Mahesh Arumugam

linkedin.com/in/aumahesh   github.com/aumahesh
aumahesh@gmail.com   aumahesh.github.io   +1 (408) 904-9746   San Jose, CA

Seeking a software architect role in distributed systems that focusses on data and learning insights from data

Technical professional with over 17 years of diverse and progressively responsible experience in distributed systems, analytics, security, infrastructure as code, research and development, team leadership, product development, engineering, problem-solving and mentoring. Excellent verbal and written communication skills, attention to detail, customer focus, and a strong commitment to work. Strong research and analytic skills that are backed up by the ability to deliver results within time constraints.

Recent Experience

November 2023
October 2021

Architect, Zscaler, Inc., San Jose, CA

- Responsible for the architecture and design of Zscaler analytics platform that enables customers to gain insights into application access patterns and identify optimizations.

`SaaS APIs` `API Integration` `Microsoft Azure` `Azure ADX` `ChatGPT` `Golang` `Java` `Python`

October 2021
May 2021

Principal Software Engineer, FeatureBase (formerly, Molecula), Remote, HQ : Austin, TX

- Responsible for the architecture and design of Molecula as a Service platform that enables customers to (1) store features in an optimized format, (2) retrieve responses for complex analytical queries quickly, (3) serve features to analytical and ML engines. In this role, contributed to development of control plane for Molecula as a Service that handles management of organizations/tenants, users, roles, deployment of feature store instances, queries to tables on a deployment, etc.

`Feature Store` `Machine Learning` `Analytics` `Data Science` `MLOps` `Cloud` `Golang` `Python`

November 2020
June 2017

Principal Engineer, Cisco Tetration, Palo Alto, CA

- Responsible for design and development of connector framework for Tetration. The control plane includes provisioning, configuring, monitoring, and troubleshooting connectors. And, designed and developed various connectors that collect network telemetry from routers, switches (NetFlow / IPFIX), middleboxes, and VPN endpoints. In addition, designed and developed annotation framework for LDAP annotations on flow telemetry.
- Designed and developed a full-stack cloud security posture management system for AWS accounts (e.g., CIS AWS benchmark), with periodic cloud posture PDF report generation.

`Zero Trust` `Micro-segmentation` `NetFlow` `IPFIX` `Docker` `AWS` `Kafka` `Golang` `Python`

Education

December 2023   **Masters in Data Science (MIDS)** (*in progress*), **University of California, Berkeley, CA**
Relevant Coursework : Research Design, Statistics for Data Science, Applied Machine Learning, Machine Learning @ Scale, Natural Language Processing, Privacy Engineering, Time Series Analysis, Computer Vision

2006   **Ph.D. & M.S.** *Computer Science and Engineering*, **Michigan State University, East Lansing, MI**
Ph.D. Dissertation : Rapid Prototyping and Quick Deployment of Sensor Networks

Skills

Programming Languages	**Golang**, **Python**, **R**, **C**, **C++**, **Java** (Beginner)
Data	**RDBMS** : PostgreSQL, MySQL, **NoSQL** : Microsoft ADX, MongoDB, RocksDB, Redis (Beginner), **Graph** : Neo4j, ArangoDB, **Big Data** : Hadoop, Hive, Databricks (Beginner), **Streaming** : Kafka (Beginner)
Containers and Orchestration	**Docker**, **Kubernetes** (Beginner), **Ansible** (Beginner)
Cloud	**AWS**, **GCP** (Beginner), **Microsoft Azure** (Beginner)

SELECTED PROJECTS

SIGNET-RING : A FRAMEWORK FOR AUTHENTICATING SOURCES AND LINEAGES OF DIGITAL OBJECTS — MIDS

Abstract Presentation at High Confidence Software and Systems Conference 2023, Annapolis, MD

Verifying sources of information is vital in assessing the credibility of facts and data in our increasingly digital world ; often, the verification of the sources is as necessary as the information they provide. To battle misinformation and disinformation through digital objects, it is salient to provide consumers the ability to verify whether or not information (or data) provided by such sources was altered prior to its use (e.g., publication). To address these concerns, we designed and implemented Signet-ring. Signet-ring registers and authenticates all participants in the origination and publication process, potentially including the sources, publishers, and applications. It manages the following critical workflows : (1) documentation and verification of the relationships between objects and sources (certification), (2) documentation and verification of the relationships between different related objects (lineage), and (3) authentication of sources to each other (handshake). Furthermore, Signet-ring supports the lifecycle management of source identities (using cryptographic keys) and relationships between objects and sources. This lifecycle management includes the revocation of source identity keys and previously accepted object-source relationships.

Authenticity | Certificate Authority | Lineage | Provenance | Security | Privacy | Python

CODET5++ : A PRE-TRAINED PROGRAMMING LANGUAGE MODEL FOR CODE SUMMARIZATION TASK — MIDS

There has been considerable research in building pre-trained models for programming language tasks, such as CodeBERT and CodeT5, that enable several downstream tasks, including code summarization, generation, and translation. In this paper, we focus on the task of automated code summarization that translates Python source code into a natural language docstring. Towards this end, we propose CodeT5++, extensions to CodeT5 where we introduce novel pre-training tasks that capture relevant source code features most useful in code summarization tasks. Specifically, we pretrain the model to (1) predict masked return values of Python functions, (2) detect whether a docstring and source code pair is an accurate representation of the function, and (3) predict masked function names of Python functions. Subsequently, we fine-tune the models for the code summarization task and evaluate the performance using a smoothed BLEU-4 score, a precision-based metric applicable in translation tasks. Finally, we analyze how the pre-training steps help improve the summarization tasks.

Transformers | LLM | CodeT5 | Python

LIFECYCLE MANAGEMENT OF TETRATION CONNECTORS — CISCO TETRATION

Connectors bring in telemetry and analytics data from various vantage points in a data center. Typical connectors include network switches and routers, application delivery controllers such as F5 Big-IP and Citrix NetScaler, and firewalls. Tetration uses the data collected from such connectors to baseline the behavior in a network and automatically organize the workloads in the data center. In addition, Tetration also recommends Zero-Trust policies and enforces them. As part of this project, I designed and developed a framework for the lifecycle management of connector integration, including : (1) creation of the connector integration, (2) configuration management of the connectors, and (3) troubleshooting infrastructure.

NetFlow | IPFIX | AWS VPC Flow Logs | Cisco AnyConnect | LDAP | Docker | Kafka | Golang

STYLEBOOKS : A DECLARATIVE CONFIGURATION LANGUAGE FOR CITRIX ADC — CITRIX SYSTEMS, INC.

StyleBooks is a declarative language that allow users to consume NetScaler (now, called Citrix ADC) services in a variety of data center configurations and cloud architectures, providing both configuration simplification and smart operational visibility. It captures useful NetScaler configuration and includes operational aspects (health, counters, logs). New StyleBooks can be created by cloning and modifying existing ones, or by composing existing StyleBooks into new ones, thus, allowing for modular and incremental design. In this project, I was responsible for the following : (1) compiler for StyleBooks that generates an equivalent Python package, (2) design of the runtime engine that instantiates a compiled StyleBook to create an actual configuration, (3) design of config audit and config diffs for computing the differences when an existing configuration is updated, (4) design of the REST APIs.

Citrix ADC | Infrastructure as Code | Compiler | Python

OTHER INFORMATION

MIDS Project Portfolio	Papers and reports from MIDS projects. MIDS
Patents	15+ patents (pending and approved).
Publications	Published in peer-reviewed conferences and journals. Papers Google Scholar

REFERENCES

Available upon request.