



BEHAVIORAL ANALYTICS COULD EXTEND IDENTITY MANAGEMENT SECURITY

Interview: Stephen Coggeshall
Mar 19, 2009

The traditional tools of identity management — user names, passwords, ID cards and biometric information — can fall short of ensuring that people are who they say they are. Behavioral analytics and, more recently, trusted reference networks are now widely used in the financial industry as a third tool for managing information access.

Stephen Coggeshall, chief technology officer at ID Analytics, has been a pioneering proponent of the use of advanced mathematical analytics in information management at Morgan Stanley, Citicorp, MasterCard and the Internal Revenue Service. He recently spoke with Government Computer News Editor-In-Chief Wyatt Kash about the state of identity analytics and their potential for government-run networks.

Q: One of the challenges of identity management is not only verifying who is accessing a network but identifying their roles in accessing specific information. How is behavioral analytics being used to address those needs?

A: It depends on the sensitivity and level of complexity of the transaction. For highly sensitive information, we need to move beyond a two-factor approach — things that you know, [such as] user name, passwords, and things that you have, physical things such as secure ID cards, biometrics that restrict access — and move toward how you behave.

We do that with an identity score that tells you not just in this transaction but in other events [associated] with this identity: What's the level of risk around this identity? Perhaps that person put in a user name and password, and that password was compromised. Or the physical device may have been compromised. So there's another way of measuring how dangerous this transaction may be by bringing in an external score.

What we do is gather information about how an identity is interacting in the world — how, for example, that Social Security number is being used in multiple transactions or events across different industries. Is that number being used in conjunction with a variety of names or just one name? What are the address dynamics around that particular identity — phone numbers, e-mails and IP addresses? We aggregate information on how the components of that identity are connected and use a fairly advanced algorithm to come up with a score that [defines] the likelihood that there's risk around that identity.

Q: What sources are you using as references?

A: It's less about the specific sources of the data and more about how the data is combined. We have built a trusted network where generally our customers contribute data about events going on with the customers in their portfolio. We gather that data across a lot of industries — financial services, telecommunications, mortgages, retail consumer products — and we aggregate it with auxiliary data, such as phone book data, Census Bureau data, other demographic data. We combine it in an intelligent way to form the basis of this data network we use to make these decisions.

So for the government, for example, building this kind of topology, they would take this same general approach, but they would combine data into this trusted network from across their different organizations.

Q: How do you define and verify a person's role in a given transaction?

A: The primary methodology [for making those decisions] is set by the business and business use. For example, if a parent is accessing a brokerage account of a child and just wants to see the balances, then the level of authentication would be a certain level. But if that person wants to transfer money in or out, then the risk score threshold may be set higher. So the [network manager] decides what level of authentication and how wide the access to particular actions may be and what additional actions may be required.

Q: How far have tools like behavior analytics and fuzzy logic come in terms of reliability for identity management?

A: They continue to get better. This kind of behavioral scoring started 30 or so years ago, first in credit scoring. They were fairly rudimentary linear algorithms. There have been substantial advances in business uses with consumer behavior scoring, for segmentation, for product offers. The IRS became an early adopter, using it for fraud protection with algorithms that looked for unusual returns and, in the early 1990s, for tax preparer fraud.

Q: What needs to happen next for those tools to become more effective or widely used?

A: We have constant improvements in our algorithms and data. The big breakthrough, emerging now, is the way to combine all three methodologies — what you know, what you have and what you do — in a blended solution.

Q: What other methods do you see evolving to speed identity authentication?

A: There are two primary needs to make substantial improvements. One is this three-pronged approach we've been talking about. The second is the architecture of the data and data communications in this trusted-networks concept that will help maximize the privacy protection, at the same time giving [network managers] the ability to make these important decisions.

Q: What safeguards are being developed to ensure that privacy laws and protections are still being followed?

A: There are several really important principles that have to be followed. We need restricted data access — who has access to what information — and [we need] to minimize that access. Another is to make sure data is end-to-end encrypted. Even at rest on disk, we keep all our data encrypted. You need to have careful controls and processes around releasing data. In the trusted-networks model, you don't need to release data. And around all these things you have to be [Payment Card Industry] compliant.

Q: How would you compare the government's progress using those tools versus what's being tried in the financial industry?

A: Financial services tends to be a technology leader, while government has tended to lag in adopting these technologies. But I do see more enlightened leaders in certain agencies — Homeland Security, for instance — and some of the other three-letter agencies have been leaders in being savvy and technology-driven.

These agencies have done well at bringing together vast amounts of data. What they've needed are algorithms to sift through very large quantities of largely unstructured data — text, video or voice — and find the needles in the haystacks.

I see two very broad trends in technology right now. One is this explosion of data and the need to work with unstructured data. The second is the ever-increasing necessity to understand who people are, how they're connected and defined, and how they exhibit themselves in the world, and then what kinds of access they should be permitted to have to various places.

Q: Which lessons are transferable and which aren't?

A: This three-pronged approach for identity intelligence we've talked about is very applicable in the government world. Associated with that is this concept of trusted networks. There's a major paradigm shift going on in the business community, where we're moving away from shipping around very sensitive data and more toward pooling information into a trusted network. And then granular-level data doesn't need to be released from that.

[Under] the old paradigm...in order to make decisions about a person, [organizations] would ship off sensitive data to some data broker. And that broker would send additional sensitive data back to the decision-maker. So there was a lot of very sensitive data being sent around the country, being gathered and even sold. There was a whole food chain that led to the selling of people's personal information. Privacy has become a huge concern because of that.

In this trusted-networks concept, people send us this sensitive data, and what they want is decisions. They don't need the granular-level data. We send back high-level information, like a score or some reason codes, that allows that business to take an action without having to ship back sensitive data. So this emerging paradigm is completely transferable to the government.