Identity-First Security

Identity is the New Perimeter for the Zero Trust Future
No concept in security has had more deaths than the perimeter. And yet, we continue to talk about it. Afterall, what is dead can never die.

Over the past 20 years, we have shifted from work being something that happened in the office on the LAN to a context where infrastructure, apps, and basically every aspect of work are all managed in the cloud. The boundaries between the inside and outside of the perimeter have been erased.

No more assumptions about who should or should not be trusted. No more defining attributes like location, devices, or time zone. Work happens everywhere, all the time. Simply put, we have outgrown the network. And with it, any lingering relevance that the network perimeter may have had left.

Identity has replaced the network as the new perimeter, leading organizations to adopt an Identity-First Security posture for protecting their assets.

To better understand why the industry is moving towards Identity-First Security as the new standard, it can be helpful to look back at the changes from the past few years in how we think about protecting our assets.
Defining Identity-First Security

In intentionally broad terms, the legacy network perimeter model of security posited that everyone inside the network should be trusted by right of them being inside the perimeter. Alternatively, everyone outside of said perimeter on the open web was highly suspect.

Ideally, the high walls of the perimeter were supposed to be enough to keep attackers at bay and the assets inside safe. However, as we have seen time and again, malicious actors have no problem getting past the perimeter and reaching the valuable assets inside. If these assets do not have additional layers of protection, then they are easy pickings for these vultures of the interwebs.

This understanding led to the development of Zero Trust principles that took the “never trust, always verify” approach. In this model, everyone and every request is constantly being challenged. Forget the perimeter. In Zero Trust we assume that the bad guys are already inside the network perimeter and put in place controls to limit their ability to do damage while continuously working to detect them.

But what happens when we no longer control the network?

Over the past decade, SaaS apps and then IaaS and PaaS — or collectively referred to as XaaS — have taken a more dominant role in the way that organizations get work done. In 2020, the average mid-market company was using 137 SaaS apps. This number more than doubles to 288 for enterprises.

This evolution into cloud-based work has significantly altered the dynamics in terms of control over security. Namely, that in the cloud, organizations are no longer in control of the network.

To be fair, this lack of control has been a major incentive for moving to the cloud in the first place. It allows your organization to scale as needed without taking on the capital expenditures and IT burden that it would take to set up and run that kind of operation.

It is wonderful. But every step forward comes with its own sets of concerns that need to be addressed. In this case, it means that no organization is fully in control of their infrastructure or apps. It’s like beer. You don’t buy it, you rent it. And the way that we access these rented services is through proving that we are the folks with the rights to use said services.
When the network was what we used to control access to our assets, we looked for ways to make our use of the network more secure. Now that the network’s role in controlling access has been replaced by identity, we need to think about how we utilize those identities in the securest way possible — implementing the same principles that set forth for Zero Trust, but now focused on identity instead of the network.

Google’s BeyondCorp framework is probably one of the best-known early examples of Zero Trust principles being applied to identity. It came in response to the 2009 Operation Aurora and sought to move away from the idea of network segregation as the primary method of protecting valuable assets.

So how does this work in practice? In the earliest eras, our approach was to set up a wide perimeter around many assets at once. Now with Identity-First Security, we invert this model and look to secure those assets with their own perimeters. Ideally from the tightest rings possible, expanding out with additional layers for added protection.

The perimeter isn’t dead. It’s just been flipped inside out. Long live the Identity-First Security perimeter.

With this recent history in mind, the shift to an Identity-First Security posture makes good sense. And many organizations were already heading in that direction. But the past year and a half kicked this evolution into overdrive.
Cloud adoption has been growing over the past decade, mostly for business practicality reasons. The cloud is scalable, has generally good up time, and is available from just about anywhere. As mentioned before, just about every application is offered as SaaS now, with on-prem available for special cases.

Then COVID-19 hit, and Work from Home (WFH) went from being a nice perk to a near default for nearly every organization. Unless you were showing up to work in a SCIF, you were logging into your SaaS apps or other cloud services from home.

As access to assets moved almost entirely to identity, hackers turned even more of their attention to attacking identity. It’s like the joke about why criminals rob banks. Because that’s where the money is.

Looking at many of the breaches that hit the headlines over the past year, think SolarWinds and Colonial Pipeline just to name a few, the identity layer has been under attack. Even identity security tools like ForgeRock have been themselves targeted for use in hacks. The stakes here are high. As we saw in these cases, the hackers compromised privileged identities to reach far inside their targets.

In response to the targeting of identities, the Biden Administration specifically called for the Federal Government to adopt a Zero Trust architecture in its Executive Order back in May. The EO cited the need for improved, automated access controls that included continuous monitoring of key Identity-First Security processes like authentication and authorization.

This is a step in the right direction. The Federal Government has a lot of influence in setting standards. Any vendor that wants to work with the Fed must conform to its rules or be left out in the cold. It’s like how California sets emissions standards for cars.

Given the rate and scale of cybersecurity incidents from the past year, it is likely that the administration will expand some portions of its regulations beyond the government sector and into many private ones as well. But that may take some time to get passed and implemented.

In the meantime, there are several steps that organizations can take to gain a competitive advantage by better protecting themselves and mitigating their risk. No need to wait for regulations in order to get started.
Identity-first security is a process of verifying that the person is who they say they are and that they have the right entitlements to be accessing what they want to access.

Getting this process right starts with implementing strong authentication controls and tools. Let’s look at the lowest of hanging fruit.

### 3 Guidelines for Implementing Identity-First Security

Building out a strong identity-first security posture is an ongoing process that organizations can and should invest in. However, like other aspects of cybersecurity, a lot can be accomplished by starting with the basics. Here below are three areas that organizations should prioritize as they start their process.

---

**A**

**Use MFA Everywhere You Can**

The first component here should be multi-factor authentication (MFA). This means requiring the person attempting to login to show that they not only know something (in this case the username and password), but also have access to a secondary channel or device. In some cases, this can be an email where a code is sent to, or more likely to the user’s mobile device.

According to an estimate by Microsoft, 99% of attacks can be prevented by using good MFA practices.

Far too often, these codes are sent via SMS, which is a highly insecure way of transmitting this information. SIM swapping attacks or other methods of intercepting the code can undermine the MFA process. Instead, use apps like Google’s Authenticator, Microsoft’s Authenticator, Duo, Authy, or one of the myriads of other fantastic services that rotate codes for MFA use.

In some cases, these services eliminate the need for entering a code by using a push notification to approve or deny access requests. All that said, it is still far better to use SMS verification than nothing at all. Start there and work your way to something better.

---

**B**

**Strict Password Policies**

Poor password hygiene is one of the leading causes behind breaches. One of the worst offenses here is password reuse. This is where folks will use the same password across multiple sites, mostly because it is easier for them than having to remember unique passwords for every site or service. Given the number of accounts that we have, it can be annoying to have a different password for each one.

The problem though is that if one of those accounts gets popped, then the attackers can look around to see where else those users might be using those passwords. These stolen passwords can be bought and sold on Dark Web forums. This is called password spraying, and it is disturbingly effective. Furthermore, it’s not just in use by script kiddies.

According to reports, the crew behind the SolarWinds hack with its beautifully executed Golden SAML attack also made plenty of use of brute force password guessing/spraying to great effect.

Hackers will do whatever works to get them to their target, and sometimes the down and dirty methods are the right fit for the job.
Eventually, organizations should look for better alternatives to passwords. This is for both security and user compliance reasons. As we have seen, passwords are far from the most secure way to manage access. Moreover, there are few if any folks out there that love dealing with passwords. They are a necessary layer of friction that keep us and hopefully attackers from our assets.

As they grow to be mid and large size companies, organizations adopt single sign-on (SSO) tools from vendors like Okta and Ping Identity that help them to reduce some of that friction. These solutions allow users to login to the applications or services that they use with a single set of credentials. This allows for more centralized control of identity, with the vendor doing the heavy lifting managing the connections and verifications.

But SSO is just the first step. Players in this space are working hard to eliminate the password. Instead, they look at biometrics, devices, and other attributes to determine if the user is in fact who they claim to be.

Another alternative solution and an excellent form of MFA is the hardware token. The YubiKey is the most common application of this technology. Instead of receiving an MFA code from the service provider, users simply plug the key into their device or place it near their device to authenticate that they are the one trying to sign on. These nifty little keys hang off your key ring and make it difficult to carry out a man-in-the-middle attack or social engineer the MFA code from the user.
Detect and Monitor Accounts and Identities

The tools and practices outlined above are where organizations should start with their authentication strategy. But authentication is just the first part of the challenge. It is not enough to know who a given identity is when they are attempting to access a given resource. The next step is whether they should be able to access the asset, for how long, and what sorts of actions they should be allowed to take in their role.

This is the authorization stage, and it can get complicated.

A

Establish a Foundation for Continuous Authorization

If authentication is a binary question that asks for a simple yes or no, then authorization is its multi-layered and complicated cousin that never stops asking questions.

Authorization is the step where we ask, “now what?”

“Now that I’ve proven my identity, what can I have access to, and what level of access?”

Access to assets allows us to work with them and do our jobs. But it also means risk. Every time we provide a user with access to a different resource, we expand the probability that that resource might be accessed by malicious actors.

Ideally, we want to provide just enough privilege for just enough time to allow our people to be effective, but not any more than that. Authorizations should be continuously evaluated and adjusted to match the dynamic pace of changes within the organization.

The theory behind the Principle of Least Privilege

More access equals a wider threat surface. So, if you minimize the threat surface, then you reduce the probability of your assets being compromised.

Authorization plays a critical role in the Zero Trust and Identity-First Security approaches. If we assume that any identity in our organization can be compromised — and we assume that there are already attackers inside our organization that have taken over one or more identities — then we want to have controls in place that will limit their ability to cause damage.

In the shared responsibility model of the cloud environment, authorization is our strongest tool for mitigating risk to our assets. This is largely because authorization is within our organization’s control, and up to the XaaS provider to enforce. It allows us to get as close as possible to segmentation in the cloud environment as we are likely to get.

But first we have to know what we are working with before we can start securing them.
One of the primary authorization challenges is creating a baseline knowledge of your identities and assets. This is harder than it sounds as the number of identities and assets spread out across multiple environments has mushroomed with the adoption of cloud services.

Looking at a real-world example of one of Authomize’s mid-size (~4.3k employees) customers, they have:

- 46.5k Identities
- 7.5M Assets (not including files)
- 6.75M Identity/Asset Security Policies
- 65M Access actions taken per month

Keeping track of, let alone understanding, which identities have access to which assets and vice versa is challenging to say the least.

Adding to the scale of this management challenge is that most Identity Access Management (IAM) tools lack the necessary visibility across environments to provide security teams with actionable insights. This is because each of the elements, identities, infrastructure providers (AWS, Azure, GCO), and apps (O365, Salesforce, GitHub) are all siloed from one to another.

While federation tools have been developed to help handle the authentication of users accessing different XaaS services, legacy IAM solutions are incapable of identifying authorizations to the assets. In practice, this means that in order to understand who is authorized to access a given asset, you would have to delve into each individual service’s platform and search for that information.

This is just the surface of the task here because there are numerous levels of authorizations within an asset as you move down the resolution chain. A user may be authorized to access certain files within a specific folder or bucket but be denied access to others within the same folder. Perhaps they are part of a group that has authorization to a resource, but that resource might have a deny rule that prohibits access to all users that are not in a designated allow list.

Further adding to the complexity of gaining visibility across XaaS services is the fact that each service is set up a little bit differently, using its own terminology and structure that can make it difficult to compile its data with those of other services. For example, AWS uses “groups” for managing authorizations to assets for multiple users, while Salesforce uses “profiles” for the same function. Correlating the data from across different XaaS services and compiling it into a centralized control set requires gaining a deep understanding of how each of these products operates and matching their data with the corresponding data from the other environments.

If an organization does succeed in mapping out all their identities and assets, they will still have to regularly repeat this process. These identities are dynamic, constantly changing as employees join, move, and leave the organization. The same holds true for assets as new ones are created, deleted, or have their policies changed. Keeping pace with these changes while maintaining control requires a continuous process that tracks changes as they occur in real-time or risks dealing with a massive review process during periodic assets. As well as running the risk that there are exploitable exposures in between audits.
One of the top priorities while mapping your identities and assets is to identify your crown jewels. These are your organization's most sensitive assets. Their compromise through any of the security parameters (confidentiality, integrity, or availability) would have a significant impact on your organization. Standard examples of crown jewels are data pertaining to intellectual property like source code, financial data, personally identifiable information (PII), or sensitive customer information.

While mapping your identities and assets is an important first step towards understanding what needs to be secured, your Crown Jewels should be your top priority for identification and protection.

Once the Crown Jewels have been identified, the work of protecting them begins. The most effective way to keep them out of reach of a malicious actor in the XaaS environment is by limiting who is allowed to access them.

If we assume in the Zero Trust model that the attackers are already within our environments, then authentication becomes a less relevant factor in our strategy. It is now up to having the most restrictive authorizations policies possible to limit access to these valuable assets.

Ask questions such as:

- **Who within your organization absolutely needs to have access to these assets in order to fulfill their duties?**
- **Are there currently identities that can have their existing access revoked?**
- **What minimal level of access do they need to have? Do they need to be an admin with write privileges or is it read enough?**
- **Do they need to have open-ended access, or can we put in place an expiration date?**

Use your findings from these queries to set policies and then implement them as security policies. Ideally you should have tools in place that will enforce these security policies continuously throughout the authorization lifecycle.

Note that these policies will have to be updated as changes occur.
Perform Periodic Access Reviews of Authorizations

Establishing a process to review authorization policies on a periodic basis is a must for all organizations. This process allows for an opportunity to confirm that everyone in the organization has the right level of authorizations that they require. They are supposed to ensure that changes are properly accounted for and that necessary adjustments are made. Especially in the case of regulated industries, auditors will request proof that the organization has undertaken a review to confirm that all accounts have the appropriate authorizations to assets.

Standard practice for access reviews consists of the IT or Security team sending lists of names and whatever access privileges that they have on record to managers or asset owners who are then requested to confirm or revoke the listed employees’ privileges to those assets. This is often a troubled and wildly inefficient process.

A common challenge that many organizations encounter is that those tasked with approving or revoking privileges do not have any real basis upon which to base their decisions. This lack of information often leads them to simply “rubber stamp” that everyone has the right level of privileges, undermining the process. This problem is so rampant, that those responsible for managing these access reviews can end up expending more time and effort proving that there has not been rubber stamping at play than actually carrying out the review.

Even the basic methods these reviews are handled by the campaign manager face difficulties as they are often simply lists of Excels and email chains. Lacking a real system for collecting proof of approvals, they will send auditors screen shots of the approvals, hoping to show that they have made the effort.

These are also taxing endeavors as the campaign managers must go through the process from scratch periodically throughout the year, updating the names and privileges as they go through the motions.

This method of access reviews is far from ideal.

Instead, access review campaigns should be managed through a centralized platform that provides the managers or asset owners with data for making their decisions. Campaign managers should be able to produce documentation that shows auditors, and management internally, that all decisions are justified and that no rubber stamping has taken place.

Maintaining the Principle of Least Privilege is a must for managing people within an organization. It is even more essential for dealing with external accounts.
Restrict External Exposure to Assets

Permitting access to external accounts outside the organization is a normal part of business. But extra efforts and sensitivity must be paid to ensure that the organization's risks are limited from unnecessary, and prolonged exposure.

One of the primary risks of dealing with external accounts is that you lack much of the visibility that you would have with accounts belonging to your organization. With accounts belonging to your organization, you can draw data from your SSO provider to fill in part of the overall picture regarding their authorizations. However, in the case of external accounts, all we have to go on is the information provided by our assets about which accounts they are in contact with.

All too often, an external account will be granted access to assets but then the asset owner will forget to revoke the access when the access is no longer necessary. Lacking the visibility to detect these authorizations from the identity side, organizations will lose track of who has external access to their assets, creating a sizable level of risk.

If the organization is not continuously monitoring usage of their assets, chances are that they will not uncover this exposure on their own.

Detecting external access to the organization's assets should be a top priority, especially if those assets include one or more Crown Jewels.

Moreover, best practices call for asset owners to revoke access to external accounts – and sometimes even for internal accounts as well – when they are no longer required for a given project or task. This is true even if you think that access might be needed again in the future. It is always easier to request access again than deal with the fallout from an incident because an unmonitored external account was permitted to hold onto their privileges.

Managing authorizations in a dynamic environment is difficult but is doable with a secure and well-organized process.
Strengthen Your JML and Permission Request Lifecycle Process

The Joiner, Mover, Leaver (JML) Lifecycle is the process that manages the provision and revocation of authorizations within the organization. New employees are onboarded into the organization and granted access to the accounts and assets that they need to do their jobs. These authorizations are later adjusted when they change their roles or revoked when they leave the organization.

This process can be very inefficient, slow, and in many cases, inaccurate to the point of being risky.

Privileges are often provisioned based on “look alike” practices that attempt to match employees with similar roles. This is a mistake because while roles are often similar, they are generally not one-to-one matches. For example, a new Content Marketing hire to the marketing department may need access to the HubSpot account but is unlikely to need access to the Salesforce account like the Marketing Manager who has to track leads. They certainly do not need to have admin level permissions.

Furthermore, the process of getting those authorizations can take a significant amount of time. If the Head of Marketing requests that the new Content Manager have access to the HubSpot account on the day that they arrive, it can take up to two weeks for the request to go through the IT and Security departments, before making its way to others who have to sign off on the request before it can be granted. In some cases, the IT person handling the request will have to track down who are the people that have to approve the request. That means lost time that the new hire is not doing his or her job.

In many cases, employees are left with inappropriate levels of authorizations that are a mix of being simultaneously too high and too low. The result is both insufficient access and unnecessary levels of risk.

Identity Governance and Administration (IGA) tools can help to manage this process more efficiently by understanding the needs of different roles and matching them with their required permissions. The problem with many tools though is that they are only really effective with helping the organization in making the appropriate access provisions moving forward and do not retroactively adjust pre-existing permissions. They are also often aimed solely on the side of keeping permissions low. While this is definitely an important goal, this is less than optimal for providing insights about where authorizations should be higher in accordance with the needs of the role.

Most of these IGA tools are also less than effective when it comes to service accounts. Also referred to as robotic identities, these are the non-human identities that are used for automated tasks between services among their other purposes. They can be assigned authorizations and be members in groups or roles.
Often left out of the discussion around identity lifecycle management, we should care a lot more about robotic identities because they play a much bigger role in our organizations than we may be aware of.

According to our research, robotic identities:

- Comprise **20%** of the identities in an organization
- **80%** are not active
- **55%** have unused (read over-privileged) authorizations
- **30%** have admin privileges

These identities can be at higher risk because they are often unmonitored and not part of the JML. They also are authenticated by use of keys or tokens, so stronger security tools like MFA are not useful here. If those keys or tokens are compromised, then the attacker can make free use of the robotic identity for their activities with low risk of anyone picking up on their actions in time to block them. They can be used for privilege chaining, allowing the attacker to reach more sensitive areas within the target.

We saw this method used by the compromise of a robotic identity in the SolarWinds attack to great effect. Mitigating the risks from robotic identities requires monitoring and restricting them as you would the human identities in your organization.

**Best practices include the following:**

- Make sure that you have alerts for the creation of new robotic identities.
- Decommission them when they are no longer being actively used. If you need a new one, then spin it up. Do not let them stick around without a good reason as they will expand your threat surface.
- Assign them owners so that they are not orphans that can be misappropriated without someone being made aware of it.
- Right-size their permissions to match their needs in line with Least Privilege Principles.
- Further limit their access by creating allow lists that designate which applications these identities are permitted to access.
- Do not hardcode keys or tokens into your applications. These can be used to compromise your robotic identities and are a painful self-inflicted wound that you are better off avoiding.
Final Takeaways for Practicing Identity-First Security

As the diminishing role of the network has given way to dominance of the cloud environment, new tools and practices have become necessities for mitigating security risks to your organization’s valuable assets.

Identity is the key to accessing the resources and tools that organizations depend on, making these assets accessible from anywhere at any time. This shift requires that we embrace an Identity-First Security approach in all its components.

This means implementing strong authentication and authorization tools that will help ensure that attackers do not continue exploiting identities for carrying out their malicious actions. These tools must be continuous and automate many of the processes that are currently being managed manually today, allowing Security teams to better prioritize their attention on those challenges that truly require their talents and time.

For more information on how your organization can implement a Zero Trust Identity-First Security architecture to meet the challenges of all your cloud environments, please contact Authomize for a free consultation and demo.

About Authomize

Authomize enables organizations to manage and secure complex and vastly different applications across hybrid environments. Our Prescriptive Analytics engine helps IT and Security teams flawlessly automate operations around authorization to prevent permission sprawl, maximize productivity, simplify identity lifecycle management, and secure the IAM plane.