2FA's costly misnomers and misconceptions

Eugene Shablygin[1], Sergey Bratus[2]

September 1, 2017

Executive summary

Two-factor authentication as currently implemented by the overwhelming majority of security products incorporates a number of weakening misconceptions. In this white paper we list these misconceptions and then discuss a way to win past them to more secure authentication schemes.

1. *E-mail should NOT be the identity presented for authentication!* E-mail is a public address of the person, used to contact the person, and left throughout a person's public Internet footprint, where it can be easily harvested by attackers. Using it as the person's identity empowers attackers to find the victim's other accounts and interactions, and to engage the victim directly.
2. *Identity should be neither public nor human-readable!* When the identity to be authenticated is a public or easily guessable string, the attacker is thereby empowered to engage the authentication system in the first step of the authentication process. There is no reason to allow this.
3. *Authentication should NOT start with the weakest and most easily compromised factor like a public or guessable username and a password.* Instead, it should start with the strongest factor, which allows the attacker least interaction with the system before being rejected. Username-and-password pairs have been given away by users to attackers over and over again, through a variety of ruses; these ruses will not stop so long as users fall for pixel-perfect copies of the login pages—that is, not anytime soon.
4. *2nd factor should NOT be an account recovery scheme!* The purpose of the "2nd factor" is NOT to stop the compromise of an identity for which the user accidentally gave away the password, or lost the password. Using it so implies that the primary factor is already easily compromised. Why use such a weak primary factor in the first place? This is poor design.

1 WWPass Corp., gene@wwpass.com

2 Dartmouth College, sergey@cs.dartmouth.edu

5. *A "first factor" credential that the user can unwittingly lose to attackers is a BAD primary credential!* When users must practice unclear, intuitive caution to not give away a credential, the game is already half lost. Why bring your weakest game with a stronger fallback if you don't mean to first lose and then salvage the game?
6. *Authentication is NOT "layered defense"!* The principle of layered defense is a favorite one in security, but it does not apply to authentication. A successful cryptographic system seeks to minimize the attacker's opportunities for interacting with it, because any allowed interaction can serve to weaken the overall scheme, e.g., by providing an oracle (cf. the so-called Cryptographic Doom Principle). Authentication must follow the same principle of minimizing attacker interaction and rejecting it early, with the best available strength.

Attacks of the 2016 election cycle, such as compromising DNC's Gmail accounts, and subsequent attacks on Google Docs showed what dedicated low-skill, low-tech attackers can achieve against an organization that uses a typical web mail or web app setup with usernames and passwords. The verdict: absolutely everything.

The main lesson of these attacks is, if users can give away their identity by some GUI action, some user will give it away. For attackers to gain access to your trove of documents, one is enough.

The take-away for defensive design is obvious: no action users can take by clicking or typing should be able to give away their identities or credentials. But what the users can't click or type, they can't lose.

Let us look at the lessons of these hacks more carefully.

Your users' email should NOT be their identity

One of the worst confusions that username-based systems peddle is that between the users' public contact information and their identities. A short, human-readable string such as an email helps others to locate and contact a user; but why use the same string—known to the attackers!—to authenticate the user himself or herself?

This string is by definition accessible to untrusted others, including attackers. User identification, on the other hand, happens between the trusted user using a trusted device and a trusted server. These devices can accommodate a identity a user need not type—and cannot easily give away. An email or a username, on the other hand, should be assumed given away by design.

This might suggest making the user's primary identity biometric, without any other intervening digital artifacts. However, biometric signatures are never precise; while great to compare against known patterns enrolled for known identities, they make poor patterns of their own. Building a large-scale identity system on the assumption that no two users will ever produce signatures that are close enough in the view of a particular sensor is a leap of faith. When a user whose biometrics are similar enough to a slightly-off readings of another is first identified in some other way, and then authenticated, the similarity does not come into to play; you will likely not be aware of the possibility of confusion. However, when these two users present their biometrics for a system to retrieve their identity, confusion is likely, and, at scale, may be intolerable.

Bringing your strongest game as a fallback is not a winning strategy

Consider a typical second-factor authentication scheme. The user first present the weakest identity he or she has, which is likely already known to the attacker; then the user presents a credential that is easily "phished". Only then the stronger factor comes into play, as an adjunct of the scheme, enrolled based on these weak primary credentials, and with additional schemes to replace that stronger factor.

These schemes occasionally prove weak in their own right. For example, NIST officially deprecated SMS-based second factor authentication because of its many insecurities[3]. Anecdotal evidence suggests these attacks are not theoretical[4].

More to the point, why not start authentication by presenting the identity that is not easily guessable, is not easily disclosed unwittingly, and is not subject to social engineering attacks? The user effort of enrolling a device for use with a non-guessable identity is hardly more than that of enrolling a trusted device as an adjunct second-factor; the attacker, by contrast, is deprived of an obvious attack point on recovery schemes of such devices.

---

3 "[Out of band verification] using SMS is deprecated, and will no longer be allowed in future releases of this guidance."—https://pages.nist.gov/800-63-3/sp800-63b.html

4 E.g., https://medium.com/@CodyBrown/how-to-lose-8k-worth-of-bitcoin-in-15-minutes-with-verizon-and-coinbase-com-ba75fb8d0bac, https://www.wired.com/2016/06/even-ftcs-lead-technologist-can-get-hacked, https://www.wired.com/2016/06/hey-stop-using-texts-two-factor-authentication