You Only Live Once Online (YOLOO): Privacy-Preserving Solution to Block and Ban Evasion

Sean Hackett
CPSC 610
December 18, 2019

Abstract
Users of today's online social networks rarely triumph over harassment and trolling. The tools for protecting oneself from unwanted interaction, namely blocking and banning, are easily evaded. This paper tracks the history of these tools back to the multi-user online worlds of the early 90s to demonstrate that block and ban evasion much a problem today as they were 25 years prior. By surveying recent incidents of cyber-harassment, I argue for the importance of effective banning and blocking. I propose an alternative to real-name registration that lever's David Chaum's blind signatures achieve privacy-preserving real-name registration. This model registration model is used to restrict users to only one account (or some small, finite number of accounts) on an online platform. With one account per user, blocking and banning decisions are tied to real-world identities, even though the real-world identity associated with any online account cannot be involuntarily revealed.

PART I: BACKGROUND

The Toad That Wasn't

This virtual world, LambdaMOO, invited users to interact in a grid of connecting virtual rooms. Avatars could move between rooms, speak, emote, and create and manipulate objects. The text-only interface streamed text down one's screen, reading like a screenplay written in real-time. In the spring of 1993, some thirty virtual avatars convened in a private chamber [1]. Their topic of discussion was governance [1]. Three days prior, a user committed a virtual atrocity, later recounted in Julian Dibbell's "A Rape in Cyberspace" [1]. Decisions were to be made on the fate of the aggressor and on the handling of future misbehavior [1].

The polity considered a range of replies to cyber-rape, up to involving real-world police to bring the miscreant, whose IP was traced to an NYU computer, to justice [1]. The users would narrow their consideration to virtually enforced punishments [1]. At the time, everyday users had just one method for restricting their interaction with others: gagging. Invoking the @gag command, any action or speech by a specified user would become invisible to oneself. The wizardry, LambdaMOO's aristocracy, could invoke the special @toad command to permamently destroy a character.

On the evening of the town hall, a wizard present toaded the perpetrator [1]. Discussions of cyber-governance would go on, but Mr. Bungle, the pseudonym of the aggressing user, was gone. No command - @gag, @toad, or otherwise – would, however, prevent Mr. Bungle's reincarnation. The person behind Mr. Bungle's actions remained untouched by the toading. He would return to LambdaMOO, joining the world as easily as any new user, this time choosing the alias Dr. Jest [1]. And Dr. Jest would go on to relive Mr. Bungle's legacy, terrorizing the community until his eventual toading [1]. To some in the community, the readily evaded toading process had failed the community, leaving it at the mercy of Mr. Bungle, to reincarnate and terrorize as he pleased.

Blocking and Banning Today

The online social networks (OSNs) of today, including Facebook and Twitter, dole out punishments as impermanent and easily evaded as those in the virtual world of LambdaMOO. The Mr. Bungles of today may have a different modus operandi. They are more likely to dox a female user or threaten her with real-world rape rather than simulate a rape online. All the same, they inflict real emotional pain and go further to threaten and incite physical harm [2]. It is with the same ease that they evade efforts to limit their interaction on social networks. In the more than twenty-five years since Mr. Bungle's rampage, the social networks have not adapted to give users fuller reign over their online interactions.

OSNs today typically offer users with two methods for protecting themselves from interaction with another user: blocking and banning. Blocking, analogous to @gag on LambdaMOO and implemented in Twitter, Facebook, Instagram, and Reddit, allows users to shield themselves from activity by another user. This may mean shielding oneself from private messages sent by another user or shielding oneself from all posts made by another user, depending on the platform and the user's preference. Whether or not a blocked user is made aware of the block also varies by platform. Banning, analogous to @toad on LambdaMOO, is done through more formal methods on today's OSNs than on LambdaMOO. Users on the four mentioned platforms can submit reports of abuse through easily accessed forms. The platforms have employees tasked with reading reports of abuse. If it is determined that a user violated the terms of service of the platform, the employees may temporarily suspend or permanently ban the transgressing account.

Creating accounts on Twitter, Facebook, Instagram, and Reddit requires little effort. Reddit asks only for your desired username and password. Twitter, Facebook, and Instagram all require a valid email, while Twitter additionally requires a valid phone number. Per its official policy, Facebook requires that users use their legal names on the platform. No evidence of using one's legal name is required to create an account, however. Further, though it violates the terms and services of all of the platforms, you can purchase accounts on the platforms. These accounts may be aged for upwards of three years and can come with thousands of followers, in the case of Twitter, or thousands in Karma, in the case of Reddit, all of which would help to protect against automatic detection as a sockpuppet account. The lax account creation policies make blocking and banning ineffective measures for preventing harassment on OSNs and feed two forms of harassment in particular: chaining and dogpiling.

Chaining

Chaining occurs when a person uses a series of accounts to harass a target [3]. Chaining may be seen as synonymous with block or ban evasion. Bailey Poland, author of *Haters: Harassment, Abuse, and Violence Online*, was the subject of repeated harassment by a single actor. She writes of her experience:

> For more than a year I have received periodic rape and death threats from one anonymous individual, always on Twitter. I have reason to believe I know who it is, but his accounts are usually deleted or suspended too quickly for Twitter to track….After a year without me responding to him directly, he escalated the behavior, publicly posting what he believed was my home address on the Gamergate hashtag and inviting people to come rape me. Shortly after that he posted an additional rape threat accompanied by a photograph taken of me then I was in the eighth grade [4].

Poland's experience is shared by other female users. A series of surveys and interviews conducted on women in the UK in 2015 documents additional accounts of harassment via chaining:

> It was defamatory and aimed at getting me fired. It was also relentless from a person I have blocked. Felt invasive and intrusive - they are monitoring me even though I've blocked them. (Respondent 67)
>
> He sets up accounts to start discussions under pseudonyms so he can abuse me, incites other people to abuse me, emails to tell me he is watching me. (Respondent 116) [2]

This form of harassment is made possible by block and ban evasion. Poland and Respondents 67 and 116 are not concerned with particular Twitter pseudonyms - they are concerned with the humans behind many pseudonyms. Security for these users means not superficial blocking and banning at the level of pseudonym but at the level of the human actor.

Dogpiling

Dogpiling occurs when many harassers team up to carry out a harassment campaign [3]. This is what happened in #GamerGate, a harassment campaign leveled against females in the video gaming community [5]. Beginning in August 2014, the movement focused on female video game developer Zoë Quinn [5]. Before the campaign, Quinn began recording all instances of harassment directed towards her online profiles [5]. By December 2014, this record had increased 1,000 fold in size as her accounts were floaded with violent and vitriolic posts [5].

Intuitively, dogpiling is less symptomatic of the failure of blocking and banning than chaining is. In the case of Poland, for example, the more than one year for which she endured violent, sexual threats may have been cut far shorter had she been able to effectively block the single person behind the attack. With dogpiling, even if block and ban evasion were curtailed, it would still be difficult to issue blocks to communities numbering in the thousands, pouring to Twitter and Facebook from sites like 4chan and 8chan.

There is evidence, however, that dogpiling, in its infant stages, may rely on block and ban evasion. Early in the #GamerGate campaign, the organizers used a network of sockpuppet accounts to spread their message. When an account was blocked or banned, the attackers would fall back on the other accounts in their network. One month into #GamerGate, the controversy had been covered in major news outlets. It had been the subject of an open letter signed by 2,495 members of the video game community [6]. Yet, according to the messages by organizers in the 4chan logs, the campaign consisted of on the order of twenty people with many sockpuppet accounts:

> Sep 06 04.56.01 <HeWhoNods> The entire movement is like twenty people with a lot of sockpuppets
> Sep 06 04.56.13 <NoWhiteKnight> That sounds about right.
> …
> Sep 06 04.56.30 <kjk> Those 20 people are actually 3 people pretending to be 20 pretending to be a movement
> ...
> Sep 06 04.56.55 <HeWhoNods> Five guys[1] and a lot of pseudonyms [7]

1 "Five guys" is a reference to accusations made by Quinn's former boyfriend and the lead harasser in the #GamerGate campaign that Quinn cheated on him with five other persons. This particular line says nothing of the number of people involved in the movement in September 2014 but is included for its explicit mention of "pseudonyms."

If these twenty or so attackers had been effectively banned from Twitter, rather than a swath of their many pseudonyms, #GamerGate may have struggled to gain the momentum needed to spiral into the massive affair that it would become. This malicious Sybil attack was enabled by block and ban evasion. Addressing block and ban evasion would serve to make the growth of dogpiling campaigns more difficult.

Existing Solutions for Block and Ban Evasion

Delete Your Account if You're Harassed

This is an obvious and obviously unsatisfying solution for avoiding harassers who manage to evade blocks and bans. Unfortunately, this is a corner into which many victims of harassment are forced, including Zelda Williams [8].

Make Your Account Private if You're Harassed

As unsatisfying as the previous solution, making one's account private forecloses one the opportunity to interact with thousands of benevolent strangers in order to avoid interaction with perhaps just one malicious person. It also, like the previous solution, allows harassers to win over their victims, forcing victims into hiding.

Require Users to Register with Real Names

If every account is tied to a real-world identity, then users can be blocked and banned by their real-world identity rather than their pseudonym. That way, if a particular, real-world person uses more than one pseudonym, all such pseudonyms will be blocked or banned from when that person is blocked or banned. This would do away with block and ban evasion as currently seen in OSNs.

After the suicide of South Korean actress Choi Jin-sil, victim of malicious rumors circulated online, in 2007, South Korea took measures to counter cyberbulling events like this and the earlier Dog Poop Girl [9]. The government's Communications Commission introduced the "Internet real-name system," which required websites with more than 300,000 users to collect the real name and state identification number of anyone posting on their site [10]. The system aimed to hold miscreants legally accountable for their online posts. Though it was not directly intended to improving blocking, the system allows for this beneficial side effect.

Real-name registration is appealing as a quick fix for online harassment. It allows users to better control their online interactions with effective blocking and it empowers law enforcement to hold harassers legally accountable. The system comes at the cost of anonymity and pseudonymity. A proper treatment of the value of anonymity and pseudonymity online is beyond the scope of this paper and subject to variation depending on one attitude towards privacy. It is worth noting, in passing, that the communities which are subject to online harassment are, in some cases, the same communities that rely on anonymity and pseudonymity to thrive. A study of anonymous Twitter accounts, for example, finds that a higher proportion of the accounts following popular gay and lesbian Twitter accounts are anonymous than accounts following popular accounts in "non-sensitive" categories [11].

PART II: INTRODUCING YOLOO

Overview

The purpose of the You Only Live Once Online (YOLOO) is to associate every online account, on some OSN, with exactly one real-world person and to associate every real-world person with no more than one online account. The key to block and ban evasion is the ability to switch to another account when your current account is blocked or banned. YOLO improves upon real-name registration by maintaining anonymity and pseudonymity. Although real-world identification is required to create an account, YOLOO guarantees users that their identities will not be linkable to their accounts.

With the anonymous and pseudonymous accounts of today's OSNs, online punishments like blocking and banning occur on the level of pseudonyms. With real-name registration, the pseudonym is stripped back, revealing the true identity of the person behind the account. The punishments then happen on the level of this true identity. YOLOO offers a solution somewhere between these two. YOLOO goes beyond pseudonyms but stops short of pulling back the figurative mask that a person might wear when expressing herself online.

This section works its way slowly to a definition of the YOLOO system by building and evaluating three more primitive approaches to account creation online.

Anonymous Account Creation

Creating an Account

To create an account, an applicant sends a *registration request* to the platform. The request might include the applicant's desired username and the public key or password with which the applicant wishes to log-in in the future. The platform creates a new account from the information in the request.

De-anonymizing users

Applicants never present their real-world identity to the platform. Therefore, the platform cannot de-anonymize its users.[2]

Real-Name Account Creation

Creating an Account

To create an account, an applicant sends her *identification*[3] and a *registration request* to the platform. The platform checks that the identity has not previously been used to create an account. If the identity check succeeds, the platform creates a new account from the information in the *registration request*.

De-anonymizing Users

---

2 It is assumed that users do not post personally identifiable information on the platform and take precautions to prevent tracing by IP address, browser fingerprinting, and the like.

3 It is assumed that persons have (1) an identity that can be sent to the platform and (2) a method of proving that they are the rightful holder of the identity.

The platform sees the applicant's *identification* and *registration request* at the same time. The platform can therefore save table with the applicant's account information alongside a copy of the applicant's *identification*. The platform can de-anonymize users any any time by looking up, in the table, the identity associated with an account.

Real-Name Account Creation, with Separate Identification and Registration Authorities

Creating an Account

The identification and registration processes can be separated. First, an applicant sends her *identification* and a *token* to the *identification authority*. The *identification authority* checks that the identity has not previously been used to create an account. If the identity check succeeds, the *identification authority* uses its private key to sign the *token* and returns the *signed token* to the applicant.

A *signed token* can be redeemed for exactly one account from the *registration authority*. The applicant sends the *signed token* and a *registration request* to the *registration authority.* The *registration authority* uses the public key of the *identification authority* to verify the signature. The registration authority checks that the *signed token* has not been previously redeemed. If the authentication and check succeeds, the *registration authority* creates a new account from the information in the *registration request*.

A *token* can be any string. Applicants would want to select long, random strings as their tokens to avoid collision with other users. The *identification authority* can use an RSA key-pair for signing and signature verification.

De-anonymizing Users

The *identification authority* sees an applicant's identity and *signed token*. The *registration authority* sees the applicant's *signed token* and account information. If the authorities collude or are hacked, the users can be de-anonymized, using the *signed token* as to link identity to account.

If one of the authorities is trusted to not collude and is secure against hacking, then users will be protected against de-anonymization. This protection is only as reliable as the authority is trustworthy and secure.

Privacy-Preserving Real-Name Account Creation

Account Creation

In the previous approach, the *signed token* served as a link between the applicant's identity and account information. If the *identification authority* were to never saw the *signed token*, then there would be no way to establish this link. Fortunately, blind signatures allow one to sign some data without ever seeing the data [12]. Before sending the *token* to the *identification authority*, the applicant first blinds the *token* by combining it with a *blinding factor*. The *blinding factor* is chosen arbitrarily by the applicant and the *identification authority* is therefore unable to compute the original *token* from the *blinded token* that it receives [12]. The *identification authority* then signs the *blinded token* and returns the *signed blinded token* to the applicant. The applicant combines the *blinding factor* with the *signed blinded token* to invert the earlier blinding [12]. By the design of the blind signature scheme, the unblinding

preserves the signature, giving the *signed token [12]*. The applicant can redeem the *signed token* with the registration authority to create an account.[4]

There are two properties of the blind signature process that are worth noting. First, for a particular *blinding factor*, the unblinding operation is unique [12]. There is no other operation known to the applicant that will modify the *signed blinded token* received from the *identification authority* while maintaining the signature [12]. This prevents the applicant from generating multiple *signed tokens* from a single *signed blinded token*. Second, it is practically infeasible to derive the *token* from the *blinded token* or the *blinded token* from the *token* without knowing the *blinding factor [12]*. The same is true for the *signed token* and *signed blinded token*.

Neither the role of the *identification authority* nor *registration authority* changes from the previous approach. The *identification authority* will still sign whatever is handed to it as a *token*, provided that the *identification* it receives is entitled to a new account. The *identification authority* does not care or necessarily know if the *token* has been blinded by the applicant or not. The *registration authority*, as before, receives a *signed token* and a *registration request*. The *registration authority* will verify that the *signed token* was signed by the *identification authority* and will create an account for the applicant if the *signed token* was not previously redeemed.

De-anonymizing Users

The *identification authority* sees the applicant's *identification, blinded token,* and *signed blinded token.* The *registration authority* sees the applicant's *signed token* and *registration request*. The *signed token* cannot be linked to either the *blinded token* or the *signed blinded token* by anyone other than the applicant. Therefore, the *identification authority* and *registration authority* cannot collude to de-anonymize the applicant. This means that the two authorities can merge into a single, untrusted entity, without any loss of privacy. That is, a single online social network provider can act as both *identification authority* and *registration authority* without being able to link your identity and account.

Summary

The privacy-preserving real-name account creation approach allows a platform to restrict people to only a single account, based on their real-world identity, without creating a link between real-world identities and accounts. Because the applicants for new accounts are responsible for blinding, they can guarantee, without trusting either authority, that their identity and account will not be linkable. This system also allows platforms to introduce addition criteria for acceptance to the platform, based on real-world identity. A platform might only allow persons living within a certain municipality or identifying with a certain gender identity to join the platform, as examples.

---

4 Upon unblinding the signed blinded token received from the identification authority, the applicant would have both the signed blinded token and the signed token. Both would be redeemable with the registration authority as they are random strings with valid signatures from the identification authority. This would allow an applicant to create two accounts, rather than the desired one. To avoid this, the registration authority will expect a token and the signature of the hash of that token. This will require that the applicant send the registration authority the blinded hashed token and receive back the signed blinded hashed token. Upon unblinding, the applicant will get the signed hashed token. The token and signed hashed token can be redeemed with the registration authority. To use the signed blinded hashed token, the applicant would need to known the string whose hash is the blinded hashed token. This would require computing the inverse of the hash function, which is practically infeasible. This approach therefore limits the applicant to only a single redeemable token, as desired. This approach would be necessary to properly implement YOLOO but is mentioned in the footnotes to make the main text more approachable.

PART III: ADDITIONAL CONSIDERATIONS

Account Reclamation

If a person creates an account through a platform that uses YOLOO and later has that account hacked, they may have no recourse to reclaim the account. The hacker may changed the public key used to log-into the account or the recovery email associated with the account. Though disappointing, today you might simply create a new account and start fresh. If you have already redeemed your one opportunity to create an account, however, then it would seem that you are locked out of the platform forever. The platform needs to stay true to its limits on persons creating multiple accounts else they create an opportunity for block and ban evasion.

To resolve this, the *identification authority* will issue *reclamation signatures.* The person whose account was hacked will send the *identification authority* their *identification* and same *blinded token* originally submitted when the account was created. The *identification authority* will check its records to determine that this is indeed the same *blinded token* used to create the account associated with the provided *identification*. The *identification authority* will then sign the *blinded token* with the private key of key-pair distinct from the key-pair used in the creation of new accounts. The former key-pair can be called the reclamation key-pair and the latter can be called the creation key-pair. After receiving and unblinding the *signed blinded token* sent back from the *identification authority*, the person will send the *signed token* to the *registration authority*. The *registration authority* will use the public reclamation key to verify the signature. If the signature checks out, the *registration authority* will allow the person to regain full control of the account associated with the provided *token*. This will allow the person to reset the password or upload a new public key to use in authentication.

This account reclamation process ensures that the identification process trumps the authentication process that the platform uses to grant a user access to an account. So long as someone can prove their identity, they can maintain control of their account, even if the account is subject to hacking. If their identity is stolen, depending on the nature of the identity, they will presumably have some recourse outside of the platform to reclaim that identity.

Account Renewal

In addition to having tokens re-signed when one's account is stolen, a platform might require that users get their tokens re-signed on a regular basis. A platform, for example, might require that every user submit a re-signed token at least once per year to keep their account from being deactivated. The *identification authority* will use a different key-pair for this renewal process than it does for creation and reclamation.

Consider a university forum. Students graduate every year from the university and should lose access to the forum. We would not want to require all students to recreate their pseudonymous accounts annually, as they may have built up credibility in the forum. Instead, we require that users submit their *blinded tokens* for renewal signing once per year by the *identification authority's* renewal private key. The *identification authority* will verify that the person submitting the *blinded token* should still have access to the forum in the upcoming year. After unblinding, the person will submit the *signed token* to the *registration authority,* which will make a note to not delete the account associated with the token.

Membership Leak

Users of YOLOO can be guaranteed that their identities will not be traced to their accounts. They do risk revealing their involvement with whichever platform with which they are creating an account. If the platform leaked the identities used to create accounts, even though it might be impossible for someone to find a particular person's account on the platform, it would still become known that the person uses the platform. Knowing that someone uses Twitter may not mean much. Knowing that someone uses Grindr, a social networking and dating app for the LGBTQ community, could be devastating for a person who is not open about their sexual orientation or gender identity.

Assuming different identities in different virtual communities can be important for free expression. You might not want your account on a networking site like LinkedIn to be attached to your account on a dating site, like Tinder or eHarmony. There is a trade-off to be made, however. Granular accounts allows for freer expression within each community but results in more detailed membership leak. Multi-purpose accounts leads to more restrained expression in each community but less membership leak.

Time-Stamp Attack

When an applicant for a new account contacts the *identification authority* and then the *registration authority*, the two authorities do not receive information that allows them to link the contact together. If the applicant contacts the *identification authority* and then, only a matter of seconds later, contacts the *registration authority*, the two authorities may be able to determine that they were in contact with the same person. This would allow the authorities to the the person's account to her identity, de-anonymizing her.

To address this, the applicant might wait a random amount of time after receiving her *signed blinded token* from the *identification authority* to contact the *registration authority*. Based on average rate of account creation, the applicant can choose the upper bound of her waiting time to agree with her desired degree of k-anonymity. If 1,000 new accounts are created each month and the user wishes be to 3,000-anonymous, then the user should randomly select a time between 0 and 3 months to wait before registering with the *registration authority*.

Conclusion

The Internet is an outlet for free expression and personal experimentation that might not otherwise be possible in the non-virtual world. The promise of the Internet is lost, however, for those subject to persistent harassment online. It is important to recognize that this harassment is often done at the hands of a few. This paper presents an approach for ensuring that these few are prevented from terrorizing victims with unwanted interaction and, importantly, ensuring that anonymity and pseudonymity are still protected.

Works Cited

[1] J. Dibbell, “A rape in cyberspace or how an evil clown, a Haitian trickster spirit, two wizards, and a cast of dozens turned a database into a society,” *Ann Surv Am L*, p. 471, 1994.

[2] R. Lewis, M. Rowe, and C. Wiper, “Online abuse of feminists as an emerging form of violence against women and girls,” *Br. J. Criminol.*, vol. 57, no. 6, pp. 1462–1481, 2016.

[3] J. Matias, A. Johnson, W. E. Boesel, B. Keegan, J. Friedman, and C. DeTar, “Reporting, reviewing, and responding to harassment on Twitter,” *Available SSRN 2602018*, 2015.

[4] B. Poland, *Haters: Harassment, abuse, and violence online*. U of Nebraska Press, 2016.

[5] Z. Jason, “Game of fear,” *Boston Mag.*, 2015.

[6] A. Zecher, “Open letter to the gaming community,” *Andreas Zecher*, 2014.

[7] C. Johnston, “Chat logs show how 4chan users created# GamerGate controversy,” *Ars Tech.*, 2014.

[8] S. Rosenbloom, “Dealing with digital cruelty,” *N. Y. Times*, vol. 23, 2014.

[9] C. Sang-Hun, “Korean star’s suicide reignites debate on web regulation,” *N. Y. Times Oct.*, vol. 13, 2008.

[10] C. Sang-Hun, “South Korean Court Rejects Online Name Verification Law,” *Retrieved Novemb.*, vol. 30, p. 2011, 2012.

[11] S. T. Peddinti, K. W. Ross, and J. Cappos, “On the internet, nobody knows you’re a dog: A Twitter case study of anonymity in social networks,” in *Proceedings of the second ACM conference on Online social networks*, 2014, pp. 83–94.

[12] D. Chaum, “Blind signatures for untraceable payments,” in *Advances in cryptology*, 1983, pp. 199–203.