



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

# Privacy Implications of Exposing Git Meta Data

Arne Beer

Matriculation number: 6489196

9. Juli 2018

## Table of Contents

---

1. Aggregation

2. Research

3. Conclusion and Outlook

Main topic of the thesis

Can Git metadata be used maliciously?

## Motivation

---



- Used in most projects for source control

## Motivation

---



- Used in most projects for source control
- No obvious leak of personal information

## Motivation

---



- Used in most projects for source control
- No obvious leak of personal information
- Leaked information could be used maliciously

## Leading Question and Goals

---

- Feasibility of scanning repositories on different scales
- Possible extraction of interesting information
- Analyse possible attack goals

## Git Metadata

---

```
tree      cd7d001b696db430b898b75c633686067e6f0b76
parent    c19b969705e5eae0ccca2cde1d8a98be1a1eab4d
author    Arne Beer <contact@arne.beer> 1513434723 +0100
committer Arne Beer <contact@arne.beer> 1513434723 +0100
```

### Chapter 2, acronyms

- Parent commit file reference
- Tree file reference
- Name and email
- Commit timestamp with UTC offset



# GitHub

- Largest accumulation of open-source Git repositories

# GitHub

- Largest accumulation of open-source Git repositories
- Great API

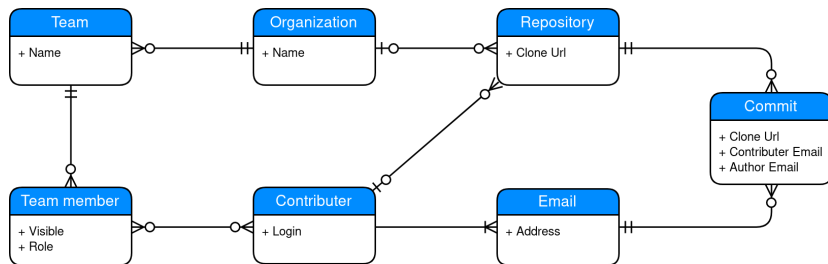
# GitHub

- Largest accumulation of open-source Git repositories
- Great API
- Organizations

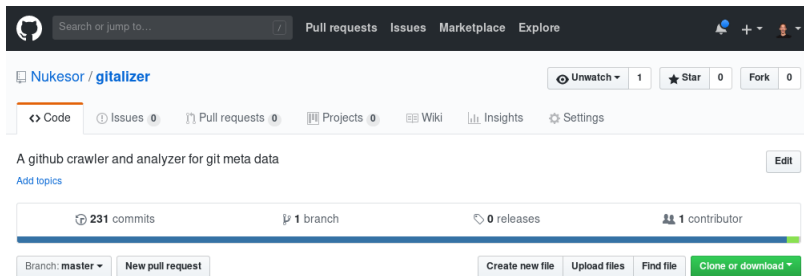
# GitHub

- Largest accumulation of open-source Git repositories
- Great API
- Organizations
- Allows exploration for collecting user repositories.
  - Stars
  - Following

# Github Data Structure



# Gitalizer



- Gathers data from Github
- Uses user or organization as entry point
- Highly optimized

## Three Chosen Attacks

---

- Holiday and sick leave detection

## Three Chosen Attacks

---

- Holiday and sick leave detection
- Sleep rhythm and working hours



## Three Chosen Attacks

---

- Holiday and sick leave detection
- Sleep rhythm and working hours
- Geographic location

## Holiday and Sick Leave: Goals

---

- Automatic detection

## Holiday and Sick Leave: Goals

---

- Automatic detection
- Accurate detection

# Example

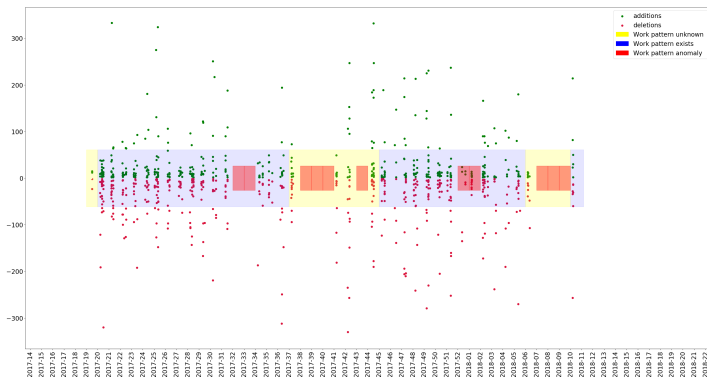


Abbildung: Holiday and Sick leave visualization

## Results

---

- Tested in a small company

## Results

---

- Tested in a small company
- Quite accurate

## Results

---

- Tested in a small company
- Quite accurate
- Needs interpretation

## Sleep Rhythm and Working Hours: Goals

---

- Detection



## Sleep Rhythm and Working Hours: Goals

---

- Detection
- Good visualization

## Sleep Rhythm and Working Hours: Goals

---

- Detection
- Good visualization
- Further implications of rhythm

# Example

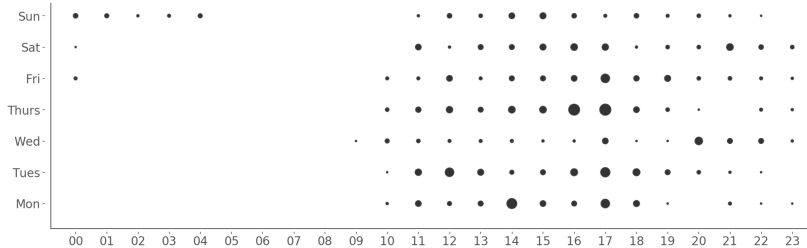


Abbildung: Regular working hour cluster

# Example

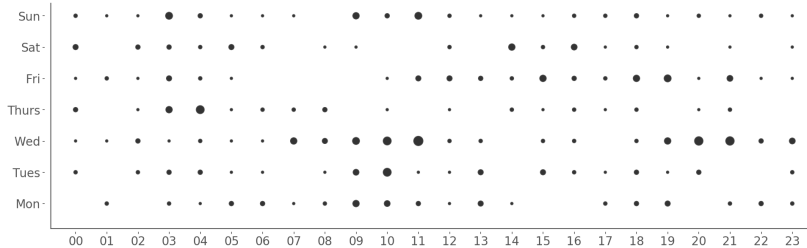


Abbildung: Person without a sleep rhythm

## Results

---

- Shows general tendency

## Results

---

- Shows general tendency
- Unsuitable for direct personal mapping

## Results

---

- Shows general tendency
- Unsuitable for direct personal mapping
- Allows to guess further information

## Geographic Location: Goals

---

- Detect holiday destinations



## Geographic Location: Goals

---

- Detect holiday destinations
- Detect home country

## Geographic Location: Goals

---

- Detect holiday destinations
- Detect home country
- Detect time periods

## Methodology

---

- Periodically check commits

## Methodology

---

- Periodically check commits
- Daylight Savings Time

## Example

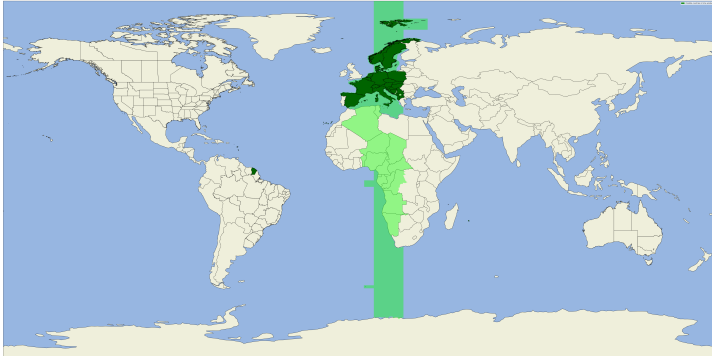


Abbildung: Home location analysis

## Results

---

- Good detection of home country

## Results

---

- Good detection of home country
- Holiday not checked

## Results

---

- Good detection of home country
- Holiday not checked
- Needs better libraries



## Conclusion

---

- Recall the goal: Is it possible to extract personal information

## Conclusion

---

- Recall the goal: Is it possible to extract personal information
- Scanning on small to middle scale

## Outlook

---

- It can become a problem

## Outlook

---

- It can become a problem
- Many more complex and promising attack vectors

## Outlook

---

- It can become a problem
- Many more complex and promising attack vectors
- Methodologies to obfuscate data

Fin

---

Thank you for your attention.