Dead Ends in Identity Management

Slepé uličky správy identít

Radovan Semančík

Radovan Semančík

Radovan Semančík studied Software Engineering at the Slovak University of Technology in Bratislava, where he also earned his PhD. He works as a Software Architect at Evolveum and was one of the founders of the company. His main areas of interest include digital identity and software system architectures. Since around 2000, he has been involved in numerous enterprise identity management (IDM, IGA) implementations. He is an active contributor to open-source projects and collaborates on large-scale international software projects. Most of his time is dedicated to leading the midPoint project, which is the largest open-source identity management system available. He has also contributed to the Apache Foundation as a committer and a member of the Project Management Committee.

Ing. Radovan Semančík, PhD. vyštudoval odbor softvérové inžinierstvo na Slovenskej Technickej Univerzite v Bratislave a na rovnakej univerzite získal titul PhD. Pracuje v spoločnosti Evolveum na pozícií softvérového architekta. Bol jedným zo zakladateľov spoločnosti Evolveum. Jeho hlavnými oblasťami záujmu sú digitálna identita a architektúry softvérových systémov. Bol zapojený do mnohých nasadení riešení pre spávu podnikových identít (IDM, IGA) približne od roku 2000. Je aktívny prispievateľ do opensource projektov a spolupracuje na rozsiahlych medzinárodných softvérových projektoch. Väčšinu svojho času venuje vedeniu projektu midPoint, ktorý je najrozsiahlejším voľne dostupným open source systémom pre správu identít. Podieľal sa na práci Apache Foundation ako committer a člen project management commitee.

Dead Ends in Identity Management

Identity and Access Management (IAM) is the core of any serious cybersecurity solution. Identity management functions are mentioned in almost every cybersecurity regulation and standard. Despite this, IAM components in cybersecurity projects are often underestimated and frequently misused.

This paper addresses recurring issues in identity governance and administration (IGA) deployment projects. We present several notorious examples of poor practices, such as:

- Inappropriate composition of IAM components,
- reliance on incorrect input data,
- incorrect sequencing of implementation steps,
- frequent issues in role-based access control (RBAC),
- problems in policy management, role, and application ownership,
- misuse of certification processes,
- unrealistic expectations regarding the benefits of artificial intelligence.

For each dead-end scenario, we present the correct approach leading to a sustainable identity management strategy. The paper presents an incremental and iterative approach to IAM deployment, based on a bottom-up methodology, respecting the current state of the organization.

Well-managed identity governance is not a sprint, but a long-term effort. This approach allows for the gradual implementation and continuous improvement of identity management within an organization.

We introduce the dynamic application of the RBAC model (policy-driven RBAC), incorporating role analysis through AI-driven role mining algorithms, which enables long-term sustainable access policy management. Finally, we demonstrate how an IGA platform, with high-level policy support and AI-driven functionalities, can form a solid foundation for cybersecurity, regulatory compliance, and adherence to industry standards.

Slepé uličky správy identít

Správa identít a prístupov (identity and access management, IAM) je jadrom každého seriózneho riešenia kybernetickej bezpečnosti. Funckie správy identít sú spomínané v takmer každej regulácií a štandarde pre kybernetickú bezpečnosť. Aj napriek tomu sú IAM komponenty v projektoch kybernetickej bezpečnosti podceňované a veľmi často sú nesprávne používané.

Príspevok sa zaoberá často sa opakujúcimi problémami v projektoch nasadenia správy identít (identity governance and administration, IGA). Uvedieme niekoľko notorických príkladov zlej praxe, ako napríklad:

- Nevhodná kompozícia IAM komponentov,
- dôvera v nesprávne vstupné údaje,
- zlé poradie krokov,
- časté problémy pri použití rolí (RBAC),
- problémy pri správe politík, vlastníkov rolí a aplikácií,
- nesprávne použitie certifikácií ako aj prehnané očakávania prínosov umelej inteligencie.

Pre každú slepú uličku si ukážeme aj správnu cestu, ktorá vedie k udržateľnému prístupu k správe identít. Príspevkom ukazujeme inkrementálny a iteratívny prístup k nasadeniu správy identít, založený na metóde „z dola nahor“, rešpektujúc aktuálny stav organizácie.

Dobre riadená správa identít nie je šprint, ale beh na dlhú trať, preto tento prístup umožňuje postupné zavedenie správy identít v organizácií a jej inkrementálne zlepšovanie.

Predstavíme dynamické použitie RBAC modelu (policy-driven RBAC), s použitím analýzy rolí pomcou algoritmov umelej inteligencie (role mining), ktoré umožní dlhodobo udržateľnú správu prístupových politík. Ukážeme, ako IGA platforma s podporou vysokoúrovňových politík a s využitím prvkov umelej inteligencie môže tvoriť pevný základ pre kybernetickú bezpečnosť a súlad s legislatívou, reguláciami a štandardami.

1 Introduction

Identity and Access Management (IAM) is a core of every decent cybersecurity solution. Identity management capabilities are mentioned in cybersecurity regulations [1] and standards [2]. As *identity* is a central concept of cybersecurity, it is almost impossible to achieve satisfactory level of cybersecurity at scale without a solid IAM foundation in place. Despite that, IAM components are often neglected and abused in cybersecurity projects. While cybersecurity solutions usually take care of authentication, overall *management* and *governance* of identities is neglected. Even in cases when identity management and governance is addressed, it is often not done properly. Identity management and governance solutions often repeat critical mistakes and follows anti-patterns that were considered bad practice as early as two decades ago. As this bad practice is still surprisingly common today, this paper provides selection of the the most common aspects of identity solutions, which are often simple, easy to understand, and completely wrong.

2 Identity Management and Governance

Terminology of identity technology can be confusing and counter-intuitive. Perhaps the first identity-related concept that comes to engineer's mind is *authentication*. While authentication, and a broader concept of *access management*, are undoubtedly the most visible parts, they are just a very small part of identity infrastructure. The underlying components implementing identity management and governance are hidden from the plain view.

Identity management and governance take care of the identity "back end". They are primarily concerned with identity inventory, lifecycle, access control policies, and especially high-level business oriented polices related to identities. While access management is primarily concerned with question "does user have access here?", primary questions of identity management and governance are "why does the user have access here?" and "who is responsible for this?". Especially identity governance is concerned with inventorization, ownership and responsibility. Identity management and governance are often referred to Identity Governance and Administration (IGA).

Identity management and governance are absolutely essential for cybersecurity at scale. Cybersecurity is all about users – their actions, permissions, the value they create and the risk they are posing. Lack of control over the users renders all other cybersecurity controls almost meaningless.

3 Dead Ends

Identity governance is far from being easy. It is complex, there are many counter-intuitive aspects and it is desperately riddled with poor data quality. This is further amplified by the fact that identity governance is often neglected and/or incorrectly implemented. Visibility is very limited, which means that identity solution can seem to work correctly, while is it not.

Following sections describe common pitfalls of identity management and governance. The description is based on decades of identity management experience. It is also supplemented by experience in designing, developing and deploying midPoint [3], an open source identity governance platform.

3.1 First Steps In Identity

The worst problems in any endeavour are likely to be created at the very beginning. Identity is no exceptions. The worst mistakes usually happen in very early phases, when the solution is in a form of rough idea. Cybersecurity professionals have a natural tendency to see identity projects from a cybersecurity perspective, usually starting with cybersecurity requirements of end users. Multi-factor

authentication is usually the first component of an identity solution. It is a start from a very wrong end. It makes little sense to implement strong authentication of poorly-managed identities.

Identity cannot be an afterthought. Identities need to be *managed* before they are used. Unmanaged identities are posing significant risks, even if the strongest authentication, authorization and cryptography is used. Identity *management* needs to be foundational part of identity solution from the very beginning.

Identity management should focus on inventorization, ownership and lifecycle of identities as the very minimum: automatically synchronize identities with authoritative sources (e.g. human resource database); link accounts to their owners (users), detect orphaned accounts; automatically de-activate identities (e.g. "leaver" users).

Recommendation: Start your identity management effort early. Prioritize it over authentication and access management. Early identity management solution may be very "light", yet it is still necessary.

3.2 My Data Are Perfect

Identity management is built on data. It can do great things as long as the data are correct. However, it is a complete disaster when the data are wrong. Identity management and governance is heavily using automation and policies, which rely on identity data. When identity data are wrong, policies are going to reach wrong decisions. Perfectly legal accounts are going to get disabled, accounts that should not exist are re-enabled, wrong privileges are applied, real alarms are muted while false alarms proliferate. The correctness and efficiency of policies relies on the data.

Even though identity data are essential for cybersecurity, quality of identity data is usually quite low. This dichotomy may seem counter-intuitive, yet it is easy to explain. Quality of any kind of data tends to be low until the data are actively validated. However, until the identity management systems is deployed, there is neither practical way nor motivation to validate identity data. Therefore identity data silently "rot" in their isolated databases.

Identity management system dramatically changes the situation. It synchronizes and compares identity data in several databases, usually the human resource (HR) database, central identity repository (Active Directory, Entra or LDAP) and application user databases. Data inconsistencies are detected, orphaned accounts are discovered, the data are corrected.

Data synchronization, correlation and clean-up is an essential step of every identity management program. However, it is by no means an easy step. Identity data owners almost always vastly over-estimate data quality. In quite a counter-intuitive way, quality of individual identity attributes varies significantly. E.g. quality of HR sunrise/sunset dates (dates person was hired or left) is usually quite good, as those data are used for payroll processing. However, quality of organizational assignment, job code and location data in the HR database is usually low, as those data are mostly informational and there is little incentive to maintain them.

Quality of identity data is usually not known before identity management effort is started. The real data quality is discovered only when the data in identity databases is compared, which often comes as an unpleasant surprise for everyone involved. Therefore, the initial identity data clean-up step usually takes much longer than expected. However, data clean-up is absolutely necessary. Application of policies on unmanaged data can be fatal, especially for policy-based access control systems (ABAC/PBAC).

Recommendation: Synchronize, correlate, compare and clean up your data early in the identity management effort. At the very minimum, compare and correlate primary data source (e.g. HR database) with the most-widely-used applications (e.g. Active Directory/Entra). Detect and remove obvious errors (e.g. orphaned accounts). Do not apply any automated access control policies until data quality is established.

3.3 Identity Project

Similarly to cybersecurity, identity management is not a *project*. It is a continuous *program*. It has a start, but it has no end. Identity management and governance needs constant maintenance and improvement. As the environment changes, the solution has to adapt: policies need to be updated to reflect organizational change, new applications need to be connected, policy exceptions need to be reviewed, new regulatory compliance requirements need to be applied and so on. The effort never ends.

The continuos nature of identity management is important organizational constraint. Attempt to deal with identity management in a form of project often lead to failures. The worst disasters tend to be created by "big bang" projects, trying to implement identity management in one big step, in a fixed-time/fixed-scope way. Identity solution cannot be purchased. It needs to be built, it needs to grow organically with the organization.

The best method to deploy identity management is to proceed iteratively and incrementally, in many small steps. E.g. start with connecting primary data source (HR database) and analyze the data. Then connect popular application (Active Directory/Entra), compare and clean-up the data. The solution can be used for visibility at this point. Next step should apply basic identity lifecycle automation. Next step could apply basic access control policies, or perhaps an access request process. Next step could connect new application to the system. And so on. Specific steps will vary for each individual organization. However, each small step should bring functional and practical solution, providing tangible value.

Recommendation: Start small. Proceed in small increments, each step improving upon the previous one. Set sustainable pace. Set up a continuous identity program, ideally aligning with cybersecurity program.

3.4 Access Control

Access control lies at the very heart of cybersecurity. Numerous access control models were proposed over the years. Two models in particular are gaining popularity recently: attribute-based access control (ABAC) and policy-based access control (PBAC). Both models are based on the principle of algorithmic policy, expressed in a form of a code or a set of precise rules. This approach is known as "policy-as-code". It is meant to be simple and efficient. However, the assumption that access control can be easily expressed in a form of algorithmic policy is much bolder than it may seem.

Writing any kind of code is surprisingly difficult task, especially considering all the possibilities and corner cases that the code needs to handle. It requires specific skills and appropriate tools. Once the code is written, it needs to be maintained, which is even more difficult task. There are likely to be policy exceptions (also known as "special cases"), temporary measures and workarounds (also known as "hacks"). The code is very likely to become unreadable, and hence unmaintainable, in quite a short time.

However, there is one surprising aspect that prohibits applicability of policy-based access control models at a large scale: in large organizations, nobody really knows what the policy is or how it should look like. Knowledge about overall policy is spread among many people, the understanding of the policy is inconsistent, often conflicting, there is huge amount of unwritten rules and established practices and decisions are often made on the basis of whether it "looks good", without any specific rules. While it may be theoretically possible to express such "policy" using algorithmic means, practical feasibility of this task is usually close to zero.

Recommendation: ABAC and PBAC model can indeed be very simple and efficient, given the right circumstances. These models can work exquisitely well as long as the policy is known, it is deterministic

and relatively stable. However, they should be used with care. Their use for organization-wide access control and governance is likely to be infeasible.

Access control models that mix declaratory and algorithmic mechanisms are much more appropriate for managing complex policies. Dynamic variations of the renowned role-based access control (RBAC) models are successfully used for this purpose. While the traditional RBAC model [4] is completely static, dynamic variants of the model have been used in identity management platforms for more than two decades. Dynamic RBAC models combine declaratory definition of roles with dynamic rules that govern automatic assignment/unassignment of roles, as well as set of privileges granted by the roles, creating a policy-driven RBAC model.

Such hybrid access control models can control both policy-based access and declaratory access (also known as "standing privilege"). The policy-based mechanisms are used for cases where the policy is known, the declaratory parts are used for case where the policy is not explicitly specified. Moreover, the models allow coexistence of the two mechanisms, and even transition from one to the other. E.g. it allows transparent management of policy exceptions, without the need to complicate policy code.

Even more importantly, the hybrid mechanism enabled a *bottom-up* approach to policy management. Even though a policy may not be known explicitly, it is often *implicitly* expressed in existing assignment of privileges. Existing privilege assignment can be analyzed to detect patterns, which can be used to specify a dynamic policy. E.g. *role mining* mechanism is a simple variation of this approach. This approach can be used to gradually built a policy starting from the bottom (privileges), going up (roles, attributes, locations, projects, organizations). This process can be iterative and incremental, aligned with continuos identity and cybersecurity program.

Recommendation: Use hybrid access control models, such as dynamic policy-driven RBAC to capture both the algorithmic and ad-hoc parts of the policy. Follow bottom-up approach to access control and policy management, proceed is manageable iterative steps.

3.5 Identity Governance

Identity management is often reduced to management of user identity data and access control policies. However, all data and policies need constant maintenance, otherwise they deteriorate and fail. Maintenance can be assured by implementing complex procedures and processes. However, even the best processes are not going to work well unless there is appropriate person *responsible* for the maintenance. *Responsibility* is the very core of governance.

The road to responsibility starts with inventory. Inventory of applications, entitlements (groups, roles, permissions) and policies is the very minimum. We cannot make sure something is properly cared for unless we know that it exists. Next step is to make sure everything has an *owner*, responsible for the maintenance. Every application, every group, role or policy needs an owner. Make sure that owners have a chance to regularly review the things that they are responsible for.

Recommendation: Assign owner for every application, role, entitlement and policy. Maintain inventory. Automatically detect objects without an owner and re-assign them.

Review campaigns (also known as access review, certification or attestation) are often used as a primary mechanism to make sure access control is maintained. The usual approach is to review (certify) assignment of all roles and entitlements on annual basis. However, this practice is vastly time-consuming and very inefficient. It creates huge number of items that need to be reviewed, assigned to reviewers that often have absolutely no idea what that specific entitlement is supposed to mean. This leads to a "rubber-stamping" practice: reviewers are confirming all accesses, afraid to remove anything, or being simply overwhelmed by the sheer amount of items for review. While this practice is still (quite regrettably) accepted as being regulation-compliant, it usually does not lead to any significant reduction of access. En-masse access review is mostly just a security theatre.

Access review need at least two improvements. Firstly, they need to be *focused*. Instead of reviewing everything, select areas that pose the greatest risk: privileged access, access to applications processing sensitive data, users with large number of entitlements, or access of users that recently changed their work position. Secondly, reviewers need *assistance* to make informed decisions. Provide additional guidance for each review item. Is it a privileged access? How sensitive is the application (classification)? Is this an outlying (uncommon) access? Reduction of review volume together with additional information leads to better decisions.

Recommendation: Make reviews efficient by focusing the campaigns, using risk-based approach and employing micro-certifications. Support reviewers by providing relevant information, including AI-based recommendation (e.g. outliers).

3.6 Artificial Intelligence

Artificial intelligence (AI) is presented in popular culture as a solution to every problem. Identity management and governance are not immune to this trend. In particular, AI is often presented as a solution to overwhelming complexity of identity-related problems, amplified by critical shortage of (human) talent and expertise. The expectations are immense.

However, AI is only as good as are the data it works on. Every organization is unique. General-purpose generative AI models have no prior knowledge that could provide groundbreaking insights about your specific situation. The AI engines can only work with the data you already have. However, if the quality of your data is low, the quality of AI insights is likely to be low as well.

Yet, there are classes of machine learning techniques that can be successfully employed. These techniques look for similarities, patterns and differences. Such algorithms are successfully employed for role mining, as well as in outlier detection mechanisms. Such mechanisms are meant to provide *assistance* to (human) user. Human expertise is still needed to make final decision. However, AI assistance can provide additional information to make better decisions faster.

Recommendation: Use proven techniques based on machine learning, such as role mining and outlier detection. AI-based assistance for reviews and approvals can be very useful, as long as the recommendations are *explainable*. Avoid over-hyped AI marketing promises.

4 Conclusion

Identity management and governance can bring a lot of value if used correctly. It is an indispensable tool for managing risks, providing visibility and maintaining regulatory compliance. However, the inherent complexity of the identity field is a breeding ground for confusion, misunderstanding and myths. Resulting anti-patterns and bad practices lead to incorrect and inefficient use of identity tools, leading projects into an expensive dead ends.

This paper describes some of the common pitfalls in identity management and governance, hoping that these mistakes will not get repeated. However, as the use of technology evolves, new pitfalls appear, and new myths are born. Therefore, the effort to document the identity myths [5] is not likely to end in any foreseeable future.

Each of the dead ends described in this paper was accompanied with a recommendation of better practice. However, one of the recommendation more important that all others. It is the recommendation of iterative approach: start small, proceed in steps, learn along the way. This is the best approach to avoid both the known as well as the unknown dead ends of identity management and governance.

References

[1] Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union (NIS 2 Directive), recital 89, 2022.

[2] ISO/IEC 27001:2022, Information security, cybersecurity and privacy protection — Information security management systems — Requirements, controls 5.15, 5.16, 5.17, 5.18, 2022.

[3] MidPoint documentation, Evolveum documentation site, https://docs.evolveum.com/midpoint/, maintained 2011-2025, retrieved 2025.

[4] INCITS 359-2012, Information Technology - Role-Based Access Control, InterNational Committee for Information Technology Standards, 2012

[5] Identity and Access Management Myths, Evolveum documentation site, https://docs.evolveum.com/iam/myths/, maintained 2014-2025, retrieved 2025.