

INNOPOLIS UNIVERSITY

# Secure Systems and Networks Research project

*Submitted By:*

Vasiliy Podtikhov,  
Bulat Saifullin,  
Timur Samigullin

*Submitted To:*

Rasheed Hussain  
Azat Safin  
Konstantin Urysov  
Kirill Saltanov

December 5, 2016

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Related Work</b>	<b>2</b>
<b>3</b>	<b>Research question</b>	<b>3</b>
<b>4</b>	<b>Methodologies</b>	<b>4</b>
<b>5</b>	<b>Common things</b>	<b>5</b>
<b>6</b>	<b>VK</b>	<b>6</b>
<b>7</b>	<b>Instagram</b>	<b>7</b>
7.0.1	Getting data from instagram . . . . .	7
7.0.2	Finding range of instagram ip . . . . .	7
7.1	Facebook . . . . .	7
7.1.1	Getting data from fb . . . . .	7
7.1.2	Find ip range of fb . . . . .	7
7.1.3	Experemental results . . . . .	7
<b>8</b>	<b>Results</b>	<b>8</b>
<b>9</b>	<b>Our programms</b>	<b>9</b>
<b>10</b>	<b>Conclusions</b>	<b>10</b>
<b>11</b>	<b>Acknowledgments</b>	<b>11</b>

## **Abstract**

In modern day social networks become widely used. Practically almost all employers using them. But they can be used to formating public opinion in way not acceptable by company, or by accident share some confidentially information. This often happened because ordinary employee don't unaware of global company goals.

In this work we will try to link a social identity to an IP address by analysis of user traffics. This will help us to establish leakage, find disgruntled employees and change company politics to prevent this situations.

# 1. Introduction

Mapping IP address to account on social network is generally believed to be difficult for an individual with no dedicated infrastructure or privileged information. Social networks owners such as Vk.com and Facebook.com have this information, but they always hide it except in the case of a legal decision. But this information may be very handy in big corporations. In average 60% of employee actively use social networks [1]. And sometime employees post trade secret in social network, usually they use fake name. But if the employee go in account while he in the corporation's network mapping IP address to account on social network, can help us to find him.

## 2. Related Work

Today we widely used Netflow analysis for security reasons [2][3]. But only recently science works was introduced whom main goal was determine users action in social networks [4][5]. Unfortunately method who helped us to identify user never was introduced. In this paper we tried to find a solution for this problem.

### 3. Research question

- Find connection between user traffic and profile changes.
- What sending data affects changes in profile?
- How to analysis user's net-flow traffics?
- How to analysis a profile in social network?

ls

## 4. Methodologies

NetFlow is a feature that was introduced on Cisco routers that provides the ability to collect IP network traffic as it enters or exits an interface. By analyzing the data provided by NetFlow, a network administrator can determine things such as the source and destination of traffic, class of service, and the causes of congestion. A typical flow monitoring setup (using NetFlow) consists of three main components:[\[1\]](#)

- Flow exporter: aggregates packets into flows and exports flow records towards one or more flow collectors.
- Flow collector: responsible for reception, storage and pre-processing of flow data received from a flow exporter.
- Analysis application: analyzes received flow data in the context of intrusion detection or traffic profiling, for example.

We analysed netflow dumps in corporation environment and tried to check if connection was established in period of time and check presence of person in this time period on site. The main purpose is to find correlations between posted time and netflow traffic.

## 5. Common things



## 6. VK

Text about subject

## 7. Instagram

Instagram is an online mobile photo-sharing, video-sharing, and social networking service that enables its users to take pictures and videos, and share them either publicly or privately on the app, as well as through a variety of other social networking platforms, such as Facebook, Twitter, Tumblr, and Flickr. [2]

### Getting data from instagram

To find correlation between posted time and netflow traffic, we need to get exact time, when user post photo.

First, all information about post was getting through API functional of Instagram web-site. For this reason third-part application should be registered at instagram development page, and access-token should be got. Request method for getting info about specific post shown bellow:

```
1 https://api.instagram.com/v1/media/shortcode/BNPjFcrh22-?access_token=3955223166.3a064fe.2562f48363ac48f8b002f713fddeae2e
```

With testing accounts everything was fine, but when I tried to access to another real account I have a API error. Instagram API has one speciality: there is a sandbox environment for testing reasons, and for getting data from every page, first, he should be invited to sandbox and he should accept requests from application, even if page open for everybody. I think, application should not ask requests for viewing user's page.

The way out is to parse raw html page, and get data from it. In source of html page I saw, that there is raw json data.

```
343e\u00441\u00442\u00430\u0043b \u0044\u0043\u00442\u00431\u0043e\u0043a\u0043 \u00441 Shodan,
343e\u00442\u0043e\u00440\u00430\u0044f \u00431\u00435\u00441\u00441\u0043c\u00435\u0043d\u0043d\u0043e \u00430
3430 \u00430\u00432\u00435 \u00442\u00443\u00442 \u00438 \u00432 \u00441\u00442\u00438\u0043c\u00435. #c
_at": 1480289313.0, "id": "17866534264032413", "user": {"username": "mrzizik", "profile_pic
/scontent.cdninstagram.com/t51.2885-19/11909134_1460692754238018_1992093853_a.jpg", "id": "1
_at": 1480429061.0, "id": "17844890545175229", "user": {"username": "realitypill", "profile
/scontent.cdninstagram.com/t51.2885-19/s150x150/13188056_947691368682955_1072891983_a.jpg",
335726297069", "date": 1480288973, "likes": {"count": 75, "viewer_has_liked": false, "nodes
", "profile_pic_url": "https://scontent.cdninstagram.com/t51.2885-19/s150x150/12530716_7390
799"}, {"user": {"username": "suck_the_fystem", "profile_pic_url": "https://scontent.cdnins
254_1423660264608783_260310832_a.jpg", "id": "1710410670"}}, {"user": {"username": "prismspe
/scontent.cdninstagram.com/t51.2885-19/11375829_1170311332994596_215096937_a.jpg", "id": "2
studios", "profile_pic_url": "https://scontent.cdninstagram.com/t51.2885-
150/10375739_1570777886553507_1095609075_a.jpg", "id": "2137303844"}}, {"user": {"username":
/scontent.cdninstagram.com/t51.2885-19/s150x150/15276481_1115274911853647_26103971301283594
{"username": "usaamaben", "profile_pic_url": "https://scontent.cdninstagram.com/t51.2885-
150/12353195_839178626202257_143040843_a.jpg", "id": "2289118381"}}, {"user": {"username": '
_pic_url": "https://scontent.cdninstagram.com/t51.2885-19/s150x150/15043634_119248181896517
="
```

It means, that I can simply get all data from specific web page even without any authorization. For parsing I used xpath and re library from python.

```
1 script = html.xpath('//script[contains(., "window._sharedData")]/text()')[0]
2 data = re.search(r"window._sharedData = (.*)?;$", script).group(1)
3 data = json.loads(data)
```

All data was stored in **data** array.

### Finding range of instagram ip

### Facebook

### Getting data from fb

### Find ip range of fb

### Experemental results

## 8. Results

## 9. Our programmms

## 10. Conclusions

## 11. Acknowledgments

# Bibliography

- [1] Hofstede, Rick; Celeda, Pavel; Trammell, Brian; Drago, Idilio; Sadre, Ramin; Sperotto, Anna; Pras, Aiko. Flow Monitoring Explained: From Packet Capture to Data Analysis with NetFlow and IPFIX". IEEE Communications Surveys Tutorials. IEEE Communications Society.
- [2] <http://www.businessinsider.com/instagram-2010-11>