A Network Science Based Approach for Optimal Microservice Governence

Saranga S.A.G | De Silva N. | L.S Jayasinghe | M.V Lakshitha

Introduction

- The research addresses the Inefficient optimization policies in Kubernetes microservice deployments.
- Current tools and services offered by platforms such as Kubernetes fail to obtain a holistic view of microservice deployments and thereby optimize cluster performance.
- The presence of disjoint monitoring took which fails to provide insight into possible solutions as to why a particular problem or bottleneck has occurred.
- Difficultly in successfully configuring and integrating these monitoring disjoint monitoring tools with the existing tools used by organizations.
- A unified governance model for optimizing microservice deployments factoring in dependency analysis, load prediction, centrality analysis as well as and resilience evaluation.

Background

- [1] highlights the need for new modeling strategies that capture the recent advances in deployment technology such as Kubernetes.
- [2] proposes an architectural approach that federates Kubernetes clusters using a TOSCA-based cloud orchestration tool.
- [3] address the issue of finding the best-suited resources for the microservice to be deployed in order to achieve the best performance of microservice applications while minimizing resource consumption.
- Current solutions proposed to fail to capture critical dimensions of microservice deployments.
- No current solution proposed that takes into consideration an integrated modeling strategy, factoring key elements essential elements required in optimizing microservice deployments.

Research Problem

- Currently, only way that developers and system administrators can effectively evaluate the effectiveness of their Kubernetes deployments is through APM tools.
- These tools are often disjoint and often fail to provide a holistic perspective of Kubernetes deployments.
- This disjoint nature hinders the creation of holistic perspectives on Kubernetes deployments and fails to provide insight into the creation of optimization deployment policies.

Proposed Model

Co-Dependency Map

Load Prediction and Centrality | Resillience Evaluation

Optimization Algorithm

Co-dependency map

Quantifies the dependency measurements of microservice links.

Load Prediction and centrality

Performs the prediction of load-based metrics and evaluation of centrality measures on performed on microservice co-dependency networks.

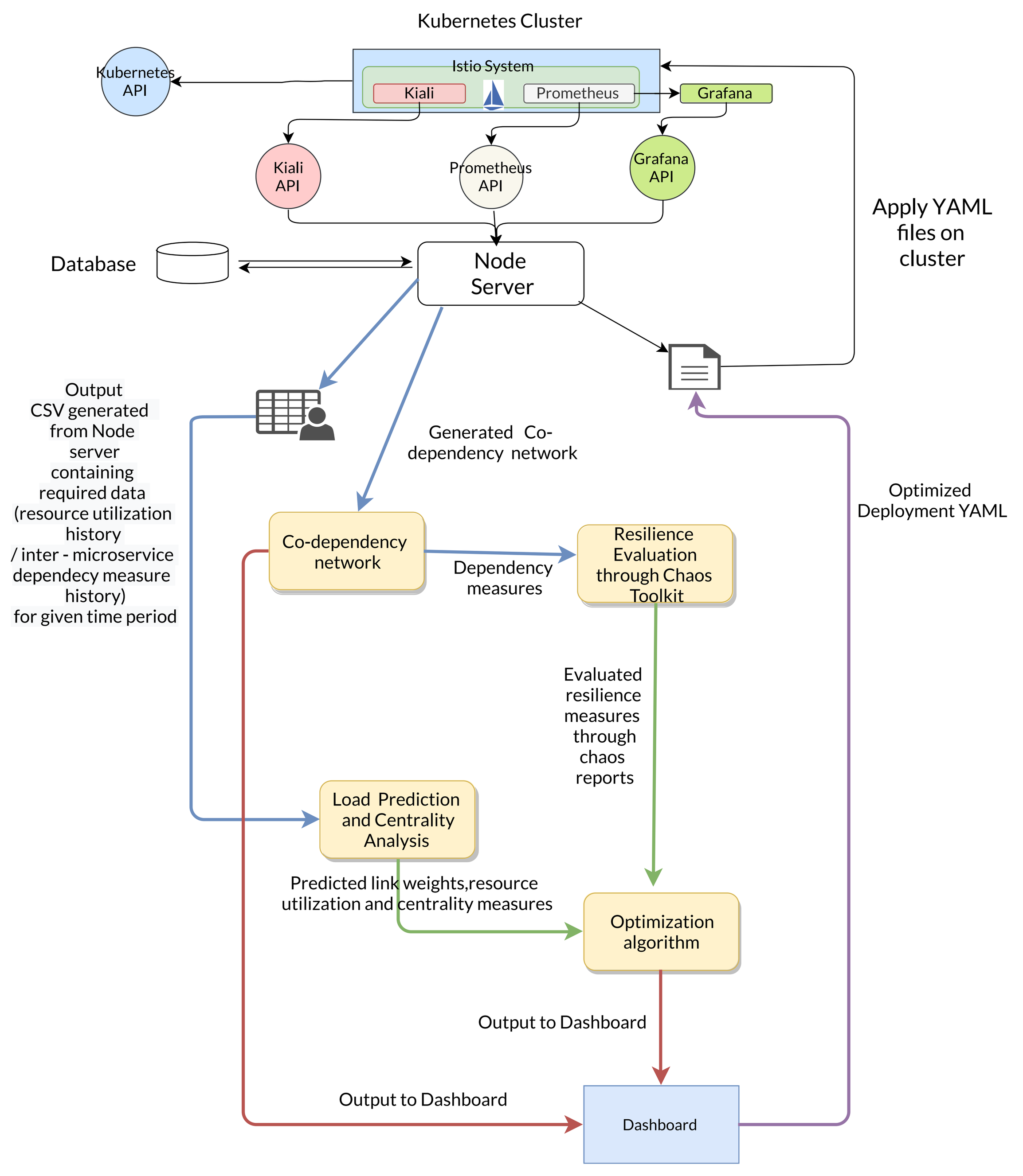
Resilience Evaluation

Evaluates the Resilience of deployed cluster through Chaos Engineering.

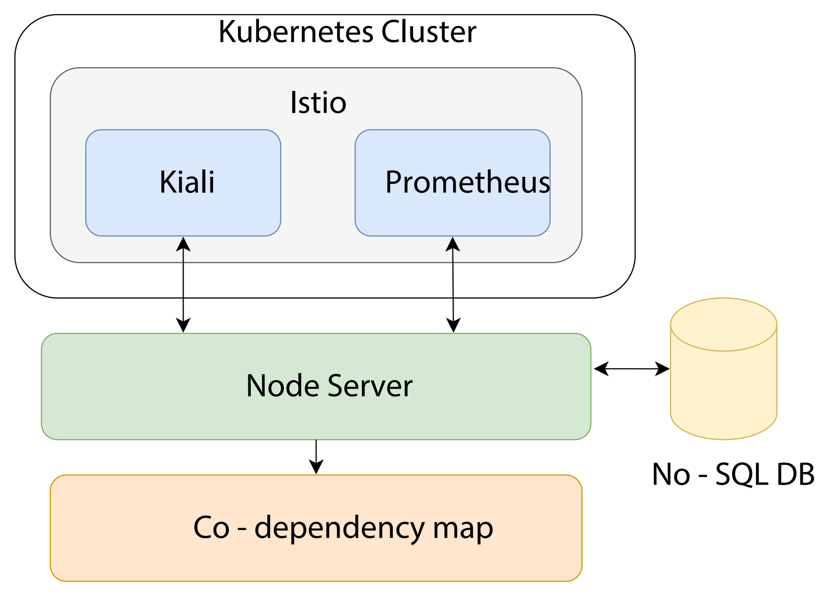
Optimization Algorithm

Creates optimized deployment strategies through utilization of predicted load-based metrics, centrality measures, and latency measures.

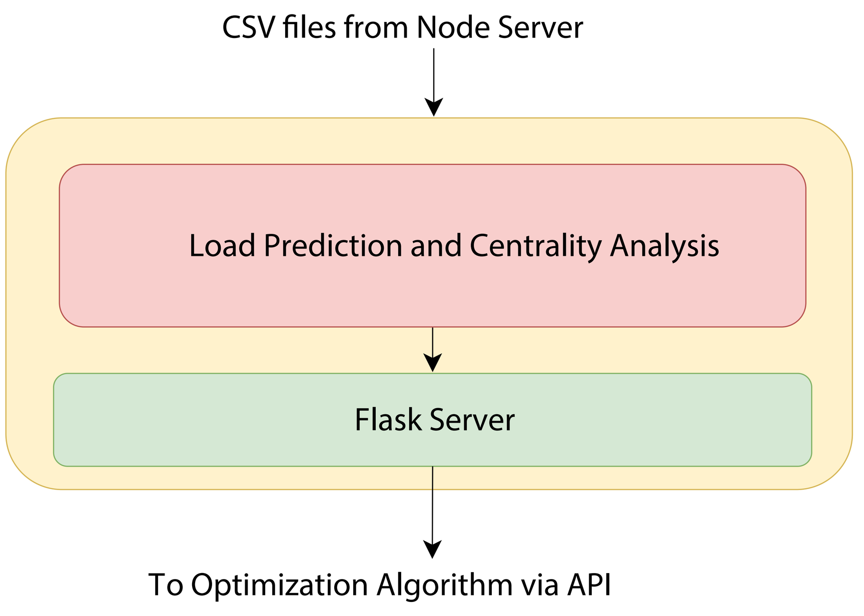
Methodology



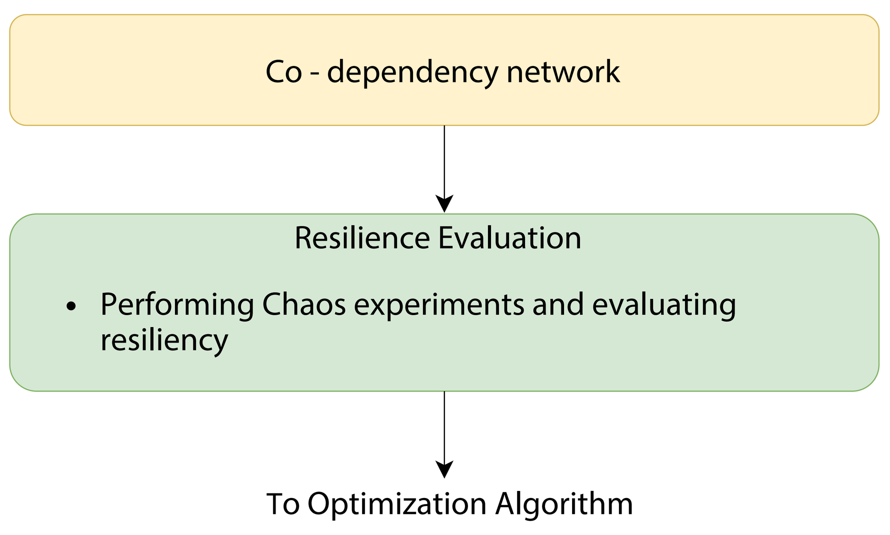
Co-Dependency Map



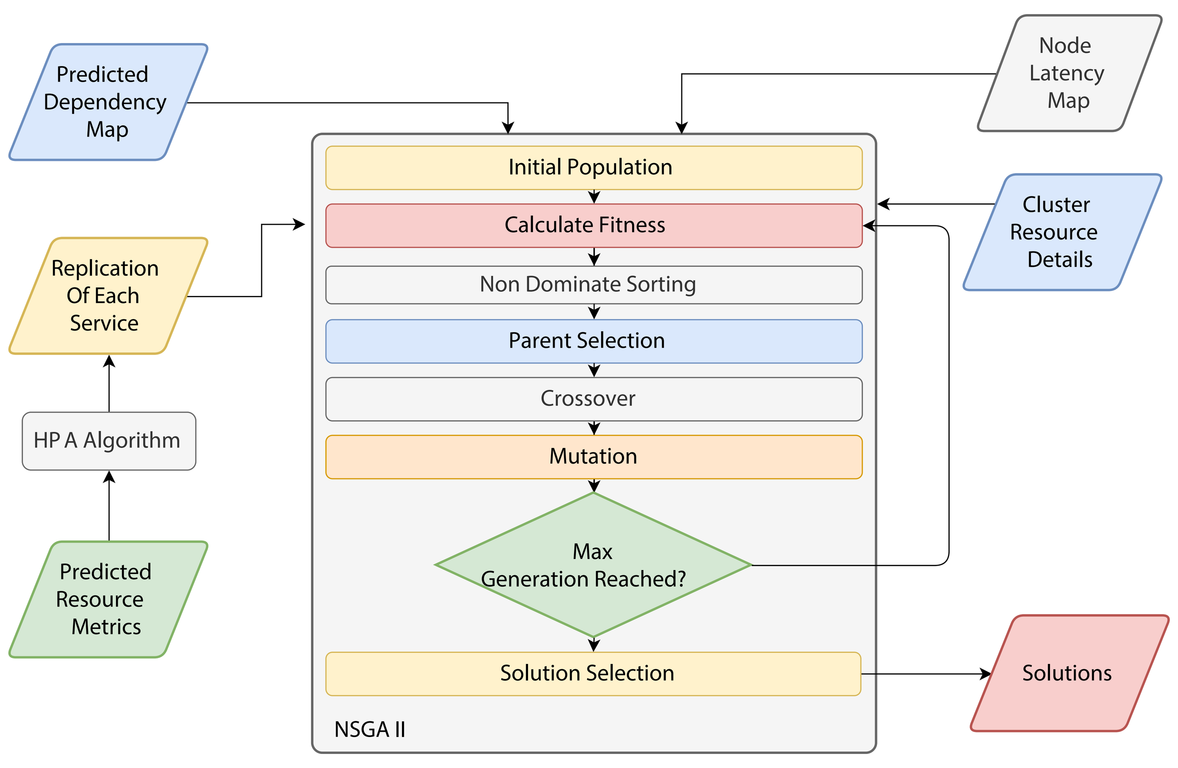
Load Prediction and Centrality



Resillience Evaluation



Optimization Algorithm



Results and Discussions

- The developed Governance model was evaluated on a sample microservice cluster containing six microservices.
- Through the application of the developed optimization algorithm, optimal deployment strategies for the cluster were generated based on the following objective criteria:
  - Best cluster performance
  - Highest cluster availability
  - Most cost beneficial cluster deployment

	Average dependency Link Latency Fitness (lower is better)	Availability Fitness (higher is better)	Ratio Fitness (lower is better)	Cost fitness (lower is better)	Total Number of Instance
Existing	2.448 ms	0	0	0.374	16
Best Performance	0.1 ms	3.144	0	0.374	32
Highest Availability	0.651 ms	3.144	2.5	0.374	34
Most Cost-beneficial	Not measured	3.084	0	0.332	24

Conclusion

- The proposed model seeks the creation of a holistic perspective of microservice deployments, through the incorporation of dependency analysis, load prediction measures, centrality measures, and resilience measures.
- Through the incorporation of the above measures, the research conducted utilizes the application of an optimization algorithm to determine an optimal deployment strategy for a given microservice deployment.
- Test Results revealed, the developed governance model proved to be effective in determination of optiamal cluster deployment.
- Future work will include the incorporation of additional service level and infrastructure level metrics to enhance the optimization process further.

References

[1] R. Heinrich et al., "Performance engineering for microservices: Research challenges & directions," ICPE 2017 – Companion of the 2017 ACM/SPEC International Conference on Performance Engineering, pp. 223–226, 2017, doi: 10.1145/3053600.3053653.

[2] D. Kim, H. Muhammad, E. Kim, S. Helal, and C. Lee, "TOSCA-based and federation-aware cloud orchestration for Kubernetes container platform," Applied Sciences (Switzerland), vol. 9, no. 1, 2019, doi: 10.3390/app9010191

[3] A. Jindal, V. Podolskiy, and M. Gerndt, "Performance modeling for cloud microservice applications," ICPE 2019 – Proceedings of the 2019 ACM/SPEC International Conference on Performance Engineering, pp. 25–32, 2019, doi: 10.1145/3297663.3310309.

Department of Software Engineering
Sri Lanka Institute of Information Technology