2022 22nd International Symposium on Cluster, Cloud and Internet Computing (CCGrid) | 978-1-6654-9956-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/CCGRID54584.2022.00037

Protecting sensitive data in the cloud-to-edge continuum: The FogProtect approach

Dhouha Ayed
Thales
Palaiseau, France

Paul-Andrei Dragan
University of Duisburg-Essen
Essen, Germany

Edith Félix
Thales
Palaiseau, France

Zoltán Ádám Mann
University of Amsterdam
Amsterdam, The Netherlands

Eliot Salant
IBM Research
Haifa, Israel

Robert Seidl
Nokia Bell Labs
Munich, Germany

Anestis Sidiropoulos
Athens Technology Center
Athens, Greece

Steve Taylor
University of Southampton
Southampton, UK

Ricardo Vitorino
Ubiwhere
Aveiro, Portugal

Abstract—**Data produced by end devices like smartphones, sensors or IoT devices can be stored and processed across a continuum of compute resources, from end devices via fog nodes to the cloud, enabling reduced latency, increased processing speed and energy savings. However, the data may be sensitive (e.g., personal data or confidential commercially sensitive information), with regulatory or other requirements for its protection.**

Protecting sensitive data in the dynamic, heterogeneous, and decentralized cloud-to-edge continuum is very challenging. This paper describes a solution: FogProtect, an integrated set of four technologies to protect data in the cloud-to-edge continuum. FogProtect addresses four concerns: (i) control and enforcement of distributed data access and usage; (ii) management of distributed data protection policies; (iii) risk assessment for data assets in the cloud-to-edge continuum; (iv) automated optimisation and adaptation to address identified risks. FogProtect operates dynamically, reacting to system changes or detected vulnerabilities to keep the data secure across the cloud-to-edge continuum.

This paper describes an overview of the FogProtect concept, discusses each of the four approaches, and illustrates their usage for the protection of data in three real-world use cases.

Index Terms—**fog computing, edge computing, data protection, security, privacy**

I. Introduction

Recent developments in cloud computing, the Internet of Things (IoT), and networking technologies have led to a continuum of connected devices. On one extreme of this continuum, there are end devices (e.g., sensors, smart wearables, cameras) that are typically resource-constrained, heterogeneous and geographically distributed. On the other extreme, there are cloud data centers, offering virtually unlimited compute and storage capacity. In between, there can be many different devices (called fog nodes or edge nodes), offering cloud-like services with limited capacity close to end devices. Applications can be partitioned among the different devices in the continuum, making optimal use of their strengths, e.g., by processing data coming from end devices in nearby fog nodes, while using the cloud for long-term data storage [1, 2].

The adoption of the cloud-to-edge continuum is driven mainly by performance and cost advantages. However, security and privacy are major concerns [3]. The cloud-to-edge continuum is frequently used for processing data that may be sensitive for various reasons. E.g., if the data relates to identifiable persons, data processing must comply with applicable law, such as the General Data Protection Regulation (GDPR) in the European Union. Or the data can be commercially sensitive, needing protection for business reasons.

Protecting data in the cloud-to-edge continuum is challenging for multiple reasons [4]. The continuum is characterized by a multitude of different stakeholders that have different roles, different interests, use different technologies, and stand in different relations to each other. This heterogeneity makes it difficult to enforce uniform security standards and to protect data throughout its lifecycle. The continuum is also subject to frequent changes (e.g., deployment of new services, failure of nodes, changes in the physical environment), which may impact the risk of data protection violations, requiring the dynamic application of appropriate countermeasures.

While there are useful technology building blocks for ensuring particular aspects of security or privacy (e.g., encryption for ensuring secrecy of data transfer between two devices), we are not aware of a technology that would ensure the end-to-end protection of sensitive data in the cloud-to-edge continuum.

This paper describes FogProtect, an integrated set of technologies to protect data in the cloud-to-edge continuum. FogProtect combines four complementary technologies, addressing four concerns. First, FogProtect controls distributed data access and data usage. Second, FogProtect manages distributed data protection policies by orchestrating different security enablers. Third, FogProtect provides data-protection-aware self-adaptation to enable automatic reaction to changes. Finally, FogProtect performs automated risk assessment for data assets in the cloud-to-edge continuum. All these technologies operate dynamically, ensuring that data is protected across the cloud-to-edge continuum and during its whole lifecycle.

The four technologies of FogProtect work closely together, based on a defined protocol. The versatility of these technologies and of their interplay allows the application of FogProtect

This work was partially supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 871525 (FogProtect). Contributions from all members of the FogProtect consortium are gratefully acknowledged.

978-1-6654-9956-9/22/$31.00 ©2022 IEEE
DOI 10.1109/CCGrid54584.2022.00037

in various scenarios. We demonstrate this for three different real-world use cases. In the Smart City use case, FogProtect protects video streams from CCTV cameras processed in smart lampposts. In the Smart Manufacturing use case, FogProtect prohibits unauthorized access to data of different sensitivity in a factory. In the Smart Media use case, FogProtect enables the secure cooperation of two organizations on sensitive data.

II. Related work

Data access and usage control. The ability for *data controllers* to meet regulations on data protection and privacy has become increasingly important, as illustrated by the European Union's enaction of the GDPR in 2016, and the California Consumer Privacy Act (CCPA) in 2018 [5]. One of the tenets in GDPR is *purpose limitation*, restricting data collection to "specified, explicit and legitimate purposes". Adding the element of *intent of use* of data may impose a layer of access restriction which cannot be handled by basic credential-based access control systems or role-based access control (RBAC) systems [6]. Additionally, there may be restrictions on exporting sensitive data based on the geographical location of the requester, or the type of data to be exported. Moreover, a data source may be composed of sensitive data which could be export-restricted based on the requesting conditions, together with data which is not restricted under the same conditions. Existing access control schemes are too course-grained for handling such requirements. In contrast, FogProtect provides fine-grained, dynamic, and flexible access and usage control.

Security policy management. Security policies provide the abstraction and formalism to enforce security requirements. A Model-based Security Toolkit was proposed in [7], which is integrated in a management framework for IoT devices, and supports specification and evaluation of security policies. The applicability of the proposed model was limited to IoT architecture. Dsouza et al. [8] designed a policy management framework for fog computing where the orchestration layer of the fog architecture is supported by a policy management component that includes a repository of rules, an attribute database, and a session administrator. The policies could be enforced at various levels, but only an enforcement point functionality was evaluated without a concrete integration in a specific fog orchestration architecture. The MUSA project[1] aimed at security-intelligent lifecycle management of distributed applications over heterogeneous cloud resources. It featured security-by-design mechanisms to allow application self-protection at runtime, and methods and tools for integrated security assurance in both the engineering and operation of multi-cloud applications. The ANASTACIA project[2] focused on providing assurance security and trustworthiness by design. It designed and implemented a security framework providing autonomous decisions using software defined networking technologies and dynamic security enforcement and monitoring methodologies and tools. ANASTACIA evolved the SECURED HSPL/MSPL proposal, adapting the model of security capabilities for network security functions. Compared to these solutions, FogProtect introduces multi-domain policy delegation and supports end-to-end security management. In addition, the security models of previous research were not oriented to cloud-native environments.

Service Management & Adaptation. Existing approaches to data protection-aware application and service management have important limitations. Some approaches only address design-time activities [9] or have limited runtime capabilities, e.g., only threat detection [10]. Other approaches focus on specific security threats, without considering the broader scope of data protection [11, 12, 13], on specific sets of constraints, e.g., physical constraints [14], or on specific solutions like virtual machine migration [15]. Adaptive application management with a focus on data protection has been studied in the ATMOSPHERE[3] project, proposing to dynamically adapt the degree of anonymity of datasets based on the measured risk of re-identification [16, 17]. Their proposed solution does not consider the architectural characteristics of the system, cannot represent issues beyond anonymity, and does not take into account other system quality factors, e.g., costs or functionality.

To address these limitations, we proposed earlier an automated model-based approach to both the detection and mitigation at runtime of complex problematic system configurations posing threats to data protection, which also considers other system goals for optimization (e.g., functionality, costs) [18]. In this work, we extend that method with new modeling primitives: the *private space* type, accounting for privacy-sensitive enclosures (both physical and logical), differentiated compute types (cloud, fog, edge), and new attributes, specific to the GDPR, to model personal and sensitive data. Moreover, we modify the adaptation proposal process to leverage the Risk Management component of FogProtect for computing the impact of adaptations on the overall risk level of the system.

Risk management. The heterogeneity of the cloud-to-edge continuum brings significant risks to data generated, stored, processed and transmitted within it. Amongst others, the Common Vulnerability Scoring System (CVSS)[4] defines three key risks to data, known as the "CIA Triad": loss of *confidentiality* (e.g. data breaches, unauthorised access or leaks), loss of *integrity* (data is corrupted maliciously or accidentally) and loss of *availability* (data is not accessible to authorised users).

Cyber security risk management is commonplace in enterprises, and certification using standardised information security assurance processes is increasingly important. ISO 27001[5] provides a certification standard for checking whether identified threats are addressed by determining security risks and specifying measures that (if correctly implemented) will address those risks. Whilst these standards form a sound basis for risk analysis, the process of analysing risks is often manual and is therefore time-consuming, expensive, and error-

[1] http://www.musa-project.eu/

[2] http://www.anastacia-h2020.eu/

[3] https://www.atmosphere-eubrazil.eu

[4] https://www.first.org/cvss/v2/guide

[5] https://www.iso.org/standard/54534.html

prone. Moreover, the results of a manual analysis are rarely reproducible, due to the human value judgements needed on the relevance of given threats. The results take the form of a document set which is difficult to consult and use when a system actually comes under attack.

The need for automation of cyber security risk management spawned research in risk modelling, analysis tools, and related methods. Automated tools such as SeaMonster [19] and SecuriCad[6] consider risk management from the perspective of attacks. Other tools such as ThreatModeler[7] adopt a software-centric perspective. FogProtect includes an automated risk management toolkit that follows the *asset-centric* ISO 27005[8] methodology for cyber security risk management, which in turn supports ISO 27001 certification. ISO 27005 considers information systems as a set of assets, which enables judgements to be made about the value of the assets and the consequent impact severity if the assets were compromised. FogProtect determines the types and likelihood of threats attacking the assets, and this likelihood, combined with the impact of compromise, leads to a risk for each threat consequence (e.g., the loss of confidentiality for a particular data asset).

Summary. The fields of access and usage control, security policy management, service management and adaptation, and risk management all contain promising building blocks towards data protection in the cloud-to-edge continuum. However, each of these fields needs further research to address the complexity, dynamicity, and heterogeneity of data protection in the cloud-to-edge continuum. Moreover, approaches from these disjoint fields need to be integrated to achieve end-to-end protection of sensitive data.

III. Technical overview of FogProtect

As shown in Fig. 1, an application may involve data processing across a variety of infrastructure nodes. To protect data processed by such applications, FogProtect adds four layers of protection (see the numbering in Fig. 1):

1) **Fybrik** is a new technology for enforcing policy-aware data access, built on top of Kubernetes.
2) **Data Protection Policy Management** is responsible for enforcing data protection policies across different technology and administrative domains. It defines a language for capturing policies and orchestrates different security enablers, including Fybrik.
3) **Service Management & Adaptation** is responsible for taking data protection into account during the management of infrastructure and application services. To ensure the continued protection of data, this layer makes automated decisions on dynamic run-time adaptations of the infrastructure, the application, and the policies.
4) **Risk Management** is responsible for continually assessing data protection risks. The computed risk information is used by Service Management & Adaptation to make adaptation decisions, thereby ensuring that the risk level stays acceptable.

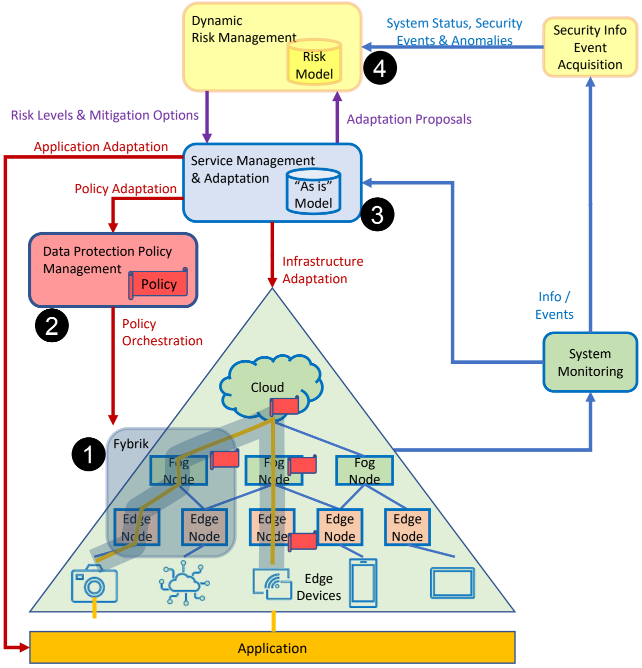


Fig. 1: FogProtect's four layers of protection

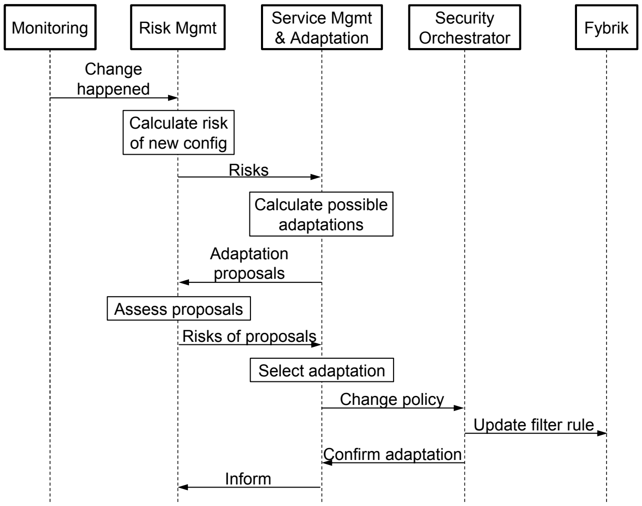


Fig. 2: Exemplary scenario

Fig. 2 shows an example of how these four layers of protection work together to maintain data protection in spite of changes in the environment. The example scenario starts with a change in the environment, e.g., a sensor reports physical tampering with a fog node. Risk Management calculates the data protection risk in the new situation. Since the tampering raises data protection risks to an unacceptable level, Risk Management informs Service Management & Adaptation. The latter determines possible adaptations for mitigation and

[6] https://www.foreseeti.com/
[7] https://threatmodeler.com/
[8] https://www.iso.org/standard/75281.html

lets Risk Management assess their implication in terms of data protection risks. This is important to avoid adaptations that introduce new risks. Based on the information received from Risk Management, Service Management & Adaptation chooses the best adaptation to mitigate the given risk. In the given example, a policy adaptation is chosen, to ensure that the tampered fog node cannot access sensitive data. The policy change is sent to the Security Orchestrator (a part of Data Protection Policy Management), which ensures that the relevant security enabler is informed. In this case, the filter rules of Fybrik are updated accordingly.

Details about each of the FogProtect components are given in the next section.

IV. Technical building blocks

A. Data usage control with Fybrik

To keep data secure, cryptographic functions can be used, for example to encrypt data-at-rest, data-in-motion (e.g. TLS) and data-in-processing (e.g. secure hardware enclaves). Applications which access data typically incorporate the usage of usernames and passwords to provide role based access control (RBAC). This approach, however, is limiting and inflexible. Not only does it put the onus of policy enforcement on each application, it also requires that each application brought into this environment (e.g. hosted environment) potentially needs to be recoded to meet the policy requirements of the environment. A change in policy means that all the applications need to be recoded and verified. Additionally, if applications need to access data in a data store which is password protected (for example, an S3 store), then the password needs to be distributed amongst all applications, posing a security risk.

Fybrik[9] is an open-source, cloud-native platform being developed by IBM to unify data access, governance and orchestration, enabling business agility while securing enterprise data. Acting as a Policy Enforcement Point (PEP), Fybrik brings together access, performance and governance for data, greatly reducing the risk of data leakage. Built on top of Kubernetes, Fybrik creates a secure fabric for the flow of data, regulating what can flow between a data user and a data source based on rules typically defined by a Data Governance Officer.

The Fybrik architecture consists of a *control plane* that takes declarative information in the form of a text file to create compute and data pipelines, and a runtime environment that encapsulates containerized workloads and intermediates the data flow in accordance with the created data pipeline. The Fybrik control plane utilizes pluggable modules that are inserted into the data flow and can handle tasks like policy-driven data redaction, credential injection to allow access to data stores, and data auditing. The configuration of data pipeline and module deployment comes from a deployment plan called a *blueprint* which is created by gathering information on:

- The data user (intent of use of data) and the required data resources.
- Data sources/assets available, taken from a data catalog.
- The policies that govern the workload and control the use of data, taken from a policy manager.
- The available infrastructure modules that can be used.

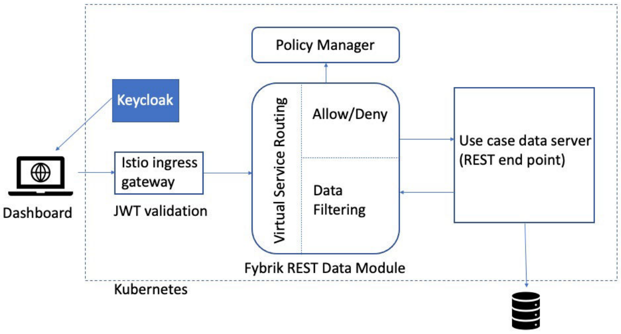


Fig. 3: FogProtect REST Fybrik Module

The FogProtect use cases use REST end points to expose their backend data stores. However, the legacy front ends (such as dashboards) have no support for access control: any user of the dashboard can access any REST end point and see all data. Using Fybrik, we were able to create a global, fine-grained access control policy to not only restrict given roles from accessing end points, but also to allow for filtering of both sensitive columns and rows in the stored data.

FogProtect provided a plugin Fybrik module which serves as a reverse proxy to the backend data source. REST end points are registered as Fybrik *Asset* resources which categorize the data (e.g., manufacturing data, HR data etc.). This categorization is used by the policies: for example, only allowing a certain role access to a specific data category. Requests from the frontend contain a JSON Web Token (JWT) which cryptographically encodes the user role (and potentially organization). The FogProtect Fybrik module verifies and decodes the JWT, and obtains the relevant policies for that user and the requested data end point. Required actions (such as blocking the end point or redacting the data stream) are then obtained from the Fybrik Policy Manager, and implemented by the module before it returns data to the requester. All REST requests for data go through a *gateway* which directs the request to the Fybrik module.

An illustration of the Fybrik REST module used in FogProtect is given in Fig. 3.

B. Data protection policy management

To provide data protection in cloud and fog environments, users' security requirements should be captured and translated into machine interpretable language. This requires a security policy formalism that supports security orchestration. This formalism needs to be abstract enough to be platform-independent to support multi-domain orchestration and needs sufficient expressiveness to specify the required security capabilities. Based on the comparative study of policy specification languages for secure distributed applications in [20] and the state of the art in approaches for policy modelling and policy

[9] https://fybrik.io/

orchestration that can be enforced in NFV (Network Function Virtualisation), IoT and SDN (Software Defined Networking) environments, MSPL (Medium Security Policy Language) [21] has been chosen in FogProtect to specify security policies, since it satisfies the mentioned requirements.

MSPL is organized by security capabilities and specifies the properties, rules, conditions, and actions associated to these capabilities. Capabilities are defined as basic features that can be configured to enforce a security policy (e.g. channel protection and filtering).

A key challenge is to automatically deploy end-to-end security policies across the network and computing continuum by selecting the *security enablers*, i.e., services or software functions that match these security policies. Categories of enablers are described in [22]. The FIWARE catalogue[10] contains examples of security enablers. Also Fybrik is an enabler. Deploying end-to-end security policies across the network and computing continuum involves the following steps.

1) Security enabler selection and composition: A set of security enablers that match the security capabilities and properties expressed in a security policy are automatically selected from a catalogue, such as authentication, authorization, channel encryption and filtering enablers. The ability of MSPL to specify dependencies between capabilities allows a security policy to be enforced by a chain of enablers.

2) Security enabler deployment: The enforcement of the security policy may require the deployment of new enablers. In that case, once the enablers corresponding to the required security capabilities are selected, the security orchestrator triggers their deployment as a chain of enablers.

3) Security enabler configuration: Deployed enablers can be selected to be part of a chain of enablers. To manage the enforcement of the security policy, enablers need to be configured (or re-configured) based on a translation of the security policy. MSPL is the policy abstraction used for expressing configurations in a platform-independent format. It is an abstract language with statements related to the typical actions of security controls (e.g., keeping track of connection status). MSPL is platform-independent and translated into a security configuration for a specific enabler. Each security enabler in the catalogue must be provided with an MSPL to platform-specific translation to automatically generate the enabler security rules and property configurations. In our proof-of-concept implementation, the deployment platform is Kubernetes, hence the orchestrator uses Kubernetes APIs for deployment and configuration of enablers. The format of the configuration pushed through the API is enabler-specific.

4) Multi-domain support: A special challenge is posed by the fragmented context where digital infrastructures are shared by several tenants with different business requirements, and operated by operators who use heterogeneous software provided by different technology providers. In that context, each operator has its orchestrator and the orchestration process is distributed. Most of the time, the deployment of micro-services is handled at the level of each cluster, and a degree of autonomy is preserved at that level. Therefore, the distribution of security orchestration is crucial to guarantee the interoperability of end-to-end security policies between operators.

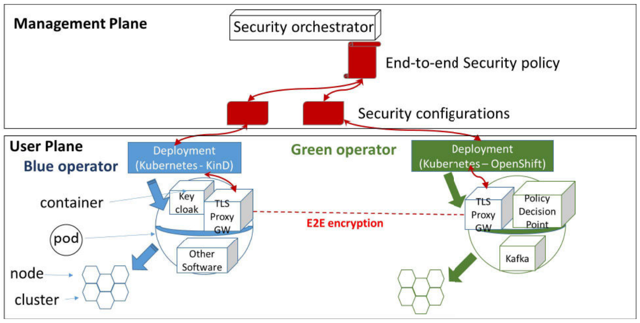


Fig. 4: End-to-end security policy orchestration

Fig. 4 shows a blue and a green operator, each of them managing a cluster of nodes, and providing an orchestrator to deploy software and configurations of a Kubernetes pod. A security policy related to a tenant service and specifying the need to deploy an end-to-end encryption capability based on two Service Mesh gateways (TLS proxies in this example) has to be orchestrated. To deploy this trans-operator security policy, a security orchestrator at the management layer deploys the two enablers through the corresponding operator orchestrators (if not yet deployed), one on the blue operator, the other on the green operator, establishes an encrypted channel to exchange data, and configures the enablers based on the security policy (certificate parameters, key size, etc.).

C. Service Management & Adaptation

The role of Service Management & Adaptation is to mitigate threats to data protection at runtime by means of automated system reconfiguration. Our approach leverages techniques from models@run.time [23], graph pattern matching and transformations, and search-based optimization to compute optimal adaptation strategies [18]. When applied to the managed system, the adaptations mitigate the threats to data protection, while leading to optimal costs, functionality, or energy consumption. The key features and building blocks are as follows.

The **meta-model**, defined at design-time, specifies the allowed node types, their attributes, and the possible relations between nodes. The meta-model is based on our extended version of the Topology and Orchestration Specification for Cloud Applications (TOSCA) modeling language [24].

The **"As is" model** is a concrete instantiation of the meta-model. The "As is" model is a continuously updated runtime artefact depicting the current state of the system managed by FogProtect and it represents the basis for automated decision-making in Service Management & Adaptation.

The **problematic configuration patterns (PCPs)**, defined at design-time, describe configurations of the managed system which can pose significant threats to data protection [25]. The PCPs are linked to the risks identified by Risk Management,

[10] https://www.fiware.org/developers/catalogue/

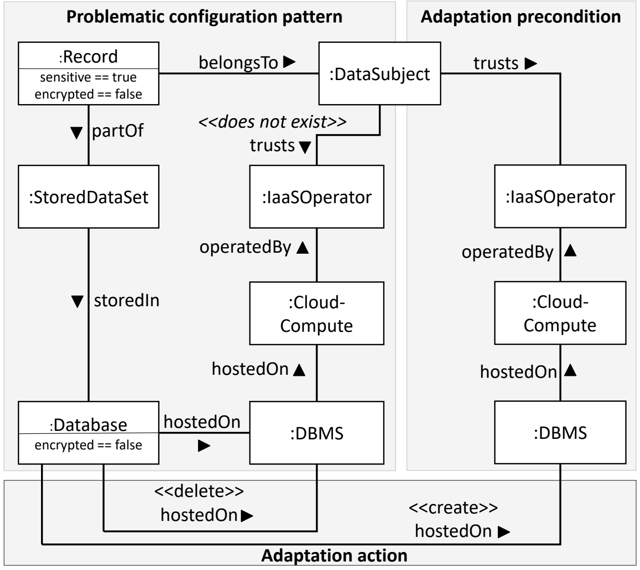
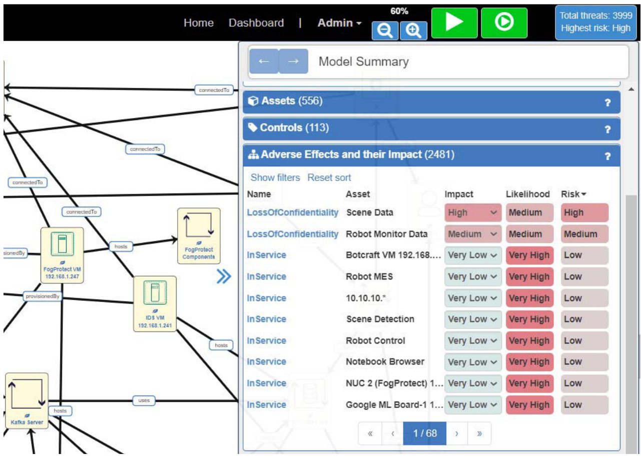


Fig. 5: Example of an adaptation rule, comprising the mitigated PCP, the adaptation precondition, and the adaptation action. The adaptation rule involves the migration of a database from an untrusted infrastructure provider to a trusted one.



Assets (556)

Controls (113)

Adverse Effects and their Impact (2481)

Show filters Reset sort

Name	Asset	Impact	Likelihood	Risk
LossOfConfidentiality	Scene Data	High	Medium	High
LossOfConfidentiality	Robot Monitor Data	Medium	Medium	Medium
InService	Botcraft VM 192.168.…	Very Low	Very High	Low
InService	Robot MES	Very Low	Very High	Low
InService	10.10.10.*	Very Low	Very High	Low
InService	Scene Detection	Very Low	Very High	Low
InService	Robot Control	Very Low	Very High	Low
InService	Notebook Browser	Very Low	Very High	Low
InService	NUC 2 (FogProtect) 1…	Very Low	Very High	Low
InService	Google ML Board-1 1…	Very Low	Very High	Low

« ‹ 1 / 68 › »

Fig. 6: Risk modelling user interface

the goal of Service Management & Adaptation being that of lowering the overall risk within the managed system by removing, through adaptation, all instances of PCPs.

The **adaptation rules**, also defined at design-time, describe the *actions* to be carried out at the level of the “As is” model and the *preconditions* under which the respective rule can be applied. Adaptation rules are associated to PCPs, with possibly more than one adaptation rule having the potential to mitigate a given PCP. Available adaptation actions include creating relations, deleting relations, and setting node attributes. Adaptation preconditions represent constraints the “As is” model must satisfy for the adaptation rule to be applicable. Preconditions may require that certain attributes match some reference values or that certain nodes and relations are present in (or absent from) the “As is” model. Fig. 5 depicts the relationship between PCPs, adaptation actions, and preconditions in the example of a virtual machine (VM) migration adaptation rule. Since adaptation rules operate at the level of the “As is” model, an additional step is required to translate adaptation actions to concrete changes in the managed system. To this end, Service Management & Adaptation features a **policy translation unit** for converting “As is” model-level changes to *security policy* changes, which are then sent to the policy orchestrator.

The **adaptation process** starts when Risk Management detects that the overall risk to data protection in the managed system is unacceptably high. Service Management & Adaptation then undertakes the following actions:

1) identifies PCP instances in the “As is” model,
2) searches for sequences of applying adaptation rules that mitigate the identified PCP instances (wherein each adaptation rule must meet its preconditions),
3) sorts the discovered adaptations in terms of their impact on system characteristics, e.g., the operating costs or functionality level of the resulting system configurations,
4) requests from Risk Management the evaluation of the potential post-adaptation system configurations,
5) applies the *best* adaptation rule sequence that lowers the overall risk to data protection to an acceptable level; if no sequence of adaptation rules can sufficiently lower the risk, Service Management & Adaptation requests the input of a human operator via a web-based frontend.

The discovery of adaptation rule sequences (Step 2 of the above process) is achieved by means of a **search-based** algorithm. The algorithm explores the “As is” model configuration space by hypothetically applying chains of adaptation rules starting from the current configuration of the “As is” model. If one of the resulting configurations is PCP-free, then the sequence of adaptation rules leading to that configuration is retained in a list, which is then used as input to Steps 3–5.

For identifying the *best* adaptation rule sequence, Service Management & Adaptation carries out an analysis of the potentially resulting “As is” model configurations. A *score* is computed for each potential configuration $M \in \mathcal{M}$:

$$Score(M) = Pref(F(M), C(M), E(M))\,. \qquad (1)$$

Here, F, C, $E : \mathcal{M} \to \mathbf{R}$ compute partial score values for functionality, cost, and energy consumption. $Pref : \mathbf{R^3} \to \mathbf{R}$ is a *value function* implementing the preference ordering between individual system-level characteristics, e.g., functionality might take priority over energy consumption and costs. Individual system characteristics, i.e. functionality, costs, and energy consumption, are computed based on structural characteristics of M or the values of certain node attributes.

D. Risk management

FogProtect uses a semi-automated approach for risk identification and analysis based on the System Security Modeller (SSM), a security risk analysis tool developed at the University of Southampton IT Innovation Centre in the OPTET project

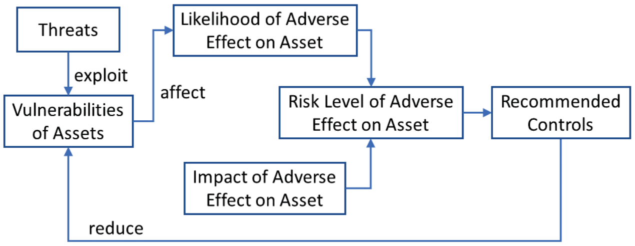

Threats
exploit
Vulnerabilities of Assets
affect
Likelihood of Adverse Effect on Asset
Risk Level of Adverse Effect on Asset
Recommended Controls
Impact of Adverse Effect on Asset
reduce

Fig. 7: Asset-based risk management

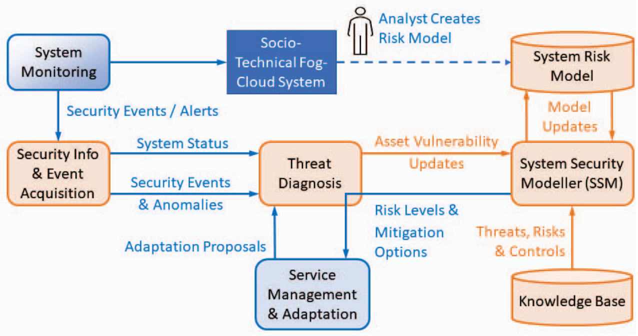

Analyst Creates Risk Model
System Monitoring
Socio-Technical Fog-Cloud System
System Risk Model
Security Events / Alerts
Model Updates
Security Info & Event Acquisition
System Status
Security Events & Anomalies
Threat Diagnosis
Asset Vulnerability Updates
System Security Modeller (SSM)
Risk Levels & Mitigation Options
Threats, Risks & Controls
Adaptation Proposals
Service Management & Adaptation
Knowledge Base

Fig. 8: Threat Diagnosis & Risk Analysis in Context

[26] and continuously enhanced ever since [27, 28, 29]. Its user interface is illustrated in Fig. 6, showing an example risk model. Individual risks (adverse effects) are shown at the lower right, with the most severe risk at the top of the list.

SSM follows the ISO27005 asset-centric risk methodology and supports modelling socio-technical systems. Assets are tangible and non-tangible items of value, e.g. software, data, machinery, services, people; and clearly data is a core focus of FogProtect. As shown in Fig. 7, assets may have vulnerabilities, which can expose them to attack by threats that cause adverse effects in the asset (unwanted, erroneous or dangerous behaviour). The risk to the asset is the severity of the adverse effect combined with the likelihood of the threat that causes it. Controls may be applied to reduce the likelihood of the threat, and therefore the risk of the associated adverse effects.

Using SSM, an analyst creates a graphical “system risk model” of socio-technical assets and their relationships. The tool combines this model with a built-in machine understandable security knowledge base to find potential threats exposed by vulnerabilities in the assets, assesses the overall risk, and recommends countermeasures to address the threats [28].

Risk assessment was previously applied at design time. The advance of FogProtect is to support dynamic risk evaluation at runtime, providing the key benefit of continuous, event-driven risk assessment for data in the cloud-to-edge continuum. This is facilitated via adjustment of parameters in the system model representing asset vulnerabilities, resulting from security alerts detected by security monitoring scanners such as Wazuh[11]. This is enabled by the Security Info & Event Acquisition (SIEA) and Threat Diagnosis components (see Fig. 8). The SIEA aggregates inputs from different security monitors & scanners and passes events from them to the Threat Diagnosis component. The Threat Diagnosis component provides a runtime API to the SSM risk management, maps security alert events to system risk model vulnerability updates, and triggers dynamic recalculation of risk based on these updates. This enables the risk analysis to be run in several scenarios:

1) At design time, when a system security model is built by an expert. The risk assessment in SSM shows the risk level and can indicate controls to bring risks to an acceptable level. The expert can decide which controls to implement and update the model accordingly.
2) At runtime, when security alerts are detected by monitors. The alerts are aggregated by the SIEA and passed to the Threat Detection, which translates them into a common representation of asset vulnerability, and triggers a risk recalculation, resulting in a new risk level.
3) At runtime, when adaptations are proposed by Service Management & Adaptation, representing different combinations of security controls that could be applied. Each option is evaluated and the resulting risk level for each is fed back to Service Management & Adaptation to inform its choice from the alternative options.
4) At runtime, when an adaptation proposal is enacted by Service Management & Adaptation. This is an actual update of the real system configuration, and the risk model is updated to reflect the changes and risk calculation is rerun to determine the new baseline risk level.

The runtime scenarios (2–4) represent a dynamic cycle of security alert detection, risk evaluation resulting from the alerts, evaluation of options to address the security alerts, and selection of one option, resulting in a new quiescent risk level.

V. Validation

A. *Smart cities use case*

A network of CCTV cameras monitors selected places of a city to obtain insights about the urban environment. Ubiwhere equips smart lampposts (modular lampposts supporting cameras, small cell antennas, EV chargers – see Fig. 9) with fog nodes that process videos recorded by the cameras to identify objects and anonymise sensitive data by blurring faces and license plates. Sensitive data is processed before sending it to the cloud, helping preserve citizens’ trust in the system. Since street furniture is vulnerable to physical attacks and other severe conditions, it is crucial to implement the right tools to protect the data within the system.

Citizens can report incidents in the urban environment using a mobile application to Ubiwhere’s Urban Platform hosted in the cloud. City operators can request the video footage of the location of the incident, with the Urban Platform getting the requested video from the relevant fog node. According to the defined policies, different users can access different types of information: the original video, the anonymised video (see Fig. 10), or inferred data (e.g. the number of people and vehicles

[11] https://wazuh.com/

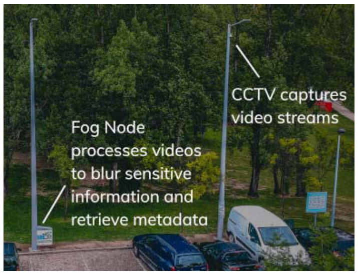

CCTV captures video streams
Fog Node processes videos to blur sensitive information and retrieve metadata

Fig. 9: Smart Cities use case scenario in FogProtect

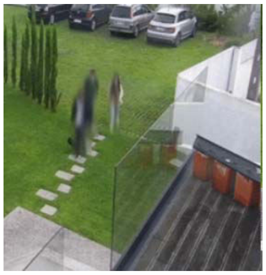


Fig. 10: Video blurred by a fog node

captured on video at the given time). Role-based access control is performed: *Law Enforcement Agents* can access all types of data; *City Managers* can access the blurred video and the metadata; *City Analysts* can only access the inferred data for their urban planning activities.

In the case of a physical attack on the fog node (e.g. someone forcing open the box containing the processing unit), the integrity of the data coming from this node can no longer be trusted. Thus, when such an attack happens, a magnetic sensor sends a message to the FogProtect system. This message is received as a monitoring event and forwarded, through the Security Information & Event Acquisition component, to Risk Management (see Fig. 2). Here, the overall risk of the system is evaluated, considering the new threat. This is reported to Service Management & Adaptation, which is responsible for devising an adaptation to react to the event. In this case, as the fog node is no longer trusted, the adaptation forbids communication with the fog node. The adaptation is communicated to Data Protection Policy Management, which is responsible for choosing the correct policy to be implemented. The policy is sent to Fybrik which, from now on, enforces the new policy. That is, Fybrik denies any request to the tampered fog node, prohibiting the use of any corrupted data.

When the fog node is repaired, an administrator triggers a "Clearance" event, which goes through all the steps previously described, but in this case it lets the FogProtect components know that the node is trustable again. As a result, FogProtect restores access to the fog node.

B. Smart manufacturing use case

This use case is about restricting the display of personal data and other critical factory data on a dashboard, based on the user who is currently logged in and on other context information. A factory production line consisting of various IoT devices, production machines like 3D printers and robots and a software stack for data storage and forwarding has been created in a laboratory, which is hosted in a 20ft freight container, called Factory-in-a-Box (FiaB, Fig. 11). FogProtect is used to retrofit a smart factory to enhance traditional network security to comply with new data protection requirements.

The data displayed on the factory dashboard is filtered by FogProtect. The applied filters depend on the role of the user currently logged in to the dashboard. *Managers* can see all data (technical robot data as well as data about persons in the factory). *Technicians* can see technical data and the total number of persons in the factory, but cannot see more fine-grained personal data, such as how many persons are in the critical zone in front of the robot. *Human Resources* cannot see technical data but has full access to data about persons in the factory.

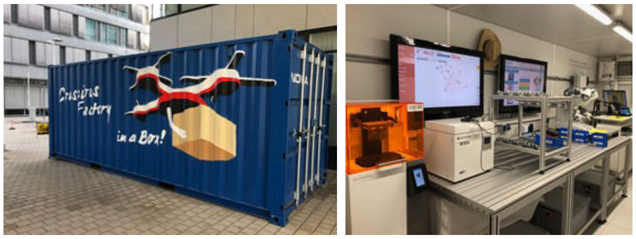


Fig. 11: Factory-in-a-Box

The use case also shows how confidentiality of factory data is protected by FogProtect. If the factory door is opened without prior authorization, FogProtect shuts down all connections between the dashboard and the data sources, to ensure that no critical data is displayed on the screen in the presence of a potential intruder. When the door is closed again, the emergency situation is considered partially resolved, and FogProtect partially restores the information flow of non-personal technical data to the dashboard. Only after a manual clearance by a security operator is the situation considered fully resolved, enabling the display of more sensitive data (personal data from video analytics), depending on the role (i.e., the authorization rights) of the user.

The "door open", "door closed" and "clearance" events are dispatched by a message broker to FogProtect's System Monitoring component, which forwards these events via Security Information & Event Acquisition to Risk Management (see Fig. 2). Risk Management and Service Management & Adaptation perform a handshake, resulting in an adaptation decision. For the "door open" event, the above mentioned emergency situation forces Risk Management to bypass most of the interactions with its System Security Modeller to reduce the latency of a mitigation decision. When a decision is made by Service Management & Adaptation, Data Protection Policy Management adapts the policy of the Fybrik instance impacted by the threat. The "door closed" and "clearance" events relax the restrictions imposed previously. In these cases, the adaptation aims not at mitigation, but at maximizing access to data endpoints without exceeding the tolerated risk threshold. That is, Service Management & Adaptation chooses an adaptation improving functionality, while ensuring that data is protected. Fig. 12 depicts the 3 possible states of the FiaB.

C. Smart media use case

In this use case, a journalist broadcasts questions, to which citizens respond. The paradigm is known from social media platforms, but is specific to news media with attendant issues

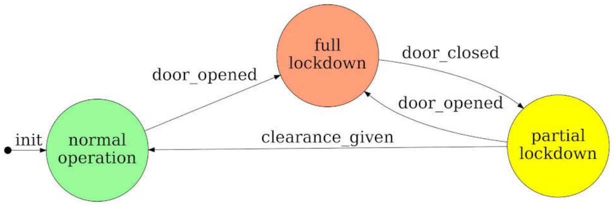


Fig. 12: FiaB states tracked by SIEA

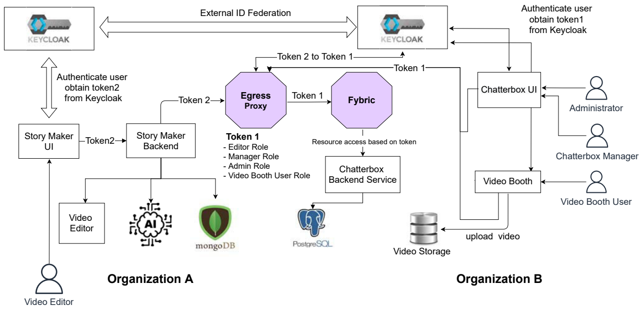
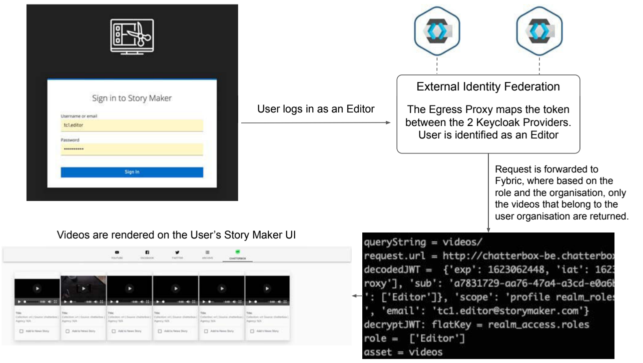


Fig. 13: Deployment overview of the Smart Media use case


Sign in to Story Maker
Username or email
tc1.editor
Password
Sign In
User logs in as an Editor
External Identity Federation
The Egress Proxy maps the token between the 2 Keycloak Providers. User is identified as an Editor
Request is forwarded to Fybric, where based on the role and the organisation, only the videos that belong to the user organisation are returned.
Videos are rendered on the User's Story Maker UI

```
queryString = videos/
request.url = http://chatterbox-be.chatterbox
decodedJWT =  {'exp': 1623062448, 'iat': 162
roxy'], 'sub': 'a7831729-aa76-47a4-a3cd-e0a6
': ['Editor']}, 'scope': 'profile realm_roles
', 'email': 'tc1.editor@storymaker.com'}
decryptJWT: flatKey = realm_access.roles
role =  ['Editor']
asset = videos
```

Fig. 14: Sequence of steps in the Smart Media use case

of provenance checking and protecting basic rights, including the secure handling of special category personal data.

An editorial team defines a survey using the Chatterbox Manager application. The survey consists of a series of questions that can be displayed in an automated sequence in a video booth. Citizens can use the equipment in the video booth (or a web application) to submit their video answers via a fog node. The submitted video answers are analysed with AI-based services. Metadata is extracted from the video and other context information, either in the cloud or in the fog node.

A video editor from another organisation can request access from the data controller of the ChatterBox data to browse using the metadata and download the video to use it in a story of their own. Access to a particular resource depends on the role of the user and the organisation the user belongs to.

Initially, the Smart Media ecosystem comprises two organisations: VRT (a public broadcaster) and ATC (a company working on media services). VRT uses some of its video booth resources on its own, and provides a video booth to ATC.

The main applications are Chatterbox Manager and Story Maker (see Fig. 13). The first one belongs to VRT and provides access to various resources (REST API / GraphQL endpoints); the latter belongs to ATC and it is an external application requesting access to resources (request of videos).

The main roles in the scenario are: VRT Administrator, VRT Manager, ATC User, and Video Booth User. Resource access control is set up so that the VRT Administrator can access all resources, both REST and GraphQL, the VRT Manager can access all REST resources and hence all video material stored on VRT premises, and the ATC User can access only the video material originating from a booth occupied by ATC.

Authentication of the users of both organisations is implemented via a respective Keycloak instance. The FogProtect Data Protection Policy Management uses Fybrik and a specific security enabler called Federation Proxy. The latter acts as external ID provider, federating the external organisation identities and assigning the user role. This information is used to form the user's unique token. The identity federation feature of the Federation Proxy enables automation of identity management processes, reduction of operational costs and complexity, and improved user experience. It also enhances enterprise security and improves business-to-business interactions.

Fybrik acts as a resource access management layer, filtering all incoming requests towards the Chatterbox Manager based on user role and organisation. Thanks to Fybrik, no changes are needed on application level to define new user roles to an existing system. Hence, granting and denying access to endpoints is decoupled from the application's logic. The addition of data filtering capabilities ensures that a client application doesn't need to be given a specific endpoint.

If a request originates from an external organisation (ATC), the Federation Proxy is triggered to federate the user identity (see Fig. 14). The user role assigned by the Federation Proxy is "editor". The request is forwarded to Fybrik which checks the user role and organisation. Based on that, it filters and returns only material belonging to ATC. Unauthorised access (unfederated user or user with undefined role) is denied.

VI. Conclusions

We presented FogProtect, an integrated approach to protect data in the cloud-to-edge continuum. FogProtect combines data usage control using Fybrik, data protection policy management, service management & adaptation, and risk assessment. FogProtect supports the whole cycle from vulnerability sensing through risk assessment to adaptations by automated update and distribution of security policies. FogProtect's wide applicability was demonstrated in three real-world use cases. The Smart City use case shows how FogProtect automatically activates controls to protect integrity after detecting a physical attack to an edge device. The Smart Manufacturing use case shows the addition of fine-grained access control based on user roles and context information to an existing factory information system. Finally, the Smart Media use case features data protection across organization boundaries.

References

[1] D. Kimovski, R. Mathá, J. Hammer, N. Mehran, H. Hellwagner, and R. Prodan, "Cloud, fog or edge: Where to compute?" *IEEE Internet Computing*, vol. 25, no. 4, pp. 30–36, 2021.

[2] Z. Á. Mann, A. Metzger, J. Prade, and R. Seidl, "Optimized application deployment in the fog," in *Intl. Conf. Service-Oriented Computing*, 2019, pp. 283–298.

[3] P. Zhang, M. Zhou, and G. Fortino, "Security and trust issues in fog computing: A survey," *Future Generation Computer Systems*, vol. 88, pp. 16–27, 2018.

[4] D. Ayed, E. Jaho, C. Lachner, Z. Á. Mann, R. Seidl, and M. Surridge, "FogProtect: Protecting sensitive data in the computing continuum," in *Advances in Service-Oriented and Cloud Computing. ESOCC 2020*, 2021, pp. 179–184.

[5] M. Kaminski, "A recent renaissance in privacy law," *CACM*, vol. 63, no. 9, pp. 24–27, 2020.

[6] P. Zhang, J. K. Liu, F. R. Yu, M. Sookhak, M. H. Au, and X. Luo, "A survey on access control in fog computing," *IEEE Comm. Mag.*, vol. 56, no. 2, pp. 144–149, 2018.

[7] R. Neisse, G. Steri, I. N. Fovino, and G. Baldini, "SecKit: A model-based security toolkit for the Internet of Things," *Comput. Secur.*, vol. 54, pp. 60–76, 2015.

[8] C. D'Souza, G. Ahn, and M. Taguinod, "Policy-driven security management for fog computing: Preliminary framework and a case study," in *IRI*, 2014, pp. 16–23.

[9] A. Alebrahim, D. Hatebur, S. Faßbender, L. Goeke, and I. Côté, "A pattern-based and tool-supported risk analysis method compliant to ISO 27001 for cloud systems," *Intl. J. Secure Softw. Eng.*, vol. 6, no. 1, pp. 24–46, 2015.

[10] L. Pasquale, S. Hanvey, M. Mcgloin, and B. Nuseibeh, "Adaptive evidence collection in the cloud using attack scenarios," *Comput. Secur.*, vol. 59, pp. 236–254, 2016.

[11] S. Iannucci and S. Abdelwahed, "Model-based response planning strategies for autonomic intrusion protection," *ACM TAAS*, vol. 13, no. 1, 2018.

[12] G. Puppala and S. K. Pasupuleti, "Dynamic security risk assessment in cloud computing using IAG," in *Progress in Computing, Analytics and Netw.*, 2018, pp. 105–116.

[13] N. Khakpour, C. Skandylas, G. S. Nariman, and D. Weyns, "Towards secure architecture-based adaptations," in *SEAMS'19*, 2019, pp. 114–125.

[14] C. Tsigkanos, L. Pasquale, C. Ghezzi, and B. Nuseibeh, "On the interplay between cyber and physical spaces for adaptive security," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 3, pp. 466–480, 2018.

[15] M. Nguyen, P. Samanta, and S. Debroy, "Analyzing moving target defense for resilient campus private cloud," in *IEEE CLOUD*, 2018, pp. 114–121.

[16] F. Brasileiro, A. Brito, and I. Blanquer, "ATMOSPHERE: Adaptive, trustworthy, manageable, orchestrated, secure, privacy-assuring, hybrid ecosystem for resilient cloud computing," in *DSN-W*. IEEE, 2018, pp. 51–52.

[17] T. Basso, H. de Oliveira Silva, L. Montecchi, B. B. N. de França, and R. L. de Oliveira Moraes, "Towards trustworthy cloud service selection: monitoring and assessing data privacy," in *Anais do XX Workshop de Testes e Tolerância a Falhas*, 2019, pp. 7–20.

[18] Z. Á. Mann, F. Kunz, J. Laufer, J. Bellendorf, A. Metzger, and K. Pohl, "RADAR: Data protection in cloud-based computer systems at run time," *IEEE Access*, vol. 9, pp. 70 816–70 842, 2021.

[19] P. H. Meland, D. G. Spampinato, E. Hagen, E. T. Baadshaug, K.-M. Krister, and K. S. Velle, "SeaMonster: Providing tool support for security modeling," in *Norsk informasjonssikkerhetskonferanse, NISK*, 2008.

[20] S. Duflos, G. Diaz, V. Gay, and E. Horlait, "A comparative study of policy specification languages for secure distributed applications," in *Intl. Workshop Distrib. Syst.: Operations and Management*, 2002, pp. 157–168.

[21] S. Sicari, A. Rizzardi, D. Miorandi, C. Cappiello, and A. Coen-Porisini, "Security policy enforcement for networked smart objects," *Computer Networks*, vol. 108, pp. 133–147, 2016.

[22] G. Arfaoui *et al.*, "A security architecture for 5G networks," *IEEE Access*, vol. 6, 2018.

[23] N. Bencomo, S. Götz, and H. Song, "Models@ run. time: a guided tour of the state of the art and research challenges," *Software & Systems Modeling*, vol. 18, no. 5, pp. 3049–3082, 2019.

[24] J. Bellendorf and Z. Á. Mann, "Specification of cloud topologies and orchestration using TOSCA: a survey," *Computing*, vol. 102, no. 8, pp. 1793–1815, 2020.

[25] S. Schoenen, Z. Á. Mann, and A. Metzger, "Using risk patterns to identify violations of data protection policies in cloud systems," in *Service-Oriented Computing – ICSOC 2017 Workshops*. Springer, 2018, pp. 296–307.

[26] N. G. Mohammadi, T. Bandyszak, A. Goldsteen, C. Kalogiros, T. Weyer, M. Moffie, B. I. Nasser, and M. Surridge, "Combining risk-management and computational approaches for trustworthiness evaluation of socio-technical systems," in *CAiSE Forum*, 2015, pp. 237–244.

[27] M. Surridge, G. Correndo, K. Meacham, J. Papay, S. C. Phillips, S. Wiegand, and T. Wilkinson, "Trust modelling in 5G mobile networks," in *Workshop on Security in Softwarized Networks*, 2018, pp. 14–19.

[28] M. Surridge, K. Meacham, J. Papay, S. C. Phillips, J. B. Pickering, A. Shafiee, and T. Wilkinson, "Modelling compliance threats and security analysis of cross border health data exchange," in *Intl. Conf. on Model and Data Engineering*. Springer, 2019, pp. 180–189.

[29] N. G. Mohammadi, L. Goeke, M. Heisel, and M. Surridge, "Systematic risk assessment of cloud computing systems using a combined model-based approach." in *ICEIS (2)*, 2020, pp. 53–66.