Project Title	Towards a Sophisticated SIEM Marketplace for blockchain-based Threat Intelligence and Security-as-a-Service		
Project Acronym	PUZZLE		
Grant Agreement No	883540	Instrument	Innovation Action
Call / Topic	H2020-SU-DS-2019 / Small and Medium Enterprises and Micro Enterprises		
Start Date	01/09/2020	Duration	36 months

D1.4 SMEs&MEs Data Chain and Open APIs Targeting at the Services b

Work Package	WP1-Architectural Design and Conceptualization		
Lead Beneficiary	UNIS		
Contributing Beneficiaries	SUITE5, FOGUS, UBITECH, MELLANOX, UBI GR, INTRA, MI, AEGIS, ICCS, FORTH		
Due Date	28.02.2022	Actual Date of Submission	28.02.2022
Version	1.0	Dissemination Level	PU

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

Disclaimer

Versioning and contribution history

Version	Date	Author	Notes
0.1	17.01.2022	Fragkiskos Samaras, Andreas Xanthos (UNIS)	1st Draft; ToC; Allocations
0.2	18.01.2022	Dr. Karagiorgou Sophia (UBITECH) Sofianna Menesidou (UBI GR)	Input at sections 3, 4 & 5
0.3	28.01.2022	Leonidas Kallipolitis (AEGIS)	Input at section 5.3
0.4	10.02.2022	ALL Technical Partners	Input integration from the tech. partners
0.5	15.02.2022	Panos Chatziadam, Nikolaos Petroulakis (FORTH)	Internal Review
0.6	16.02.2022	Manolis Karampinakis, Costas Calogiros (AEGIS)	Internal Review
0.9	25.02.2022	Fragkiskos Samaras, Andreas Xanthos (UNIS)	Comments & reviews addressed; Final version
1.0	28.02.2022	Dr.Stylianos Kazazis (UBI GR), Christina Stratigaki (UBITECH)	QA Review and submission

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

List of Acronyms

Acronym	Description
API	Application Programming Interface
cBPF	classic Berkeley Packet Filter
CPE	Common Platform Enumeration
CTI	Cyber Threat Intelligence
CVE	Common Vulnerabilities and Exposures
CVSS	Common Vulnerability Scoring System
CWSS	Common Weakness Scoring System
DBIR	Data Breach Investigations Report
eBPF	extended Berkeley Packet Filter
EU	European Commission
EUCI	EU Classified Information
IoCs	Indicators of Compromise
ISACs	Information Sharing and Analysis Centres
NVD	National Vulnerability Database
PII	Personally Identifiable Information
RRAE	Runtime Risk Assessment Engine
SMEs&MEs	Small and Medium-sized Enterprises
STIX	Structured Threat Information eXpression
TAXII	Trusted Automated Exchange of Intelligence Information
TLV	Type-length-value
UI	User Interface
WP	Work Package
XDP	eXpress Data Path

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

Table of Contents

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

List of Figures

List of Tables

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

1 Executive Summary

The objectives of this deliverable are to: a) present the data models, the information exchange schemas and the data endpoints for the cybersecurity services that the Small and Medium-sized Enterprises (SMEs&MEs) directly access via the PUZZLE Framework; entitled in this context as *Data Chains*; b) revisit selected components from the technical architecture of the PUZZLE Framework which are accessible by SMEs&MEs and contribute to the end user journey with the PUZZLE Marketplace and the integrated platform; and c) describe the data schema of each Data Chain by means of attributes, required inputs and outputs.

We present the way the various data related tasks have been evolved during the reporting period (i.e. M9-M18). This deliverable (D1.4) differentiates from its previous version for several reasons. First, during this period, the Conceptual Architecture of the PUZZLE Framework (D1.7; M12) was detailed by clarifying the Data and the Control Plane of the PUZZLE Components. In this release of D1.4, we focus on the Data Plane and specifically on the data and their schema. The latter refers to the data that are travelling from the SMEs&MEs perspective, are required for the various inputs/outputs and are visible to the end users while interacting with the PUZZLE Marketplace and the selected user-oriented functionalities of the integrated platform. Also, the specification of the Application Programming Interfaces (APIs) of the PUZZLE prototypes were reported at D5.1 Technical Integration Points, Open APIs Specification and Testing Plan by setting the basis for the technical integration of the solution. Among the technical aspects covered by D5.1, it was also covered the description of the internal interactions of the PUZZLE Components. In this release of D1.4, we focus on the data attributes and the data flows that are needed by the end users to realize their business requirements and are related with the selected cybersecurity services and the PUZZLE Marketplace which are consuming/producing inputs/outputs on their side. This report is the final deliverable of a living document which facilitated to clarify the different Data Chains by SMEs&MEs and has contributed in setting the landscape for interoperable cybersecurity services, omitting replication among the different PUZZLE deliverables and preparing a self-contained report.

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

2 Introduction

2.1 Objectives of the deliverable

This deliverable is the final release of the Data Chains, data schemas and endpoints of the selected PUZZLE Marketplace functionalities, focuses on the user perspective and the direct access to the exposed services by the Small and Medium-sized Enterprises (SMEs&MEs). With a reference to the Conceptual Architecture of the PUZZLE Framework (D1.7; M12), we present the data lifecycle for SMEs&MEs, the end users and third parties which can interact with PUZZLE. The deliverable is the direct outcome of Task 1.3 Data Chain for SMEs&MEs Targeting at the Services and Open APIs Definition under Work Package 1 (WP1) dealing with the data required by SMEs&MEs in order to directly interact with the PUZZLE Marketplace and the selected cybersecurity services which are exposed to their end users. Task 1.3 in its timeframe facilitated to define and refine the data endpoints of the cybersecurity services provided by the PUZZLE Marketplace and the selected services which are accessible by SMEs&MEs. Also, during the reporting period, Task 1.3 has monitored all the technical activities and the identified data endpoints, has concretised with the feedback of the PUZZLE Demonstrators the data flows and therefore D1.4 is differentiated to its previous release (i.e. D1.3) and D5.1. This is because we have clarified the different data attributes and schemas which are of value for SMEs&MEs, the business requirements of the end users and the user journey they will experience while using the PUZZLE Marketplace by leaving out the specificities of the internal interaction of services where WP5 is more appropriate to address. The data endpoints are still important because they facilitate to scale up the solution to third party data or solution providers setting the protocol to either on board new cybersecurity services or to use the data which will be shared or made available to further assess reporting or analytic functionalities.

2.2 Relation with the other Tasks and WPs

Task 1.3 is the starting point of data which has given input to the related technical work, which is conducted in the frame of WP2, WP3, WP4 and WP5. During the reporting period, its progress has been heavily influenced by the outcomes produced in the frame of Task 1.1 Technical Security and Privacy Requirements Analysis, Task 1.2 Supported Use Cases and MVP Definition and Task 1.4 Legal, Ethical and Compliance Requirements. The definition of the data lifecycle

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

in PUZZLE by SMEs&MEs, end users and third-party data or solution providers and the accompanying data schemas have given input to the Task 1.5 PUZZLE Framework Reference Architecture and have assisted to differentiate Task 1.3 with the Task 5.1. As a result, D1.4 focuses on the data endpoints which are accessible by the end users and contribute to the user journey while using the PUZZLE Marketplace and the selected/exposed functionalities. On the contrary, D5.1 focused on the internal interaction of the services which are serving the smooth interoperation of the Marketplace, the Orchestrator and the enforcement of policies which are not directly accessed or controlled by SMEs&MEs end users.

In the reporting period, Task 1.3 has been aligned with the technical requirements and specifications (i.e. WP1), the technical work delivered (i.e., WP2, WP3 and WP4) and the PUZZLE Dashboard (i.e., WP5) integrating the functionalities of the Marketplace and the reporting analytic services which are user-oriented and support SMEs&MEs to address the emerging threats derived by the network or the end user applications. The following figure depicts the relation among the different Work Packages (WPs).

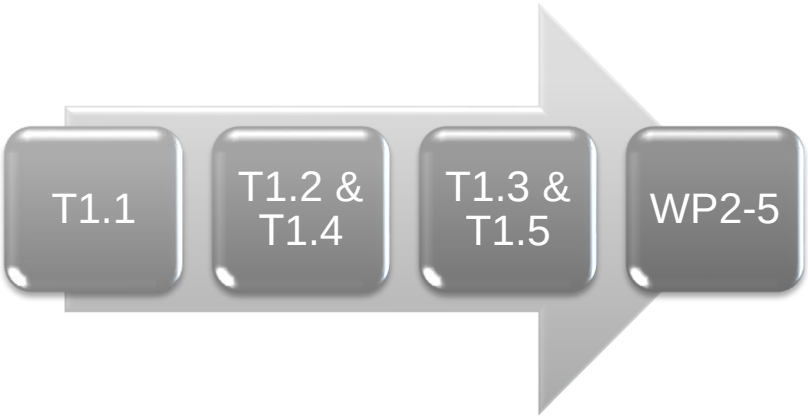

Figure 1- Relation of T1.3 and D1.3 with the other Tasks and WPs.

2.3 Structure of the document

The document is structured among the following main chapters:

- **Chapter 1** presents the Executive Summary of this deliverable;
- **Chapter 2** introduces the deliverable, its scope, its differentiation across the reporting period, the input it gets from other tasks and the contribution it makes in the other work packages;

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

- **Chapter 3** provides the data lifecycle of SMEs&MEs in the PUZZLE Marketplace and the selected cybersecurity services which are directly accessible by the end users;
- **Chapter 4** presents the High-Level Conceptual Architecture of the PUZZLE Marketplace, selects and highlights functionalities which are accessed by SMEs&MEs;
- **Chapter 5** presents the overview, the data and the schemas of the PUZZLE Marketplace and the selected services targeting the end users;
- **Chapter 6** concludes this deliverable and summarizes the outcomes of Task 1.3.

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

3 Information Exchange Schemas and Data Chain for SMEs&MEs

In the following sections, we present the latest figures and investigations about data breaches and their impact in SMEs&MEs. Also, we describe the standardised data models and schemas used in the cybersecurity which are needed to maintain and report on identified incidents, as well as the data lifecycle of SMEs&MEs in PUZZLE. The standardised information exchange schemas are suitable for threat intelligence and sharing, vulnerabilities scoring and risk quantification. End users and SMEs&MEs use them in the PUZZLE Framework while interacting with the Runtime Risk Assessment Engine (RRAE), Trust Assurance Services, Blockchain-based Threat Intelligence, Policy Enforcement Services, Edge and Cloud Analytics, PUZZLE Dashboard and PUZZLE Marketplace.

3.1 Data Lifecycle in SMEs&MEs

Data is the greatest asset for modern organizations of any size, and the data lifecycle is key in running organizations smoothly. Business data can also be one of the greatest risks when left unprotected or inadequately managed. As the volume of data within businesses grows, so does the challenge of efficiently protecting and managing it [1].

Especially for small and medium size organizations, establishing effective governance practices is increasingly critical. According to Forbes [4], SMEs&MEs have suffered 50% more cyberattack attempts per week in 2021, while the rise — partly due to Log4j [5]— helped boost cyberattack attempts to an all-time high in Q4 2021. According to Verizon's 2021 Data Breach Investigations Report (DBIR) [3], the number of breaches in smaller companies in 2020 was close to those targeting large ones, while the common reasons why SMEs&MEs overlook cybersecurity and data protections [6] are as follows:

- **Reason #1**: The first reason is that there seems to be a misconception that Cyber Criminals mainly target big businesses. In reality, 43% of all cyberattacks are aimed at SMEs, as many cyber criminals view them as easy pickings. In contrast, 54% of SMEs believe they are too small to be targeted.
- **Reason #2**: Another is that small businesses underestimate the value of their data. However, it is estimated that just a single stolen record containing Personally

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

Identifiable Information (PII) costs €130, which multiplied by thousands of records on customers or staff the true cost is larger.

- **Reason #3**: Some small business owners are also unaware of all the possible consequences of being a cyber victim. This fact may cost up to €11,000 to investigate.
- **Reason #4**: Another obstacle is the perceived difficulty and cost of becoming compliant with cybersecurity guidelines and implementing cybersecurity measures, where as many as 47% of small businesses do not understand how to protect themselves against cyber-attacks and up to 3 in 4 SMEs don't have dedicated IT security personnel.

In the context of the PUZZLE Marketplace, the lifecycle of data within an enterprise relates with the continuous monitoring of the collected data and the tasks performed by end users with low cybersecurity experience to configure and set runtime adaptations to the Runtime Risk Assessment Engine, Trust Assurance Services, Blockchain-based Threat Intelligence, Policy Enforcement Services, Edge and Cloud Analytics and PUZZLE Dashboard. The data lifecycle also relates with the usage (i.e., upload, delete, update, add metadata and executables for the cybersecurity services) of the PUZZLE Marketplace. Regarding the business-related data, the PUZZLE Marketplace focuses on providing algorithms and mechanisms for secure-by-design and private-by-design access, incorporating certification, authentication and encryption methods among the data endpoints and the underlying cybersecurity services. For this reason, in this section we focus on the identification of the various data types and the respective standards which set the baseline for the services of the Data Plane and the selected functionalities that are directly accessible by SMEs&MEs and their end users.

According to ENISA's Cooperative models [2] regarding the Information Sharing and Analysis Centres (ISACs), the type of information, which needs to be collected, monitored and exchanged, includes:

- Incidents - details of attempted and successful attacks that may include a description of information lost, techniques used, intent, and impact. The severity of an incident could range from a successfully blocked attack to a serious national security situation;
- Threats - yet-to-be-understood issues with potentially serious implications; indicators of compromise, such as malicious files, stolen email addresses, impacted IP addresses, or malware samples; or information about threat actors. Threat information can help

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

operators detect or deter incidents, learn from attacks, and create solutions that can better protect their own systems and those of others;

- Assets - they could be abstract assets (like digital services, processes, etc.), virtual assets (data, models' analysed results, etc.), physical assets (devices, hardware, cables, a piece of equipment), and primarily refer to cyber assets in the context of PUZZLE. The assets are used to quantify and assess the risk from the Runtime Risk Assessment Engine.;
- Vulnerabilities - in software, hardware, or business processes that can be exploited for malicious purposes;
- Mitigations - methods for remedying vulnerabilities, containing or blocking threats, and responding to and recovering from incidents. Common forms of such information include patches to plug vulnerabilities, antivirus updates to stop exploitation, and directions for purging malicious actors from networks;
- Vendors – the applications share a common product vocabulary allowing interoperability and the identification of products at a standardized level of granularity;
- Situational awareness - information that enables decision-makers to respond to an incident and that may require real-time telemetry of exploited vulnerabilities, active threats, and attacks. It could also contain information about the targets of attacks and the state of critical public or private networks;
- Best practices - information related to how software and services are developed and delivered, such as security controls, development and incident response practices, and software patching or effectiveness metrics;
- Strategic analysis - gathering, distilling, and analysing many types of information to build metrics, trends, and projections. It is often blended with projections of potential scenarios to prepare government or private sector decision-makers for future risks.

In D1.3 we presented the information exchange schemas (i.e. CVEs, CVSS, CPEs, STIX, TAXII, etc.) in cybersecurity which are being used in the project to serve the different technical and business data flows and interactions among the PUZZLE's probing, monitoring and analysis services, as well as the targeted SMEs&MEs applications. In the following section, we focus on the Data Plane which refers to the data that is travelling from the SMEs&MEs perspective, is required for the various inputs/outputs and is visible to the end users while interacting with

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

the PUZZLE Marketplace and the selected user-oriented functionalities of the integrated platform.

3.2 Data Lifecycle for SMES&MEs in PUZZLE

Before we delve into the details of the data models and the data chains supported by each PUZZLE Component accessible by the end users, it is important to elaborate on the data prerequisites that should be satisfied in the technical solution. The data prerequisites are coming from the human-in-the-loop actions and include the customization and filling of the Policy Template. The Policy Template can be configured at the Administrative Interface of the PUZZLE Dashboard where the end users can parameterize security policies.

3.2.1 Policy Template

The Policy Template allows to fill in and configure the actions which are taken by the end users to instantiate and provide runtime adaptations to the PUZZLE Components via the PUZZLE Dashboard. The end users can interact with the PUZZLE Dashboard in order to instantiate some of the templated security services that are hosted on the PUZZLE Marketplace. This is an important aspect that has to be clarified; Marketplace does not contain binary artefacts that are 'downloadable' and 'installable' at the worker-level. On the contrary, it contains formal descriptions of security services in the form of customizable templates that are materialized during their instantiation by the end user. Before we shed light on this 'template instantiation' we should clarify that templates are conceptually grouped in Runtime Risk Assessment Engine, Edge and Cloud Analytics, Policy Enforcement Services and Trust Assurance Services.

3.2.2 SMEs&MEs using the PUZZLE Data Plane

The artefacts specified in D1.7 can be translated by both the Control and the Data Plane of the PUZZLE Framework. The ones targeting the Control Plane are automatically instantiated without user intervention. However, there are artefacts which are related to the Runtime Risk Assessment Engine, Edge and Cloud Analytics, Policy Enforcement Services and Trust Assurance Services, which can be templated as service descriptions by the end user and contribute to the Data Plane configuration. The programmable Data Plane configuration

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

includes all the data that is traveling across the PUZZLE Framework to ground the solution with the business requirements. The files and the requirements are as follows:

- Objective files (binary) of accelerated packet replication rules. These files are injected into the eBPF engine of the worker's kernel and are mainly used for Analytics.
- Objective files (binary) of accelerated packet processing rules. These files are injected into the XDP engine of the worker's kernel and are mainly used for Analytics and Enforcement Services.
- Hashing requirements for attestation challenges, which is mainly used in Trust Assurance Services.
- Complex Event Processing rules that will be used to model complex conditions that have to trigger an action if-positive. These rules are registered to the central Expert System.
- Define a compatible analytics pipeline that can be used on top of replicated traffic. These pipelines are used in Analytics.

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

4 PUZZLE's High-Level Architecture

This section presents and highlights the components derived by the PUZZLE Logical Reference Architecture delivered through D1.7 in M12 by focusing on the services directly access by the SMEs&MEs, end users and the PUZZLE Demonstrators.

As highlighted in Figure 2, the PUZZLE technical components directly accessed by SMEs&MEs, the end users and the Use Cases Demonstrators are as follows:

1. The **Runtime Risk Assessment Engine** (RRAE) which natively quantifies risks and vulnerabilities for the SMEs software/services assets "attached" and exposed to the PUZZLE Framework. RRAE can be also configured to quantify risks and vulnerabilities for additional software and hardware assets graphs, enriched through a User Interface (UI) by the end users and administrators of the Demonstrators, external SMEs and MEs.

2. The **Trust Assurance Services** enhance traffic and service integrity and trustworthiness. Two types of services based on eBPFs are supported: a) the traffic integrity and b) the service integrity.

3. The **Blockchain-based Threat Intelligence** which incorporates all components that materialize a secure and trustworthy data sharing environment regarding threat intelligence information. It uses the Structured Threat Information eXpression (STIX) standard [7].
4. The **Policy Enforcement Services** which enforce the cybersecurity services and policies via the definition and configuration of the Policy Template by the end users. This template, during instantiation can be concretized or extended. The template has to be syntactically and logically correct and contains rules which enable the services instantiation, the introspection of variables and the configuration of user preferences.
5. **Edge and Cloud Analytics** which incorporate all algorithms and components that undertake "offline analytics" tasks i.e., tasks that are not considered near real-time, and assist on performing network data interception, packet indexing and data manipulation (i.e., replication, offloading, dropping, etc.), as well as attestation challenges by implementing the appropriate signalling for the integrity assurance of the running services.

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

6. The Administrative and Operational Dashboard, hereafter **PUZZLE Dashboard**, which assists in instantiating some of the "templated" security services that are hosted on the PUZZLE Marketplace.

7. The **PUZZLE Marketplace** which encapsulates the central reference point where "templated" security services are offered, packaged, and advertised.

In Figure 2, the PUZZLE Components along with the underlying services through which end users can interact with the entire framework are highlighted.

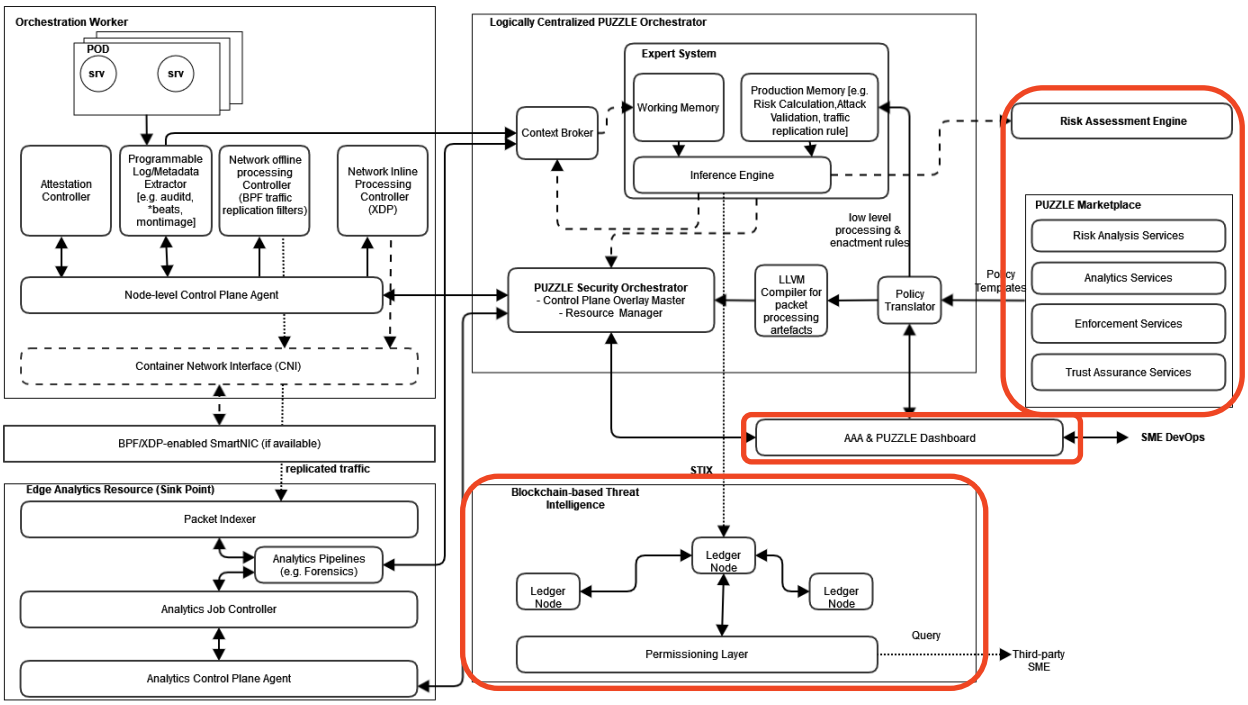

Figure 2- PUZZLE Logical Reference Architecture.

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

5 PUZZLE Components and Data Exchange Schemas for SMEs&MEs

In the following sections, we present the PUZZLE Components and the main data exchange endpoints by means of constraints, input and output data.

5.1 Runtime Risk Assessment Engine

5.1.1 Overview

The PUZZLE Runtime Risk Assessment Engine component is responsible for performing a security and privacy risk analysis of every software (i.e. information system, service, REST API, etc.) or hardware (i.e. router, switch, gateway, hub, edge device, etc.) asset that is exposed via an IP/MAC address by the SMEs/MEs infrastructure. To materialise the assessment of the risk, the engine consumes a metamodel by means of graph which represents all the alternative assets paths in a bottom-up approach, starting from the SMEs/MEs assets (i.e., devices, operating systems, services, etc.) and hierarchically propagating to SMEs/MEs serving business applications exposed to the surface Internet. The engine is multi-threaded by design since each separate risk requires different set of calculations by taking into account basic (i.e., attack vector, complexity, etc) and environmental (i.e., code maturity, remediation level, etc.) aspects from the SMEs/MEs ecosystems.

5.1.2 Description of Data Schema

The required data schema in order to build this specific Data Chain with SMEs&MEs with the Runtime Risk Assessment Engine of PUZZLE is as follows:

Table 1- Runtime Risk Assessment Engine Data Schema

Data Attribute	Data Description and Role
Assets	Specifies all the places where organization keep sensitive information (software and hardware). Consists of name, CPE name, category (e.g. network, OS, storage, software, etc.), business value, IP & port (if applied), connection with other assets and user groups that have access on the asset

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

Vulnerabilities	All known vulnerabilities from NIST (NVD). Each vulnerability is assigned by CVE identifier and describes a weakness found in software or hardware products. Also, vulnerabilities consist of scores related to negative impact to confidentiality, integrity, or availability when vulnerability is exploited. Schema format: <CVEid, accesscomplexity, accessvector, authentication, availabilityimpact, confidentialityimpact, cvssscore, exploitabilityscore, impactscore, integrityimpact, isconfirmed, legalimpact, islegal, legalavailabilityimpact, legalconfidentialityimpact, legalintegrityimpact, modified, published, description, protection>
Common Platform Enumeration	A structured naming schema for software and hardware products. CPE list is provided by open APIs and describes product, type, vendor and known vulnerabilities for this product. Schema format: <cpe_version>:<part>:<vendor>:<product>:<version>:<update>:<edition>:<language>:<sw_edition>:<target_sw>:<target_hw>:<other> The latest CPE definition version is 2.3
Business Processes	A structured and reusable activity that contains a subset of organization assets in order to accomplish a specific organizational goal

5.2 Trust Assurance Services

5.2.1 Overview

The Trust Assurance Services ensure the integrity of the traffic and the interacting services. They are responsible for performing the checks and controls about the devices' (or execution environments') state, by exploiting the eBPFs lightweight bytecodes. Attestation mechanisms will be integrated for verifying the integrity of the monitored traffic by the deployed Programmable Agents (cf. D1.7; Programmable Agents for Network Inline Processing, Offline Processing and Log or Metadata Extraction). The focus of the Trust Assurance Services is to

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

protect the sharing of this data (alongside the evidence as compiled by the attestation mechanisms). The Data Chains established by the different Programmable Agents will handle network traffic and kernel-based integrity checks as follows:

- **Network Traffic Inline Processing**: "inline" processing via lightweight bytecodes will be supported by using eXpress Data Path (XDP). XDP is a new system in the kernel that lets you write custom eBPF programs to filter network packets. In the PUZZLE Framework, we call them "XDP programs". The XDP programs run as soon as the packet gets to the network driver (so very quickly). When an XDP program is being executed, it returns and exits with either XDP_TX[1], XDP_DROP[2], or XDP_PASS[3]. PUZZLE Inline Processing policies are shipped within the Policy Templates (cf. D1.7).
- **Offline Processing**: "offline" processing refers to a kernel packet filter which allows a user space program to attach a filter program onto a socket and limit certain dataflows coming through the socket in a fast and effective way. Linux BPF originally provided a set of instructions that could be used to program a filter: this is nowadays referred to as classic BPF (cBPF).
- **Log or Metadata Extraction**: it refers to the programmability of the data plane to extract logs and metadata from the inline and the offline processing. This is going to be done through dpdk rte_flow API. A flow is built from matching and operation values. All matching and operation are TLV (Type-length-value) and build to be extended for future protocols and actions. A flow has group and priority and allow to build a hierarchal of tables. Matching can be performed on packet data (protocol headers, payload) and properties. Possible operations include dropping traffic, diverting it to specific queues, to virtual/physical device functions or ports, performing tunnel offloads, adding marks and so on.

[1] The XDP_TX action results in bouncing the received packet-page back out the same NIC it arrived on.
[2] XDP_DROP instructs the driver to drop the packet. Given this action happens at the earliest RX stage in the driver, dropping a packet simply implies recycling it back-into the RX ring queue it just "arrived" on. There is simply no faster way to drop a packet. This comes close to a driver hardware test feature.
[3] XDP_PASS means the XDP program chooses to pass the packet to the normal network stack for processing.

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

5.2.2 Description of Data Schema

The required data schema in order to build this specific Data Chain with SMEs&MEs with the Trust Assurance Services of PUZZLE is as follows:

Table 2- Trust Assurance Services Data Schema

Data Attribute	Data Description and Role
Inline processing	Policy Templates with the specifications XDP_TX, XDP_DROP, XDP_PASS
Offline processing	Set some buffer size, BPF filters, etc. to offload data and process them in a batch/offline manner
Log and Metadata	Type-length-value attributes to match rules in the support of logs and metadata extraction. XDP metadata which support a) counters: they allow admins and monitoring tools to catch and count monitoring events; b) tracepoints: they are much more flexible than counters, with the downside that errors might never be caught, if the tracepoint isn't active.

5.3 Blockchain-based Threat Intelligence

5.3.1 Overview

The PUZZLE Blockchain-based Threat Intelligence is responsible to make available selected data about cyber-incidents, Indicators of Compromise (IoCs) and detected vulnerabilities which can be shared among different SME&MEs, end users and stakeholders granted access to the Distributed Ledger infrastructure. The Blockchain-based Threat Intelligence will serialize produced indicators at STIX format [7] and the platform will offer TAXII feeds [8]. Within the platform every indicator that is identified will be also published in a Distributed Ledger in order to achieve accountability.

For this purpose, PUZZLE will support the de-facto protocols and standards in the field of Threat Intelligence storage and sharing. More specifically, PUZZLE will adopt STIX and TAXII protocols. STIX stands for Structured Threat Information Expression (STIX™) and is a language and serialization format used to exchange cyber threat intelligence. On the other hand, TAXII stands for Trusted Automated Exchange of Intelligence Information (TAXII™) and is an application layer protocol for the communication of cyber threat information in a simple and scalable manner. TAXII enables organizations to share CTI by defining an API that aligns

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

with common sharing models. TAXII is specifically designed to support the exchange of CTI represented in STIX. Therefore, the two protocols are complementary to each other.

5.3.2 Description of Data Schema

The required data schema in order to build this specific Data Chain with SMEs&MEs with the Blockchain-based Threat Intelligence of PUZZLE is as follows:

Table 3- Blockchain-based Threat Intelligence Data Schema

Data Attribute	Data Description and Role
Incidents and Data to be shared	• type: vulnerability type • name: Name to identify the vulnerability • description: Description for the content of the message • impact_level: the extent of a malware implications • likelihood: the probability of a vulnerability to be compromised • created: timestamp of the creation of the message • modified: timestamp of the last modification time of a message • pattern: the detection pattern for the vulnerability described in this message • malware_types: the types of the malware, taken from a STIX open vocabulary. • valid_from: timestamp that details the time from which this indicator is still considered valid intelligence.
Vulnerabilities	• adware: software funded by advertising that might also gather sensitive user information • backdoor: a malicious program that allows an attacker to perform actions on a remote system.

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

	• bot: a program that resides on an infected system, communicating with and forming part of a botnet. • ddos: A program that is used to perform a distributed denial of service attack. • exploit-kit: a software toolkit to target common vulnerabilities. • keylogger: a type of malware that monitors keystrokes. • ransomware: a type of malware that encrypts files on a victim's system, demanding ransom to unlock them. • rootkit: a type of malware that conceal its presence and activities. • screen-capture: a type of malware used to capture images from target systems screen • spyware: software that gathers information on a user's system without their knowledge and sends it to another party. • trojan: any malicious computer program which is used to hack into a computer by misleading users of its true nature. • virus: a malicious program that replicates by reproducing itself or infecting other programs by modifying them. • worm: a self-replicating, self-contained program that usually executes itself without user intervention.

5.4 Policy Enforcement Services

5.4.1 Overview

The PUZZLE Policy Enforcement Services support a multi-objective optimisation engine which gets as input near-optimal configurations by means of Policy Templates and give as output the deployment specifications (i.e. service descriptions) of the cybersecurity services at the SMEs/MEs private infrastructure. It also takes into account resource constraints of the underlying infrastructure along with security, configuration and user-defined constraints in order to select the optimal collection of cybersecurity services to be downloaded by the PUZZLE Marketplace and be deployed and enforced by the PUZZLE Security Orchestrator in an agnostic to the end user manner.

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

5.4.2 Description of Data Schema

The required data schema in order to build this specific Data Chain with SMEs&MEs with the Policy Enforcement Services of PUZZLE is as follows:

Table 4- Policy Enforcement Services Data Schema

Data Attribute	Data Description and Role
Policy Template	Structured and ordered list with suggested policies by means of eBPF rules which can be downloaded by the Marketplace (in JSON format) and enforced at the PUZZLE Security Orchestrator

5.5 Edge and Cloud Analytics

5.5.1 Overview

Edge and Cloud Analytics involve the analysis and reaction in real-time by harvesting data that are being collected and processed in three modes, in: batches, micro-batches and streams. Several factors need to be considered such as the available resources in the devices, identifying what data is critical and needs to be analysed in real-time, whether additional analysis needs to be done, and if there are storage requirements (e.g., for forensics, historical analysis, or satisfying legal requirements). Edge Analytics is particularly important for managing the security of high number of heterogeneous dispersed devices that would be impossible without a minimum of automation and distribution of processing and analysis tasks.

5.5.2 Description of Data Schema

The required data schema in order to build this specific Data Chain with SMEs&MEs with the Edge and Cloud Analytics of PUZZLE is as follows:

Table 5- Edge and Cloud Analytics Data Schema

Data Attribute	Data Description and Role
Events	Extracted attributes from events (e.g., network packets, network sessions, log entries, syslog, application traces or messages)
Metadata	Computed values, statistics, extracted features
Devices and Endpoints	Compromised devices or endpoints. Many different types of cyber-attacks, e.g., DDoS attacks, Malicious scanning, Man-in-the-Middle attacks, Tampering or eavesdropping of the transmitted data

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

Generic or specialised honeypots	Scanning campaigns. Zero-day attacks. Attacks using IoT devices. Botnets and Command and Control activities. Attacks on web applications.
BGP announcements	AS-level attacks. Internet routing attacks: BGP poisoning, hijacking, DNS cache poisoning, DNS spoofing, DNS hijacking, man-in-the-middle attacks, Distributed Reflected DoS.
Traceroute data	IP-level attacks. Abnormal network disruptions (delay changes, packet loss, forwarding), analysing packet routes via Internet eXchange Points (IXP), DDoS.
Indicators of Compromise (IoCs)	Indicators of Compromise in the form of a message (e.g., in JSON, STIX or CSV format) corresponding to an alarm or notifications. Includes at least information on the rule that detected it (an ID number), a short description, the source and destination IP addresses, the attributes from the packets that were used to detect it. List of extracted features from the packets and sessions corresponding to statistics, attribute values, time related values, etc. These can be used for further analysis (e.g., Complex Event Processing, Change Point Analysis, behaviour analysis based on Machine Learning).

5.6 PUZZLE Dashboard

5.6.1 Overview

The PUZZLE Dashboard shall enable SMEs&MEs and end-users, especially non-technical persons, to interact with the PUZZLE technical components and to increase their cybersecurity awareness levels regarding the status of their assets, including vulnerabilities, threats and risks, on-going attacks, as well as proposed mitigation actions. It will also facilitate the authoring, management and deployment lifecycle of the cybersecurity services provided by the PUZZLE Marketplace via an interface to configure the Policy Templates.

5.6.2 Description of Data Schema

The required data schema in order to build this specific Data Chain with SMEs&MEs with the PUZZLE Dashboard is as follows:

Table 6- PUZZLE Dashboard Data Schema

Data Attribute	Data Description and Role
Real-time data	All the data that are generated from interactions with the PUZZLE dashboard as well as any incoming data to the dashboard at real-time. Notifications and alerts are included in this data schema (predefined JSON and STIX formats will be used). These

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

	data will be transmitted over real-time messaging systems (Kafka)
Historical data	All the data that the dashboard transmits or receives from persistent data storages (predefined JSON and STIX formats will be used)

5.7 PUZZLE Marketplace

5.7.1 Overview

The PUZZLE Marketplace acts as the main entrance point of the PUZZLE ecosystem and the main portal from where SMEs&MEs and users will be able to onboard, express their requirements, provide their own services by means of eBPF mechanisms and trigger all the other PUZLLE components.

5.7.2 Description of Data Schema

The PUZZLE Marketplace will mostly provide data and metadata to other components and as such it is expected that this component will only implement GET methods. The required data schema in order to build this specific Data Chain with SMEs&MEs with the PUZZLE Marketplace is as follows:

Table 7- PUZZLE Marketplace Data Schema

Data Attribute	Data Description and Role
Policy Templates	Templates are stored in Marketplace database (MongoDB); Configured Policy Templates are sent to Orchestrator (in JSON format)
User Data	PUZZLE user related data and metadata (stored in a relational database) for user management and user authentication/authorisation by the AAA components
Infrastructure Data	Data related to the SME infrastructure (e.g., assets information) stored as part of the Application space and to be used for the risk assessment

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

6 Conclusions

The objectives of D1.4 were to: a) present the data models, the information exchange schemas and the data endpoints as the Data Chains of the cybersecurity services that the Small and Medium-sized Enterprises (SMEs&MEs) directly access via the PUZZLE Framework; b) revisit selected components from the technical architecture of the PUZZLE Framework which are accessible by SMEs&MEs and contribute to the end user journey with the PUZZLE Dashboard, PUZZLE Marketplace and the integrated platform; and c) describe the data schema of each Data Chain by means of attributes, required inputs and outputs.

We reported the final data exchange schemas according to the existing cybersecurity threat and information sharing principles and standards focusing on the selected PUZZLE Components and their underlying services. We presented the data lifecycle of SME&MEs in PUZZLE by using its Data Plane.

During the reporting period, Task 1.3 has monitored all the technical activities and the identified data endpoints across the different technical work packages (WP2-WP5) in alignment with the Conceptual Architecture of PUZZLE. The differentiations of D1.4 from its previous version is that in this release we have focused on the Data Plane of the PUZZLE Framework perceived by the end users. The programmable Data Plane refers to the data that are travelling from the SMEs&MEs perspective, are required for the various inputs/outputs and are visible to the end users while interacting with the PUZZLE Marketplace and the selected user-oriented functionalities of the integrated platform. It also refers to the templated service descriptions required by the end users in order to initiate the PUZZLE Components and perform runtime adaptations. During this period, the Conceptual Architecture of the PUZZLE Framework (D1.7; M12) was detailed by clarifying the Data and the Control Plane of the PUZZLE Components. Also, the specification of the Application Programming Interfaces (APIs) of the PUZZLE prototypes were reported at D5.1 Technical Integration Points, Open APIs Specification and Testing Plan by setting the basis for the technical integration of the solution. This final release of D1.4 concentrates on the data attributes and the data flows that are needed by the end users to realize their business requirements and are related with the selected cybersecurity services and the PUZZLE Marketplace which are consuming/producing inputs/outputs on their side. This report is the

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

outcome of a living document which has facilitated to clarify the different Data Chains by SMEs&MEs and has contributed to setting the landscape for interoperable cybersecurity services while putting the human-in-the-loop only for the actions that are required.

In the next period, we will intensify our effort to validate the Data Chains in close collaboration and alignment with the PUZZLE Pilots and the participants who will be funded with the Validation Contracts.

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

7 References

[1] Brian Curtis. Complete Framework of Data Lifecycle Management. Available Online: https://www.yourtechdiet.com/blogs/5-best-practices-to-enhance-data-lifecycle-management/

[2] ENISA: Information Sharing and Analysis Center (ISACs) - Cooperative models. Available Online: https://www.enisa.europa.eu/publications/information-sharing-and-analysis-center-isacs-cooperative-models

[3] Verizon: 2021 Data Breach Investigations Report (DBIR). Available Online: https://enterprise.verizon.com/resources/reports/2021-data-breach-investigations-report.pdf

[4] Chuck Brooks. Cybersecurity in 2022 – A Fresh Look at Some Very Alarming Stats. Forbes 2022. Available Online: https://www.forbes.com/sites/chuckbrooks/2022/01/21/cybersecurity-in-2022--a-fresh-look-at-some-very-alarming-stats/?sh=1e20f5196b61

[5] John K. Waters, Kurt Mackie. Widespread Log4j Remote Code Execution Vulnerability Could Affect Millions. Dec. 2021. Available Online: https://redmondmag.com/articles/2021/12/15/widespread-log4j-remote-code-execution-vulnerability-could-affect-million.aspx

[6] Cybersecurity & Data Breaches – 10 things every SME needs to know. Jan. 2022. Available Online: https://manageditexperts.co.uk/cybersecurity-data-breaches-10-things-every-sme-needs-to-know/

[7] Structured Threat Information eXpression (STIX). Available Online: https://stixproject.github.io/

[8] Trusted Automated Exchange of Intelligence Information. Available Online: https://oasis-open.github.io/cti-documentation/taxii/intro.html

The work described in this document has been conducted within the project PUZZLE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 883540. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.