UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR IT-SICHERHEIT

Übungsblatt 6

SQL und Datenbanksicherheit

Abgabe bis 30. Mai 2023 um 12:00 Uhr im Moodle

Cybersecurity im Sommersemester 2023

Thomas Eisenbarth, Jan Wichelmann, Anja Köhl

Einführung

Nach der kleinen Unterbrechung mit den Sicherheitsmodellen setzen Sie Ihre Analyse des Kundenportals fort. Die Verbindung ist inzwischen mit TLS 1.3 gesichert, und der Mailserver wurde ordnungsgemäß konfiguriert um die Authentizität von Rechnungs-E-Mails sicherzustellen. Aber waren das bereits alle Probleme?

Aufgabe 1 SQL-Injection (4 Punkte)

Sie öffnen das Kundenportal in Ihrem Browser und werden zur Eingabe von Benutzername und Passwort aufgefordert. Da sie keines von beiden besitzen, sollte ein Login nicht möglich sein. Sie sind jedoch kreativ und tippen als Benutzername

```
xyz' OR 1=1;-- a
```

ein, das Passwortfeld lassen Sie leer. Und, Überraschung! Sie sind plötzlich als `kunde9729` angemeldet und können dessen Daten vollständig einsehen.

Die Entwicklungsabteilung ist ob dieser scheinbar schon beinahe magischen Leistung völlig verdutzt und bittet Sie um eine Erklärung.

1. Erklären Sie die einzelnen Bestandteile Ihrer Eingabe und deren Wirkung. Nennen Sie hierzu eine mögliche SQL-Abfrage, die hinter dem Kundenlogin stehen könnte. Warum wurden Sie als `kunde9729` angemeldet?
2. Beschreiben Sie eine wirksame Gegenmaßnahme gegen derartige Angriffe.

Aufgabe 2 Blind SQL-Injection (4 Punkte) *[optional bei bestandenem Praktikum]*

Diese Aufgabe muss nicht bearbeitet werden, wenn Sie stattdessen das "SQL-Injections"-Praktikum bestehen. Bitte geben Sie Ihren Gruppennamen aus dem Praktikum an, um die Korrektur zu erleichtern.

Sich als beliebiger Kunde einzuloggen reicht Ihnen nicht. Da Sie nun schonmal angemeldet sind und noch mehr Fehler finden möchten, halten Sie Ausschau nach weiteren verwundbaren Eingabefeldern und werden schnell fündig.

Ein Kunde kann über die Website anhand einer Produktnummer ein Hostingprodukt unkompliziert hinzubuchen. Nach Eingabe der Nummer erscheint dann ein Formular, das zur Kontrolle der Rechnungsadresse auffordert. Nach einigem Probieren stellen Sie folgendes fest:

- Leider ist die Website derart schlecht programmiert, dass dieses Formular immer erscheint, auch bei ungültiger Produktnummer.
- Die Keywords `WAIT`, `DELAY` und `CASE` werden serverseitig gefiltert und haben keinen Effekt.
- Bei der Eingabe

  ```
  123' AND 1=1/(SELECT ASCII(SUBSTRING(HOST_NAME(),1,1))-100)
  ```

wird Ihnen statt der Website nur `500: Internal Server Error` angezeigt.

- Bei allen anderen Eingaben

```
123' AND 1=1/(SELECT ASCII(SUBSTRING(HOST_NAME(),1,1))-X)
```

mit `X = 0, 1, ..., 99, 101, ..., 255` erhalten Sie das normale Formular.

Erklären Sie dieses Verhalten, und wie sich dieses zur Extraktion von Daten ausnutzen lässt.

Aufgabe 3 Anonymität (12 Punkte)

Da Sie schonmal da sind, tritt Bob, der Leiter der Buchhaltung, an Sie heran: Im Rahmen der Regierungsinitiative *Digitalisierung* fordert das Statistikamt von ITS monatlich einige Kundendaten, um daraus Statistiken zur Verbreitung und Nutzung von Hostingangeboten zu erstellen. Eine begrenzte Anonymisierung ist erlaubt, um Rückschlüsse auf einzelne Kunden zu verhindern.

Im Folgenden findet sich ein Auszug aus der akkumulierten Datenbank (`kunden.csv` im Moodle):

ID	Name	PLZ	Typ	Vertragsbeginn	Anzahl Server
1	datenhändler24 AG	42424	Unternehmen	2019	5
2	Musikverein Musterstadt e.V.	42424	Verein	2015	1
3	Sportverein Elend e.V.	38875	Verein	2015	2
4	Carl Coder	23562	Privatperson	2017	2
5	Werbeagentur Witz GmbH	42424	Unternehmen	2016	1
6	Wäscherei Dieter Rogen GmbH	23556	Unternehmen	2017	9
7	Paul's Pizzaservice GmbH	40789	Unternehmen	2016	3
8	Laugengebäckfreunde Lübeck	23562	Verein	2012	1
9	Mega Marzipan GmbH & Co. KG	23560	Unternehmen	2016	2
10	Friseursalon Oberhaarz GmbH	38875	Unternehmen	2012	2

Die ID wird automatisch von der Datenbank vergeben und dient hier nur der Nummerierung, erlaubt also keine weiteren Rückschlüsse auf einen bestimmten Datensatz.

Können Sie Bob helfen, den Datensatz ausreichend zu anonymisieren?

1. Welche Spalten sind *Identifier*, welche Spalten sind *Quasi-Identifier*, welche sind *sensitive Attribute*? Begründen Sie Ihre Antwort.
2. Unterteilen Sie die Datensätze anhand der *Quasi-Identifier* in Anonymitätsmengen, und entfernen Sie dabei die *Identifier*. Bestimmen Sie anschließend den aktuellen Wert für k.
3. Modifizieren Sie die *Quasi-Identifier* mittels *Generalisierung*, sodass die Datenbank 3-Anonymität erreicht, und geben Sie die resultierende anonymisierte Datenbank an. Versuchen Sie, den Genauigkeitsverlust hierbei zu minimieren.

   Sie können annehmen, dass das Statistikamt primär am Kundentyp interessiert ist, und die genaue Postleitzahl (PLZ) eher nachrangig ist.
4. Durch eine andere Datenerhebung erfährt Statistikamt-Sachbearbeiter Stefan, welche Unternehmen in der anonymisierten Datenbank vertreten sind, allerdings kennt er nicht die Zuordnung von Datensätzen zu Unternehmen. Außerdem weiß er, dass die *datenhändler24 AG* im Jahr 2018 gegründet wurde.

   Welche Aussagen kann Stefan mit diesem Wissen über die *datenhändler24 AG* treffen? Welche Auswirkungen hat das auf die Anonymität der anderen Kunden in derselben Äquivalenzklasse?
5. Für eine andere statistische Erhebung zur Nutzung von Hostingangeboten wird ITS gebeten, über eine Web API direkten Zugriff auf die nicht anonymisierte Originaldatenbank bereitzustellen, um genauere Abfragen zu ermöglichen. Um dennoch einen gewissen Schutz zu erreichen, müssen Abfragen an diese Datenbank immer mindestens drei Datensätze zurückliefern, über deren Serveranzahl dann ein Durchschnittswert berechnet wird. Als Abfragefelder sind nur Postleitzahlen und der Vertragsbeginn zugelassen (Kombinationen und Abfragen über Bereiche sind erlaubt).

   So erfasst die Abfrage „Alle Kunden in PLZ-Bereich 4****“ vier Datensätze, als Durchschnittswert wird entsprechend $(5+1+1+3)/4 = 2.5$ zurückgegeben. Die Anfrage „Alle Kunden mit Vertragsbeginn zwischen 2010 und 2013“ wird hingegen abgelehnt, da diese nur auf zwei Datensätze zutrifft.

Zeigen Sie, dass diese Art der Anonymisierung unzureichend ist, indem Sie zwei Abfragen konstruieren, aus deren Ergebnissen sich die Serveranzahl der Firma *Mega Marzipan GmbH & Co. KG* eindeutig ableiten lässt.

Sie dürfen annehmen, dass die Anzahl der betroffenen Zeilen zusammen mit dem Durchschnittswert ausgegeben wird.