Realtime
publishers

Strengthening Application Security with Identity Fraud Prevention Systems

The Essentials Series

Incorporating Identity Verification and Knowledge-Based Authentication into Your Applications

sponsored by

ACXIOM®
WE MAKE INFORMATION INTELLIGENT™

Ken Hess

Introduction to Realtime Publishers

by Don Jones, Series Editor

For several years now, Realtime has produced dozens and dozens of high-quality books that just happen to be delivered in electronic format—at no cost to you, the reader. We've made this unique publishing model work through the generous support and cooperation of our sponsors, who agree to bear each book's production expenses for the benefit of our readers.

Although we've always offered our publications to you for free, don't think for a moment that quality is anything less than our top priority. My job is to make sure that our books are as good as—and in most cases better than—any printed book that would cost you $40 or more. Our electronic publishing model offers several advantages over printed books: You receive chapters literally as fast as our authors produce them (hence the "realtime" aspect of our model), and we can update chapters to reflect the latest changes in technology.

I want to point out that our books are by no means paid advertisements or white papers. We're an independent publishing company, and an important aspect of my job is to make sure that our authors are free to voice their expertise and opinions without reservation or restriction. We maintain complete editorial control of our publications, and I'm proud that we've produced so many quality books over the past years.

I want to extend an invitation to visit us at http://nexus.realtimepublishers.com, especially if you've received this publication from a friend or colleague. We have a wide variety of additional books on a range of topics, and you're sure to find something that's of interest to you—and it won't cost you a thing. We hope you'll continue to come to Realtime for your educational needs far into the future.

Until then, enjoy.

Don Jones

Introduction to Realtime Publishers i

Incorporating Identity Verification and Knowledge-Based Authentication into Your Applications 1

Identity Proofing: A Two-Step Process 1

Verification of Account Data 1

Knowledge-Based Authentication 2

Non–User-Generated Data Elements 3

How Are Non–User-Generated Data Assets Better Than the Data I Already Have? 3

Understanding Data Breadth and Depth 3

Integrating Knowledge-Based Authentication into Existing Applications 5

Summary 5

Copyright Statement

Incorporating Identity Verification and Knowledge-Based Authentication into Your Applications

What are the implications of using non–user-generated data elements in an application? More data points and data points not generated by the user are more secure because they don't depend on user input or a set of standard questions. Data service customers need to know what's required to integrate this service into their applications so that the service is transparent to the end user. Potential customers also want to know what the additional data points mean to them in terms of value and security.

Identity Proofing: A Two-Step Process

Identification is the process of verifying that an identity is bound to the person (claimant) making the claim to that identity. The greatest challenge in this process is that identification must occur remotely (not face to face). Remote verification must ensure, with reasonable assurance, that the identity requested matches the claimant.

This verification process is the first step in positively connecting an identity to the person who claims it. The second step in the process requires that the person making the identity claim provide knowledge-based information that assumes that only the verified identity (person) would possess. The two steps are verification of account data and knowledge-based authentication.

Verification of Account Data

Identity verification is your first defense against identity theft and fraud. The user enters data known to her at this first level to begin the process. The information used at this point generally excludes all other possible users and generates a confidence level number to qualify that the claimant information and the information match. Once the claimant positively verifies herself, she proceeds to the next step.

Traditional Two-Factor Authentication Explained

The most popular form of multi-factor authentication is two-factor authentication. When most people think of authentication, they think of secret passwords or answers to questions such as “What is your mother’s maiden name?” But, these “factors” may also take the form of physical attributes, bits of information, or tokens. Security experts have broken down these factors into simple terms so that everyone can grasp this elusive concept:

- Something You Are—Fingerprints, voice, retina, and so on
- Something You Know—Passwords
- Something You Have—Challenge-response lists, smart cards, security tokens, and so on

Higher-level (more secure) systems combine these approaches into *multi-factor authentication*.

Currently, a huge amount of research in the category of “something you are,” or biometrics, confirms the need for higher-level authentication methods. The theory propelling this biometric research frenzy is that biometric data assumes that physical characteristics are more difficult, if not impossible, to replicate by a fraudulent claimant attempting to gain access to personal information or accounts.

Note

You might see identity proofing referred to as multi-factor authentication or e-authentication.

Knowledge-Based Authentication

Knowledge-based authentication is the second step in the identity proofing process that typically consists of answering questions based on previously given answers to “canned” questions. Standard questions make it easier for fraudulent users to bypass an otherwise secure system. This is the “proofing” step where, based on information provided in the first step, the claimant offers tokens (bits of knowledge) to prove her claimed identity.

For example, if you answered 20 questions about yourself, your family, your hobbies, your work, and your general interests, would a criminal have your password to your bank account or your online shopping account? It’s likely that he would. It’s also likely that he would have no trouble gaining access to your other private information with these few bits of data. A well-conceived authentication scheme avoids the probability of guessed information from the 20 questions interview.

There is a way to confidently protect a legitimate claimant and her private information without using standard questions and responses. It's known as using non–user-generated data elements.

Non–User-Generated Data Elements

By using data elements that aren't user generated, you effectively deter fraudulent users from gaining access to your account and other private information. These data elements produce randomly-generated questions from stored data assets and require that the user possess an in-depth knowledge of her history and personal information. These randomly-generated questions are based on data gathered from a variety of public and private sources and require no input from the user.

How Are Non–User-Generated Data Assets Better Than the Data I Already Have?

These non–user-generated data assets comprise hundreds of data points and elements that, while known to the data owner (individual), aren't tied to any set of standard questions posed to the individual or her responses. These randomly-generated data elements provide an almost infinite array of prompts to which the user responds and authenticates successfully.

The value of any data asset collection is proportional to the depth and breadth of its elements. Some data asset providers have created databases with more than 900 possible individual data elements for each identity in the database. And, each one of those elements potentially contains several records. You can grasp how the records of more than 300 million individuals could swell to contain more than six billion records.

Data Flux

Consumer data is constantly changing. Think of the number of marriages, divorces, births, deaths, address changes, employment changes, and bankruptcies that occur in a single year to gain an idea of how quickly data becomes obsolete.

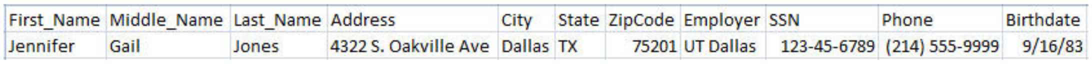
When selecting data assets, you want the most records available and the largest number of individual data points within those records. You also want fresh data that contains regularly updated information.

Understanding Data Breadth and Depth

As briefly covered in the previous section, data breadth and depth are extremely important when using data assets for identity-proofing applications. Data breadth refers to the number of possible data points gathered and encompasses a wide range of topics. Data depth is defined as the number of associated records for each of those collected data points. However, even terabytes of data points are useless without the ability to link and cross-reference them in ways that make it possible to uniquely identify an individual for accurate knowledge-based authentication. This cross-referencing and building viable knowledge-based authentication exams is a difficult process. The technology to do so, coupled with a broad and deep data set, is a significant data provider selection point.

Take your current home address as an example. Your address is a data element. The list of former addresses over the past 20 years would give you the depth of that element. If you've had six addresses during that time period, they would all appear in the database, and you could query them for use in an authentication exam.

Consider the identity record in Figure 1.1. Now imagine this record being more than 900 columns wide and having several rows of associated information. That is the reach of the information described here.



First_Name	Middle_Name	Last_Name	Address	City	State	ZipCode	Employer	SSN	Phone	Birthdate
Jennifer	Gail	Jones	4322 S. Oakville Ave	Dallas	TX	75201	UT Dallas	123-45-6789	(214) 555-9999	9/16/83

Figure1.1: An example of data elements in a database.

The following is an example of how these data assets are used in a real-world situation.

> Jennifer G. Jones of Dallas, TX decides to purchase a new car and calls the 800 number for a car loan application. The Customer Service Representative (CSR) for ABC Auto Finance asks Jennifer for her address. Jennifer reports her current address as 4322 S. Oakville Ave, Dallas, TX, 75201.
>
> The CSR verifies that, indeed, Jennifer G. Jones lives at the reported address with a very high confidence number that the information is accurate. Jennifer also has no alerts or flags associated with her name. She also isn't on any government "watch" list.
>
> The CSR proceeds to authenticate Jennifer by asking her more pointed questions, such as her Social Security Number, employer name, salary, number of years at her current address, and a few other data points to see how well they match with the CSR's data and to qualify her for a loan. The questions asked are a mixture of those to qualify Jennifer for the loan and to satisfy the lender that the CSR is speaking to Jennifer G. Jones of 4322 S. Oakville Ave, Dallas, TX, 75201, a 26-year old math professor at a local university and not an imposter posing as Jennifer.
>
> When Jennifer arrives at the auto dealership to purchase her new car, her identity is then physically matched to the information given to the lender by phone.

Now you have a better understanding of how an individual identity could have hundreds of associated data points. In addition, you understand how using a database that contains a variety of elements and their histories would prove valuable in your efforts to accurately match claimants and identities through varied and difficult-to-correctly-guess exam questions.

It would be difficult for an imposter to pose as Ms. Jones, falsely qualify for the car loan, and then sign for that loan. The CSR has too much information at her disposal to allow a falsified loan application to be processed. If those data assets weren't broad enough or have enough depth, the confidence level in that loan wouldn't exist.

Integrating Knowledge-Based Authentication into Existing Applications

If your data asset provider allows you to integrate their data into your applications, they should supply you with the real-time API with which you'll use to query their database. Some providers also offer consulting services to assist you in these efforts.

Your consultants will work with you to create your XML via the Web Services Description Language (WSDL) so that the actual database schema won't matter to you. This method greatly simplifies your database queries and your integration efforts.

Another focus of the consulting services is to help you create an authentication exam strategy. You'll have to decide how many questions to use for authentication exams, success levels, number of possible responses, and whether to give a claimant the opportunity to repeat a failed exam.

Five Questions to Ask a Prospective Data Asset Provider

- How many individuals are represented in your database?
- How often are the records refreshed in your database?
- Do you use non–user-generated data?
- Do you provide an API for application integration or do we have to use your portal?
- How current is your data?

Summary

Identity proofing imparts a high degree of confidence to banks, insurance companies, and credit card companies in mitigating the risk of fraudulent loans and stolen assets. Identity proofing also protects the consumer from identity theft. It is of extreme importance to provide the maximum level of safety to the data asset buyer, as well as the consumer, by ensuring that the data breadth and depth, data accuracy, data delivery, and data update frequency are of the highest quality possible.

The second article in this series covers security and compliance issues surrounding data assets, their gathering, and use. And, the third article presents an evolutionary view of data delivery into the future of computing and the special risks of cloud and mobile-based interactions.