Troy Hunt
[redacted]
Surfers Paradise
QLD 4217
Australia

ATTN: The Honorable Morgan Griffith

Dear Mr. Griffith:

Thank you for the opportunity to testify before Congress in the “Identity Verification in a Post-Breach World” hearing. It was a privilege and I hope I was able to lend a valuable voice to the dialogue.

Please find following the answers to your questions sent following the hearing:

1. **In your testimony, you talked about how data is often "irrevocable" once it's been compromised. In other words, there really isn't a way for a consumer, or even a business, to find their stolen information and "take it back."**
    a. **Once information has been stolen from an organization, where does it typically end up? Is it someone's personal computer, a hosting service, somewhere else?**

    Stolen data may exist in all of these locations. Individuals will keep personal copies on their PCs (their intentions may vary from benign curiosity to malicious use) and hosting services are often used to redistribute this data further afield. Peer to peer torrent services are also frequently used, a perfect example of which appeared only the week after my testimony via a thread on Reddit: https://www.reddit.com/r/pwned/comments/7hhqfo/combination_of_many_breaches/

    Here we have a 593GB torrent of literally hundreds of different data breaches in one handy download. The context of that Reddit thread was that an individual had then taken those breaches, extracted the email addresses and passwords (removing the cryptographic protection that was provided to many of them) and turned it into another torrent of 41GB with 1.4 *billion* credentials. This data is now being actively used to compromise the accounts of victims where they’ve reused their password across other services.

    b. **Do malicious actors looking to sell this kind of compromised data sell it more than once?**

    Yes. A notable example was the sale of the LinkedIn data breach in May 2016 via the seller known as “peace_of_mind” who sold the data multiple times over for as much as 5 BTC each time (about US$2.2k at the time). In fact, as the data was sold over and over again, the value dropped as the data began circulating more: https://motherboard.vice.com/en_us/article/53ddqa/linkedin-finally-finished-resetting-all-the-passwords-leaked-in-2012

c. Based on your testimony and reporting that we've seen, compromised information, once it becomes well-known that a service has suffered a breach, seems to become much more widely available. Is this true?

Yes. When a data breach is unknown, the victims have no impetus to protect themselves from this specific risk, for example they wouldn't proactively change their passwords. Once known, a breached organisation will frequently force password resets thus protecting their members. They'll also notify members of the incident which then prompts them to change that same password on other services where they've reused it thus decreasing the value of the data to malicious parties. As that value decreases, there is less value in holding the data and it tends to begin circulating more broadly.

d. So, after this whole process, how many copies of a single breached database or set of information might exist?

Once data begins circulating, it's simply impossible to say. In a case like Ashley Madison where the data was intentionally redistributed as broadly as possible, there would be *at least* tens of thousands of copies of the data and it continues to replicate to this day.

e. Even with these multiple copies of information floating around, what makes it so difficult for organizations to find this data and "take it back?"

It's very dependent on the nature of the breach. Some data breaches may be difficult to find because the data is being quite tightly held; the original attacker may not have shared it or only done so within a small, trusted circle. But then in cases like the aforementioned LinkedIn and Ashley Madison data breaches, that data remains very easily discoverable to this day and both those organisations would have obtained copies of it very early on in order to assess their risk posture.

"Taking it back", however, is a very different story. Digital theft is unlike physical theft in that a stolen item can't simply be retrieved because there is always the risk that other copies remain. I've been involved in data breaches cases in the past where all parties known to have the data have made commitments that they've removed all copies (for example, the Australian Red Cross Blood Service data breach), but this ultimately relies on trusting an unidentified third party that they've kept their word and not redistributed the data or made additional copies. This is why my testimony referred to "there's no putting the data breach genie back in the bottle" because there's (usually) no guarantee that data in unauthorised hands hasn't been further distributed.