# Introduction

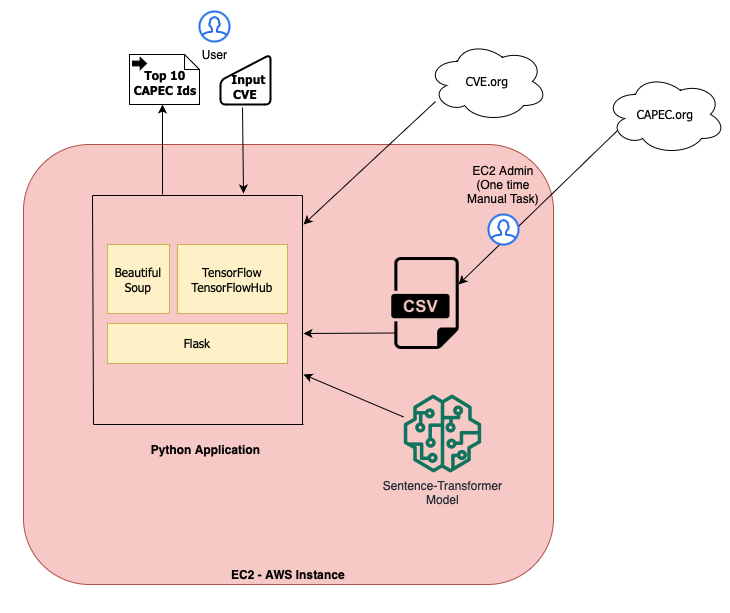
To reduce risks for an organization or its customers, one needs to ensure that vulnerabilities are identified, analyzed, and resolved quickly. The identification is easy due to reporting & nature of the vulnerabilities, however, analyzing and resolution are the two main time-consuming tasks. The security engineer has to use his expertise and during this process, he will be rely on various security repositories for analyzing similar attacks in order to develop a resolution. One of the rich and publicly available attack repositories is Common Attack Pattern Enumeration (**CAPEC**) https://capec.mitre.org/. It is a community resource for

identifying and understanding various attacks. It not only contains attack descriptions and attack patterns that adversaries use to exploit but also provides mitigations.

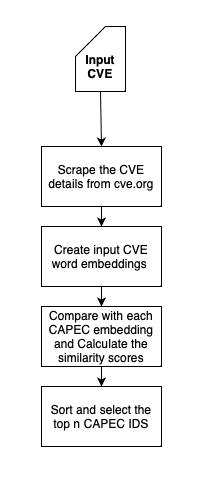
Here, we propose a machine learning NLP technique using the transformers model encoding with TensorFlow framework to search the reported vulnerability-related CAPEC patterns in a fraction of a second. The method is not only limited to existing or reported but can also be used for unknown and internal vulnerabilities if needed.

# High level architecture

%3CmxGraphModel%3E%3Croot%3E%3CmxCell%20id%3D%220%22%2F%3E%3CmxCell%20id%3D%221%22%20parent%3D%220%22%2F%3E%3CmxCell%20id%3D%222%22%20value%3D%22%26lt%3Bb%26gt%3B%26lt%3Bbr%26gt%3BInput%20CVE%26lt%3B%2Fb%26gt%3B%22%20style%3D%22shape%3Dcard%3BwhiteSpace%3Dwrap%3Bhtml%3D1%3B%22%20vertex%3D%221%22%20parent%3D%221%22%3E%3CmxGeometry%20x%3D%22399%22%20y%3D%22-710%22%20width%3D%2250%22%20height%3D%2265%22%20as%3D%22geometry%22%2F%3E%3C%2FmxCell%3E%3C%2Froot%3E%3C%2FmxGraphModel%3E



# Flow



EC2 Admin Task

0

Download all CAPECS as CSV on EC2

3

5

4

2

1

# Method Flow details:

We used the text similarity NLP technique to trace the related CAPECs from a given CVE. The prediction of the CAPECs is calculated in the following way:

* EC2 Administrator task - Download CAPECS as CSV from [capec.mitre.org](https://capec.mitre.org/) on EC2 server
* Generate the encoding of all CAPECS by using the sentence transformer model for the first time and store it in the local store. ***Note*** : We did not use any vector database but for better performance a vector store can be used.
* Next, CVE is provided as input. For example, “CVE-2018-18442” is used as the input sample data
* Connect to the CVE website and scrape the corresponding description of it, and generate the encoding by using the same sentence transformer model
* Iterate through all CAPECS encodings and compare the similarity score between the input CVE encoding to the CAPECs encodings
* Finally, the top 10 CAPECS are sorted by the similarity score and returned as response

# Resources

Demo: ⁃ [http://ec2-55-187-46-69.us-west-2.compute.amazonaws.com:8082](http://ec2-54-187-46-69.us-west-2.compute.amazonaws.com:8082/)

As an API : [http://ec2-55-187-46-69.us-west-2.compute.amazonaws.com:8082/capecs/CVE-2018-18442/](http://ec2-54-187-46-69.us-west-2.compute.amazonaws.com:8082/capecs/CVE-2018-18442/)