TKPERM: Cross-platform Permission Knowledge Transfer to Detect Overprivileged Third-party Applications

Faysal Hossain Shezan, Kaiming Cheng, Zhen Zhang, Yinzhi Cao, Yuan Tian
Permission-based Access Control

Android

Camera
- take pictures and videos

Contacts
- read your contacts
- find accounts on the device
- modify your contacts

Location
- access approximate location (network-based) only in the foreground

Chrome

IFTTT needs permission to access:
- Connect MESH
- Connect Google Drive
- Create IFTTT Account

Already have an account?
Not now
Allow

Android

Chrome

IFTTT
Permission Correlation with Description

Android App

Permission Correlation with Description

Android App

Uber Technologies, Inc.

Showing permissions for all versions of this app

Location
  approximate location (network-based)

Requested Permission

Permission Correlation with Description

The app uses your location so your driver knows where to pick you up.

Uber Description

Location Permission

Requested Permission

Android App

The app uses your location so your driver knows where to pick you up.
What is Overprivileged?

GamingHub
(Chrome Extension)

GamingHub- https://chrome.google.com/webstore/detail/gaminghub/eafoakfmpnpdecnhhaallihdhhkgjin
What is Overprivileged?

GamingHub (Chrome Extension)

Location Permission

Requested Permission

GamingHub - https://chrome.google.com/webstore/detail/gaminghub/eafoaklfmpnpedcnnhaalihkdbhkug
What is Overprivileged?

GamingHub (Chrome Extension)

Location Permission

Requested Permission

GamingHub Description

Primary Features:
1. Quick & Easy Access to popular web games
2. Minimalist & Elegant Design
3. Hand Picked High Quality Wallpapers that change according to mood
4. New & Exciting ways for accessing Online Content
5. Let us know what you'd like, more to come soon!
What is Overprivileged?

GamingHub (Chrome Extension)

Location Permission

Requested Permission

No Explanation for the Usage of Location Permission

GamingHub Description

GamingHub - Instant & Elegant Access to Online Web Games

GamingHub enables you to quick & elegant access to some of the most popular web games to date. It does so by displaying them as quick access links on your New Tab Page, which, if you like your games, makes for a quick access with a few simple clicks.

Coming soon: We are working hard on delivering the ability to pick & choose which games will be presented or quick access in order to deliver a more personalized experience.

Primary Features:
1. Quick & Easy Access to popular web games
2. Minimalist & Elegant Design
3. Hand Picked High Quality Wallpapers that change according to mood
4. New & Exciting ways for accessing Online Content
5. Let us know what you’d like, more to come soon!

GamingHub- https://chrome.google.com/webstore/detail/gaminghub/eafokfmpnpeanhaalihdbkgh
Challenges

Number of publicly known "IoT Platforms" (2015-2019)

Taken from: https://iot-analytics.com/iot-platform-companies-landscape-2020/
Challenges

Extensive data labeling and parameter tuning on new platforms

Some platforms have limited data

Taken from: https://iot-analytics.com/iot-platform-companies-landscape-2020/
Key Insights

Android App

FlySmart
FlySmart mobile application  Travel & Local
Everyone

⚠️ You don't have any devices.
You can share this with your family. Learn more about Family Library

Permission Knowledge
Location

Chrome App

Oplao weather
offered by oplao.com

⭐⭐⭐⭐⭐ (477)
5,017 users

Overview
Accurate weather forecast. Local to Global.
The best Chrome weather extension. 5 star rated. Easy to use. Oplao weather plugin for Google Chrome contains status bar icon, current weather, detailed forecast, 3 day forecast, fast locations change button (up to 7 locations).
Solution - Transfer Learning
Goal

General framework to detect unexpected permissions
Research Questions

1. What knowledge to transfer? (e.g., what original domain should we select, what permissions in Android should we use)?

2. How to minimize the amount of labeled data needed?
System Overview of TKPERM
System Overview of **TKPERM**

**Source Platform**
- Read Contacts
- Access Coarse Location
- Access Fine Location
- ..........
- Camera

→ Domain Selection
System Overview of **TKPERM**
System Overview of TKPERM

Source Platform
- Read Contacts
- Access Coarse Location
- Access Fine Location
- Camera

Source Model Training

Domain Selection

Source Model
System Overview of TKPERM

Source Platform
- Read Contacts
- Access Coarse Location
- Access Fine Location
- Camera

Target Platforms
- Chrome Geolocation
- Chrome Proxy
- Chrome Content Settings
- SmartThings Switch

Source Platform:
1. Domain Selection
2. Source Model Training
3. Source Model
System Overview of **TKPERM**

**Source Platform**
- Read Contacts
- Access Coarse Location
- Access Fine Location
- Camera

**Source Model**
- Training

**Target Platforms**
- Chrome Geolocation
- Chrome Proxy
- Chrome Content Settings
- SmartThings Switch

1. Domain Selection
2. Source Model Training
3. Source Model
4. Data Selection
5. \( + \)
System Overview of TKPERM

Source Platform
- Read Contacts
- Access Coarse Location
- Access Fine Location
- Camera

Source Model Training

Domain Selection

Source Model

Target Model Training

Data Selection

Target Platforms
- Chrome Geolocation
- Chrome Proxy
- Chrome Content Settings
- SmartThings Switch
System Overview of **TKPERM**

**Source Platform**
- Read Contacts
- Access Coarse Location
- Access Fine Location
- Camera

**Target Platforms**
- Chrome Geolocation
- Chrome Proxy
- Chrome Content Settings
- SmartThings Switch

1. Domain Selection
2. Source Model Training
3. Source Model
4. Data Selection
5. Target Model Training
6. Target Model
7. +
8. +

Source Model

Target Model
System Overview of **TKPERM**

Source Platform:
- Read Contacts
- Access Coarse Location
- Access Fine Location
- Camera

Target Platforms:
- Chrome Geolocation
- Chrome Proxy
- Chrome Content Settings
- SmartThings Switch

1. Domain Selection
2. Source Model Training
3. Source Model
4. + 5. Data Selection
6. 7. Target Model Training
7. Target Model
8. Source Model
Domain Selection

Research Question: What knowledge to transfer?

Greedy Selection Approach

Compute and aggregate source domain(s) performs
Domain Selection

Research Question: What knowledge to transfer?

Greedy Selection Approach

Compute and aggregate source domain(s) performs

Remove source domain(s) which work worst
Domain Selection

Research Question: What knowledge to transfer?

Greedy Selection Approach

- Compute and aggregate source domain(s) performs
- Remove source domain(s) which work worst
- Find the best combination of the source domain(s)
Domain Selection

Greedy Selection Approach

- Compute and aggregate source domain(s) performs
- Remove source domain(s) which work worst
- Find the best combination of the source domain(s)

Research Question: What knowledge to transfer?
Data Selection

Research Question: How to minimize the amount of labeled data needed?

Use source model to rank the unlabeled document
Data Selection

Research Question: How to minimize the amount of labeled data needed?

- Use source model to rank the unlabeled document
- Pick the top 20 documents from the target domain
Data Selection

Research Question: How to minimize the amount of labeled data needed?

- Use source model to rank the unlabeled document
- Pick the top 20 documents from the target domain
- Ask human annotator to label data
Data Selection

Use source model to rank the unlabeled document

Pick the top 20 documents from the target domain

Ask human annotator to label data

✔ Research Question: How to minimize the amount of labeled data needed?
Dataset

- Android: 36,193 sentences
- Chrome: 4,705 sentences
- SmartThings: 292 sentences
- IFTTT: 666 sentences

AutoCog: Measuring the Description-to-permission Fidelity in Android Applications, Qu et al. (CCS 2014)
Evaluation

Question 1. What is the end-to-end performance of TKPERM?
Question 2. What is the performance of each component in TKPERM?
Question 3. What is the computation overhead of TKPERM?
Evaluation

Question 1. What is the end-to-end performance of TKPERM?
Question 2. What is the performance of each component in TKPERM?
Question 3. What is the computation overhead of TKPERM?
Evaluation (Effectiveness)

**Source Domain Selection:** H-divergence v/s Greedy Selection in IFTTT Platform

<table>
<thead>
<tr>
<th>Target Domain</th>
<th>Source Selection</th>
<th>Source Domain(s)</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evernote</td>
<td>H-Divergence</td>
<td>Read Calendar</td>
<td>75.86%</td>
</tr>
<tr>
<td></td>
<td>Greedy Selection</td>
<td>Coarse Location + Fine Location + Camera</td>
<td>83.13%</td>
</tr>
<tr>
<td>BMW Lab</td>
<td>H-Divergence</td>
<td>Read Contact</td>
<td>92.30%</td>
</tr>
<tr>
<td></td>
<td>Greedy Selection</td>
<td>Send SMS + Record Audio</td>
<td>95.24%</td>
</tr>
<tr>
<td>Facebook</td>
<td>H-Divergence</td>
<td>Read Calendar</td>
<td>76.09%</td>
</tr>
<tr>
<td></td>
<td>Greedy Selection</td>
<td>Camera</td>
<td>88.09%</td>
</tr>
<tr>
<td>Google Calendar</td>
<td>H-Divergence</td>
<td>Read Calendar</td>
<td>91.30%</td>
</tr>
<tr>
<td></td>
<td>Greedy Selection</td>
<td>Read Calendar + Coarse Location</td>
<td>92.30%</td>
</tr>
<tr>
<td>Google Contact</td>
<td>H-Divergence</td>
<td>Read Contacts</td>
<td>99.20%</td>
</tr>
<tr>
<td></td>
<td>Greedy Selection</td>
<td>Read Contacts</td>
<td>99.20%</td>
</tr>
</tbody>
</table>
## Evaluation (Effectiveness)

### Source Domain Selection: H-divergence v/s Greedy Selection in IFTTT Platform

<table>
<thead>
<tr>
<th>Target Domain</th>
<th>Source Selection</th>
<th>Source Domain(s)</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evernote</td>
<td>H-Divergence</td>
<td>Read Calendar</td>
<td>75.86%</td>
</tr>
<tr>
<td></td>
<td>Greedy Selection</td>
<td>Coarse Location + Fine Location + Camera</td>
<td>83.13%</td>
</tr>
<tr>
<td>BMW Lab</td>
<td>H-Divergence</td>
<td>Read Contact</td>
<td>92.30%</td>
</tr>
<tr>
<td></td>
<td>Greedy Selection</td>
<td>Send SMS + Record Audio</td>
<td>95.24%</td>
</tr>
<tr>
<td>Facebook</td>
<td>H-Divergence</td>
<td>Read Calendar</td>
<td>76.09%</td>
</tr>
<tr>
<td></td>
<td>Greedy Selection</td>
<td>Camera</td>
<td>88.09%</td>
</tr>
<tr>
<td>Google Calendar</td>
<td>H-Divergence</td>
<td>Read Calendar</td>
<td>91.30%</td>
</tr>
<tr>
<td></td>
<td>Greedy Selection</td>
<td>Read Calendar + Coarse Location</td>
<td>92.30%</td>
</tr>
<tr>
<td>Google Contact</td>
<td>H-Divergence</td>
<td>Read Contacts</td>
<td>99.20%</td>
</tr>
<tr>
<td></td>
<td>Greedy Selection</td>
<td>Read Contacts</td>
<td>99.20%</td>
</tr>
</tbody>
</table>
## Evaluation (Effectiveness)

### Data Selection: Comparison of With & Without Data Selection

<table>
<thead>
<tr>
<th>Platform</th>
<th>Performance</th>
<th>Configuration</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No Transfer</td>
<td>Without Data Selection</td>
<td>With Data Selection</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F1 Score</td>
<td>84.25%</td>
<td>91.08%</td>
<td>91.83%</td>
<td></td>
</tr>
<tr>
<td>IFTTT</td>
<td>Improvement</td>
<td>-</td>
<td>6.83%</td>
<td>7.58%</td>
<td></td>
</tr>
<tr>
<td>Chrome</td>
<td>F1 Score</td>
<td>70.60%</td>
<td>84.36%</td>
<td>89.13%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Improvement</td>
<td>-</td>
<td>13.76%</td>
<td>18.53%</td>
<td></td>
</tr>
<tr>
<td>SmartThings</td>
<td>F1 Score</td>
<td>72.80%</td>
<td>84.65%</td>
<td>89.1%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Improvement</td>
<td>-</td>
<td>11.85%</td>
<td>16.3%</td>
<td></td>
</tr>
</tbody>
</table>
### Evaluation (Effectiveness)

**Data Selection**: Comparison of With & Without Data Selection

<table>
<thead>
<tr>
<th>Platform</th>
<th>Performance</th>
<th>Configuration</th>
<th>No Transfer</th>
<th>Without Data Selection</th>
<th>With Data Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFTTT</td>
<td>F1 Score</td>
<td></td>
<td>84.25%</td>
<td>91.08%</td>
<td>91.83%</td>
</tr>
<tr>
<td></td>
<td>Improvement</td>
<td></td>
<td>-</td>
<td>6.83%</td>
<td>7.58%</td>
</tr>
<tr>
<td>Chrome</td>
<td>F1 Score</td>
<td></td>
<td>70.60%</td>
<td>84.36%</td>
<td>89.13%</td>
</tr>
<tr>
<td></td>
<td>Improvement</td>
<td></td>
<td>-</td>
<td>13.76%</td>
<td>18.53%</td>
</tr>
<tr>
<td>SmartThings</td>
<td>F1 Score</td>
<td></td>
<td>72.80%</td>
<td>84.65%</td>
<td>89.1%</td>
</tr>
<tr>
<td></td>
<td>Improvement</td>
<td></td>
<td>-</td>
<td>11.85%</td>
<td>16.3%</td>
</tr>
</tbody>
</table>
Evaluation (Effectiveness)

**TKPERM Performance Analysis (Metric: F1 Score)**

<table>
<thead>
<tr>
<th>Platform</th>
<th>Target Domain</th>
<th>Source Domain</th>
<th>Transfer</th>
<th>No Transfer</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFTTT</td>
<td>Evernote</td>
<td>Coarse Location + Fine Location + Camera</td>
<td>83.13%</td>
<td>79.78%</td>
<td>3.35%</td>
</tr>
<tr>
<td></td>
<td>BMW Lab</td>
<td>Send SMS + Record Audio</td>
<td>95.24%</td>
<td>85.71%</td>
<td>9.53%</td>
</tr>
<tr>
<td></td>
<td>Facebook</td>
<td>Camera</td>
<td>88.09%</td>
<td>75.00%</td>
<td>13.09%</td>
</tr>
<tr>
<td></td>
<td>Google Calendar</td>
<td>Read Calendar + Coarse Location</td>
<td>94.30%</td>
<td>83.54%</td>
<td>10.76%</td>
</tr>
<tr>
<td></td>
<td>Google Contact</td>
<td>Read Contact</td>
<td>98.41%</td>
<td>97.22%</td>
<td>1.19%</td>
</tr>
<tr>
<td>Chrome</td>
<td>Geolocation</td>
<td>Fine Location + Coarse Location + Read Contact</td>
<td>88.29%</td>
<td>62.50%</td>
<td>25.79%</td>
</tr>
<tr>
<td></td>
<td>Proxy</td>
<td>Send SMS + Fine Location</td>
<td>93.78%</td>
<td>89.69%</td>
<td>4.09%</td>
</tr>
<tr>
<td></td>
<td>Content Settings</td>
<td>Fine Location + Read Contact</td>
<td>85.31%</td>
<td>59.61%</td>
<td>25.70%</td>
</tr>
<tr>
<td>SmartThings</td>
<td>Lock</td>
<td>Write Setting</td>
<td>85.71%</td>
<td>75.00%</td>
<td>10.71%</td>
</tr>
<tr>
<td></td>
<td>Motion Sensor</td>
<td>Read Contact</td>
<td>87.10%</td>
<td>53.33%</td>
<td>33.77%</td>
</tr>
<tr>
<td></td>
<td>Switch</td>
<td>Send SMS + Read Calendar</td>
<td>94.39%</td>
<td>90.09%</td>
<td>4.30%</td>
</tr>
</tbody>
</table>
### Evaluation (Effectiveness)

#### TKPERM Performance Analysis (Metric: F1 Score)

<table>
<thead>
<tr>
<th>Platform</th>
<th>Target Domain</th>
<th>Source Domain</th>
<th>Transfer</th>
<th>No Transfer</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFTTT</td>
<td>Evernote</td>
<td>Coarse Location + Fine Location + Camera</td>
<td>83.13%</td>
<td>79.78%</td>
<td>3.35%</td>
</tr>
<tr>
<td></td>
<td>BMW Lab</td>
<td>Send SMS + Record Audio</td>
<td>95.24%</td>
<td>85.71%</td>
<td>9.53%</td>
</tr>
<tr>
<td></td>
<td>Facebook</td>
<td>Camera</td>
<td>88.09%</td>
<td>75.00%</td>
<td>13.09%</td>
</tr>
<tr>
<td></td>
<td>Google Calendar</td>
<td>Read Calendar + Coarse Location</td>
<td>94.30%</td>
<td>83.54%</td>
<td>10.76%</td>
</tr>
<tr>
<td></td>
<td>Google Contact</td>
<td>Read Contact</td>
<td>98.41%</td>
<td>97.22%</td>
<td>1.19%</td>
</tr>
<tr>
<td>Chrome</td>
<td>Geolocation</td>
<td>Fine Location + Coarse Location + Read Contact</td>
<td>88.29%</td>
<td>62.50%</td>
<td>25.79%</td>
</tr>
<tr>
<td></td>
<td>Proxy</td>
<td>Send SMS + Fine Location</td>
<td>93.78%</td>
<td>89.69%</td>
<td>4.09%</td>
</tr>
<tr>
<td></td>
<td>Content Settings</td>
<td>Fine Location + Read Contact</td>
<td>85.31%</td>
<td>59.61%</td>
<td>25.70%</td>
</tr>
<tr>
<td>SmartThings</td>
<td>Lock</td>
<td>Write Setting</td>
<td>85.71%</td>
<td>75.00%</td>
<td>10.71%</td>
</tr>
<tr>
<td></td>
<td>Motion Sensor</td>
<td>Read Contact</td>
<td>87.10%</td>
<td>53.33%</td>
<td>33.77%</td>
</tr>
<tr>
<td></td>
<td>Switch</td>
<td>Send SMS + Read Calendar</td>
<td>94.39%</td>
<td>90.09%</td>
<td>4.30%</td>
</tr>
</tbody>
</table>
## Evaluation (Effectiveness)

**TKPERM Performance Analysis (Metric: F1 Score)**

<table>
<thead>
<tr>
<th>Platform</th>
<th>Target Domain</th>
<th>Source Domain</th>
<th>Transfer</th>
<th>No Transfer</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFTTT</td>
<td>Evernote</td>
<td>Coarse Location + Fine Location + Camera</td>
<td>83.13%</td>
<td>79.78%</td>
<td>3.35%</td>
</tr>
<tr>
<td></td>
<td>BMW Lab</td>
<td>Send SMS + Record Audio</td>
<td>95.24%</td>
<td>85.71%</td>
<td>9.53%</td>
</tr>
<tr>
<td></td>
<td>Facebook</td>
<td>Camera</td>
<td>88.09%</td>
<td>75.00%</td>
<td>13.09%</td>
</tr>
<tr>
<td></td>
<td>Google Calendar</td>
<td>Read Calendar + Coarse Location</td>
<td>94.30%</td>
<td>83.54%</td>
<td>10.76%</td>
</tr>
<tr>
<td></td>
<td>Google Contact</td>
<td>Read Contact</td>
<td>98.41%</td>
<td>97.22%</td>
<td>1.19%</td>
</tr>
<tr>
<td>Chrome</td>
<td>Geolocation</td>
<td>Fine Location + Coarse Location + Read Contact</td>
<td>88.29%</td>
<td>62.50%</td>
<td>25.79%</td>
</tr>
<tr>
<td></td>
<td>Proxy</td>
<td>Send SMS + Fine Location</td>
<td>93.78%</td>
<td>89.69%</td>
<td>4.09%</td>
</tr>
<tr>
<td></td>
<td>Content Settings</td>
<td>Fine Location + Read Contact</td>
<td>85.31%</td>
<td>59.61%</td>
<td>25.70%</td>
</tr>
<tr>
<td>SmartThings</td>
<td>Lock</td>
<td>Write Setting</td>
<td>85.71%</td>
<td>75.00%</td>
<td>10.71%</td>
</tr>
<tr>
<td></td>
<td>Motion Sensor</td>
<td>Read Contact</td>
<td>87.10%</td>
<td>53.33%</td>
<td>33.77%</td>
</tr>
<tr>
<td></td>
<td>Switch</td>
<td>Send SMS + Read Calendar</td>
<td>94.39%</td>
<td>90.09%</td>
<td>4.30%</td>
</tr>
</tbody>
</table>
## Evaluation (Effectiveness)

**TKPERM Performance Analysis (Metric: F1 Score)**

<table>
<thead>
<tr>
<th>Platform</th>
<th>Target Domain</th>
<th>Source Domain</th>
<th>Transfer</th>
<th>No Transfer</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFTTT</td>
<td>Evernote</td>
<td>Coarse Location + Fine Location + Camera</td>
<td>83.13%</td>
<td>79.78%</td>
<td>3.35%</td>
</tr>
<tr>
<td></td>
<td>BMW Lab</td>
<td>Send SMS + Record Audio</td>
<td>95.24%</td>
<td>85.71%</td>
<td>9.53%</td>
</tr>
<tr>
<td></td>
<td>Facebook</td>
<td>Camera</td>
<td>88.09%</td>
<td>75.00%</td>
<td>13.09%</td>
</tr>
<tr>
<td></td>
<td>Google Calendar</td>
<td>Read Calendar + Coarse Location</td>
<td>94.30%</td>
<td>83.54%</td>
<td>10.76%</td>
</tr>
<tr>
<td></td>
<td>Google Contact</td>
<td>Read Contact</td>
<td>98.41%</td>
<td>97.22%</td>
<td>1.19%</td>
</tr>
<tr>
<td></td>
<td>Chrome</td>
<td>Geolocation + Fine Location + Read Contact</td>
<td>88.29%</td>
<td>62.50%</td>
<td>25.79%</td>
</tr>
<tr>
<td></td>
<td>Proxy</td>
<td>Send SMS + Fine Location</td>
<td>93.78%</td>
<td>89.69%</td>
<td>4.09%</td>
</tr>
<tr>
<td></td>
<td>Content Settings</td>
<td>Fine Location + Read Contact</td>
<td>85.31%</td>
<td>59.61%</td>
<td>25.70%</td>
</tr>
<tr>
<td></td>
<td>SmartThings</td>
<td>Lock Write Setting</td>
<td>85.71%</td>
<td>75.00%</td>
<td>10.71%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Motion Sensor Read Contact</td>
<td>87.10%</td>
<td>53.33%</td>
<td>33.77%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Switch Send SMS + Read Calendar</td>
<td>94.39%</td>
<td>90.09%</td>
<td>4.30%</td>
</tr>
</tbody>
</table>

12.93% improvement compared to No Transfer
Evaluation

Question 1. What is the end-to-end performance of TKPERM?
Question 2. What is the performance of each component in TKPERM?
Question 3. What is the computation overhead of TKPERM?
Evaluation

Question 1. What is the end-to-end performance of TKPERM?
Question 2. What is the performance of each component in TKPERM?
Question 3. What is the computation overhead of TKPERM?
Evaluation (Scalability)

**Computation Overhead** (Run in Amazon Elastic Compute Cloud (EC2), NVIDIA Tesla V100)

<table>
<thead>
<tr>
<th>Platform</th>
<th>Target Domain</th>
<th>Time (hh:mm:ss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFTTT</td>
<td>Evernote</td>
<td>33:27:03</td>
</tr>
<tr>
<td></td>
<td>BMW Lab</td>
<td>14:08:40</td>
</tr>
<tr>
<td></td>
<td>Facebook</td>
<td>22:57:20</td>
</tr>
<tr>
<td></td>
<td>Google Calendar</td>
<td>15:15:18</td>
</tr>
<tr>
<td></td>
<td>Google Contact</td>
<td>18:40:17</td>
</tr>
<tr>
<td>Chrome</td>
<td>Geolocation</td>
<td>07:37:28</td>
</tr>
<tr>
<td></td>
<td>Proxy</td>
<td>06:54:01</td>
</tr>
<tr>
<td></td>
<td>Content Settings</td>
<td>09:42:45</td>
</tr>
<tr>
<td>SmartThings</td>
<td>Lock</td>
<td>03:47:59</td>
</tr>
<tr>
<td></td>
<td>Motion Sensor</td>
<td>04:09:44</td>
</tr>
<tr>
<td></td>
<td>Switch</td>
<td>14:11:08</td>
</tr>
</tbody>
</table>
# Evaluation (Scalability)

**Computation Overhead** (Run in Amazon Elastic Compute Cloud (EC2), NVIDIA Tesla V100)

<table>
<thead>
<tr>
<th>Platform</th>
<th>Target Domain</th>
<th>Time (hh:mm:ss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFTTT</td>
<td>Evernote</td>
<td>33:27:03</td>
</tr>
<tr>
<td></td>
<td>BMW Lab</td>
<td>14:08:40</td>
</tr>
<tr>
<td></td>
<td>Facebook</td>
<td>22:57:20</td>
</tr>
<tr>
<td></td>
<td>Google Calendar</td>
<td>15:15:18</td>
</tr>
<tr>
<td></td>
<td>Google Contact</td>
<td>18:40:17</td>
</tr>
<tr>
<td>Chrome</td>
<td>Geolocation</td>
<td>07:37:28</td>
</tr>
<tr>
<td></td>
<td>Proxy</td>
<td>06:54:01</td>
</tr>
<tr>
<td></td>
<td>Content Settings</td>
<td>09:42:45</td>
</tr>
<tr>
<td>SmartThings</td>
<td>Lock</td>
<td>03:47:59</td>
</tr>
<tr>
<td></td>
<td>Motion Sensor</td>
<td>04:09:44</td>
</tr>
<tr>
<td></td>
<td>Switch</td>
<td>14:11:08</td>
</tr>
</tbody>
</table>
Measurement Result

114 Chrome Apps
(35.73%)
Measurement Result

- 135 IFTTT Apps (28.36%)
- 114 Chrome Apps (35.73%)
Measurement Result

135 IFTTT Apps (28.36%)

114 Chrome Apps (35.73%)

80 SmartThings Apps (32.9%)
Measurement Result

- 135 IFTTT Apps (28.36%)
- 114 Chrome Apps (35.73%)
- 80 SmartThings Apps (32.9%)
- 329 Overprivileged Apps (32.33%)
Conclusion

1. General Framework

General framework to detect Overprivileged applications in new platforms
Conclusion

1. General Framework

Just got the VR why do some apps ask for so many permissions?

For example some games want permission for Camera "Take pictures and video" Phone "Read phone status and identity" Microphone "Record audio" the game in question is singleplayer. Storage "Modify or delete SD card contents", most games only ask for read content on SD card.

It seems fishy at least that some games ask for so many permissions and to be honest I stayed away from games like this.
1. General Framework

**Conclusion**

TKPERM works well (90.02% F1 score on avg.)

How to identify overprivileged application in VR (new platform)?

Posted in: https://www.reddit.com/r/GearVR/comments/5ga1na/just_got_the_vr_why_do_some_apps_ask_for_so_many/
Conclusion

1. General Framework

Just got the VR why do some apps ask for so many permissions?

For example some games want permission for Camera "Take pictures and video" Phone "Read phone status and identity" Microphone "Record audio" and Storage "Modify or delete SD card contents".

It seems fishy at least that they have so many permissions. So I be honest I stayed away from games like this.
Conclusion

1. General Framework
2. Result

TKPERM works well
(90.02% F1 score on avg.)
Conclusion

1. General Framework
2. Result
3. Public Dataset

Dataset Available at: https://drive.google.com/drive/u/1/folders/1Yfnz-Zp8LfIYldM6jTH-QKE88NcSX
Thank You!

Contact: Faysal Hossain Shezan (Email-fs5ve@virginia.edu)
Conclusion

1. General Framework
2. Result
3. Public Dataset
Backup Slides
<table>
<thead>
<tr>
<th>Plat.</th>
<th>Perform.</th>
<th>Word Embedding</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Target</td>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>IFTTT</td>
<td>F1 score</td>
<td>86%</td>
<td>91.83%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Improv.</td>
<td>-</td>
<td>5.83%</td>
<td></td>
</tr>
<tr>
<td>Chrome</td>
<td>F1 score</td>
<td>88%</td>
<td>89.13%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Improv.</td>
<td>-</td>
<td>1.13%</td>
<td></td>
</tr>
<tr>
<td>Smart Things</td>
<td>F1 score</td>
<td>85%</td>
<td>89.1%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Improv.</td>
<td>-</td>
<td>4.1%</td>
<td></td>
</tr>
<tr>
<td>Permission</td>
<td>Performance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acc.</td>
<td>Prec.</td>
<td>Rec.</td>
<td>F1</td>
</tr>
<tr>
<td>Fine Location</td>
<td>85%</td>
<td>73%</td>
<td>84%</td>
<td>78%</td>
</tr>
<tr>
<td>Coarse Location</td>
<td>84%</td>
<td>53%</td>
<td>84%</td>
<td>65%</td>
</tr>
<tr>
<td>Camera</td>
<td>88%</td>
<td>80%</td>
<td>89%</td>
<td>85%</td>
</tr>
<tr>
<td>Read Calendar</td>
<td>89%</td>
<td>87%</td>
<td>89%</td>
<td>88%</td>
</tr>
<tr>
<td>Read Contact</td>
<td>92%</td>
<td>92%</td>
<td>90%</td>
<td>91%</td>
</tr>
<tr>
<td>Record Audio</td>
<td>84%</td>
<td>83%</td>
<td>83%</td>
<td>83%</td>
</tr>
<tr>
<td>Write Settings</td>
<td>87%</td>
<td>69%</td>
<td>86%</td>
<td>77%</td>
</tr>
<tr>
<td>Send SMS</td>
<td>93%</td>
<td>93%</td>
<td>100%</td>
<td>97%</td>
</tr>
<tr>
<td>Write APN</td>
<td>92%</td>
<td>88%</td>
<td>97%</td>
<td>94%</td>
</tr>
<tr>
<td>Total</td>
<td>88.22%</td>
<td>79.78%</td>
<td>89.11%</td>
<td>84.20%</td>
</tr>
</tbody>
</table>
System Overview (Domain Selection)

### Algorithm 1: Source Domain Selection using Greedy Selection Algorithm

**Input:** Source Domain Data List, \([D_S]\); Target Domain Data, \(d_t\)

**Output:** Aggregated Source List, \([A_S]\)

1. **procedure** SELECTSOURCEDOMAINS
2. \([A_S] \leftarrow \emptyset\)
3. \(P_{\text{best}} \leftarrow -\infty\)
4. \(P_{\text{current}} \leftarrow\) initialize to zero
5. \(\{D_S, d_{f1}\} \leftarrow \text{computealsd}_{f1}([D_S], d_t)\)
6. **while** \(\text{size}([D_S, d_{f1}]) \geq 0\) **do**
7. \(d_s \leftarrow \text{highest}_{f1}([D_S, d_{f1}])\)
8. remove \(d_s\) from \([D_S, d_{f1}]\)
9. add \(d_s\) to \([A_S]\)
10. \(P_{\text{current}} \leftarrow \text{computeds}_{f1}([A_S], d_t)\)
11. **if** \(P_{\text{current}} < P_{\text{best}}\) **then**
12. remove \(d_s\) from \([A_S]\)
13. **break**
14. **end if**
15. \(P_{\text{best}} \leftarrow P_{\text{current}}\)
16. **end while**
17. Return \([A_S]\)
18. **end procedure**
System Overview (Data Selection)

Algorithm 2 Selecting fine-tune dataset for target model using Data Selection Module.

Input: Source Model, $\mathcal{M}_S$; Unlabeled Target Domain Dataset, $[A_U]$

Output: Fine-tune Dataset, $[D_R]$

1: procedure SELECTFINE_TUNEDATASET
2: for each document, $A \in [A_U]$ do
3:     for each sentence, $d_t \in A$ do
4:         pred $\leftarrow$ prediction($d_t$, $\mathcal{M}_S$)
5:         if pred = 1 then
6:             $\mathcal{R}_A \leftarrow \mathcal{R}_A + 1$
7:         end if
8:     end for
9:     add $\{A, \mathcal{R}_A\}$ to $[\mathcal{D}_R]$
10: end for
11: $[\mathcal{D}_R]^* \leftarrow$ sorteddesc($[\mathcal{D}_R]$)
12: $[D_F] \leftarrow$ top20($[\mathcal{D}_R]^*$)
13: Return $[D_F]$
14: end procedure

Ranking:

$$\mathcal{R}_A = \sum_{j=1}^{\text{len}(A)} \left[ \mathcal{P}(y_j | x_j) \right] = 1$$

\(\forall\) sentence, $x_j \in \text{document}, A$

Selection:

$$[D_F] = \left[ \text{sort}_{\text{desc}} \{A_i, \mathcal{R}_A_i \}_{i=1}^{m} \right]_{j=1}^{n}$$
Permission-based Access Control
Goal

G1. Semantic Knowledge
Goal

G1. Semantic Knowledge
G2. Permission Knowledge
System Overview

Source Platform

- Read Contacts
- Access Coarse Location
- Access Fine Location
- ........
- Camera
System Overview

Challenge: What knowledge to transfer?
Domain Selection

Greedy Selection Approach

Aggregate source domain(s) which performs best

Remove source domain(s) which work worst

Find the best combination of the source domain(s)
System Overview

Source Platform

- Read Contacts
- Access Coarse Location
- Access Fine Location
- Camera

Domain Selection

1
System Overview

- Read Contacts
- Access Coarse Location
- Access Fine Location
- Camera

Source Platform

Source Model Training

Domain Selection
System Overview

Source Platform

- Read Contacts
- Access Coarse Location
- Access Fine Location
- Camera

Source Model

Source Model Training

Domain Selection

1

2

3
System Overview

Source Platform
- Read Contacts
- Access Coarse Location
- Access Fine Location
- Camera

Source Model
- Training

Domain Selection

Source Model

Target Platforms
- Chrome Geolocation
- Chrome Proxy
- Chrome Content Settings
- SmartThings Switch
System Overview

Challenge: How to minimize the amount of labeled data needed?
Data Selection

- Use source model to rank the document
- Rank unlabeled documents from the target domain
- Pick the top 20 documents from a target domain
- Ask human annotator to label data
System Overview

Source Platform

1. Read Contacts
2. Access Coarse Location
3. Access Fine Location
4. Camera

Source Model Training

Source Model

Data Selection

Target Platforms

1. Chrome Geolocation
2. Chrome Proxy
3. Chrome Content Settings
4. SmartThings Switch
System Overview

Source Platform

- Read Contacts
- Access Coarse Location
- Access Fine Location
- Camera

Source Model Training

Domain Selection

Source Model

Target Platforms

- Chrome Geolocation
- Chrome Proxy
- Chrome Content Settings
- SmartThings Switch

Data Selection

Target Model Training

1. ......
2. Source Model Training
3. ......
4. + 5.
6. + 7.

Source Model
System Overview

Source Platform
- Read Contacts
- Access Coarse Location
- Access Fine Location
- Camera

Target Platforms
- Chrome Geolocation
- Chrome Proxy
- Chrome Content Settings
- SmartThings Switch

1. Domain Selection
2. Source Model Training
3. Source Model
4. Data Selection
5. Target Model Training
6. Target Model
7. +
8. +

Source Model
Target Model

……….
Conclusion

• IFTTT – 135 apps
• Chrome Extension – 114 apps
• SmartThings – 80 apps