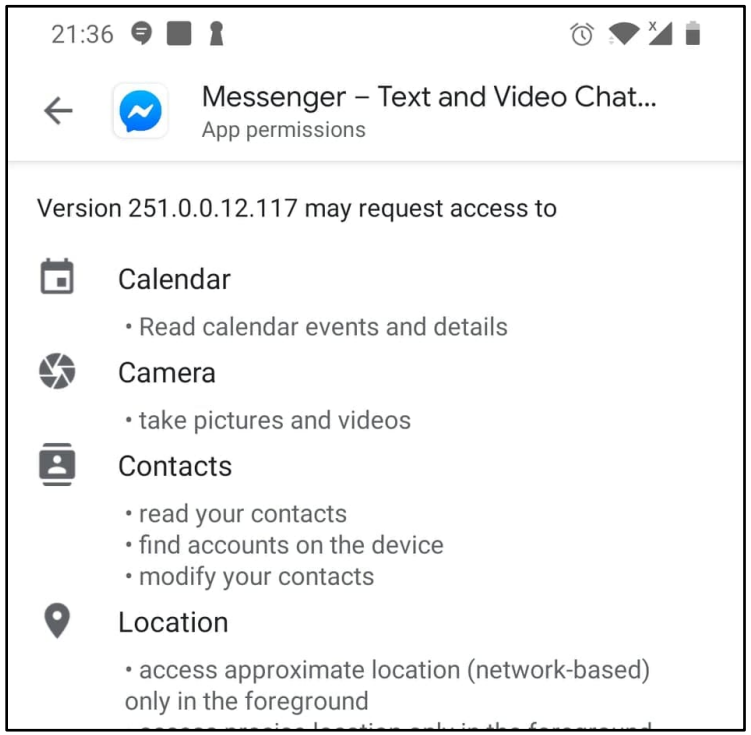
TKPERM: Cross-platform Permission Knowledge Transfer to Detect Overprivileged Third-party Applications

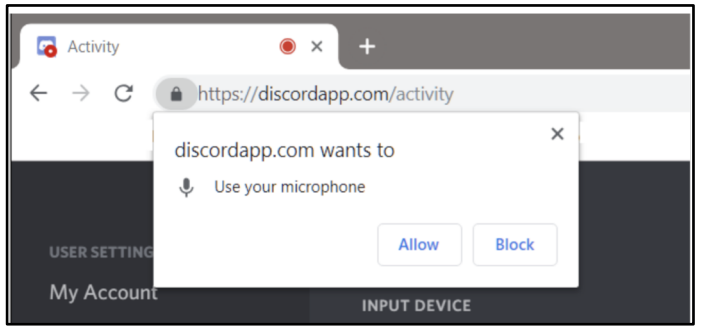
Faysal Hossain Shezan, Kaiming Cheng, Zhen Zhang,

Yinzhi Cao, Yuan Tian

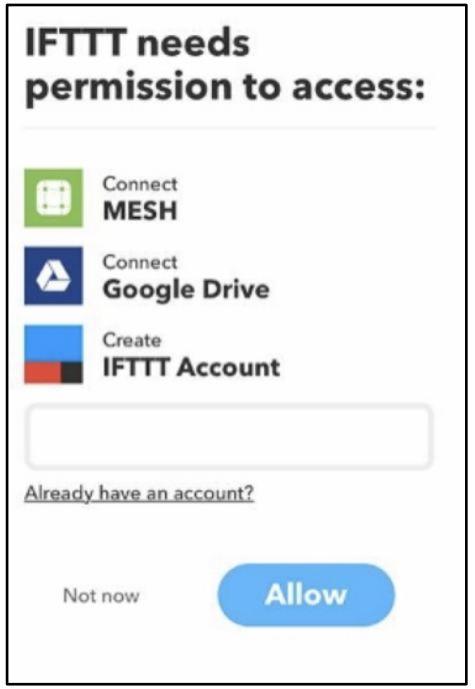
Permission-based Access Control



Android



Chrome



IFTTT

Permission Correlation with Description

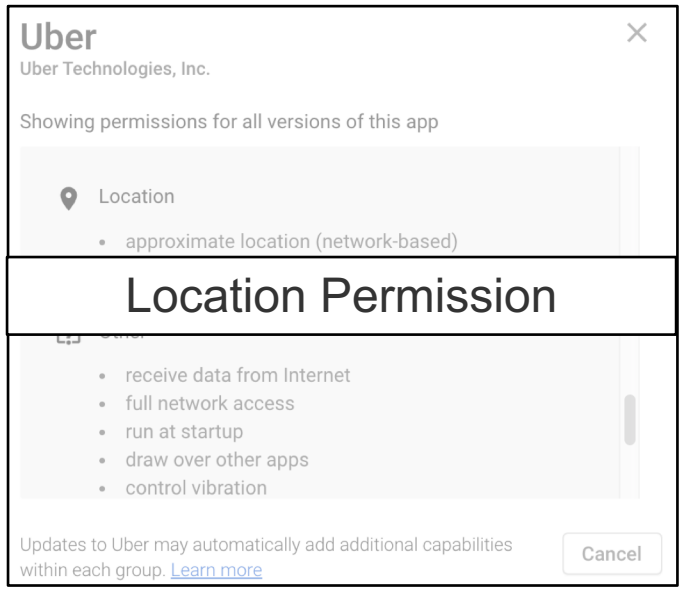
Android App

UBER- https://play.google.com/store/apps/details?id=com.ubercab&hl=en_US

Permission Correlation with Description

UBER

Android App

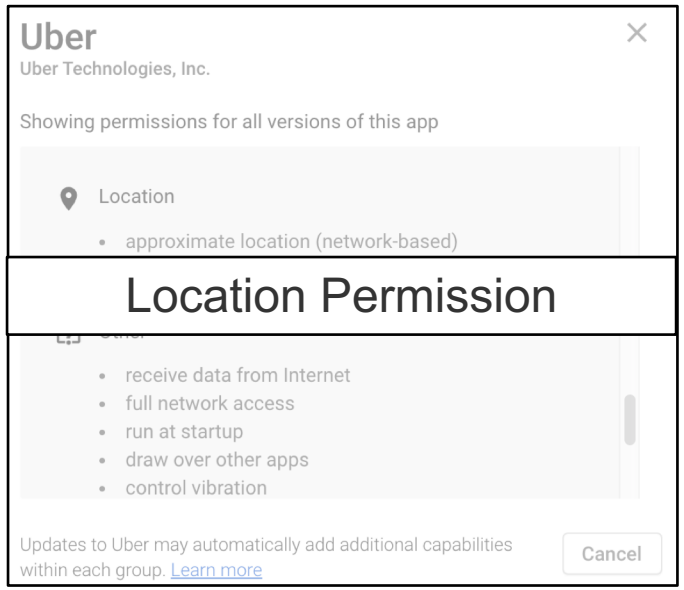


Requested Permission

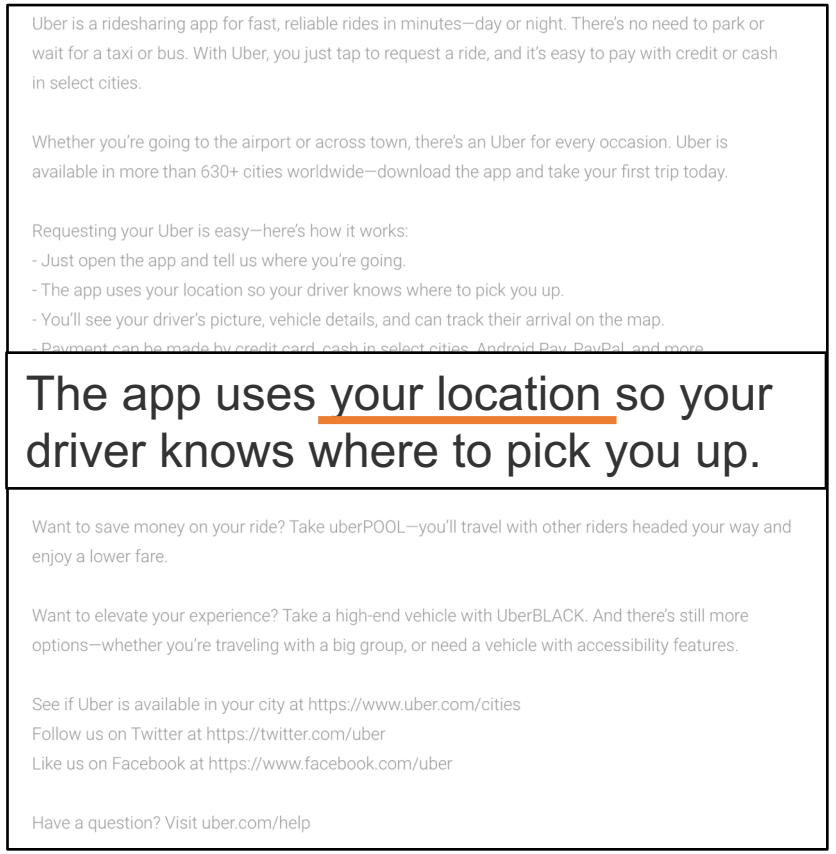
UBER- https://play.google.com/store/apps/details?id=com.ubercab&hl=en_US

Permission Correlation with Description

Android App



Requested Permission

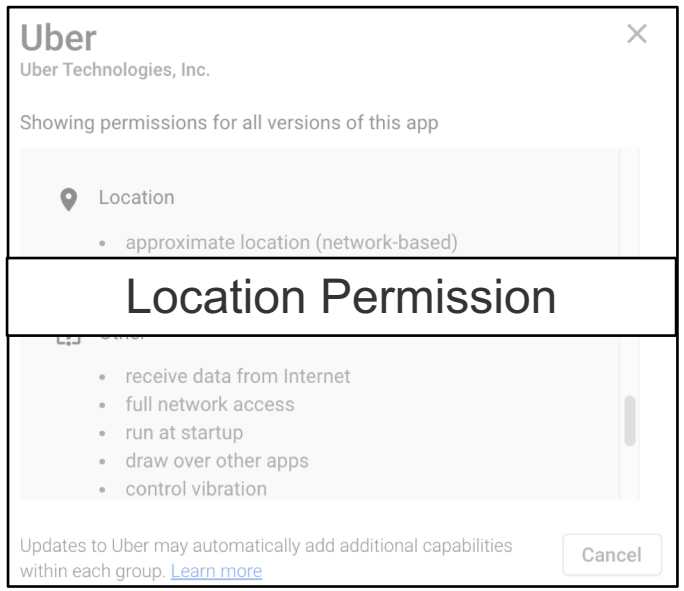


Uber Description

UBER- https://play.google.com/store/apps/details?id=com.ubercab&hl=en_US

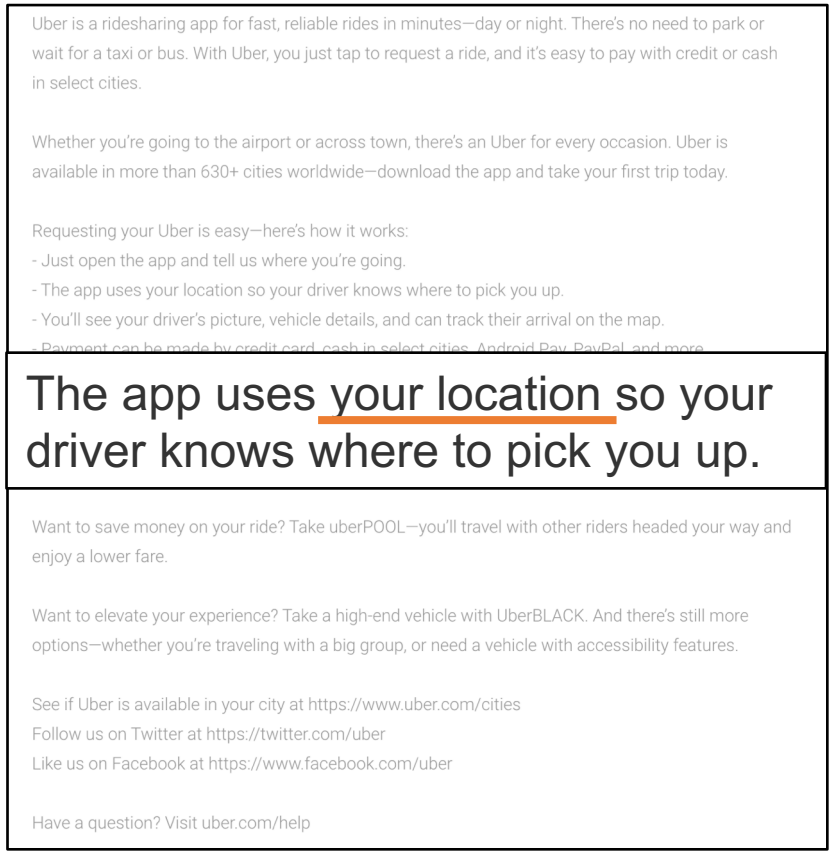
Permission Correlation with Description

Android App



Requested Permission

Consistent



Uber Description

UBER- https://play.google.com/store/apps/details?id=com.ubercab&hl=en_US

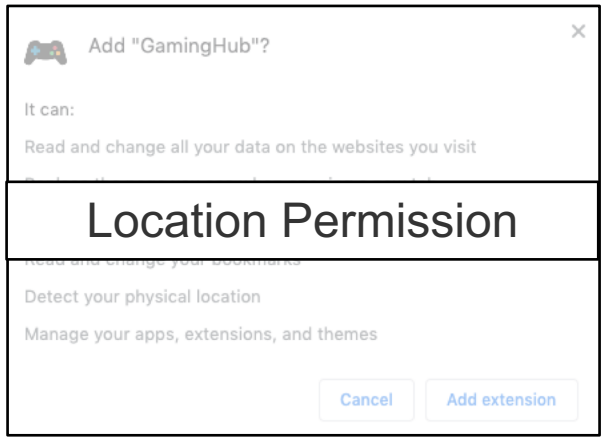
What is Overprivileged?

GamingHub
(Chrome Extension)

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

What is Overprivileged?

GamingHub
(Chrome Extension)

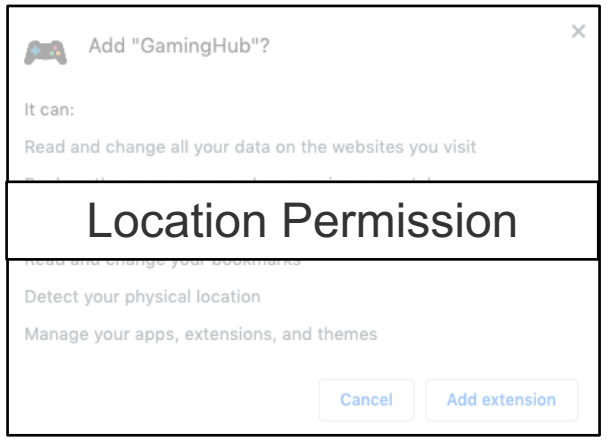


Requested Permission

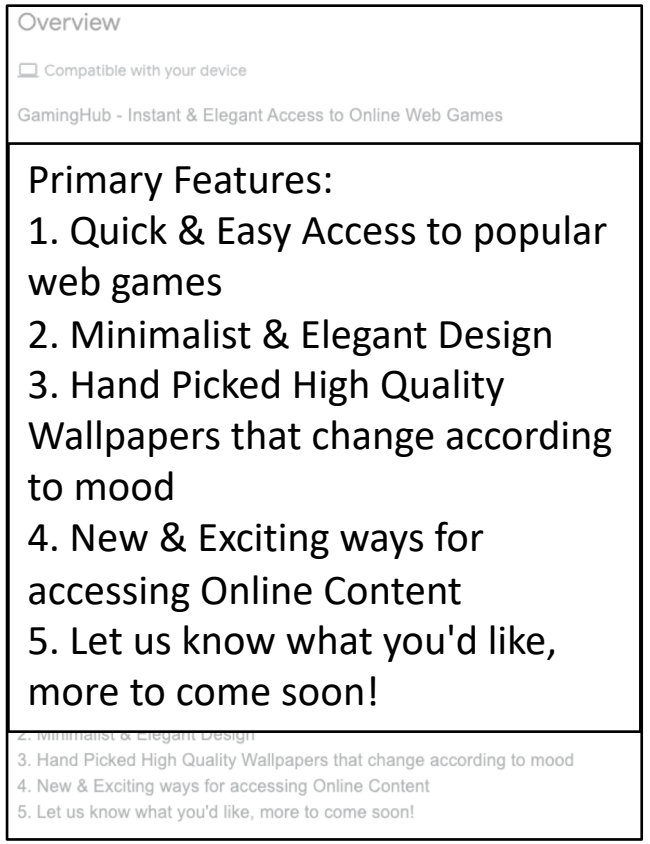
GamingHub- https://chrome.google.com/webstore/detail/gaminghub/eafoaklfmpnpdecnhhaailihkdbhkgin

What is Overprivileged?

GamingHub
(Chrome Extension)



Requested Permission

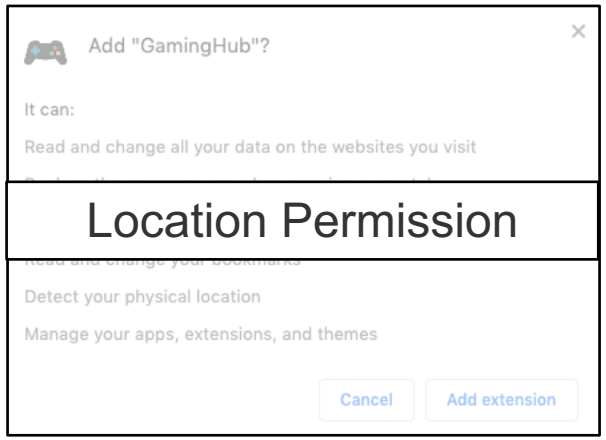


GamingHub Description

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

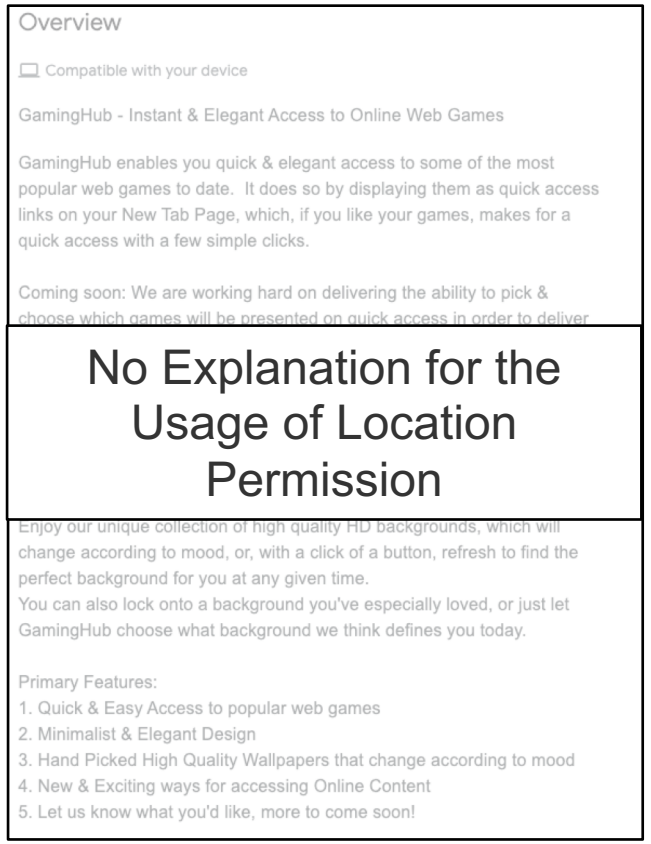
What is Overprivileged?

GamingHub
(Chrome Extension)



Requested Permission

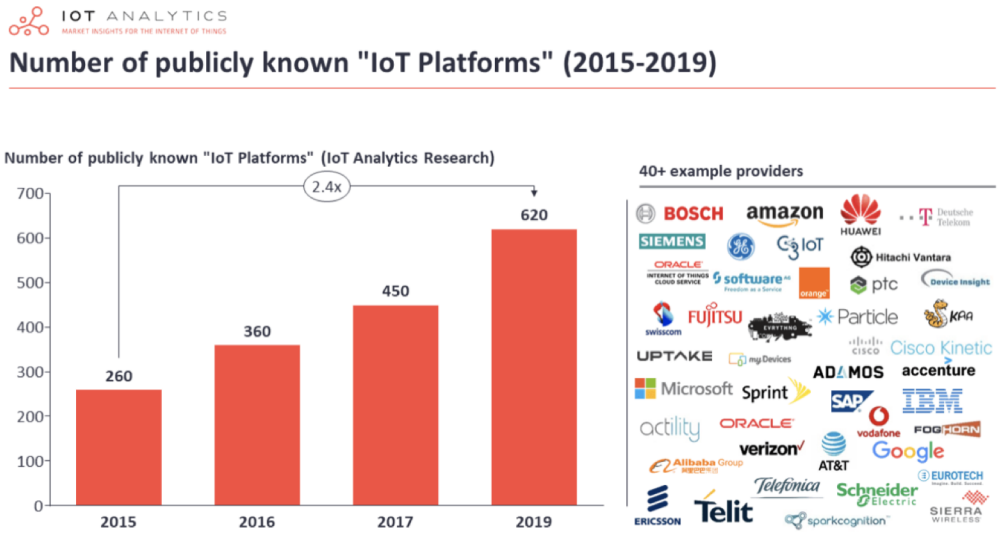
No Match



GamingHub Description

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

Challenges

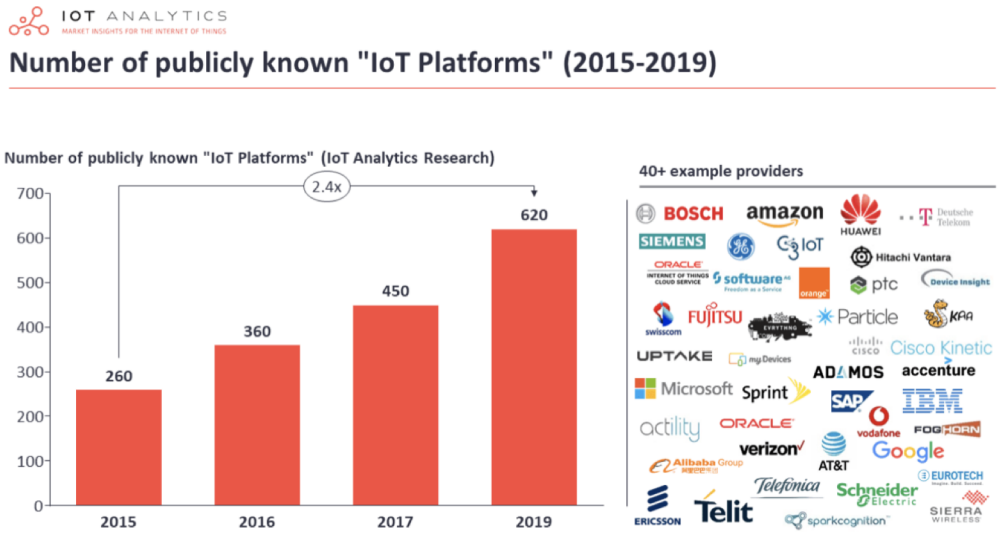


Taken from: https://iot-analytics.com/iot-platform-companies-landscape-2020/

Challenges

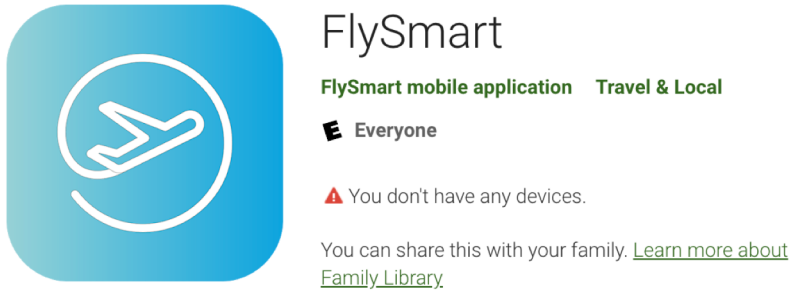
Extensive data labeling and parameter tuning on new platforms

Some platforms have limited data



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

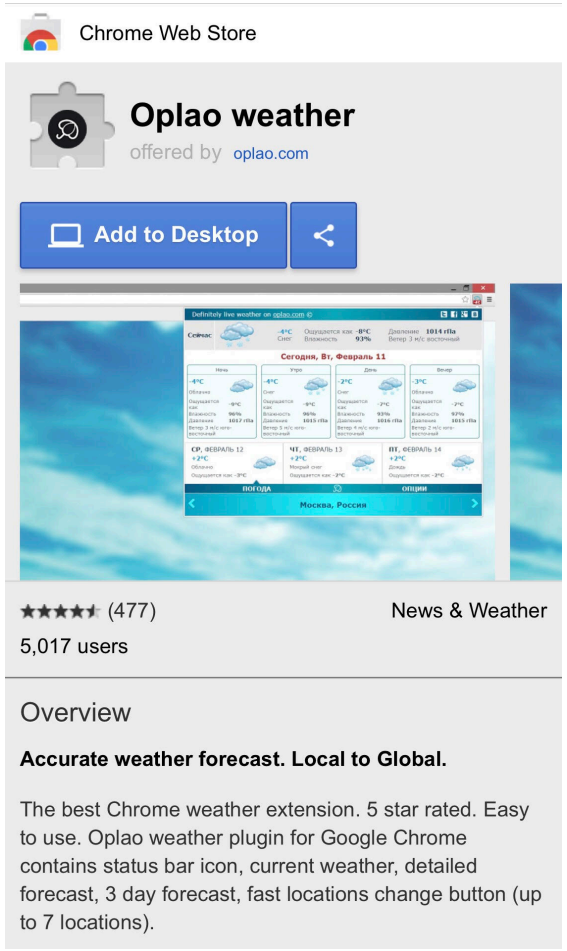
Key Insights



Android App

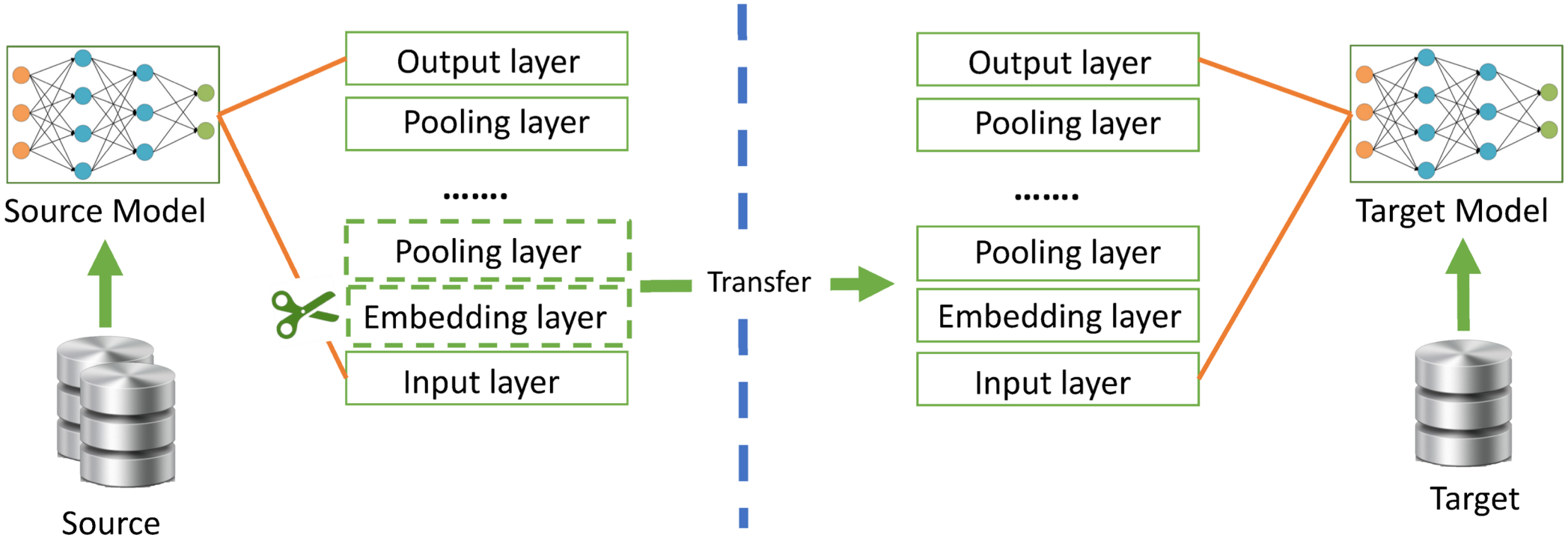
Permission Knowledge

Location



Chrome App

Solution- Transfer Learning



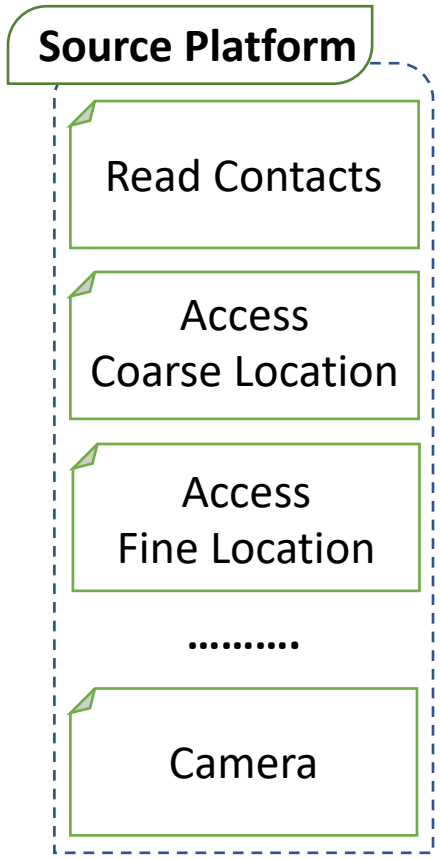
Goal

General framework to detect unexpected permissions

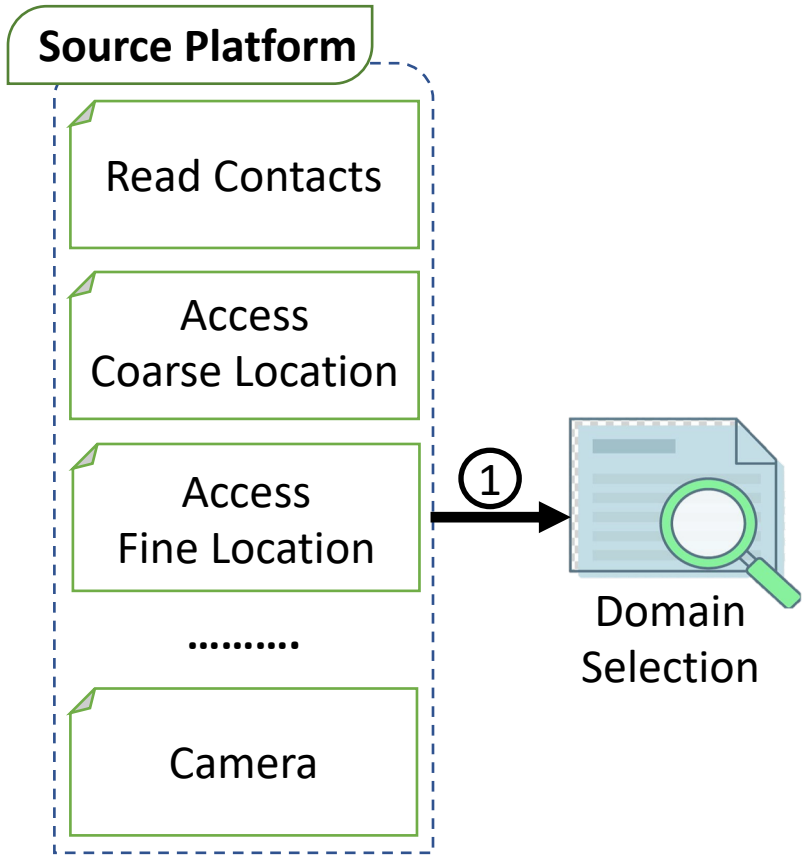
Research Questions

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?

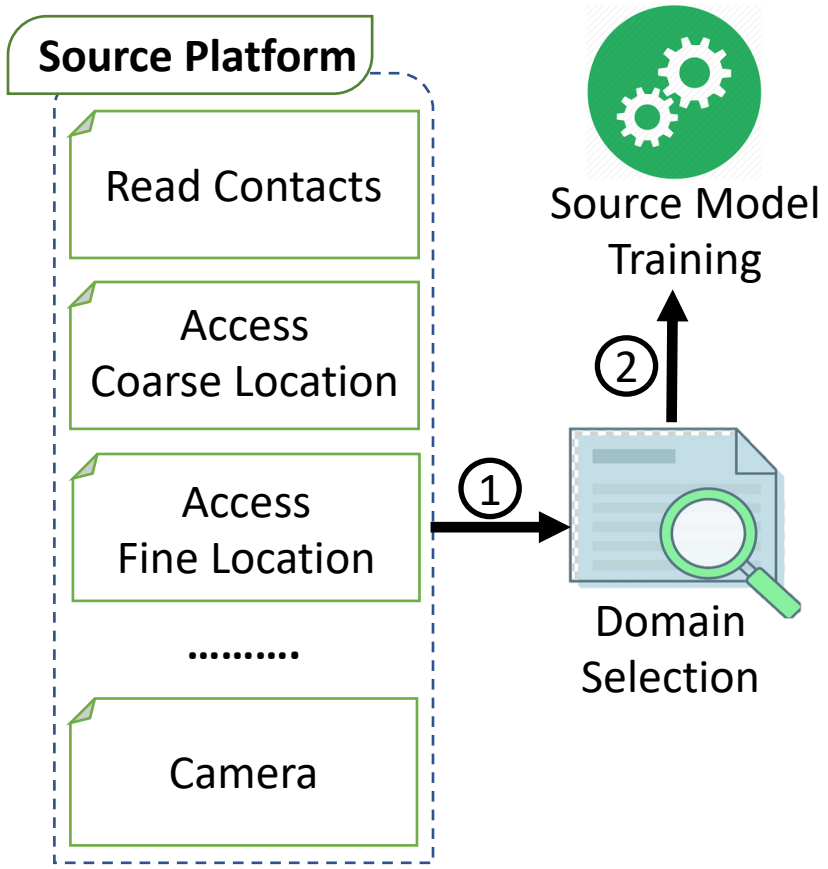
System Overview of TKPERM



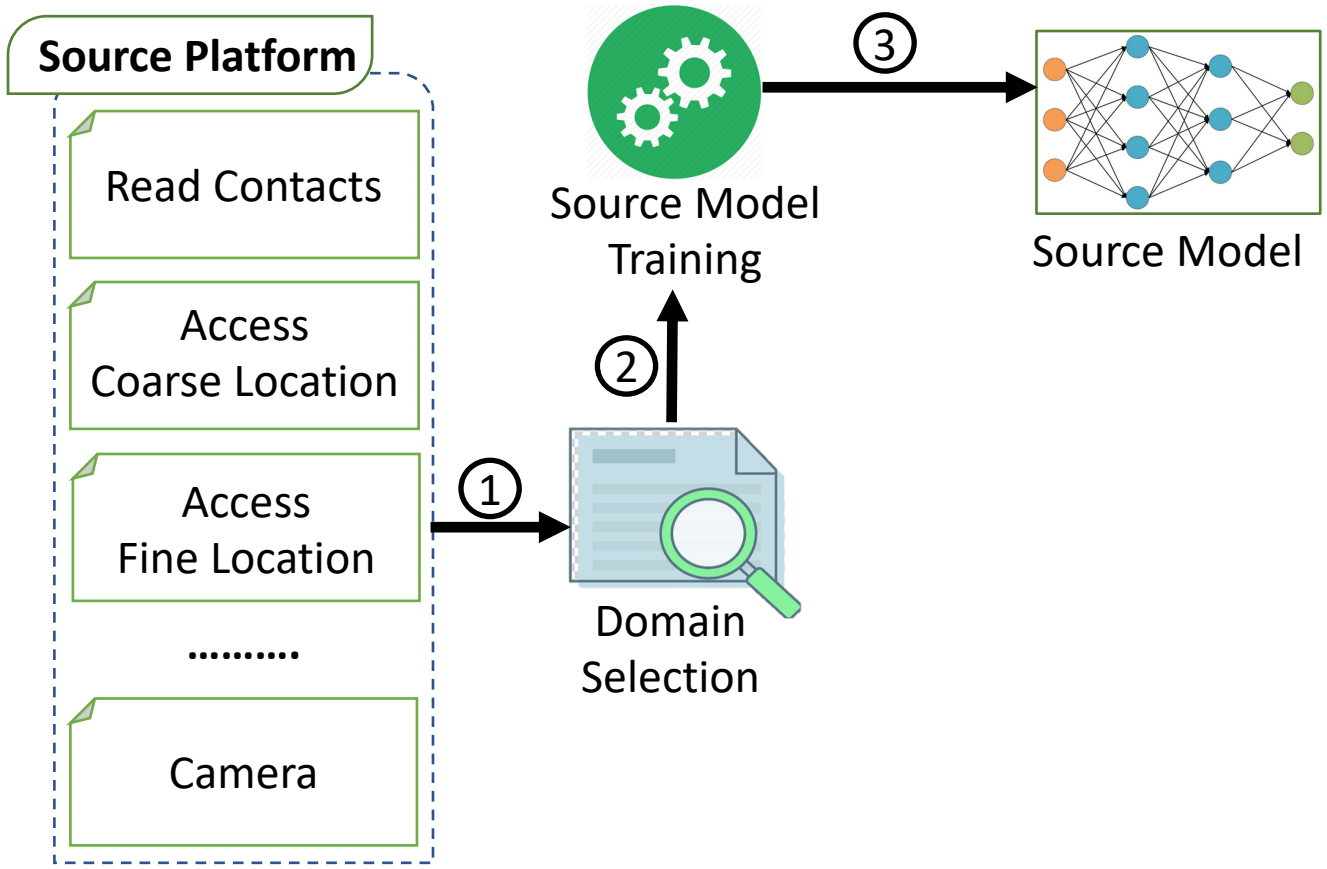
System Overview of TKPERM



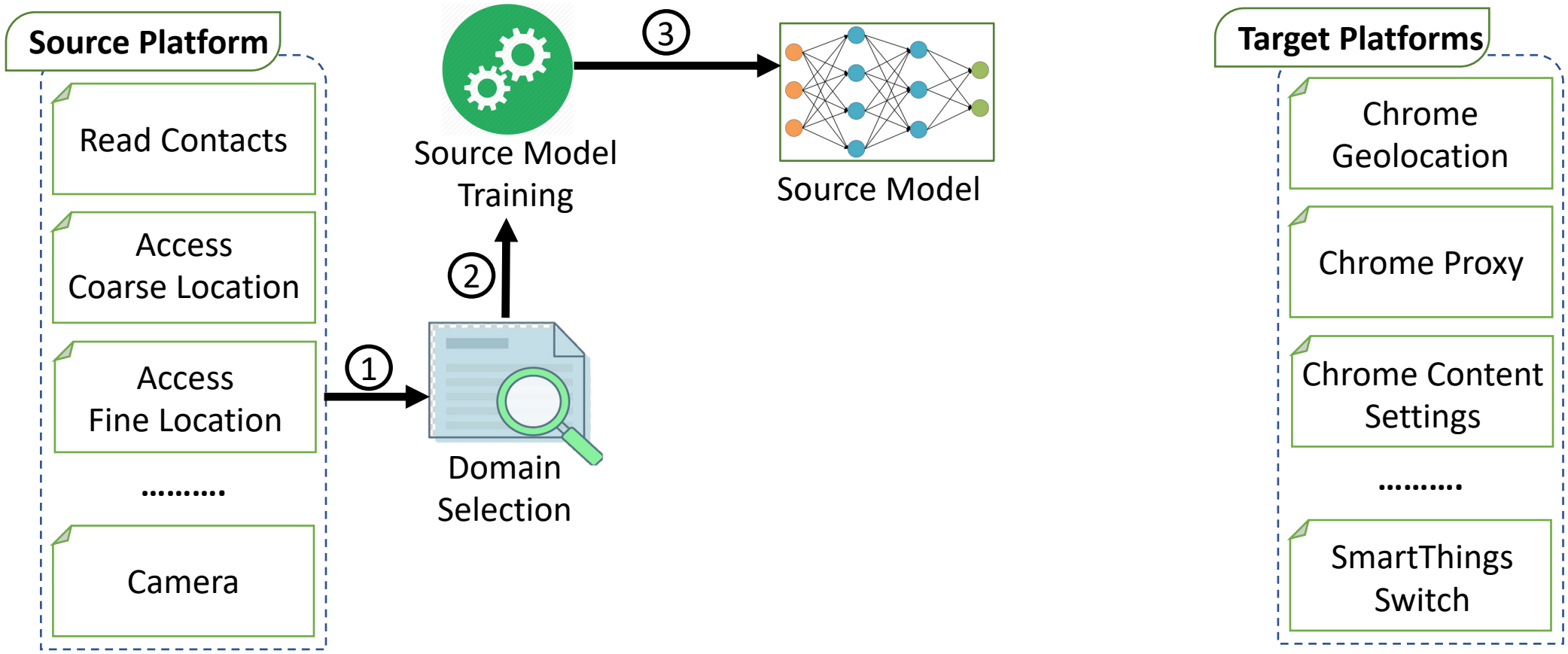
System Overview of TKPERM



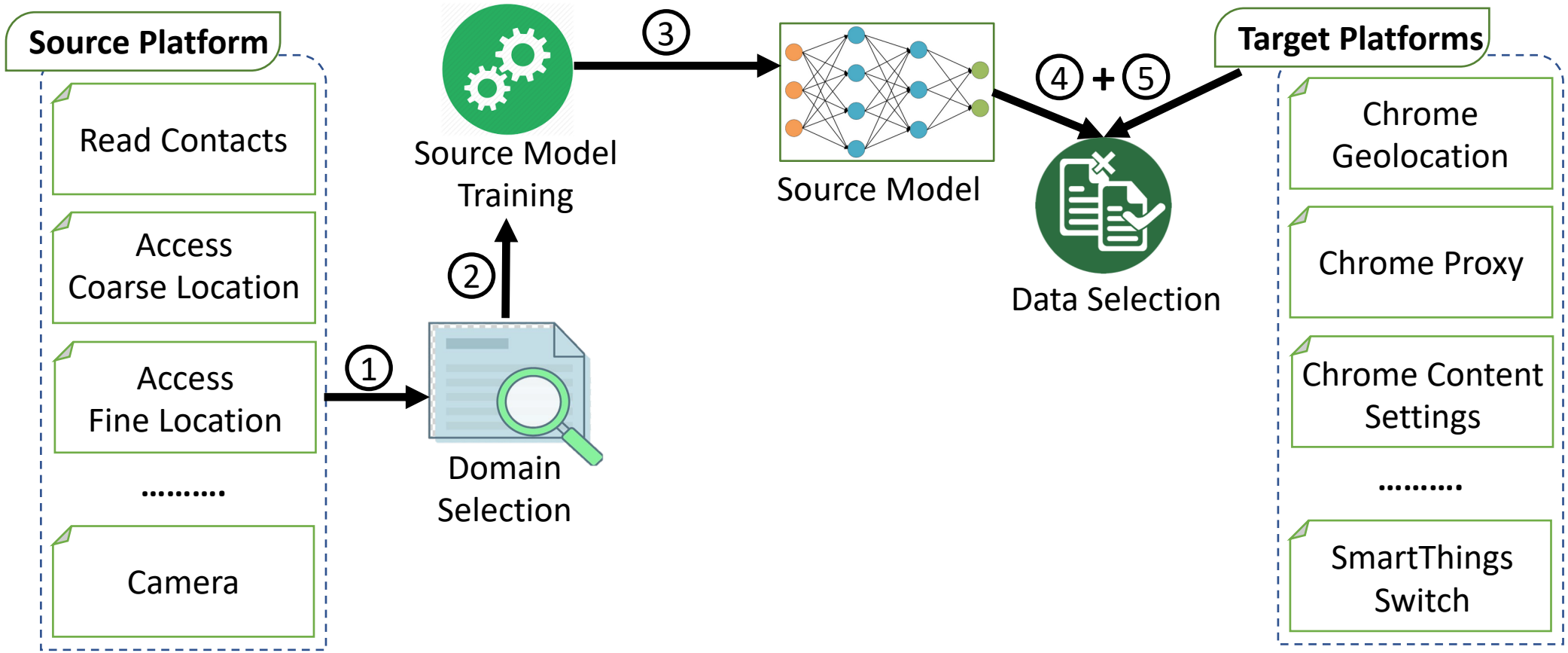
System Overview of TKPERM



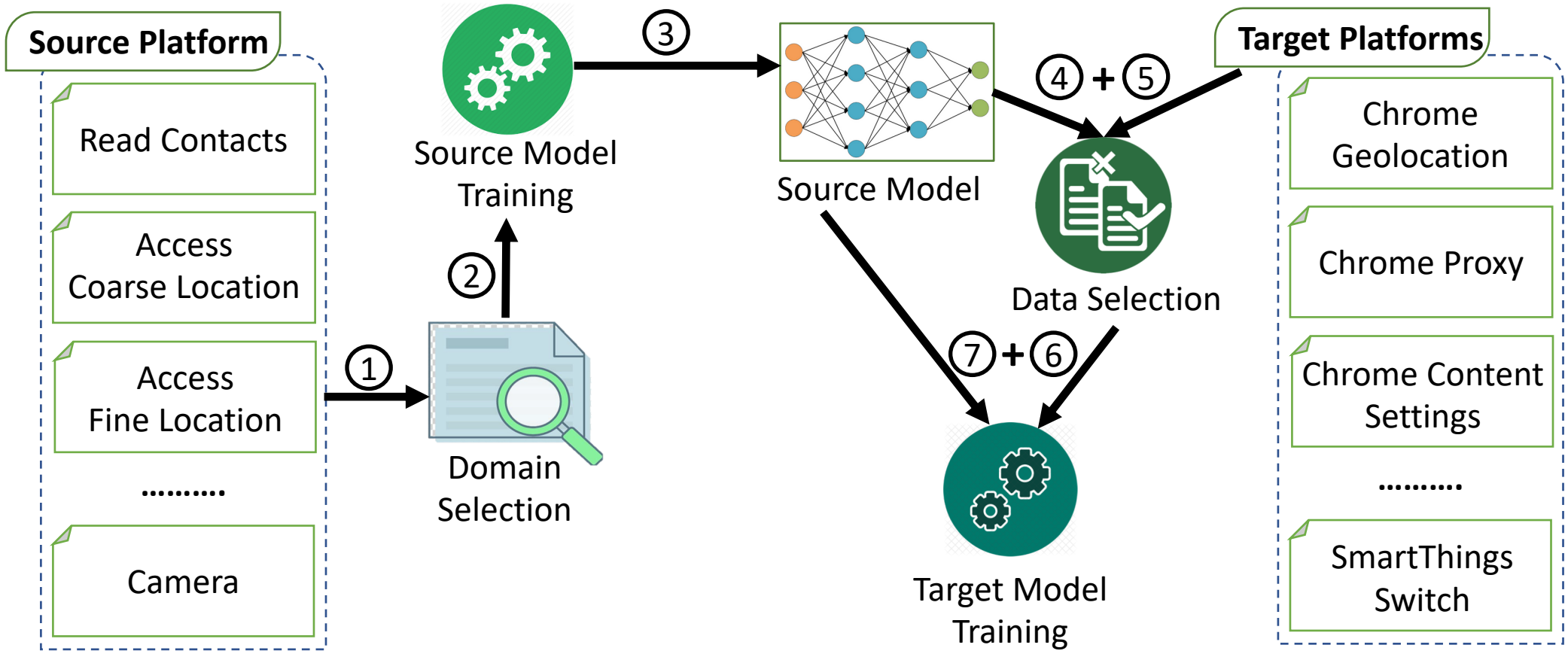
System Overview of TKPERM



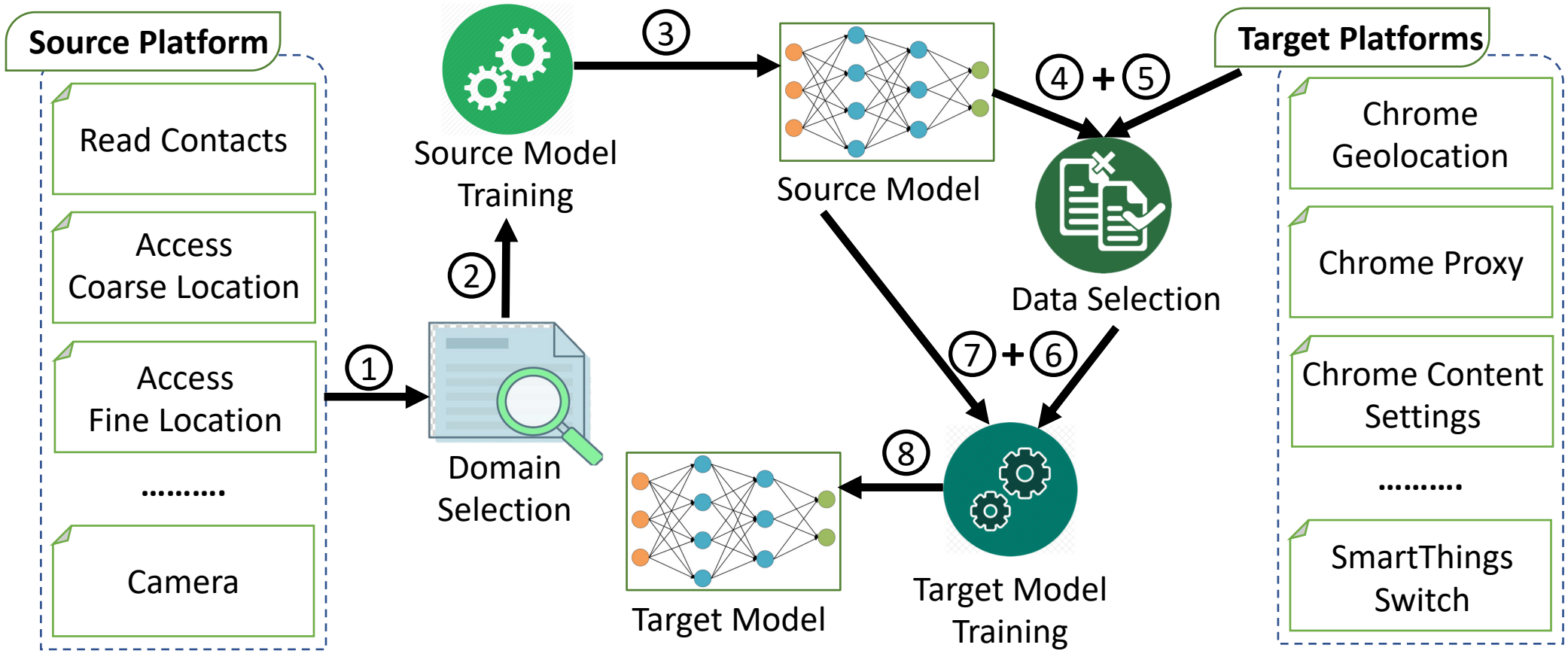
System Overview of TKPERM



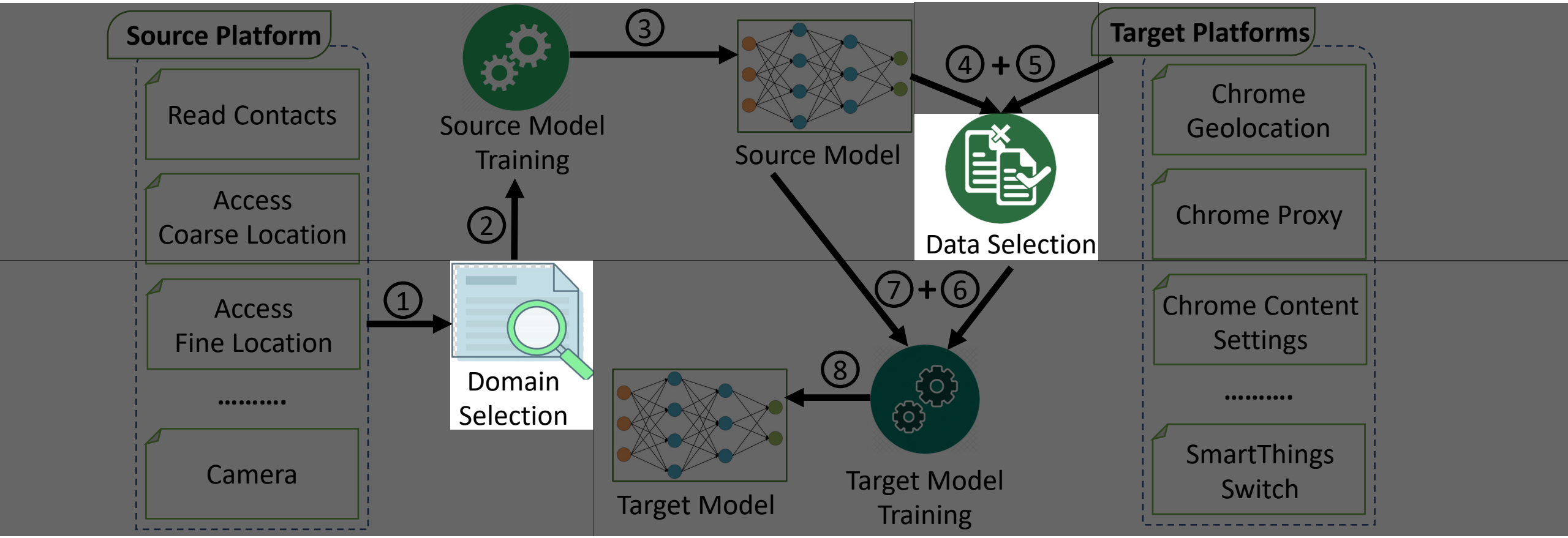
System Overview of TKPERM



System Overview of TKPERM



System Overview of TKPERM



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

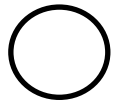
Ask human annotator to label data

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

Dataset

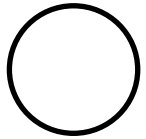
Android

Chrome



SmartThings

292
Sentences

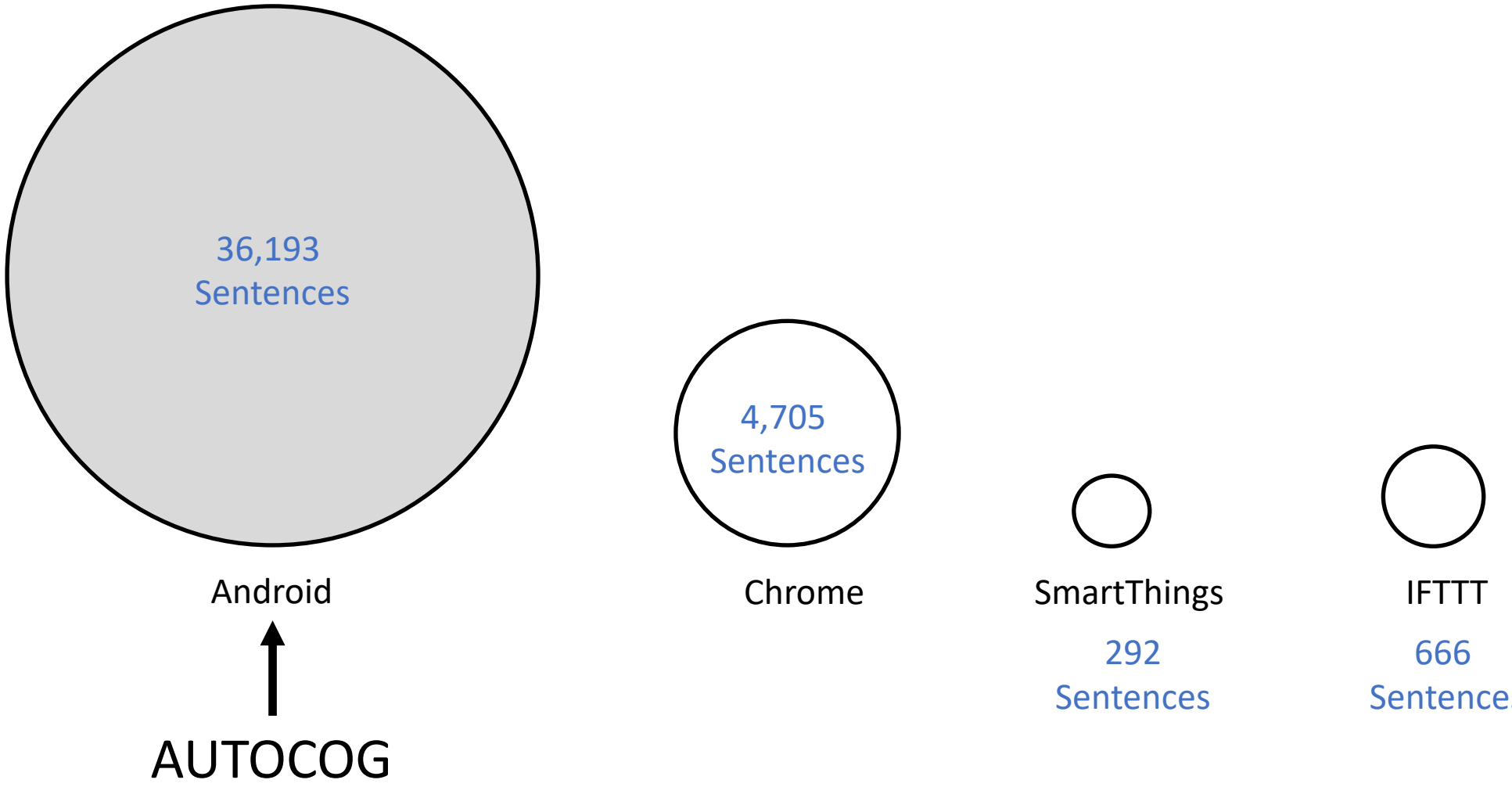


IFTTT

666
Sentences

Available at: https://drive.google.com/drive/u/1/folders/1Yfnz-ZpBpL8lftYIdM6JtH-QKE88NcSX

Dataset



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?

Effectiveness

Question 3. What is the computation overhead of TKPERM?

Evaluation (Effectiveness)

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

Target Domain	Source Selection	Source Domain(s)	F1
Evernote	H-Divergence	Read Calendar	75.86%
	Greedy Selection	Coarse Location + Fine Location + Camera	83.13%
BMW Lab	H-Divergence	Read Contact	92.30%
	Greedy Selection	Send SMS + Record Audio	95.24%
Facebook	H-Divergence	Read Calendar	76.09%
	Greedy Selection	Camera	88.09%
Google Calendar	H-Divergence	Read Calendar	91.30%
	Greedy Selection	Read Calendar + Coarse Location	92.30%
Google Contact	H-Divergence	Read Contacts	99.20%
	Greedy Selection	Read Contacts	99.20%

Evaluation (Effectiveness)

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

Target Domain	Source Selection	Source Domain(s)	F1
Evernote	H-Divergence	Read Calendar	75.86%
	Greedy Selection	Coarse Location + Fine Location + Camera	83.13%
BMW Lab	H-Divergence	Read Contact	92.30%
	Greedy Selection	Send SMS + Record Audio	95.24%
Facebook	H-Divergence	Read Calendar	76.09%
	Greedy Selection	Camera	88.09%
Google Calendar	H-Divergence	Read Calendar	91.30%
	Greedy Selection	Read Calendar + Coarse Location	92.30%
Google Contact	H-Divergence	Read Contacts	99.20%
	Greedy Selection	Read Contacts	99.20%

Evaluation (Effectiveness)

Data Selection: Comparison of With & Without Data Selection

Platform	Performance	Configuration		
		No Transfer	Without Data Selection	With Data Selection
IFTTT	F1 Score	84.25%	91.08%	91.83%
	Improvement	-	6.83%	7.58%
Chrome	F1 Score	70.60%	84.36%	89.13%
	Improvement	-	13.76%	18.53%
SmartThings	F1 Score	72.80%	84.65%	89.1%
	Improvement	-	11.85%	16.3%

Evaluation (Effectiveness)

Data Selection: Comparison of With & Without Data Selection

Platform	Performance	Configuration		
		No Transfer	Without Data Selection	With Data Selection
IFTTT	F1 Score	84.25%	91.08%	91.83%
	Improvement	-	6.83%	7.58%
Chrome	F1 Score	70.60%	84.36%	89.13%
	Improvement	-	13.76%	18.53%
SmartThings	F1 Score	72.80%	84.65%	89.1%
	Improvement	-	11.85%	16.3%

Evaluation (Effectiveness)

TKPERM Performance Analysis (Metric: F1 Score)

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

Evaluation (Effectiveness)

TKPERM Performance Analysis (Metric: F1 Score)

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

Evaluation (Effectiveness)

TKPERM Performance Analysis (Metric: F1 Score)

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

Evaluation (Effectiveness)

TKPERM Performance Analysis (Metric: F1 Score)

Platform	Target Domain	Source Domain	Transfer	No Transfer	Improvement
IFTTT	Evernote	Coarse Location + Fine Location + Camera	83.13%	79.78%	3.35%
	BMW Lab	Send SMS + Record Audio	95.24%	85.71%	9.53%
SmartThings	Lock	Write Setting	85.71%	75.00%	10.71%
	Motion Sensor	Read Contact	87.10%	53.33%	33.77%
	Switch	Send SMS + Read Calendar	94.39%	90.09%	4.30%

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?

Effectiveness

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?

Effectiveness

Scalability

Evaluation (Scalability)

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

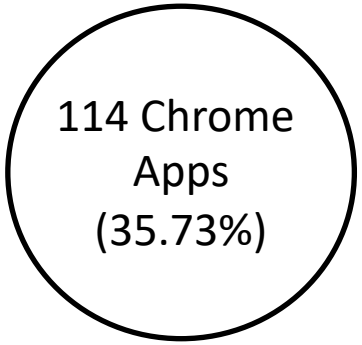
Platform	Target Domain	Time (hh:mm:ss)
IFTTT	Evernote	33:27:03
	BMW Lab	14:08:40
	Facebook	22:57:20
	Google Calendar	15:15:18
	Google Contact	18:40:17
Chrome	Geolocation	07:37:28
	Proxy	06:54:01
	Content Settings	09:42:45
SmartThings	Lock	03:47:59
	Motion Sensor	04:09:44
	Switch	14:11:08

Evaluation (Scalability)

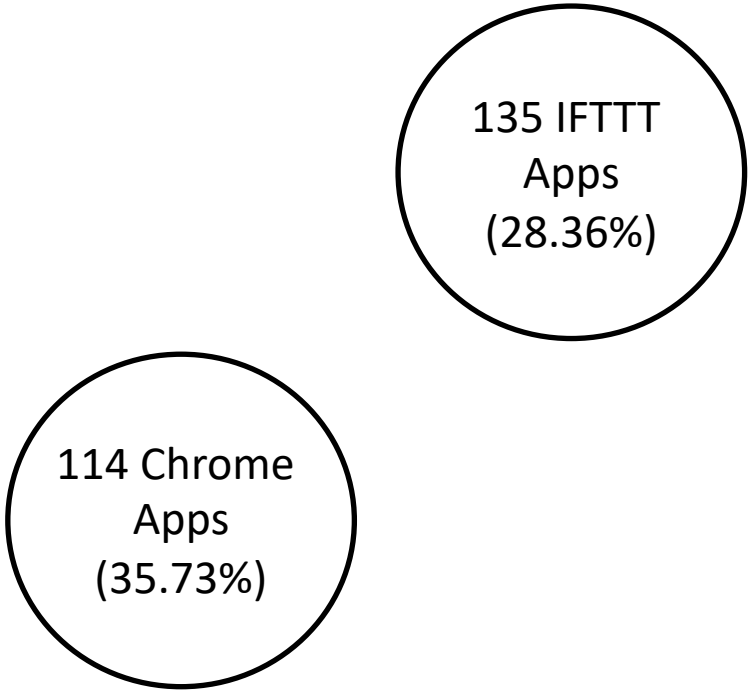
Computation Overhead (Run in Amazon Elastic Compute Cloud (EC2), NVIDIA Tesla V100)

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

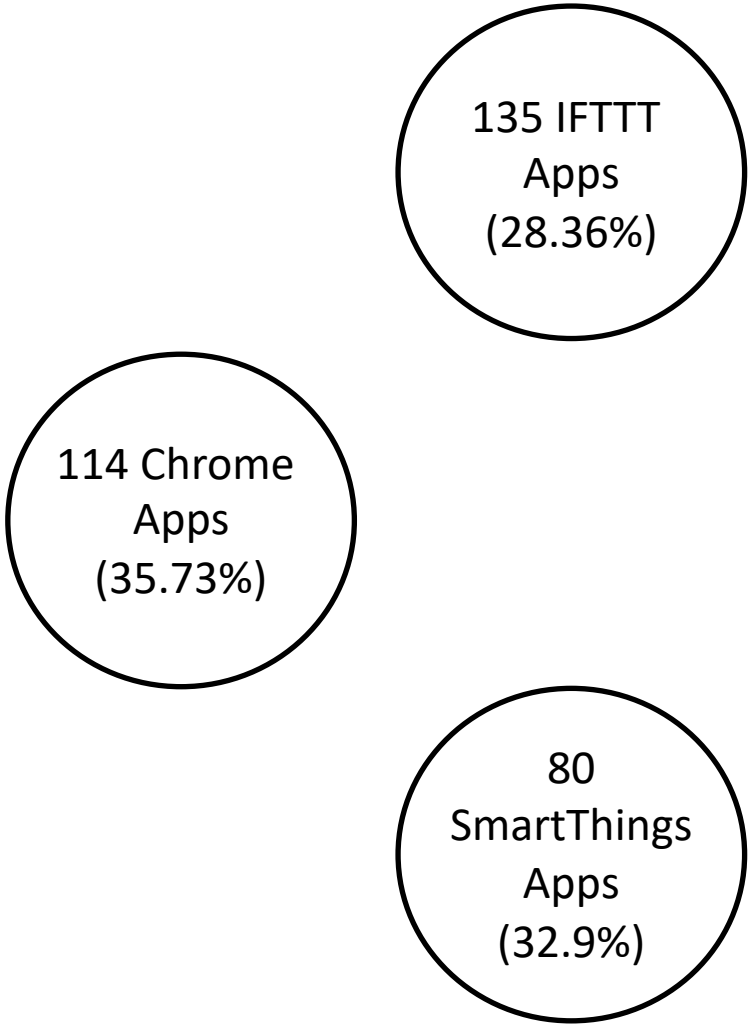
Measurement Result



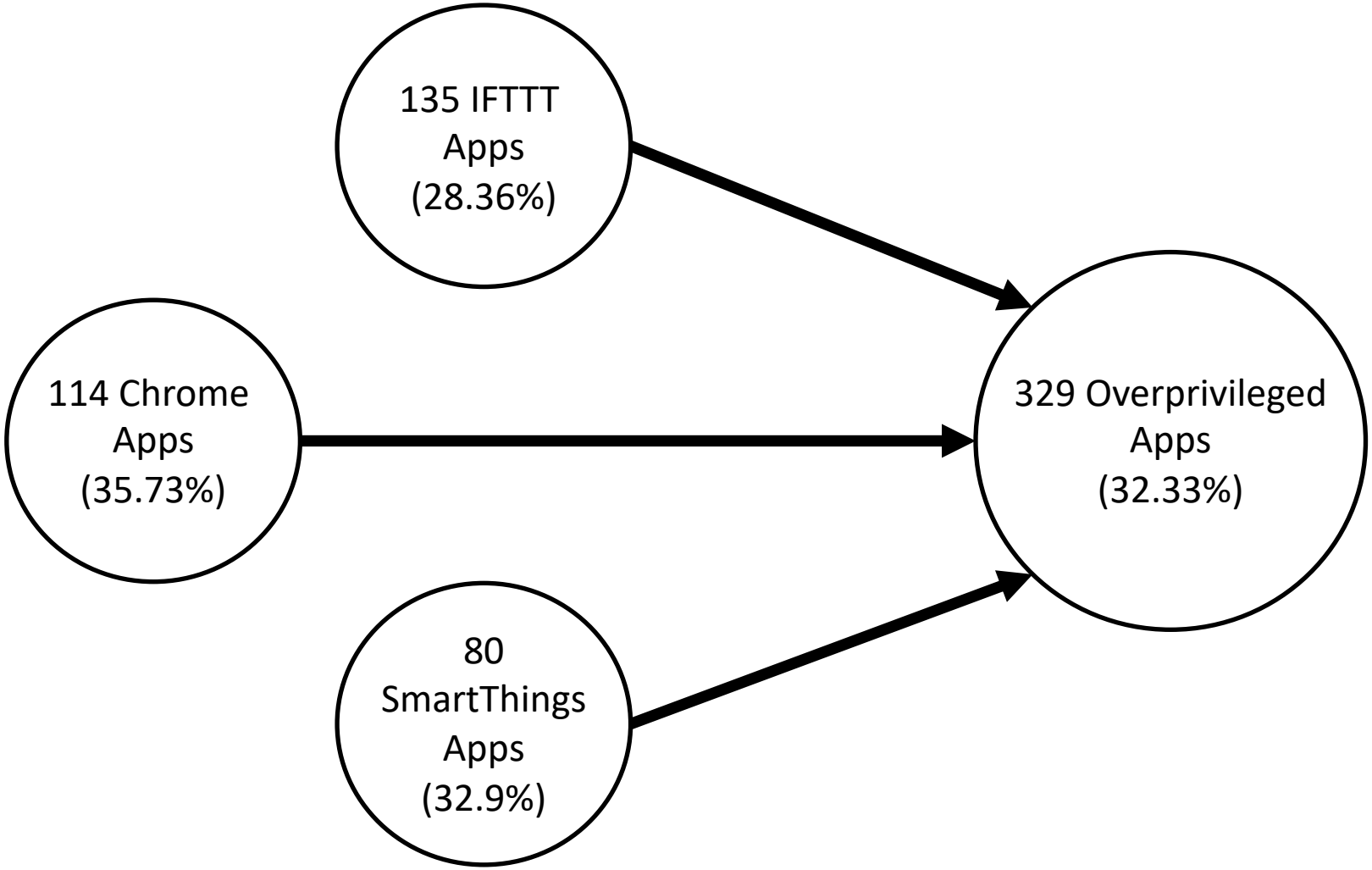
Measurement Result



Measurement Result



Measurement Result



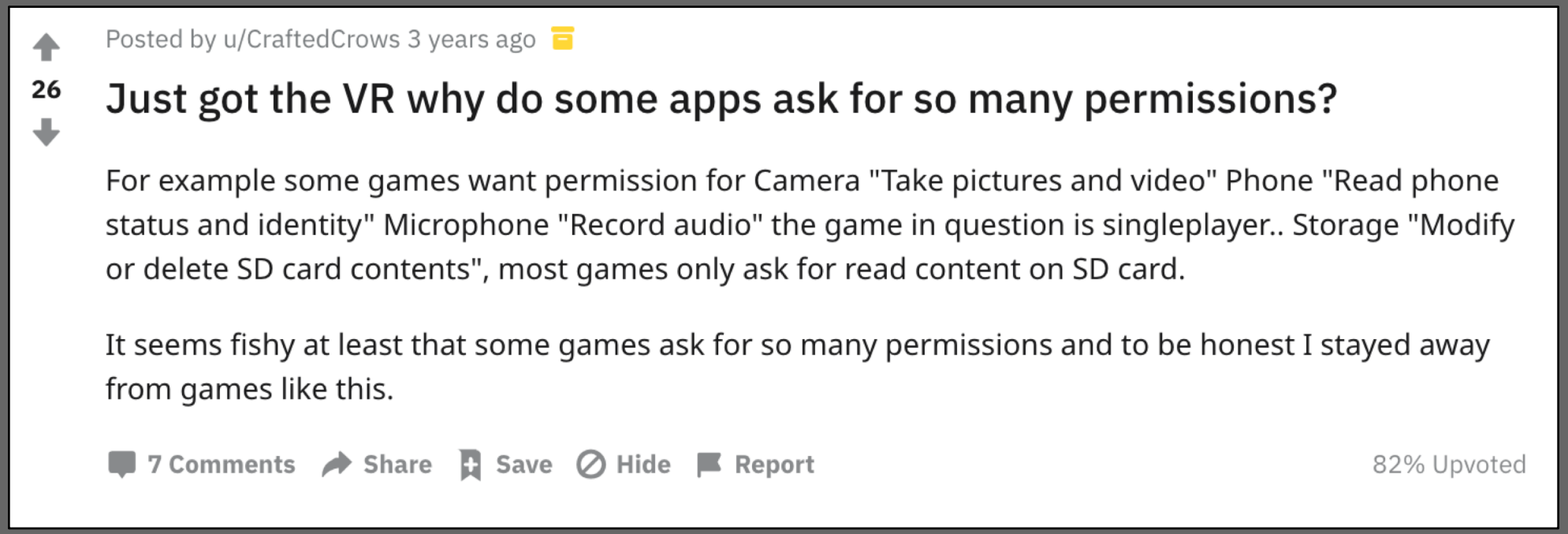
Conclusion

1. General Framework

General framework to detect
Overprivileged applications in new platforms

Conclusion

1. General Framework



Posted by u/CraftedCrows 3 years ago

26

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.

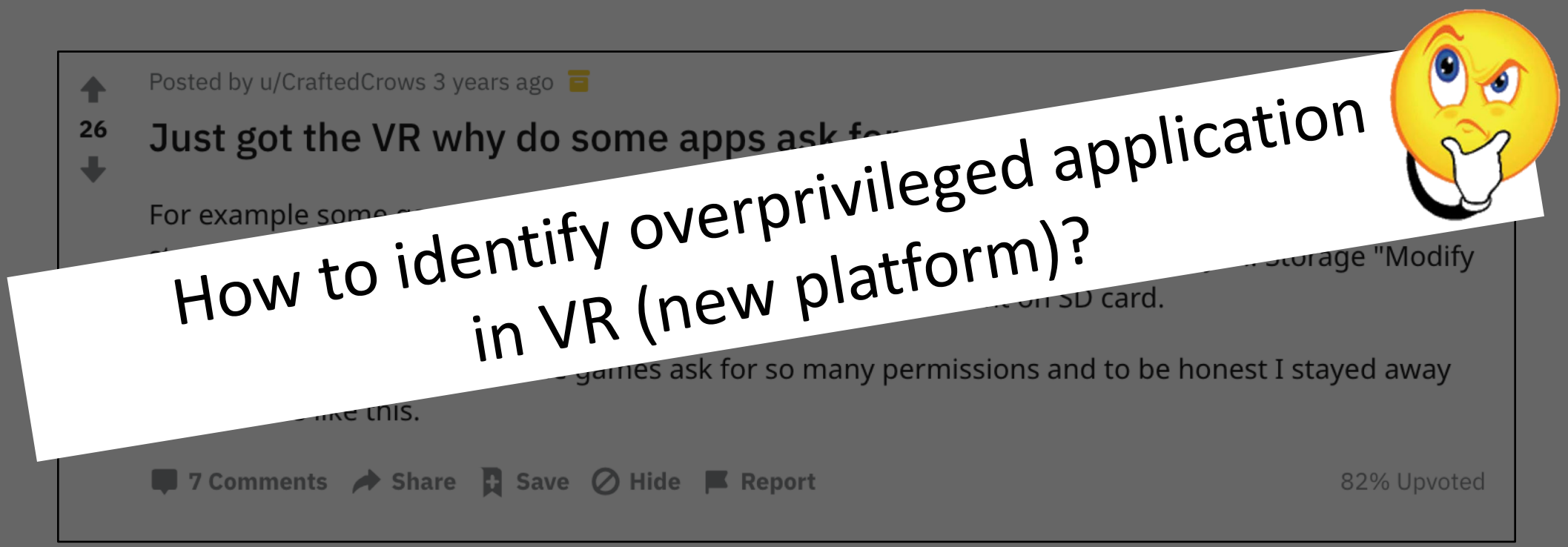
It seems fishy at least that some games ask for so many permissions and to be honest I stayed away from games like this.

7 Comments Share Save Hide Report 82% Upvoted

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



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

Posted by u/CraftedCrows 3 years ago

26

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" Microp eplayer.. Storage "Modify or delete SD card contents"

It seems fishy at least that e honest I stayed away from games like this.

7 Comments Share Save Hide Report 82% Upvoted

TKPERM!!!

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
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-ZpBpL8lftYIdM6JtH-QKE88NcSX

Thank You!

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

Conclusion

1. General Framework
2. Result
3. Public Dataset

Dataset

Email: fs5ve@virginia.edu

Dataset Available at: https://drive.google.com/drive/u/1/folders/1Yfnz-ZpBpL8lftYIdM6JtH-QKE88NcSX

Backup Slides

Plat.	Perform.	Word Embedding	
		Target	Source
IFTTT	F1 score	86%	91.83%
	Improv.	-	5.83%
Chrome	F1 score	88%	89.13%
	Improv.	-	1.13%
Smart Things	F1 score	85%	89.1%
	Improv.	-	4.1%

Permission	Performance			
	Acc.	Prec.	Rec.	F1
Fine Location	85%	73%	84%	78%
Coarse Location	84%	53%	84%	65%
Camera	88%	80%	89%	85%
Read Calendar	89%	87%	89%	88%
Read Contact	92%	92%	90%	91%
Record Audio	84%	83%	83%	83%
Write Settings	87%	69%	86%	77%
Send SMS	93%	93%	100%	97%
Write APN	92%	88%	97%	94%
Total	88.22%	79.78%	89.11%	84.20%

System Overview (Domain Selection)

Algorithm 1 Source Domain Selection using Greedy Selection Algorithm

Input: Source Domain Data List, $[\mathcal{D}_S]$; Target Domain Data, d_t

Output: Aggregated Source List, $[\mathcal{A}_S]$

procedure SELECTSOURCEDOMAINS
  $[\mathcal{A}_S] \leftarrow \emptyset$
  $P_{best} \leftarrow -\infty$
  $P_{current} \leftarrow$ initialize to $zero$
  $[\{D_S, d_{f1}\}] \leftarrow computealldsf_1([D_S], d_t)$
  while $size([\{D_S, d_{f1}\}]) > 0$ **do**
    $d_s \leftarrow highestf_1([\{D_S, d_{f1}\}])$
    remove d_s from $[\{D_S, d_{f1}\}]$
    add d_s to $[\mathcal{A}_S]$
    $P_{current} \leftarrow computedsf_1([\mathcal{A}_S], d_t)$
    if $P_{current} < P_{best}$ **then**
      remove d_s from $[\mathcal{A}_S]$
      break
    end if
    $P_{best} \leftarrow P_{current}$
  end while
  Return $[\mathcal{A}_S]$
end procedure

System Overview (Data Selection)

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

Input: Source Model, $\mathcal{M}_{\mathcal{S}}$; Unlabeled Target Domain Dataset, $[\mathcal{A}_{\sqcup}]$

Output: Fine-tune Dataset, $[\mathcal{D}_{\mathcal{F}}]$

procedure SELECTFINETUNEDATASET
  for each document, $\mathcal{A} \in [\mathcal{A}_{\sqcup}]$ **do**
    for each sentence, $d_t \in \mathcal{A}$ **do**
      $pred \leftarrow$ prediction$(d_t, \mathcal{M}_{\mathcal{S}})$
      if $pred = 1$ **then**
        $\mathcal{R}_{\mathcal{A}} \leftarrow \mathcal{R}_{\mathcal{A}} + 1$
      end if
    end for
    add $\{\mathcal{A}, \mathcal{R}_{\mathcal{A}}\}$ to $[\mathcal{D}_{\mathcal{R}}]$
  end for
  $[\mathcal{D}_{\mathcal{R}}]^* \leftarrow sorted_{desc}([\mathcal{D}_{\mathcal{R}}])$
  $[\mathcal{D}_{\mathcal{F}}] \leftarrow top_{20}([\mathcal{D}_{\mathcal{R}}]^*)$
  Return $[\mathcal{D}_{\mathcal{F}}]$
end procedure

Ranking:

$$\mathcal{R}_{\mathcal{A}} = \sum_{j=1}^{len(\mathcal{A})} 1 \mid \lfloor \mathcal{P}(y_j|x_j) \rceil = 1$$

$$\forall\, sentence, x_j \in document, \mathcal{A}$$

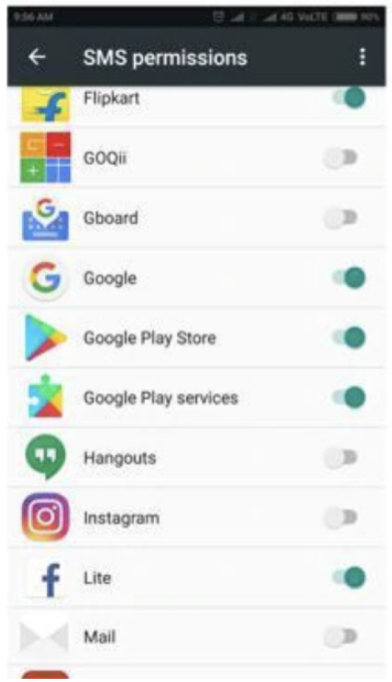
Model

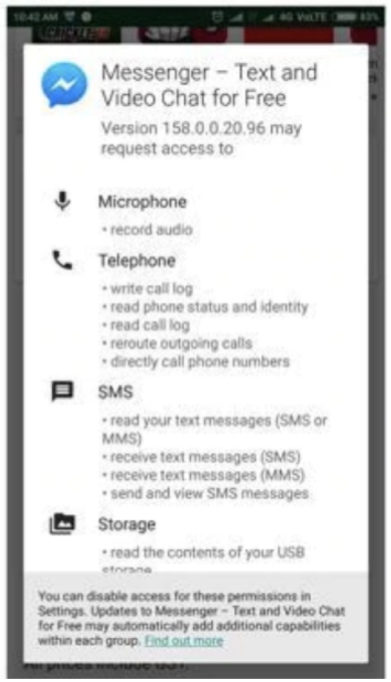
Chrome Content

Selection:

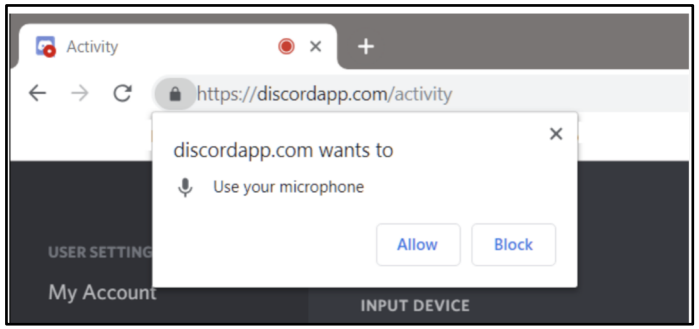
$$[\mathcal{D}_{\mathcal{F}}] = [sort_{desc}[\{\mathcal{A}_i, \mathcal{R}_{\mathcal{A}_i}\}_{i=1}^{m}]]_{j=1}^{n}$$

Permission-based Access Control

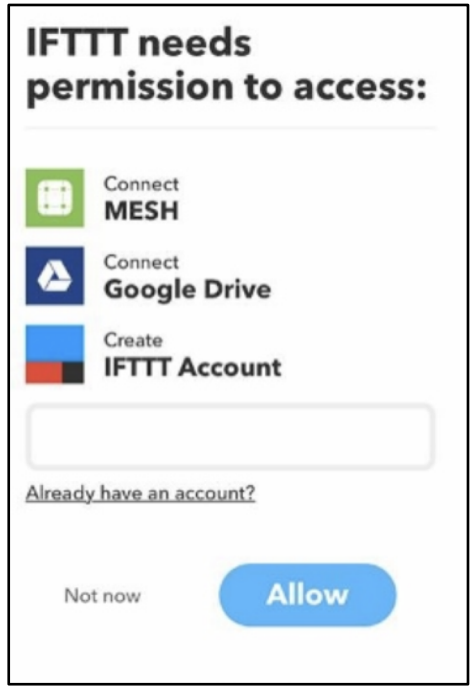




Android

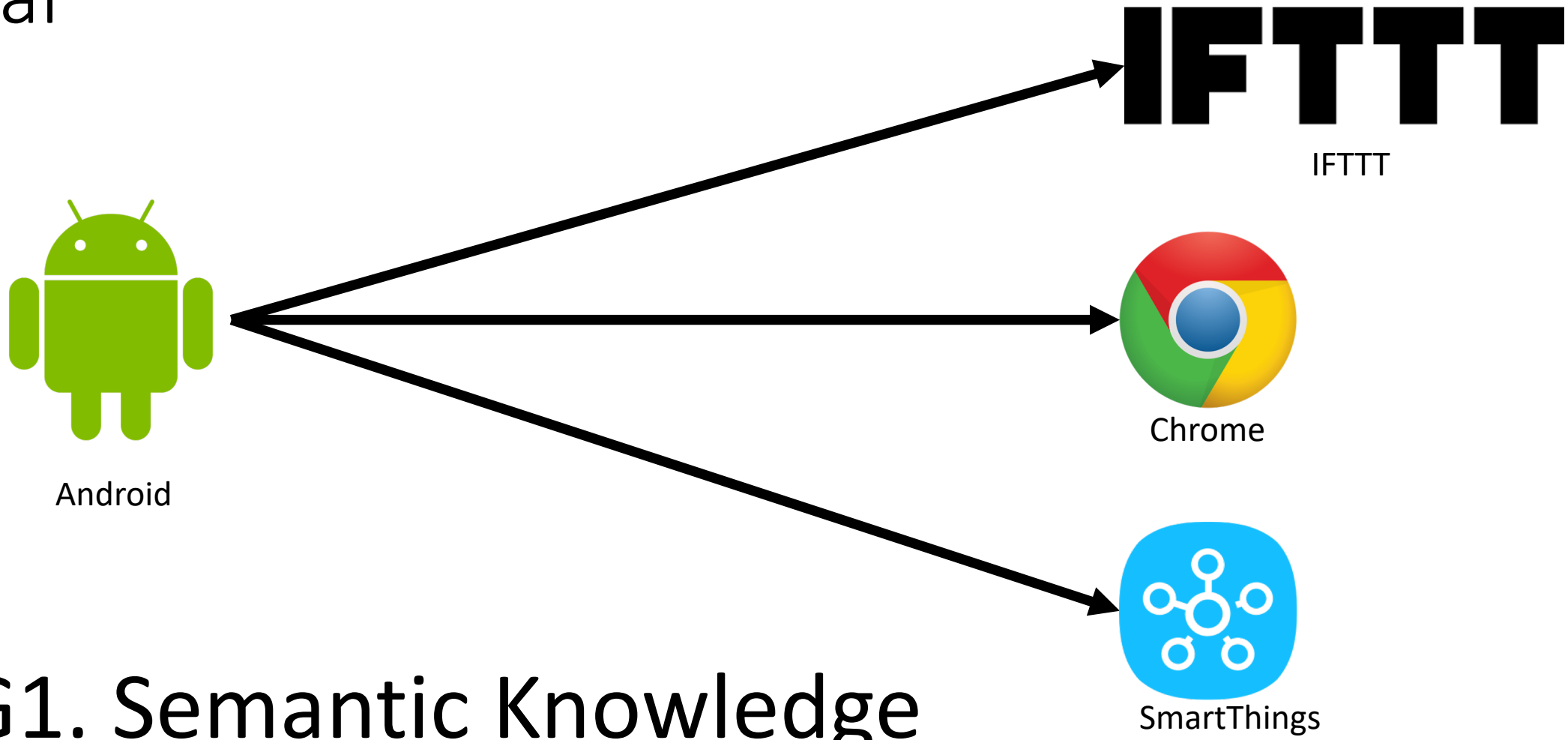


Chrome



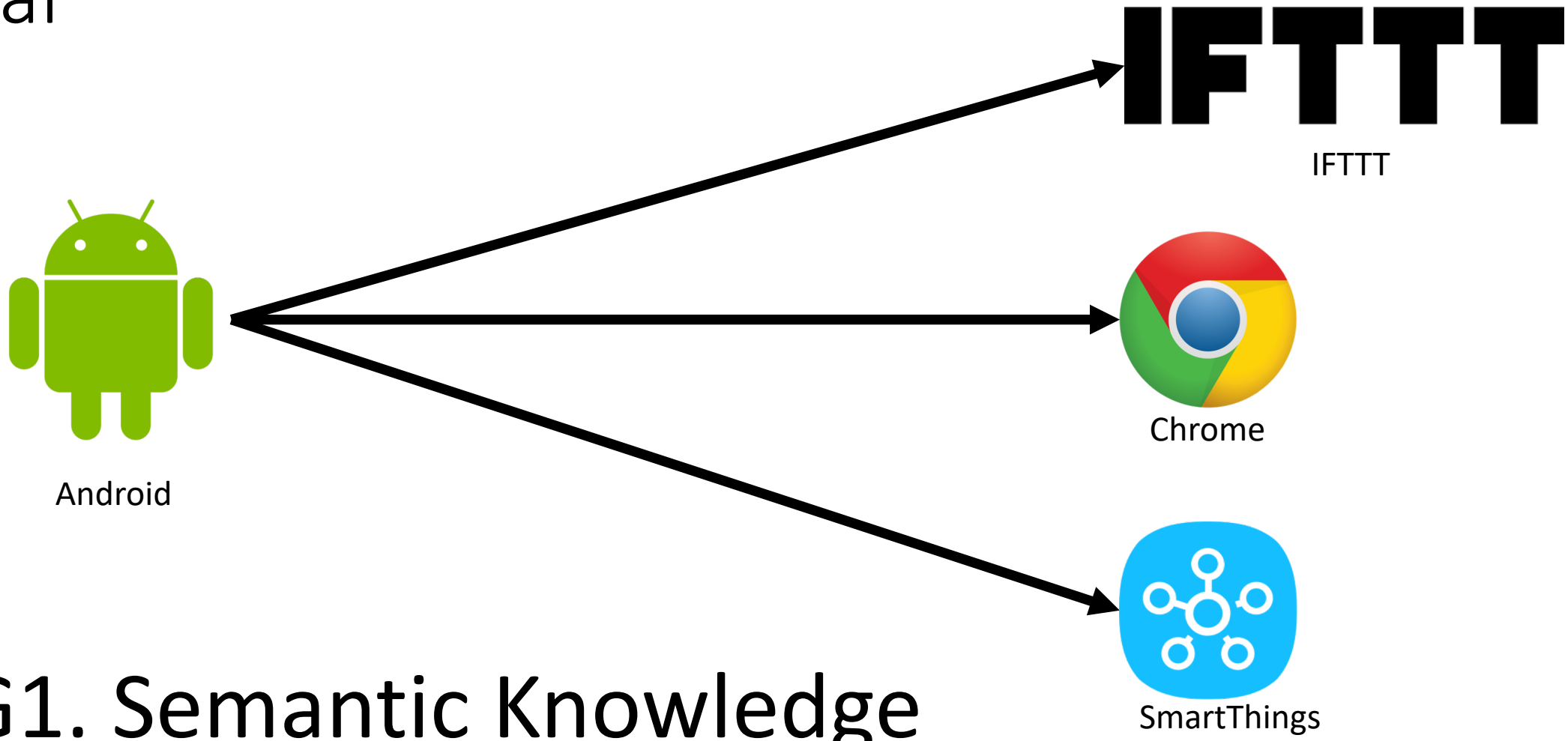
IFTTT

Goal



G1. Semantic Knowledge

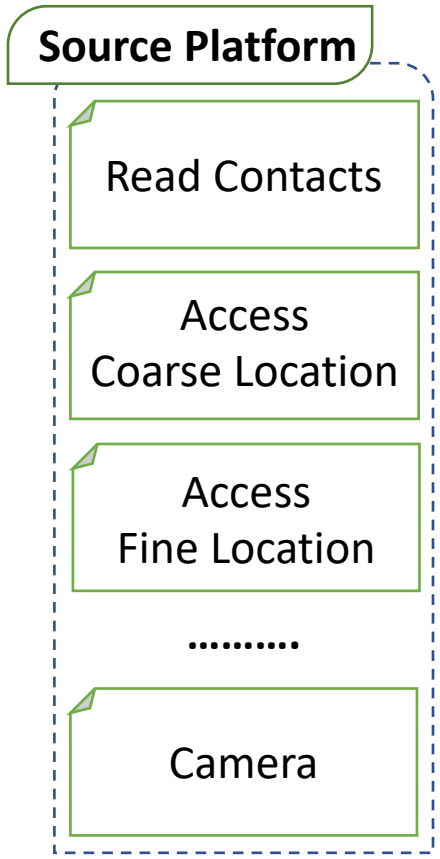
Goal



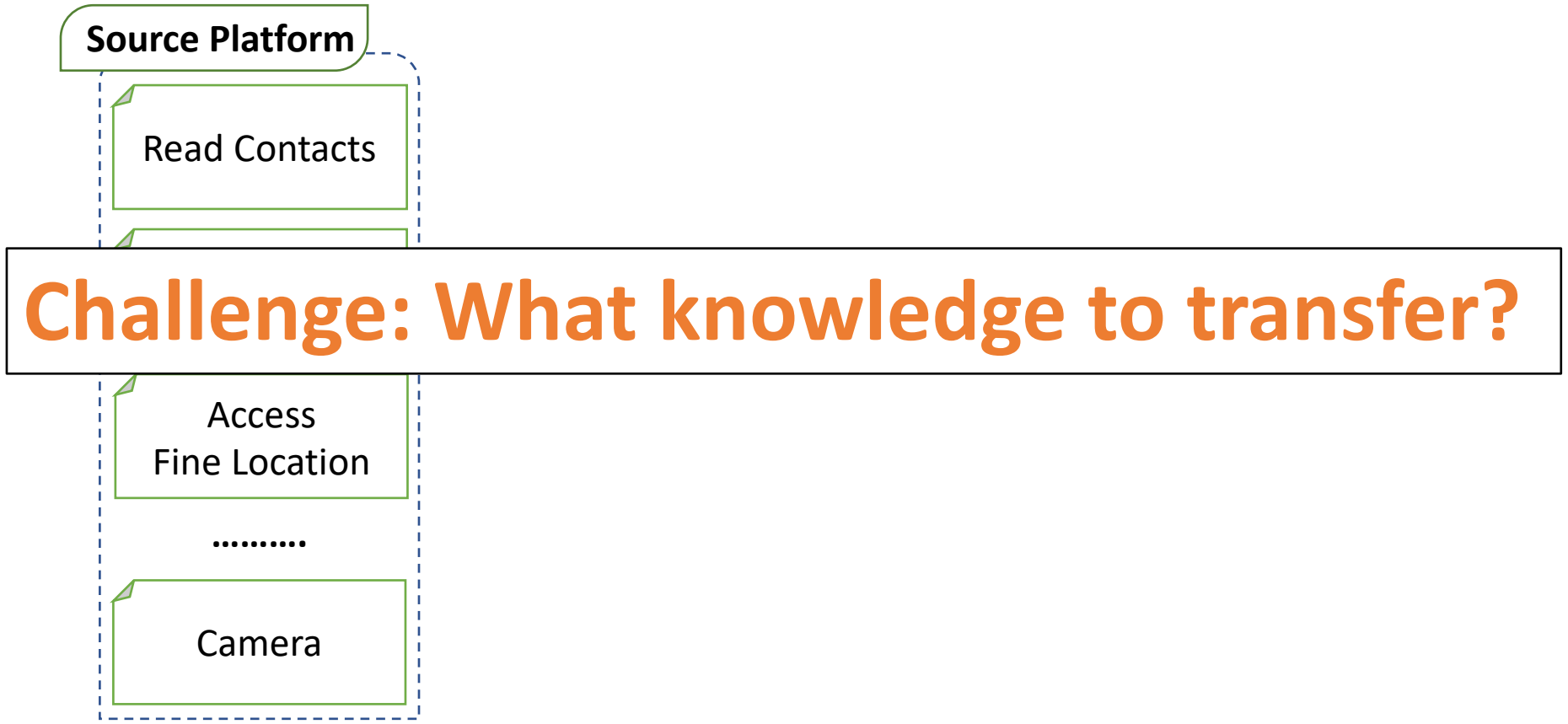
G1. Semantic Knowledge

G2. Permission Knowledge

System Overview



System Overview



Domain Selection

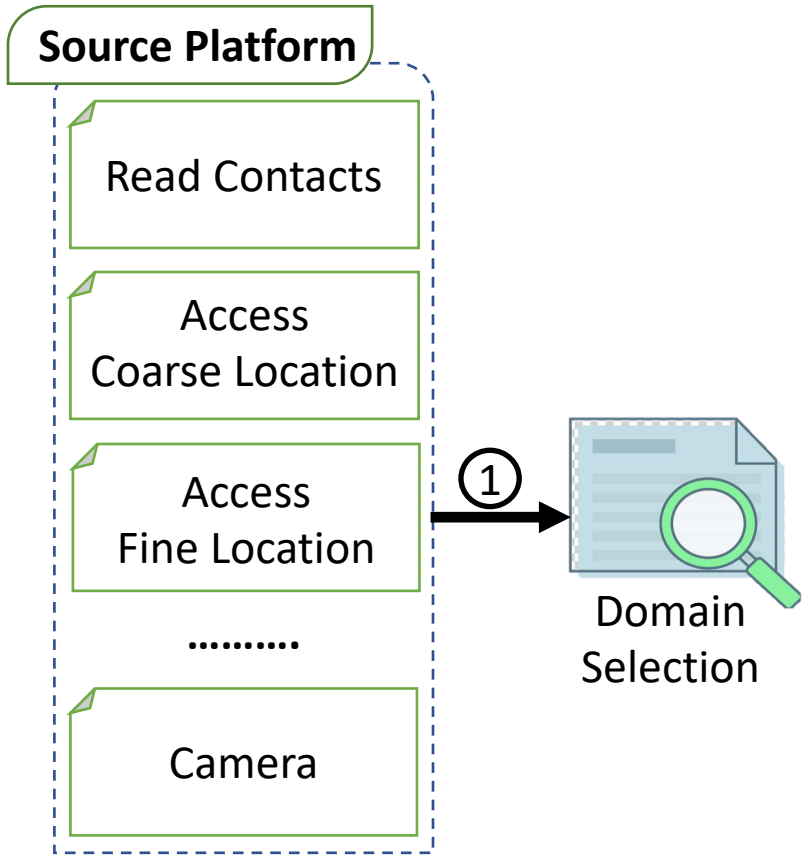
Greedy Selection Approach

Aggregate source domain(s) which performs best

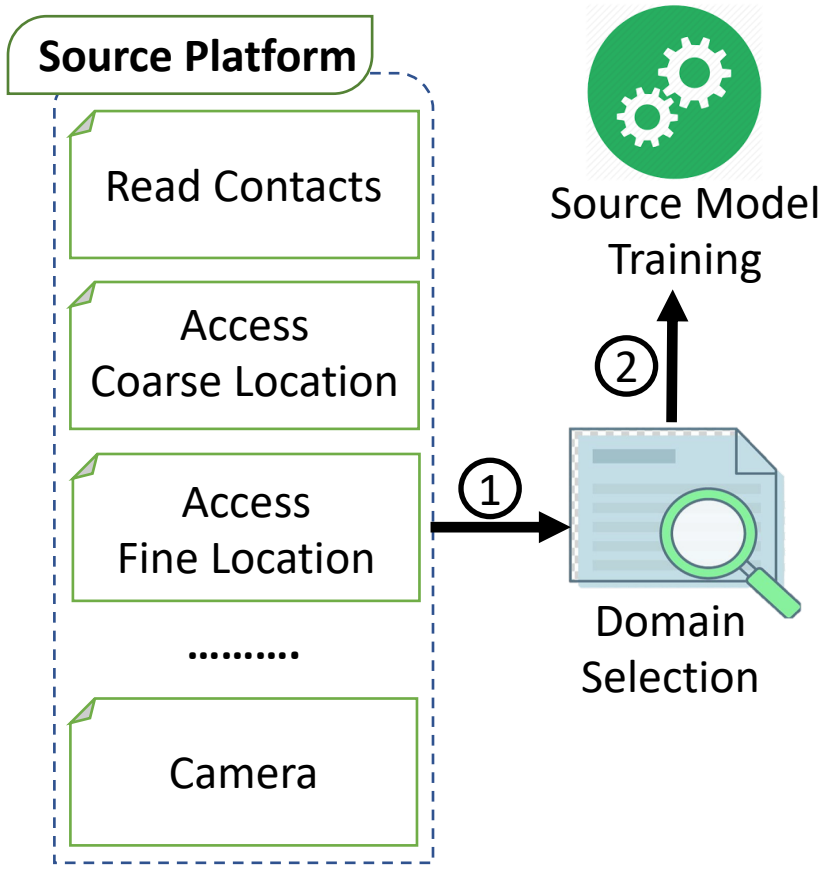
Remove source domain(s) which work worst

Find the best combination of the source domain(s)

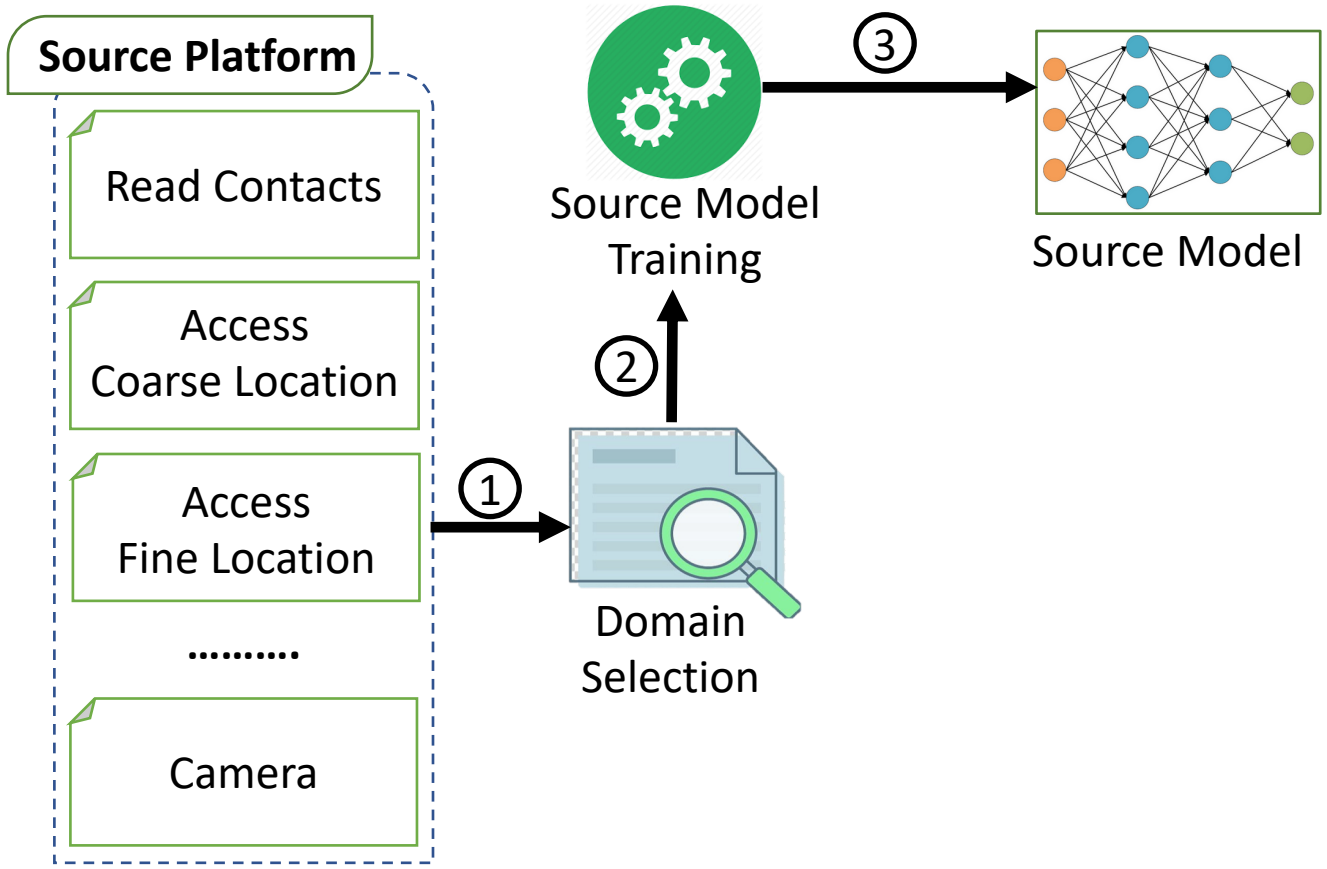
System Overview



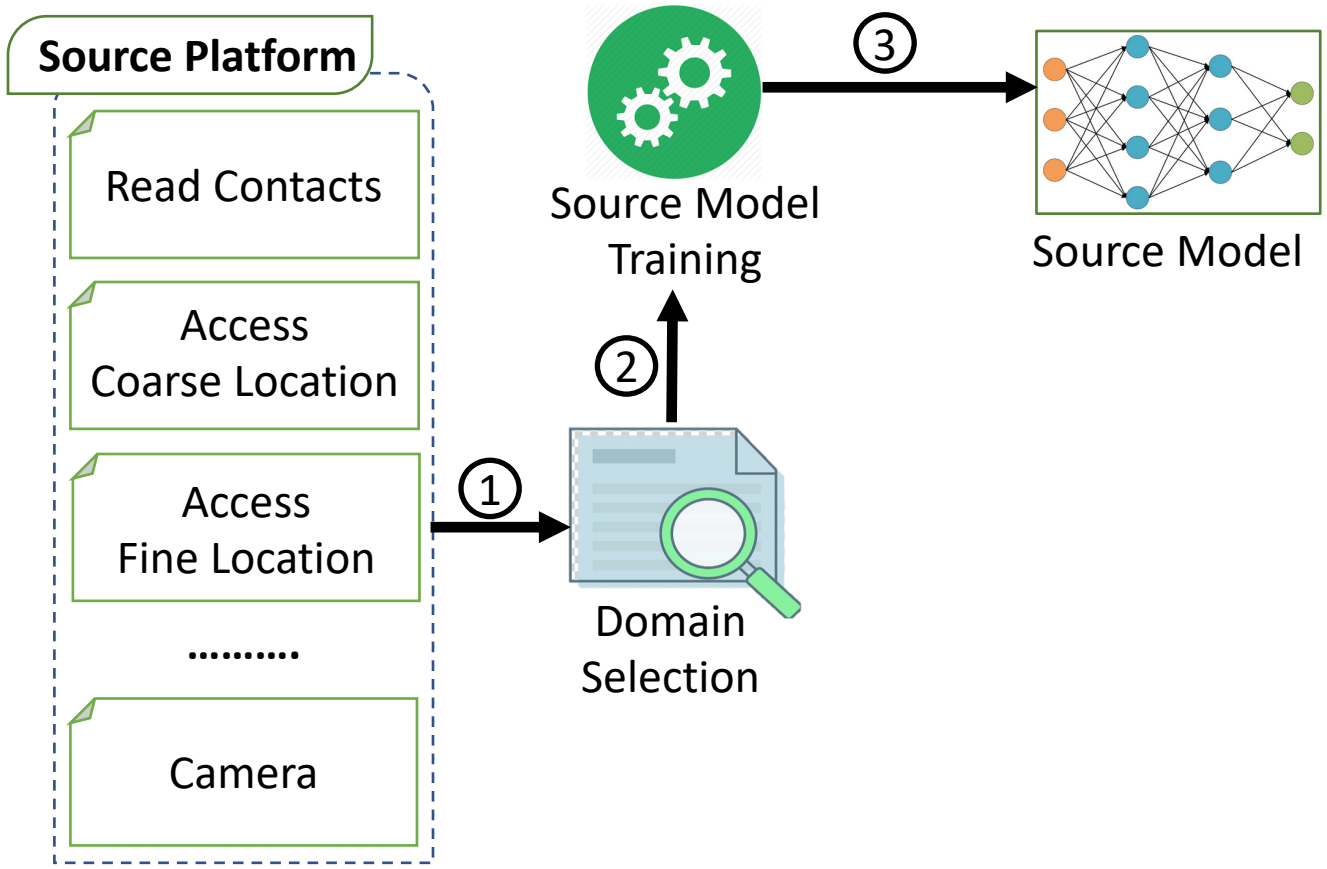
System Overview



System Overview



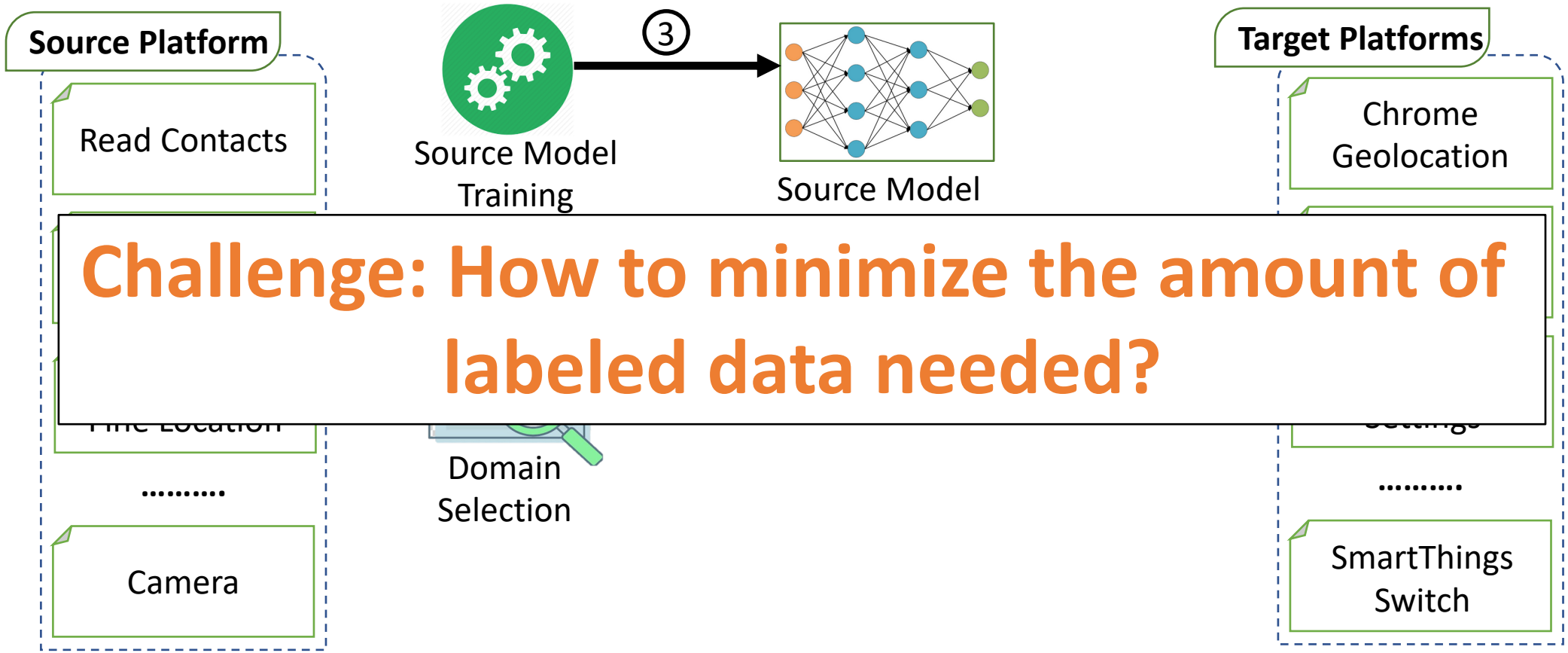
System Overview



Target Platforms

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

System Overview



Data Selection

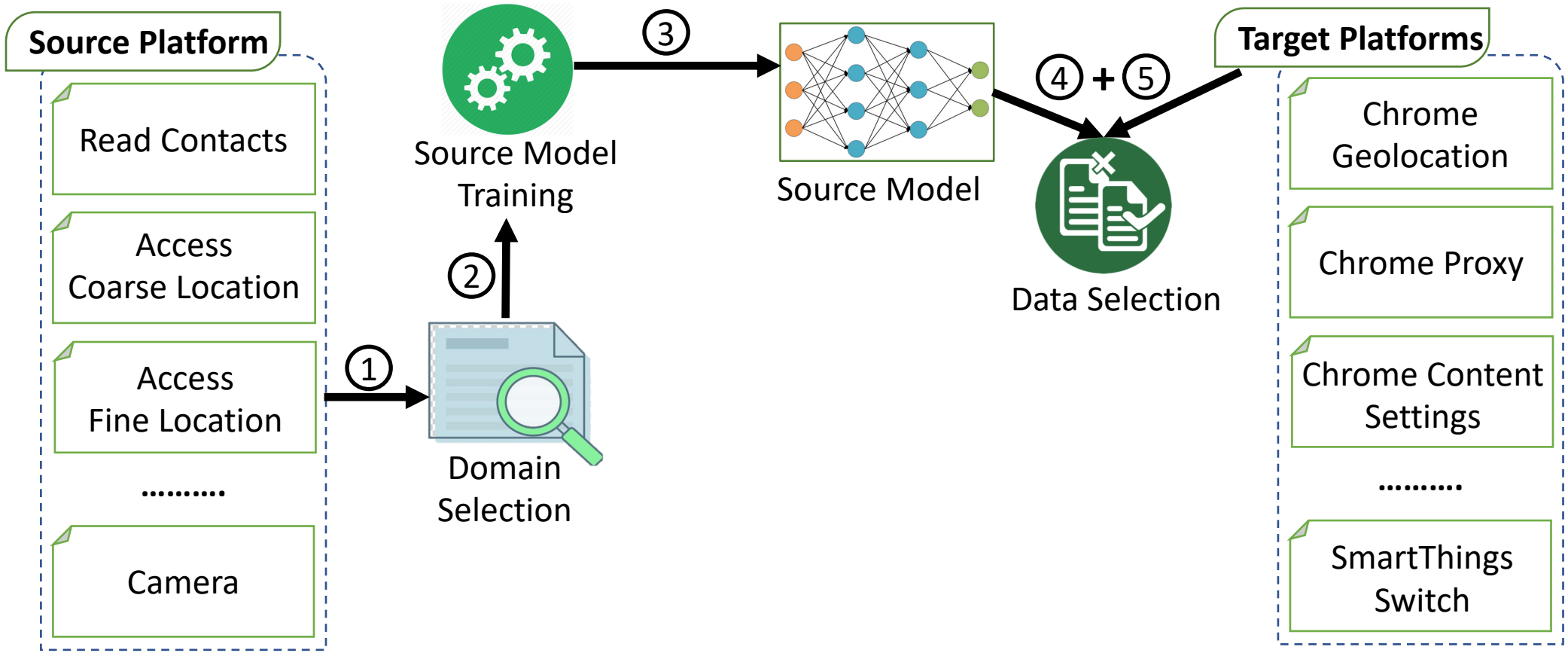
Use source model to rank the document

Rank unlabeled documents from the target domain

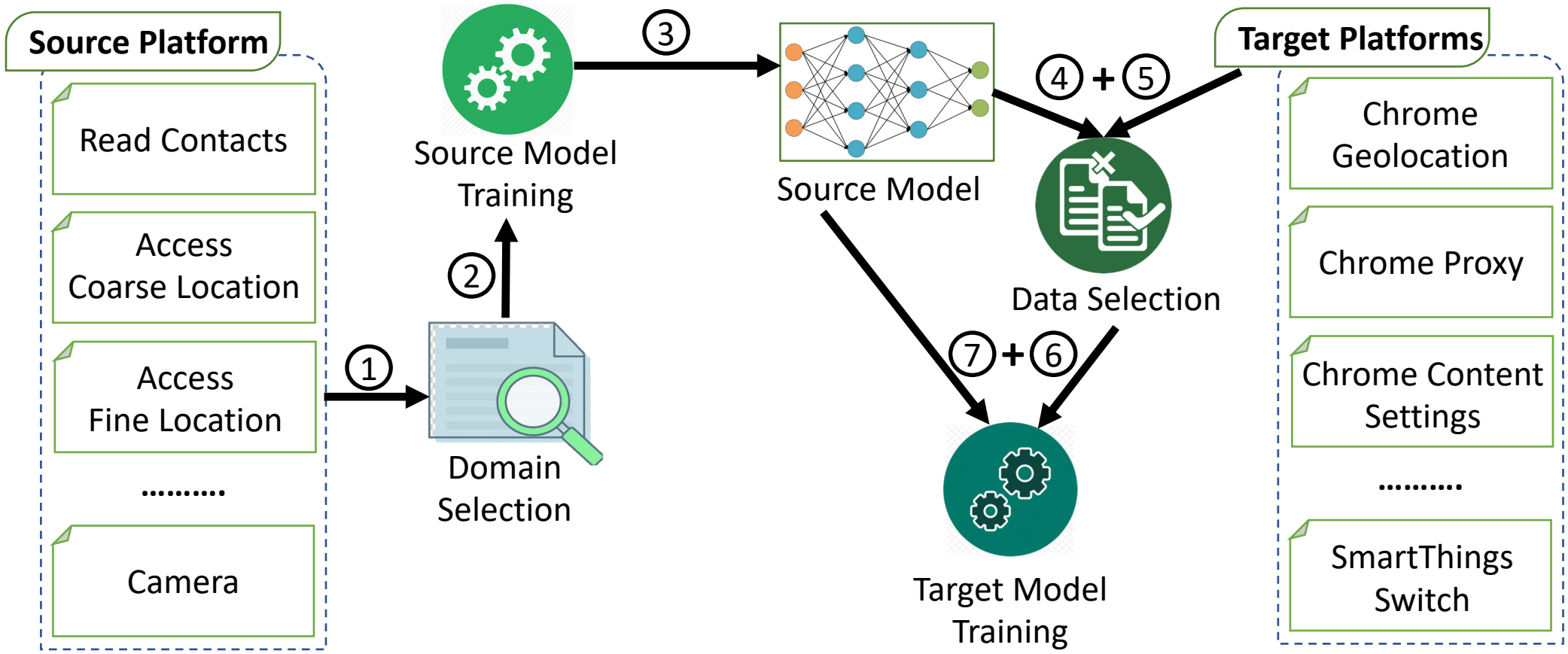
Pick the top 20 documents from a target domain

Ask human annotator to label data

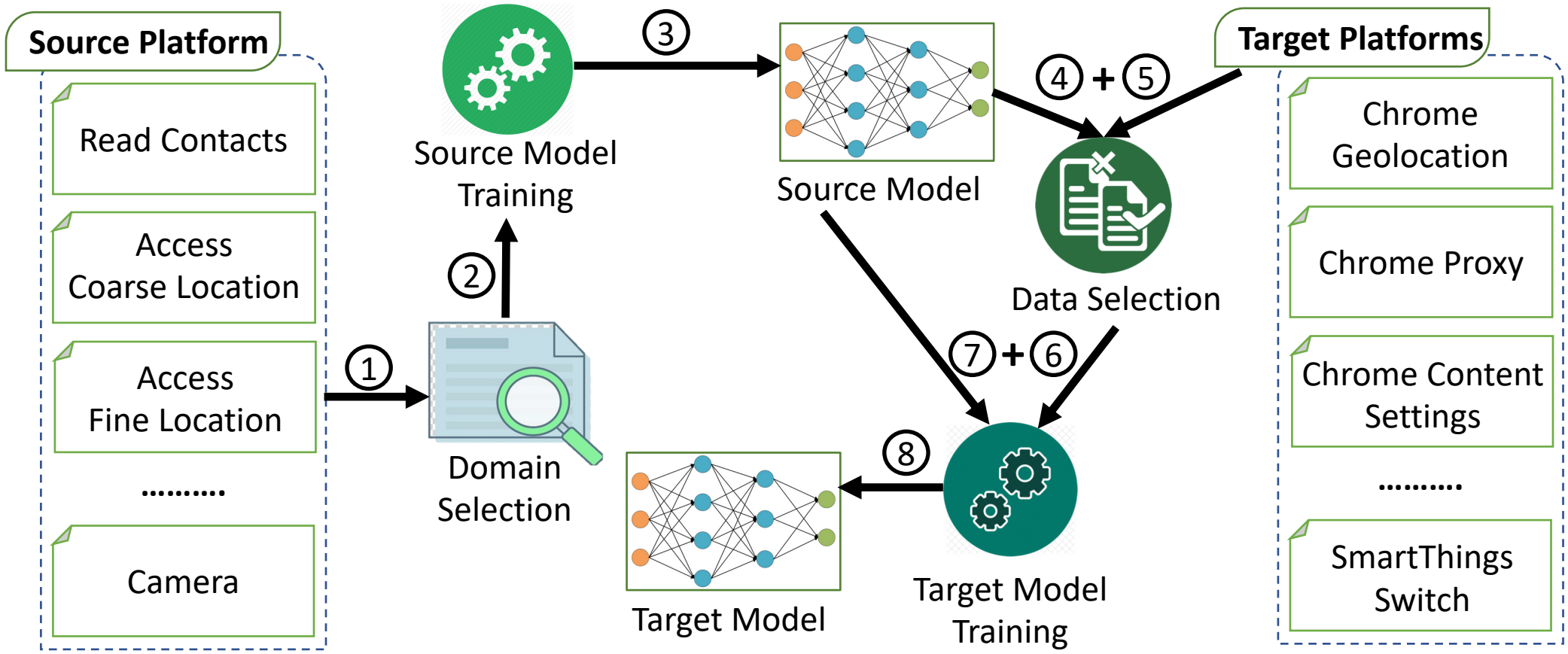
System Overview



System Overview



System Overview



Conclusion

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