

Popular third-party genetic genealogy site is vulnerable to compromised data, impersonations

October 29 2019, by Sarah McQuate



DNA testing services are making it easier for people to learn about their heritage. People can also use their genetic testing results to connect to potential relatives in their family trees by using third-party sites, like GEDmatch, where they can compare their DNA sequences to others in the database. Credit: [Marco Verch/Flickr](#)

DNA testing services like 23andMe, Ancestry.com and MyHeritage are making it easier for people to learn about their ethnic heritage and genetic makeup. People can also use genetic testing results to connect to potential relatives by using third-party sites, like GEDmatch, where they can compare their DNA sequences to others in the database who have uploaded test results.

But a less happy ending is also possible. Researchers at the University of Washington have found that GEDmatch is vulnerable to multiple kinds of security risks. An adversary can use only a small number of comparisons to extract someone's sensitive genetic markers. A malicious user could also construct a fake genetic [profile](#) to impersonate someone's relative.

The team [posted its findings](#) Oct. 29. The researchers have also had this research accepted at the [Network and Distributed System Security Symposium](#) and will present these results in February in San Diego.

"People think of genetic data as being personal—and it is. It's literally part of their physical identity," said lead author Peter Ney, a postdoctoral researcher in the UW Paul G. Allen School of Computer Science & Engineering. "This makes the privacy of genetic data particularly important. You can change your credit card number but you can't change your DNA."

The mainstream use of genetic testing results for genealogy is a relatively recent phenomenon. The initial benefits may have obscured some underlying risks, the researchers say.

"When we have a new technology, whether it is smart automobiles or medical devices, we as a society start with 'What can this do for us?' Then we start looking at it from an adversarial perspective," said co-author Tadayoshi Kohno, a professor in the Allen School. "Here we're

looking at this system and asking: 'What are the [privacy issues](#) associated with sharing genetic data online?'"



UW researchers found that an adversary can use only a small number of comparisons on GEDmatch to extract sensitive genetic markers for someone and construct a fake genetic profile to impersonate someone's relative. Shown here is a genetic pedigree outline of two parents with two kids. Then another child (red) falsely claims to be related to the father. Credit: Rebecca Gourley/University of Washington

To look for security issues, the team created a research account on GEDmatch. The researchers uploaded experimental genetic profiles that they created by mixing and matching genetic data from multiple databases of anonymous profiles. GEDmatch assigned these profiles an ID that people can use to do one-to-one comparisons with their own

profiles.

For the one-to-one comparisons, GEDmatch produces graphics with information about how much of the two profiles match. One graphic is a bar for each of the 22 non-sex chromosomes. Each bar changes length depending on how similar the two profiles are for that chromosome. A longer bar shows that there are more matching regions, while a series of shorter bars means that there are short regions of similarity interspersed with areas that are different.

The team wanted to know if an adversary could use that bar to find out a specific DNA sequence within one region of a target's profile, such as whether or not the target has a mutation that makes them susceptible to a disease. For this search, the team designed four "extraction profiles" that they could use for one-to-one comparisons with a target profile they created. Based on whether the bar stayed in one piece—indicating that the extraction profile and the target matched—or split into two bars—indicating no match—the team was able to deduce the target's specific sequence for that region.

"Genetic information correlates to medical conditions and potentially other deeply personal traits," said co-author Luis Ceze, a professor in the Allen School. "Even in the age of oversharing information, this is most likely the kind of information one doesn't want to share for legal, medical and mental health reasons. But as more genetic information goes digital, the risks increase."

Next the researchers wondered if an adversary could use a similar technique to acquire a target's entire profile. The team focused on another GEDmatch graphic that describes how well the profiles match by showing a line of colored pixels that mark how well each DNA segment in the query matches the target: green for a complete match, yellow for a half match—when one strand of DNA matched but not the

other—and red for no match.

Then the team played a game of 20 questions: They created 20 extraction profiles that they used for one-to-one comparisons on a target profile that they created. Based on how the pixel colors changed, they were able to pull out information about the target sequence. For five test profiles, the researchers extracted about 92% of a test's unique sequences with about 98% accuracy.

NUMBER OF MATCHES DETERMINES LENGTH OF BAR



COMPLETE MATCH



↑
**DOESN'T
MATCH**

For the one-to-one comparisons, GEDmatch produces a bar for each of the 22 non-sex chromosomes that changes length depending on how similar the two profiles are for that chromosome. Shown here is an example of this graphic. A longer bar shows that there are more matching regions (top), while a series of shorter bars means that there are short regions of similarity interspersed with areas that are different (bottom). Credit: Rebecca Gourley/University of Washington

"So basically, all the adversary needs to do is upload these 20 profiles and then make 20 one-to-one comparisons to the target," Ney said.

"They could write a program that automatically makes these comparisons, downloads the data and returns the result. That would take 10 seconds."

Once someone's profile is exposed, the adversary can use that information to create a profile for a false relative. The team tested this by creating a fake child for one of their experimental profiles. Because children receive half their DNA from each parent, the fake child's profile had their DNA sequences half matching the parent profile. When the researchers did a one-to-one comparison of the two profiles, GEDmatch estimated a parent-child relationship.

An adversary could generate any false relationship they wanted by changing the fraction of shared DNA, the team said.

"If GEDmatch users have concerns about the privacy of their [genetic data](#), they have the option to delete it from the site," Ney said. "The choice to share data is a personal decision, and users should be aware that there may be some risk whenever they share data. Security is a difficult problem for internet companies in every industry."

Prior to publishing their results, the researchers shared their findings with GEDMatch, which has been working to resolve these issues, according to the GEDmatch team. The UW researchers are not affiliated with GEDmatch, however, and can't comment on the details of any fixes.

"We're only beginning to scratch the surface," Kohno said. "These discoveries are so fundamental that people might already be doing this and we don't know about it. The responsible thing for us is to disclose our findings so that we can engage a community of scientists and

policymakers in a discussion about how to mitigate this issue."

More information: Genotype Extraction and False Relative Attacks: Security Risks to Third-Party Genetic Genealogy Services Beyond Identity Inference. dnasec.cs.washington.edu/genetic-genealogy/

Provided by University of Washington

Citation: Popular third-party genetic genealogy site is vulnerable to compromised data, impersonations (2019, October 29) retrieved 26 April 2024 from <https://techxplore.com/news/2019-10-popular-third-party-genetic-genealogy-site.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--