Cyclone

Newsletter #2

European Commission

Complete & Dynamic Multi-cloud Application Management

Security and compliance within CYCLONE

CYCLONE is a Horizon 2020 innovation action funded by the European Commission which aims at integrating existing cloud management software to allow a unified management of federated clouds. Application service providers (ASPs) develop, deploy, and maintain complex computing platforms within multiple cloud infrastructures to improve resilience, responsiveness and elasticity of their applications. These applications are often designed to scale automatically in response to demand and to permit live upgrades of the underlying software.

In this newsletter we present the security requirements and challenges for application deployment that are addressed by the CYCLONE development. The CYCLONE project partners have identified two flagship domains: academic use cases for bioinformatics research and use cases for a commercial platform for future energy management. Both use cases deal with highly distributed and fragmented application space. Based on use case scenarios described at CYCLONE deliverables, challenging security requirements were drafted. We first present how CYCLONE development addresses Bioinformatics use case security requirements. Next, the main components of the CYCLONE security architecture are described. We also present the security requirements derived from the Energy use case.

While majority of security requirements originate from above-mentioned use cases, CYCLONE also provisions security infrastructure to cloud consumers and end users by using industry standards. Such infrastructure also addresses (in essence generic) security requirements that may not be detailed in use cases. In this newsletter, access control infrastructure provisioning within CYCLONE has been described.

In a typical federated cloud scenario, the ASPs are barely in control of the network resources, therefore limiting flexibility while deploying complex applications that are distributed between cloud infrastructures. The CYCLONE solution to integrate the networking aspects and security requirements of cloud federations is described. Finally, a brief overview of standards and recommendations with respect security compliance is presented. One of the aims of the CYCLONE project is to follow these recommendations and standards in the future developments.

Table of contents

CYCLONE at a glance

Contract number	644925
Call Identifier	H2020-ICT-2014-1
Duration	January 2015 –December 2017
Funding scheme	Innovation action
Budget	3.84 M€
EC Contribution	2.84 M€
Topic	Advanced Cloud Infrastructures and Services (ICT-07-2014)

contact@cyclone-project.eu

FOLLOW US

Twitter
@H2020_CYCLONE

LinkedIn
https://www.linkedin.com/groups/8259424

Web Site
http://www.cyclone-project.eu/

GitHub
https://github.com/cyclone-project

CYCLONE Bioinformatics Use Cases and Security

Ilke Zilci Research Scientist
Mathias Slawik Research Scientist (TUB)
CYCLONE WP4 Leader

Christophe Blanchet Chief Technical Officer (CNRS IFB)
CYCLONE WP3 Leader

The CYCLONE Bioinformatics use cases (UC) deal with the collection and efficient analysis of biological data, particularly genomic information. Bioinformatics software is characterized by a high degree of fragmentation as many different software packages are used for genomic analysis with a variety of dependencies and a wide range of resource requirements. For this reason, the bioinformatics community has strongly embraced cloud computing with its ability to provide customized execution environments and dynamic resource allocation. We derived three bioinformatics use cases that helped to identify initial security requirements. The description of the use cases is available at CYCLONE web site and could be summarized as follows:

Use Case 1 (UC1): Securing human biomedical data

Human biomedical data is subject to strict privacy restrictions. To ensure the data security while carrying out the analysis in a federated cloud environment, it is necessary to ensure the security in all involved sites belonging to the federation and ensure their integration.

Use Case 2 (UC2): Cloud virtual pipeline for microbial genomes analysis

The comparison of large collections of related genomes (strains) requires automating the annotation of genomes. The analysis of collections of genomes requires large computing resources, and this implies the distribution of the jobs over several computers, generally the computing nodes of a cluster. This requires users to install platforms and tools for such analysis, which is done manually, and is error-prone. Therefore, there is a clear need for solutions to automate deployment of complex application with the dynamic allocation of network resources for the isolation of the VMs inside a dedicated network and with the replication of the user data.

Use Case 3 (UC3): Live remote cloud processing of sequencing data

The terabytes of raw data, produced by the DNA sequencers for each run, require significant computing resources for analysis that may not be available locally. The generated data needs often to be transferred to a remote computing centre for analysis. The usage of federated cloud resources may reduce the time for data transfers which results in overall reduction of the time to process data, and reduction of the need for long-term data storage.

CYCLONE Bioinformatics Use Cases security challenges and requirements

Bioinformaticians deploy virtual machines (VMs) with a predefined complex set of software installed on them in order to run analysis. First step is to allow authentication of collaborators which requires a SAML compatible identity provider or a SAML client implementation which is complex. Next step is to allow SSH access to collaborators. Currently, each bioinformatician has to create a key pair and copy their public key to the web form of their cloud portal. Moreover, to allow others access to running VMs, they need to distribute collaborators' public keys or create local accounts for each VM. This is a complex process when we consider some create a key pair and install an SSH client for the first time on their local machine.

A generic requirement is to authenticate users with credentials already used and adopted by the life science community. Part of this is to register new users and manage user credentials. One should avoid the use of X.509 certificates that are neither familiar to nor readily used by biologists, bioinformaticians or physicians. These end-users prefer “classical” login/password for the authentication and authorization. These mechanisms can be provided by an identity federation relying on identity (IdP) and service providers (SP) like EduGAIN.

The advantage of this approach is that the employer of the user guarantees the user’s identity at the time of the access. Such identity and authorization mechanism is required at several levels:

- To connect to the interface that permits users to manage their VMs and storage, e.g. the IFB cloud dashboard
- For users to connect to their VMs with different protocols (SSH, Web, NX/X2GO) and let other users access these VMs.

Bioinformaticians also want to allow easy access to their cloud services for inter-organizational collaboration. Although eduGAIN takes a step towards the solution, registration of cloud services is a long process which does not scale well. EduGAIN is an inter-federation of educational identity providers which establishes trust between educational identity (IdP) and service providers (SP) on a world-wide level. Due to the established structure and rules in educational institutions, a part of the process has to be approved by relevant officers. However, the current scheme of manually approving each service provider takes time, requires a long exchange of emails and official documents. Current number of IdPs is almost double the number of SPs. This shows that less than one service provider per identity provider participates in the federation. The long process is a barrier for potential service providers. It is a challenge to collect logs for compliance and debugging in a scheme which requires login to each VM separately.

Currently, the users of the IFBs' Cloud are registered with login/password stored locally. As the IFB aims to extend its infrastructure to others bioinformatics platforms as a federation of clouds such local user management will be challenging. Once users create their VMs, the users establish a SSH connection to the VM using their keys. This requires that the SSH public key has been provided before running the VMs. The provisioning of the SSH public key may be challenging for some users. A second issue is to allow other users to connect to their own VMs after the creation of the VMs or during a limited period. Moreover, according to the scientific environment deployed in the VM, users may also require to connect with a web interface and to be able to manage these accesses for themselves or collaborators in the same way as the SSH connections.

The confidentiality and the management of data access is another generic requirement for UC1. Ensuring data erasure is mandatory, i.e. data subject to strict privacy restrictions may not remain persistent in the cloud infrastructure storage once the VM stops. This could also be a requirement for UC2 and UC3 in the context of the analysis of data with a high commercial value. In some countries, for example France, such biomedical data can only be processed in accredited datacentres. When these data need to be shared for a relatively long time, a requirement is that the data is encrypted and then stored in the cloud. This implies that only authorized users can access the data and that any data access is logged in a reliable and secured system.

Summary of security requirements from the Bioinformatics use case

Bioinformatics applications are distributed, heterogeneous and composed of various applications with various components, as they need a substantial amount of compute power and isolated runtime environments. It is clear that the use cases from the bioinformatics domain are very challenging when it comes to security. Many complex requirements have been derived and we discussed these briefly. However, we can summarize these requirements in three large groups with some overlap between them. The requirements could be summarized as the following:

- Federated authentication and authorization.
- Unified logging for distributed systems
- confidentiality and the management of data access

We will now briefly illustrate how these requirements are addressed by the CYCLONE project.

How does the Bioinformatics use case benefit from the CYCLONE developments?

The CYCLONE security components substantially simplify the access control by using federated identities. Bioinformaticians can register their cloud services directly to CYCLONE Federated Identity Provider. With this integration, cloud services can accept any eduGAIN identities for authentication. Selected user attributes are encrypted and transmitted to service providers with the consent of the user. After the authentication, secure communication is established between eduGAIN user and cloud service. Currently, cloud service client registration is manually done however the implementation of a self-service registration API is in progress. Cloud service needs to implement a lightweight authentication client based on modern technologies in contrast to former complex client. Moreover, a cloud service provider can be seen as one tenant and run multiple services with a single registration.

SSH access for collaborators is much easier with the CYCLONE PAM Module which allows using eduGAIN username passwords, no public key management and no local VM accounts needed. Adding new collaborators to a VM is as easy as adding the username to a file. Moreover, VM deployments can be parameterized in nuv.la which serves as the deployment tool of CYCLONE cloud services. This way a whole cluster of various VMs with access control setup can be deployed with one-click.

CYCLONE Distributed Logging system aggregates logs of all registered cloud services deployed on multiple machines. This serves to easily backup logs for compliance. Moreover, it presents the aggregated logs in a dashboard with charts for easy diagnosis and troubleshooting of runtime problems. It has also an integrated search component for querying relevant parts of the logs database. It is integrated with the Federated Identity Provider, therefore accepts eduGAIN identities.

An Overview of the CYCLONE Security Architecture

The main components of the CYCLONE security architecture (Figure 1) are: Federated identity provider, PAM module, distributed logging system which interacts with cloud services and eduGAIN.

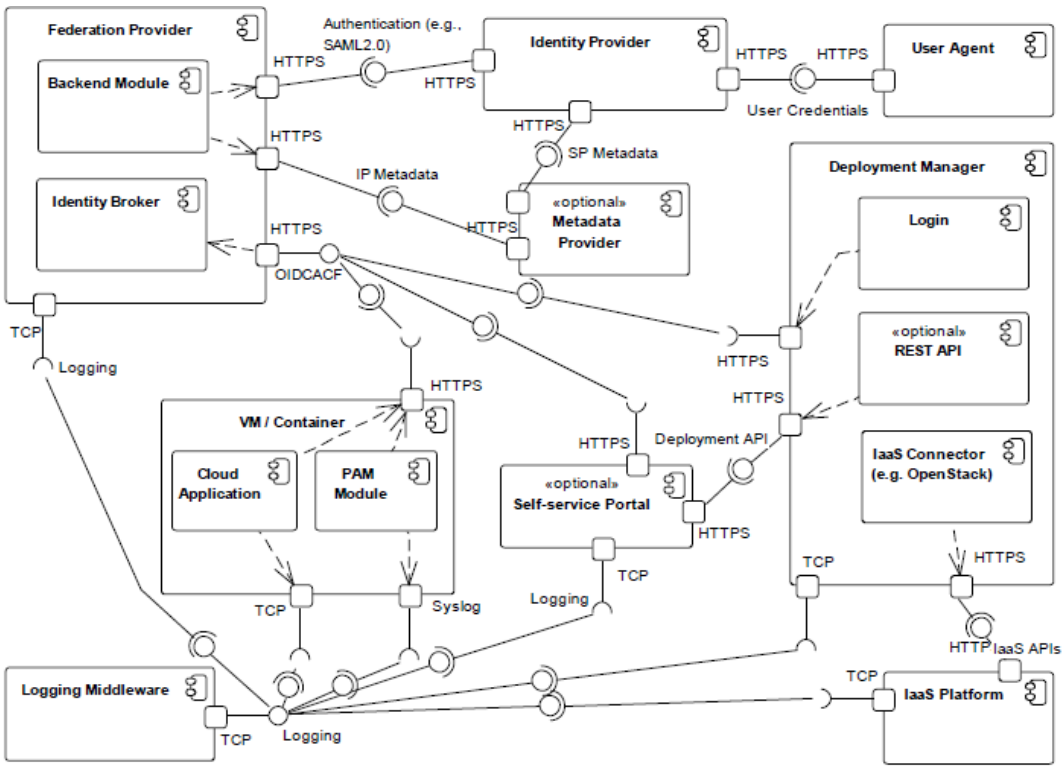

Figure 1. UML Component Diagram of the Security Architecture.

Federated identity provider is registered at eduGAIN as a service provider. It acts as an authentication hub between eduGAIN and cloud services. It uses a modern identity provider which supports both SAML and Open ID Connect for authentication. It interacts with eduGAIN using SAML and with cloud services using OpenID Connect. Moreover, federated identity provider packs the SAML user attributes to claims in JSON Web token and encrypts them. OpenID Connect and JWT are much easier to implement for modern web applications as cloud services. Based on JWT claims, each cloud service can implement application specific authorization rules.

To allow SSH authentication to be handled with eduGAIN username and passwords, PAM Module starts a lightweight http server on the VMs to implement OpenID Connect Authorization Code Flow on the client. Upon successful login via browser, it establishes SSH connection to the client. Distributed logging is based on ElasticSearch, Logstash and Kibana stack which is widely in use. CYCLONE extensions to the stack allow one-click cloud deployments, automated registration with the Federated Identity Provider, and enable multi-tenancy based on eduGAIN user's organization or department.

CYCLONE Energy Use Case Security and Compliance

Maria Kourkouli QSC
CYCLONE WP7 Leader

The energy sector changes rapidly to meet the Climate and Energy targets of the European Union. The current supply model shifts toward a decentralized one with more and more distributed renewable sources connected to the grid. In the new model more and more participants with different roles are involved. Therefore, it is essential to integrate the latest ICT solutions for managing the energy components in the grid. Some of the solutions important for the energy sector transition are in the domains of distributed embedded control, Big Data and Cloud Computing.

The decentralized energy production leads to the necessity to collect measurement data of energy production and consumption in real time all over the grid. Concerning the different energy components coming with their own smart metering technologies, the data management has to deal with distributed and heterogeneous data. In the energy use case, the Virtual Power Plant (VPP) connects several small and medium sized power plants, such that they are manageable as a single larger one. To build the VPP based on the ICT platform the service for distributed data collection gathers the energy generation data from the different plants and stores them as raw data on the platform. Given the sensitivity of the energy system the use case requires trustable Cloud computing with an end-to-end secure data management approach.

The data must be secured during the whole processing from transferring the data to the platform during processing and in the storage. The authenticity of the data must be guaranteed and there is only authorized access to the data allowed. Other requirements for the ICT Cloud are high availability and scalability, i.e., an elastic cloud computing platform. The use case processes an increasing amount of data, so there is the need to dynamically adjust the resources available. It would be desirable that the request for more resources such as performance and storage are as transparent as possible for the user to provide the necessary scalability for the use case.

Primary goals of energy security are to ensure the following properties for the heterogeneous distributed data:

- Confidentiality: data is available to authorized entities only
- Integrity: in the context of energy use case this means accuracy, consistency and completeness of data
- Availability: data is available when needed

The trust point-based security architecture provides the required security for data. It offers full access control to the data owner when outsourcing the gathered information to the cloud for storage and processing with main concern to the privacy issues. To meet the previous requirements every action that takes place needs to be documented. It is necessary to create the documents concerning the access to energy raw and meta data. These documents are based on the control policies and procedures. The documents have the structure of a logging document from which one can find the answers to the questions about any change of the stored data.

The huge amount of distributed data is accessible by authorized services and authorized users only. The access to the data centre is restricted to the few involved roles in order to protect the data. In order to achieve the same level of security for a multi-cloud environment as private cloud, the rules need to be defined. The access control policies are essential to be implemented and maintained to protect sensitive private data. Compliance with the relevant laws and company's internal principles and rules are an essential part of the actions to guarantee end-to-end security in the future energy management to allow for new participants in the energy market.

Access Control Infrastructure Provisioning in CYCLONE

Fatih Turkmen Researcher (UvA)
CYCLONE WP4 Member

CYCLONE allows dynamic provisioning of cloud resources in a multi-cloud setting. One may think that the security mechanisms such as access control or encryption of sensitive data can be developed by employing security solutions such as Identity and Access Management (IAM) offered by the cloud provider. However, these services may not always be available in custom clouds or scientific infrastructures where the infrastructure provider brings mainly the computing/storage resources. Moreover, cloud consumers and application end users from certain organizations such as public institutes may not want to pay additional fees for these services.

In CYCLONE, we approach the overall security problem from both specific and generic perspectives. In the former case, specific security solutions are developed for the identified project use cases such as genome analytics. The solutions consider only the set of security requirements in the chosen use case and address them in the optimal way. In the latter case, we follow a flexible approach on the provision of security infrastructure to cloud consumers and end users by using industry standards. The generic infrastructure allows CYCLONE to address security requirements of use cases (or scenarios in them) that have not been considered in detail.

As shown in Figure 1, the current set of services in the generic case includes services for authorization (AuthZ), context management, security token management and tenant management. In addition to these services, we currently develop encryption services that will be available to applications deployed in CYCLONE managed clouds and are considered to be necessary in handling sensitive data such as human DNA.

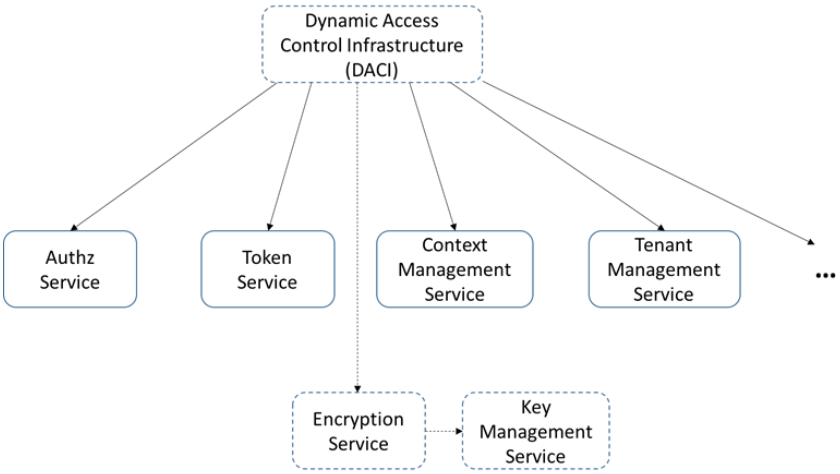

Figure 2. CYCLONE Provisioned Security Services

The security services are made available to a cloud developer in the form of reusable components that are deployed to provisioned VMs at selected providers during application deployment. The authorization standard eXtensible Access Control Markup Language (XACML) is at the core of security services. Since XACML allows authorization policies to be specified by using attributes, it also allows the use of these attributes in building context information.

The use of XACML as the central component allows easier management of permissions and better separation of concerns. For instance, bioinformatics researchers and admins can get an overview of permissions by looking at the policies and obtain insight about current authorizations to genome data, analysis results or available cloud resources such as VMs. XACML also allows dynamic authorization decisions based on context information. A change in the attributes of a user or a resource is sufficient to have an impact on authorizations.

CYCLONE Virtual Private Network service

José Aznar, Project Manager (i2CAT)
CYCLONE WP5 Leader

The OpenNaaS- Network Services Manager/Orchestrator (CNSMO) component has been designed to provide with the network and concrete security services. This provisioning is according to the CYCLONE use cases demands and specific CYCLONE platform needs. One of these services is the Virtual Private Network (VPN) service, which holds a twofold purpose: First, it is in charge of providing the secure connectivity between all client VMs even when these are provisioned by different cloud service providers. Second, the VPN service is considered to establish a secured connection between the CYCLONE user (ASP) and the VMs that put together the application being deployed.

The OpenNaaS-CNSMO is integrated as part of the SlipStream catalogue. It is instantiated and deployed by SlipStream as part of the overall application deployment to make the network services available to the ASPs. Therefore, it has been fundamental to ensure a proper integration of CNSMO with the SlipStream logic.

CYCLONE VPN service components

In order to deploy the VPN Service, the following components are required:

- **VPN Server**: This service expects to receive the certificates and configurations of the VPN Server, and provides API to launch the OpenVPN Server software.
- **VPN Clients**: This Service expects the certificates and configurations of the VPN Clients from the Orchestrator and provides API to manage the OpenVPN client software.
- **VPN Configurator**: This service is in charge to provide the configuration files required for the Server and the Clients and generate the required certificates and keys. As input, this service expects the generic configuration of the VPN like VPN Server Port, IP ranges used by the Clients, etc.
- **VPN Orchestrator**: It interacts with the rest of the services in order to configure and run the VPN as a whole. In order to be able to deploy the service, the orchestrator is subscribed to the other components and it won't start to deploy anything before checking that the status of the rest of the components is ready.

VPN service logic deployment and bootstrapping

Starting from the premise that CNSMO relies on SlipStream to deploy the network services in cloud federated environments, the general bootstrap process, illustrated at Figure 3 is as follows:

Step 1: The CYCLONE user (the ASP) specifies the application profile and requirements. To this end, the ASP must agree with the user of the application on the concrete aspects of the application and has to map them to a recipe in which the details of the required deployment are specified, including the details of the VPN service.

Step 2: Once the application recipe is ready, the user is able to run the application deployment.

Step 3: SlipStream prepares and deploys the requested VMs to run the user's application based on the requested recipe and bootstraps OpenNaaS CNSMO. Thus, CNSMO relies on the VMs deployed by SlipStream to incorporate the VPN service. Each time an application is deployed by SlipStream, it builds component images, runs those images in VMs and, afterwards, runs deployment recipes of each VM.

Step 4: Once CNSMO is launched, the CNSMO SlipStream application contains a deployment recipe with appropriate instructions for CSNMO how to deploy VPN service. The recipe is run by SlipStream inside the CNSMO VM. This recipe gathers information of the deployment from SlipStream and calls the CNSMO API to deploy the VPN service. The recipe then uses the SlipStream to announce to the rest of components that the VPN service has been set up, so SlipStream can resume their deployment.

Step 5: The deployment information is used by CNSMO to determine which services are deployed in which components. In the case of the VPN service, it adds a VPN server VM and allocates ***dockerized*** VPN clients at each of the application VMs.

Step 6: The application deployment is finalized and the application is ready to be used (including the VPN service).

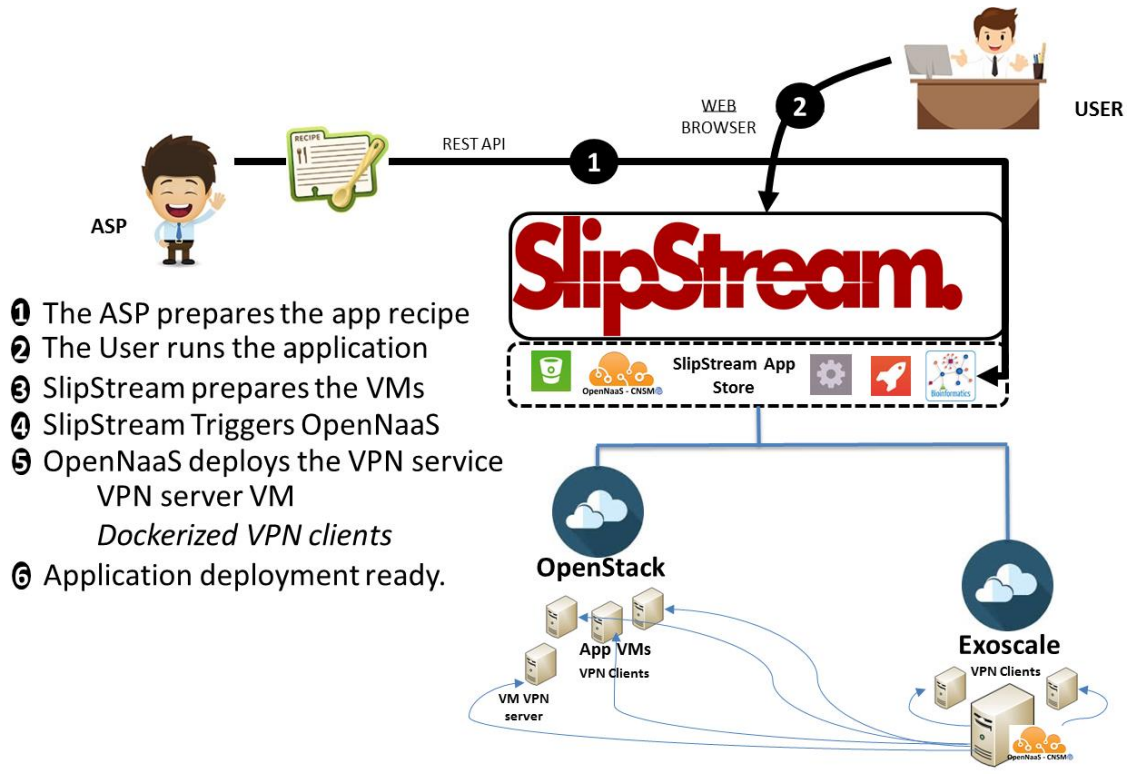

Figure 3. OpenNaaS-CNSMO bootstrapping process to deploy the VPN service.

Thus, the VPN service components are deployed by SlipStream as explained before in a set up consisting of the following VMs:

- **Server VM:** A VM that contains the VPN server CNSMO context, the VPN configurator CNSMO context and the VPN Orchestrator CNSMO context. The role of this VM is to stablish the link between the entire client VMs of the target VPN.
- **Client VMs:** The role of these VMs is just to run the user applications and have a VPN Client CNSMO context. The client VMs are the VMs deployed to run CYCLONE use cases' applications.

Compliance and Security - Existing Standards and Recommendations

Fatih Turkmen Researcher
Yuri Demchenko Senior Researcher (UvA)
CYCLONE WP2 Leader

Compliance and security are related and in some cases interchangeable. ***Security*** is commonly defined as a set of technical, physical, and administrative controls in order to ensure normal operation of a system or application. ***Compliance*** is a certification or confirmation that the system or an organization meets the requirements of the specified standards, established legislation, regulatory guidelines or industry best practices that can be jointly defined as compliance framework. The compliance framework should map different requirements to internal controls and processes. CYCLONE investigates and employs (where applicable) compliance standards and recommendations during applications' design and development. Table 1 provides a list of such standards pertinent to use cases considered in the project. Our particular interest is on the legislations (GDPR and HIPAA/HiTech) and the recommendations (ISO/IEC 27001**,** GINA and GDS) which deal with the protection of sensitive data such as genomic data and the management of security infrastructure in the cloud computing setting. The ultimate goal of the CYCLONE project is to ensure the CYCLONE products are compliant with EU legislations and meet the best practices.

Standard/Regulation name (description)	Acronym
ISO/IEC 27001: Specifies the requirements for establishing, implementing, maintaining and continually improving an information security management system within the context of the organization. Link: https://www.iso.org/obp/ui/#iso:std:iso-iec:27001:ed-2:v1:en	ISO/IEC 27001
NIST SP 800-144 Guidelines for Security and Privacy in Cloud Computing: Provides security and privacy guidelines about outsourcing computation and data handling to/in the cloud. Link: http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-144.pdf	NIST SP 800-144
Cloud Security Alliance, Security Guidance for Critical Area of focus in Cloud Computing: Presents best security practices in the government and operation of cloud from 14 different domains such Security as a Service. Link: https://downloads.cloudsecurityalliance.org/initiatives/guidance/csaguide.v3.0.pdf	CSA guidance
ENISA Cloud Computing Security Risk Assessment: Presents a list of organizational and technical risks related to security in cloud-based systems to allow risk assessment during cloud adoption. Link: https://www.enisa.europa.eu/publications/cloud-computing-risk-assessment	ENISA
General Data Protection Regulation: A regulation by which the European Commission intends to strengthen and unify data protection for individuals within the European Union (EU). It also addresses export of personal data outside the EU. Link: http://ec.europa.eu/justice/data-protection/reform/index_en.htm	GDPR
Health Insurance Portability and Accountability Act: Title II of HIPAA defines policies, procedures and guidelines for maintaining the privacy and security of individually identifiable health information as well as outlining numerous offenses relating to health care and sets civil and criminal penalties for violations. Link: https://en.wikipedia.org/wiki/Health_Insurance_Portability_and_Accountability_Act	HIPAA
Health Information Technology for Economic and Clinical Health Act [8]: The HITECH Act sets meaningful use of interoperable Electronic Health Record (HER) adoption in the health care system as a critical national goal and incentivized EHR adoption. In the first part, meaningful use of EHRs is defined. The second part defines the privacy cases. Link: https://en.wikipedia.org/wiki/Health_Information_Technology_for_Economic_and_Clinical_Health_Act	HITECH
FISMA Certification and Accreditation [9] : Describes security requirements which US federal agencies expect to be in place for the protection of information and information systems. Link: https://en.wikipedia.org/wiki/Federal_Information_Security_Management_Act_of_2002	FISMA
Genetic Information Nondiscrimination Act [10] : GINA is an Act of Congress in the United States designed to prohibit the use of genetic information in health insurance and employment. Link: https://en.wikipedia.org/wiki/Genetic_Information_Nondiscrimination_Act	GINA
Genomic Data Sharing Policy [11] : To foster research and collaboration on genomic data National Institutes of Health (NIH) issues GDS policy for the sharing of genome data. Link: https://gds.nih.gov/	GDS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 644925