Bild: Thomas Kuhlenbeck

Rausgerutscht

Datenleck in Österreich durch Azure-Workflow

Interne Links einer Microsoft-Azure-Anwendung sind in den Suchindex von Bing geraten – personenbezogene Daten waren dadurch öffentlich einsehbar. Eine Spurensuche offenbart, wie schnell vermeintlich geheime Links zum Sicherheitsrisiko werden – nicht nur bei Azure.

Von Jan Mahn

Das soll so nicht sein, dachte sich ein c't-Leser Mitte Juni, als er in seiner bevorzugten Suchmaschine DuckDuckGo nach etwas ganz anderem gesucht hatte und die Suchergebnisse sah. In Kombination mit dem Suchwort „ifa“, unter dem man eher Treffer zur Funkausstellung in Berlin erwarten würde, fand er DIN-A4-Seiten im Querformat, betitelt mit „Klientenkarte“.

Die Bögen enthielten je ein Foto sowie diverse personenbezogene Daten wie Geburtsdatum, Sozialversicherungsnummer und Verwandtschaftsverhältnisse. Auch ein Bezug zum ursprünglichen Suchbegriff war vorhanden: Die Dokumente waren jeweils mit einer individuellen Nummer namens IFA versehen. Was es damit auf sich hatte, verrieten die Dokumente nicht. Unser Leser verfeinerte seine Suchanfrage und konnte DuckDuckGo sowie Microsofts Suchmaschine Bing (aus deren Daten sich auch DuckDuckGo bedient) dazu bringen, eine ganze Ergebnisseite mit solchen ominösen Klientenkarten auszuspucken. Mit diesen Erkenntnissen und der Suchanfrage wandte er sich an die c't-Redaktion.

Wir konnten das Problem bestätigen und machten uns auf die Suche nach dem Verantwortlichen für diese offenbar nicht geplante Veröffentlichung. Doch es war gar nicht so einfach, einen Betreiber auszumachen. Wo eine IFA als eindeutige Nummer verwendet wird, konnten wir mit einer Websuche nicht ergründen, und auch die Domain half nicht weiter: Die Klientenkarten lagen auf einer Subdomain von westeurope.logic.azure.com und waren Bestandteile von Microsofts Produkt „Azure Logic Apps“. Damit können Firmen und Behörden Arbeitsabläufe digitalisieren – Formulare, Freigabeprozesse, Listen und Auswertungen.

Vieles sprach dafür, dass irgendjemand seine Klienten über einen solchen Azure-Workflow verwaltet und die veröffentlichten A4-Dokumente ein End- oder Zwischenergebnis eines Verwaltungsprozesses sind. Ein Name des Betreibers war in der Azure-Adresse und in den Dokumenten nicht enthalten, wir mussten also weitersuchen.

Doch wer nennt seine Kunden schon „Klienten“, ein Anwalt vielleicht? Auffällig war die Nationalitäten der Klienten – vor allem Syrer und Afghanen. Das deutete darauf hin, dass es sich um ein Verfahren zur Registrierung von Geflüchteten handeln könnte. Also kontaktierten wir die für uns naheliegendste Adresse, das deutsche Bundesamt für Migration und Flüchtlinge. Vielleicht hatten wir eines ihrer Systeme gefunden. Und selbst wenn nicht, könnten sie vielleicht immerhin verraten, was eine IFA-Nummer ist und zu welcher Organisation sie gehört.

Die Pressestelle des Bundesamtes meldete sich umgehend telefonisch und berichtete von einer eifrigen internen Suche nach der Herkunft der Datensätze. Das Fazit der Recherche: Aus Deutschland könnten die Datensätze sicher nicht stammen, das Bundesamt nutze keine Azure-Cloud-Produkte und auch eine zehnstellige Sozialversicherungsnummer auf den

Dokumenten passe nicht nach Deutschland. Einen heißen Tipp hatte man doch für uns: Der Begriff „Klient“ und die Sozialversicherungsnummer könnten nach Österreich gehören. Die deutschsprachige Schweiz hatten wir zuvor bereits ausgeschlossen, weil die Dokumente ein ß enthielten, und so folgten wir der Spur nach Wien.

Spurensuche in Österreich

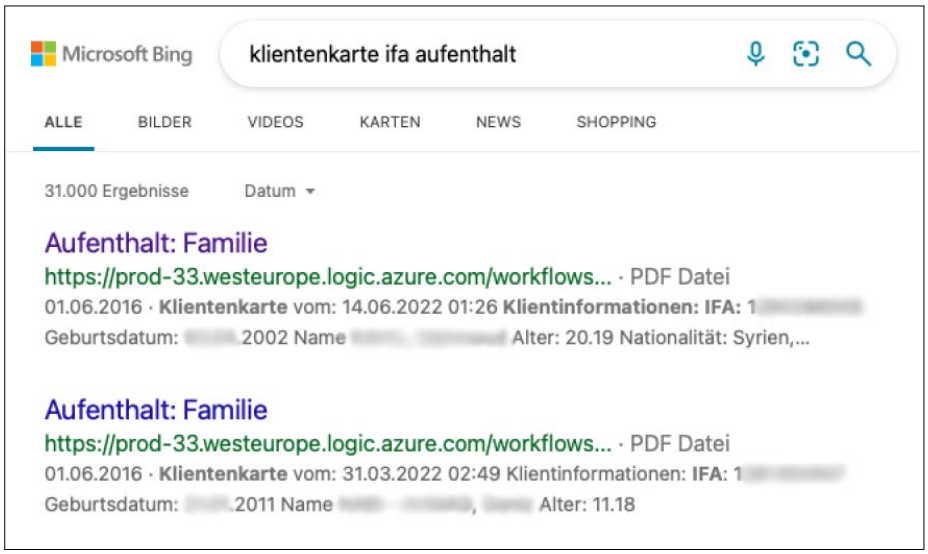
Unsere nächste Anfrage ging daher ans österreichische Bundesministerium für Inneres (BMI), dem das Bundesamt für Fremdenwesen und Asyl unterstellt ist. Volltreffer: Am nächsten Tag meldete sich der Pressesprecher telefonisch und bestätigte, dass man den Verursacher ausgemacht habe, und zwar die Bundesagentur für Betreuungs- und Unterstützungsleistungen (die BBU GmbH mit der Republik Österreich als einziger Gesellschafterin). Sie ist verantwortlich für die Betreuung und Beratung von Asylbewerbern und hat zur Registrierung ein Online-Verfahren auf Basis von Azure entwickeln lassen.

Nach unserem Hinweis begann das Unternehmen direkt mit der Fehlerbeseitigung und bestätigte uns später: „Wir haben von den einsehbaren Daten am Mittwoch, den 21. Juni 2022 um 09:40 Uhr erfahren und konnten diese Einsicht bereits am selben Tag um 10:18 Uhr schließen.“ Das deckt sich mit unseren Beobachtungen. Doch ein Problem blieb bestehen: Die Suchmaschinen Bing und DuckDuckGo zeigten die Treffer weiterhin an, auch wenn die verlinkten Seiten nicht mehr erreichbar waren. Und in der Vorschau der Suchergebnisse standen ausgerechnet alle personenbezogenen Daten der Betroffenen als Fließtext hintereinander.

Am Telefon schilderte uns der Sprecher der BBU, was hinter den Kulissen passierte: Das Unternehmen, das den Azure-Workflow eingerichtet hatte, betreue noch andere Azure-Workflow-Kunden und stehe schon in Kontakt mit Microsoft. Kern der Untersuchung sei die Frage, ob es vielleicht ein generelles Problem gebe und über eine problematische Microsoft-interne Abkürzung massenhaft vertrauliche Links bei Microsofts Suchmaschine Bing landen. Eine Robots.txt-Datei, über die Suchmaschinen für gewöhnlich an Links kommen und die Seiten in den Index aufnehmen, konnte man nicht finden, auch keine anderen Anzeichen für systematisches Indexieren von Azure-Workflow-Links. Der Kontakt zu Microsoft war aber für ein anderes Problem nützlich: Die verwaisten Einträge im Bing-Suchindex verschwanden nach zwei Tagen spurlos. Für das Entfernen von Daten aus Suchmaschinenindexen ist das ein rasantes Tempo.

Arbeitshypothese

Einen grundsätzlichen Konfigurationsfehler konnten die Forschungen von BBU und der Microsoft-Partnerfirma nicht aufdecken, nur eine recht plausible Arbeitshypothese liefern. Die von der Suchmaschine



Waren nicht für die Öffentlichkeit bestimmt: Bei Bing tauchten Klientenkarten mit personenbezogenen Daten auf. Laut Domain waren sie Teil eines Workflows bei Microsoft Azure.

verpetzten Links zeigten nicht nur eine Klientenkarte an, sie dienten auch als Trigger, um den nächsten Schritt im Registrierungsprozess anzustoßen. Eigentlich waren sie nicht dafür gedacht, im Browser geöffnet zu werden. Weil Mitarbeiter der BBU diese Schritte aber ab und zu per Hand auslösen mussten, könnten sie die Adressen, so die Theorie, manuell im Browser geöffnet haben – konkret in Microsoft Edge, dem Standardbrowser der BBU. Genau dieser Schritt könnte das Leck verursacht haben.

Wie wir nachstellen konnten, reicht ein einziges Zeichen vor dem https:// einer URL, und Edge interpretiert eine Eingabe in der Adresszeile nicht als URL, sondern als Suchanfrage für Bing. Ein solches Zeichen ist beim händischen Kopieren und Einfügen schnell falsch kopiert und genug solcher Suchanfragen könnten Bing veranlassen, die eigentlich vertrauliche Adresse zu indexieren.

Kurzerhand probierten wir selbst, eine bisher garantiert von niemandem indexierte URL mit der Beschreibung eines eigens erfundenen Fantasietiers auf diesem Weg in den Bing-Index zu schleusen. Doch auch wiederholtes Suchen nach der URL mit mehreren Kollegen und verteilt über mehrere Tage konnte Bing nicht dazu bringen, die Seite zu indexieren. Die Arbeitshypothese der BBU, dass der Suchschlitz von Edge die undichte Stelle war, können wir damit weder bestätigen noch widerlegen; sie wirkt aber durchaus plausibel, weil nur Bing und nicht Google diese Daten fand.

Herausfinden konnten die Betreiber am Ende der Analyse, wie viele Datensätze betroffen waren: „Die anschließende Untersuchung durch ein internationales Expertenteam hat ergeben, dass aufgrund des spezifischen Verhaltens eines Workflows tatsächlich Daten von insgesamt 35 Asylsuchenden auf einigen Suchmaschinen zu finden waren. Für weitere Fälle wurden keine Indizien gefunden“, schrieb uns die BBU in ihrer abschließenden Stellungnahme. Die österreichische Datenschutzbehörde habe man über den Abfluss der Daten informiert.

Grundsatzproblem

Der Fall wirft ein Schlaglicht auf ein vielfach genutztes technisches Konzept und zeigt seine Schwächen auf: URLs, die etwas auslösen, Daten generieren oder anzeigen und die nur durch einen „geheimen“ Adressbestandteil geschützt sind, gibt es nicht nur bei Azure Logic Apps. Bei vielen Dateiablageplattformen einschließlich Nextcloud und Google Drive gibt es eine Möglichkeit, Links mit lesendem oder schreibendem Zugriff zu erzeugen und an andere zu verschicken.

Das Prinzip: Wer den Link kennt, darf zugreifen, man muss ihn also wie ein Geheimnis behandeln. Abgesichert ist das Konzept nur dadurch, dass der geheime Adressbestandteil ausreichend lang und zufällig generiert ist. Diese Bedingung gilt auch für die gefundenen Azure-Links, die nicht zu erraten waren.

Die Tücken lauern allerdings an vielen Stellen: Erste Schwachstelle ist der Mensch, der die vertraulichen Links vielleicht nicht allzu vertraulich behandelt. Sie werden (wie möglicherweise in diesem Fall) in Suchschlitze von Suchmaschinen eingetippt, an andere weitergeleitet, in cloudsynchronisierten Lesezeichenlisten gespeichert oder aus Bequemlichkeit aus der Firma ans private Mobiltelefon geschickt. Außerdem landen sie im Verlauf von Browsern und in Caches und können dort schlimmstenfalls von Unbefugten gelesen werden. Kurzum: Mit Links stellen viele Nutzer Dinge an, die sie mit ihren Passwörtern eher nicht machen würden.

Was tun?

Sollte man Adressen mit integrierten Geheimnissen also verteufeln und sich für ihre Ächtung einsetzen? Ganz so drastisch muss man vielleicht nicht vorgehen. Wer als Admin oder Entwickler Software mit solchen URLs bereitstellt, sollte aber gut abwägen, ob es sicherere Alternativen gibt und diesen den Vorzug geben. Sofern möglich sollten Abfragen immer nur nach Anmeldung mit Benutzername und Passwort gelingen, am besten abgesichert mit einem zweiten Faktor. Sobald personenbezogene Daten im Spiel sind, ist deren Schutz wichtiger als der Komfortgewinn durch eine gesparte Anmeldung.

Auch aus dem Grundschutz-Kompendium des BSI (siehe ct.de/y8ct) kann man eine Pflicht ableiten, solche URLs mit Geheimnissen durch eine Benutzeranmeldung zu ersetzen. Im Abschnitt ORP.4 zu „Identitäts- und Berechtigungsmanagement“ heißt es: „Der Zugang zu schützenswerten Ressourcen einer Institution ist auf berechtigte Benutzer und berechtigte IT-Komponenten einzuschränken. Benutzer und IT-Komponenten müssen zweifelsfrei identifiziert und authentisiert werden.“ Mit anonymen Links, die womöglich von mehreren Mitarbeitern genutzt und fleißig umher kopiert werden, ist das definitiv nicht gewährleistet.

Wenn für eine Anwendung wirklich nur eine solche „geheime“ URL infrage kommt, kann man das Risiko für Missbrauch immerhin dadurch minimieren, dass man die Links nur in dem überschaubaren Zeitfenster funktionieren lässt, in dem sie unbedingt gebraucht werden – und keine Minute länger. *(jam@ct.de)* c't

BSI-Grundschutzkompendium:
ct.de/y8ct