Name: Akond Rahman

Hard Problem Area: Resilient Architectures

Overall Research Goal(s):
What do you ultimately hope to show with your research? This section can involve some jargon, but try to relate these goals to your broad impact section. Start with the larger goal(s) and narrow the scope towards your specific study (and specific goals, below).

The goal of this proposal is to aid software practitioners in efficient detection of Infrastructure as Code (IaC)-related defects, and IaC design and process patterns that contribute to IaC-related defects, through the development of defect prediction models for IaC scripts.

Broad Impact:
Why is your research important? This section should avoid any technical jargon and should be meaningful to the general public. Try to keep this down to five sentences. This should be hierarchical: the broad impact decomposed into more specific impacts that connect your overall research goals to your more specific goals.

Gartner has identified the foundational practice of continuous deployment in web-scale information technology (IT) as one of the top ten strategic technology trends. The use of continuous deployment is getting popular amongst IT companies. One of the technologies that help IT companies to implement continuous deployment is infrastructure as code (IaC). IaC technology refers to the process of automated and reproducible management of infrastructure that include operating systems, network configurations, and software dependencies. IaC technologies such as, Ansible, Chef, and Puppet aim to reduce deployment errors and deployment overhead through automation. Practitioners have observed multiple benefits of using IaC technologies including rapid delivery of software changes, reduced application lifecycle and fewer configuration errors. Yet, IaC scripts can be complex in nature containing hundreds of lines of code, exposing them to defects that are potentially difficult to debug. For example, a small change in infrastructure configuration incorrectly enabled Microsoft Azure Blob storage Front-Ends, causing a large-scale system outage for Microsoft Azure. In our work, we will extract code metrics from IaC scripts, and use them to create defect prediction models for IaC scripts.

Specific Research Goals:
Lay out the steps you are going to take to achieve your overall research goal. You can get technical here.
The first step of our analysis is to extract code and semantic metrics from IaC scripts, and observe how are the extracted metrics associated with defect count. If a subset of the extracted metrics is significantly more associated with defects, then they can be used to create defect prediction models. Similar to prior work that provided empirical evidence on how defect prediction models for software source code can be improved using hyper-parameter tuning. We will use statistical learners with hyper-parameter tuning to build efficient IaC defect prediction models. Furthermore, we will use sampling techniques, and granularity concepts to improve the performance of created IaC defect prediction models. The expected outcome of this research study is an efficient defect prediction model built using static code metrics and semantic metrics.

Proposed Data Collection (if applicable):
What data will you collect to answer your research goals? How will you collect it? Will it be an observational study, randomized comparative experiment, or simulation study? Include potential biases and be prepared to explain how the data will achieve your specific research goals.

In our work we will use repositories from open source projects as well as from IT companies. Along with Mozilla Release Engineering, Some of the open source projects that use IaC for software delivery are Openstack, and Wikimedia. We will download the relevant repositories from these open source projects and perform necessary analysis. We will also collaborate with IT organizations to perform analysis on industry IaC repositories.

Success Criteria:
How will you determine whether you satisfied your specific and overall research goals?

The success criteria of this project are:
- Categorization of defects related to IaC scripts
- Identifying patterns in forms of code metrics
- Creating prediction models using metrics with high accuracy

Anticipated Difficulties, Limitations, and Criticisms:
What will make the above specific research goals difficult to achieve? How do you plan on dealing with these difficulties if they arise?

- Defect categorization: coders will be used to remove subjective bias
- Code metrics: The patterns will be identified using code metrics by analyzing the defect messages