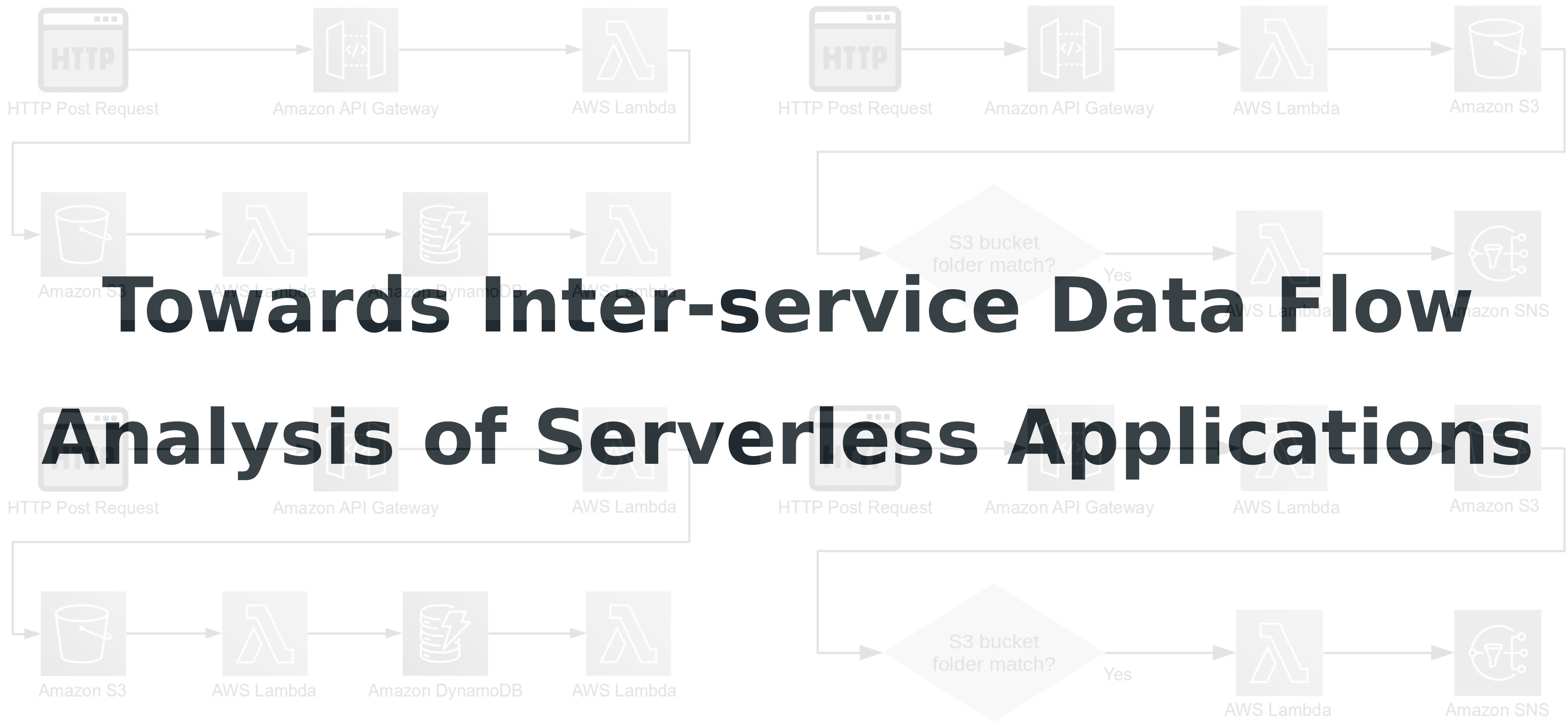
Towards Inter-service Data Flow Analysis of Serverless Applications



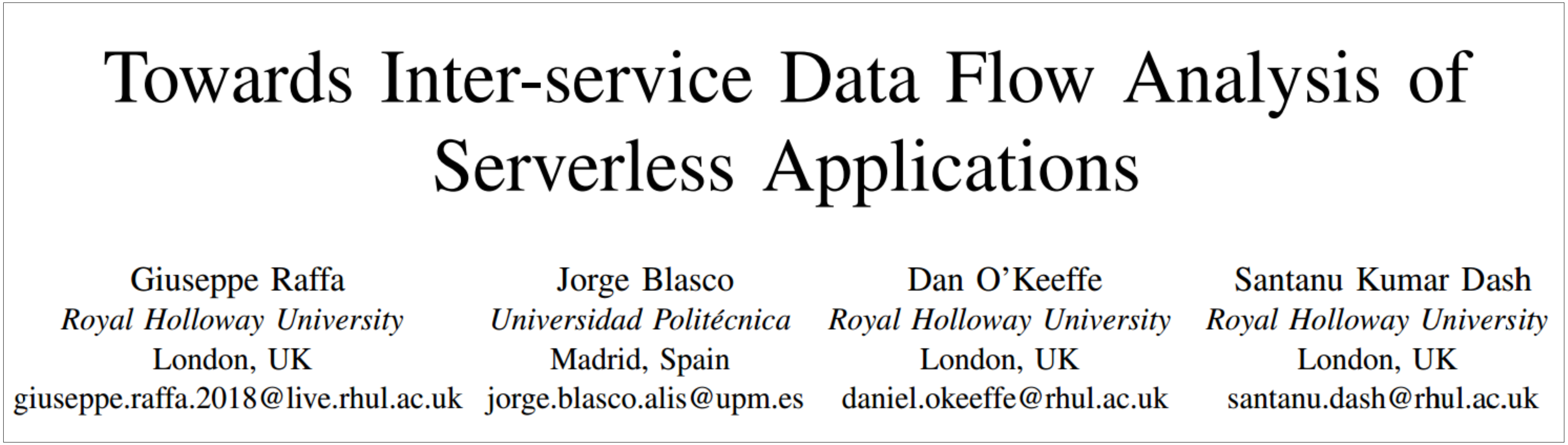
Giuseppe Raffa, J. Blasco, D. O’Keeffe, S. K. Dash

MISCS Doctoral Conference, 23rd April 2024

Contact email: giuseppe.raffa.2018@live.rhul.ac.uk

ROYAL HOLLOWAY UNIVERSITY Of LONDON

SANER 2024 Paper



Towards Inter-service Data Flow Analysis of Serverless Applications

Giuseppe Raffa
Royal Holloway University
London, UK
giuseppe.raffa.2018@live.rhul.ac.uk

Jorge Blasco
Universidad Politécnica
Madrid, Spain
jorge.blasco.alis@upm.es

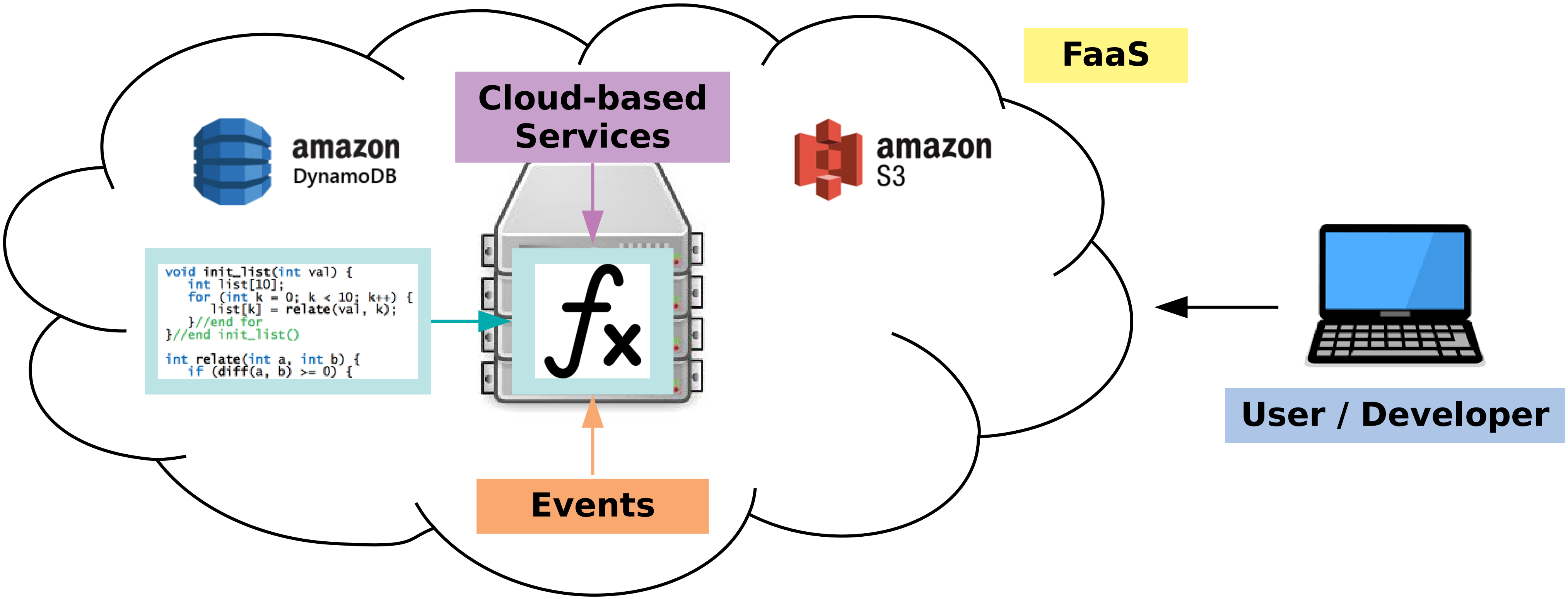
Dan O'Keeffe
Royal Holloway University
London, UK
daniel.okeeffe@rhul.ac.uk

Santanu Kumar Dash
Royal Holloway University
London, UK
santanu.dash@rhul.ac.uk

SANER 2024 Early Research Achievement (ERA) Track

https://github.com/giusepperaffa/serverless-security-microbenchmarks

Serverless Computing Model



```
void init_list(int val) {
   int list[10];
   for (int k = 0; k < 10; k++) {
      list[k] = relate(val, k);
   }//end for
}//end init_list()

int relate(int a, int b) {
   if (diff(a, b) >= 0) {
```

- **Advantages**
  - Cost-effectiveness
  - No infrastructure management
- **Disadvantages**
  - Debugging
  - Execution-related limits

Motivation & Challenges

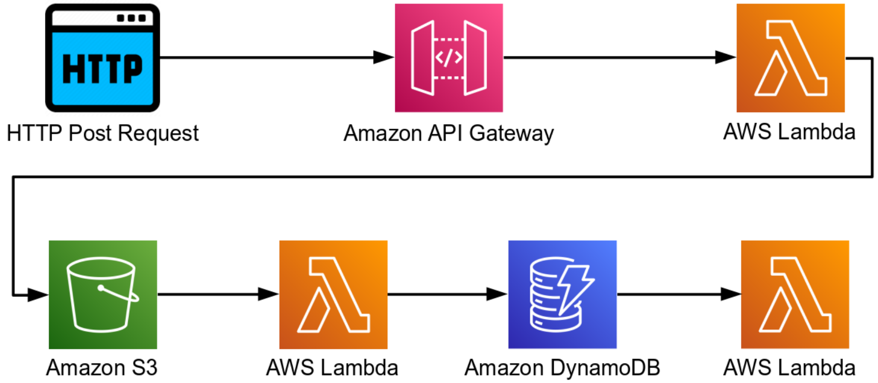
- **Why static data flow analysis?**
  - Most of serverless security tools rely on dynamic analysis
  - Static analysis is an effective supplement
- **What are the challenges?**
  - Information from infrastructure and application code
  - Variety of sources and events
  - Black-box nature of platform services
- **Our work**

Suite of security-oriented microbenchmarks

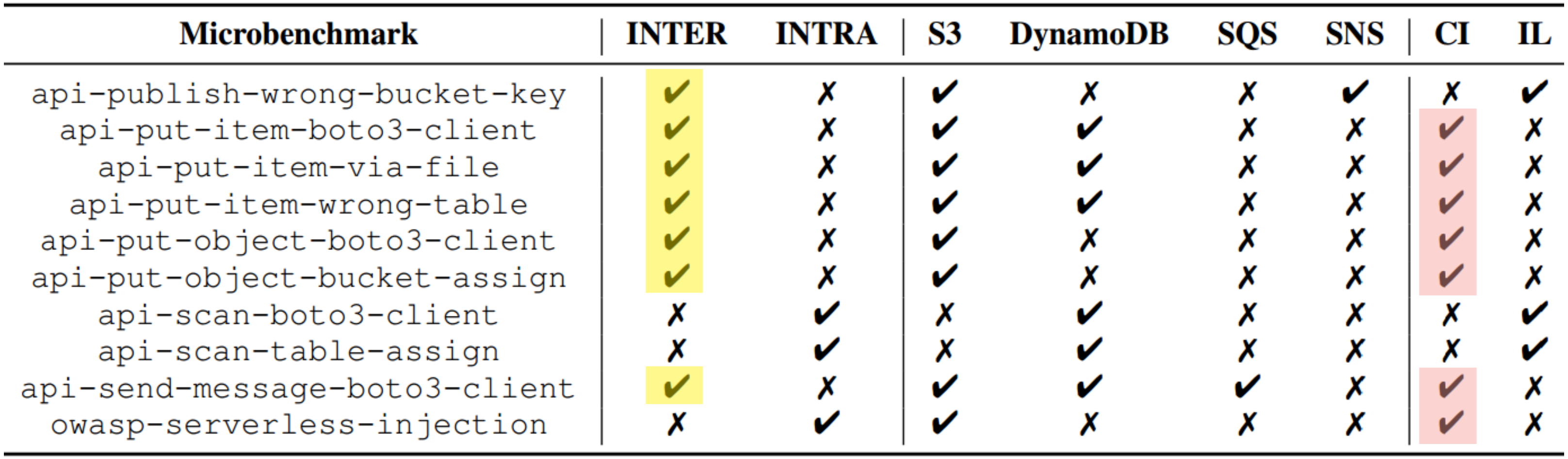
Approach to detecting security-sensitive data flows

Microbenchmarks Suite

- **Design approach**
  - Code injection and information leakage vulnerabilities
  - AWSomePy dataset characterization

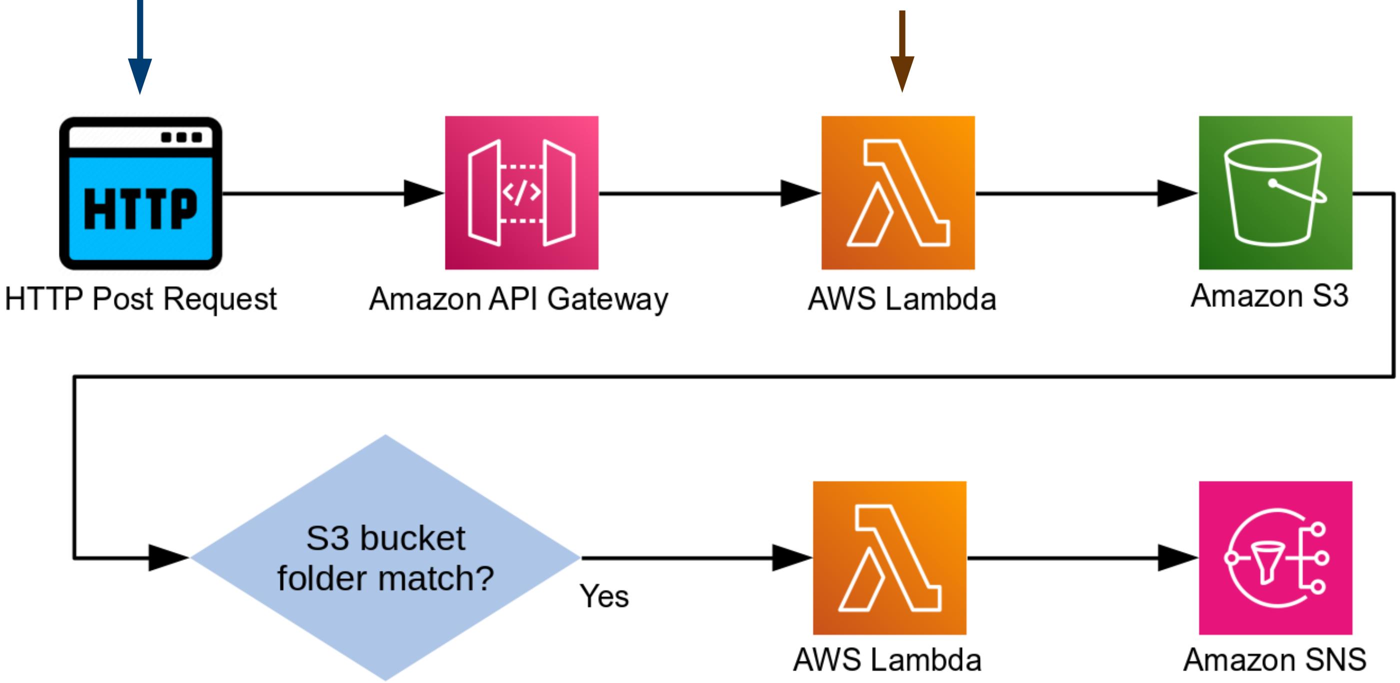


- **Summary**



	Flow		Services				Vuln.	
Microbenchmark	INTER	INTRA	S3	DynamoDB	SQS	SNS	CI	IL
api-publish-wrong-bucket-key	✔	✗	✔	✗	✗	✔	✗	✔
api-put-item-boto3-client	✔	✗	✔	✔	✗	✗	✔	✗
api-put-item-via-file	✔	✗	✔	✔	✗	✗	✔	✗
api-put-item-wrong-table	✔	✗	✔	✔	✗	✗	✔	✗
api-put-object-boto3-client	✔	✗	✔	✗	✗	✗	✔	✗
api-put-object-bucket-assign	✔	✗	✔	✗	✗	✗	✔	✗
api-scan-boto3-client	✗	✔	✗	✔	✗	✗	✗	✔
api-scan-table-assign	✗	✔	✗	✔	✗	✗	✗	✔
api-send-message-boto3-client	✔	✗	✔	✔	✔	✗	✔	✗
owasp-serverless-injection	✗	✔	✔	✗	✗	✗	✔	✗

Information Leakage Example



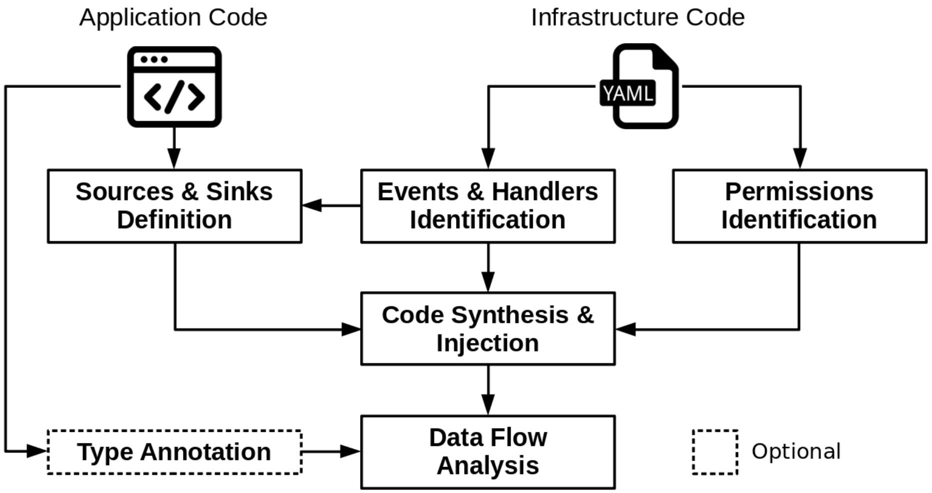
User-controlled entry point of the application

Request inspection and file uploaded to a S3 bucket

Email notification with user-provided information

Prototype Analysis Framework

- **Analysis approach**
  - Infrastructure and application code processed
  - Code instrumented to obtain synchronous equivalent



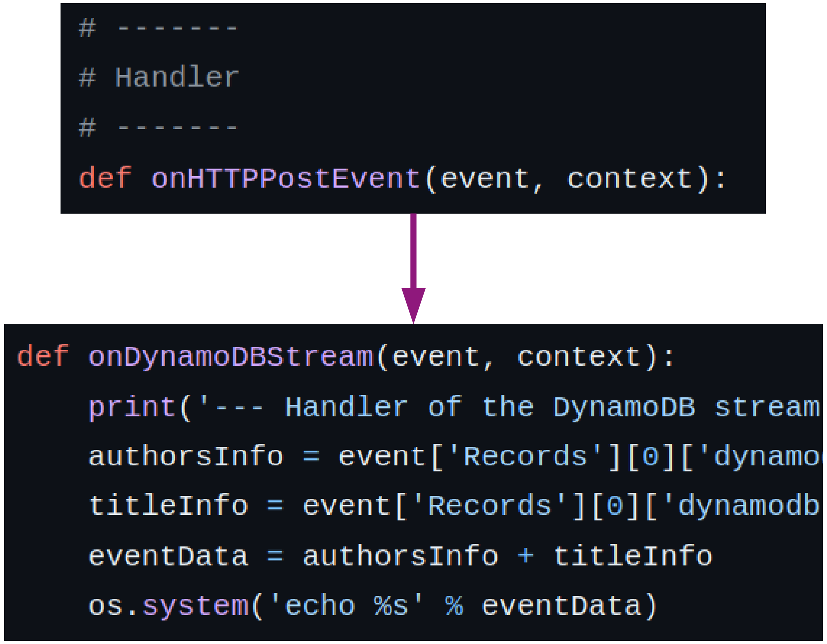
Application Code
Infrastructure Code
YAML
Sources & Sinks Definition
Events & Handlers Identification
Permissions Identification
Code Synthesis & Injection
Type Annotation
Data Flow Analysis
Optional

- **Implementation**
  - Code modified semi-automatically
  - Data flow analysis with Pysa
- **Evaluation**

7 true positives	2 false positives	1 false negative

Related Work

- **Obetz et al. [1], [2]**
  - Main objective:
    - Call graph generation
- **Our work**
  - Main objective:
    - Security-sensitive data flows identification



```
# -------
# Handler
# -------
def onHTTPPostEvent(event, context):
```

```
def onDynamoDBStream(event, context):
    print('--- Handler of the DynamoDB stream
    authorsInfo = event['Records'][0]['dynamo
    titleInfo = event['Records'][0]['dynamodb
    eventData = authorsInfo + titleInfo
    os.system('echo %s' % eventData)
```

[1] M. Obetz et al. 2019. Static Call Graph Construction in AWS Lambda Serverless Applications. In 11th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 19).

[2] M. Obetz et al. 2020. Formalizing Event-Driven Behavior of Serverless Applications. In European Conference on Service-Oriented and Cloud Computing (ESOCC 2020).

Conclusion & Future Work

- **Key takeaways**

Security-sensitive data flows

New suite of microbenchmarks

Studied approach is feasible

- **Future work**
  - Fully automated analysis pipeline
  - Improvement of infrastructure code processing
  - Support for higher number of cloud services and APIs

https://github.com/giusepperaffa/serverless-security-microbenchmarks