Soulbound

2022 Jan 26 See all posts

One feature of World of Warcraft that is second nature to its players, but goes mostly undiscussed outside of gaming circles, is the concept of *soulbound* items. A soulbound item, once picked up, cannot be transferred or sold to another player.

Most very powerful items in the game are soulbound, and typically require completing a complicated quest or killing a very powerful monster, usually with the help of anywhere from four to thirty nine other players. Hence, in order to get your character anywhere close to having the best weapons and armor, you have no choice but to participate in killing some of these extremely difficult monsters yourself.

The purpose of the mechanic is fairly clear: it keeps the game challenging and interesting, by making sure that to get the best items you have to actually go and do the hard thing and figure out how to kill the dragon. You can't just go kill boars ten hours a day for a year, get thousands of gold, and buy the epic magic armor from other players who killed the dragon for you.

Of course, the system is very imperfect: you could just pay a team of professionals to kill the dragon with you and let you collect the loot, or even outright buy a character on a secondary market, and do this all with out-of-game US dollars so you don't even have to kill boars. But even still, it makes for a much better game than every item always having a price.

What if NFTs could be soulbound?

NFTs in their current form have many of the same properties as rare and epic items in a massively multiplayer online game. They have social signaling value: people who have them can show them off, and there's more and more tools precisely to help users do that. Very recently, Twitter started rolling out an integration that allows users to show off their NFTs on their picture profile.

But what exactly are these NFTs signaling? Certainly, one part of the answer is some kind of skill in acquiring NFTs and knowing which NFTs to acquire. But because NFTs are tradeable items, another big part of the answer inevitably becomes that NFTs are about signaling wealth.

CryptoPunks are now regularly being sold for many millions of dollars, and they are not even the most expensive NFTs out there. Image source here.

If someone shows you that they have an NFT that is obtainable by doing X, you can't tell whether they did X themselves or whether they just paid someone else to do X. Some of the time this is not a problem: for an NFT supporting a charity, someone buying it off the secondary market is sacrificing their own funds for the cause and they are helping the charity by contributing to others' incentive to buy the NFT, and so there is no reason to discriminate against them. And indeed, a lot of good can come from charity NFTs alone. But what if we want to create NFTs that are not just about who has the most money, and that actually try to signal something else?

Perhaps the best example of a project trying to do this is POAP, the "proof of attendance protocol". POAP is a standard by which projects can send NFTs that represent the idea that the recipient personally participated in some event.

Part of my own POAP collection, much of which came from the events that I attended over the years.

POAP is an excellent example of an NFT that works better if it could be soulbound. If someone is looking at your POAP, they are not interested in whether or not you paid someone who attended some event. They are interested in whether or not *you personally* attended that event. Proposals to

put certificates (eg. driver's licenses, university degrees, proof of age) on-chain face a similar problem: they would be much less valuable if someone who doesn't meet the condition themselves could just go buy one from someone who does.

While transferable NFTs have their place and can be really valuable on their own for supporting artists and charities, there is also a large and underexplored design space of what *non-transferable* NFTs could become.

What if governance rights could be soulbound?

This is a topic I have written about ad nauseam (see [1] [2] [3] [4] [5]), but it continues to be worth repeating: **there are very bad things that can easily happen to governance mechanisms if governance power is easily transferable**. This is true for two primary types of reasons:

- If the goal is for governance power *to be widely distributed*, then transferability is counterproductive as concentrated interests are more likely to buy the governance rights up from everyone else.
- If the goal is for governance power *to go to the competent*, then transferability is counterproductive because nothing stops the governance rights from being bought up by the determined but incompetent.

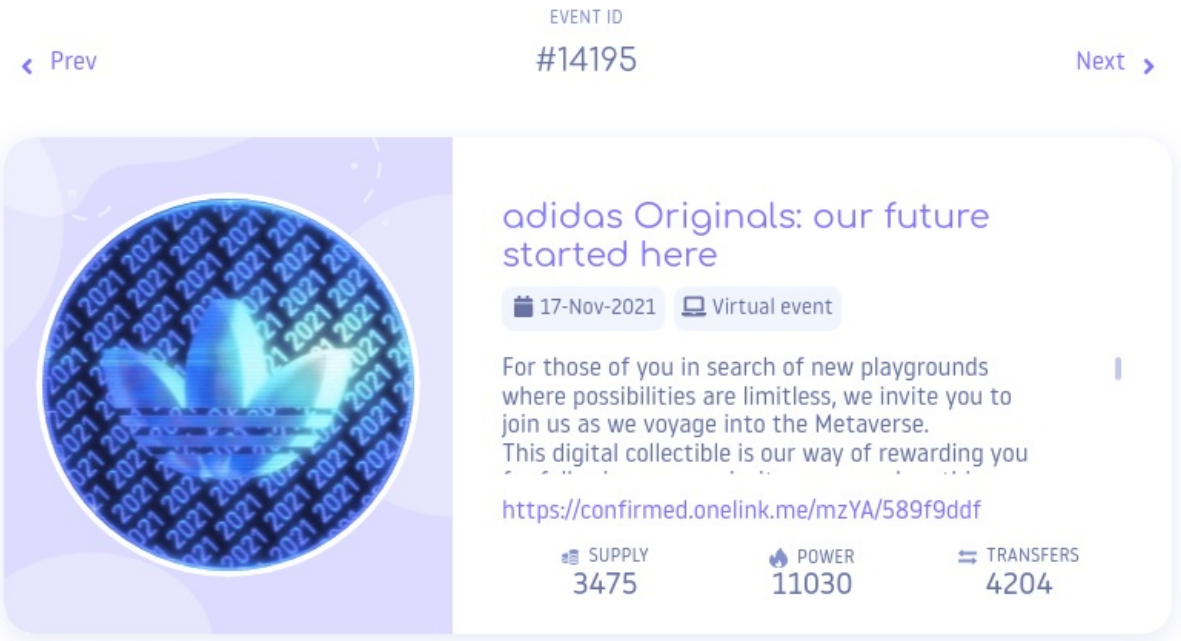
If you take the proverb that "those who most want to rule people are those least suited to do it" seriously, then you should be suspicious of transferability, precisely because transferability makes governance power flow away from the meek who are most likely to provide valuable input to governance and toward the power-hungry who are most likely to cause problems.

So what if we try to make governance rights non-transferable? What if we try to make a CityDAO where more voting power goes to the people who actually live in the city, or at least is reliably democratic and avoids undue influence by whales hoarding a large number of citizen NFTs? What if DAO governance of blockchain protocols could somehow make governance power conditional on participation? Once again, a large and fruitful design space opens up that today is difficult to access.

Implementing non-transferability in practice

POAP has made the technical decision to not block transferability of the POAPs themselves. There are good reasons for this: users might have a good reason to want to migrate all their assets from one wallet to another (eg. for security), and the security of non-transferability implemented "naively" is not very strong anyway because users could just create a wrapper account that holds the NFT and then sell the ownership of that.

And indeed, there have been quite a few cases where POAPs have frequently been bought and sold when an economic rationale was there to do so. Adidas recently released a POAP for free to their fans that could give users priority access at a merchandise sale. What happened? Well, of course, many of the POAPs were quickly transferred to the highest bidder.



More transfers than items. And not the only time.

To solve this problem, the POAP team is suggesting that developers who care about non-transferability implement checks on their own: they could check on-chain if the current owner is the same address as the original owner, and they could add more sophisticated checks over time if deemed necessary. This is, for now, a more future-proof approach.

Perhaps the one NFT that is the most robustly non-transferable today is the proof-of-humanity attestation. Theoretically, anyone can create a proof-of-humanity profile with a smart contract account that has transferable ownership, and then sell that account. But the proof-of-humanity protocol has a revocation feature that allows the original owner to make a video asking for a profile to be removed, and a Kleros court decides whether or not the video was from the same person as the original creator. Once the profile is successfully removed, they can re-apply to make a new profile. Hence, if you buy someone else's proof-of-humanity profile, your possession can be very quickly taken away from you, making transfers of ownership non-viable. Proof-of-humanity profiles are de-facto soulbound, and infrastructure built on top of them could allow for on-chain items in general to be soulbound to particular humans.

Can we limit transferability without going all the way and basing everything on proof of humanity? It becomes harder, but there are medium-strength approaches that are probably good enough for some use cases. Making an NFT bound to an ENS name is one simple option, if we assume that users care enough about their ENS names that they are not willing to transfer them. For now, what we're likely to see is a spectrum of approaches to limit transferability, with different projects choosing different tradeoffs between security and convenience.

Non-transferability and privacy

Cryptographically strong privacy for transferable assets is fairly easy to understand: you take your coins, put them into tornado.cash or a similar platform, and withdraw them into a fresh account. But how can we add privacy for soulbound items where you cannot just move them into a fresh account or even a smart contract? If proof of humanity starts getting more adoption, privacy becomes even more important, as the alternative is all of our activity being mapped on-chain directly to a human face.

Fortunately, a few fairly simple technical options are possible:

- Store the item at an address which is the hash of (i) an index, (ii) the recipient address and (iii) a secret belonging to the recipient. You could reveal your secret to an interface that would then scan for all possible items that belong to your, but no one without your secret could see which items are yours.
- Publish a hash of a bunch of items, and give each recipient their Merkle branch.

- If a *smart contract* needs to check if you have an item of some type, you can provide a ZK-SNARK.

Transfers could be done on-chain; the simplest technique may just be a transaction that calls a factory contract to make the old item invalid and the new item valid, using a ZK-SNARK to prove that the operation is valid.

Privacy is an important part of making this kind of ecosystem work well. In some cases, the underlying thing that the item is representing is already public, and so there is no point in trying to add privacy. But in many other cases, users would not want to reveal everything that they have. If, one day in the future, being vaccinated becomes a POAP, one of the worst things we could do would be to create a system where the POAP is automatically advertised for everyone to see and everyone has no choice but to let their medical decision be influenced by what would look cool in their particular social circle. Privacy being a core part of the design can avoid these bad outcomes and increase the chance that we create something great.

From here to there

A common criticism of the "web3" space as it exists today is how money-oriented everything is. People celebrate the ownership, and outright waste, of large amounts of wealth, and this limits the appeal and the long-term sustainability of the culture that emerges around these items. There are of course important benefits that even financialized NFTs can provide, such as funding artists and charities that would otherwise go unrecognized. However, there are limits to that approach, and a lot of underexplored opportunity in trying to go beyond financialization. Making more items in the crypto space "soulbound" can be one path toward an alternative, where NFTs can represent much more of who you are and not just what you can afford.

However, there are technical challenges to doing this, and an uneasy "interface" between the desire to limit or prevent transfers and a blockchain ecosystem where so far all of the standards are designed around maximum transferability. Attaching items to "identity objects" that users are either unable (as with proof-of-humanity profiles) or unwilling (as with ENS names) to trade away seems like the most promising path, but challenges remain in making this easy-to-use, private and secure. We need more effort on thinking through and solving these challenges. If we can, this opens a much wider door to blockchains being at the center of ecosystems that are collaborative and fun, and not just about money.