

Apps Data Management

**University of Texas at Dallas
Computer Science**

Big Data Management and Analytics Lab



**This project is supported by NSF.
It was conducted under the supervision of Professor Latifur Khan.**

Table of Contents

Table of Contents	1
Overview	2
Part 1: Login	5
Part 2: Navigating the Main Page.....	8
Part 3: Dynamic Analysis	11
Part 4: Data Analytics	14
Part 5: Documentation	17
Part 6: About	18

Overview

With the recent emergence of mobile platforms capable of executing increasingly complex software and the rising ubiquity of such platforms in sensitive domains such as banking, military operations, healthcare, and law enforcement, user privacy and data security have become major concerns. This attracted researchers to study various methods to perform Traffic Analysis (TA) on mobile apps that connect to remote hosts or app servers through cloud service providers.

The focus of this project is to identify an app (e.g. a bank app) by listening to the encrypted TCP traffic as illustrated in Figure 1. The attacker passively gathers encrypted packets, extracts features, and uses data analytics/machine learning to predict the name of the app. This is known as **App Fingerprinting**, where attacker can gain information that negatively affects the peace and security of the end users or the general population at large. It is important to note that the app server IP address cannot be used to reveal the name of the app as multiple apps may be hosted on the same cloud server with the same IP.

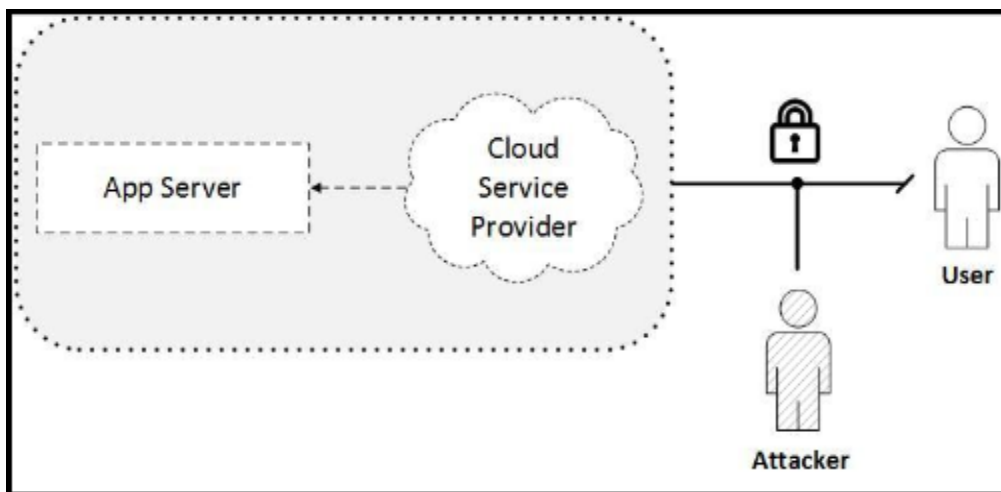


Figure 1. An illustration of App Fingerprinting.

To help researchers study the effect of such adversarial attacks, the web application tool introduced in this manual has two objectives, mobile **Dynamic Analysis** and **Data Analytics**.

The first part of our project is **Dynamic Analysis** which has been implemented through apps **Data Collection**. We build our own data set by executing multiple Android apps. As part of the process, the tool allows web browser clients to demo data collection using an Android emulator.

Figure 2 illustrates the actual data collection process over a period of one month using an Android phone.

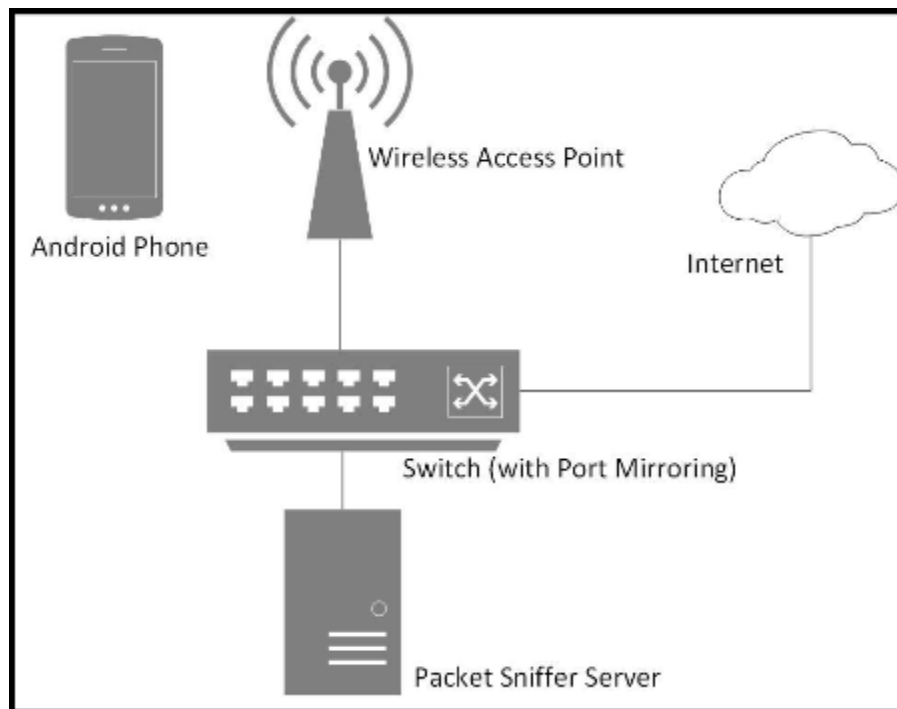


Figure 2. Data Collection Process.

The dataset was collected by executing multiple Android apps on a Samsung Galaxy S device, running Android version 4.3.1. We randomly select 2000 financial apps from Google Play Store. We then install and launch these apps on the phone which is connected to the Internet via a wireless router. Each trace per app is collected over a 30-sec period passively using a mirroring switch at the wireless router. These traces from such apps are then used to perform data analytics. It is important to note that we uninstall each app as soon as we complete capturing a trace to avoid any background noise during further trace generation. Using this web application, the clients will experience this dynamic analysis process as will be explained in more detail in subsequent sections.

The second part of the project is **Data Analytics**. We use the traces from Data Collection to determine patterns which are used for app fingerprinting. Various features from network packets are extracted and then used alongside machine learning techniques to train a model that can identify an app from its network traffic. We evaluate this method in various environments to test its efficacy. The web application will help the client experience the effect of various machine learning algorithms with different feature extraction techniques.

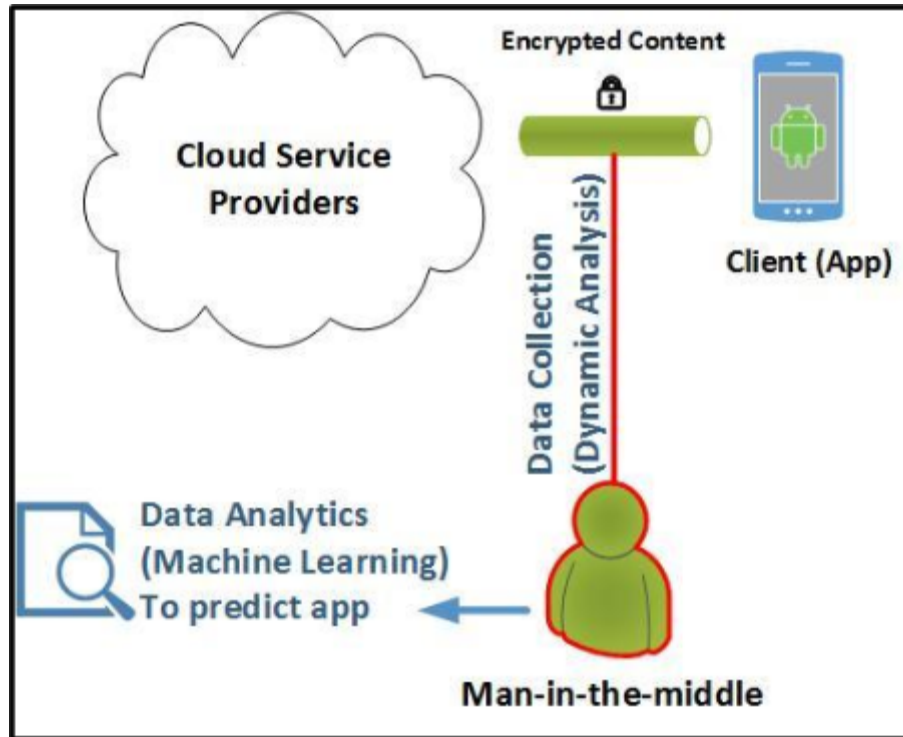


Figure 3. Dynamic Analysis and Data Analytics.

Figure 3 depicts the big picture of the app fingerprinting attack scenario. The attacker (man-in-the-middle) eavesdrops passively and collects app traces. Using feature extraction methods, these traces are converted to vectors that machine learning algorithms take as input for the classification/prediction task.

This manual shows how to implement the data collection/dynamic analysis process using an Android emulator. It also shows how to implement the data analytics part using various machine learning approaches for the fingerprinting task.

Part 1: Login

When you first access the website, the login screen appears.



Big Data Management and Analytics Lab
Computer Science

Smart Phone Apps Data Management

Directed by: Prof. Latifur Khan

Sign in



Username

Password

Login

[Create an account](#)

[Forgot Password?](#)

Copyright © 2016. Big Data Management and Analytics Lab. The University of Texas at Dallas. dml.utd@gmail.com. All rights reserved.

1. If you have an account, login. If you do not, follow the steps bellow.
2. Click the *create an account* link.



Big Data Management and Analytics Lab
Computer Science

Smart Phone Apps Data Management

Directed by: Prof. Latifur Khan

Sign in



Username

Password

Login

[Create an account](#)

[Forgot Password?](#)

2 →

Copyright © 2016. Big Data Management and Analytics Lab. The University of Texas at Dallas. dml.utd@gmail.com. All rights reserved.

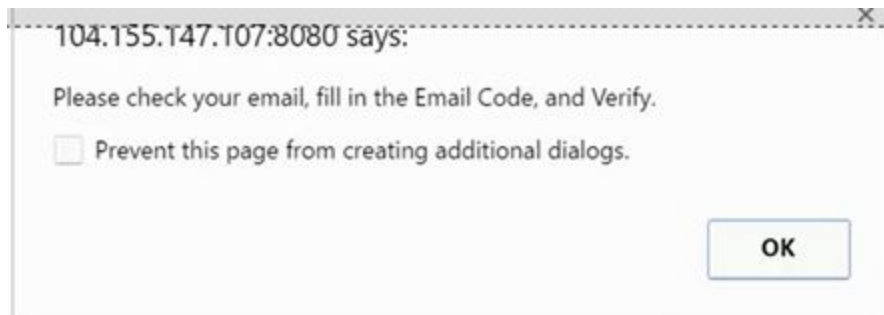


3

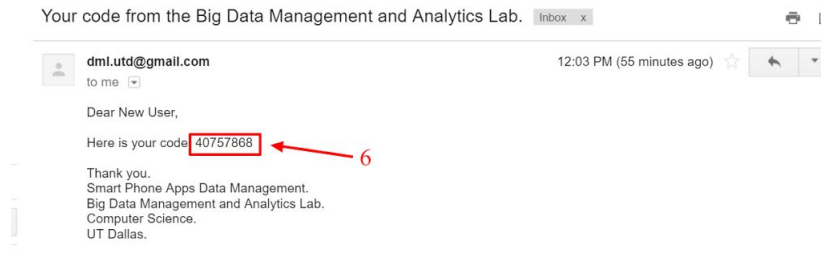
4

Copyright © 2016. Big Data Management and Analytics Lab. The University of Texas at Dallas. dml.utd@gmail.com. All rights reserved.

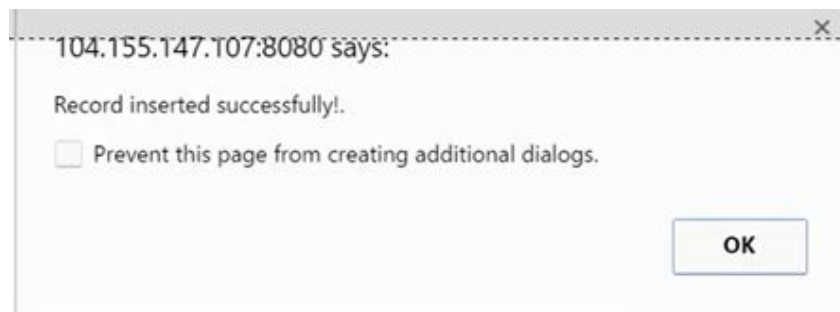
3. Fill in the fields. Make sure to enter a valid email address as you will need this to complete your account creation.
4. Click the Submit button when you are done.
5. If all the fields were entered correctly and the username and email address you selected have not been already used, then a message will appear letting you know to check your email. If there are any errors, a message will appear letting you know, and the fields that need to be corrected will be outlined in red.



6. Check your email and get the access code. The code will expire within 10-20 minutes.



7. Paste the code into the box on the website. A message will appear letting you if it was successful.



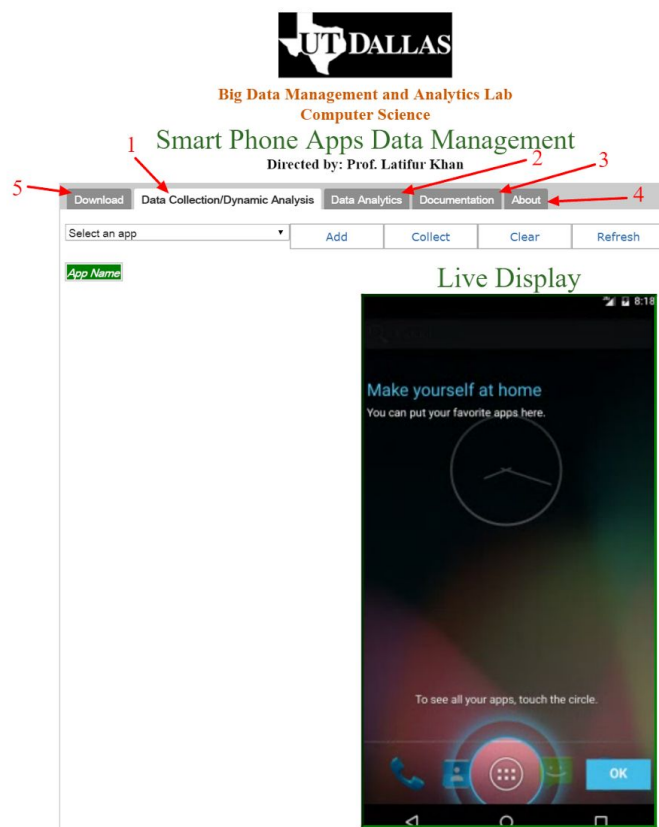
8. From here you will be redirected to the login screen. Enter your username and password and click the Login button.



Part 2: Navigating the Main Page

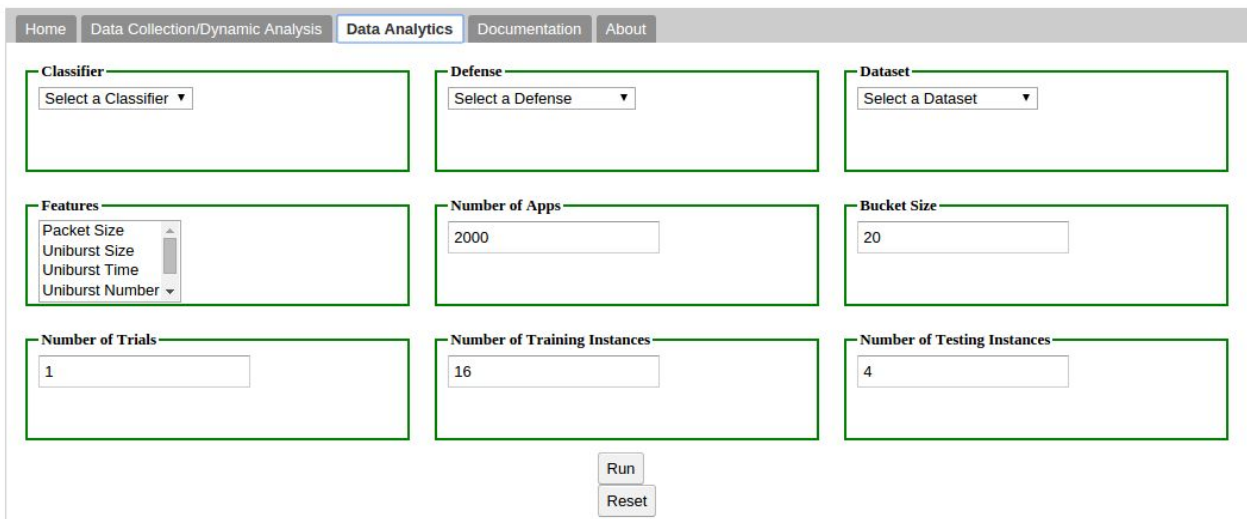
Once logged in, the main page loads. All functions can be accessed from this page.

1. The “**Data Collection / Dynamic Analysis**” tab is open when the page loads. This hosts the first part of the project. This is where you can run tests to determine what websites an app accesses when it runs. As part of the process, the tool allows you to demo data collection using an Android emulator.



[Logout](#)

- The “**Data Analytics**” tab is where you can run app fingerprinting tests by setting specific parameters. The web application will help you experience the effect of various machine learning algorithms with different feature extraction techniques.



The screenshot shows the 'Data Analytics' tab of the web application. It features a navigation bar with tabs: Home, Data Collection/Dynamic Analysis, Data Analytics (selected), Documentation, and About. The main content area contains several input fields for configuring the app fingerprinting test:

- Classifier:** A dropdown menu with the text 'Select a Classifier'.
- Defense:** A dropdown menu with the text 'Select a Defense'.
- Dataset:** A dropdown menu with the text 'Select a Dataset'.
- Features:** A list of features with checkboxes: Packet Size, Uniburst Size, Uniburst Time, and Uniburst Number.
- Number of Apps:** A text input field containing the value '2000'.
- Bucket Size:** A text input field containing the value '20'.
- Number of Trials:** A text input field containing the value '1'.
- Number of Training Instances:** A text input field containing the value '16'.
- Number of Testing Instances:** A text input field containing the value '4'.

At the bottom of the form, there are two buttons: 'Run' and 'Reset'.

- The “**Documentation**” tab contains helpful links. These include relevant papers, website documentation, etc.



The screenshot shows the 'Documentation' tab of the web application. It features a navigation bar with tabs: Home, Data Collection/Dynamic Analysis, Data Analytics, Documentation (selected), and About. The main content area displays a list of links:

- [User Manual](#)
- [Paper: Peek-a-Boo, I Still See You: Why Efficient Traffic Analysis Countermeasures Fail](#)
- [Paper: P2V: Effective Website Fingerprinting Using Vector Space Representations](#)

4. The “**About**” tab contains a brief overview of the project.



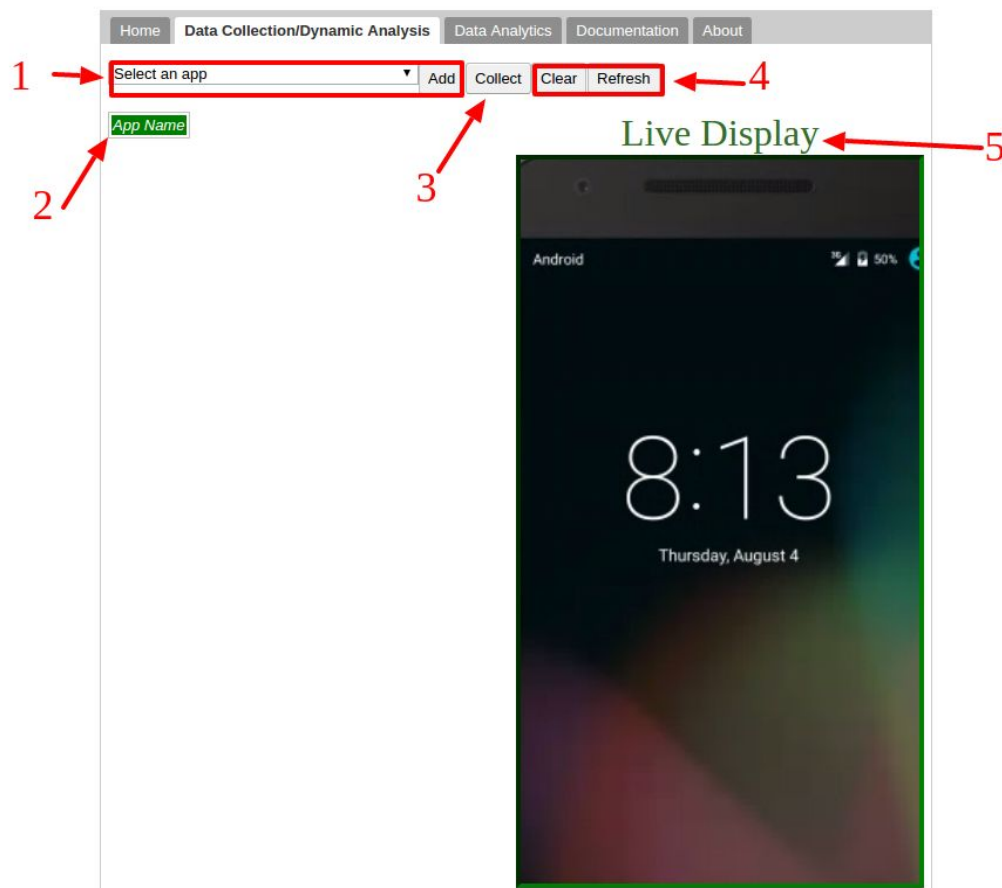
5. The “Download” tab hosts a mechanism to download files we collect during tests. Ignore this for now.
6. The “**Logout**” link at the bottom of the page allows you to logout and return to the login screen. It is available with any tab selected.

Part 3: Data Collection/Dynamic Analysis

This tab hosts the mechanism to run app specific tests.

When apps run, they access a number of remote services in the background that a user is not aware of. These tests determine which services are being accessed.

Data Collection enables the user to install mobile apps, open them, and collect traces (as TCP packets). As part of the process, the tool allows you to demo data collection using an Android emulator.




1. The drop down menu is used to select an app to test. Click on the app in the menu and click the “**Add**” button. The apps selected will be tested to determine what remote services they access.
2. Once an app has been added, its name will appear in the table under the green “**App Name**” heading. You can add/test multiple apps.

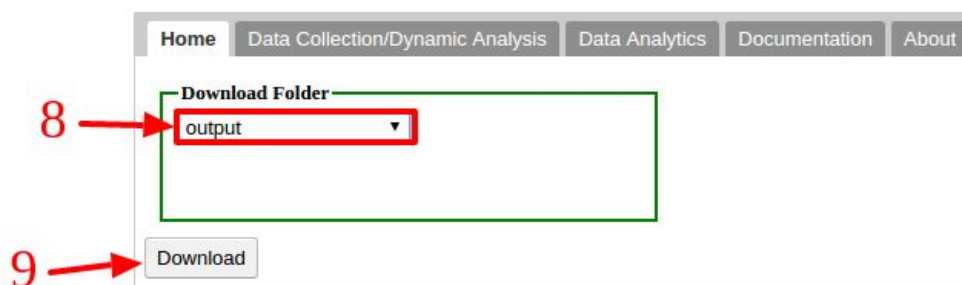
App Name
abshahin.soft.valyuta.apk
ackmaniac.currencyfxrates.apk
acore.tauxascal.apk

3. When you are done adding apps, click the “**Collect**” button. The tests will start running. A message will appear letting you that the process has begun.

App Name
abshahin.soft.valyuta.apk
ackmaniac.currencyfxrates.apk
acore.tauxascal.apk

 The request is being processed. Please wait.

4. If you would like to clear the list of apps you want to test, hit the “**Clear**” button. If you want to refresh the tests, hit the “**Refresh**” button.
5. The “Live Display” screen shows the process that occurs when the tests are running. Ex: apps opening, closing, etc.
6. When the process is complete, a pop up window will appear.
7. To obtain the results, go to the “**Home**” tab.



8. In the drop down menu in the “**Download Folder**” box, select “**output**”, or the relevant folder (date and time).
9. Click the “**Download**” button. A link will appear at the bottom left of your screen like the image below.



10. Click the link and follow the file path
`/home/dml_u td/tomcat/webapps/appfin/appsData/output/`
 and the results will be there.

Part 4: Data Analytics

This tab hosts the mechanism to run feature specific app tests.

When a user opens an app, it accesses remote servers. Sometimes, these servers cannot be readily identified and certain features they have must be extracted and used to identify them. Such traces are used as input vectors for Data Analytics where different machine learning algorithms can be applied. The web application will help you experience the effect of various machine learning algorithms with different feature extraction techniques.

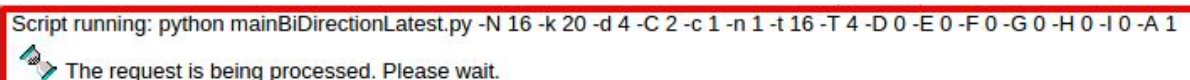
The screenshot shows the 'Data Analytics' tab of a web application. The interface includes a navigation bar with tabs: Home, Data Collection/Dynamic Analysis, Data Analytics (selected), Documentation, and About. The main content area contains several input fields and buttons, each highlighted with a green border and a red number:


- 1**: Classifier dropdown menu.
- 2**: Defense dropdown menu.
- 3**: Dataset dropdown menu.
- 4**: Features multi-select bar (showing Packet Size, Uniburst Size, Uniburst Time, Uniburst Number).
- 5**: Number of Apps input field (value: 2000).
- 6**: Bucket Size input field (value: 20).
- 7**: Number of Trials input field (value: 1).
- 8**: Number of Training Instances input field (value: 16).
- 9**: Number of Testing Instances input field (value: 4).
- 10**: Run button.
- 11**: Reset button.

1. Use the drop down menu to select a **Classifier** under which the tests will be run.
2. Use the drop down menu to select a **Defense**, if any, that will be applied to the apps that are being being fingerprinted.
3. Use the dropdown menu to select a **Dataset** from which websites will be selected for the tests.
4. Use the multi-select **Features** bar to select the features, if any, you want the tests to take into account. These features are only accessible if you choose the

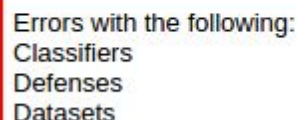
Bi-Directional classifier. To select multiple features, press and hold the ctrl key as you select features. The features you select will be the ones used to fingerprint the websites and will be the focus of the tests.

5. Enter the **Number of Apps** you want the test to use. The maximum is 2000.
6. Enter the **Bucket Size** you want the test to use. This number must be less than the number of apps you choose.
7. Enter the **Number of Trials** you want to tests to have. This must be less than 100.
8. Enter the **Number of Testing Instances** you want to tests to have.
9. Enter the **Number of Training Instances** you want to tests to have. The number of testing and training instances must add up to 20.
10. If you want to reset the fields, click the “**Reset**” button.
11. When you are done filling out all the fields, click the “**Collect**” button.
12. If all the fields were correctly filled in, a message will appear letting you know that the process has begun, as shown in the image below.



Script running: python mainBiDirectionLatest.py -N 16 -k 20 -d 4 -C 2 -c 1 -n 1 -t 16 -T 4 -D 0 -E 0 -F 0 -G 0 -H 0 -I 0 -A 1
 The request is being processed. Please wait.

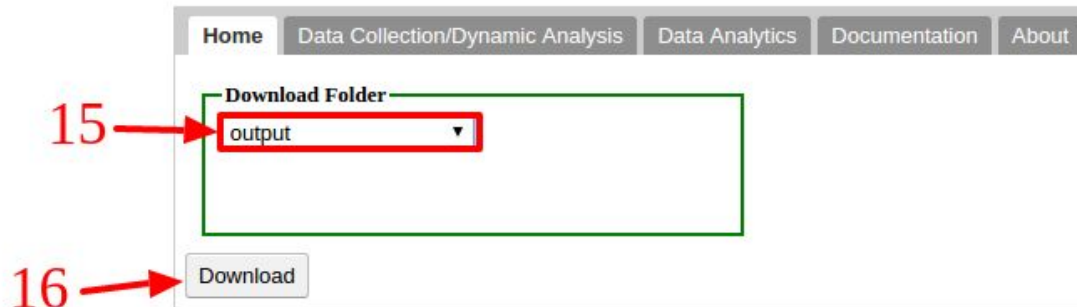
If they were not correctly filled in, a message will appear letting you know which ones to correct, as shown in the image below. Correct them, and hit “**Collect**” again.



Errors with the following:
 Classifiers
 Defenses
 Datasets

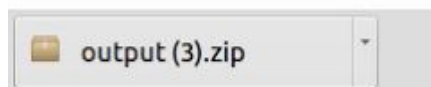
13. When the tests are done running, a pop up window will appear.

14. To obtain the results, go to the “**Home**” tab.



15. In the drop down menu in the “**Download Folder**” box, select “**output**”, or the relevant folder (date and time).

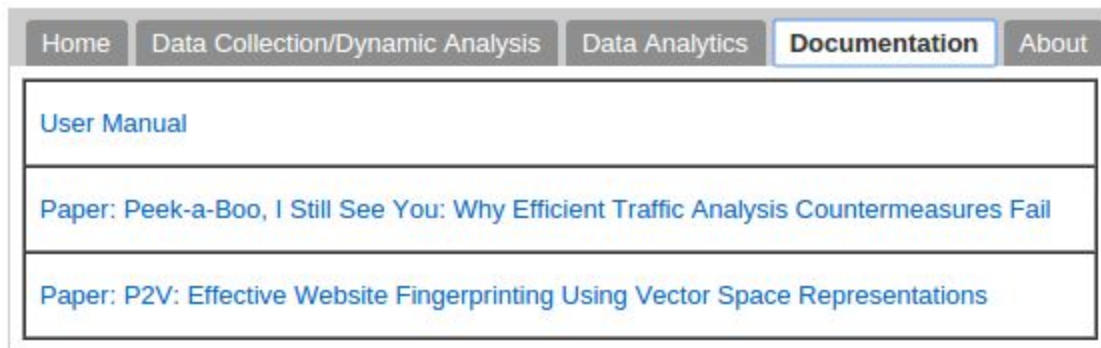
16. Click the “**Download**” button. A link will appear at the bottom left of your screen like the image below.



17. Click the link and follow the file path
`/home/dml_u td/tomcat/webapps/appfin/appsData/output/`
 and the results will be there.

Part 5: Documentation

This tab hosts useful links.



There are links to useful papers relative to app fingerprinting, user manual, etc.

Part 6: About

This tab has a brief description of the project.



There is a link to Dr. Khan's page and a brief overview of the project. For more information, look to the Overview section of this manual.