Why Data Science is Foundational for an Advanced Cyber Program

Most CISOs are likely to agree: Data Science *is* the future of cybersecurity. This assertion is simply based on the growing number of vendors that advertise their data science and machine learning bona fides combined with the use of algorithms and deep learning in their products. All the enterprise needs to do is buy software or even update what they already have to the latest version, and their departments are using machine learning algorithms to influence and drive front line security controls (EDR, email filtering, log file correlation, PAM, etc.).

For me, data science is much more foundational to the way I do security control design. In my last two positions as CISO, first at Aetna and now at MassMutual, the first person I hired was a Chief Data Scientist. Initially, I thought I would get better analytics, which I did. We were able to identify patterns and produce analytic results to help make better decisions on how to allocate scarce resources to the highest risk. I thought this was the ultimate goal.

What I didn't know seven years ago when I started down this path was that by running data models against data sources in real time using streaming technology, it's possible to use those models to segment both customers and internal users based on behavioral attributes at a very granular level. Connecting this segment at a point in time to a specific treatment executed in an orchestration or workflow engine (IAM platform for provisioning/de-provisioning, DLP solution, CASB, etc.) enables an action (allow access, limit access, monitor privilege) for specific behavioral patterns. This orchestration represents real time decisions driven by behaviors and what I refer to as *model-driven security.*

Passwords are Obsolete

One prime example of model-driven security is continuous authentication. For 60 years, the enterprise has used a primary set of credentials -- a username and password -- as the primary means of authentication. It's worked remarkably well, so well that the idea that authentication is an event with a binary outcome is literally wired into the brains of security professionals. Once a consumer provides the correct key, the consumer is trusted in the network. In fact, the problem today isn't the keys, it's us humans. We just have too many digital assets that require authentication to remember passwords for, so we compensate by using the same combinations over and over. Even when using password managers, people often use repeat combinations rather than the long strings of random characters that are automatically generated for them. And multi-factor authentication only adds another layer of binary controls to the authentication event, at the cost of substantial friction to the user. In some cases, 30-50% of users simply opt out of using MFA in digital systems because they see the cost outweighing the benefit. MFA is creating a poor consumer digital experience that consumers are avoiding.

Cyber criminals and fraudsters have long since figured out that it is infinitely easier to get the keys to the house and walk in the front door than to find other ways to break in. With data science and automation, they harvest billions of user id/password combinations to millions of websites and put them up for sale or exchange on the

dark web. Using tools like Sentry MBA, buyers can try out those credentials on active websites. On an average list of 10,000 credentials, the hit rate is 2%. It might not sound like much, but that's 200 accounts with potentially valuable information which can then be aggregated and re-sold, linked to money-mule accounts, used to make fraudulent purchases, or used as legitimate email addresses from which to send phishing emails.

This hits not just the consumer's bottom line, but the enterprise as well. Today, 50-90% of all digital log-in attempts are criminals attempting credential stuffing – which means that most of the cost of your IT infrastructure is paying for criminals attempting to log into your site. If you told a board member with fiduciary responsibility that, you can be sure they'd ask you to find a better way.

With model-driven security, there is a better way. Here's how it works:

1. **Capture:** The enterprise captures many independent behavioral attributes such as location, time of use, commonly used applications, and even how the user holds the device.
2. **Algorithm:** From this, a pattern of behavior for that specific attribute is represented as a number or algorithm (mathematical representation of an event). This becomes a baseline reference to be compared against the attribute data captured in real-time during a web or mobile session.
3. **Deviation Score:** This results in a deviation score for each attribute at that point in time. Combining this with many deviation scores from multiple attributes can be represented by a single aggregated score that determines a confidence level.
4. **Confidence Level:** That confidence level score can be fed to any application continuously, in real time, to enable it to take action within pre-determined threshold levels. If the confidence level is high, then full access to the site or app is provided. If the confidence level decreases beyond the threshold, access is restricted.

To be clear, this is does not require highly sophisticated AI. Financial services firms have been writing rules into processing systems for 20 years. You take a pattern of behavior, represent it mathematically, use that as a cornerstone measure to measure real time behavior against the pattern, come up with a deviation score, and then assign a treatment based on the score. If the deviation is too large, revoke access.

Better, Faster, Cheaper

This approach has clear benefits. First, the user experience is much better – the user literally does nothing; no passwords, tokens or third party authenticator apps are necessary. Second, the cost is lower. You save all the time spent on authenticating and resetting usernames and passwords, and all that cost of your digital infrastructure being used by criminals. And third, the security is better. Letting behavioral analytics continuously drive access decisions means the human is removed from the process, so the human can step back and study trends and focus on controls design as opposed to implementation. With the right infrastructure investment and

skills, it becomes a closed loop model that scales from the smallest to the largest enterprise. Simply put: this is a game-changer.

The catch, beyond the initial investment in the people and systems, is that moving to this system requires a paradigm shift in the minds of the security team. They must unlearn the fundamental tenet of cybersecurity that authentication is a binary event with a beginning and end. This turns out to be harder than it sounds, which is part of the reason we still have passwords when they are clearly marching towards obsolescence. The other reason we do not yet see this at wide scale across the industry is that with large firms who have hundreds of apps in use both internally and externally, this kind of transformation is a years-long process. But it is happening at the most sophisticated firms. And it will happen across many more security processes, across the industry.

The Insight

Data science is fundamentally transforming cybersecurity. By establishing behavioral models and continuously measuring them against user actions, anomalies can be detected and treated in real time, without human intervention. The result is a better user experience for the customer and better security at a cheaper cost. This change will require a paradigm shift by security teams, but it can and will happen at every scale, across the industry.