Discourse on the attack vectors that arise from the Rules of "The DAO 1.0"

THIS IS A WIP DRAFT (v0.1). PLEASE TO NOT DISTRIBUTE WITHOUT PERMISSION.
Please Send comments on content and formatting to (dino at smartwallet dot org)

CONTENTS:

FOREWARD:

Over the past 3 weeks, a Distributed Autonomous Organization (DAO) that exists in a Smart Contract on the Ethereum blockchain has 'raised' 11.5 million Ether, an amount equal to $161 million at the time of writing. This particular DAO creatively decided to call itself "The DAO, a name that is either ambitious or presumptuous depending on your point of view. Yet, it is one that has lived up to its name. So far.

In this discourse we analyze the structure of 'The DAO', outline biases that arise from technical implementation details of the said structure, and present potential technical and social solutions to remedy them.

PART 1: Introduction and definition of terms (Rules, Actors, Classes)

First, we start by identifying the two major factors that are responsible for how money moves out of 'The DAO'.

Those factors are i) The Rules and ii) The Actors.

The Rules of "The DAO" were defined by the programmers who wrote the solidity code of its smart contract. In general, a DAO is likely to operate successfully if the programmers make good Rules. The Rules of 'The DAO' will be described in English in PART 2.

Next, we define the Actors. Actors are entities that can interact with the DAO in a finite set of ways that are based on the Rules. For the purposes of this post we assume they are human beings who belong to one of the three following classes:

1. The Technical Class

This is the class of token holders who have read and understood the actual Solidity smart contract that controls The DAO (located at https://github.com/TheDAO/DAO-1.0/blob/master/DAO.sol). We assume this class makes up a very small minority of participants.

*2. The Semi-informed Class***

This class of token holders understands, in laymans terms, the description of the Rules of the DAO as described to them by the technical class. Thus, they must rely on the trustworthiness of of the technical class to tell them the truth on how the smart contract operates (i.e. what the Rules are), and also be confident that the technical class has made no errors themselves.

3. The Naive Class

This class of token holders neither understands the technical implementations nor is semi-informed. One of the motivations for the Naive class to buy tokens is "fear of missing out". They see the first two classes doing buying tokens, and assume that since the other classes are more informed they must be making an educated bet, so they are just following on. Others in this class may also be uninformed short term speculators (See Wikipedia: Crowd Psychology)

The types of individuals within Classes:

In any of the three classes, there may exist the following individual types: Nominal, Whale, and Attacker.

Nominal individuals are those whose actions have no net bias, assuming all else equal, and thus may be found in any class.

Whales are individuals who hold a large amount of tokens and thus control a large amount of votes. They are more likely to be found in the Technical or Semi-Informed class, and less likely to be part of the Naive class. Whales can either

1) Act beneficially to The DAO by providing additional liquidity for good investments, and act rationally in the long run to add any net bias.

2) Act as an attacker who attempts to rob the bank (i.e. try to use a large voting block to pay themselves out on a bad proposal.) Even without an outright majority, Whales could still create a cartel of other whales or large holders to join them in a vote to rob the Naive class, who are least likely to vote against them.

3) Act as defense against an attacker.

Attackers are individuals who have identified weaknesses in the Rules themselves (i.e. a code exploit), or seek to take advantage of structural and economic weaknesses due to a set of Rules that result in biased voting, or some other outcome that allows them to steal or extort money from other actors (i.e. the Stalker attack). They are most likely to be part of the Technical Class. An Attacker who is also a Whale (an ***Attacking Whale***) is a

particularly dangerous threat to any DAO that has implemented bad Rules.

PART 2: Description of the Rules of The DAO.

Creation Phase

- The DAO creation phase is 27 days long. During this period, 'The DAO' Tokens (TDT) can be purchased by sending Ether to the following Ethereum address: 0xbb9bc244d798123fde783fcc1c72d3bb8c189413
- The price of TDT varies during the creation phase. First, it starts at 1.00 Ether for 100TDT for the first 14 days. Then, an increase of 0.05Ether per 100TDT for the following 10 days, then a final 3 day period at 1.50Ether per 100TDT.
- Latecomers who paid more than 1.00Ether per 100TDT have their additional Ether above 1.00 locked up as a 'late penalty' in the 'extraBalance' account. For example, if a token holder paid 1.05 Ether for 100TDT, they can only withdraw 1.00 Ether. The extra 0.05 Ether will stay locked in the extraBalance and can only be moved after an amount equal to the extraBalance has been spent on proposals. Thus, individual token holders can't withdraw funds from the extraBalance account.

Withdrawing Funds

- Token holders can convert their TDT into Ether by 'Splitting' from the DAO, a process that takes 34 days in total to complete. The amount of Ether they receive per TDT is equal to 1 divided by the total balance of Ether left in the DAO, minus the Ether in the extraBalance account. Thus, token holders that successfully split from The DAO before any proposals have been funded will receive 1.00ETH per 100TDT.
- All TDT that is converted back into Ether is destroyed.

The Curator

- The Curator account is responsible for adding addresses to the 'allowedRecipients' whitelist. (Currently the 'Curator' account is actually a multi-signature address with keys held by 10 individuals (LINK TO CURATOR LIST)
- Addresses on the whitelist are the only addresses that can be funded by the DAO. Proposals that want funding from the DAO must to ask the Curator to add their address to the whitelist.

Proposals and Voting

- When a proposal has its address whitelisted by the Curator, token holders can then vote on whether or not they want to fund the proposal.
- The only voting types are YES and NO.

- There is a minimum voting period of 14 days.
- A simple majority of YES votes is required for a proposal to be successfully funded.
- If a token holder votes either yes or no on a proposal, they cannot withdraw their Ether by splitting from the DAO until the voting period has ended. If the proposal succeeds, they can only withdraw their share of whatever the remaining Ether balance is left in The DAO after the proposal has been funded.
- Token holders that don't vote can withdraw their Ether up to 7 days before any proposal voting ends, without any risk that some of their Ether will be spent to fund the proposal.
- A minimum quorum of voters is required in order for a vote to be valid. The minimum quorum varies between 20 and 53% depending on the size of the proposal. Very large proposals will require a 53% quorum, while small ones only need 20%. The exact formula for the minimum quorum can be found in the white paper (LINK TO WHITE PAPER).
- There is no limit to how many proposals can be active at one time. In order to prevent 'proposal spam' there is a non-refundable listing fee of X Ether.

Changing the Curator, Splitting, and Withdrawing

- Any token holder can initiate a vote at any time to change the Curator. This voting period lasts for 7 days.
- If the vote fails, the token holder that initiated the vote can decide to Split from The DAO by themselves, while taking the Ether equivalent of all the TDT that voted YES with them. If they elect to do this, a new DAO will be created that will hold their Ether, and Curator of the new DAO will be the address they used when they initiated the vote.
- When a token holder splits from The DAO through the above mechanism, there is a 27-day creation period for the new DAO. This means that it actually takes 7+27=34 days in total to initiate a Curator Change vote, split from The DAO, and then wait for the new DAO to be formed.
- When a token holder has successfully split into their own new DAO, they can create a proposal to pay themselves out the full balance of all the Ether left in the new DAO.

Transferability of TDT.

- TDT are fully transferrable to any valid Ethereum address, and therefore can be sold immediately on exchanges or over the counter. Thus, if a token holder does not want to wait 34 days to Split from the DAO and withdraw their Ether, they can just sell their TDT tokens directly on exchanges for Ether, or perhaps even Bitcoin.

— ANY MORE RULES TO ADD?

PART 3: How Rational Actors in each Class will act, based on the characteristics of various proposals, and how attackers can exploit those actions.

In order to determine how rational Actors will behave, we define and describe different types of Proposals and their properties.

First, we assume all proposals have the following properties:

1. (**etherCost**) -The amount of Ether that the DAO will fund proposal with. This property is defined by the proposal itself. For example, a proposal may ask for 1000 Ether to make 1000 T-Shirts.

2. (**etherReturn**) - The amount of Ether that the proposal will return to the DAO. This property will usually be estimated by the proposal itself, and will also be debated by token holders. For example, a proposal may estimate that they will sell 1000 T-Shirts at a profit of 5 Ether each, and thus estimate they will return 5000 Ether to The DAO.

3. (**P_success**)The chance that the proposal succeeds in returning the amount of ether determined by property #2. This property is the most difficult to determine, and will be subject to all sorts of debate by token holders during the debating and voting periods.

Next, we use these properties to define types of proposals

Positive EV Proposals
This type of proposal is one that creates positive expected value (+EV) for The DAO. A proposal is +EV if EtherReturn*P_success > EtherCost. Rational actors will vote YES on proposals they believe are +EV.

Negative EV Proposals
This type of proposal is one that creates negative expected value (-EV) for The DAO. A proposal is -EV if either EtherReturn < EtherCost, or EtherReturn*P_success < EtherCost. Rational actors SHOULD vote NO on Proposals they believe are -EV, but because of the structure of the Rules, they are at a disadvantage to do so. Since all token holders who vote on a proposals are cannot split from the DAO until the outcome of the vote is determined, they would be better off splitting from the DAO with their Ether risk free, as opposed to voting NO and taking the risk that the proposal will succeed. If the proposal fails, they can always buy back TDT on exchanges later.

Absurd Proposals
This type of proposal has either etherReturn < etherCost, or a very low value of P_success.

Attacking Proposals

An Attacking Proposal is a special type of Absurd proposal that has more YES votes than NO votes during any phase of the voting period. Attacking Proposals can manifest themselves in three different types of attacks; The Robbery Attack, the Token-Value Attack, the extraBalance Attack.

The Robbery Attack is one in which an Attacking Whale that votes YES stands to benefit from the Ether that will be sent to the Attacking proposal address if the proposal is successful. This type of attack is very difficult to detect because Attacking Whales will only vote YES at the very last minute, leaving no time for The DAO token holders to withdraw their funds, and leaving them blindsided that such an Absurd proposal was actually funded. We have identified a potential Attacking Whale who invested 888,888 Ether into The DAO from the following address: 0x04c973aff06f64b880524f16ae8c821928233ee5 . This Whale currently owns 7.7% of all outstanding votes in The DAO. This means that in a vote that achieves only a 20% minimum quorum, this Whale already has 77% of the required YES votes to pass the proposal. The Whale would only need to conspire with 3.3% of remaining token holders to vote YES on an Absurd proposal, in return for paying the conspirators out from the stolen funds. It is also possible that this Whale also controls a number of smaller addresses.

The Token-Value Attack is one in which an Attacking Whale stands to benefit by driving the the TDT below book value, and then purchasing them in the open market. A Token-Value attack is most successful if the Attacker can i) Incentivize a large portion of token holders NOT to split, but instead sell their TDT directly on exchanges, and ii) incentivize a large portion of the public NOT to purchase TDT on exchanges. An Attacker can achieve (i) by implementing the Stalker attack on anyone who splits, and then making that attack public on reddit, the forums, and in the media. The Attacker does not even need to do the stalker attack on a real person, but could make many fake accounts and have them all post “OH NOES!?! I got stalked!! 0mg guys, don’t try to split from this attack, just sell your TDT on Poloniex as fast you you can!! omogmogm!!111!”. Even though the Stalker attack can be mitigated and the victim can eventually recover their money, at the time of writing, only the Technical Class is able to defend themselves from a Stalker attack, as no GUI tools have been written to help the Semi-Informed and Naive classes defend themselves. This alone is enough to make (i) quite effective.

An Attacking Whale can achieve (ii) By making an Absurd proposal, waiting for the 6th day before voting ends, and then voting YES on it with a large block of votes. At this point no rational market actors would then want to buy TDT tokens, as they would not be able to make any arbitrage profits by converting the TDT back into Ether because it takes 7 days to split; the attacking proposal will end in 6 days, and if it succeeds it will be too late to split, and the Ether will be gone. The combined result of (i) and (ii) means that the asks on the order book will be very heavy, and the bids will be very light. The net result of this is that the TDT will trade well below book value. The Attacking Whale can

then buy up all the TDT on exchanges for a risk free profit, because the attacking whale is the only TDT buyer who has no risk if the Attacking proposal actually manages to pass.

The extraBalance Attack is one in which an Attacking Whale tries to scare all token holders into splitting from The DAO so that book value of TDT increases. The book value of TDT increases because token holders who split can not recover any extraBalance, so as holders split, the extraBalance becomes a larger percentage of the total balance, thus increasing the book value of the TDT. Currently the extraBalance is $3,000,000, which means the book value of TDT should be 1.02. If the Attacker can scare away half the token holders, the TDT will increase in value to 1.04. If the Attacker can scare away 95% of the token holders, the book value of the remaining TDT will be 2.00. In this attack, the Attacking Whale would do the opposite of the Token-Value Attack by trying to incentivize token holders to split. This can be achieved by creating an Absurd proposal and then immediately voting YES on it with a large voting block of TDT, scaring all the token holders, and then giving them 14 days until the end of the voting period so they have more than enough time to split. In this scenario, since there is more than enough time to split from The DAO, rational Actors will be much better off splitting than voting NO, since splitting will be risk free, and voting NO will result in losses if the attackers have enough yes votes.

PART 4: Potential solutions to mitigate these attacks.

Implement a post-vote grace period.
This would allow any token holders a grace period of n days to withdraw their Ether in the case that an Absurd proposal succeeds.

Implement a new voting option called 'NO_AND_WITHDRAW_IF_VOTE_SUCCEEDS'
This option is likely better than the grace period option, as it would give more information to all the other DAO token holders on the intentions of each voter.

Rational YES votes believe a proposal is +EV.

Rational NO votes believe the proposal is -EV, BUT are still willing to stay in the DAO even if it passes, perhaps because they have voted YES on a different proposal that they believe would make up for the losses of the -EV proposal.

Rational NO_AND_WITHDRAW_IF_VOTE_SUCCEEDS votes indicate that the voter believes this proposal would cause long-term damage to the value of the TDT and no longer wants to be part of The DAO if it succeeds.

Implement instant, direct withdrawals to Ether addresses
Rather than forcing TDT holders to go through the complicated and convoluted splitting

process, allow them to withdraw Ether instantly and directly to any Ethereum address. This change would neutralize most of the economics behind the Token-Value attack.

PART 5: Suggestions to "The DAO" token holders and the Curators.

Given the above discourse, we believe it would be irresponsible for the Curators to whitelist any proposals until the DAO is upgraded to "Version 1.1" that implements improvements to the Rules in order to mitigate the potential attacks described in this paper. We suggest that the Curators reach a consensus on the way forward, and make an announcement to all "The DAO" token holders as soon as possible.

It is our opinion that if no actions are taken to mitigate any of the attacks described in this paper, all "The DAO" token holders should split from the DAO immediately and not participate in voting on any proposals.

NOTE: THIS IS NOT FINANCIAL ADVICE. DON'T LISTEN TO US. WE ARE NOT, AND WILL NOT BE HELD RESPONSIBLE FOR YOUR FINANCIAL DECISIONS.

Special thanks to Vlad Zamfir, who spent quite some time discussing this paper.

If you have any comments or questions on this paper please direct them to dino at smartwallet dot org