Selecting Which Visits to Track via Filters

Abstract

Filtering is a powerful option to restrict the data Google Analytics collects in order to focus on a subset of users. If improperly implemented, a filter can also severely limit a library's ability to accurately analyze the data. Chapter 4 of Library Technology Reports *(vol. 49, no. 4) "Maximizing Google Analytics: Six High-Impact Practices" presents an in-depth example of using a filter to exclude data from library staff computers to focus on click data generated by actual website users.*

Filters are a great way to identify specific types of visitors, and focusing on actual website users is a useful high-impact practice. As more and more libraries strive to achieve a data-driven culture of assessment, it's important to ensure that the data we're relying on is reporting what we think it is about our users. While library staff are users too, they're superusers, and their behavior in our online environments is not typical of the majority of our patrons. So how can we be sure that we're getting accurate data from web analytics software reports to improve the end-user experience? The answer is filtering staff use out of our website analytics data. In this chapter, we'll walk through the process, step by step, of setting up filters in Google Analytics. The general process, in three steps, is

- determine an appropriate range of staff IP addresses to exclude;
- establish a master profile in Google Analytics and create a new profile to filter; and
- set up a filter that will exclude the IP range of staff users on your new profile.

Note that creating filters requires administrative access to Google Analytics, so if you do not have administrative access, you may need to communicate your intentions and desires to those who do.

Finding a Useable IP Range to Filter

First and foremost, you'll need to determine an effective IP range to use to filter out library staff (or other audience) use. This can be tricky because analytics software divines the "where?" information about user locations from computer IP addresses. Depending upon your library's network architecture (including how the library's network fits into a parent institution, if applicable) and how IP ranges are defined for all of your employee and public workstations, it's unlikely that you will be able to exactly, precisely filter out all staff use in a way that provides only end-user data. However, the shortcomings of using IP addresses to determine visitor and visit information is inherent in all analytics measurements, and while it's not precise and perfect, you will still likely glean useful information about end-user behavior by setting up a staff/public filter. Let's consider two examples, a good one and a better one.

Good: Internal Library Use, External Library Use

Many network administrators use Dynamic Host Configuration Protocol, commonly called DHCP, to dynamically assign IP addresses to individual computers across networks, including staff workstations, labs, classrooms, and public computing areas. While DHCP makes the task of managing large and complex networks easier and more efficient, it can throw a wrench

into the analytics process simply because IP addresses change on a regular basis: what is an IP address for a staff computer one day may be reassigned to a public or classroom workstation the next. That doesn't make your quest for user-only web data hopeless; you just need to understand how you're defining what you're asking the analytics software to provide. In this scenario, data collected on traffic external to the library would be a closer approximation of student use, and internal traffic of library staff use.

Better: Staff Use, Public Use

No two networks are the same, however; if you're lucky, there's a chance that your network administrators have separated staff and public use, perhaps even on two separate networks, depending upon your network architecture. This second "separate network" scenario would obviously provide a more accurate picture of customer versus staff use of the library's websites. But even this is not 100 percent perfect; if library staff work from home, their visitor and visit data for those sessions on the website will be recorded in reports with patron data because the IP address provided will belong to their home Internet provider. An exception to this would be if staff working remotely use a virtual private network, or VPN, which would assign them a work IP address. If you're uncertain how to determine an IP range, explain your goal of separating your audiences via IP to your network administrators and ask what they would advise.

Setting Up a Master Profile and Creating a Profile to Filter

Once you've arrived at a range of staff IP addresses to exclude in a filter, you're ready to make changes in Google Analytics. Remember, applying filters requires administrative access. One very important thing to know about filters is that they are destructive to data: this means that once a filter is added, the information that you're asking to be omitted via the filter will no longer be collected. This certainly seems less than desirable; while you may want to distinguish between staff and public use, you may not be willing to do without a complete data picture. Enter profiles, which are "defined view[s] of visitor data from a property."[1] Google Analytics provides up to fifty profiles per site, and your account already has one by default. Start small and add just one profile, or optionally two profiles, for the project at hand. First, establish a master profile; second, create a new a profile on which to place a filter; and finally, apply the filter. The master profile will collect all of your data, from everyone everywhere, all of the time, and the filtered profile will slice out the end-user-only analytics data.
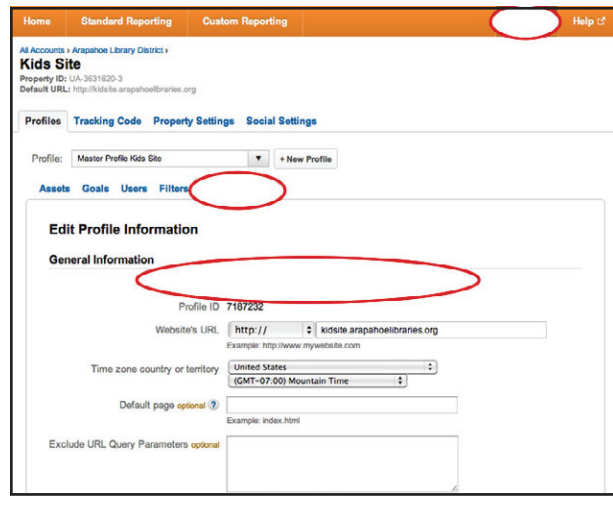

Figure 4.1
Establishing a master profile, Arapahoe Library District Kids website

Creating a Master Profile

Assuming that you have not added any additional profiles beyond the default profile that is standard with a new Google Analytics account, you may set the default as your master profile. Log in to Google Analytics with an account with administrative access and click the Admin button in the upper right corner. (If you track more than one site and have multiple web properties, you may need to click on the one you'd like to filter here.) Click Profile Settings in the second tab row, and in the Profile Name field, type in "Master Profile" or another descriptive name of your choosing. Leave the rest of the form as is, click the Apply button, and behold, the master profile (see figure 4.1)!

If you find that multiple profiles are already established, it would be wise to consult with your colleagues first before renaming them, creating any new ones, or filtering any existing ones. If the default profile is already in use by someone else, or if you're not sure if it is, follow the instructions above to establish a new profile to use as your master, and then create a second one to use as your filtered profile.

Creating a Profile to Filter

Now that you have a master profile set up to catch all of your data, all of the time, it's time to create a new profile to filter (see figure 4.2). From the Property Settings screen—where you left off above—click on the Profiles tab, and then click on the + New Profile button.

Then, as shown in figure 4.3, enter a descriptive name, such as "Public Use (No Staff)," for the profile in the Profile Name field. Click the Create Profile button.
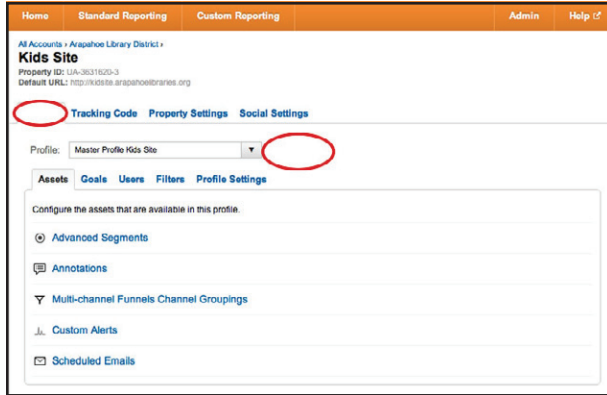
Figure 4.2
Creating a new profile, Arapahoe Library District Kids website

Figure 4.3
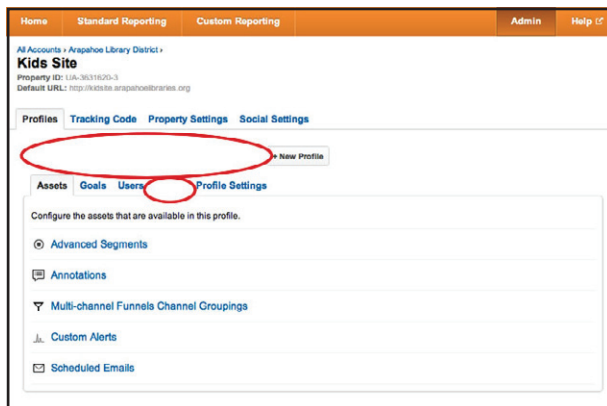Naming a new profile, Arapahoe Library District Kids website

Figure 4.4
Selecting the profile to filter, Arapahoe Library District Kids website

Adding the Filter

Finally, add a filter to that new profile. You should still be on the page for your newly created profile; make sure that it's selected in the Profile drop-down menu (figure 4.4), and then click on Filters, and after that, the + New Filter button (figure 4.5).

Fill out the form, which is dynamic and will change

Figure 4.5
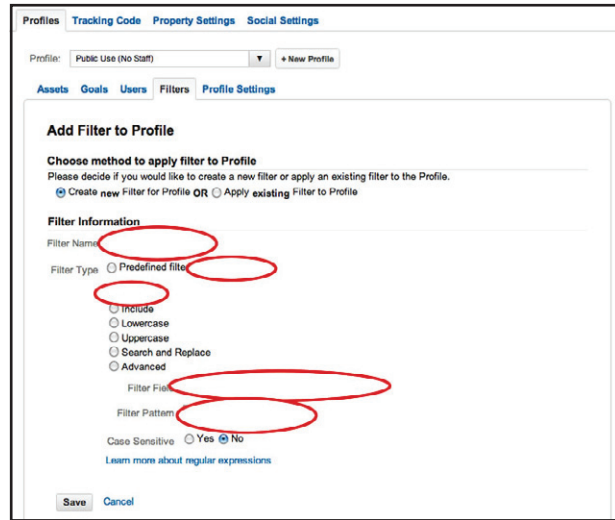Applying a filter, Arapahoe Library District Kids website

Figure 4.6
Defining filter properties, Arapahoe Library District Kids website

as you enter data (see figure 4.6). Type whatever name you like for the filter, for example, "Public Use (No Staff)." Next, change the Filter Type to Custom Filter, and then select Exclude from the list of options, setting the Filter Field to Visitor IP Address from the drop-down menu.

To establish the Filter Pattern, we recommend using Google Analytics's IP Address Range Tool, which is available in the online documentation. Open the IP Address Range Tool in another web browser tab or window, type in the first and last IP addresses in the range you established earlier, and click the Generate RegEx button. Copy and paste the results into the Filter Pattern field back on the Google Analytics form. If you have more than one range of IP addresses, read "More Tips on IP Address Filtering" in the online documentation below the IP Address Range Tool.

Google Analytics: IP Address Range Tool
http://support.google.com/analytics/bin/answer.py?hl=en&answer=1034771

Double-check that the filter was added to the new profile, not the master profile.

If you have more than one site, lather, rinse, and repeat until you have profiles and filters set as desired on all of them. You could also create a Staff Only filtered profile if you'd like to track staff-only use of the website; just follow the above steps to create another profile and filter to add to it, but use an Include custom filter type when creating the filter on the new Staff Only profile. If you've set up custom reports, or would like to, you can set them on your filtered profiles if you'd like to focus on user-only or staff-only use.

You may be tempted to jump right over and check out what your site data looks like with library staff data filtered out, but unfortunately, you'll have to wait for a few hours. Filters collect data only going forward from the time they were set, which is yet another reason that the master profile is important. Remember: filters, while very helpful for the sake of analysis, are destructive to data. The master profile will retain all of your historical data, and going forward, the filtered profile will slice out the audience-specific data for you.

If other colleagues have access to Google Analytics, be sure to communicate any changes you've made so that they understand the purpose of the profiles and which would best suit their needs. For example, even though the web team would likely want to focus its analysis on the user-only filtered profile, library administration would still likely want to report the statistics to the library's stakeholders from the master profile, as it is the sum total of web use.

Conclusion

So, in summary:

- Determine an effective range of IP addresses to include or exclude from Google Analytics reports.
- Establish a master profile in Google Analytics that will track all user data (library staff and users) and retain your historical data.

- Create a new profile for public users.
- Apply a filter to include or exclude a range of IP addresses, as appropriate.
- Optionally, create a second filtered profile for staff use.
- Wait a few hours for the separated user or staff data to roll in!

After these settings are implemented, be aware of any major network changes that could possibly affect the accuracy of your IP ranges. If you notice any sudden, unexplained, or drastic differences in your data or reports, or if your filters stop working altogether, check with your network administrators to determine if the ranges you are using are still accurate and the most effective available.

Additional Resources

- "Create/Manage Profile Filters," Google Analytics website, http://support.google.com/analytics/bin/answer.py?hl=en&answer=1034823.
- "Exclude Internal Traffic," Google Analytics website, http://support.google.com/analytics/bin/answer.py?hl=en&answer=1034840.

Note

1. "Accounts, Users, Properties, and Profiles," Google Analytics website, accessed March 4, 2013, http://support.google.com/analytics/bin/answer.py?hl=en&answer=1009618.