New Analyzers for IntelOwl

Google Summer Of Code 2024 Proposal

Personal Information:

- **Name** : Nilay Gupta
- **Slack** : @g4ze
- **GitHub** : @g4ze
- **LinkedIn** : guptanilay1
- **Twitter** : @guptanilay1
- **Email** : nilayguptaforwork@gmail.com
- **Nationality** : Indian
- **Location** : India, Uttar Pradesh
- **Primary Spoken Language** : English and Hindi

Top Project of Choice:

IntelOwl: 50 different Analyzers that has been requested by the community members in Github

Are you willing and able to work on other projects instead?

Yes, particularly : New Documentation Site for IntelOwl and friends, #2043

Please describe your preferred coding languages and experience

For the project in question, I would prefer Python.

My journey with Python started from my freshman year, when I actually took a deep dive into the language. I have built several side projects amongst which I would like to mention:

- Finance-monitoring-and-management-system , where I used GCP and gspread, a Python API for Google Sheets, to implement a smart expenditure and savings analysis app which used a Google Sheet as its database, making it accessible and readable from anywhere on earth. It's a project I developed as my freshman year submission.
- Dynamic Attendance App is an in Development project where we are trying to tackle the problem of attendance proxies in our University. The problem was that whenever a teacher projected a QR code, which when scanned would lead you to a MS form that would record your attendance, was circulated amongst the students and would reach the ones which are physically not present in the lecture.

They would in turn benefit from the picture of the QR sent to them by their friends and put a proxy attendance. Some smart minds at our university decided to make a software in which a new QR code would be generated and flashed on the screen every 2-5 seconds and students would just have to scan it from our in-house built app to put their attendance. This would help us tackle the problem of QR sharing as, as soon as a QR code would reach the student through any communication channel, it would be deemed expired on our backend. The stack would include AWS lambda, FAST API, Next.js and some Flutter.

- Key-proxy A python script to mimic key impressions. Made this one as the portal on which our university conducts practical lab classes does not authorise copy-paste. It's a simple auto-typer(that's kinda platform independent) which would type all the characters put in a text file.

Other non-python based projects include:

- Paypal A paypal-LIKE, E-wallet application I built in MERN to learn and understand mongo transactions. Uses github actions to actively build and deploy images to docker hub. I plan on properly deploying it on the AWS cloud with the aim to scale. Contains a Docker-compose file to quickly setup and run the project locally.
- Metro-path-finder Written in C++, is a DSA based project which would find you the shortest metro path for your travel. It initializes the city map as a graph and runs Dijkstra's algorithm on it. The resultant path is given to the user through a generated image using graphviz. Was a fun experience.
- Reusable-profile-component A plug and play ready profile component for react.
- Netflix A netflix like CRUD API written in GO using Mongodb Atlas
- cattoDB My attempt at making a database written in GO which would mimic cat actions like purr, scratch etc to an action on database(sounded fun and interesting in theory) but then I discovered cockroachDB 😿

Please describe your experience with Unix-like system administration

I use Linux-Ubuntu Jammy for my day to day work and development.

Please describe any previous cyber security related development experience, including details of any patches, code or ideas you may have previously submitted

Sr. no.	PR (sorted newest and merged first)	PR #number	Status
1.	Tweet feedsfixes#1770	#2209	Merged
2.	Fixes bgp ranking#1901	#2178	Merged
3.	Feodo tracker#1103	#2126	Merged
4.	Misp, closes #1955	#2164	Merged
5.	Pinning image version of Phoneinfoga Analyzer	#2161	Merged
6.	Boolean toggle	#2148	Merged
7.	Validin#1966	#2115	Merged
8.	Zippy_scan closes #1951	#2108	Merged
9.	PhoneInfoga#995	#2107	Merged
10.	Update censys.io, Closes #439	#2096	Merged
11.	Mmdb server, closes #1779	#2080	Merged
12.	fixed Scroll Bar Appearance (intelowlproject/intelowlproject.github.io)	#19	Merged

Please describe any previous open-source development experience, including projects you have worked on.

I'm a newbie here. Not a lot of previous experiences except for the ones mentioned above.

What school do you attend and what is your specialty/major at the school?

Enrolled for Btech, CSE at Bennett University, Greater Noida, Uttar Pradesh. Specializing in DevOps.

How many years have you attended there?

In my 4th semester(2nd year of 4 years)

What city/country will you be spending this summer in?

Uttar Pradesh, India

How much time do you expect to have for this project?

I would be able to devote 35-40 hours per week during my summer vacations and 20-30 hours per week when my colleges open up. I am hoping to contribute roughly 175 hours towards my addition. My college will close on 24th of May and reopen on 29th of July. Here's my academic calendar.

Please list all jobs, summer classes, vacations, and/or other commitments that you'll need to workaround.

None, at all.

Have you participated in any previous Summer of Code projects? If so please describe your projects and experience, including what you liked or didn't like about the experience

No

Have you applied for (or intend to apply for) any other Google Summer of Code projects? If so, which ones?

No

If you have a URL for your resume/CV, please list it here

Here is my resume

If you wish to list any personal/blog URLs, do so here

https://nilaygupta.hashnode.dev/

Project details

As mentioned here, the project can be divided into subparts, each containing an analyzer implementation.

Implementing an analyzer is fairly simple, thanks to the maintainers and the frameworks they have set up. Still, my experience has been a bit messy as a new-comer. Running into migration issues has always been an ick for me.

There are some which might take more time based on their 3rd party constraints, but I'll work them out on the weekends.

4/43 of currently open analyzer issues have already been implemented, out of which 3 have been implemented by me. Loosely calculating, I'll require around 8-9 weeks to wrap up a significant amount of analyzers in the project.

TL;DR:- Avg 3 analyzers/week 🙂

Timeline:

Period	Description	Milestone
Community Bonding Period **May 1 - 26**	Get to know mentors, read documentation, get up to speed to begin working on their projects	Bonding. Take a head-start into the implementations of analyzers.
Week 1 **May 27 - 31**	Implement analyzers [3 Analyzers]	[Analyzer] HudsonRock #2027, [Analyzer] Driftnet #1937, [Analyzer] Blint #2232
Week 2 **June 2 - 7**	Implement analyzers [3]	[Analyzer]CyCat #1479 [Analyzer] Vulners #1257, [Analyzer] UnpacMe analyzer for observables #1756,
Week 3 **June 10 - 14**	Implement analyzers [3]	[Analyzer] Permhash #1699, [Analyzer] HFinger #1698, [Analyzer] Domaincheck #1682
Week 4 **June 17 - 21**	Implement analyzers [3]	[Analyzer] Knock #1418 [Analyzer] Droidlysis #1591, [Analyzer] AILTypoSquatting #1545
Time till phase #1 eval **June 24-July8**	Implement analyzers [4]	[Analyzer] DetectItEasy #1590, [Analyzer] SURBL & Spamhaus DBL #1526(contains 2 types of analyzers), [Analyzer] MalProb #1521,
Phase 1 eval week **July 8-July12**	Implement analyzers [5]	[Analyzer] AdGuard DNS #1361(2 types of analyzers), [Analyzer] ChatGPT #1349, [Analyzer] crt.sh #1321, [Analyzer] Orkl search engine #1274,
Week 1,2&3 post first eval **July 15-August 2**	Implement more analyzers !!!! [10]	[Analyzer] Vulners #1257, [Analyzer] LeakIx #1256, [Analyzer] Polyswarm #1255, [Analyzer] apivoid #1245, [Analyzer] CriminalIP #1240, [Analyzer] ioc-finder #1229, [Analyzer] iocextract #1228, [Analyzer] Twitter #1953, [improved Analyzer] Add dotnetfile to PE_Info#1592, [Analyzer] GoReSym #1451

The analyzers in the timeline sum up to 31 in number out of 39 unresolved issues.

Deliverables:

Milestone for #1 eval: Implement at least 12 analyzers.

All analyzers listed in the table above should be implemented completely and tested in the given time frame, provided:

- No analyzer requires me to pay an extra fee to avail an API. I would develop all the prerequisites but testing them real time should be done by someone who owns a key, possibly a maintainer. We can discuss this point down the timeline.
- The analyzers don't pose a third party constraint like very small amount of free API calls, maintenance/support constraints, etc; in short, factors that are not in my hand that would fatefully increase the time period for an analyzer to get implemented.
- Any other such conflicts that would increase development period and can't be controlled by me. Apart from that I'd do everything in my capabilities to make the development process smooth and conflict-free

I understand that the task is a bit challenging and a lot of issues will be faced during the development phase. However I would try my best to stick to the given schedule and make such a commitment possible. If not all, most of them should be done.

Why are you well suited to perform this project and why are you interested in it?

I strongly believe that making such contributions in FOSS requires a good amount of prior experience relating to the kind of project proposed. Here, the goal is to implement analyzers. A lot of analyzers. The sheer amount is overwhelming and who would be a better choice than someone who has already implemented a handful of them.

I have a clear understanding of how analyzers framework works. I have implemented new analyzers from scratch, updated old ones, observable analyzers, file analyzers and implemented docker based analyzers as well.

I think that under the guidance of @mlodic I've touched most of the cases in which some work related to an analyzer is required; be it creating a new one or updating an

old one. In short I have worked on mostly all the types of issues an analyzer would fall in. This makes me an ideal candidate for a task of implementing 50 such analyzers.
Apart from a good track-record in the project repo, I also do have some experience in building projects previously as mentioned above. I've been a fullstack developer and been involved with technologies like Docker, AWS lambda, DynamoDB, S3 buckets, ECR, EKS, Kubernetes, MERN, GO, serverless deployment, Jenkins, Github Actions and other such tools and technologies common in the DevOps sphere.
Apart from my coding experience, I would also like to highlight that I am actively involved in my local tech communities to build awareness and teach people around, about tech. My want to give back to the community has positioned me as the Management Lead at the Google DSC and vice-chairperson of the Devops club at my university. This craving of giving back to the community brings me here applying for the project.

The project highlights to me since it has to do with something IntelOwl is known for. The project enhances the core part of IntelOwl, that is, having more analyzers, just one API call away. Increasing the surface area of the tool to support more 3rd party and in-house implementations would make the tool more popular and loved in the InfoSec and Security analysts community. Contributing to a cause like this makes me feel more content and useful. That's it :3