IEEE SecDev|2021

IEEE COMPUTER SOCIETY

Analyzing OpenAPI Specifications for Security Design Issues

Carmen Cheh and Binbin Chen

Motivation

T-Mobile Alerts 2.3 Million Customers of Data Breach Tied to Leaky API

(Threatpost, 2018)

Don't Put It on the Internet: Tesla Backup Gateway Edition

Nov 17, 2020 | 13 min read | Derek Abdine

(Rapid7, 2020)

DAN SALMON SECURITY 06.26.2019 09:00 AM

I Scraped Millions of Venmo Payments. Your Data Is at Risk

Opinion: Venmo makes sending and receiving money a social affair. But those emoji-laden payment descriptions leave you exposed to cyberattacks.

(Wired, 2019)

Related Work

State-of-the-Practice Tools

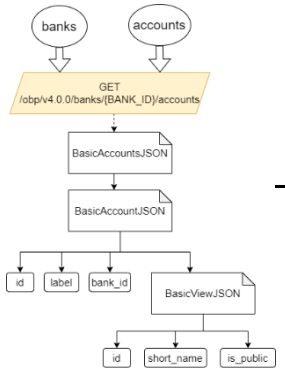
State-of-the-Art Research

- Focused on discovering design flaws
- Atlidakis et al. [1] (2020)
  - Four security rules concerning design and operations on data objects
  - Vulnerability can result in privilege escalation or sensitive data exposure
  - Stateful API fuzzer with property checker
- Iversen [2] (2018)
  - Dynamic analyses check HTTP status code and response body
  - Static analyses finds patterns in specification (e.g., keywords)

[1] V. Atlidakis, P. Godefroid, and M. Polishchuk, "Checking security properties of cloud services REST APIs," in *2020 IEEE 13th International Conference on Software Testing, Validation and Verification,* pp. 387-397.
[2] P. Iversen, "Specification-based security analysis of REST APIs," Master's thesis, NTNU, 2018.

Our Contribution: Security Analyses Approach

OpenAPI Graph Model



IDENTIFY SENSITIVE FIELDS

Shortlist high-frequency fields

Here are fields familiar to you. Which are sensitive or public info?

X and Y are sensitive.

Expand list of sensitive/non-sensitive fields

We think Z and A are also sensitive. Is this true?

We agree Z is sensitive.

List of sensitive and non-sensitive fields

INFER SECURITY DESIGN ISSUES

X, Y, and Z are sensitive.

Identify insecure & high-risk API call

API calls B, C, and D are potentially insecure.

Calculate exposure level

Here are the exposure levels for X, Y, and Z – and B, C, and D.

Case Study: Open Bank Project (OBP)

- Global standard and open source API solution for open banking

- Statistics:
  - 304 API calls
  - 142 resources
  - 345 schemas
  - 402 data fields

- Available documents: **[Retrieved May 1, 2021]**
  - OpenAPI specification
  - Glossary – Richer details about API calls

We use this as a source of ground truth

OBP: OpenAPI Specification

Sample API call Specification:

GET /obp/v4.0.0/banks/{BANK_ID}/accounts

```
'/obp/v4.0.0/banks/{BANK_ID}/accounts':
  get:
    security:
      - directLogin: [ ]
    description: 'Returns the list of accounts at BANK_ID that the user has access to. Authentication is Mandatory.'
    parameters:
      - in: path
        name: BANK_ID
        required: true
        type: string
    responses:
      '200':
        schema:
          $ref: '#/definitions/BasicAccountsJSON'
      '400':
        schema:
          $ref: '#/definitions/ErrorBankNotFound'
```

```
BasicAccountsJSON:
  required:
    - accounts
  properties:
    accounts:
      type: array
      items:
        $ref: '#/definitions/BasicAccountJSON'
```

```
BasicAccountJSON:
  required:
    - id
    - label
    - bank_id
    - views_available
  properties:
    id:
      type: string
    label:
      type: string
    bank_id:
      type: string
    views_available:
      type: array
      items:
        $ref: '#/definitions/BasicViewJson'
```

Sample HTTP 200 Response


```
{
  "accounts": [
    {
      "id": "8cde9f0",
      "label": "NoneLabel",
      "bank_id": "GENODEM1GLS",
      "views_available": [
        {
          "id": "1",
          "short_name": "HHH",
          "is_public": true
        }
      ]
    }
  ]
}
```


SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

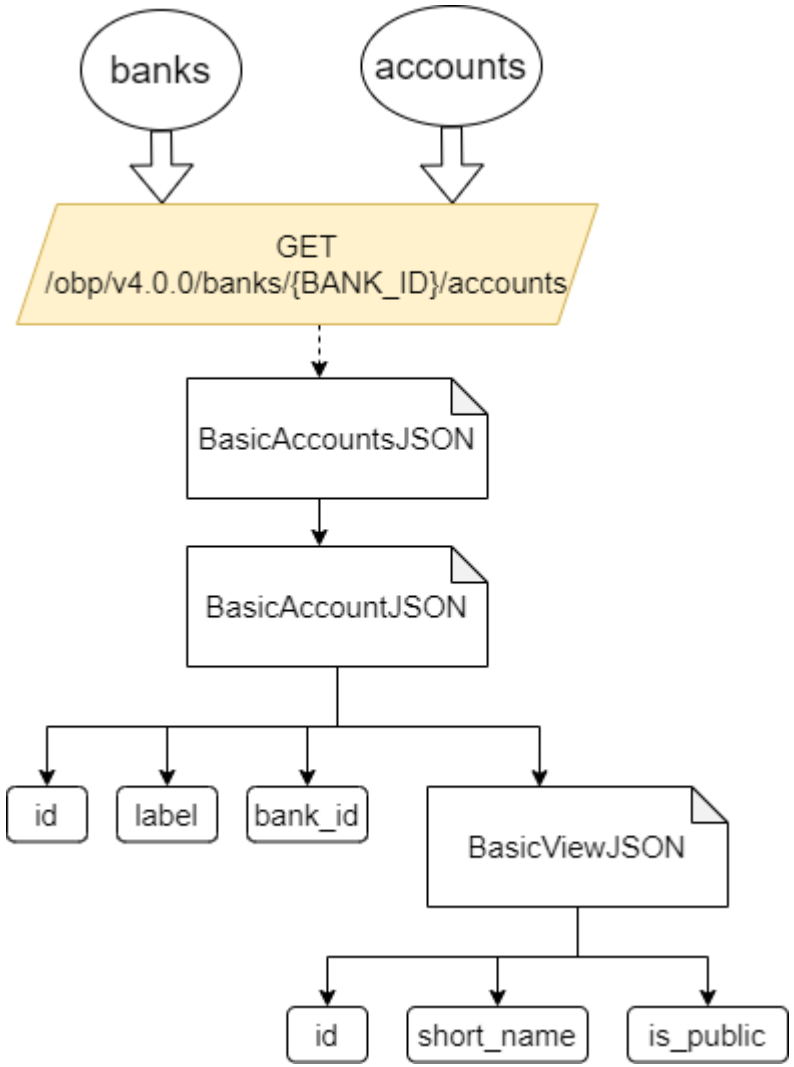
OpenAPI Graph Model

Resources

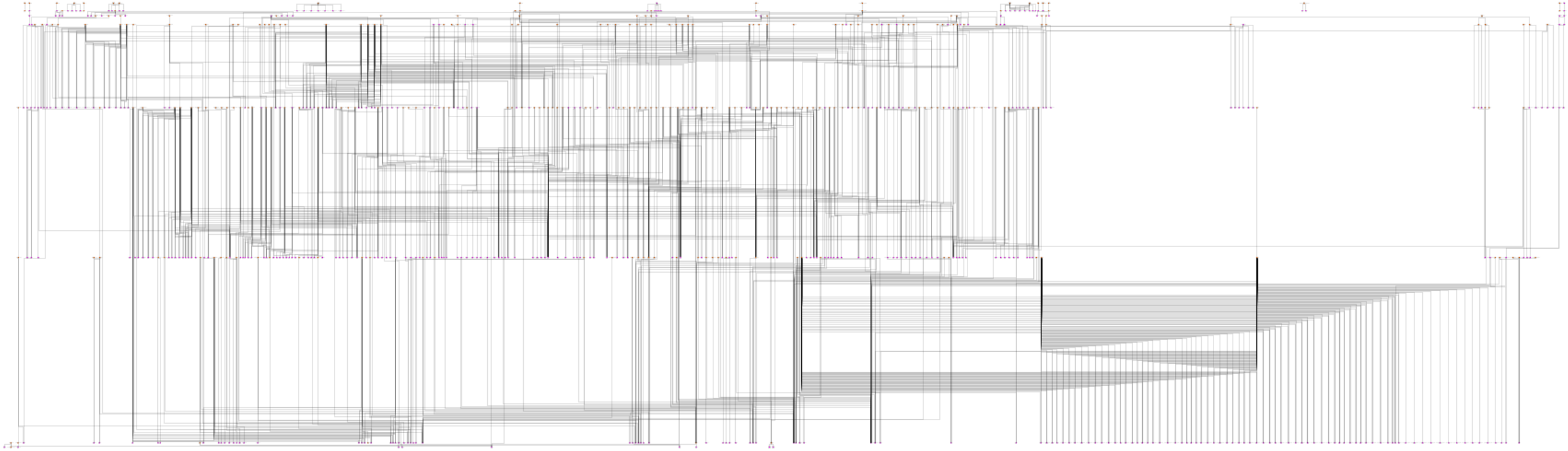
API Call

Schemas

Data Fields

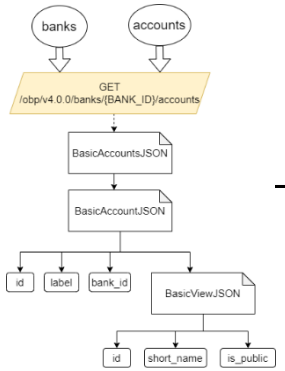


OBP: Full OpenAPI Graph Model



Identify Sensitive Fields: OBP

OpenAPI Graph Model



IDENTIFY SENSITIVE FIELDS

Shortlist high-frequency fields

Here are fields familiar to you. Which are sensitive or public info?

X and Y are sensitive.

Expand list of sensitive/non-sensitive fields

We think Z and A are also sensitive. Is this true?

We agree Z is sensitive.

List of sensitive and non-sensitive fields

INFER SECURITY DESIGN ISSUES

Identify insecure & high-risk API call

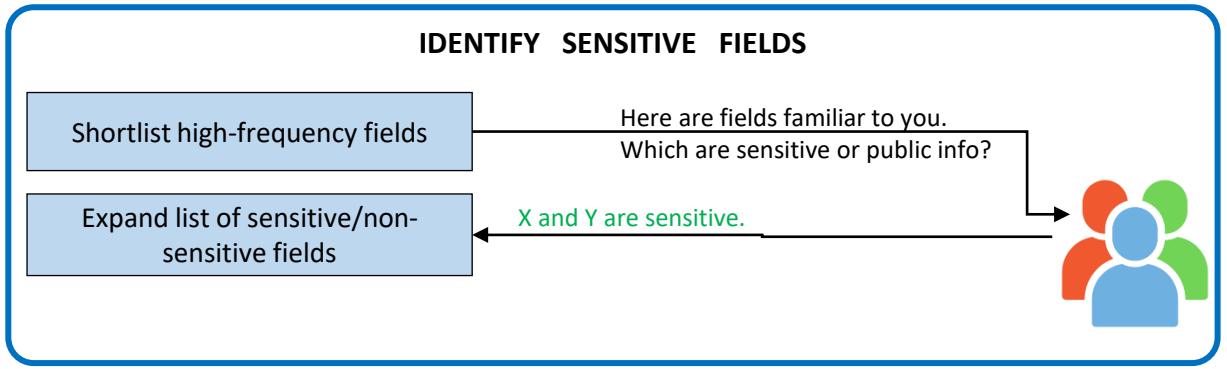
X, Y, and Z are sensitive.

API calls B, C, and D are potentially insecure.

Calculate exposure level

Here are the exposure levels for X, Y, and Z – and B, C, and D.

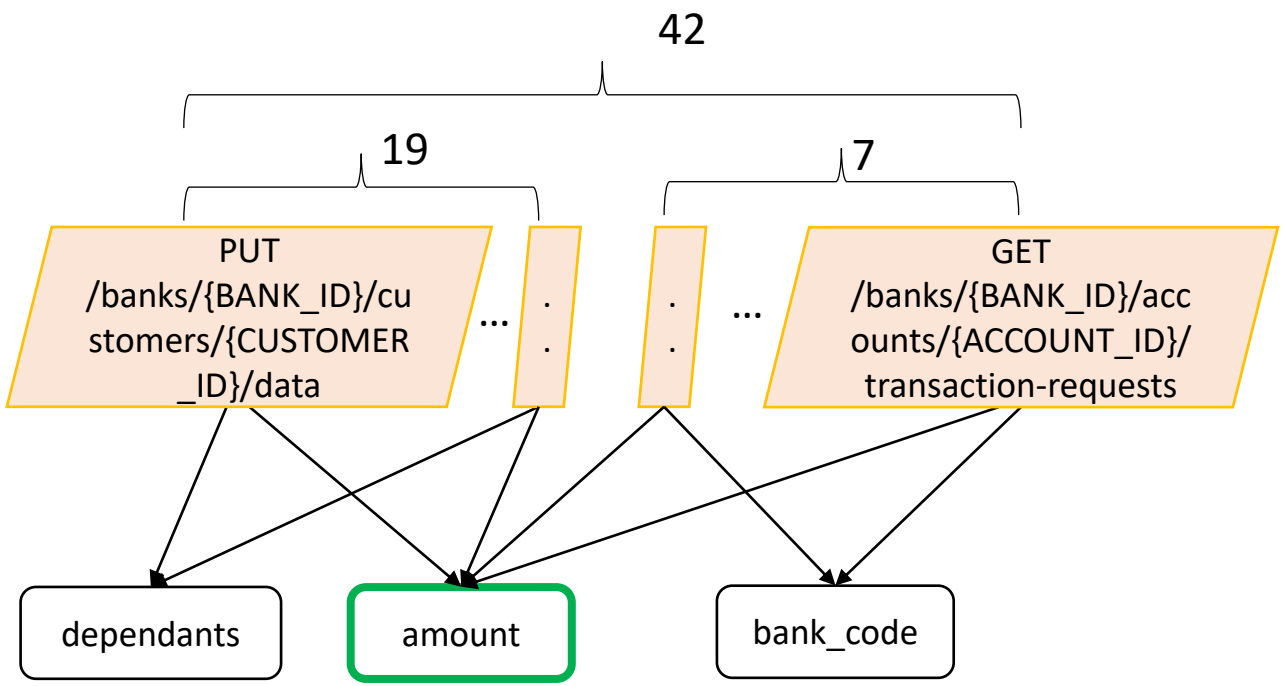
Identify Sensitive Fields: Results

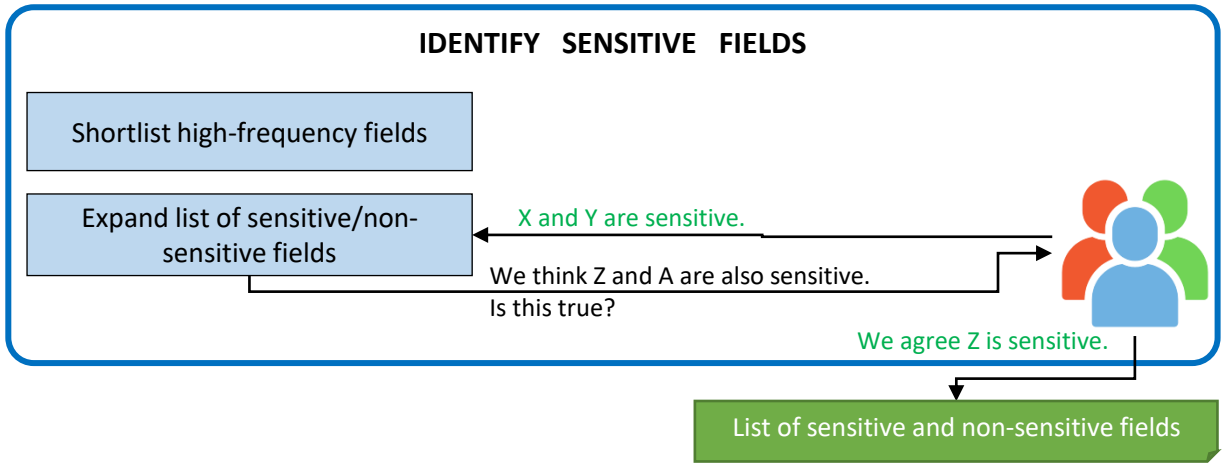


How many resources use this data field?

Field	Num of Resources
bank_id	66
jsonString	47
date	43
currency	39
amount	39
email	33
type	32
provider	31
customer_id	27
legal_name	26
date_of_birth	26
mobile_phone_number	26

Identify Sensitive Fields: Results





How close is this data field to the sensitive data field?

$$distance(amount, dependants) = 1 - \frac{19}{42} = 0.55$$

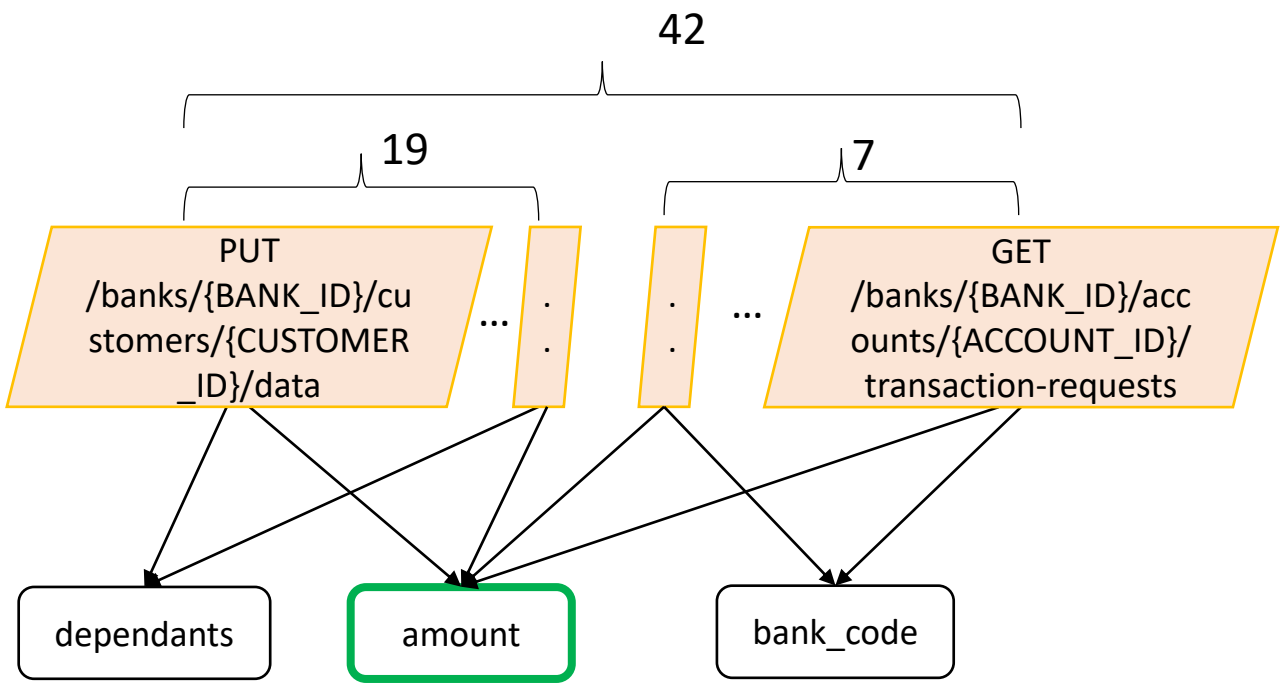
$$distance(amount, bank\_code) = 1 - \frac{7}{42} = 0.83$$

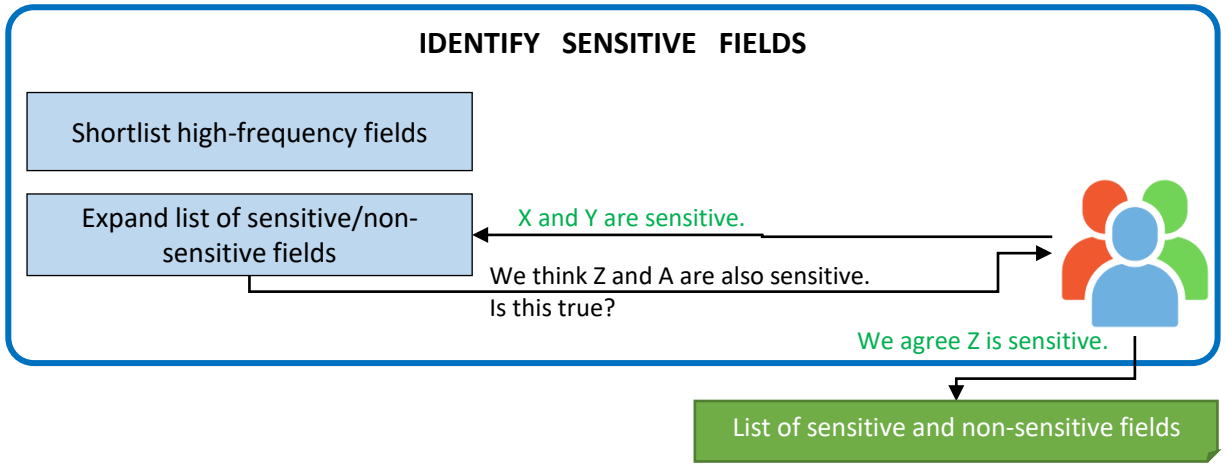
Threshold

$25th\ percentile\ of\ all\ distance(amount, \cdot)$

0.805

Identify Sensitive Fields: Results





How close is this data field to the sensitive data field?

customer_id	legal_name	last_ok_date	rating	kyc_status
account_id	mobile_phone_number	challenge_type	bank_id	username
amount	branch_id	customer_number	type	provider_id
currency	transaction_ids	user_id	dob_of_dependants	role_name
provider	created	relationship_status	highest_education_attained	entitlement_id
date_of_birth	date	dependants	employment_status	password

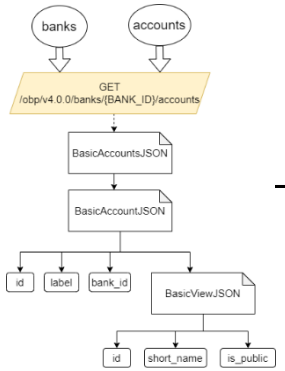
...

Manually identified sensitive fields based on glossary

$$\frac{22}{32}$$

Infer Security Issues: OBP

OpenAPI Graph Model



IDENTIFY SENSITIVE FIELDS

Shortlist high-frequency fields

Here are fields familiar to you. Which are sensitive or public info?

X and Y are sensitive.

Expand list of sensitive/non-sensitive fields

We think Z and A are also sensitive. Is this true?

We agree Z is sensitive.

List of sensitive and non-sensitive fields

INFER SECURITY DESIGN ISSUES

X, Y, and Z are sensitive.

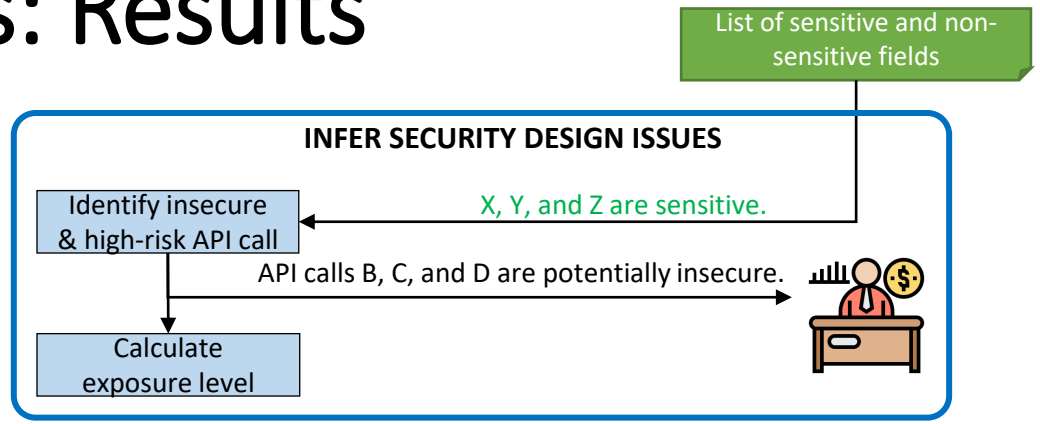
Identify insecure & high-risk API call

API calls B, C, and D are potentially insecure.

Calculate exposure level

Here are the exposure levels for X, Y, and Z – and B, C, and D.

Infer Security Issues: Results



What unauthenticated API calls return sensitive data?

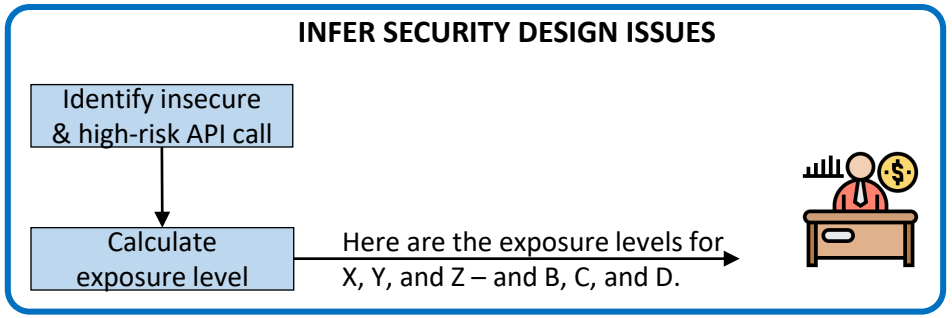
GET /obp/v4.0.0/customers/CUSTOMER_ID/kyc_documents — *Returns a list of documents that affirm the identity of the customer (e.g., passport, driving licence).*

GET /obp/v4.0.0/banks/{BANK_ID}/balances — *Get the balances for the accounts of the current user at a bank.*

What API calls return both sensitive data and non-sensitive data?

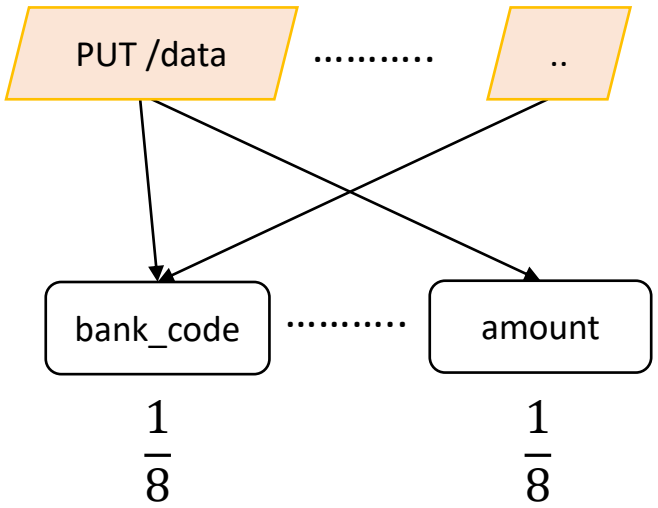
34 API calls found

Infer Security Issues: Results

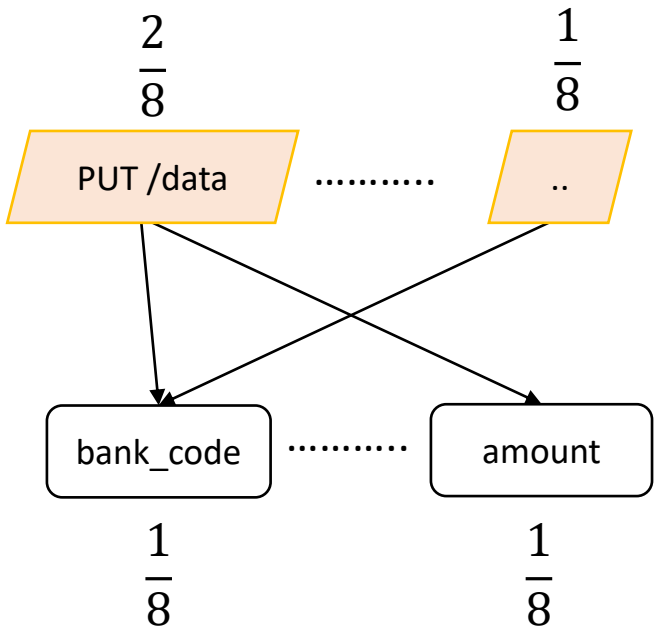


How exposed is this data field and API call to the public?

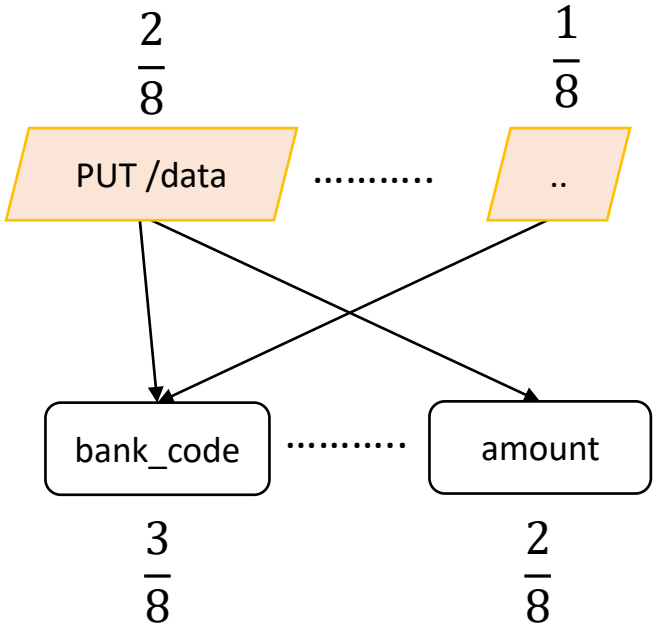
Iteration 0



Iteration 1



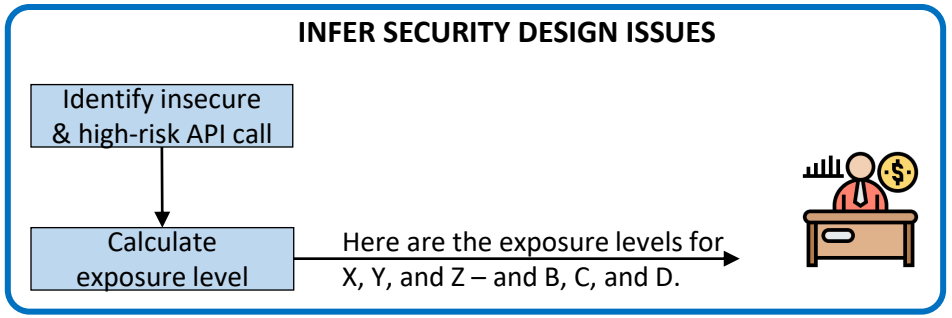
Iteration 2



Infer Security Issues: Results

How exposed is this data field to the public?

Field	Exposure
customer_id	37.3
account_id	32.3
amount	27.7
currency	27.7
provider	25.8
date_of_birth	13.2
legal_name	13.2
mobile_phone_number	13.2
transaction_ids	9.01
customer_number	4.64
user_id	4.33
relationship_status	4.25
dependants	4.25
rating	4.25
dob_of_dependants	4.25
highest_education_attained	4.25
employment_status	4.25
username	4.02



How exposed is this API call to the public?

API Call
PUT /management/banks/BANK_ID/cards/CARD_ID
POST /management/banks/BANK_ID/cards
GET /management/banks/BANK_ID/cards
GET /banks/BANK_ID/accounts/ACCOUNT_ID/VIEW_ID/account
GET /banks/BANK_ID/accounts/ACCOUNT_ID/permissions

Conclusion and Future Work

- Analyzing OpenAPI specifications using the relationships between API calls and data fields allow us to:
  - **Identify sensitive data**
  - **Highlight potentially insecure or high-risk API calls**

- **Future research area:** Identifying additional security design issues
  - **Potential solutions:** Looking at other components in OpenAPI specifications (e.g., data field type)