

Next Generation De-duplication Technologies for Marketing Research

A recent study suggests that there are 7.8 connected Internet devices per household in the U.S.

This poses a challenge to survey security and deduplication that most companies do not fully understand.

Assuring respondents' answers are not duplicated is a fundamental principle of data collection, and as the number of connected devices grows, so too does the complexity of assuring respondents are unique.

High "survey N" sizes, low incidences, short timelines, and niche markets drive the need for multi-faceted fielding tactics. Employing a robust deduplication solution will ensure data integrity, reduce costs, and eliminate uncomfortable client discussions. Unfortunately, many of today's solutions haven't kept up with market needs.

Let's look at some of the more prevalent deduplication technologies in use today and see where they fall short.

IP Deduplication

"IP," also commonly referred to as an Internet address, facilitates online communication. In the same way that someone needs your phone number to give you a call, a remote computer needs your IP address to communicate with your computer.

IP Deduplication technologies catalogue the IP addresses of the computers asking to enter a survey. If a participant tries to enter with the same IP, they are considered a duplicate and refused entry. This appears to be a great solution, however, it has drawbacks.

1. IPs addresses are not always static – an IP address can change.
2. Multiple devices and users share the same IP address.
3. If you travel, your home or work IP doesn't travel with you.

IP Deduplication in and of itself is not a bad technology. However, having a survey with a unique IP address does not guarantee a duplicate-free project. Conversely, a survey with duplicated IP addresses does not necessarily signify duplication issues.

"Cookie" Drop Deduplication

Cookies are small files stored in a computer's browser directory or program data subfolders. In short, they are used to make online experiences go as smoothly as possible.

Cookie Deduplication Technologies function by requesting information that can be retrieved at a later time. This sounds like a great solution. However, there are several disadvantages to Cookie Drop Deduplication Technologies.

1. Many Internet users browse 'privately' or 'incognito' and reject the use of cookies.
2. Most browsers allow users to delete cookies after each online session.
3. Each browser on each device a respondent uses has its own cookie repository.

Browser Fingerprint/Digital Fingerprint Deduplication

A browser or device fingerprint is information collected about a computing device for the purpose of identification. Comparable to a human fingerprint, a computing device has several values that can determine uniqueness.

Unfortunately, browser-fingerprinting isn't fool-proof. It mistakenly assumes points of identification distributed among users are random, when in fact, mainstream devices fall into a more homogenous ecosystem. For example, the number of iPhone and iPad versions and models are not diverse and will more often incidentally fingerprint identically between two unique survey participants.

Another environment to consider is the corporate network. Here, you are likely to find many dozens or hundreds of potential participants with the exact same browser, plugins, fonts etc., leading to false-positive results.

Other Considerations

Testing Bypass

Standard practice is to ensure those invited to a survey are correctly re-routed and credited. In this case, security is often temporarily bypassed; there is a security “on/off switch” that the survey host system can “flip.” An ideal security solution will eliminate accidental errors in security bypass.

Resetting IDs – Allowing disqualifies to attempt again

In small universe or low qualification studies, it may become necessary to relax survey qualification points. In these conditions, there is a need to allow past participation re-entry without the requirement to modify, script or turn off security. A flexible solution will have this ability.

Blacklisting/Whitelisting

Customized deduplication solutions will facilitate the concept of blacklisting and whitelisting of IP addresses, cookies, and digital profiles.

Flexibility

The ideal tool will allow adjustments for more or less selective detection. For example, a consumer-based survey may require unique IP addresses, but a healthcare survey may allow respondents with the same IP (i.e. unique respondents at the same location) to participate.

Fraud Prevention

A robust deduplication solution will prevent fraud. Participants who try to obscure or hide their identity should be disallowed.

A deduplication solution should employ all reasonable techniques and methods together. In other words, it is best to develop a systematic way of knowing all the devices, IP addresses, and digital fingerprints of any single survey participant.

At SHC Universal, we are committed to quality and Perfect Data. Our years of experience have taught us how valuable, and indeed, necessary, a robust deduplication solution is. This solution needs to solve all of the challenges inherent with current deduplication tools and we look forward to working with industry participants to draft a roadmap for creating one.

Jason Freeman – COO, SHC Universal

Phil Sofia – President, SHC Universal

Disclaimer: The views and opinions expressed in this feature are those of the author and may not reflect the official policy or position of the BHBIA. The BHBIA have not verified any of the information quoted and do not accept any responsibility for its accuracy, or otherwise.