Security and Data Protection at Sema4

Last update: 25 March 2024

Introduction

When automating business critical processes, the automations often have to deal with sensitive data that must be protected appropriately. Questions and concerns around a Sema4 deployment often revolve around security and data privacy - and rightly so.

This whitepaper describes Sema4 approach to cybersecurity and data protection, and we also take a look at the design of some key features.

We are committed to ensuring that

- Data handled in our tools remains confidential
- Sema4 Control Room can be used in a secure manner, and
- Users can leverage powerful security features to protect data

On a high level, our cybersecurity initiatives can be structured to the following areas:

1. Compliance with selected programs
2. Company culture and operational security
3. External penetration testing
4. Product security

We'll be covering each of the areas in this document.

Compliance

Compliance-wise we maintain SOC 2 Type II and HIPAA on an annual basis. These two audit programs ensure we have all the basics covered along with a comprehensive set of corporate policies that are compatible with typical information security requirements also in tightly regulated industries such as finance and healthcare. Our customer success team is happy to share the audit reports upon request. Behind the scenes, we use Vanta as a platform for continuous monitoring of our compliance posture.

Security culture and operational security

While compliance and audits are more about checking the required boxes, our security-minded culture and operational security are something we are truly passionate about. Security is central to Sema4's mission - we acknowledge the concerns and risks

many organizations have around putting data to cloud, and we take take our role in enabling safe deployments seriously. Our interests are completely aligned with our customers, as our reputation is on the line.

Since day one, we have fostered a security-conscious culture inside the company. We empower our staff to work in a security aware manner, from onboarding to regular company-wide security training. We also leverage both internal and external security knowledge to test and verify our solution. Operational security consists of the technical measures we take to make sure security and data protection is in place and possible misuses can be detected.

Security and privacy training for employees

All of our employees undergo thorough security and privacy training as part of the onboarding process. Additionally, the trainings are refreshed annually. We also host regular company-wide and team-specific training whenever shifts in our ways of working happen or new risks are identified.

Operational security

Our security motto is to follow the principle of least privilege. We restrict both human and programmatic access to services and data. We remain vigilant for any potential misbehavior in our systems. If we encounter anything that implies a security-related incident has occured, we have a process in place to conduct a thorough security incident response.

Penetration testing

Per our corporate policies, we conduct third-party penetration testing annually, or whenever we introduce significant changes to the architecture - whichever comes first. Given that we want to evolve our product fast, we have also decided to opt for a tighter penetration testing cadence. Currently (Q1/2024) we have a continuous penetration testing agreement with a trusted partner and we get a fresh report every quarter.

If a vulnerability is discovered, we prioritize its fix above other development work and roll out a patch as soon as possible.

We understand that security is an ongoing commitment and despite our great efforts to secure our tools, sometimes what is required to discover a problem is to have multiple pairs of eyes looking at the same solution. That’s why we also have a responsible disclosure policy for security researchers in case they happen to find a vulnerability in our tools.

Product security

Guiding principles

The following principles have been an integral part of our product design since day one.

1. **No security through obscurity.** In practice this means that the security of our product does not rely on the design or the implementation staying secret. We are confident in discussing any part of the solution design - around security or otherwise. We are also happy to engage in any Information Security reviews or audits, because everything is based on solid fundamentals and we have nothing to hide.
2. **Principle of least privilege**. Our product has a modular architecture, and it is composed of several different components or services. Each of these components has a tightly controlled, specific set of permissions that allow them to do their job but nothing else. This means that in the unfortunate case of a security issue in a subsystem, the impact would be also be contained into the area the subsystem is responsible for.
3. **Tenant separation** is rooted deep in the architecture. All data is logically separated by tenant, and all data access - including access from internal services - always requires the tenant identifier to scope the request.

Sema4 Control Room

Sema4 Control Room is used to manage and operate automations that can be run either directly in the cloud with the usage of cloud containers or on local machines and servers with the Worker application.

We use Amazon Web Services (AWS) as our datacenter provider, and we heavily leverage AWS managed services for building our product. In fact, the overwhelming majority of the product runs entirely on managed services - often times referred to as a “serverless” stack. In practice this means AWS is responsible for securing most of the infrastructure all the way up to application layer.

Whenever data travels to Sema4 Control Room we assure that it is:

- Encrypted in transit
- Encrypted at rest using industry standard encryption algorithms
- Data stored in Control Room Vault is also encrypted in application-level

Sema4 product accessible via the self-service interface at https://cloud.robocorp.com is hosted in AWS European data centers and all data processed by customer workloads is therefore located in Europe.

What comes to **data processed by the automations** running on our platform, **we never ever send it to any third parties** apart from infrastructure on AWS. Additionally, we utilize a

tiny, carefully chosen set of additional vendors to provide our services. An up-to-date list of subprocessors can be found at: https://robocorp.com/dpa-subprocessors

Data required in Control Room

The Control Room has been designed to operate with workload-agnosic, minimal control data being exchanged between Workers running on customer infrastructure and the Control Room. The control data includes items such as:

- Workspace (tenant) identifier
- Process and Step identifier and name
- Task package identifier
- Periodic status update containing the current status of the worker and information about available resources such as memory and disk space on the host machine
- Sensitive data such as the “run command” containing instructions and credentials for starting a new run is encrypted on the application layer using the public key of the recipient. This guarantees only the intended recipient is able to decrypt the message.

While the above list is not guaranteed to be exhaustive at the time of you reading this document, the same fundamentals appy to all data potentially added to the interface. The worker control channel is implemented using a websocket protocol with JSON messages, and all communication can be logged and inspected by the customer on the local machine any time.

In addition to the minimum required control data, the Control Room offers additional opt-in features where other types of data is transferred to the Control Room. The customer can always choose which features to utilize depending e.g. on the sensitivity of the data in question.

Enterprise offering

Sema4 Enterprise offering includes alternatives that may be better suited for deployments with strict regulatory or compliance requirements such as:

- Control Room within other geographical regions
- Control Room without any third-party subprocessors apart from AWS
- Virtual Private Control room not accessible via public internet
- Single-tenant Control Room

Get in touch with our sales team for more inforrmation on the Enterprise offering.

Sema4-hosted Cloud Workers

Automations can be run with zero infrastructure setup in a Sema4-provided container. The technology in use is Docker and for each individual run we deploy a short lived container to

provide an environment for duration of the run. The containers are segregated from one another and all data is removed within minutes after completion of the run.

Data Encryption

Communications are always secured using TLS version 1.2 or better, and we adjust our configuration if the best practices happen to change.

For storing and encrypting customer data we primarily rely on AWS managed services such as S3, DynamoDB and Relational Database Service with keys managed by AWS Key Management Service.

In cases where we implement encryption in our application, KMS is still utilized for managing the keys. Under the hood data is typically encrypted using AES with 256-bit keys and whenever we control the mode, we prefer the Galois-Counter Mode.

Additionally, application layer encryption is employed for specifically sensitive data such as Vault content, API keys in database or any payloads containing access tokens. This guarantees that accidental logging or a leaked database backup will not compromise the confidentiality.

Control Room Vault

Hard-coded credentials and secrets in source code are a major security risk in case the code is exposed to third parties accidentally. The Vault is a feature intended to securely manage credentials, API tokens or other secrets needed by automations. Everything stored in the Vault is encrypted at rest. In addition to this, the contents are encrypted two-fold during transit - both by strong HTTPS encryption and by application-level encryption using an envelope encryption pattern. This means that even if the HTTPS traffic were to be exposed by e.g. logging, the secrets within the HTTP message body are not exposed. Organization owner can select whether user can view the secret values in Control Room, or whether the secrets are accessible only to the automations.

API keys in database

When API keys or other credentials are stored in a database, we employ the envelope encryption pattern described along with the Vault implementation.