

Echt anonymisiert

Softwaretests stellen Unternehmen vor eine kritische Entscheidung: Soll der fehlerfreie Ablauf in der Praxis gewährleistet sein, müssen in den Probeläufen realitätsnahe Datensätze zum Einsatz kommen. Allerdings wären dabei echte Kundendaten nicht ausreichend geschützt.

> Gemäß einer Untersuchung der Nationalen Initiative für Internet-Sicherheit (NIFIS) verwenden 64 Prozent aller Unternehmen echte Kundendaten in ihren Test- und Entwicklungsumgebungen. Diese Vorgehensweise ist nachvollziehbar. Denn nur Echtdateien liefern Testergebnisse, die unangenehme Überraschungen im Praxisbetrieb ausschließen. Aus Sicht des Datenschutzes ist dieses Vorgehen allerdings mehr als fahrlässig. Denn die

„Die Verwendung von Kundendaten in einer schwach gesicherten Testumgebung verstößt gegen geltendes Recht.“

Verwendung von Kundendaten in einer schwach gesicherten Testumgebung verstößt gegen geltendes Recht, wie beispielsweise das Bundesdatenschutzgesetz. Empfindliche Strafen für das Unternehmen drohen.

Noch schwerer wiegt aber der Vertrauensbruch gegenüber Kunden und Geschäftspartnern. Denn mit den Daten sind immer Rechte Dritter, Persönlichkeitsrechte oder vertrauliche Informationen von Unternehmen verknüpft. Kunden und Geschäftspartner verlassen sich zu

Recht darauf, dass sensible Informationen über sie mit höchster Vertraulichkeit behandelt werden. Dem steht entgegen, dass Entwicklungssysteme gegen Datendiebstahl bei weitem nicht so gut abgesichert sind wie Produktionssysteme. Das hat meist folgende Gründe:

- In Entwicklungssystemen müssen Rechte großzügiger vergeben werden, um effizient arbeiten zu können.

- Im Entwicklungsstadium kann eine Software noch nicht so stabil laufen, dass ein versehentlicher Verlust der Vertraulichkeit sicher ausgeschlossen werden kann.

- In Testumgebungen werden vorwiegend Fehler provoziert, um Sicherheitslücken aufzudecken. So ist es für einen Datendieb kein Problem, die Daten am Testsystem zu kopieren.

Unternehmen, die eine neue Software entwickeln oder einführen, stecken also in einem Dilemma: Einerseits brauchen sie Echtdateien für effektive Tests. Andererseits können sie die Sicherheit der Daten in einer Testumgebung nicht gewährleisten. Mehrere Auswege bieten sich an: Eine Möglichkeit ist es, weiterhin Produktionsdaten in Test- und Entwicklungssystem zu verwenden. Um ihre Sicherheit zu gewährleisten, wird die Testumgebung aber nach den gleichen Standards abgesichert wie Produktionssysteme. Unter diesen Bedingungen würde die Softwareentwicklung je-

Auszug aus...



Ausgabe 05/2010



Softwareentwicklungssysteme sind gegen Datendiebstahl bei weitem nicht so gut abgesichert wie Produkivsysteme.

doch massiv behindert und viele Tests wären nicht durchführbar.

Echtbetrieb wird zum Lotteriespiel

Das Risiko, dass Daten in die falschen Hände geraten, wäre reduziert. Allerdings stiege das Risiko, mit einer unausgereiften Software in den Echtbetrieb zu gehen. Eine weitere radikale Lösung wäre es, Produktionsdaten im Entwicklungs- und Teststadium grundsätzlich nicht zu verwenden. Aber auch hier überwiegen die Nachteile. Zwar wären die echten Datensätze absolut sicher, doch sind viele Test-szenarien oder Massentests ohne die Verwendung von Produktionsdaten nicht aussagekräftig genug. Der Echtbetrieb würde zum Lotteriespiel. Auch diese Option ist daher nicht empfehlenswert. Die dritte Möglichkeit heißt Testdaten-anonymisierung. Dabei kommen zwar echte Daten zum Einsatz, sie werden aber so verfremdet, dass die originalen Datensätze daraus nicht rekonstruiert werden können. Oberste Priorität hat auch dabei, dass die verwendeten Daten keinerlei Rückschlüsse auf die Originale erlauben. Trotzdem sollten sie repräsentativ sein, also die Aussagen und Beziehungen der echten Daten beibehalten. Zudem müssen die anonymisierten Daten jederzeit reproduzierbar sein, um das Laufzeitverhalten der Daten realistisch

beurteilen zu können. Bislang war es üblich, die Daten zu maskieren, also mit Phantasiewerten zu überdecken, die aber die wesentlichen Aussagen (Geschlecht, Alter etc.) wiedergeben. Der Nachteil dieser Vorgehensweise liegt darin, dass es je nach Unternehmen auf unterschiedlichste Strukturelemente ankommen kann. Für eine Branche ist die Geschlechterverteilung interessant, eine andere benötigt eher die exakte Altersstruktur. Es ist also eine diffizile Aufgabe, die Daten so zu maskieren, dass ihre statistische Aussagekraft in jeder relevanten Hinsicht erhalten bleibt. Der Anbieter DV-Ratio hat eine Methode entwickelt, die keine statistische Verfälschung zulässt. Nach einem intelligenten System werden Daten des originalen Datenbestandes gemischt und neu kombiniert. Die Inhalte der einzelnen Datenfelder bleiben unverändert. Ausgeklügelte Algorithmen tauschen aber die Inhalte von Feldern unter Einhaltung von statistischen Verteilungen. Aus den Feldern mehrerer echter Daten ergibt sich ein neuer, scheinbar zufällig zusammen gewürfelter Datensatz. So ist der Ursprungszustand nicht mehr rekonstruierbar. Trotzdem sind die Testdaten jederzeit reproduzierbar. Die statistische Aussage bleibt dabei erhalten, da die einzelnen Felder der Originaldaten in ihrem ursprünglichen Zustand bleiben. <

WILHELM GEIGER