Leveraging Dynamic Information for Identity and Access Management:
An Extension of Current Enterprise IAM Architecture

Alexander Puchta1, Sebastian Groll2 and Günther Pernul2

1Nexis GmbH, Franz-Mayer-Str. 1, 93053 Regensburg, Germany
2Chair of Information Systems, University of Regensburg, Universitätsstraße 31, 93053 Regensburg, Germany

Keywords: Identity and Access Management, IAM, Access Control, Architecture, Big Data, Stream Data, Real Time Analysis.

Abstract: Identity and access management (IAM) functions as a core component for today’s enterprises managing digital identities and their access to resources. However, IAM systems are quite isolated from other applications with useful information resulting in individual data pots. By interconnecting these systems, important information on relevant IAM entities like criticality or usage information can be additionally gathered for further improvement. Current IAM landscapes within enterprises are not prepared for such challenges as the data needs to be harmonised, analysed, and verified. Within this work a state-of-the-art IAM architecture in enterprises and existing shortcomings are defined. Based on these, an extended IAM architecture scheme is proposed and described in detail. Key component is the integration of additional information sources for mutual benefit in IAM and external applications. Finally, the approach is applied to two use cases based on real data. They originate from our conducted IAM projects and show the feasibility of the proposed architecture.

1 INTRODUCTION

Identity and access management (IAM) has increasingly become a core component in today’s enterprise IT as it is vital to fulfil internal and external requirements and obligations. SOX (SOX, 2002) or Basel III (Basel Comittee on Banking Supervisions, 2010) require the application of various processes, policies and technical solutions. This also includes a change of the current IAM architecture as it is mainly centralised around one IAM system being connected to managed applications. IAM currently is rarely and if then only partially linked to other important systems like SIEM or real time information systems. However, there potentially exists a great mutual benefit in tighter connections of IAM with further information sources. On the one hand IAM systems may retrieve additional information which is helpful for existing IAM processes like access reviews, also known as recertification (Osmanoglu, 2013). On the other hand analytical findings inside the IAM system are also valuable to be automatically exported into other systems to enhance quality and process duration.

Within this work, we want to define how existing information can be integrated in IAM, analysed and exported back into other systems if needed. In order to reach this goal, we focus on the following three research questions:

- **RQ-1.** What does a current IAM architecture look like and which shortcomings exist?
- **RQ-2.** What could be an extended IAM architecture for better integration of dynamic information from external (= non-IAM related) systems?
- **RQ-3.** How can the extended IAM architecture be implemented and is there a practical applicability?

By answering these research questions our work provides two main contributions. We propose a generic IAM architecture based on scientific research and experience from IAM practitioners which describes how the integration of external information (= information from external systems) can be carried out. Besides that, we show exemplary solutions how the architecture can be implemented and demonstrate the feasibility by applying anonymised real world IAM data to the approach.

The remainder of this work is structured as follows. Section 2 introduces the relevant background on IAM and related work as well as the definition of a current IAM architecture (RQ-1). Within Section 3 we define the extended architecture and give an overview of the relevant building blocks to answer
Section 4 gives a more fine-grained view of each building block (RQ-3). We demonstrate the feasibility of our approach within Section 5 with several use cases based on real world data from our IAM project work (RQ-3). Section 6 concludes our work and highlights possible future research directions as well as current limitations.

2 BACKGROUND & RELATED WORK

Within this section the key concepts of IAM are defined. Furthermore, we introduce a typical state-of-the-art IAM architecture which can be commonly found in practice and show related work regarding the integration of dynamic information into IAM. We refer to dynamic information in this work as all kind of information being not exclusively IAM data from applications. This comprises static data as well as continuous information flows (e.g. data streams).

2.1 Background on IAM and IAM Architecture

When speaking of IAM it primarily supports two purposes: Managing identities and granting them access to resources (e.g. files or applications). Within the scope of this paper we are assuming an enterprise focused IAM limited to a single company’s context.

Based on their status in the identity management process, access control is applied to identities to regulate access to documents, applications or other information (Samarati and de Vimercati, 2000). In order to achieve this, various access control models are enforced in IAM. One of the most common types is role-based access control (RBAC) (Sandhu et al., 1996). Today it is one of the most commonly found access control models in IAM systems but remains a quite static approach (Kunz et al., 2019). By leveraging identities’ attributes and predefined policy rules, attribute-based access control (ABAC) enables enterprises to apply a more dynamic access management (Hu et al., 2014). Please note that there exist various other access control models. However, for the sake of clarity we focus only on the core models RBAC and ABAC as these are also the most common models in practice.

Enterprises employ large pieces of software, so called IAM systems like SailPoint or OneIdentity to manage thousands of identities. As such IAM systems are a key component of IAM, enterprises often leverage a centralised system architecture. The connection to other systems being no part of the IAM cosmos is typically rather loose. For example it may only consist of one monitoring system connecting to the IAM system database to read out logs. Thus, the company can react in the case of a failure (e.g. data could not be loaded into the IAM system).

However, such approaches leave IAM isolated from other relevant information systems like SIEM systems or data streams of connected applications. Within this paper we want to propose a new architecture and procedure to leverage such existing information. This improves the system as a whole. On the one hand additional information relevant for IAM can be retrieved from the IAM system. On the other hand critical information which may be generated within IAM systems via data analysis is vital for other processes can be exported to connected systems. Thus, SIEM systems may get information of a possible data breach faster, resulting in an overall improved performance of the company’s IT processes.

2.2 Related Work

Regarding access control a large body of literature can be found. When it comes to defining various access control models, like the previously mentioned RBAC (Sandhu et al., 1996) or ABAC (Hu et al., 2014) models, there is a huge variability in existence. However, as this is only one part of IAM, our work is not predominantly focused on access control itself. When it comes to IAM-specific topics not as much literature can be found. Literature is mainly focused on specific problems within IAM, like role mining or modelling (Colantonio et al., 2012; Elliott and Knight, 2010). Regarding the integration of dynamic information there already exists some fundamental work, however it is not focused on defining an overarching architecture but is more focused on the introduction of attribute-based approaches in enterprises (Hummer et al., 2015; Hummer et al., 2016; Kunz et al., 2015).

To the best of our knowledge none of these or other existing approaches define a suitable architecture for enterprise IAM in order to connect IAM with external applications or mechanisms to gain mutual benefits by leveraging external information. Our architecture comprises of several building blocks with different use cases in mind. In advance to this work we already published related work on various of these building blocks (Kunz et al., 2019; Nuss et al., 2018; Puchta et al., 2019). Within this work we aim to span the arch around these individual topics to generate a unified architecture to interconnect IAM with external systems.
3 ARCHITECTURE DEFINITION & OVERVIEW

In this section we define the current shortcomings of IAM architecture and today’s IAM in general more formally as this represents the baseline for our improved approach. Based on this, we give an overview of our proposed extension of an IAM architecture and the underlying building blocks.

3.1 Current Shortcomings of IAM Architecture

As mentioned before, current IAM architecture is centred around an IAM system connected to relevant applications. This connection is predominantly limited to database access to retrieve or change IAM relevant data (e.g. account and entitlement information). There may also be a loose connection to other systems not in the IAM cosmos. However, based on our practical experience this is mainly a one-way connection. Additionally, we conducted interviews with three different IAM consultants having four to 15 years experience with IAM projects in order to verify this structure.

Such architecture results in several shortcomings for current IAM solutions when it comes to integration of external information. IAM systems have quite static procedures as they are mainly based on an import and export cycle of assignments to the respective connected systems. This often results in relevant information missing or being erroneous in IAM systems (Kunz et al., 2019). Examples for this would be the lack of data on responsibilities or criticality of entitlements. Besides that, a further relevant information is often not present within current IAM systems, namely usage of entitlements per (individual) account. Such usage information is especially of high importance, when it comes to decision making within IAM (e.g. access reviews) as shown in Section 5. Because decisions are often made from managers responsible for employees or entitlements they need to rely on quickly comprehensible information.

Besides the points stated above, further literature can be found defining shortcomings or challenges within the current IAM generation. In one of our previous works we already analysed existing IAM literature and practice to derive relevant IAM challenges (Puchta et al., 2019). Please note that we do not refer to each individual underlying literature source for the sake of brevity. The relevant literature per challenge can be taken from the referenced work. These challenges may also be interpreted as shortcomings of current IAM architecture. Namely the missing priority of privacy within IAM, heterogeneity of IAM data, data quality management, and the transformation from an RBAC centred approach to an attribute based one.

All these shortcomings indicate that the approach to IAM in general needs to be revised. When speaking of enterprise IAM this can mainly be achieved via a profound architecture as a basis for the technical implementation.

3.2 Overview of the Extended IAM Architecture

Our proposed extended IAM architecture can be found in Figure 1. We added a central component to handle the input of IAM and external data, in the following it is called InterConnectivity Module (ICM). The arrows indicate the information flow within our generic IAM architecture. Please note that the ICM could also be an integral part of the IAM system and must not be a standalone component. However, for the sake of clarity we treat them as two different parts of the architecture. The ICM is used as a mediator between connected IAM applications, the IAM system, and external applications. Here information from connected and external applications is harmonised to allow for a homogeneous data flow to the IAM system. Furthermore, data quality approaches can be applied to the imported data (Kunz et al., 2019).

As data may need to be combined, the connected applications are not directly connected to the IAM system but are rather communicating via the ICM with the IAM system. The ICM may need to convert this information into a suitable format. Where possible, information is represented as an attribute of the respective IAM entity to enable the application of ABAC (Hu et al., 2015). However, when it comes to pure IAM operations (e.g. provisioning of entitlements), the IAM system is still able to perform such actions independently from the ICM to guarantee performance of the basic IAM operations.

When exporting information to external applications the ICM handles these operations as well. On the contrary to current IAM architecture there is a bidirectional interface in existence. This also enables the IAM system to receive a response from external applications. For this, a use case would be an automatically started emergency process to deactivate a specific user in case a SIEM system detects an insider attack (Anderson et al., 2007).

Regarding user experience nothing will change as users like analysts or decision makers are still accessing the front end of the IAM system. The only change they will notice, is the enhancement of information available or additional tasks to be available like fur-
ther data analysis steps (c.f. Section 4.2) or verification of information to be exported (c.f. Section 4.3).

As the ICM consists of different building blocks the internal structure is described in the following.

4 ARCHITECTURAL BUILDING BLOCKS OF THE ICM MODULE

Figure 2 shows the internal structure of the ICM. We created four different steps while the arrows indicate the information flow within the ICM:

- **Data Harmonisation & Quality Monitoring.** IAM data is integrated with external information. Furthermore, privacy and data quality related procedures are applied.
- **IAM Data Analysis.** Information is provided for the IAM system for analysis. Valuable information for external systems is generated.
- **Result Verification.** Results from the IAM data analysis can be verified by experts and marked for export to external applications.
- **Data and Result Storage.** The relevant combined data sets are immutably stored if proof is needed.

Within the next sections we describe each building block more detailed and hint at possible implementations. Please note that we provide a generic architecture which shall be suitable for as many enterprises as possible. Thus, the mentioned implementations are not exhaustive measures but rather exemplary ones. Some enterprises may even decide not to fully implement all procedures for individual reasons.

4.1 Data Harmonisation & Quality Monitoring

As soon as new dynamic information is sent to the ICM module the first step is started. First the information from the connected external applications need to be harmonised with the already existing IAM data set. In order to ensure compatibility with existing IAM systems, our approach integrates the external information as attributes in the context of ABAC. Thus, for each information a suitable data type can be found (e.g. boolean attribute, string attribute or numerical attribute). Based on already existing IAM data models this results in attribute connections with identities, accounts, permissions or roles (Kunz et al., 2019). Within this work we extend the scope such that external information can also be linked to an assignment of two IAM entities (e.g. information on the usage value of a specific assignment of one permission to an account). For better understanding Table 1 shows suitable examples which external information may be relevant per entity.

<table>
<thead>
<tr>
<th>IAM entity</th>
<th>Potential external information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity</td>
<td>Business function or trust score</td>
</tr>
<tr>
<td>Account</td>
<td>Necessity of account</td>
</tr>
<tr>
<td>Permission</td>
<td>Responsible person, criticality</td>
</tr>
<tr>
<td>Role</td>
<td>Usage, business relevance</td>
</tr>
<tr>
<td>Assignment</td>
<td>Usage of assignment</td>
</tr>
</tbody>
</table>

During this approach additional privacy measures are to be applied. In order to preserve the possibility to export identified anomalies back to external applications like a SIEM system, pseudonymisation
methods instead of anonymisation are to be preferred within our approach. Thus, relevant identities or permissions can be depseudonymised during the export process or after the transmission to a SIEM system. One method for pseudonymisation is the encryption of privacy related attributes (e.g. first and last name of identities). As many enterprises already have a key infrastructure in existence, suitable public and private keys can be generated and distributed to the relevant systems. Although existing IAM literature does not cover this topic yet, please note that this work does not focus on an in-depth procedure how to apply pseudonymisation within IAM.

Finally, data quality measures are applied to the harmonised data set in order to improve the information base used by IAM analysts and decision makers. As data from various different sources are used data quality problems are imminent. One example would be several different location attribute values indicating the same location (e.g. Munich, MUC, and München). There are already existing measures for general data quality management as well as measures specifically suited for data quality within IAM which can be used for implementation (Batini and Scannapieco, 2016; Kaiser et al., 2007; Kunz et al., 2019).

4.2 IAM Data Analysis

Within this step the harmonised IAM data set is provided to the IAM system for further analysis. On the one hand additional information from external applications can support several already existing IAM data analysis measures. Examples of this extensive collection would be visual analytics, role mining or policy management (Colantonio et al., 2012; Puchta et al., 2019; Fuchs and Pernul, 2008; Hummer et al., 2016). Besides that additional information also has major impact on the conduction of access reviews (Osmanoglu, 2013; Reinwarth, 2019). Further information can be displayed for the decision maker to improve the result quality of access reviews (i.e. improved rate of removed assignments).

On the other hand new automated analysis procedures can be established. Having information on the usage of specific permissions or their criticality for business processes related requests can be analysed from the IAM system and automatically denied (e.g. if an identity requests a specific permission to be added to her account although none of her coworkers have used this permission). Furthermore, anomalies can be identified when comparing permission usage within specific identity groups.

4.3 Result Verification & Data Export

All information that needs to be exported to external systems is verified within this step. Especially when it comes to automatically identified anomalies, a human domain expert should be responsible to verify if the anomaly is indeed potential threat information or if there exist explainable reasons for the anomaly (e.g. administrative work during the weekend).

After successful verification, the data can be exported to the respective external application. Within this work, we are focusing on threat information as an example. Thus, a suitable application for this use case would be a SIEM system. The ICM module has to prepare the data by including all relevant information from IAM or other applications and converting it into a suitable data format. When speaking of threat information there exist various languages for threat exchange as well as an in-depth analysis (Menges and Pernul, 2018). Based on this outcome, STIX\textsuperscript{1} or the more generic MISP\textsuperscript{2} may be suitable for our approach. Especially MISP allows for an export appropriate for our use case as individual MISP objects suitable for IAM can be defined.

4.4 Data and Result Storage

As the integration of data is an ongoing process, we decided to introduce a storage module specifically for the ICM besides the fact that IAM systems already have their own databases. Reasons for that are the possibilities to integrate information as soon as they are sent to the IAM as well as the usage of immutable and more flexible storage technologies. To solve this redundancy of storage, the IAM system could also access the result storage provided by the ICM module. However, in reality this may be quite challenging as established IAM systems typically have their individual database structure needed for fully functionality.

Furthermore, the ICM is able to harmonise all information on its own such that the IAM system only receives completely integrated and privacy-respecting data. Thus, no additional data operations need to be conducted within the IAM system. Additionally, as potential threat information is to be exported via the ICM module we want to ensure immutability of existing records as well as the possibility to depseudonymise affected IAM entities.

The technical implementation could be done via traditional SQL databases. However, if enterprises focus on immutability of stored records as well as flexi-
bility upcoming storage technologies like blockchain may be a suitable implementation as well (Nuss et al., 2018; Dunphy and Petitcolas, 2018).

5 REAL-LIFE APPLICABILITY OF THE ARCHITECTURE

Within this section we will show the feasibility of our proposed approach. However, as our approach also has generic elements included, we focus on evaluating the applicable parts of our research. For this we are using aforementioned access reviews as the outcome heavily relies on the quality of information given (Osmanoglu, 2013). Furthermore, within access reviews, decision makers in IAM can either confirm or remove an assignment which leads to binary results when measuring the results. Thus, we want to show that the quality of results of reviews is significantly improved by leveraging the extended IAM architecture. Typically within IAM projects, one way to measure quality of access reviews would be to look at the numbers of removed assignments. The better the quality of information provided, the higher the confidence of decision makers to remove assignments instead of confirming them (Osmanoglu, 2013).

The foundation of this section is represented by two anonymised real-life use cases from our IAM project experience. Firstly, FinCorp working within the finance & insurance sector, and secondly EngineCorp active in the engineering sector. During these projects we (partly) applied our proposed architecture resulting in having external information integrated into the IAM system. Subsequently, we conducted the reviews together with the individual enterprise by leveraging parts of the existing IAM analytics and governance platform Nexis Controle\(^3\). After concluding the reviews during 2019 and 2020, we extracted and analysed the decisions made by the individual employees. As the rate of removed assignments may greatly differ depending on the individual enterprise and its culture we conducted two access reviews within each use-case with only one of them including external information. Thus, by having a control group per company, we can ensure comparability of our results. Furthermore, we can secure that the decision makers for both access reviews mostly remain the same. Based on these reasons, we only compare the results within each company.

Table 2 shows the overview of our results. Per anonymised enterprise we used one additional external information during the reviews. Both enterprises have more than 5,000 employees and respectively reviewed more than 20,000 assignments in the time period 2019 to 2020. When looking at the indicator of removed assignments our results for each enterprise show an extensively higher confidence to remove assignments when having additional external information integrated via our approach. We also applied a \(\chi^2\) hypothesis test to both of our use cases to show the significance of the results and thus to proof that our approach is linked with the rate of removed assignments. With FinCorp having \(\chi^2 \approx 1328.12\) and EngineCorp having \(\chi^2 \approx 178.25\), both use cases show a statistically significance with a significance level of \(\alpha = 0.001\).

In the following the individual use cases are described more precisely.

5.1 FinCorp

Our first use case covers a finance and insurance company having a centralised IAM system as well as various connected applications (e.g. Microsoft Active Directory, SAP, and databases). As a member of this sector, FinCorp needs to comply with various internal as well as external regulations regarding the management of IT systems. The company currently is in the final steps of a role modelling project. Additional information of external applications proved also to be a critical success factor for that. During the process additional information from the individual departments was gathered in role modelling sessions (e.g. if an resource is still needed). Subsequently, we integrated the information into the IAM system using parts of Nexis Controle as our ICM resulting in additional attributes for entitlements. Therefore, the domain of an entitlement could be described (e.g. accounting entitlement or customer service entitlement) as well as its necessity. Furthermore we applied attribute quality mechanisms by identifying duplicate values or missing key attributes.

Parallel to the role modelling, access reviews including this external information were conducted from 03/2019 to 06/2020. Various managers had to recertify the entitlements assigned to employees and not already included in modelled roles. The previously assessed information on permission necessary was added as external information to the review. The information was expressed as binary value, if the permission was still needed or not. In total, 23,260 assignments could be reviewed while 21.62% of these assignments were removed.

The second access review examined the existing roles, their included employees and entitlements as well as the assignments of functional accounts used

\(^3\)https://www.nexis-secure.com
for trainees. These accounts were not recertified during the first review. No additional external information was included during this review and 9,006 assignments were reviewed. However, only 4.67% of these assignments were removed. Although this is quite a high number compared to other IAM projects from our experience, there is clearly a significant difference when comparing the results with the first review.

By applying our approach during the role modelling and review process FinCorp was able to improve the quality of their modelled roles by integrating additional information in the process. Furthermore, existing assignments beside their role model could be effectively removed.

5.2 EngineCorp

EngineCorp develops and builds machine parts within the engineering sector. Access reviews are conducted on a regular yearly basis with a focus on SAP role hierarchies and account permission assignments. For the review conducted in 2019, we added further external information for the recertification of the SAP composite roles. In detail we included the information if a SAP single role within the SAP composite role was used by any of the SAP composite role members during the last 6 months. Thus, decision makers are able to quickly identify the single roles which may not be required by the users. We did not include this information when it comes to the access review of individual account permission assignments (= second review).

The usage information itself can be gathered within the SAP application. We extracted the data from several SAP tables and aggregated it to show only the usage within SAP composite roles but not for individual accounts. In this use case SAP functions both as connected application and as external application with regards to our extended IAM architecture. We are aware of potential privacy issues when it comes to SAP composite roles assigned to only one account. After discussion with EngineCorp, these composite roles should be reviewed as well, however we did not include the list of the accounts assigned to the respective composite role during the review.

In total, 3,635 SAP role hierarchy assignments were reviewed including the usage information. 7.21% of these assignments were removed by the decision makers. Similar to our first use case there is a significant difference to the review conducted without having external information. Although, 17,212 account permission assignments were audited, only 2.72% of these assignments should be removed. The reviews were conducted in parallel while the decision makers for both recertifications remained the same for the most part.

6 CONCLUSION

Enterprises constantly have to face a rising number of internal and external regulations. While IAM has already proven its value for these scenarios, it lacks of interconnections with other IT systems for mutual benefits. Within this work we carved out the look of current IAM architecture as well as potential shortcomings (RQ-1). However, the extended IAM architecture should still be compatible with already existing approaches. By adding the ICM and structuring the information flow we proposed a new IAM component where information from external applications can also be considered (RQ-2). The introduced ICM was further detailed with four different building blocks and descriptions with a more practical point of view for each of them (RQ-3). To show the practical applicability of our approach two real-life use cases from our IAM project experience were introduced. By conducting two reviews per use case with one having applied our procedure we were able to show a significant improvement regarding the rate of removed assignments.

However, as our use cases are based on real data, we could not grant an identical setup of our reviews per company. Exemplary, EngineCorp reviewed assignments of SAP roles with external information and account assignments in the control group review. Also the number of reviewed assignments differed. This is a natural limitation when applying real enterprise data. It would be uneconomical behaviour if different employees should conduct the same review while adding external information to one review would be the only difference. Additionally, we made the reviews as similar as possible by using the same graphical interface, the same emails, and the same support
documents. Finally, the comparably high $\chi^2$ values per use case indicate a clear difference when applying our proposed work.

In the future we want to further apply our architectural approach to more IAM use cases in order to collect even further data. By using our extended architecture, enterprises can achieve the integration of such additional information. As we provide a more generic approach for this, it has to be specifically tailored to the individual enterprise. This is a step not covered within our work and needs to be done within an IAM project in cooperation with the respective business. One aspect of this is also the appliance and extent of privacy measures or IAM processes in general.

**ACKNOWLEDGEMENTS**

This research was supported by the Federal Ministry of Education and Research, Germany, as part of the BMBF DINGfest project (https://dingfest.ur.de).

**REFERENCES**


