small Trusted Execution Environment (TEE) that provides a unique sequential identifier generator (USIG) to implement Reliable Broadcast with $n > f$ processes and show that the quorum-critical proofs still hold when adapting the quorum size to $\lfloor \frac{n}{2} \rfloor + 1$. This first USIG-based ABAB preserves the simplicity of DAG-Rider and serves as starting point for further research on TEE-based ABAB.

2012 ACM Subject Classification Security and privacy → Distributed systems security

Keywords and phrases Byzantine Fault Tolerance, Trusted Execution Environments, Asynchrony

Digital Object Identifier 10.4230/LIPIcs.DISC.2023.43

Funding This work was supported by funding from the topic Engineering Secure Systems of the Helmholtz Association (HGF).

1 Introduction

Atomic Broadcast primitives play a crucial role for Byzantine-fault tolerant (BFT) State Machine Replication (SMR). A prominent example for BFT SMR in the partially synchronous model is PBFT [6]. By use of small Trusted Execution Environments (TEE) that generate and sign unique sequential identifiers on each process, called USIGs, Veronese et al. [12] showed that PBFT's communication complexity can be reduced and the fault tolerance can be increased to $n \geq 2f + 1$ while still tolerating Byzantine faults. The authenticity/integrity of TEEs can be verified remotely and, thus, TEEs are assumed to only fail by crashing. However, as shown by Miller et al. [11], Asynchronous Byzantine Atomic Broadcast (ABAB) outperforms approaches based on the partially synchronous model particularly under faults and tends to show a simpler design. While it is known that any asynchronous crash fault-tolerant algorithm can be compiled to withstand Byzantine faults using TEEs [3, 7], the proposed compilers show either a polynomial or an exponential overhead in runtime. We are interested in a simple and straightforward design of a USIG-enhanced ABAB that does not add further message rounds and only adds a constant number of bits to each message (essentially a counter value and a signature). To this end, we adapt DAG-Rider [10] to provide ABAB with $n \geq 2f + 1$ processes. We give a quick recap on DAG-Rider and explain the adaption *TEE-Rider*. Besides using TEE-based Reliable Broadcast and changing the required quorums from $2f + 1$ to $\lfloor \frac{n}{2} \rfloor + 1$, we leave DAG-Rider unchanged. We show that the quorum-based arguments of DAG-Rider still hold for TEE-Rider.



© Marc Leinweber and Hannes Hartenstein;
licensed under Creative Commons License CC-BY 4.0
37th International Symposium on Distributed Computing (DISC 2023).
Editor: Rotem Oshman; Article No. 43; pp. 43:1–43:7



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

2 TEE-Rider: Transforming DAG-Rider to $n \geq 2f + 1$

We make use of the following definition of Atomic Broadcast for a set of processes $P, n := |P|$. The processes communicate over authenticated point-to-point links with eventual delivery.

► **Definition 1** (Atomic Broadcast). *Each process $p_i \in P$ receives client transactions t via events `clientRequest(t)`. Correct processes deliver tuples (t, r, p_i) , where t is a client transaction, $r \in \mathbb{N}_0$ a round number, and $p_i \in P$ the process that initially received t , satisfying the following properties:*

Agreement: *If a correct process $p_i \in P$ delivers (t, r, p_j) , then every other correct process $p_k \in P, k \neq i$ eventually delivers (t, r, p_j) with probability 1.*

Integrity: *For each round $r \in \mathbb{N}_0$ and process $p_j \in P$, a correct process $p_i \in P$ delivers (t, r, p_j) at most once.*

Validity: *If a correct process $p_i \in P$ receives an event `clientRequest(t)`, then every correct process $p_k \in P$ eventually delivers (t, r, p_i) with probability 1.*

Total Order: *Let m_1 and m_2 be any two valid tuples that are delivered by any two correct processes $p_i, p_j \in P$. If p_i delivers m_1 before m_2 , then p_j delivers m_1 before m_2 .*

2.1 Changes in Assumptions, Building Blocks, and Setup

In addition to the assumptions of DAG-Rider, we assume that each process is equipped with a USIG [12] that may only fail by crashing. It implements a signature service that binds a unique counter value to each signature it produces. The USIG is used for Reliable Broadcast with a fault tolerance of $n > f$ as, e.g., implemented in [8, Algorithm 1]: it is a “single echo” algorithm with USIG-signed messages and attached counter value. Correct processes relay a message once and reject messages with invalid USIG signatures or with counter values already received which prevents equivocating messages for the same counter. An instance of the Reliable Broadcast abstraction has two functions: `broadcast(r, m)` to reliably broadcast exactly one arbitrary message m for round r to all processes in P , and `delivered()` which returns all messages that were received by the instance since the last call to `delivered()`. We expect the Reliable Broadcast abstraction to fulfill the following properties:

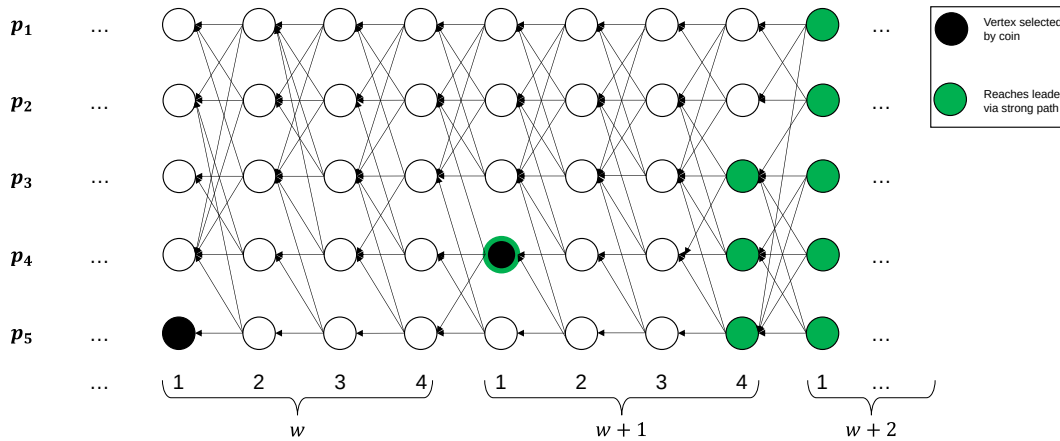
► **Definition 2** (Reliable Broadcast). *A sender $p_s \in P, n := |P|$ can call `broadcast(c, m)`. Correct processes deliver (c, m) where $c \in \mathbb{N}_0$ and m an arbitrary message satisfying the following properties:*

RB-Agreement: *If a correct process $p_i \in P$ delivers (c, m) , then every other correct process $p_k \in P, k \neq i$ eventually delivers the same (c, m) .*

RB-Integrity: *For each $c \in \mathbb{N}_0$, a correct process $p_k \in P$ delivers (c, m) at most once.*

RB-Validity: *If a correct sender calls `broadcast(c, m)`, then every correct process $p_i \in P$ eventually delivers (c, m) .*

Additionally, we assume an asynchronous common coin, e.g. as defined by Cachin et al. [5], that produces a uniformly distributed common random number p out of $\{p \mid p \in \mathbb{N}_0 : p < n\}$ for all correct processes and a name $i \in \mathbb{N}_0$ as soon as $f + 1$ processes invoked `toss(i)`; repetitive calls with same the i yield the same p . We further assume a trusted setup of the common coin and the USIGs using a public key infrastructure (to set up the common coin’s threshold signature scheme, dealerless variants [4] and those with an asynchronous setup [1] exist).



■ **Figure 1** Example DAG for $n = 5$ processes of which at maximum 2 may be faulty. Shown is the “global” state of the graph, i.e., after every process eventually received every vertex. For simplicity, all vertices are valid and weak edges are left out. The direct commit rule is not fulfilled for any process for wave w ; it is fulfilled for processes p_3, p_4 , and p_5 for wave $w + 1$. The green coloring highlights the effect of the direct commit rule as proven in Lemma 3. The direct commit rule ensures that a correct process can commit a wave retrospectively if it was not able to commit when it finished the wave. Since the leader vertex of wave $w + 1$ has a strong path to the leader vertex of wave w , wave w will be committed retrospectively.

2.2 The Algorithm

DAG-Rider uses n Reliable Broadcast instances to disseminate process messages and to construct a directed acyclic graph (DAG) that captures the communication history of all processes. In a second step, each process derives consensus on the order of transactions using a Common Coin based on the graph structure. The adapted DAG-Rider algorithm executed by a correct process $p_i \in P$ is shown in Algorithm 1. Utility functions are listed in Algorithm 2. The core of the approach is the construction and interpretation of a (local) DAG that captures received transactions and the observed communication sequence between processes. The DAG is structured in rounds and a round contains at maximum one vertex per process, i.e., n vertices. Rounds are addressed in an array style and the local view of a process p_i on the DAG is indicated by an index i . The very first round $DAG_i[0]$ is initialized with n hard-coded “genesis” vertices. A vertex in round r has two types of edges: strong edges point to vertices of round $r - 1$ and weak edges point to vertices of any round $r' \leq r - 2$. As soon as p_i received $\lfloor \frac{n}{2} \rfloor + 1$ *valid* vertices for a round r , i.e., $\lfloor \frac{n}{2} \rfloor + 1$ vertices referencing $\lfloor \frac{n}{2} \rfloor + 1$ vertices of round $r - 1$ as strong edges (*v.strong*, l. 11), for which it also knows its predecessors (l. 15), p_i will *complete* round r and transition to round $r + 1$. Now, as soon as p_i receives a client transaction, it will become the payload of a vertex v which is created and broadcast by p_i for round $r + 1$ (ll. 44 and 21-27). The vertex v connects to all vertices p_i received for round r (l. 23). If p_i received vertices u for older rounds that are not reachable from the newly created vertex using the transitive closure of strong and weak edges (a “path”), u will become a weak edge of v (*v.weak*, l. 26). The new vertex is broadcast using Reliable Broadcast instance i to all processes (l. 27). Every fourth round a so-called wave, consisting of four rounds, is completed (l. 19) and the DAG structure is used to derive a total order on the transactions (ll. 28-42). Each wave w has exactly one wave leader v which is chosen calling *coin.toss(w)* from the vertices of w ’s first round($w, 1$). The random number is used to select the process whose vertex is to be used as wave leader. If v was not

(yet) received or there are no $\lfloor \frac{n}{2} \rfloor + 1$ vertices in w 's fourth round($w, 4$) that have v in their transitive closure of strong edges (a “strong path”), i.e. the *direct commit rule* is not fulfilled, the wave cannot be committed (l. 30). If wave w can be committed, process p_i checks first if there are wave leaders of waves w' between the last wave that was committed (variable *decidedWave*) and the current wave w that were received in the meantime and are connected to the leader of the wave $w' + 1$ (retrospective commit, ll. 33-36). The wave leaders are used as the root for a deterministic graph traversal to determine the total order of transactions (ll. 38-42). An example for a resulting graph with $n = 5$ processes, i.e. $f \leq 2$, is shown in Figure 1.

3 Correctness Argument

Lemmas 1 and 2 of the original DAG-Rider publication [10] are crucial for Total Order and Agreement and rely on quorum intersection arguments. The following Lemmas 3 and 4 show the corresponding results for a quorum size of $\lfloor \frac{n}{2} \rfloor + 1$. Results for Integrity and Validity simply follow from the original paper.

► **Lemma 3.** *If a correct process $p_i \in P$ commits the wave leader v of a wave w when it completes wave w in round($w, 4$), then any valid vertex v' of any process $p_j \in P$ broadcast for a round $r \geq \text{round}(w + 1, 1)$ will have a strong path to v .*

Proof. Since p_i commits v in round($w, 4$), the direct commit rule is fulfilled (l. 30): $\exists U \subseteq \text{DAG}_i[\text{round}(w, 4)]: |U| \geq \lfloor \frac{n}{2} \rfloor + 1 \wedge \forall u \in U: \text{strongPath}(u, v)$. A valid vertex must reference at least $\lfloor \frac{n}{2} \rfloor + 1$ distinct vertices of the previous round with a strong edge (l. 11). Thus, a process $p_j \in P$ broadcasting a valid vertex v_j for round($w + 1, 1$) selected at least $\lfloor \frac{n}{2} \rfloor + 1$ vertices of round($w, 4$) as strong edges for v_j . Any two subsets of size $\lfloor \frac{n}{2} \rfloor + 1$ of a superset of size n intersect at least in one element. Thus, every valid vertex of a process broadcast for round($w + 1, 1$) must have at least one edge to a vertex of U , and, via U to v . As every valid vertex of round($w + 1, 1$) has a strong path to v , and every valid vertex of round($w + 1, 2$) connects to at least $\lfloor \frac{n}{2} \rfloor + 1$ vertices of round($w + 1, 1$), by induction, any valid vertex v' of any process $p_j \in P$ broadcast for a round $r \geq \text{round}(w + 1, 1)$ has a strong path to v . ◀

► **Lemma 4.** *When a correct process $p_i \in P$ completes round($w, 4$) of wave w , then $\exists V_1 \subseteq \text{DAG}_i[\text{round}(w, 1)], V_4 \subseteq \text{DAG}_i[\text{round}(w, 4)]: |V_1| \geq \lfloor \frac{n}{2} \rfloor + 1 \wedge |V_4| \geq \lfloor \frac{n}{2} \rfloor + 1 \wedge (\forall v_1 \in V_1, \forall v_4 \in V_4: \text{strongPath}(v_4, v_1))$.*

Proof. By use of Reliable Broadcast and validity checks in ll. 11 and 15, faulty processes are limited to omission faults. Thus, the *get-core* argument of Attiya and Welch [2, Sec. 14.3.1] still holds [2, Sec. 14.3.3]: Let $A \in \{0, 1\}^{n \times n}$ be a matrix that contains a row for each possible vertex of round($w, 3$) and a column for each possible vertex of round($w, 2$). Let $A[j, k] = 1$ if the vertex of process p_j of round($w, 3$) has a strong edge to the vertex of process p_k of round($w, 2$) or p_j sends no vertex (or an invalid one) but p_k sends a valid vertex for round($w, 2$). As there are at least $\lfloor \frac{n}{2} \rfloor + 1 \leq n - f$ correct processes, each row of A contains at least $\lfloor \frac{n}{2} \rfloor + 1$ ones and A contains at least $n(\lfloor \frac{n}{2} \rfloor + 1)$ ones. Since there are n columns, there must be a column l with at least $\lfloor \frac{n}{2} \rfloor + 1$ ones. This implies there is a vertex v_l by process p_l in round($w, 2$) s.t. $\exists V_3 \subseteq \text{DAG}_i[\text{round}(w, 3)]: |V_3| \geq \lfloor \frac{n}{2} \rfloor + 1 \wedge \forall v_3 \in V_3: \text{strongPath}(v_3, v_l)$. As at most f vertices in V_3 belong to faulty processes that may commit send omission faults for round($w, 3$) and $\lfloor \frac{n}{2} \rfloor + 1 \geq f + 1$, by quorum section at least one vertex of V_3 is received by any correct process $p_j \in P$ before it sends its vertex for round($w, 4$). Thus, every valid vertex in $\text{DAG}_i[\text{round}(w, 4)]$ has at least one strong edge to a vertex of V_3 . Since v_l must be valid

■ **Algorithm 1** TEE-Rider pseudocode for process $p_i \in P, n := |P|, n \geq 2f + 1$.

```

1: state DAG: array of sets of vertices, DAG[0] initialized with “genesis” vertices
2: state r:  $\mathbb{N}$ , initialized with 1
3: state decidedWave:  $\mathbb{N}_0$ , initialized with 0
4: state transactionsToPropose: queue of client transactions t, initialized empty
5: state buffer: set of vertices, initialized empty
6: state rb: array of n Reliable Broadcast instances with delivered() and broadcast(r, m)
7: state coin: common coin instance with toss(w)
8: while True do
9:   for k  $\leftarrow$  0 up to n - 1 do
10:    for m = (r', v)  $\in$  rb[k].delivered() do
11:     if  $|v.\text{strong}| < \lfloor \frac{n}{2} \rfloor + 1$  then continue
12:     v.source  $\leftarrow$  pk; v.round  $\leftarrow$  r'; v.delivered  $\leftarrow$  False
13:     buffer.add(v)
14:   for v  $\in$  buffer do
15:    if v.round > r  $\vee \exists u \in v.\text{strong} \cup v.\text{weak}: u \notin \cup_{r' \geq 0} \text{DAG}[r']$  then continue
16:    DAG[v.round].add(v)
17:    buffer.remove(v)
18:   if  $|\text{DAG}[r]| < \lfloor \frac{n}{2} \rfloor + 1$  then continue
19:   if r mod 4 = 0 then waveReady( $\frac{r}{4}$ )
20:   r  $\leftarrow$  r + 1
21:   wait until  $\neg \text{transactionsToPropose.is Empty}()$ 
22:   v  $\leftarrow$  new vertex
23:   v.block  $\leftarrow$  transactionsToPropose.dequeue(); v.strong  $\leftarrow$  DAG[r - 1]
24:   for r'  $\leftarrow$  r - 2 down to 1 do
25:    for u  $\in$  DAG[r'] do
26:     if  $\neg \text{path}(v, u)$  then v.weak.add(u)
27:   rb[i].broadcast(r, v)
28: function waveReady(w)
29:   v  $\leftarrow$  getWaveLeader(w)  $\triangleright \perp$  if round(w, 1) vertex of chosen process is not in DAG
30:   if  $v = \perp \vee |\{u \mid u \in \text{DAG}[\text{round}(w, 4)]: \text{strongPath}(u, v)\}| < \lfloor \frac{n}{2} \rfloor + 1$  then return
31:   leadersStack  $\leftarrow$  new stack; leadersStack.push(v)
32:   for w'  $\leftarrow$  w - 1 down to decidedWave + 1 do
33:    u  $\leftarrow$  getWaveLeader(w')
34:    if  $u \neq \perp \wedge \text{strongPath}(v, u)$  then
35:     leadersStack.push(u); v  $\leftarrow$  u
36:   decidedWave  $\leftarrow$  w
37:   while  $\neg \text{leadersStack.is Empty}()$  do
38:    v  $\leftarrow$  leadersStack.pop()
39:    verticesToDeliver  $\leftarrow$   $\{u \mid u \in \cup_{r' > 0} \text{DAG}[r']: \text{path}(v, u) \wedge \neg u.\text{delivered}\}$ 
40:    for u  $\in$  verticesToDeliver in deterministic order do
41:     u.delivered  $\leftarrow$  True; deliver (u.block, u.round, u.source)
42: upon clientRequest (t)
43:   transactionsToPropose.enqueue(t)

```

Algorithm 2 Utility functions pseudocode.

```

1: function path( $v, u$ ) : boolean
2:   return exists a sequence of vertices  $(v_1, v_2, \dots, v_k) \in \cup_{r' \geq 0} DAG[r']$  such that
3:    $v_1 = v \wedge v_k = u \wedge \forall i \in [2, k]: v_i \in v_{i-1}.strong \cup v_{i-1}.weak$ 
4: function strongPath( $v, u$ ) : boolean
5:   return exists a sequence of vertices  $(v_1, v_2, \dots, v_k) \in \cup_{r' \geq 0} DAG[r']$  such that
6:    $v_1 = v \wedge v_k = u \wedge \forall i \in [2, k]: v_i \in v_{i-1}.strong$ 
7: function getWaveLeader( $w$ ) : vertex or  $\perp$ 
8:    $p_j \leftarrow coin.toss(w)$ 
9:   if  $\exists v \in DAG[\text{round}(w, 1)]: v.source = p_j$  then return  $v$ 
10:  return  $\perp$ 
11: function round( $w, i$ ) :  $\mathbb{N}$ 
12:  return  $4(w - 1) + i$ 

```

and thus has a strong edge to each vertex of a set $V_1 \subseteq DAG_i[\text{round}(w, 1)], |V_1| \geq \lfloor \frac{n}{2} \rfloor + 1$, any valid vertex of rounds $r \geq \text{round}(w, 4)$ has a strong path to every vertex, including V_1 , reached by v_i via strong paths. Please note that the construction of the set V_1 is valid for all correct processes that complete the wave and, thus, represents the “common core”. ◀

4 Discussion and Conclusion

DAG-Rider shows the power of causal order broadcast to implement consensus. The adaption for TEEs preserves the simplicity of DAG-Rider while increasing its fault tolerance and reducing the communication effort (i.e., from “double echo” to “single echo” Reliable Broadcast). The ease of adaption of DAG-Rider for TEEs make it a perfect textbook example for TEE-based ABAB. Follow-up work to DAG-Rider, Tusk [9], addresses a major deployability issue, namely garbage collection, shortens the wave length, and replaces the underlying Reliable Broadcast with a communication scheme that leverages the graph structure to achieve linear communication complexity in the happy case. Additionally, to the best of our knowledge, there exists no TEE-based, dealerless, and asynchronous common coin primitive. In summary, investigating a TEE-based dealerless setup as well as transforming the follow-ups of DAG-Rider for empirical studies to investigate the assumed superiority of asynchronous TEE-based approaches, e.g., in comparison to MinBFT [12], is a promising line of research.

References

- 1 Ittai Abraham, Marcos Kawazoe Aguilera, and Dahlia Malkhi. Fast asynchronous consensus with optimal resilience. In Nancy A. Lynch and Alexander A. Shvartsman, editors, *Distributed Computing*, volume 6343 of *Lecture Notes in Computer Science*, pages 4–19. Springer, 2010. doi:10.1007/978-3-642-15763-9_3.
- 2 Hagit Attiya and Jennifer L. Welch. *Distributed computing - fundamentals, simulations, and advanced topics (2. ed.)*. Wiley, 2004. doi:10.1002/0471478210.
- 3 Naama Ben-David, Benjamin Y. Chan, and Elaine Shi. Revisiting the power of non-equivocation in distributed protocols. In Alessia Milani and Philipp Woelfel, editors, *PODC '22: ACM Symposium on Principles of Distributed Computing, Salerno, Italy, July 25 - 29, 2022*, pages 450–459. ACM, 2022. doi:10.1145/3519270.3538427.

- 4 Alexandra Boldyreva. Threshold signatures, multisignatures and blind signatures based on the gap-diffie-hellman-group signature scheme. In Yvo Desmedt, editor, *Public Key Cryptography — PKC 2003*, volume 2567 of *Lecture Notes in Computer Science*, pages 31–46. Springer, 2003. doi:10.1007/3-540-36288-6_3.
- 5 Christian Cachin, Klaus Kursawe, and Victor Shoup. Random oracles in constantinople: Practical asynchronous byzantine agreement using cryptography. *J. Cryptol.*, 18(3):219–246, 2005. doi:10.1007/s00145-005-0318-0.
- 6 Miguel Castro and Barbara Liskov. Practical byzantine fault tolerance and proactive recovery. *ACM Trans. Comput. Syst.*, 20(4):398–461, 2002. doi:10.1145/571637.571640.
- 7 Allen Clement, Flavio Junqueira, Aniket Kate, and Rodrigo Rodrigues. On the (limited) power of non-equivocation. In Darek Kowalski and Alessandro Panconesi, editors, *ACM Symposium on Principles of Distributed Computing, PODC '12, Funchal, Madeira, Portugal, July 16-18, 2012*, pages 301–308. ACM, 2012. doi:10.1145/2332432.2332490.
- 8 Miguel Correia, Giuliana Santos Veronese, and Lau Cheuk Lung. Asynchronous byzantine consensus with $2f+1$ processes. In Sung Y. Shin, Sascha Ossowski, Michael Schumacher, Mathew J. Palakal, and Chih-Cheng Hung, editors, *Proceedings of the 2010 ACM Symposium on Applied Computing (SAC), Sierre, Switzerland, March 22-26, 2010*, pages 475–480. ACM, 2010. doi:10.1145/1774088.1774187.
- 9 George Danezis, Lefteris Kokoris-Kogias, Alberto Sonnino, and Alexander Spiegelman. Narwhal and Tusk: a DAG-based mempool and efficient BFT consensus. In Yérom-David Bromberg, Anne-Marie Kermarrec, and Christos Kozyrakis, editors, *EuroSys '22: Seventeenth European Conference on Computer Systems, Rennes, France, April 5 - 8, 2022*, pages 34–50. ACM, 2022. doi:10.1145/3492321.3519594.
- 10 Idit Keidar, Eleftherios Kokoris-Kogias, Oded Naor, and Alexander Spiegelman. All you need is DAG. In Avery Miller, Keren Censor-Hillel, and Janne H. Korhonen, editors, *PODC '21: ACM Symposium on Principles of Distributed Computing, Virtual Event, Italy, July 26-30, 2021*, pages 165–175. ACM, 2021. doi:10.1145/3465084.3467905.
- 11 Andrew Miller, Yu Xia, Kyle Croman, Elaine Shi, and Dawn Song. The honey badger of BFT protocols. In Edgar R. Weippl, Stefan Katzenbeisser, Christopher Kruegel, Andrew C. Myers, and Shai Halevi, editors, *Proc. 2016 ACM SIGSAC Conf. on Computer and Communications Security, Vienna, Austria, 2016*, pages 31–42. ACM, 2016. doi:10.1145/2976749.2978399.
- 12 Giuliana Santos Veronese, Miguel Correia, Alysson Neves Bessani, Lau Cheuk Lung, and Paulo Veríssimo. Efficient byzantine fault-tolerance. *IEEE Trans. Computers*, 62(1):16–30, 2013. doi:10.1109/TC.2011.221.