object p because private data is read from sources. Sinks are associated with the constraint c because private data is sent out by sinks. In the source mapping, we map p by its category into a predefined set of source APIs. In the sink mapping, we parse the sink list from Susi [35] to generate a lookup table $H(K) = T$, where K is a keyword set and T is the sink set. The lookup table maps a constraint c to a set of sinks $\bigcup_i t_i$, where each t_i is a sink function. We have $k \in c$ and $H(k) = \bigcup_i t_i$, where $k \in K$ and $t_i \in T$. Our approach returns all possible sinks if it cannot detect any matched constraint c.

Taint Flow Analysis. We generate rewriting specifications based on the taint flow analysis. The taint flow analysis is used to identify private data leak vulnerabilities. Reachability between a source and a sink is evaluated in the G_{icfg} and the output is a taint flow graph $G(V, E, S, T)$. A taint flow path $f = \{v_0 \rightsquigarrow v_n\}$ represents a potential data leak from v_0 to v_n. We generate rewriting specifications based on a taint flow path f. We identify the unit u in $R_r = \{u, op\}$ as $u = v_n \in f$. The unit u is uniquely identified by matching three types of signatures: the class signature, the method signature and the statement signature. The class signature represents the class of a sink. The method signature represents the method in the class. The statement signature represents the exact line of code in a method.

5. AUTOMATIC REWRITING

The purpose of rewriting is to automatically modify the code based on a rewriting specification R_r. We develop a new rewriting framework to recognize rewriting specifications. Although there exist some rewriting solutions on Smali bytecode (e.g., I-ARM-Droid [13] and RetroSkeleton [12], we choose to reimplement our rewriting framework on Jimple for compatibility and reliability. Jimple is a three-address typed intermediate representation (IR) transformed from Smali by the Soot [23], which allows creation and reorganization of registers. Rewriting on Jimple is less error-prone than the direct modification on Smali. The modified IR is re-compressed as a new APK file and signed with a new certificate. The new APK remains the original logic and functions. Only sensitive sinks are rewritten for privacy protection enforcement.

In this paper, we only demonstrate the use case of rewriting for privacy protection. The rewriting framework can be used for general purposes, e.g., repackaging and vulnerability mitigation. We propose three basic operations in our rewriting framework.

Unit Insertion. We would insert new local variables, new assignments, Android framework APIs and calls to user defined functions into a method. We first need to inspect the method's control flow graph (CFG) and existing variables. We generate new variables for storing parameter contents and new units (e.g., add expression, assignment expression) for invoking additional functions. These new variables and units are inserted before u in R_r.

For Android framework APIs, we generate a new *SootMethod* with the API signature, and insert it as a callee in the app. We generate an invoking unit as a caller to call the callee in the method. To guarantee the consistence of parameters of the API, we inspect the types of local variables and put them into the caller function in order.

For user defined functions in Java code, we generate a new *SootClass* with the class and method signature. We insert the *SootClass* as a new class. We invoke a user defined function in the method by generating a new invocation unit. The Java code is compiled into Jimple IR and attached to the app as a third-party library.

Unit Removal. Unit removal is more straightforward, as we directly remove the unit u in the original function. However, we need to guarantee that the removal unit has no-violation of the control-/data-flow integrity on the rest code. We analyze the CFG of the function to guarantee the validation of removing a unit.

Unit Edition. Unit edition modifies the unit u by changing its parameters, callee names or return values. We decompose u into registers, expressions and parameters, and modify them separately. For example, for a unit as a invocation $u = r_0.sendHttpRerquest(r_1...r_n)$, we decompose it as a register r_0, a function call *sendHttpRerquest* and the parameters $\{r_1, ...r_n\}$. *sendHttpRerquest* is replaced with *myOwnHttp* by implementing a customized checking function. We are able to replace parameter r_0 with a new object r_{new} for monitoring network traffic.

We utilize basic operations to generate complex rewriting strategies. An if-else-condition can be decomposed as a unit insertion (initialize a new variable as the condition value), a unit edition (get condition value) and two unit insertions (if unit and else unit). Users can customize rewriting strategies in rewriting. In our approach, $op \in R_r$ is dependent on the sentimental value $s \in T_s$. If $s = -1$, we define our rewriting strategy is to insert a check function *check*() before u. The check function terminates the invocation of u if privacy violation happens to u at runtime. This rewriting strategy enforces privacy protection with dynamic interruption.

The complexity of rewriting an app is $O(MN)$, where M is the number of units for rewriting and N is the lines of code. Given a rewriting specification $R_r = \{u, op\}$, the searching algorithm is linear by matching the unit in an app. The time for operation op is constant.

6. LIMITATION AND DISCUSSION

In our design, we mainly focus on the technical challenges for developing a natural language interface to programs for rewriting. User interfaces and legal issues (e.g., copyright restrictions) are out of the scope of discussion.

NLP Limitation. Our study shows higher precision in understanding user intentions than the state-of-the-art approach for app rewriting. In our prototype, we expect users express the sentence in a structured format that can be parsed by our prototype. However, our approach still introduces some false positives and incorrect inferences. The inaccuracy most comes from abnormal ways of user presentations, e.g., ambiguous grammars and complex expressions. Also, the standard dependence parsing cannot correctly infer the semantic relations for complex sentences sometimes. The inaccuracy of dependence parsing further introduce wrong identifications of security objects and the sentimental value. In the future work, more efforts are required to better resolve above limitations.

Taint Analysis Limitation. Our prototype is built on the static program analysis framework [6]. The inaccuracy of our prototype comes from the inherent limitations of the static analysis. Static analysis cannot handle dynamic code obfuscation. Static analysis often over-approximates taint flow paths. The detected taint flow paths may not be feasible at runtime. We mitigate limitations of static analysis

with an efficient rewriting mechanism. Our rewriting framework can terminate a function's invocation only when privacy violation happens. We plan to extend our prototype with a more capable hybrid analysis. The hybrid analysis enhances the resistance of dynamic obfuscations.

Rewriting Limitation. Our rewriting framework provides multiple operations for app customization. The research prototype is based on the Soot infrastructure with the Jimple intermediate representation. The Jimple IR is extracted from Android Dex bytecode by app decomposition. We recompile Jimple IR into Dex bytecode after rewriting. ART and Dalvik virtual machines [8] are compatible to run Dex bytecode. The efficiency of rewriting relies on app decomposition. Current decomposition frameworks [30, 16] cannot handle native code and dynamic loaded code. Similar to static analysis, we cannot rewrite bytecode that is not compatible to Jimple. How to detect dynamically loaded code is studied recently in [33]. In the future work, we plan to enhance rewiring by increasing decomposition capacity.

Deployment Limitation. Our research goal is building a generic tool to support flexible natural language inputs. Everyone could use the tool as an interface to specify security policies and customize apps. However, the usability of our prototype is limited to current techniques, e.g., the performance bottleneck of taint flow analysis. The mitigation of limitations can be resolved by choosing a low-precision taint flow analysis configuration, e.g., context/flow-insensitive module. The prototype only demonstrates a possible language interface to programs for app rewriting. More efforts are needed to improve the precision in both language understanding and program analysis.

7. EXPERIMENTAL EVALUATION

We present the evaluation of our approach in the section. We list four research questions in our evaluation. **RQ1:** What is the accuracy of *IronDroid* in identifying user intentions? (precision, recall and accuracy) **RQ2:** How flexible is *IronDroid* in rewriting apps? (success rate and confirmation of modified behaviors) **RQ3:** What is the performance overhead of *IronDroid* ? (size overhead and time overhead) **RQ4:** How to apply *IronDroid* for practical security applications? (security applications)

To answer the research questions, we evaluate the accuracy of user intention extraction in Section 7.2. We evaluate the feasibility of rewriting in Section 7.3. We measure the performance of our approach in Section 7.4 We provide three case studies in Section 7.5.

7.1 Experiment Setup

We implement our research prototype by extending StanfordNLP library for user intention analysis. We utilize PCFG parsing [22] for structure parsing and Stanford-Typed dependencies [11] to detect semantic relations. We use WordNet [32] similarity to compute the relatedness between two words. Our taint flow analysis is based on the cutting-edge program analysis tool FlowDroid [6]. We extend FlowDroid to support natural language transformation and rewriting specification generation. We implement our own Android rewriting framework with Jimple IR based on Soot.

To evaluate the accuracy of user intention analysis, we evaluate totally 153 sentences collected from four users with different security backgrounds. Users express their concerns on the personal information. The average length of a sentence is 10. 30.1% of sentences contain subordinate clauses. We have two volunteers to independently annotate each sentence in the corpus. Each sentence is discussed to a consensus. We use the corpus as ground truth for user intention analysis. For the taint flow analysis and app rewriting, we evaluate benchmark apps (total 118 apps) from DroidBench. We dynamically trigger sensitive paths in rewritten apps for validation. We compute time and size overhead for the performance evaluation. In security applications, we show how to potentially apply our approach to privacy protection. We also demonstrate how to achieve a more fine-grained location restriction. The location restriction is beyond the current Android dynamic permission mechanism.

7.2 RQ1: User Intention Accuracy

The output predicates $T_s = \{p, s, c\}$ is a set of tuples. We mostly focus on the accuracy of the user object p and the user sentimental value s. These two concepts are most important to express user concerns and attitudes. The constraint c has a huge variance and is hard to perform standard statistics evaluation. We compare precision with following models:

- **StandfordNLP** includes a basic sentiment model for sentence-based sentimental analysis. In our problem, StanfordNLP does not provide the security object identification.
- **Keyword-based Searching** identifies the security object and the sentimental value by keyword matching in the sentence. Keyword-based searching ignores the structure and semantic dependencies. Keyword-based searching is regarded as the state-of-the-art analysis in user intention identification.
- ***IronDroid*** captures the insight of sentence structure and semantic dependence relations. *IronDroid* identifies the object and the sentimental value via graph mining from two stages of parsing trees. *IronDroid* also utilizes the semantic similarity for word comparison.

For a scientific comparison, the keyword-based searching and *IronDroid* share the same predefined object set. We identify categories of objects: {*phone, calendar, audio, location*}. The four categories present a high coverage of private data for mobile app users. We measure the precision, recall and accuracy of user sentimental value identification. TP(true positive), FP(false positive), TN(true negative), and FN(false negative) are defined as:

1. TP: the approach correctly identifies the user sentimental value as negative ("not sharing").
2. FP: the approach incorrectly identifies the user intention value as negative ("not sharing").
3. TN: the approach correctly identifies the user intention value as positive ("sharing").
4. FN: the approach incorrectly identifies the user intention value as positive ("sharing").

[8] https://goo.gl/EvdMdO

We further define precision (P), recall (R) and accuracy (Acc) as:

$$P = \frac{TP}{TP+FP}, R = \frac{TP}{TP+FN}, Acc = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

IronDroid achieves the highest accuracy than the other two approaches. Table 3 presents the sentimental analysis of three different approaches, our approach achieves 95.4% accuracy in identifying user sentimental value. StandfordNLP achieves very low accuracy 68.6% in the experiment. We believe the standard NLP model is not suitable for our problem. In our problem, we compute the sentimental value towards the security object. StandfordNLP computes the sentimental value for the whole sentence. There is no direct correlation between two different sentimental levels. For example, standfordNLP identifies "I do care, please block the location information" and "it is okay to share my location information" with the same optimistic/positive attitude. However, the first sentence presents a negative intention (blocking) for the object "location". Therefore, we cannot directly apply standfordNLP for our problem. Keyword-based searching fails because of the lack of insight in sentence structure and semantic dependence. For example, "do not share my location if I am not at home" is misclassified by the keyword-based searching. The main reason is that "not" in the subordinate clause has not direct semantic dependence on the main clause. Our approach has more insights in capturing sentence structures and semantic dependence relations. In summary, our approach is very accurate in understanding user intentions.

Table 4 presents the improvement of *IronDroid* comparing with the keyword-based searching for the object identification. For each category, we define the improvement in precision, recall and accuracy as:

$$\begin{aligned} \triangle P &= P_{Ucer} - P_{keyword} \\ \triangle R &= R_{Ucer} - R_{keyword} \\ \triangle Acc &= Acc_{Ucer} - Acc_{keyword} \end{aligned} \quad (2)$$

How to compute the precision, recall and accuracy for each object can be easily referred as the user sentimental evaluation. Our results show that, comparing with keyword-based searching, *IronDroid* effectively identifies objects with the average increase in precision, recall and accuracy of 19.66%, 28.23% and 8.66%. The improvement comes from two major reasons: word semantic similarity to increase true positives and dependence analysis to reduce false positives. Keyword-based searching introduces large false positives by over-approximating security objects. For example, in the sentence "I do not want share my calendar information if it is sent to other devices", keyword-based searching identifies both "calendar" and "device" as security objects because of the lack of structure analysis.

Table 5 presents the overall accuracy for user intention identification. The evaluation is based on the combination of the sentimental intention s and the object p. We mark the result $\{p, s\}$ true only when both s and p is identified correctly. *IronDroid* achieves 90.2% accuracy rate, while keyword-based approach only achieves 53.6% accuracy rate. Our approach has nearly 2-fold improvement than the state-of-the-art approach. The experimental results validate that our approach is effective in identifying user intentions from natural languages.

	TP	FP	TN	FN	P(%)	R(%)	Acc(%)
StandfordNLP	55	18	50	30	75.3	64.7	68.6
Keyword-based	64	9	63	17	87.7	79.0	83.0
Ours	68	5	78	2	**93.2**	**97.1**	**95.4**

Table 3: Sentimental analysis accuracy for *IronDroid* , StanfordNLP and keyword-based searching. *IronDroid* achieves higher accuracy, precision and recall than the other two approaches. our approach effectively achieves the improvement in 19.66% for precision, 28.23% for recall and 8.66% for accuracy on average.

Object	△P %	△R%	△Acc%
Location	19.23	14.52	11.11
Phone	20.00	38.15	9.80
Calendar	26.09	28.57	7.84
Audio	13.33	31.66	5.88
Average	**19.66**	**28.23**	**8.66**

Table 4: object detection improvement comparing with keyword-based approach. Our approaches have the average increase in precision, recall and accuracy of 19.66%, 28.23% and 8.66% for four categories of objects.

	Keyword-based	*IronDroid*
Accuracy	53.6%	**90.2%**

Table 5: The overall accuracy for understanding sentimental and the object. Our approach has a significant improvement than the keyword based searching approach.

7.3 RQ2: Rewriting Robustness

DroidBench is a standard test suite for evaluating the effectiveness of Android program analysis [20, 6]. DroidBench is specifically designed for Android apps. We tested our rewriting capability based on the benchmark apps. We use the example "never share my device information, if it is sent out in text messages" as the input for user intention analysis. We generate the user intention predicate T_s as $p =$ "Phone", $s = -1$ and $c =$"sent out in text messages". In the user-intention-guided taint flow analysis, we map T_s to a list of code inputs. Table 6 presents how we map "Phone" into source APIs and "text" into sink APIs. We perform the taint flow analysis in 13 different categories among 118 apps. Each category represents one particular evasion technique (e.g., callback). Without loss of generality, we randomly choose an app for rewriting in each category. The tested apps have a high coverage to mimic real-world apps. We identify a taint flow from a source to a sink within the taint flow graph. We generate the rewriting specification R_r by extract the unit u of the sink. For the rewriting operation op, we use unit removal to remove the sensitive sink API. We use unit insertion to add a dynamic logging function. The sink API is the leaf node in the ICFG, the unit removal has no impact on the rest code. To evaluate the feasibility of the rewriting, we install all the rewritten apps in a real-world device Nexus 6P with Android 6.0.1. We aim to answer two major questions:

1. Whether these apps remain valid program logics (do not break the functionality)?
2. Whether we would observe the modified bahavior at runtime (protect privacy)?

To answer the first question, we install rewritten apps on the device. We utilize *monkey* tool to generate a pseudo-

Category	Mapping
Device info	getDeviceId(),getSubscriberId(), getSimSerialNumber(), getLine1Number()
Sent by text	sendTextMessage(),sendDataMessage(), sendMultipartTextMessage(),sendMessage()
Never share	unit removal (sink unit), unit insertion (log unit)

Table 6: Mapping the sentence to analysis inputs for rewriting in the experiment. The sentence is from the example "Never share my device information, if it is sent out in text messages".

Category	AppName	PA time	RE time	Run	Logging Confirm
Alias	Merge1	1.74	6.40	✓	✓
ArrayList	Array Access1	1.13	4.34	✓	✓
Callbacks	Button1	2.52	4.48	✓	✓
FieldandObjctSensitivity	Inherited Objects1	1.27	4.24	✓	✓
InterAppCom	N/A	-	-	-	-
InterComponentCom	Activity Communication1	1.08	4.51	✓	✓
LifecyCycle	ActivityLifecycle2	1.15	4.20	✓	✓
GneralJava	Loop2	1.14	4.18	✓	✓
Android Specific	DirectLeak1	1.11	4.10	✓	✓
ImplicitFlows	N/A	-	-	-	-
Reflection	Reflection1	1.22	4.27	✓	✓
Threading	N/A	-	-	-	-
Emulator Detect	PlayStore1	1.90	6.46	✓	✓
Summary	10			10	10

Table 7: Runtime scalability for testing benchmark apps. N/A means we did not find any app matched the user command with in the taint analysis. Run w. monkey means the rewritten apps can run on the real-world device Nexus 6P. We use *monkey* to generate 500 random inputs in one testing. Confirmation means we detect the modified behavior using the adb *logcat*. PA is short for taint-flow-based program analysis time, RE is short for automatic rewriting time.

random stream of user events. We generate 500 random inputs in testing one app. We perform the stress test on the rewritten app. To answer the second question, we need to trigger taint flow paths and observe the modified behaviors. We use *logcat* to record the app running state. The dynamic logging function *Log.e()* can be captured by *logcat* at runtime. To increase the coverage of user interactions, We also manually interact with the app, e.g., clicking a button. If we observe such log information, the behavior of an app is modified and we validate our rewriting on the app.

Table 7 presents the robustness of our rewriting. Our approach successfully runs and detects all the rewritten apps. For example, we detect a taint flow path as $f = \{getDeviceId() \rightsquigarrow sendTextMessage()\}$ in $Button1.apk$. Our user-intention-guided taint flow analysis successfully captures the sensitive taint path triggered by a callback *onClick()*. Our rewriting technique identifies the taint flow path f and the rewriting unit $sendTextMessage()$. We modify the code in the app without violates its control and data dependence. These steps guarantee the success of our rewriting technique. In summary, our approach is very efficient in rewriting apps.

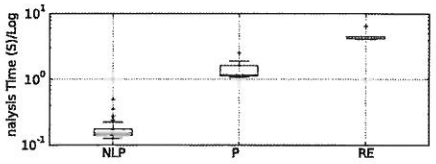


Figure 4: Time overhead for analysis, the average time for NLP analysis is 0.16 second, the average time for PA analysis is 1.43 seconds, the average for RE is 4.68 seconds.

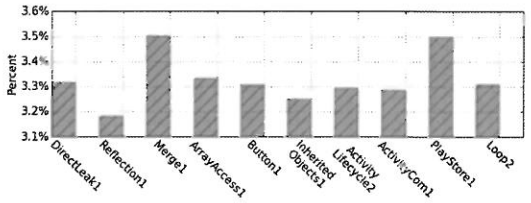


Figure 5: Size Overhead for benchmark apps, our rewriting introduces 3.3% percent overhead on file size for benchmark apps.

7.4 RQ3: Performance Overhead

We compare the runtime overhead of natural language processing (NLP), taint flow analysis (PA) and rewriting (RE) in Figure 4. Experiments were performed over on a Linux machine with Intel Xeon CPU (@3.50GHz) and 16G memory. Figure 4 presents the three runtime distributions in log scale. The average time for NLP is 0.16 second. The average time for PA is 1.43 seconds. The average time for RE is 4.68 seconds. The average runtime of RE is larger than that of NLP and PA. RE needs to decompile each class for inspection and recompile the app after rewriting.

Figure 8 presents the size overhead of rewriting benchmark apps. The size overhead comes from two major sources: *1)* the complexity of rewriting specifications; *2)* the number of impacted code in rewriting. In the experiment, the average file size of the benchmark apps is 305.23 KB. Our approach achieves 3.3% size overhead on average, which is relatively negotiable. In summary, our approach is very practical in rewriting apps.

7.5 RQ4: Security Applications

In this section, we show how to extend *IronDroid* for real-world problems. We present three security applications that cover *1)* proactive privacy protection, *2)* mitigating ICC vulnerability and *3)* location usage restriction. These security applications demonstrate the feasibility of our rewriting framework. Table 8 presents the statistics of three demo apps.

7.5.1 Proactive Privacy Protection

The purpose of this demo is to show the rewriting flexibility for normal users. *com.jb.azsingle.dcejec* is an e-electronic book app. The app passes all the 56 anti-virus tools for vetting screening. The user concerns on the data leakage to the storages. Users can proactively restrict the data usage. The sentence is that "do not reveal my phone information to outside storages". We use this example to show how a user could use *IronDroid* for a proactive privacy protection. Based on the our NLP analysis, we identify user in-

Package	Report	Label
com.jb.azsingle.dcejec	0/56	benign
com.xxx.explicit1	1/57*	grayware
com.bdsmartapp.prayerbd	0/56	benign

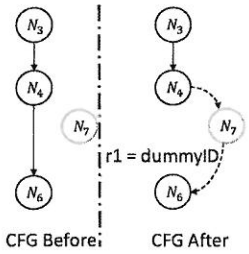
Table 8: All three apps pass the vetting screening of anti-virus tools. * means the alert mentions it contains a critical permission *READ_PHONE_STATE* without any additional information. We identify the second app as grayware [4]. Users have the preference for customizing these apps for personalized security.

tention predicate T_s = {"phone", -1, "outside storage"}. The user-intention-guided taint flow analysis identifies a critical taint flow graph f_1 in a class *com.ggbook.i.c*. The taint flow represents that the phone information is written into a file *KA.xml*. The sink unit is a sensitive function *putString*(). *KA.xml* is located in the device storage. The taint flow path f_1 is presented as below:

- f_1: [r3 = invoke r2.<TelephonyManager: String getDeviceId()>(), r4 = r3, return r4, <com.ggbook.c: String z> = r1, r1 = <com.ggbook.c: String z>, r7 = invoke r7.<SharedPreferences: putString()>("KA", r1)]

Our rewriting is based on the unit that invokes the *putString*() function. The unit is identified as r7 = invoke r7.<SharedPreferences: putString()>("KA", r1). In rewriting, we implement a dynamic checking function *IDcheck*(). *IDcheck*() accepts the parameter *r*1 from the unit. The type of *r*1 is Java.lang.String. *IDcheck*() dynamically checks *r*1 by comparing the value with the actual device ID p = *getDeviceId*(). If p = *getDeviceId*() $\in$ *r*1, *IDcheck*() returns a dummy ID and overwrite the parameter *r*1. The dummy ID can be further used without violating data flow dependencies. Figure 6 presents the code snippet and the CFG after we insert the *IDcheck*() function.

```
① com.ggbook.b(…){
② super.onCreate()
③ $r3=r2.getDeviceId()
④ $r1=$r3
⑤ $r7=SharedPrefer()
⑥ *$r7.putString("KA",$r1)
}
① IDcheck(p){
  if isDeviceId(p)
    return dummyID
  else return p}
```



(a) A simplified code structure of the method. The IDcheck function is inserted after rewriting.

(b) The before and after CFG. The CFG is changed. The privacy is protected by rewriting.

Figure 6: The demo of code snippet and the modified control flow graph (CFG). We insert a IDcheck() function as a new node in the original CFG.

To validate our rewriting feasibility and dummy ID insertion. We run the app with *monkey* on a real device. We manually find the shared preference file in the app file directory. The shared preference file is stored in /data/data/com.jb.azsingle.dcejec/shared_prefs, we identify the dummy ID as 12345678910. The entry *IsImei* = *true* suggests that the dummy ID is successfully inserted. The private phone data is protected from outside storages.

```
<map><string name="KA">12345678910</string>
<boolean name="IsImei" value="true"/></map>
```

7.5.2 *Mitigating ICC Vulnerability*

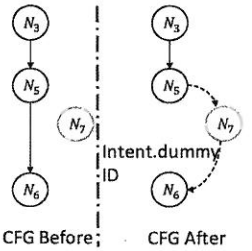
The purpose of this demo is to present the rewriting feasibility on mitigating vulnerabilities. In this example, we focus on preventing inter-component communication (ICC) vulnerability. Existing detection solutions aim at detecting vulnerable ICC paths [14, 29] with static program analysis. How to avoid realistic ICC vulnerabilities at runtime is still an open question.

We apply *IronDroid* to showcase a practical mitigation solution to prevent ICC vulnerability. *com.xxx.explicit*1 is an app in ICCBench[9]. The app contains a typical ICC vulnerability. The sensitive phone data (IMEI) is read from a component. The IMEI is put into a data field in an intent. The data is sent out to another component by the intent. The receiver component has access to IMEI without acquiring for the permission READ_PHONE_STATE. The ICC vulnerability results in a privilege escalation for the receiver component.

Our rewriting is based on a user command: "do not share my device information with others". The user-intention-guided taint flow analysis identifies a critical taint flow as f_2 in a class *explicit*1.*MainActivity*. The taint flow path f_2 is presented as:

- f_2: [r5 = invoke r4.< TelephonyManager: getDeviceId() > (), invoke r2.< putExtra(String,String)>("data",r5), invoke r0.< startActivity(android.content.Intent) >(r2)]

```
① onCreate(…){
② super.onCreate()
③ $r5=r4.getDeviceId()
④ $r2=new Intent()
⑤ $r2.putExtra('data',$r5)
⑥ *startActivity($r2)
}
① Intentcheck(intent){
  if checkIntent()
    modifyIntent()
    //insert dummyID
```



(a) A simplified code structure of the method. The IDcheck function is inserted after rewriting.

(b) The before and after CFG. The CFG is changed and the vulnerability is mitigated by rewriting.

Figure 7: The demo of code snippet and the modified control flow graph (CFG). We insert a Intentcheck() function as a new node in the original CFG.

Our rewriting specification focuses on the unit that invokes the *startActivity* function. *startActivity* is used for triggering an ICC. The intent *r*2 is an object that consists of the destination target (e.g., the package name) and data information (e.g., extras). Replacing intent would cause runtime exceptions. The communication of components is based on intents. The component that can receive the intent is defined in the package field of an intent. We aim to protect user privacy without interrupting communication channels. In rewriting, we implement a hierarchy checking function *Intentcheck*(). *Intentcheck*() accepts the parameter of an intent and modify the intent in-place. *Intentcheck* can recursively check all the fields in an intent and rewrite the intent. In this example, we rewrite an intent by inserting dummy data in the data field. Figure 7 presents the code snippet and the CFG after we insert the *Intentcheck*() function.

[9]https://goo.gl/EPjb3n

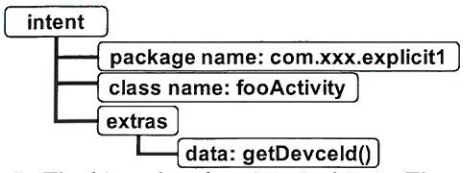


Figure 8: The hierarchy of an intent object. The privacy data getDdeviceId() is stored with a key *data* in a HashMap in *extras* filed.

To validate the modified intent, we run the app with *monkey* on a real device and manually intercept the receiver component. We successfully detect the dummy ID as 12345678910 in the data field of *getIntent*().

The above example demonstrates how a normal user could use *IronDroid* to reduce ICC vulnerability. For better privacy, our rewriting is also capable of redirecting an intent. One possible way to redirect an intent for security is changing an implicit intent to an explicit intent. In the explicit intent, the destination of the intent can be verified by our rewriting. We plan to extend our approach to mitigate other vulnerabilities, e.g., adversarial advertisement libraries.

7.5.3 *Location Usage Restriction*

The location is a primary concern for most users. *com.bdsmartapp.prayerbd* is an app to provide location-based praying service. It is unclear whether the app would abuse the location and send it to untrusted websites. The purpose of this demo is to show a fine-grained control of location usage. Current Android dynamic permission mechanism only provides turning on and off of the location permission. Our approach achieves extended functionalities: *1) Target inspection*, a user can inspect a targeted function (e.g., to web servers, by text messages and to other apps). *2) Destination inspection*, a user can inspect whether the packet is sent to an untrusted destination. *3) Sensitivity restriction*, a user can protect sensitive locations by sending dummy locations.

We extend our approach to providing a fine-grained restriction with a *LocationCheck* function. We rewrite the app based on the user sentence "do not send my home location to untrusted websites". We extracted user intention predicate T_s ={"Location", -1, "to untrusted website" }. The security object p = {"Location"} is mapped to APIs for reading locations (e.g., *getLatitude*() and *getLongitude*()). The the constraint c = {"to untrusted website" } is mapped to APIs for generating network requests (e.g., *URLConnect*()). In addition, we implement a function *Locationcheck* to determine *1)* the home location and *2)* the trustworthiness of a website. Advertisement websites (e.g., Admob, Flurry) are regarded as untrusted. The home location is defined by a location range. If the URL contains untrusted addresses or the current location is in the location range, the URL is modified by *Locationcheck* for privacy protection.

At runtime, a location request googleapis.com/maps/api/geocode/json?latlng=xxx,xxx&sensor=true is captured by *Locationcheck*. The location request is sent to the Google map server with *googleapis*. *Locationcheck* inspects the latitude and longitude in the location request. *Locationcheck* replaces the original location with a dummy latitude and longitude. A dummy location is shown on the screen layout when we test the rewritten app. In this demonstration, *IronDroid* protects user private location by supporting function-level restriction. More advanced version of our approach can be combined with anomaly data detection solutions.

8. RELATED WORK

NLP for Security and Privacy Compared with NLP applications in other areas (e.g., document analyzing), NLP has only been recently introduced in solving security problems. Prior works utilized NLP to discover online security threats [25] and analyze web privacy policies [44]. In the mobile security, Whype [31] and Autocog [34] used NLP to infer permission-related descriptions. Supor [21], Uipicker [28] and UiRef [9] detected sensitive user inputs from Android app layouts. Lei *et al* [24] generates inputs for program testing from input format specifications. Our work showcases a new application by integrating NLP with app customization. We demonstrate that new NLP techniques are required for personalized security challenges. Our approach requires innovative natural language analysis to understand user intentions for rewriting. We expect to extend our prototype with commercial voice control products (e.g., Siri, Alexa) in the future.

Taint Flow Analysis Researchers proposed taint flow analysis to identify data-flow paths from sources to sinks [10]. CHEX [26] and AndroidLeaks [19] detected data flows to find suspicious apps. DroidSafe [20] used a point-to graph to detect sensitive taint flows. FlowDroid [6] proposed a context-, object- and flow-sensitive static analysis to identify taint flows. Our research prototype extends FlowDroid for a precise taint flow analysis. Unlike these approaches aim to identify malware, we use taint flow analysis for personalized security. Our taint flow analysis is used to generate rewriting specifications for app customization.

Android Rewriting The app-retrofitting demonstration in RetroSkeleton [12] aimed at updating HTTP connections to HTTPS. Aurasium [42] instrumented a low-level library for monitoring functions. Reynaud *et al.* [36] rewrote an app security verification function to discovered in-app billing vulnerabilities. Fratantonio *et al.* [18] enforced secure usage of the Internet permission. The rewriting targets and goals in these approaches are specific. Many rewriting targets can be identified through parsing. Our rewriting approach requires more substantial code modification (e.g., parameter inspection, flow analysis). We demonstrate that more capable rewriting techniques are required for personalized security challenges. Our rewriting framework is general and compatible for complex app customization.

ARTist [8] is a compiler-base rewriting tool to modify Android runtime virtual machine. However, compiler-based rewriting relies on a particular Android version and needs modification of the Android system. Our bytecode rewriting modifies an app's bytecode and is compatible for all the Android versions. Bytecode rewriting is more suitable for our scenario for rewriting apps with a natural language interface.

Defense of Vulnerabilities Pluto [38] discovered the vulnerabilities in advertisement libraries. TaintDroid [15] adopted dynamic taint analysis to track the potential misuse of sensitive data. Anception [17] proposed a mechanism to defend the privilege escalation by deprivileging a portion of the kernel with sensitive operations. DeepDroid [40] applied dynamic memory instrumentations to enforce security policies. Instead of modifying the kernel or Android framework, our approach focuses on the app-level rewriting. We demon-

strate promising security applications of app customization beyond defending vulnerabilities.

9. CONCLUSIONS AND FUTURE WORK

We investigated the problem of app customization for personalized security. We proposed a user-centric app customization framework with natural language processing and rewriting. Our preliminary experimental results show that our prototype is very accurate in understanding user intentions. The prototype is also very efficient in rewriting. We show its applications in providing more fine-grained location control and mitigating existing vulnerabilities. Our approach makes impressive progress toward personalized app customization. More efforts are needed to improve the usability of personalized app customization. For future work, we plan to improve the usability of our prototype with more engineering efforts.

10. REFERENCES

[1] Amazon Alexa. https://developer.amazon.com/alexa.

[2] Android dynamic permission. https://developer.android.com/training/permissions/requesting.html.

[3] Apple Siri. http://www.apple.com/ios/siri/.

[4] Andow, B., Nadkarni, A., Bassett, B., Enck, W., and Xie, T. A study of grayware on Google Play. In *Proc. of MoST* (2016).

[5] Arp, D., Spreitzenbarth, M., Hübner, M., Gascon, H., Rieck, K., and Siemens, C. Drebin: Effective and explainable detection of Android malware in your pocket. In *Proc. of NDSS* (2014).

[6] Arzt, S., Rasthofer, S., Fritz, C., Bodden, E., Bartel, A., Klein, J., Le Traon, Y., Octeau, D., and McDaniel, P. FlowDroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for Android apps. In *Proc. of PLDI* (2014).

[7] Bach, N., and Badaskar, S. A review of relation extraction. In *Literature review for Language and Statistics II* (2007).

[8] Backes, M., Bugiel, S., Schranz, O., von Styp-Rekowsky, P., and Weisgerber, S. ARTist: The Android runtime instrumentation and security toolkit.

[9] Benjamin Andow, Akhil Acharya, D. L. W. E. K. S., and Xie, T. Uiref: Analysis of sensitive user inputs in android applications. In *Proc. of WiSec* (2017).

[10] Bugiel, S., Davi, L., Dmitrienko, A., Fischer, T., Sadeghi, A.-R., and Shastry, B. Towards taming privilege-escalation attacks on Android. In *Proc. of NDSS* (2012).

[11] Cer, D. M., De Marneffe, M.-C., Jurafsky, D., and Manning, C. D. Parsing to stanford dependencies: Trade-offs between speed and accuracy. In *Proc. of LREC* (2010).

[12] Davis, B., and Chen, H. RetroSkeleton: Retrofitting Android Apps. In *Proc. of MobiSys* (2013).

[13] Davis, B., Sanders, B., Khodaverdian, A., and Chen, H. I-ARM-Droid: A rewriting framework for in-app reference monitors for android applications.

[14] Elish, K. O., Yao, D. D., and Ryder, B. G. On the need of precise inter-app icc classification for detecting Android malware collusions. In *Proc. of IEEE Mobile Security Technologies (MoST)* (2015).

[15] Enck, W., Gilbert, P., Han, S., Tendulkar, V., Chun, B.-G., Cox, L. P., Jung, J., McDaniel, P., and Sheth, A. N. TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS)* (2014).

[16] Enck, W., Octeau, D., McDaniel, P., and Chaudhuri, S. A study of android application security. In *Proc. of USENIX Security* (2011).

[17] Fernandes, E., Aluri, A., Crowell, A., and Prakash, A. Decomposable trust for Android applications. In *Proc. of DSN* (2015).

[18] Fratantonio, Y., Bianchi, A., Robertson, W., Egele, M., Kruegel, C., Kirda, E., Vigna, G., Kharraz, A., Robertson, W., Balzarotti, D., et al. On the security and engineering implications of finer-grained access controls for Android developers and users. In *Proc. of DIMVA* (2015).

[19] Gibler, C., Crussell, J., Erickson, J., and Chen, H. AndroidLeaks: Automatically detecting potential privacy leaks in Android applications on a large scale. In *Proc. of TRUST* (2012).

[20] Gordon, M. I., Kim, D., Perkins, J., Gilham, L., Nguyen, N., and Rinard, M. Information-flow analysis of Android applications in DroidSafe. In *Proc. of NDSS* (2015).

[21] Huang, J., Li, Z., Xiao, X., Wu, Z., Lu, K., Zhang, X., and Jiang, G. SUPOR: Precise and scalable sensitive user input detection for android apps. In *proc. of USENIX Security* (2015).

[22] Klein, D., and Manning, C. D. Accurate unlexicalized parsing. In *Proc. of ACL* (2003).

[23] Lam, P., Bodden, E., Lhoták, O., and Hendren, L. The soot framework for java program analysis: a retrospective. In *Proc. of CETUS* (2011).

[24] Lei, T., Long, F., Barzilay, R., and Rinard, M. C. From natural language specifications to program input parsers. In *Proc. of ACL* (2013).

[25] Liao, X., Yuan, K., Wang, X., Li, Z., Xing, L., and Beyah, R. Acing the ioc game: Toward automatic discovery and analysis of open-source cyber threat intelligence. In *Proc, of CCS* (2016).

[26] Lu, L., Li, Z., Wu, Z., Lee, W., and Jiang, G. CHEX: statically vetting Android apps for component hijacking vulnerabilities. In *Proc. of CCS* (2012).

[27] Nadeau, D., and Sekine, S. A survey of named entity recognition and classification. *Lingvisticae Investigationes* (2007).

[28] Nan, Y., Yang, M., Yang, Z., Zhou, S., Gu, G., and Wang, X. Uipicker: User-input privacy identification in mobile applications. In *Proc. of USENIX Security* (2015).

[29] Octeau, D., Jha, S., Dering, M., McDaniel, P., Bartel, A., Li, L., Klein, J., and Le Traon, Y. Combining static analysis with probabilistic models to enable market-scale Android inter-component analysis. In *Proc. of POPL* (2016).

[30] Octeau, D., Jha, S., and McDaniel, P. Retargeting android applications to java bytecode. In *Proc. of FSE* (2012).

[31] Pandita, R., Xiao, X., Yang, W., Enck, W., and Xie, T. WHYPER: Towards automating risk assessment of mobile applications. In *Proc. of USENIX Security* (2013).

[32] Pedersen, T., Patwardhan, S., and Michelizzi, J. Wordnet:: Similarity: measuring the relatedness of concepts. In *Demonstration papers at HLT-NAACL* (2004).

[33] Poeplau, S., Fratantonio, Y., Bianchi, A., Kruegel, C., and Vigna, G. Execute this! analyzing unsafe and malicious dynamic code loading in android applications. In *Proc. of NDSS* (2014).

[34] Qu, Z., Rastogi, V., Zhang, X., Chen, Y., Zhu, T., and Chen, Z. Autocog: Measuring the description-to-permission fidelity in android applications. In *Proc. of CCS* (2014).

[35] Rasthofer, S., Arzt, S., and Bodden, E. A machine-learning approach for classifying and categorizing Android sources and sinks. In *Proc. of NDSS* (2014).

[36] Reynaud, D., Song, D. X., Magrino, T. R., Wu, E. X., and Shin, E. C. R. Freemarket: Shopping for free in Android applications. In *Proc. of NDSS* (2012).

[37] Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., Potts, C., et al. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proc. of EMNLP* (2013).

[38] Soteris Demetriou, Whitney Merrill, W. Y. A. Z., and Gunter, C. A. Free for all! assessing user data exposure to advertising libraries on Android. In *Proc. of NDSS* (2016).

[39] Tian, K., Yao, D. D., Ryder, B. G., and Tan, G. Analysis of code heterogeneity for high-precision classification of repackaged malware. In *Proc. of MoST* (2016).

[40] Wang, X., Sun, K., Wang, Y., and Jing, J. DeepDroid: Dynamically enforcing enterprise policy on Android devices. In *Proc. of NDSS* (2015).

[41] Wüchner, T., Ochoa, M., and Pretschner, A. Robust and effective malware detection through quantitative data flow graph metrics. In *Proc. of DIMVA* (2015).

[42] Xu, R., Saïdi, H., and Anderson, R. Aurasium: Practical policy enforcement for Android applications. In *Proc. of USENIX Security* (2012).

[43] Zhang, M., Duan, Y., Yin, H., and Zhao, Z. Semantics-aware Android malware classification using weighted contextual api dependency graphs. In *Proc. of CCS* (2014).

[44] Zimmeck, S., and Bellovin, S. M. Privee: An architecture for automatically analyzing web privacy policies. In *Proc. of*