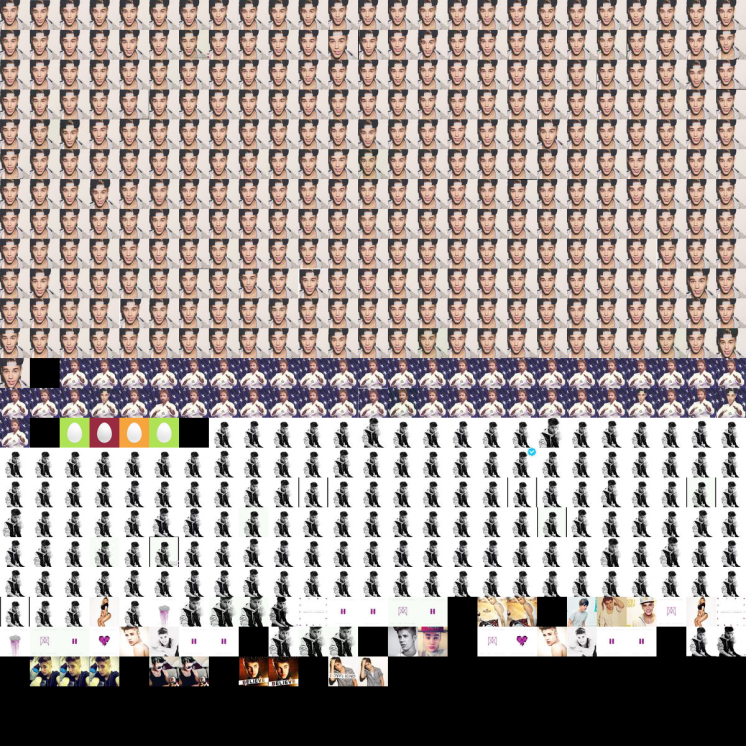
Ben Petroski

Computer and Network Security

**Identifying Fake Twitter Accounts with the Same Username via Profile Picture Comparisons**

Twitter is a social media platform that provides users with a method of interacting and acquiring information. Due to the large number of users on Twitter, and therefore potential audience, the creation of fake accounts has become more prevalent. These fake accounts are created to amass a significant number of followers, which can then be sold for a profit. Such accounts can be an annoyance to users and may convey information that is inaccurate or unnecessary. Additionally, these fake accounts, which comprise roughly 10% of all accounts, can be abusive to the back-end systems used to store twitter data. For these reasons, we take an approach to identifying accounts which may be fake using a data set of 659 profiles with identical usernames.

 From this data set, we are able to extract URLs for profile pictures of users, which will be the basis of our approach for identifying possibly fake profiles. A simple Python script was used to compare each image to another without any duplicate comparisons. The Manhattan norm was calculated between each pair of images to determine the level of similarity. After a few test runs, a Manhattan norm less than 45 was determined to signify that two images were very similar and was set to be the cutoff for generating lists of matches. Using this list generation method, we are able to generate several lists containing the index of the image URL for pairs that are similar to one another. Additionally, a list containing the index of the image URL was created for image pairs that have a Manhattan norm of zero as it was noticed that most of these pairs have identical profile picture filenames. Lastly, a mosaic of the images in the matching list is created. This was used as a method to verify that the Manhattan norm matched as many identical images as possible. This mosaic is generated as an output file, which is stored as “mosaic.bmp”. To distinguish between lists, black squares are inserted as a divider.

We are able to successfully generate 14 lists, consisting of 536 URLs, with remarkable accuracy and thus have determined that this approach is a viable method for identifying potentially fake profiles.

1. Sastanin. "How Can I Quantify Difference between Two Images?"StackOverflow. N.p., 14 Oct. 2010. Web. 27 Apr. 2014. (<http://goo.gl/ntkCie>)
2. Gullen, Tom. "Algorithm to Compare Two Images." Stack Overflow. N.p., 10 Aug. 2010. Web. 27 Apr. 2014. (<http://goo.gl/LVPU6n>)