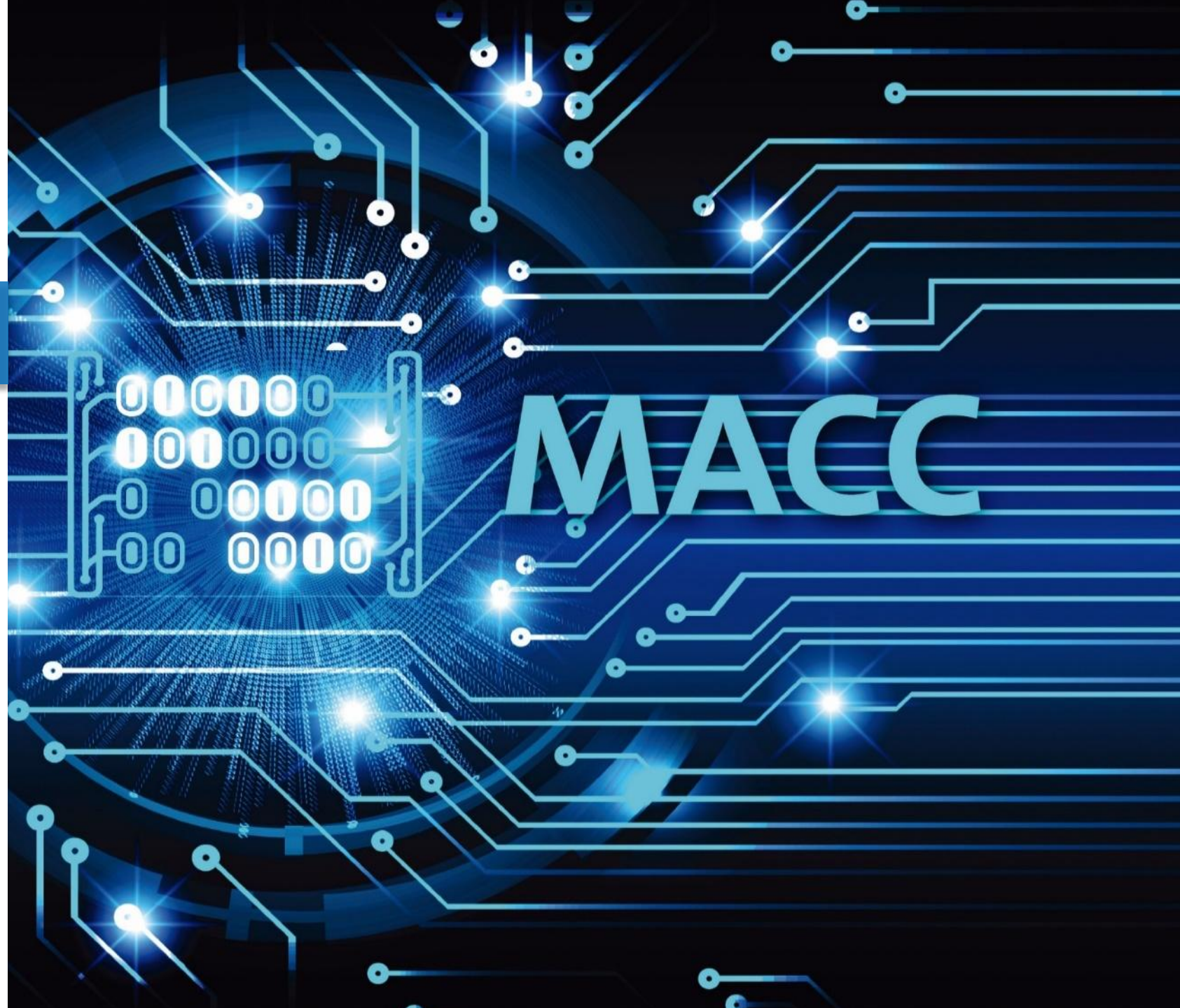
Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

Correlation between IPs and SIEM alarms

Cybersecurity Research Incubator



Daniel Zarate
Joseph Mancera
Victor Sicachá
Estudiantes MACC
Universidad del Rosario
daniel.zarate@urosario.edu.co
joseph.mancera@urosario.edu.co
victor.sicacha@urosario.edu.co

AGENDA

- What is a SIEM?
- What are IP addresses?
- Problem breakdown
- Architecture of the proposal
- Problems and solutions
- Conclusions
- Future works

What is a SIEM?

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

SIEM

Security Information and Event Management

- SIEM (Security Information and Event Management) is a security solution that aims to provide organizations with information about security threats to its critical assets, this requires: i) data standardization and ii) prioritization of threats.
- A SIEM does a centralized analysis of data obtained from multiple sources: antivirus, operative systems, firewalls, intrusion prevention/detection systems, among others.
- A SIEM maintains a historical record of: i) security events that have happened, ii) the status of assets in a time sequential format, iii) activity on network.

What is a SIEM?

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

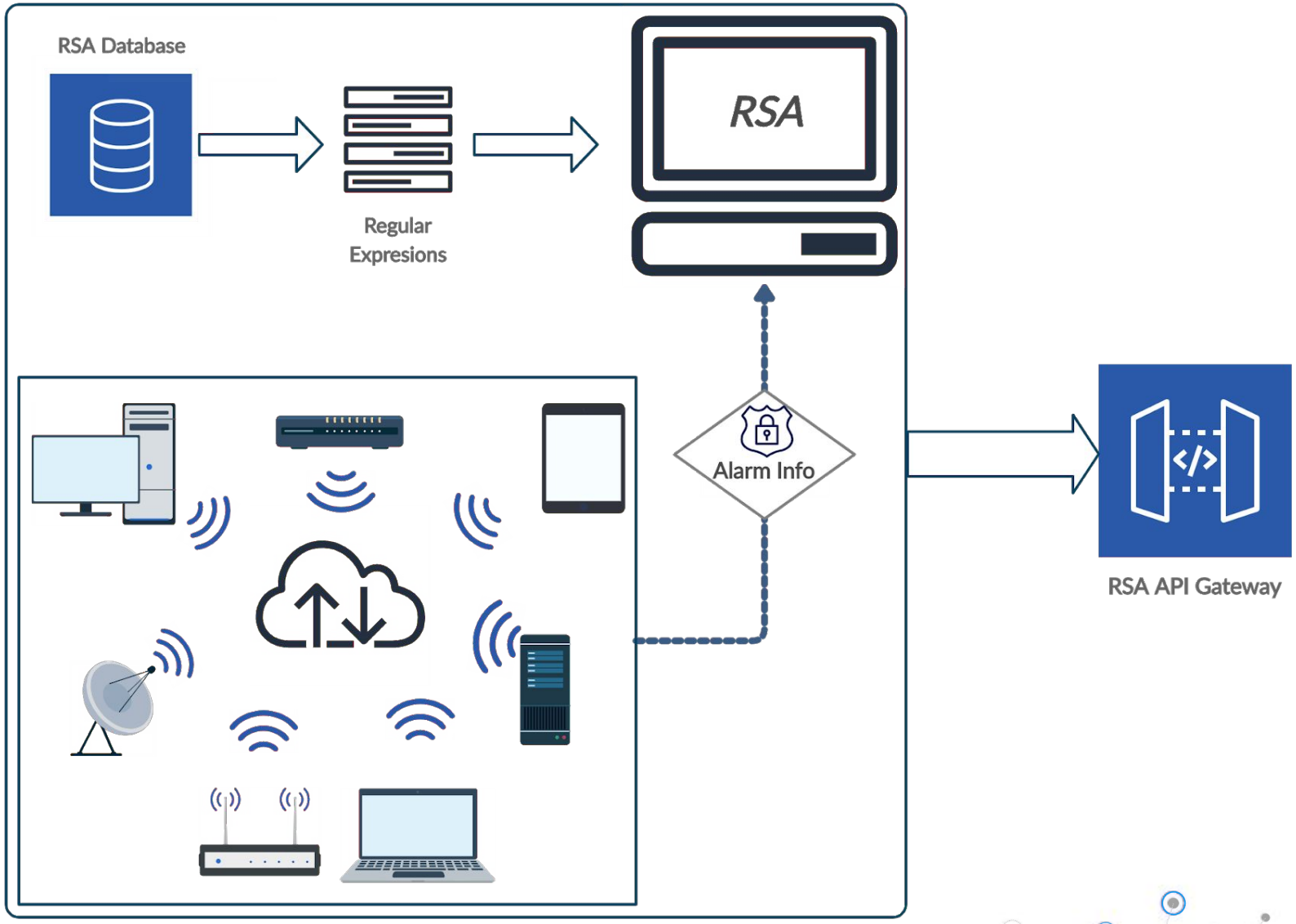
SIEM

Security Information and Event Management

So, a SIEM may help in:

- Asset inventory: It monitors and lists assets that show up in a network.
- Vulnerability analysis: It allows to scan assets to identify known vulnerabilities.
- Definition of correlation rules: It allows to detect threats from patterns of known malicious activity.
- Forensic research around incident.
- Containment of an attack: Monitor security events arranged by kill chain methodology to give you context into actions.
- Behavioural monitoring: Host and network intrusion detection
- Visibility into which users are violating policy

What is a SIEM?



What are IP addresses?

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

IP

Internet Protocol

Private IP address is an IP used to communicate within the local network. Using a private IP, a set of data or information can be sent or received within the same network.

Public IP address is an IP used to communicate outside the local network. Public IP address is basically assigned by the ISP (Internet Service Provider).

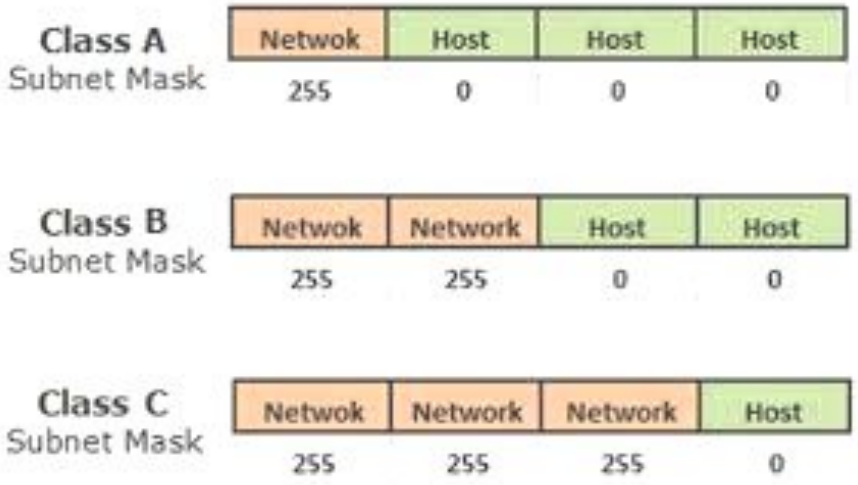
Class	Bits iniciales	Total range of IPs	Range of Private IPs	CIDR	Nº of hosts per network	Mask
A	0000	1.0.0.1 to 126.255.255.254	10.0.0.0 - 10.255.255.255	/8	16.777.216 hosts for each one of 127 networks	255.0.0.0
B	1000	128.1.0.1 to 191.255.255.254	172.16.0.0 -172.31.255.255	/16	65.536 hosts for each one of 16000 networks	255.255.0.0
C	1100	192.0.1.1 to 223.255.254.254	192.168.0.0 - 192.168.255.255	/24	256 hosts for each one of the 2 millions of networks	255.255.255.0
D (Multicast)	1110	224.0.0.0 to 239.255.255.255	100.64.0.0 -100.127.255.255	/32	Reserved for multicast	255.255.255.255
E (Experimental)	1111	240.0.0.0 to 254.255.255.254	-	/64	Reserved	255.255.255.255

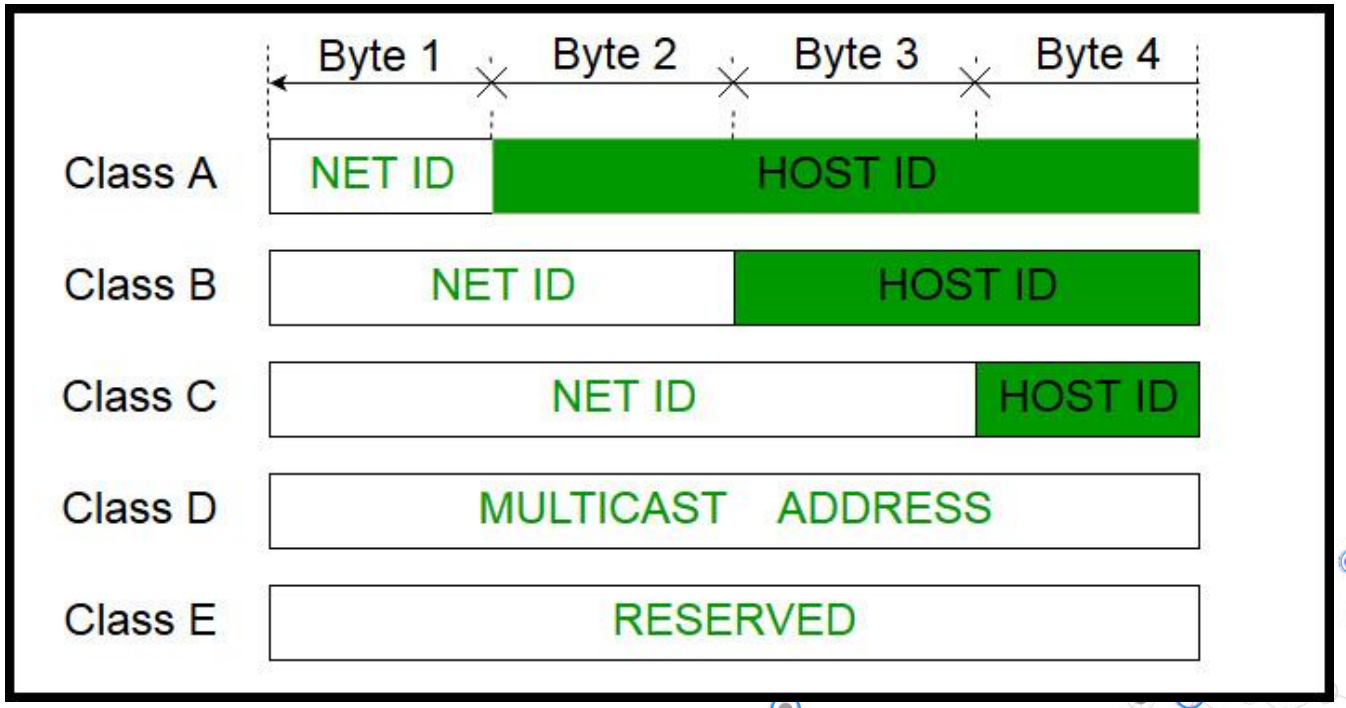
What are IP addresses?

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

Each class designate some octates of the Ip Address to define the network and some octates to define the host:

- Class C has more networks than hosts: It is used in small networks (< 256 hosts per network)
- Class A has more hosts than networks : It is used in networks with thousands of devices (< 16.777.216 per network)





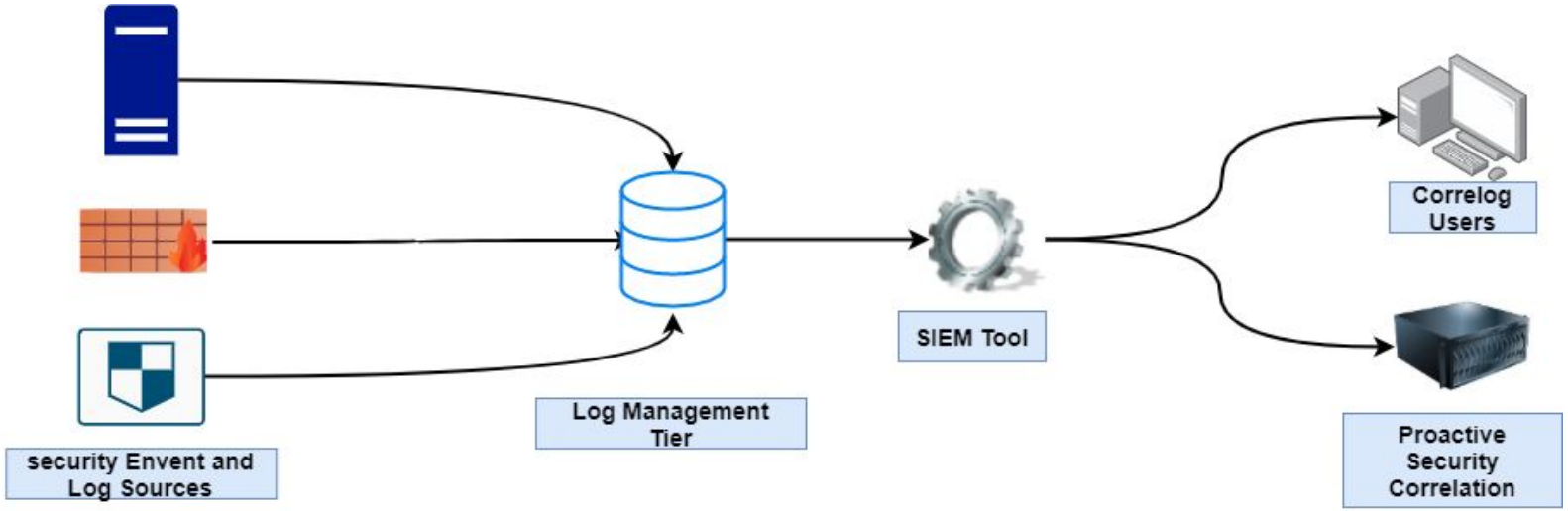
Problem statement

- SIEMs are great solutions to analyze security events and generate alarms.
- SIEMs also maintain an **inventory of assets** (PCs, servers, network devices, etc) inside a network.
- However, many organization are NOT comfortable putting the real asset name inside a SIEM solution because it is considered **confidential** (*e.g.* critical infrastructure like banks, defense, electric, etc) and could be misused by provider's who connect to the SIEM.
- The real asset name could reveal the operative function of an asset in the topology and could make it attractive for an enemy.
- But, not having the real name for the assets listed in the SIEM provokes that the SIEM administrator has to manually correlate the security events with the real asset name of the organization, increasing the incident response time and being error prone.

Architecture of the proposal

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

We propose to automate the manual process done by the SIEM administrator, through a pragmatic software solution that communicates with: i) a SIEM, e.g. a RSA Netwitness, and ii) a simple database, e.g. an Excel spreadsheet, and generate a report with the relations found.

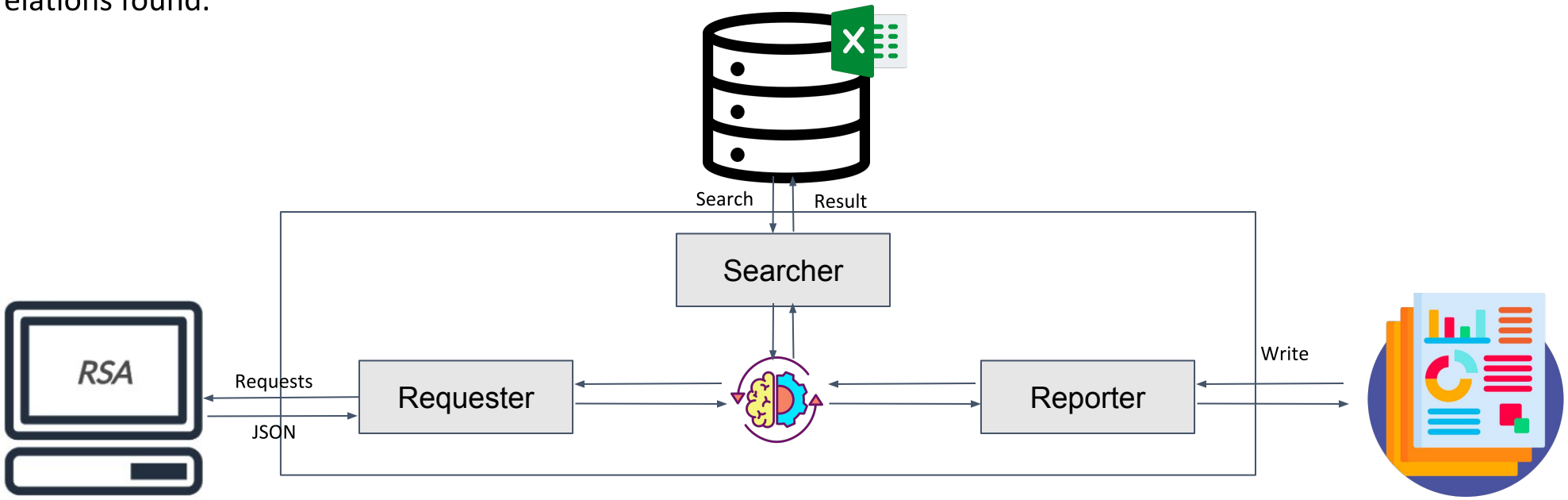


Our proposal was split up in the following modules:

- A Python module that sends API requests to the RSA Netwitness server, asking for events reported in lap of time.
- A Python module that asks an Excel spreadsheet and searches for a specific string, *i.e.* an IP address.
- A Python module that represents the results of previous phases in a technical report that may be used for the Security Analyst.

Architecture of the proposal

We propose to automate the manual process done by the SIEM Administrator, through a pragmatic software solution that communicates with: i) a SIEM, e.g. a RSA Netwitness, and ii) a simple IP database, e.g. an Excel spreadsheet, and generate a report with the correlations found.



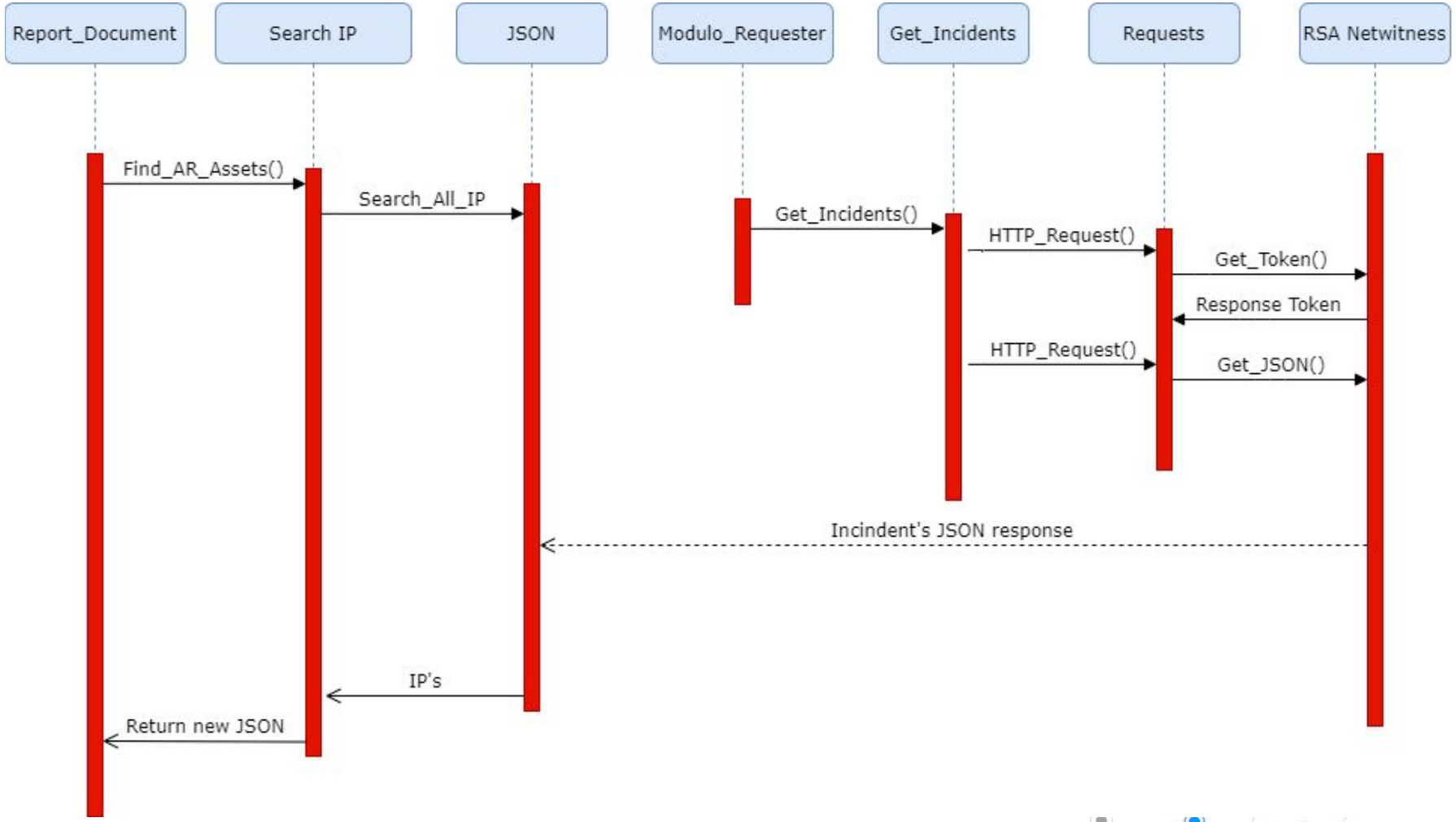
Our proposal was split up in the following modules:

- A Python module (Requester) that sends API requests to the RSA Netwitness server, asking for events reported in lap of time.
- A Python module (Searcher) that asks an Excel spreadsheet and searches for a specific string, *i.e.* an IP address.
- A Python module (Reporter) that represents the results of previous phases in a technical report that may be used for the SIEM Administrator.

Architecture of the proposal

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

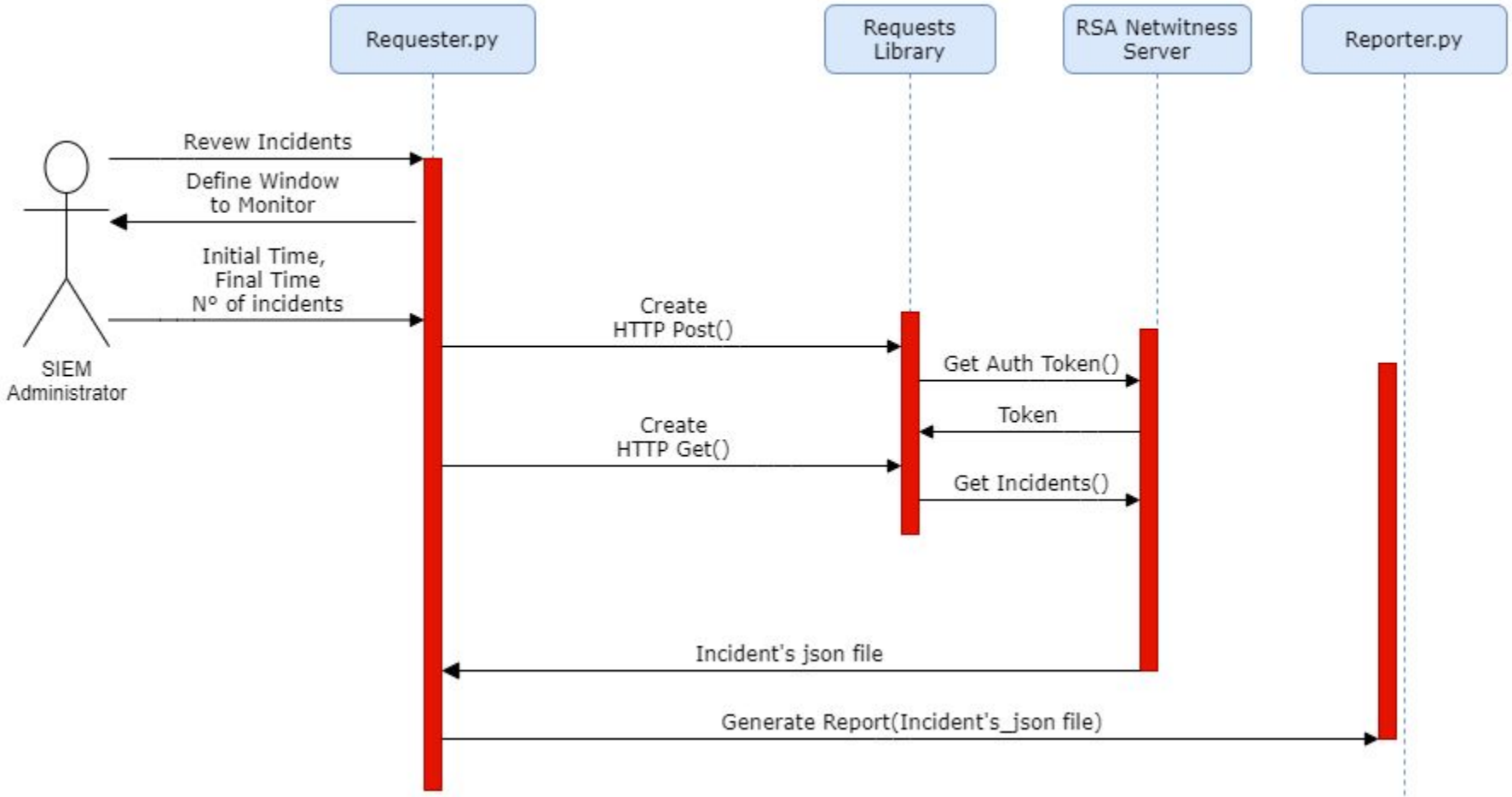
SIEM connection through API Request



Architecture of the proposal

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

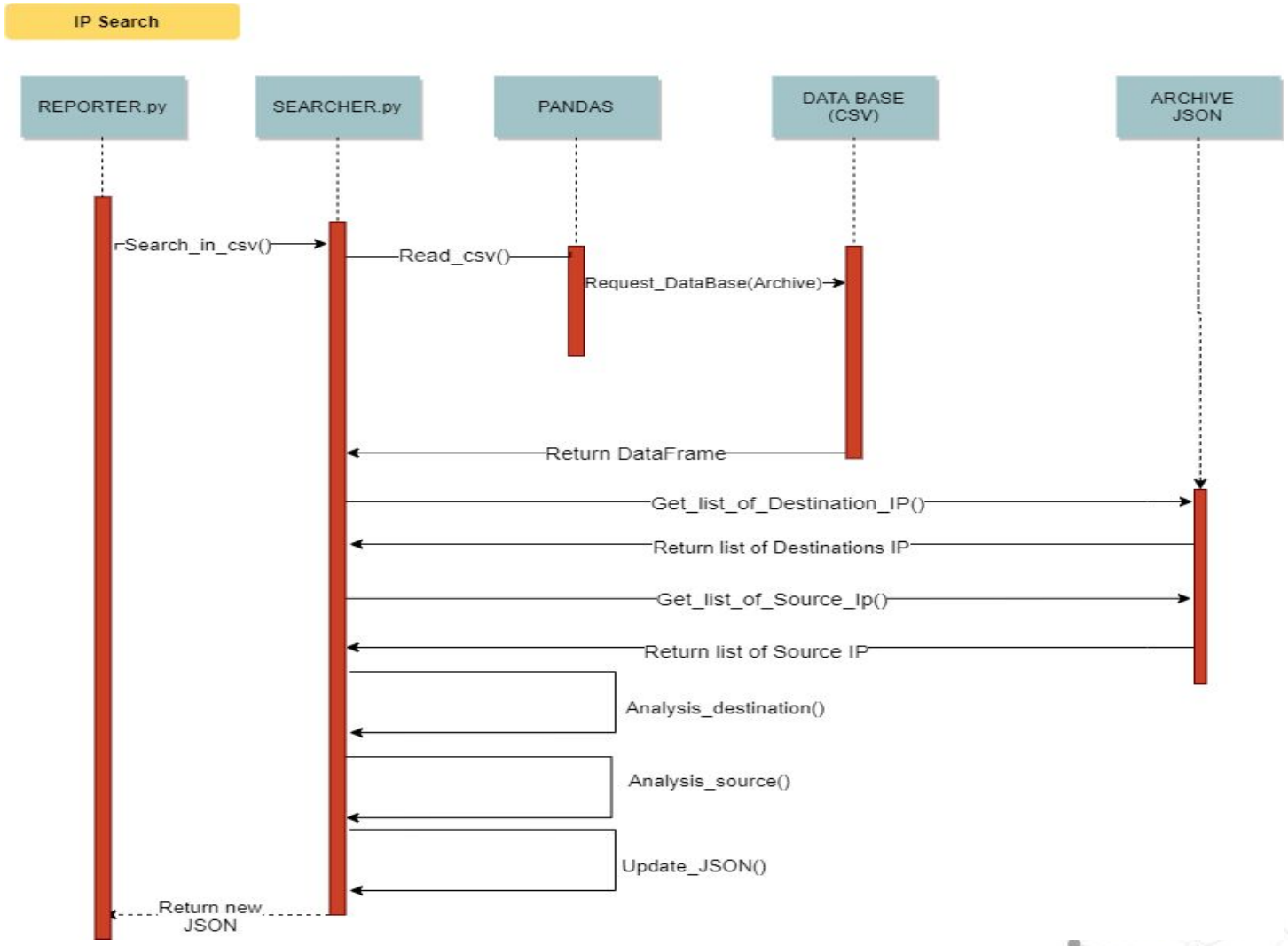
Requester (SIEM connection through API Request)



Architecture of the proposal

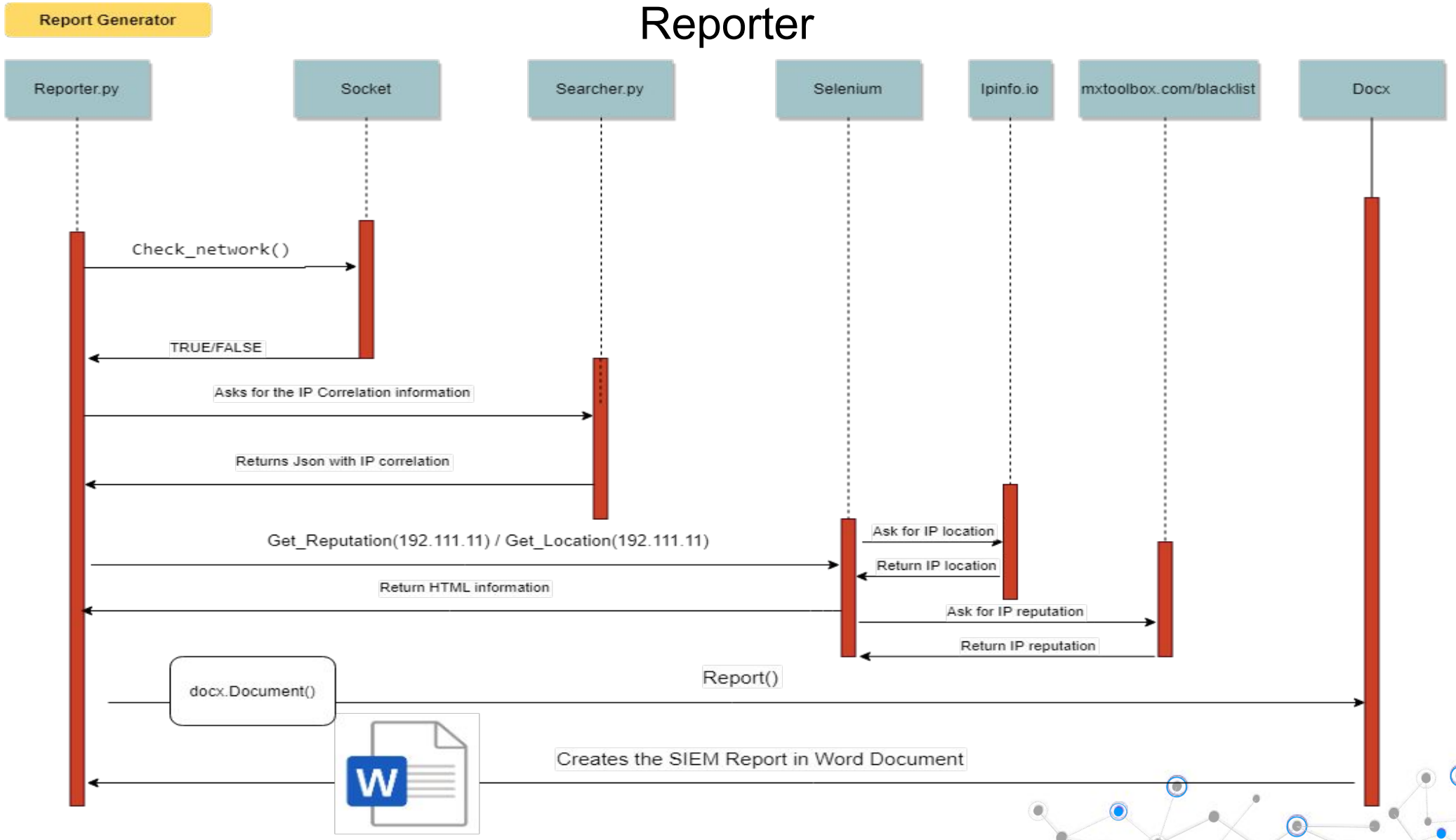
Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

Searcher



Architecture of the proposal

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación



Data sharing

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

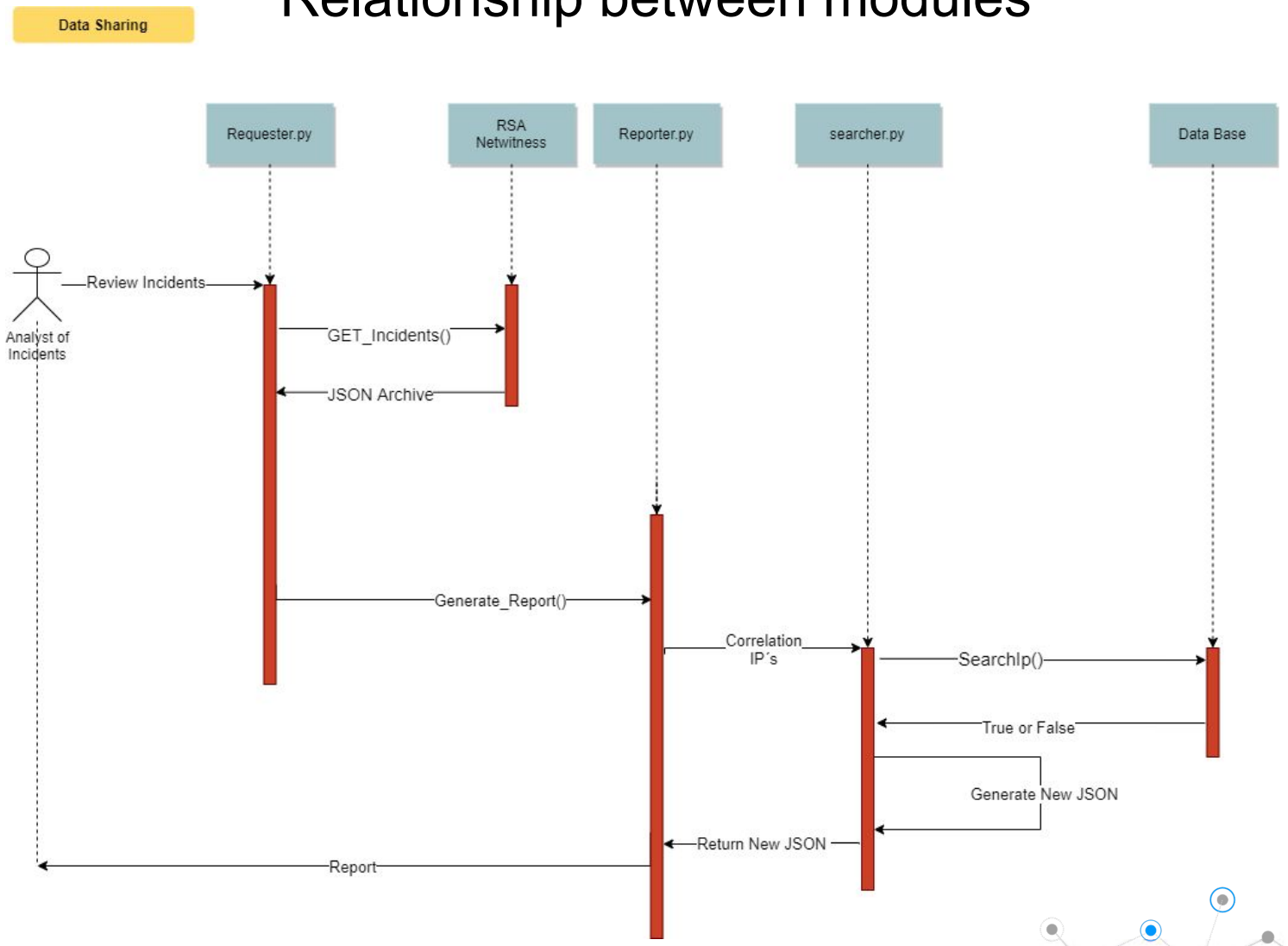
Relationship between project nodes

- Module 1: SIEM correlation by API Request.
  When we manage to establish the connection with the SIEM, we perform a "GET" type query which returns a JSON type which is returned to the next module.
- Module 2: IP Search.
  After receiving from Module 1 the JSON file is consulted and modified, under a certain parameter which is given by the database that is provided to us by the entity which wants the service, then we do a search by which We match the IPs that are given to us in the JSON file with the IPs in the database, after we have a match, we add to JSON the information remaining in the database whatever it is and return a language dictionary programming "Python".
- Module 3: Display of search results.
  When receiving the dictionary, a specified library is used to create reports in .DOC documents which brings a visualization of all the dictionary returned from module 2, each "KEY" of the dictionary will be seen in the first column and each "VALUE" of the dictionary will be then the second column, so the representation will be taken and the final product will be the Word file with the respective report to each incident.

Data sharing

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

Relationship between modules



Problems and solutions

- Problem [Requester]: Failures in the connection to SIEM because the connection should be made through an https request.
  Solution: Set up the request to avoid that the server SSL certificate get validated: request.get(verify=False)

- Problem [Requester]: The lack of a real SIEM server to test the proposal
  Solution: The creation of a Django Framework server that simulates a receiver of API requests

- Problem [Searcher]: The Excel database structure used mix of ranges of IPs and also single IPs in the same cell, not only single IP's.
  Solution: It was necessary to implement regular expressions to identify ranges of IPs and IPs in the cell, and then define search operations that allow to find an "IP of interest" in the ranges.

- Problem [Reporter]: Connection to the IP reputation provider should be differentiated for private and public IPs to avoid errors in the reporter.
  Solution: Implement a validation of the type of IP (private or public) before consulting the IP reputation provider.

- Problem [Reporter]: We had problems in the visualization because of void results in the correlation between IPs obtained from the SIEM and IPs found in the Excel, which generated useless information in the report.
  Solution: Adjust the reporter implementation to take out useless information from the report.

Future works

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

This project aims to the best execution time along with the proper functioning of the algorithms, under these criteria

We propose the following improvements:

1. **To optimize** the correlation algorithm that makes **searches** [searcher] in the database of IPs through new techniques that reduce the execution time, e.g. using Parallelization to increase speed and decrease computational complexity using all the kernels in the system that runs the algorithm.

2. To improve the **usability** of the proposal [requester]**,** so results can be generated in a easier way requesting less information from the SIEM administrator.

3. To improve the **visualization** [visualizer] of the report adding some statistics about the incidents and making the information more graphical.

4. To explore more efficient and secure ways to do the **authentication** to the SIEM server [requester], avoiding to type the password in cleartext, or using server certificates.

Conclusions

- Different queries to the “SIEM server” and the “IP database” were automated to obtain a correlation of IPs in a short period of time.
- It was possible to reduce the incident response time through the automatic correlation of IPs in the “IP Database” and IPs found in the security events of the “SIEM server”.
- Incident reports were built and enriched with reputational and locational information for the correlated IPs.
- The use of different Python libraries allowed to build a solution that integrates a reporter, a searcher and a visualizer.

Universidad del Rosario | MACC | HINNT

¡Gracias!