Datenschutz- und Sicherheitsanalyse von Mobilen Learning Apps

Sunny Dass, Michael Kreutzer, Linda Schreiber und Hervais Simo Fhom

Zusammenfassung

In den letzten Jahren hat die Popularität von mobilen Learning Apps für Kinder und Jugendliche stark zugenommen – insbesondere im Kontext der aktuell herrschenden Corona-Pandemie, in der der Präsenz-Schulbetrieb mehrfach stark eingeschränkt werden musste. Im Rahmen dieses Beitrags untersuchen wir, inwieweit Android Learning Apps vor dem Hintergrund der DSGVO die Privatheit ihrer Nutzenden (i. d. R. Minderjährige) gewährleisten bzw. Anforderungen an Datensicherheit erfüllen. Die Datengrundlage für die Untersuchung besteht aus 199 Learning Apps aus dem Google Play Store. Die Analyse unterteilt sich in zwei Schritte: die grobgranulare und die feingranulare Analyse. Die grob-

Diese Forschungsarbeit wurde vom Bundesministerium für Bildung und Forschung (BMBF) und vom Hessischen Ministerium für Wissenschaft und Kunst (HMWK) im Rahmen ihrer gemeinsamen Förderung für das Nationale Forschungszentrum für angewandte Cybersicherheit ATHENE unterstützt.

M. Kreutzer · L. Schreiber · H. Simo Fhom (✉)
Fraunhofer SIT, Darmstadt, Deutschland
E-mail: Hervais.SimoFhom@sit.fraunhofer.de

M. Kreutzer
E-mail: Michael.Kreutzer@sit.fraunhofer.de

L. Schreiber
E-mail: Linda.Schreiber@sit.fraunhofer.de

S. Dass
Technische Universität Darmstadt, Darmstadt, Deutschland
E-mail: sunny.dass@stud.tu-darmstadt.de

M. Friedewald et al. (Hrsg.), *Selbstbestimmung, Privatheit und Datenschutz,*
DuD-Fachbeiträge, https://doi.org/10.1007/978-3-658-33306-5_11

granulare Analyse befasst sich mit Beobachtungen und statistischen Erkenntnissen, welche direkt aus den bereits gesammelten Metadaten der Apps ersichtlich sind. Weiterhin werden die Ergebnisse hinsichtlich Datenschutz und Datensicherheit kritisch hinterfragt, indem Metadaten bezüglich Datenschutzerklärung und Berechtigungen eingestuft werden. Die feingranulare Analyse baut auf der grobgranularen Analyse auf. Hierbei wird das Android Package (APK) mittels Tools zur statischen und dynamischen Analyse genauer betrachtet. Des Weiteren werden das Vorhandensein und die Qualität von Maßnahmen zur Absicherung des Datenverkehrs der ausgewählten Apps untersucht und bewertet. Wir stellen fest, dass viele Learning Apps Datenschutzrichtlinien oder sichere Datenübertragung bieten, die Apps auf unsichere Weise implementiert sind und oft eine zum Teil niedrige Codequalität vorweisen, was auf zusätzliche Cybersicherheits- und Datenschutzrisiken schließen lässt.

Schlüsselwörter

Learning Apps • Privatheit • Datenschutz

1 Einführung

In der aktuellen Corona-Pandemie scheinen Learning Apps, welche von Kindern und Jugendlichen jederzeit und von überall benutzt werden können, eine besonders attraktive Alternative zu sein, um ihre schulischen Fertigkeiten weiterzuentwickeln Tengler et al. (2020) [12]. Die Vorteile solcher Learning Apps zeichnen sich vor allem durch die Bereitstellung eines vielfältigen digitalen Lernangebots aus. Trotz steigender Attraktivität bleibt jedoch die Frage offen, wie es mit Daten- und Privatheitschutz in Mobile Learning Apps steht. Dabei ist insbesondere die Frage zu beantworten, ob und wie privatheitsinvasiv und anfällig für Cyberangriffe Mobile Learning Apps sind.

Das Hauptziel dieser Arbeit ist es, eine umfassende Untersuchung der Datenschutz- und Sicherheitsproblematik im Zusammenhang mit Learning Apps aus dem Google Play Store durchzuführen. Die vorliegende Arbeit behandelt diese Frage auch vor dem Hintergrund der DSGVO, die beispielsweise mit Erwägungsgrund 38 einen besonderen Schutz für Kinder und Minderjährige vorschreibt (siehe Abschn. 2). Dies wirft weitere Fragen auf, denen wir im Rahmen unserer Studie nachgegangen sind, u. a.: Wie hoch ist der Anteil von Learning Apps ohne eine Datenschutzerklärung? In welchem Land sitzt der Anbieter? Welche Zusatzbibliotheken für Tracking und Werbung (die sog. Third-Party Libraries) sind im App-

Code eingebettet? Welche Berechtigungen werden i. d. R. angefordert? Welche/wie viele Sicherheitslücken und Fehlkonfigurationen im Quellcode gibt es? Um eben erwähnte Fragen beantworten zu können, wurde im Rahmen der Studie eine Vorgehensweise konzipiert, die aus zwei aufeinanderfolgenden und abgestimmten Schritten besteht: 1) Aufbauen eine Datengrundlage und 2) Durchführung der Analyse.

Als Datengrundlage, siehe Abschn. 3, werden 199 Android Learning Apps aus dem Google Play Store Germany in unsere Analyse einbezogen, 167 kostenlose Apps und 32 kostenpflichtige Apps. Pro App werden die sog. Android Package (APK-Datei) und entsprechenden Metadaten gesammelt, 33 Metadaten insgesamt. Zu den Metadaten gehören u. a. die Datenschutzerklärungs-URL, die Adresse des Entwicklers und die Gesamtanzahl der Bewertungen der App auf Google Play. Um festzustellen, inwieweit diese Learning Apps eine Bedrohung für die Privatheit und den Datenschutz ihrer Nutzenden darstellen, konzipieren wir eine Analyse bestehend aus zwei Stufen: eine grobgranulare Analyse (Abschn. 4) und eine feingranulare Analyse (Abschn. 5). Die grobgranulare Auswertung zielt auf eine Beurteilung der Privatheitsbedrohung einer Learning App ausschließlich auf Grundlage einer statistischen Interpretation ihrer Metadaten ab. Darunter fallen Statistiken zu einzelnen Metadaten, wie die Entwickleradresse bei der Ursprungslandsanalyse oder die durchschnittliche Bewertung, Anzahl an Installationen oder Kommentaren, welche bei der Popularitätsanalyse erstellt worden sind. Weiterhin werden die Ergebnisse hinsichtlich Datenschutz und Datensicherheit kritisch hinterfragt. Somit folgt eine Datenschutzerklärungsanalyse bei der die Verteilung zwischen Apps mit bzw. ohne Datenschutzerklärung zusammen mit anderen Metadaten in Verbindung gebracht wird. Zuletzt folgt eine Berechtigungsanalyse, wobei Berechtigungen hinsichtlich normaler und gefährlicher Berechtigungen statistisch dargestellt und analysiert werden. Die feingranularen Analysen konzentrieren sich dagegen darauf, die APK-Datei der gegebenen Learning App auf mögliche Schwachstellen und Fehlkonfigurationen zu überprüfen. Dabei werden zur Charakterisierung ihres Laufzeitverhaltens und ihrer Datensammlungspraktiken sowohl statische als auch dynamische Analysemethoden und Tools verwendet.

2 Learning Apps im DSGVO-Kontext

Für die vielfältigen Risiken der digitalisierten Welt sind insbesondere Kinder und Jugendliche in ihrer Entwicklungsphase anfällig. Zudem sind sich Kinder den Folgen der Verarbeitung von personenbezogenen Daten oftmals weniger bewusst oder

sie können diese schlechter kontrollieren.[1] Diese besondere Schutzbedürftigkeit wird auch von der Datenschutz-Grundverordnung (DSGVO) anerkannt.

Gemäß Erwägungsgrund 38 sollte ein besonderer Schutz „insbesondere die Verwendung personenbezogener Daten von Kindern für Werbezwecke oder für die Erstellung von Persönlichkeits- oder Nutzerprofilen und die Erhebung von personenbezogenen Daten von Kindern bei der Nutzung von Diensten, die Kindern direkt angeboten werden, betreffen". Da das Merkmal „direkt angeboten" nicht weiter eingeschränkt wird, ist davon auszugehen, dass hiervon nicht nur Angebote erfasst sind, die sich ausschließlich an Kinder richten, sondern auch Angebote, die sich sowohl an Kinder als auch an Erwachsene richten.[2]

Unter der DSGVO besteht eine grundsätzliche Verpflichtung zur Information betroffener Personen zum Zeitpunkt der Datenerhebung durch den Verantwortlichen. Gemäß Art. 13 Abs. 1 DSGVO umfasst dies unter anderem den Kontakt der verantwortlichen Stelle, die Zwecke der Verarbeitung sowie Informationen über die Rechte Betroffener. Die Vorschriften zur Ausgestaltung dieser Datenschutzinformationen ergeben sich aus Art. 12 Abs. 1 DSGVO. Demnach hat der Verantwortliche alle Informationen „in präziser, transparenter, verständlicher und leicht zugänglicher Form in einer klaren und einfachen Sprache zu übermitteln; dies gilt insbesondere für Informationen, die sich speziell an Kinder richten." Die Artikel-29-Datenschutzgruppe präzisiert die Anforderung der „klaren und deutlichen Sprache", indem Informationen so einfach wie möglich und unter Vermeidung komplexer Satz- und Sprachstrukturen zu formulieren sind. Darüber hinaus sollte eine übermäßig juristische, technische oder fachspezifische Terminologie vermieden werden. Zur Voraussetzung der Verständlichkeit gehört, dass der Verantwortliche den jeweiligen Kenntnisstand des Zielpublikums berücksichtigt und die Informationen so gestaltet, dass sie von einem durchschnittlichen Mitglied dieses Publikums verstanden werden. Wenn ein App-Anbieter also Kinder als Zielgruppe hat oder sich bewusst ist oder sein sollte, dass das Angebot insbesondere von Kindern genutzt wird, sollte er sicherstellen, dass Vokabular, Tonfall und Stil der verwendeten Sprache für Kinder geeignet sind.[3] Ein Beispiel für eine auf Kinder ausgerichtete Sprache nennt die Artikel-29-Datenschutzgruppe der „UN Convention on the Rights of the Child in Child Friendly Language".[4] Die Datenschutzinformation kann durch andere Maß-

[1] Ausführlich hierzu Roßnagel (2020) [8].

[2] Vgl. Heckmann/Paschke, Art. 8, in: Ehmann/Selmayr [5], Rn. 21/22; Buchner/Kühling, Art. 8, in: Kühling/Buchner [6], Rn. 16.

[3] Article 29 Working Party (2016) [1], S. 8–10.

[4] https://sites.unicef.org/rightsite/files/uncrcchilldfriendlylanguage.pdf. Stand: 06.07.2020.

nahmen wie Comics, Animationen und Piktogramme ergänzt werden.[5] Die sprachlichen Voraussetzungen implizieren auch, dass die Datenschutzinformation in der Sprache des Landes zur Verfügung gestellt werden, in dem die betroffenen Personen sind, an die sich die App richtet.[6]

3 Datengrundlage

Für die Erfassung der relevanten Learning Apps beschränken wir uns in dieser Arbeit auf den Google Play Store Germany als Datenquelle. Die Wahl von Google Play Germany lässt sich durch dessen hohen Marktanteil unter den globalen App Stores begründen.

Vorgehensweise bei der Datensammlung Anhand von passenden Sucheinstellungen und Suchbegriffen werden Apps aus dem Google Play Store selektiert und tabellarisch aufgeführt. Hierdurch ergibt sich zunächst eine Zwischenliste, welche anschließend aufgrund von Inkonsistenzen in einem Cleaning-Prozess aussortiert werden muss. Dies führt zu mehreren Iterationen im Suchprozess der Apps, wobei die Suchbegriffe konstant bleiben und lediglich die Sucheinstellungen variiert werden.

Sucheinstellungen Da wir mit Google Play Store Germany im deutschsprachigen Bereich nach unseren Apps suchen möchten, wählen wir als Sucheinstellung auf der Google Play Store-Webseite im Browser die Header-Sprache Deutsch. Weiterhin wollen wir in der Sektion Apps auf Google Play nach unseren Learning Apps suchen, sodass sich folgende URL als unsere Grundlage zum Datensammeln ergibt: https://play.google.com/store/search?q&c=apps&hl=de. Da wir außerdem nutzerbezogene Ergebnisse, welche durch Anmeldungen mithilfe eines Google-Kontos oder IP-bezogene Präferenzen auftreten können, vermeiden möchten, nutzen wir den Play Store ohne Anmeldung als Google Nutzer. Anders als beim Google Play Store in Smartphones werden im Browser keine Apps als Anzeige von Google geworben, sodass diese nicht in die Datenerfassung einbezogen werden. Alle Apps in unserem Datensatz wurden durch die Einstellung „Alle Preise“ in der Sektion „Apps“ ausgewählt, sodass sowohl kostenfreie als auch kostenpflichtige Anwendungen berücksichtigt worden sind.

[5] Article 29 Working Party (2016) [1], S. 12.

[6] Dix, Art. 12, in: Simitis/Hornung/Spiecker gen. Döhmann (2018) [10], Rn. 15.

Verwendete Suchbegriffe & Suchergebnisse Für unsere Suche werden 12 Suchbegriffe verwendet, welche semantische Verbindungen mit Learning Apps haben. Dazu zählen vor allem Begriffe mit Bezug zu Schulfächern und allgemeines Lernen. Da hierbei verschiedene Bereiche in Betracht gezogen worden sind, gelten die Ergebnisse der einzelnen Suchbegriffe als repräsentativ für die Gesamtheit von Learning Apps. Die verwendeten Suchbegriffe sind: *Learning, Education, School, M Learning, Mobile Learning, Language Learning, Sprachen lernen, Mathe, Deutsch, Hausaufgaben, Übungsaufgaben und Lernen*. Abgesehen vom Suchbegriff „Übungsaufgaben" wurden bei allen Suchbegriffen exakt 250 Apps gefunden, weil Google Play Store Germany je Suchbegriff nicht mehr Ergebnisse im Browser anzeigt. Fasst man alle Apps aus diesen 12 Zwischenergebnissen zusammen, so ergibt sich eine Gesamtheit von 2997 Apps.

Cleaning Aus den gesammelten 2997 Apps ergeben sich die folgenden drei Herausforderungen, welche bewältigt werden müssen, um einen konsistenten Datensatz für die Analyse zu sichern: 1) Duplikate (Eine App kann in mehreren Suchbegriff-Ergebnislisten vorkommen); 2) Große Ergebnismenge – (Weil die Analyse einer solche Menge an Apps den Rahmen unserer Studie sprengen würde, sind lediglich die Top 30-Apps pro Suchbegriff in den Datensatz eingeflossen); 3) Falsch zugeordnete Learning Apps (Apps, die keine Learning features vorweisen und lediglich das Etikett des Suchbegriffs tragen, sind irrelevant für die Analyse). Nach den Cleaning-Schritten erhalten wir eine Liste von 167 Apps.

Zusätzliche Betrachtung kostenpflichtiger Apps Auffällig bei dieser Datenmenge aus 167 Apps ist jedoch, dass bei keinem der zugrundeliegenden Suchbegriffe kostenpflichtige Apps in den Top 30 Ergebnissen aufgelistet sind. Bei den zuvor betrachteten Suchergebnissen traten kostenpflichtige Apps erst jenseits der Top 30 Grenze auf, da sie von Nutzern anscheinend seltener benutzt werden. Für unsere Analyse sollen aber auch kostenpflichtige Anwendungen explizit betrachtet werden. Hierfür wird lediglich die Einstellung „Alle Preise" zu „kostenpflichtig" in der Suche bei Google Play Store Germany umgewandelt. Die restlichen Einstellungen verbleiben unverändert. Der Vorgang der Datenerhebung wird hier genauso mit denselben Suchbegriffen vorgenommen. Jedoch definieren wir die Top 10 je Suchbegriff als angemessenen Umfang für die Datenerfassung kostenpflichtiger Apps. Somit erhalten wir mit einem identischen Vorgehen wie bei der ersten Suche eine Gesamtheit von 120 Apps. Nach einer Bereinigung der so erhobenen Datenmenge (analog zum Cleaning-Prozess für die Datenmenge der kostenlosen Apps) erhalten wir 32 kostenpflichtige Learning Apps. Unser Datensatz ist nach erfolgreicher Aussortierung der irrelevanten Datenelemente nun vollständig und enthält 199 Learning

Apps für Kinder und Jugendliche – 167 kostenlose Apps und 32 kostenpflichtige Apps.

Erfasste Metadaten Für eine ausgewogene Analyse reichen die zunächst gesammelten Angaben zu den Apps wie etwa App-Namen und App-URLs nicht aus, sodass weitere Informationen, sog. Metadaten, gesammelt werden müssen. 33 potenziell nützliche Metadaten (33 werden in dieser Studie verwendet), welche aus dem Google Play Store Germany frei verfügbar sind, werden berücksichtigt.

Datenerfassungsinstrumente Für die Datengrundlage werden pro App die APK-Datei und die entsprechenden Metadaten gesammelt. Für die Datenerfassung im Rahmen unserer Studie sind entsprechend zwei Instrumente entwickelt und eingesetzt worden: ein Metadaten-Crawler und ein APK-Crawler. Mittels beider Instrumente erhalten wir folglich eine Liste von 167 kostenlosen und 32 kostenpflichtigen Apps zusammen mit ihren gecrawlten Metadaten als Datengrundlage unserer Analyse.

4 Grobgranulare Analyse

Nachfolgend stellen wir unsere Ergebnisse der Ursprungslands-, Popularitäts- und Datenschutzerklärungsanalyse vor.

4.1 Ursprungslandsanalyse

Wir definieren das Ursprungsland als das Land, welches den Standort des Entwicklers laut Google Play Store bezeichnet. Somit wird das Ursprungsland einer App aus dem gesammelten Metadatenattribut „Entwickler-Adresse“ gewonnen. Es wird ein zusätzliches Attribut für die Apps aufgegriffen, um diese Ursprungsländer einer Ursprungsregion zuzuteilen. Wir wählen hierbei Regionen bzw. Länder als Ursprungsregionen aus, welche wir für diese Analyse hinsichtlich Datenschutz und Cybersicherheit als relevant erachten. Somit ergeben sich folgende 6 Ursprungsregionen: Europäische Union (EU), Vereinigte Staaten von Amerika, Kanada, Russland, China und Sonstige. Bei der EU sind auch Länder wie die UK (United Kingdom/Vereinigtes Königreich) enthalten, welche zwar nicht mehr in der EU-Mitglied sind, jedoch immer noch ein Datenschutzabkommen mit der EU einhalten (Stand: 2020).

Die Werte aus Abb. 1 zeigen deutlich, dass die meisten Apps sowohl bei den kostenlosen als auch kostenpflichtigen als Ursprungsland Deutschland oder die Vereinigten Staaten aufweisen. Bei den kostenlosen Apps fällt allerdings noch auf, dass 8 Apps aus Indien und 10 Apps aus Spanien stammen. Leider geben 33 der kostenlosen Apps keine Information zur Adresse des Entwicklers in Google Play Store an, sodass ihre Ursprungsländer für diese Analyse nicht verfügbar sind. Nachdem die einzelnen Länder den entsprechenden Ursprungsregionen zugeordnet worden sind, erhalten wir eine Verteilung der Ursprungsregionen wie folgt: Etwa 46 % der kostenlosen und 59 % der kostenpflichtigen stammen aus der Ursprungsregion EU. Interessant für die weitere Analyse ist hierbei, inwiefern die Apps die strengen Datenschutzrichtlinien der EU einhalten. Abgesehen von den sonstigen Ursprungsländern sind die Vereinigten Staaten von Amerika am zweithäufigsten vertreten. Auffällig bei den kostenlosen Apps ist vor allem, dass 25 bzw. 33 der Apps zu den sonstigen Ursprungsregionen bzw. keinem Ursprungsland zugeordnet worden sind, wobei die kostenpflichtigen lediglich ein bis zwei Apps dieser Eigenschaft enthalten.

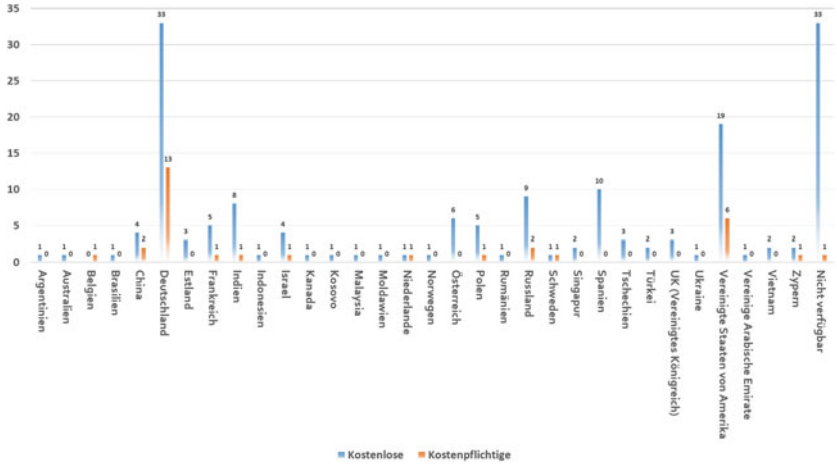

Abb. 1 Anzahl der Apps pro Ursprungsland

4.2 Popularitätsanalyse

Unter Popularität ist der Bekanntheitsgrad einer Learning App gemeint. Dies ist nicht zu verwechseln mit der Beliebtheit einer App, da eine App auch durch negative Erfahrungen bekannt und folglich nicht als beliebt bezeichnet werden kann. In unserer Popularitätsanalyse werden Metadaten betrachtet, welche vor allem die Bekanntheit einer App implizieren. Hierzu zählen wir die Metadaten: Installationen (minimale Anzahl an Installationen auf Endgeräten durch Nutzer), Bewertungen (die exakte Anzahl der Nutzerbewertungen, wobei eine Bewertung durch das Verteilen von Sternen in Google Play Store abgegeben wird), Kommentare (die Anzahl der vom Nutzer verfassten Kommentare unter einer App in Google Play Store), durchschnittliche Bewertung und Android-Versionen.

In Tab. 1 werden einige wichtige Durchschnittswerte zu entsprechenden Metadaten dargestellt. Hierbei zeigen die Werte deutlich, dass die kostenpflichtigen Apps im Durchschnitt weitaus weniger Kommentare, Installationen und Bewertungen aufweisen als die kostenlosen Apps. Dies ist nicht überraschend, da die meisten App-Store-Nutzer kostenfreie Apps vorziehen und daher diese Diskrepanz in den Metadaten entsteht. Überraschend ist jedoch, dass die Nutzerzufriedenheit bei kostenpflichtigen Apps bei einer durchschnittlichen Bewertung von 3,6 nicht besonders gut ist, obwohl die Entwickler Geld für den Erwerb dieser Apps fordern. Die Werte aus der Tabelle implizieren, dass die kostenlosen Apps populärer sind als die kostenpflichtigen, da sie weitaus mehr Kommentare, Installationen und Bewertungen im Durchschnitt ausweisen. Je höher diese Werte sind, desto höher ist auch der Bekanntheitsgrad einer App. Aus Abb. 2 geht hervor, dass vor allem Learning Apps, welche die Vereinigten Staaten als Ursprungsregion bzw. -land haben, eine relativ hohe Anzahl an Installationen aufweisen im Vergleich zu den übrigen Ursprungsregionen. Zwar scheint auch die Ursprungsregion Kanada eine beträchtliche Anzahl an Installationen aufzuweisen, jedoch zeigte die Ursprungslandsanalyse, dass Kanada

Tab. 1 Durchschnittliche Werte für Metadaten aus unserem Datensatz

–	Kostenlose Apps	Kostenpflichtige Apps
Ø Anzahl an Installationen	3.155.822	14.850
Ø Anzahl an Bewertungen	90.070	835
Ø Anzahl an Kommentaren	35.970	343
Ø Anzahl an Berechtigungen	7,37	4,5
Ø Bewertung	4,01	3,6

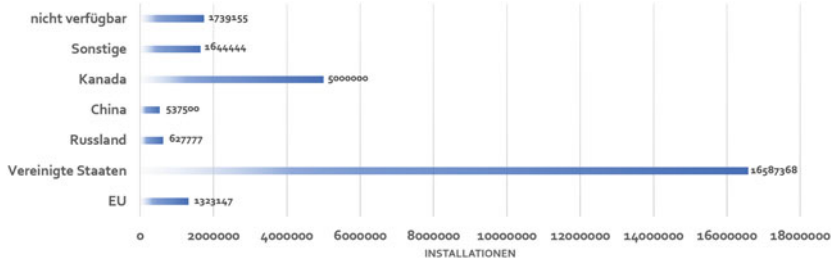

Abb. 2 Relation zwischen Ursprungsregion und zugehöriger durchschnittlicher Anzahl an Installationen bei den kostenlosen Apps

als Region lediglich eine App als Datenelement beinhaltet und somit kein ausgewogener Vergleich zu den Regionen Europa und Vereinigten Staaten durchgeführt werden kann. Die Europäische Union weist in dieser Grafik keine besonderen Merkmale auf, jedoch beträgt die Anzahl der durchschnittlichen Installationen für Apps aus dieser Region weniger als die Hälfte des Mittelwerts aller kostenlosen Apps, wie ein Vergleich mit Tab. 1 zeigt. Überraschend zu diesem Ergebnis ist der Vergleich mit den kostenpflichtigen Apps in Abb. 3. Hier liegt der Peak an Installationen bei der Region China, wobei auch hier erwähnt werden muss, dass diese Region nur 2 Apps als Elemente aufweist. In dieser Grafik belegen die Apps mit der Ursprungsregion Vereinigten Staaten den vorletzten Platz vor den sonstigen Regionen. Im Vergleich zu den kostenlosen Apps liegt die durchschnittliche Anzahl an Installationen bei den europäischen, kostenpflichtigen Apps viel näher bei dem Mittelwert aus Tab. 1.

4.3 Bewertungen

Abb. 4 zeigt, dass der Großteil der Apps eine durchschnittliche Bewertung von 4 Sternen enthält. Unter 167 kostenlosen Apps existieren jedoch auch 7 Apps, welche keine Bewertung zum Zeitpunkt der Datenerfassung hatten (z. B. die *OLLIS:Mathe 3* App). Besonders auffällig hierbei ist, dass keine der Apps eine Bewertung oder Nutzerkommentare aufweist, obwohl die Anwendungen teilweise seit 2017 in Google Play Store veröffentlicht worden sind. Unter den kostenpflichtigen existieren 5 Apps, welche ebenfalls keine Bewertungen von Nutzern im Google Play Store erhalten haben.

In Abb. 5 sind die durchschnittlichen Bewertungen pro Ursprungsregion für kostenlose und kostenpflichtige Apps abgebildet. Wir stellen fest, dass Europa als

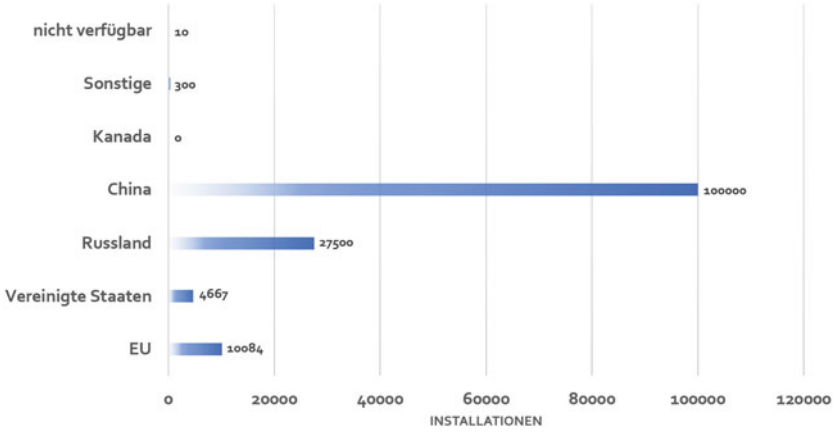

Abb. 3 Relation zwischen Ursprungsregion und zugehöriger durchschnittlicher Anzahl der Installationen bei den kostenpflichtigen Apps

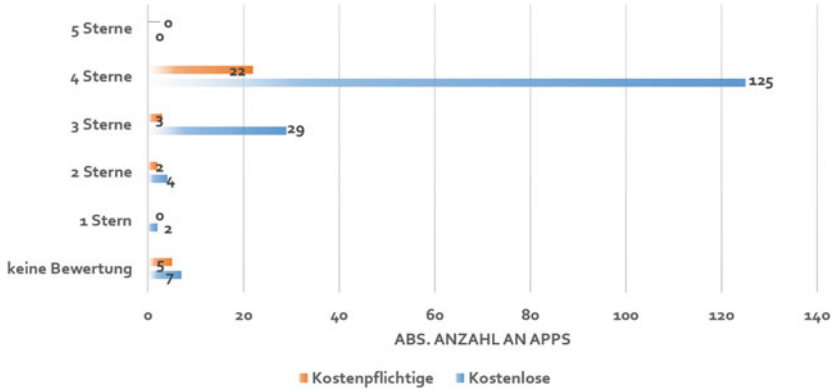

Abb. 4 Anzahl der Sternbewertung und zugehörige Anzahl an Apps; die Sternbewertung ist auf die nächste ganze Zahl abgerundet worden

Ursprungsregion für beide Datensätze eine wesentlich schlechtere Durchschnittsbewertung aufweist als die übrigen Regionen.

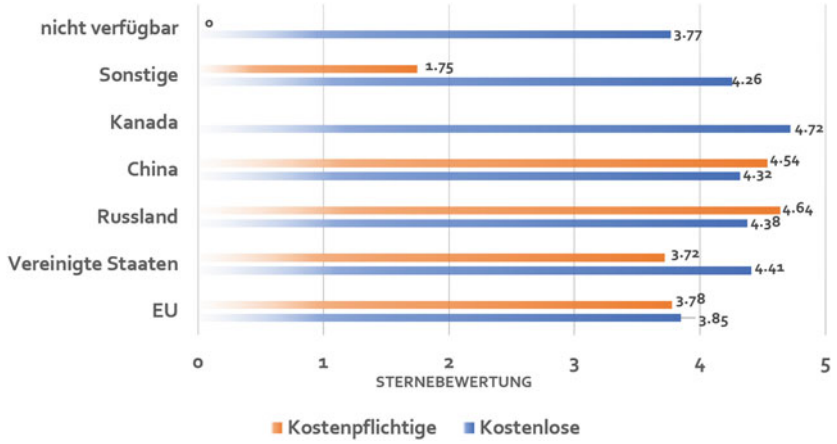

Abb. 5 Relation zwischen Ursprungsregion und zugehöriger durchschnittlicher Bewertung

4.4 Datenschutzerklärungsanalyse

Entwickler sind gesetzlich dazu verpflichtet eine Datenschutzerklärung zu erstellen und für ihre Nutzer im Google Play Store öffentlich verfügbar zu machen, falls sie personenbezogene Daten vom Nutzer verarbeiten (siehe Abschn. 2). Wenn also Learning Apps keine Datenschutzerklärung bereitstellen, so dürfen sie beispielsweise persönliche Informationen von ihren Nutzern weder anfordern noch speichern. Dies muss je nach Anwendungsfall auf Implementierungsebene überprüft werden. Wenn bestimmte sensitive Berechtigungen angefragt werden, welche auf persönliche Informationen vom Smartphone des Nutzers zugreifen können, so muss eine Datenschutzerklärung vorliegen.[7] Im Google Play Store ist die Datenschutzerklärung, sofern diese vom Entwickler veröffentlicht worden ist, bereits vor der Installation über einen Link verfügbar, welcher im Crawling-Prozess erfasst worden ist. Es gibt jedoch auch Apps wie etwa *Mathe-Formeln – Offline26,* welche einen Link zu einer Datenschutzerklärung in Google Play Store angeben[8], wobei dieser Link jedoch zu keiner gültigen Webseite oder Dokument führt. Solche Apps werden ebenfalls zu Apps ohne Datenschutzerklärung hinzugezählt, da dem Nutzer zur Zeit der Überprüfung keine Erklärung angeboten werden konnte. Von den 167 kostenlosen Apps besitzen 13 keine Datenschutzerklärung, welche auf Google Play

[7] Article 29 Working Party (2016) [1].

[8] http://thanhhangfood.com/elearning_privacy_policy.txt

Tab. 2 Vergleich zwischen Learning Apps mit und ohne Datenschutzerklärung

–	Mit Datenschutzerklärung	Ohne Datenschutzerklärung
Ø Anzahl an Installationen	3.407.930	169.315
Ø Anzahl an Bewertungen	97.500	2055
Ø Anzahl an Kommentaren	38 940	777
Ø Anzahl an Berechtigungen	7,44	6,54
Durchschnittliche Bewertung	4,00	4,06

Store öffentlich verfügbar ist. So hat beispielsweise die App *Mathe Arena – Mathematik für Abitur & Matura27* keine Datenschutzerklärung. Auffällig bei einigen dieser Anwendungen ist jedoch, dass sie im Durchschnitt 6,5 Berechtigungen und einige Elemente sogar bis zu 20 Berechtigungen fordern. Die Verteilung von Apps mit und ohne Datenschutzerklärungen ist bei den kostenpflichtigen Apps weiter kompakter als bei den kostenlosen. So geben nur 3 von 32 kostenpflichtigen Apps keine Datenschutzerklärung an. Nennenswert hierbei ist vor allem, dass nur eine davon Berechtigungen anfordert (Tab. 2).

Der Vergleich zwischen Apps mit und ohne Datenschutzerklärung in Tab. 5 zeigt, dass Anwendungen ohne eine Datenschutzerklärung deutlich weniger Installationen, Bewertungen und Nutzerkommentare aufweisen als die übrigen Anwendungen. Die Anzahl der angeforderten Berechtigungen ist im Durchschnitt bei Apps ohne Datenschutzerklärung nur leicht höher. Die Bewertungen der Apps sind bei beiden Vergleichsgruppen sehr nah beisammen.

Ursprungsregionen der Apps ohne Datenschutzerklärung 38 % (5 von 13) dieser Apps geben keine Entwickleradresse an, sodass wir keine Ursprungsregion für diese Elemente feststellen können. Besonders auffällig bei den Daten sind die 4 Apps ohne Datenschutzerklärung, welche der EU als Ursprungsregion angehören. Von diesen Apps fordern 2 Apps 20 Berechtigungen bzw. 7 gefährliche Berechtigungen[9] an, obwohl sie keine gültige Webseite für die Datenschutzerklärung anbieten. Gefährliche Berechtigungen indizieren meist den Gebrauch von sensitiven Informationen des Nutzers. Im Vergleich hierzu gehören 2 der 3 kostenpflichtigen Apps ohne Datenschutzerklärung der Ursprungsregion EU und das übrige Datenelement gehört der Region Russland an. Eine der europäischen Apps fordert hierbei 2 gefährliche Berechtigungen an.

[9] Android Developers: Protection levels: https://developer.android.com/guide/topics/permissions/. Stand: 06.07.2020.

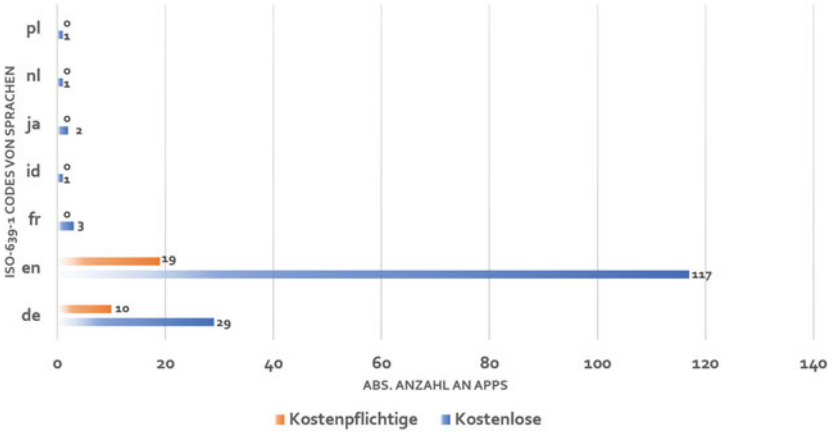

Abb. 6 Sprachen der Datenschutzerklärungen

Sprachen der Datenschutzerklärungen Datenschutzerklärungen auf Deutsch und Englisch sind mit großem Vorsprung am stärksten vertreten, siehe Abb. 5. In den Datenschutzerklärungen treten neben Deutsch (29 kostenlose und 10 kostenpflichtige), Englisch (117 kostenlose und 19 kostenpflichtige) die Sprachen Französisch (3 kostenlose und 0 kostenpflichtige), Indonesisch (1 kostenlose und 0 kostenpflichtige), Japanisch (2 kostenlose und 0 kostenpflichtige), Niederländisch (1 kostenlose und 0 kostenpflichtige) und Polnisch (1 kostenlose und 0 kostenpflichtige) auf (Abb. 6).

Die Ergebnisse in Abb. 7 zeigen, dass Learning Apps mit Datenschutzerklärungen in den Sprachen Deutsch und Französisch die höchste Anzahl an Berechtigungen im Schnitt aufweisen, wobei Französisch zusätzlich eine sehr schlechte durchschnittliche Bewertung mit 1,35 Sternen besitzt. Die Apps mit deutschen Datenschutzerklärungen besitzen lediglich eine Durchschnittsbewertung von 3,23 Sternen.

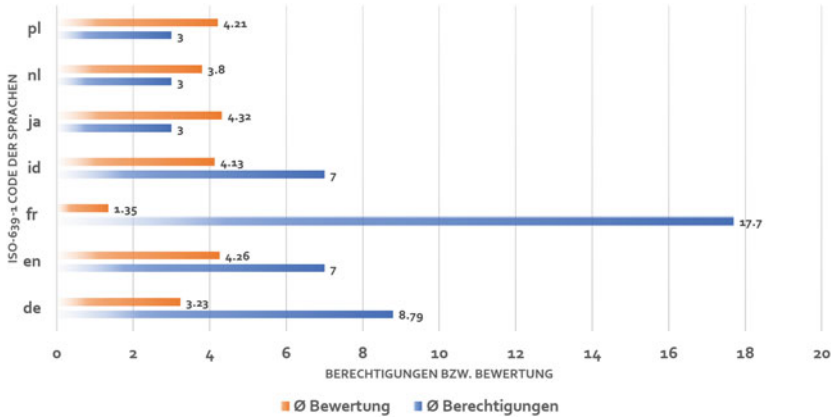

Abb. 7 Sprache der Datenschutzerklärung und zugehörige Statistik bezüglich durchschnittlicher Bewertung und angeforderte Berechtigungen bei den kostenlosen Apps

5 Feingranulare Analyse

5.1 Berechtigungsanalyse

Das Android-Betriebssystem bietet ein Berechtigungsmodell zur Steuerung des Zugangs zu persönlichen Daten (z. B. Standort) und sensiblen Systemressourcen (z. B. Mikrofon)[10]. Für die Berechtigungsanalyse stützen wir uns auf den **„Protection levels“** von Google, welche jede einzelne Berechtigung in **normale, gefährliche und Signatur Berechtigungen** unterteilt.[11] Hierbei wird zusätzlich zwischen den sogenannten Laufzeit-Berechtigungen und Berechtigungen, welche zur Installationszeit angefordert werden, unterschieden. Bei letzteren fordert die App vor der Installation die Zustimmung des Nutzers für alle in Google Play Store aufgelisteten Berechtigungen an. Als **normale Berechtigungen** definiert Google Berechtigungen, welche benötigt werden, um auf Daten außerhalb der Sandbox einer App zuzugreifen. Hierbei besteht, verglichen mit gefährliche Berechtigungen, jedoch nur ein geringfügiges Risiko für die Privatsphäre des Nutzers oder Operationen anderer Apps. Diese Berechtigungen werden der App zur Installationszeit gewährt, sofern diese im Android-Manifest der App deklariert worden

[10] https://developer.android.com/guide/topics/permissions/overview. Stand: 06.07.2020.

[11] Android Developers: Protection levels: https://developer.android.com/guide/topics/permissions/. Stand: 06.07.2020.

sind, und können vom Nutzer nicht widerrufen werden. Ein Beispiel für eine normale Berechtigung wäre *android.permissions.BLUETOOTH*, welche einer App die Möglichkeit gewährt, sich mittels Bluetooth mit anderen Bluetooth-Geräten zu verbinden. ***Signatur-Berechtigungen*** werden einer App ebenfalls zur Installationszeit gestattet, jedoch nur, falls die angeforderte Berechtigung mit demselben Zertifikat signiert ist wie die App, welche die Berechtigung definiert. So kann beispielsweise eine App mittels der Signatur-Berechtigung *android.permission.BATTERY_STATS* Informationen zum Status der internen Batterie sammeln. ***Gefährliche Berechtigungen*** hingegen ermöglichen es einer App, auf private Daten und Informationen des Nutzers zuzugreifen oder die Operation von anderen Apps zu beeinträchtigen. Außerdem können Apps Kontrolle über das Gerät erhalten und sich somit negativ auf den Nutzer auswirken. Der Nutzer muss jedoch explizit den aufgelisteten, gefährlichen Berechtigungen zustimmen, damit die App Zugriff auf die betreffenden Funktionen erhält. Ohne dessen Zustimmung kann die App die Funktionen, welche die gefährliche Berechtigung benötigt, nicht ausüben. So wird beispielsweise die gefährliche Berechtigung *android.permission.CAMERA* benötigt, falls eine App Zugriff auf die Kamera des Gerätes fordert.

Angeforderte Berechtigungen per Apps Die Anzahl der angeforderten Berechtigungen für alle Apps variiert sehr stark – zwischen 0 und 28 Berechtigungen pro App. Abb. 8 stellt eine Übersicht über die Verteilung der Berechtigungen über alle untersuchten Learning Apps unseres Datensatzes dar. Der Graph weist 3 Peaks bei 3, 8 und 15–20 Berechtigungen bei den kostenlosen Apps auf. Bei den kostenfreien hingegen sind die Balken relativ ausgeglichen, jedoch liegen die größeren Balken hier bei vergleichsmäßig niedrigeren Berechtigungszahlen als bei den kostenlosen. Diese Erkenntnis spiegelt sich auch in den Durchschnittswerten aus der Übersicht in Tab. 1 wieder.

Angeforderte Berechtigung per Protection Level Insgesamt fordern die 199 Apps 1412 normale, 267 gefährliche und 79 Signatur-Berechtigungen in den Android-Manifesten an. Die daraus resultierenden Durchschnittswerte sind in Abb. 9 dargestellt.

Berechtigungen & Sprache der Datenschutzerklärung Bereits mit Abb. 7 wurden die hohen Berechtigungswerte bei den Apps mit französischer Datenschutzerklärung festgestellt. Dieser Befund spiegelt sich auch in Abb. 10 wider: Da die gefährlichen Berechtigungswerte mit den Berechtigungen allgemein in Relation zu stehen scheinen, ist die Feststellung nicht überraschend, dass die Apps mit französi-

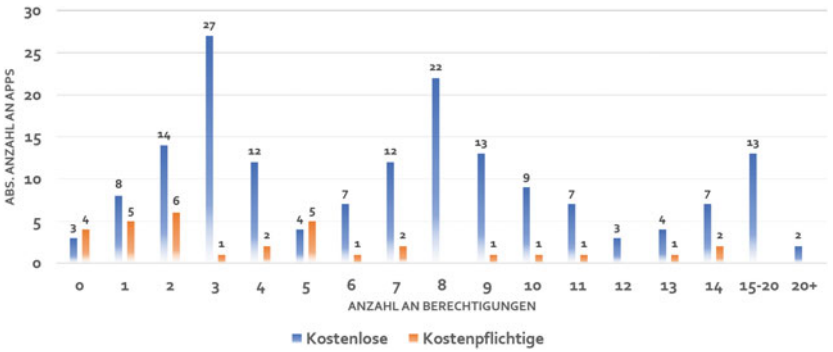

Abb. 8 Relation zwischen Apps und Berechtigungen

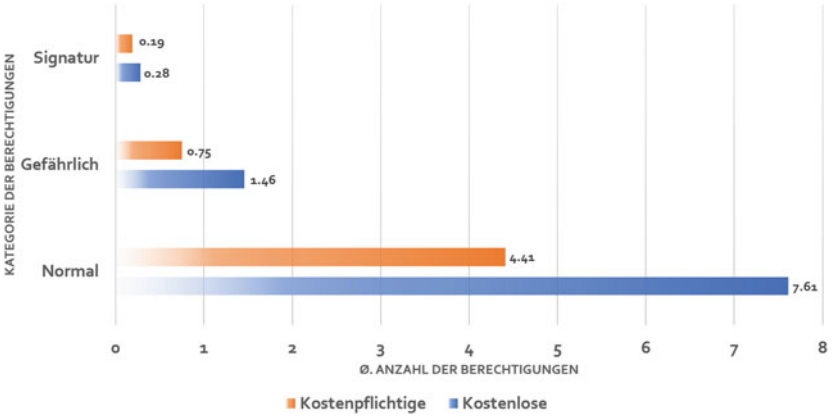

Abb. 9 Verteilung der Berechtigungen in ihre Kategorien und ihr durchschnittlicher Wert unter allen Apps

scher Datenschutzerklärung auch sehr viele gefährliche Berechtigungen anfordern. Bei den übrigen Sprachen sind keine besonderen Merkmale erkennbar.

Verteilung gefährlicher Berechtigungen Aus einer genauen Untersuchung geht hervor, dass 12 einzigartige, gefährliche Berechtigungen im Datensatz der kostenlosen Apps bzw. 5 im Datensatz der kostenpflichtigen auftreten. Wir führen hierbei eine Häufigkeitsanalyse durch, um festzustellen, für welche Funktionen Learning

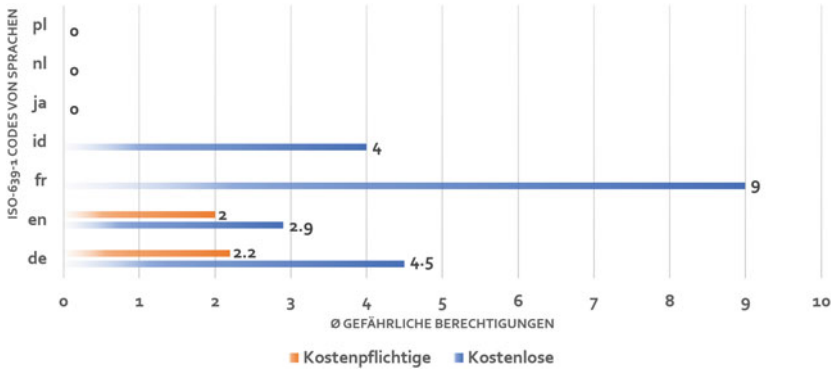

Abb. 10 Durchschnittliche Anzahl an gefährlichen Berechtigungen für Apps mit bestimmter Sprache in der Datenschutzerklärung

Apps gefährliche Berechtigungen im Allgemeinen anfordern. Die Ergebnisse dieser Analyse, in Abb. 11 zusammengefasst, zeigen deutlich, dass die meisten Apps in beiden Datensätzen überwiegend Zugriff auf Speicherdaten durch Berechtigungen wie *WRITE_EXTERNAL_STORAGE* oder *READ_EXTERNAL_STORAGE* erhalten. Berechtigungen wie etwa

RECORD_AUDIO oder *CAMERA,* welche für Gerätefunktionen wie das Mikrofon oder die Kamera benötigt werden, nehmen untere Plätze in der Häufigkeitsanalyse ein. *ACCESS_FINE_LOCATION* und *ACCESS_COURSE_LOCATION* wer-

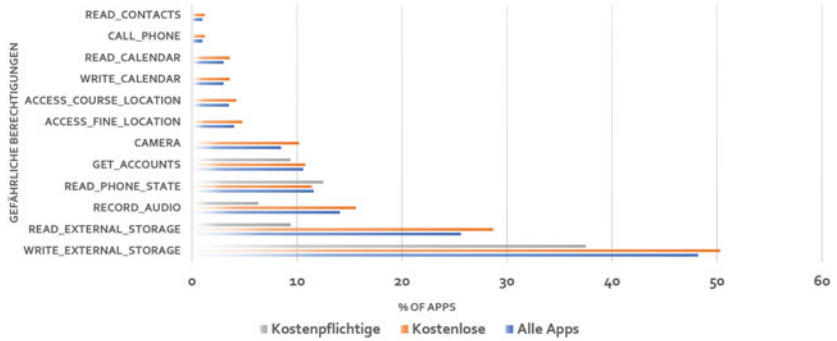

Abb. 11 Häufigkeit der angeforderten gefährlichen Berechtigungen

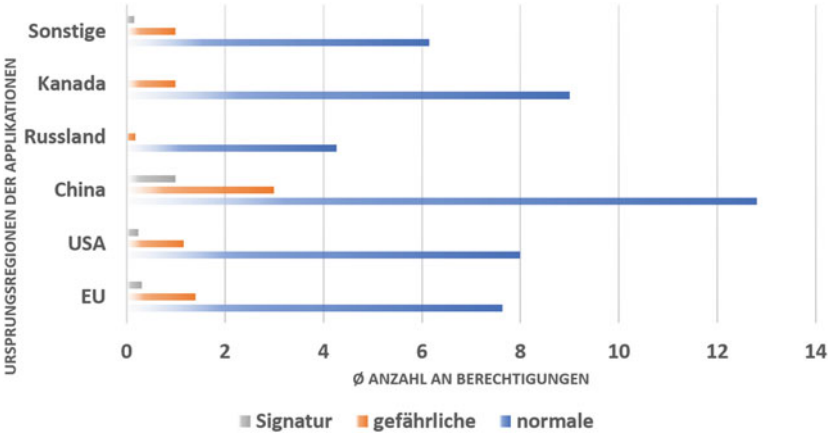

Abb. 12 Verteilung der durchschnittlichen Berechtigungen je Kategorie nach dem Google Berechtigungsmodell unter den Ursprungsregionen aller Apps

den für die Standortermittlung des Gerätes benötigt, sind jedoch nur von weniger als 5 % aller Learning Apps angefordert worden. Auffällig hierbei ist vor allem, dass Berechtigungen für die Nutzung der Standort-Funktion, Kamera- oder Kalenderinhalte von keiner der kostenpflichtigen Apps genutzt wird.

Berechtigungen & Ursprungsregionen Ein weiterer Teil unserer Analyse zielte auf den Vergleich zwischen einzelnen Ursprungsregionen aller Apps in Bezug auf die durchschnittlichen Berechtigungen in jeder Kategorie des Google Berechtigungsmodells für die 165 der 199 Apps, welche einer Ursprungsregion durch eine Angabe des Entwickler-Standorts zugeordnet werden konnten. Die Ergebnisse sind in Abb. 12 zusammengefasst. Hierbei wird deutlich, dass vor allem Apps aus den Regionen EU, Vereinigte Staaten und China mehr gefährliche und Signatur-Berechtigungen anfordern als die übrigen Ursprungsregionen. Sehr auffällig ist jedoch der Peak für alle drei Berechtigungskategorien in der Ursprungsregion China. Dieses Muster ist auch in den Top 10-Apps nach gefährlichen Berechtigungen erkennbar: zwei dieser Top 10-Apps weisen die Ursprungsregion China auf und 2 weitere liefern keinerlei Informationen zum Entwickler-Standort. Außerdem sind darunter 2 Apps enthalten, welche eine Datenschutzerklärung in der Sprache Französisch darlegen, obwohl im gesamten Datensatz der 199 Apps nur 3 Apps eine französische Datenschutzerklärung besitzen. In Abb. 13 sind nicht die Sprachen, sondern die Ursprungsregionen mit den gefährlichen Berechtigungen in Relation

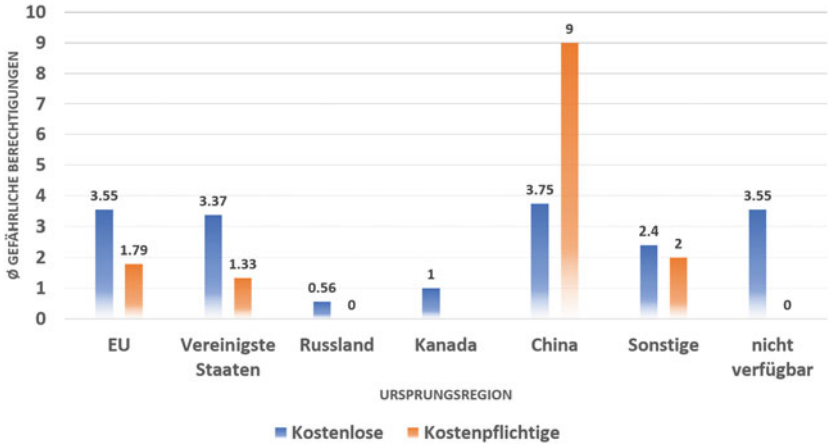

Abb. 13 Relation zwischen Ursprungsregion und zugehöriger Anzahl an gefährlichen Berechtigungen

gebracht worden. Hierbei stellen wir einen Peak bei der Ursprungsregion sowohl für die kostenlosen als auch kostenpflichtigen Apps fest. Learning Apps, deren Entwickler ihren Standort in China haben, scheinen mehr gefährliche Berechtigungen anzufordern als die übrigen Apps.

5.2 Third-party Library Analyse

Um Entwicklungs- und Wartungskosten für Apps gering zu halten, greifen Entwickler häufig auf Bibliotheken von Drittanbietern, die sog. Third-party Libraries, zurück. Mittels der Third-party Libraries ist es nicht nur möglich neue Funktionen in der App zu integrieren, sondern auch mit Werbung die Kosten aus Entwicklung und Wartung wieder einzunehmen. Eingeteilt werden können die Third-party Libraries in drei verschiedene Kategorien: Werbenetzwerke, soziale Netzwerken und Entwicklungswerkzeuge. Dies entspricht der Kategorisierung durch die Statistik-Webseite appbrain[12] Unsere Analyse baut ebenfalls auf Exodus Pri-

[12] AppBrain: Android library statistics. https://www.appbrain.com/stats/libraries. Stand: 06.07.2020.

vacy[13] auf. Dabei findet ein Abgleich zwischen der Library-Liste von Exodus mit ihren Code-Signaturen und einer Liste von eingebetteten Java-Klassennamen aus der untersuchten App statt, sodass ein positiver Vergleich die Existenz der Libraries in der Anwendung beweist. Folglich erhalten wir eine Liste aller Libraries, welche im Datensatz der 199 Apps verwendet worden sind, und wie oft die Information darüber, darin vorkommen. Die durchschnittliche Anzahl an Third-party Libraries in Learning Apps beträgt bei den kostenlosen etwa 4 und 1,6 bei den kostenpflichtigen. Die in den Apps verwendeten Libraries sollen nun genauer untersucht werden. Hierfür werden die von Exodus Privacy ausgegebenen Libraries aufgelistet und hinsichtlich ihrer Verteilung über alle Apps alysiert.

Top 25-Libraries unter allen kostenlosen und kostenpflichtigen Apps Da der Durchschnittswert für die Libraries bei etwa 4 liegt, ist es nicht verwunderlich, dass 60 % aller Apps 3 oder weniger Libraries verwenden. Circa 23 % der Learning Apps verwenden mehr als 5 Libraries. Das Auftreten der Top 25-Libraries in allen Apps wird in Abb. 13 prozentual untereinander verglichen. Wir gehen hierbei separat auf die kostenlosen und kostenpflichtigen ein. Besonders auffällig zeigen sich die Libraries von Google, welche mit 6 namenhaften Libraries wie Google Firebase Analytics oder Google Analytics in den Top 10 Libraries aufgelistet sind. Dies ist jedoch nicht verwunderlich, da Libraries von Google wie beispielsweise Firebase Analytics in etwa 61,4 % aller Android Apps enthalten sind. Die überwiegende Mehrheit der weites verbreitetsten Libraries werden von Internet Giganten angeboten. Erst die Libraries ab Platz Nummer 12 in Abb. 14 weisen einen anderen Namen als Google oder Facebook auf. Diese als nächsthäufigste verwendeten Libraries *Unity3dAds, Moat* und *AppLovin* werden genauso, wie Google Ads und Facebook Ads hauptsächlich für Werbeanzeigen benutzt. Aus unserer Analyse ist zu entnehmen, dass etwa 14 % aller Apps in unserem Datensatz Libraries für soziale Netzwerke benutzen. 38 % bzw. 48 % der Apps beinhalten Libraries für Werbeanzeigen bzw. Libraries für Entwicklungstools. Der hohe Anteil von Libraries der Kategorie Werbeanzeigen (etwa 38 %) deutet darauf hin, dass ein Großteil von Learning Apps dazu verwendet werden, um gezielt Werbung auf dem Gerät des Nutzers anzuzeigen. Die am häufigsten vertretene Kategorie sind jedoch die Entwicklungstools, welche in circa 48 % aller Apps in unserem Datensatz auftreten.

[13] Exodus Privacy: Trackers. https://reports.exodus-privacy.eu.org/en/info/trackers/. Stand: 06.07.2020.

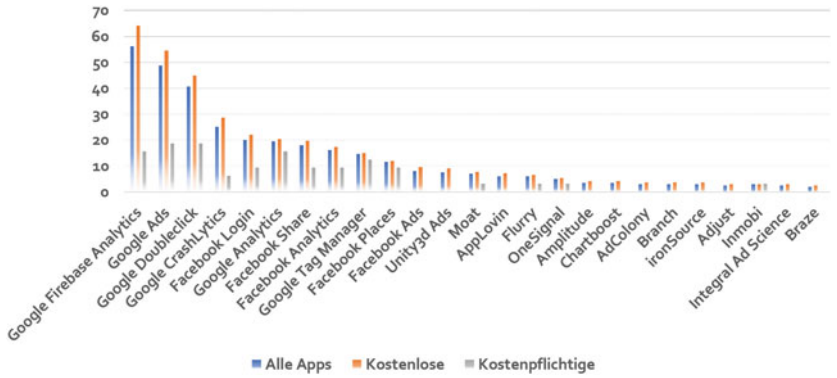

Abb. 14 Aufkommen der Top 25-Libraries unter allen kostenlosen und kostenpflichtigen Apps

Eingebettete Top 25-Libraries unter den Top 10-Apps nach Anzahl der Installationen Betrachten wir nun die Ergebnisse für die Top 10-Apps nach Installationen hinsichtlich ihrer eingebetteten Libraries, zusammengefasst in Tab. 3, so stellen wir fest, dass Libraries für Werbeanzeigen und Libraries für Entwicklungstools etwa gleichermaßen vertreten sind.

Der Schwerpunkt der am häufigsten benutzten Third-party Libraries liegt hauptsächlich bei den Top 4-Libraries von Google *(Google Firebase Analytics, Google Ads, Google Doubleclick* und *Google CrashLytics).*

Top 10-Apps nach meisten Third-party Libraries Zum Vergleich werden die Top 10-Apps nach Anzahl der eingebetteten Libraries in Tab. 3 dargestellt. Hierbei sind vor allem die Top 9-Libraries, welche den Unternehmen Google und Facebook zuzuschreiben sind, am häufigsten verwendet worden.

Top 10-Apps nach Anzahl der eingebetteten Libraries und korrespondierende Informationen bezüglich Datenschutzerklärung und Ursprungsregion Vergleichen wir die Apps mit und ohne Datenschutzerklärung hinsichtlich der durchschnittlichen Anzahl an eingebetteten Libraries, siehe Tab. 5, so stellen wir fest, dass Apps ohne Datenschutzerklärung weniger Libraries im Code enthalten als diejenigen mit einer Datenschutzerklärung. Diese in Abb. 15 illustrierten Ergebnisse lassen darauf schließen, dass Entwickler von Apps ohne Datenschutzerklärung dazu neigen, weniger Libraries zu verwenden. Wie bereits erläutert, können Libraries dazu eingesetzt werden, um nutzerbezogene Daten von der App zu erheben und abzuspei-

Tab. 3 Eingebettete Top 25-Libraries unter den Top 10-Apps nach Installationen. A1 – Duolingo; A2 – Photo-math; A3 – Google Classroom; A4 – Neuro Nation; A5 – Mathe Einmal-ein; A6 – Quizlet; A7 – Coursera: online courses; A8 – ABC Spiele; A9 – Class-Dojo; A10 – Mathe-Spiele

–	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Google Firebase Analytics	✓	✓		✓		✓	✓	✓	✓	✓
Google Ads	✓			✓	✓	✓		✓	✓	✓
Google Double-click				✓	✓			✓		✓
Google CrashLytics	✓	✓		✓		✓	✓		✓	
Facebook Login	✓	✓		✓		✓	✓			
Google Analytics				✓		✓				
Facebook Share	✓	✓		✓		✓	✓			
Facebook Analytics	✓			✓			✓			
Google Tag Manager				✓		✓				
Facebook Places	✓						✓			
Facebook Ads	✓									✓
Unity3d Ads						✓				
Moat						✓				
AppLovin						✓				
Flurry						✓				
OneSignal						✓				
Amplitude						✓				
Chartboost						✓				
AdColony						✓				
Branch						✓				
ironSource								✓		
Adjust	✓			✓						
Inmobi										
Integral Ad Science										
Braze						✓				

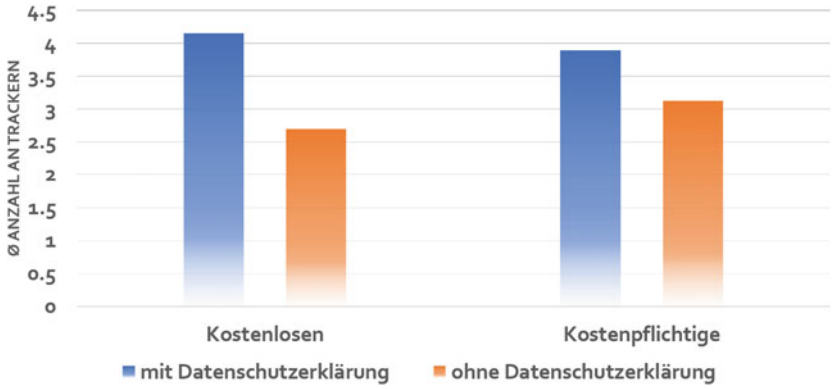

Abb. 15 Relation zwischen App mit und ohne Datenschutzerklärung bezüglich durchschnittlicher Anzahl an Trackern

chern, was gegen die Datenschutzrichtlinien verstoßen würde, falls die Entwickler dies nicht in einer Datenschutzerklärung dargelegt haben.

Auch bei den Top 10-Apps nach Libraries untersuchen wir, ob sie eine Datenschutzerklärung haben und welcher Ursprungsregion sie zugewiesen werden können. Die Ergebnisse hierzu sind in Tab. 4 abgebildet. Auch hier besitzen alle Apps eine Datenschutzerklärung, und diese sind bis auf eine französische Datenschutzerklärung in Deutsch bzw. Englisch verfasst. Bei den Ursprungsregionen sind jedoch einige Besonderheiten anzumerken: 3 Apps gehören der Ursprungsregionskategorie Sonstige an, wobei zwei davon dem Ursprungsland Indien und die übrige Apps Moldawien zuzuschreiben ist. Eine App gibt sogar keinerlei Informationen zum Entwickler-Standort an, sodass eine Ursprungsregion nicht abgeleitet werden kann. Die übrigen Apps stammen aus den Ursprungsregionen EU oder Vereinigte Staaten. Die Ursprungsregion der App scheint kaum Einfluss auf die Anzahl an eingesetzten Libraries zu haben.

5.3 Dynamische Code Analyse

In diesem Abschnitt sollen ausgewählte Learning Apps aus unserem Datensatz mittels einer werkzeugunterstützten dynamischen Codeanalyse unter die Lupe genommen werden. Die Vorgehensweise ergänzt die oben beschriebene statische Codeanalyse und zielt primär auf die Qualität des Quellcode, das Netzwerkverhalten und potenzielle Datenlecks in dem von einer App initiierten Datenaustausch über das Internet (Tab. 5).

Tab. 4 Eingebettete Top 25-Libraries unter den Top 10-Apps nach Anzahl der Libraries. A21 – Kinderzimmer Rätsel und Spiel; A22 – Neuro Nation; A23 – Kinder lernen zu schreiben; A24 – Einmaleins lernen & üben; A25 – Mathematik Logik spiele; A26 – Kostenlos Deutsch lernen mit FunEasy Learn; A27 – Skillshare Online Kurse; A28 – Homework Helper & Solver; A29 – M-Learningo; A210 – Kinderspiele f. Kinder

-	A21	A22	A23	A24	A25	A26	A27	A28	A29	A210
Google Firebase Analytics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Google Ads	✓	✓	✓	✓	✓	✓		✓		✓
Google Double-click	✓	✓	✓	✓	✓			✓		✓
Google CrashLytics		✓		✓	✓	✓	✓	✓	✓	
Facebook Login	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Google Analytics		✓			✓	✓	✓	✓	✓	
Facebook Share	✓	✓	✓	✓	✓	✓	✓	✓		
Facebook Analytics	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Google Tag Manager		✓				✓	✓	✓	✓	
Facebook Places	✓	✓	✓				✓			
Facebook Ads				✓	✓	✓				
Unity3d Ads	✓		✓	✓	✓					✓
Moat	✓			✓						✓
AppLovin						✓				
Flurry	✓		✓			✓		✓		
OneSignal					✓	✓				
Amplitude						✓				
Chartboost				✓		✓				
AdColony	✓					✓				
Branch						✓				
ironSource					✓					✓
Adjust	✓									
Inmobi										✓
Integral Ad Science										✓
Braze						✓	✓			

Tab. 5 Top 10-Apps nach Anzahl der eingebetteten Libraries und korrespondierende Informationen bezüglich Datenschutzerklärung und Ursprungsregion

App-Name	Datenschutz-erklärung	Sprache	Ursprungsregion
Kinderzimmer Rätsel und Spiel	✓	Englisch	Sonstige (Indien)
NeuroNation	✓	Deutsch	EU
Kinder lernen zu schreiben	✓	Englisch	Sonstige (Indien)
Einmalleins lernen & üben	✓	Englisch	EU
Mathematik – Logikspiele	✓	Englisch	EU
Kostenlos Deutsch lernen mit FunEasyLearn	✓	Englisch	Sonstige (z. B. Moldawien)
Skillshare Online-Kurse	✓	Englisch	Vereinigte Staaten
Homework Helper & Solver	✓	Englisch	Nicht verfügbar
M-Learning	✓	Französisch	EU
Kinderspiele für Kinder ab 3	✓	Deutsch	Vereinigte Staaten

Hierbei soll eine typische Interaktion des Nutzenden mit der App mittels eines Emulators simuliert und der daraus resultierende Datenverkehr untersucht werden. Der von der App initiierte Datenaustausch über das Internet wird dabei abgefangen und hinsichtlich Datenschutz- und Datensicherheitsaspekten analysiert. Die gesammelten Datenpakete sollen Aufschluss darüber geben, wie personenbezogene und personenbeziehbare Daten durch die App verwendet bzw. an wen weitergeleitet werden. Das Hauptaugenmerk liegt dabei bei dem Einsatz von Third-party Libraries, Datenlecks in initiiertem Datenverkehr und (un)sichere Kommunikation. Aufgrund der Größe unseres Datensatzes ist eine solche feingranulare Analyse für alle 199 Apps nur mit enormen Zeitaufwand möglich. Wir beschränken daher den Umfang der zu analysierenden Apps auf 10 Kandidaten. Die Untersuchungsgegenstände für die werkzeugunterstützte dynamische Codeanalyse und korrespondierende Informationen bezüglich Datenschutzerklärung, Ursprungsregion und Installationen sind in Tab. 6 zusammengefasst.

5.3.1 Datenleaks

Bei dieser feingranularen Analyse soll die Vertraulichkeit der übermittelten Daten, welche potenziell auch persönliche Daten des Nutzers sein können, überprüft werden. Ob ein Datenpaket explizit eine persönliche Information enthält, kann nur eindeutig bestätigt werden, falls die Informationen als Klartext verschickt worden sind oder ein verschlüsseltes Datenpaket zuerst entschlüsselt und anschließend auf persönliche Daten untersucht wird. Um die verschlüsselte Übertragung von Daten

Tab. 6 Untersuchungsgegenstand für die dynamische Analyse

App-Name	Datenschutz-erklärung	Sprache	Ursprungsregion	Installationen (Mio.)
Duolingo	✓	Englisch	Vereinigte Staaten	100
Google Classroom	✓	Deutsch	Vereinigte Staaten	50
NeuroNation	✓	Deutsch	EU	10
Mathe-Einmaleins	✓	Englisch	Nicht verfügbar	10
Coursera: Online courses	✓	Englisch	Nicht verfügbar	10
Quizlet	✓	Deutsch	Vereinigte Staaten	10
Jitsi Meet	✓	Englisch	Nicht verfügbar	5
Mathematik-Logikspiele, Übungen für das Gehirn	✓	Englisch	EU	5
SoloLearn	✓	Englisch	Vereinigte Staaten	5
GeoGebra	✓	Englisch	Nicht verfügbar	5

zu verifizieren, untersuchen wir alle Datenpakete auf die benutzten Protokolle der Datenübertragung. Anhand des verwendeten Protokolls kann beurteilt werden, ob die Daten im Paket verschlüsselt oder als Klartext verschickt worden sind. Falls wir anhand der benutzten Protokolle verifizieren können, dass Daten sicher verschlüsselt übertragen werden, werden insofern Datenlecks nicht berücksichtigt. Daher stützen wir die Argumentation über Datenlecks im App-Datenverkehr auf eine Kombination zwischen dem von Razaghpanah et al. (2015) [7] vorgestellten App Lumen Privacy Monitor und den im Rahmen des Forum Privatheit entwickelten Tools, u. a. PISA Conrad/Simo (2018) [3]. Beide Tools ermöglichen es, Datenpakete lokal auf dem Nutzergerät zu untersuchen und Datenlecks aufdecken. Wir nutzen die Ergebnisse der Analyse durch Lumen Privacy Monitor und PISA, um potenzielle Datenlecks der untersuchten Apps in 3 Risikostufen zu unterteilen: niedriges, mittleres und hohes.

Niedriges Sicherheitsrisiko Unter den gefundenen Sicherheitslücken, welche ein niedriges Risiko für die persönlichen Informationen des Nutzers darstellen, fallen unter anderem Informationen zum Gerät des Nutzers oder etwa Verbindungseinstel-

lungen. Für die vorliegende Studie übernehmen wir die in Lumen Privacy Monitor vordefinierte Spezifikation von niedrigen Sicherheitsrisiken. Diese umfassen zum einen die Marke *(Brand),* das Modell und den Hersteller *(Device Model),* das Betriebssystem und die Android Version *(Build Fingerprint)* des verwendeten Gerätes. Außerdem stellt eine App ein geringfügiges Sicherheitsrisiko für die Daten des Nutzers dar, wenn sie Verbindungseinstellungen *(Connectivity)* einlesen kann. Alle untersuchten Apps weisen bei fast allen niedrigen Risiken einen positiven Befund durch Lumen Privacy Monitor auf.

Mittleres Sicherheitsrisiko Unter mittleren Sicherheitsrisiken werden Datenlecks wie *InstalledApps, Keyword:secret, UnprotectedLocationQueries* und *UntransparentCalenderAccess* zusammengefasst. Installed Apps definiert ein Risiko/ Bedrohung mittleren Grades, da die untersuchte Learning App Zugriff auf die übrigen auf dem Gerät installierten und aktuell ausgeführten Apps erhält und diese Informationen weiterleitet. Die so gewonnenen Informationen können von Entwicklern und Werbetreibenden genutzt werden, um u. a. Persönlichkeitsmerkmale zu inferieren Seneviratne et al. (2014) [9] und Personalisierung als Teil ihrer Marketing-Strategie umzusetzen. Zu den Apps mit dem Datenleck Installed Apps zählen *Duolingo* und *Mathematik – Logikspiele, Übungen für das Gehirn*. Die zweite Lücke *Keyword:secret* weißt auf das Vorhandensein des zuvor eingestellten Suchworts secret im Datenverkehr auf. Diese Tatsache wird als mittelschweres Datenleck ausgedrückt. Wir verwenden während der Simulation die Zeichenkette *„secret“* für Login-Daten bzw. Benutzernamen. Wenn also z. B. die Lumen App diese Zeichenkette im Datenfluss der untersuchten Learning-App ausfindig machen kann, so besteht die Gefahr, dass ein Angreifer ebenso eine solche sensitive Information aus dem Datenfluss ermitteln und ausnutzen kann. Daher stellt dies ein großes Risiko dar. Unter den zehn untersuchten Learning Apps könnte diese Problematik ausschließlich bei der App *JitsiMeet* festgestellt werden. Unter *UntransparentCalenderAccess* fassen wir Datenleaks zusammen, die dadurch entsteht, wenn die untersuchte App durch Query-Schnittstellen und ohne explizite Interaktion mit dem Nutzenden (z. B. Betätigung eines OK-Button) auf Kalendereinträge zugreifen kann. Bei einer der untersuchten Learning-Apps konnte dieses Datenleak festgestellt werden. *UnprotectedLocationQueries* weist auf das Risiko einer aus Sicht des Nutzenden unbeabsichtigen bzw. intransparenten Offenlegung von Standort- oder Web-Abfragedaten hin. Hier fokussieren wir insbesondere auf Leaks, die entweder durch unsichere Kommunikation mit (Dritt-)Dienstanbietern wie z. B. Google Maps oder einer Interpretation von WiFi- und Access-Point-Informationen wie z. B. die WiFi-MAC Adresse Cunche (2014) [4] entstanden sind. Unsere Befunde zeigen, dass 10 % der untersuchten Apps Google-Maps-Anfragen über unsichere Kanäle stellen.

Angreifer können dementsprechend Standort-Informationen aus den übermittelten Anfragen extrahieren. 30 % der Apps geben Details über WiFi- und Access-Point preis. Aggregiert betrachtet sind es 30 % der Apps die das *UnprotectedLocation-Queries* Risiko darstellen – d. h. die Möglichkeit bieten, unautorisiert auf Standorte zuzugreifen und Bewegungsprofile zu erstellen.

Hohes Sicherheitsrisiko Hohe Sicherheitsrisiken umfassen den Zugriff auf und die Weiterleitung von sensitiven Identifiers, welche Werbenetzwerken und Libraries die Möglichkeit gibt, den Nutzer über mehrere Anwendungen und Plattformen hinweg zu beobachten bzw. zu verfolgen. Dies stellt eine massive Beeinträchtigung der Privatsphäre dar, da der Nutzer dadurch zum Opfer von Tracking, Profilbildung und gezielter unerwünschter Werbeanzeigen werden kann. Die Hälfte der im Rahmen der dynamischen Analyse betrachteten Learning-Apps greifen und teilen mindesten einen eindeutigen Identifier (IMEI/MEID in 20 % der Fälle und MAC-Adresse in 30 % der Fälle). Darüber hinaus erfassen alle diese Apps zusätzlich die sog. *„Google advertising ID“,* eine Kennung für Werbetreibende, die es ihnen ermöglicht, die Werbeaktivitäten von Benutzern auf Android-Geräten anonym zu verfolgen. Ein Zugriff auf eindeutige Bezeichner, auch in Kombination mit der Advertising ID, stellt eine klare Verletzung der Google Play-Programmrichtlinien für Entwickler dar.[14]

5.3.2 Third-party Library-Analyse

Gemäß der EU-Gesetzgebung dürfen ohne elterliche Zustimmung keine Tracking-Aktivitäten bei Apps für Kinder stattfinden Stapf et al. (2021) [11]. In den USA werden ähnliche Anforderungen im Rahmen der COPPA (Children's Online Privacy Protection Act) US Congress (1998) [2] formuliert. Bei einigen der untersuchten Learning Apps, konnten keine Hinweise auf Google Play gefunden werden, die darauf abzielen, Eltern – oder den Nutzenden – über eine mögliche Tracking-Aktivität durch die App zu informieren. Im Durchschnitt beinhalten untersuchte Apps ca. 5,6 Third-party Libraries. Zu den am meisten angebundenen Drittparteien gehören *Google Firebase Analytics* (bei 90 % der Apps), *Google Tag Manager* (bei 90 % der Apps) und *Facebook* (bei 80 % der Apps). Unsere Ergebnisse bestätigen die zuvor im Rahmen der statischen Analyse gemachte Feststellung: in Learning Apps sind Third Party Libraries in hohen Maßen vorhanden, siehe Abschn. 5.2.

[14] https://support.google.com/googleplay/android-developer/answer/6048248?hl=de. Stand: 06.07.2020.

5.3.3 Sichere Kommunikation

Zielknoten Zunächst wollen wir die erfassten Zielknoten der einzelnen Verbindungen innerhalb des analysierten Datenflusses betrachten. Die meisten Zieladressen im Datenverkehr der untersuchten Apps liegen in den Vereinigten Staaten, wobei Deutschland am zweithäufigsten vertreten ist. Eine der untersuchten Learning Apps leitet Daten an einen sich in der Volksrepublik China befindlichen Server weiter. Ein Großteil der Apps in dieser Analyse leitet Daten zu IP-Adressen weiter, welche 6 oder weniger verschiedenen Organisationen zuzuordnen sind. Jedoch stechen die Apps *Mathe-Einmaleins, Quizlet* und *Coursera: online courses* besonders hervor, da sie IP-Verbindungen zu mehr als 9 verschiedenen Organisationen aufbauen. Nach unserer Analyse interagiert die App *Coursera* sogar mit mehr als 16 verschiedenen Organisationen.

Verwendete Sicherheitsprotokolle Wie die Ergebnisse der Untersuchung des Datenverkehrs durch Wireshark[15] zeigen, enthalten alle Apps Datenpakete in ihrem Datenfluss, welche das Verschlüsselungsprotokoll Transport Layer Security (TLS) verwenden. Lediglich eine App zeigt Anfälligkeiten und zwar in Bezug auf einen fehlerhaften benutzerdefinierten SSL/TLS-Vertrauensmanager. Hier wurde die Vertrauensverwaltung für Socket-Kommunikation auf unsichere Weise modifiziert. Durch die Anwendung des TLS-Protokolls wird sowohl die Verschlüsselung der übermittelten Daten als auch deren Integrität gewährleistet. Betrachtet man die Verteilung von TLS-Datenpaketen innerhalb des analysierten Datenverkehrs der Apps, so stellen wir fest, dass im Durchschnitt etwa 33 % des Datenflusses in den populärsten Apps mit TLS verschlüsselt ist. Die Abb. 16 stellt die Verteilung der TLS Datenpakete aller untersuchten Apps prozentual dar. Die Apps *Google Classroom* und *JitsiMeet* fallen hierbei besonders auf, da mehr als 50 % des Datenverkehrs mittels TLS verschlüsselt übertragen worden ist. Unter den übrigen Apps wie etwa *NeuroNation, Mathe–Einmaleins* oder *Coursera: online courses* weisen die Apps aber nur weniger als 20 % verschlüsselten TLS-Datenverkehr auf. Gehen wir näher auf die verwendeten TLS-Protokolle ein, so stellen wir fest, dass die Apps hauptsächlich TLS in der Version 1.2 (v1.2) bzw. Version 1.3 (v1.3) benutzen, wobei v1.3 eine neuere und damit fortschrittlichere Version darstellt. In TLS v1.3 wurden einige Sicherheitslücken, welche in der älteren Version 1.2 auftreten, behoben. Betrachten wir die Verbreitung der besser gesicherten TLS v1.3, so stellen wir fest, dass 6 der 10 untersuchten Apps weniger als 1 % des Datenflusses mit dieser TLS Version absichern. Die TLS Version 1.2 hingegen weist im Durchschnitt etwa 28 %

[15] Wireshark. https://www.wireshark.org/. Stand: 06.07.2020.

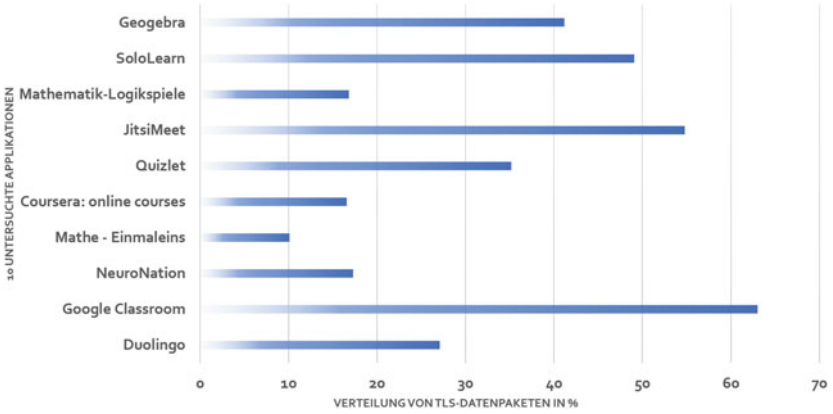

Abb. 16 Aufkommen von TLS-Datenpaketen im gesamten Datenfluss während der Datenverkehrsanalyse der untersuchten Apps

des gesamten Datenflusses auf und ist somit weitaus öfter vertreten als die neuere Version. Obwohl die untersuchten Apps zu den populärsten in unserem Datensatz gehören, so verwenden die Verbindungen dennoch nur selten die TLS v1.3, welche bereits seit 2018 veröffentlicht ist und dem Stand der Technik entspricht.

6 Fazit

Learning Apps gelten inzwischen als wichtige Instrumente zur Bereitstellung schulischer und außerschulischer Bildungsangebote und zur Vermittlung grundlegenden Wissens an Kinder und Jugendliche. Durch Learning Apps können Lernangebote personalisiert und interaktive Lernmethoden bereitgestellt werden. Diese können wiederum jederzeit und von überall konsumiert werden. Eine derartige Digitalisierung des Lern- und Wissensvermittlungsprozesses geht jedoch mit der Notwendigkeit einher, die damit verbundenen Risiken für die informationelle Selbstbestimmung und die Sicherheit der Daten der Minderjährigen zu erkennen und einzuordnen.

Die vorliegende Arbeit untersuchte, inwieweit Learning Apps privatheitinvasiv und anfällig für Cyber-Angriffe sind. Unsere umfassende Analyse von 199 Android Learning Apps belegt, dass ein signifikanter Anteil der untersuchten Apps zahlreiche Anfälligkeiten hinsichtlich Datenschutz und Cybersicherheit aufweisen. Mit den Ergebnissen unserer Studie zielen wir darauf ab, Pädagogen, Minderjährige,

Datenschutzbehörden, Verbraucherschutzorganisationen und App-Entwickler über Datenschutz- und Cybersicherheitsproblematiken im Zusammenhang mit Learning Apps zu sensibilisieren.

Literatur

1. Article 29 Working Party: Guidelines on transparency under Regulation 2016/679. WP260 rev.01. Brussels (2016). https://ec.europa.eu/newsroom/article29/redirection/document/51025
2. Commission, F.T., et al.: Children's online privacy protection rule ("coppa"). Retrieved on September **16** (2016)
3. Conrad, B., Simo, H.: Pisa - towards detecting tracking services via privacy-preserving mobile traffic analysis. Forum Privatheit Tech, Report (2018)
4. Cunche, M.: I know your mac address: targeted tracking of individual using wi-fi. J. Comput. Virol. Hack. Tech. **10**(4), 219–227 (2014)
5. Ehmann, E., Selmayr, M. (Hrsg): Datenschutz- Grundverordnung. Kommentar. 2. Aufl. C.H. Beck, München (2018)
6. Kühling, J., Buchner, B. (Hrsg): Datenschutz- Grundverordnung/BDSG. Kommentar. 2. Aufl. C.H. Beck, München (2018)
7. Razaghpanah, A., Vallina-Rodriguez, N., Sundaresan, S., Kreibich, C., Gill, P., Allman, M., Paxson, V.: Haystack: In situ mobile traffic analysis in user space. arXiv preprint arXiv:1510.01419 S. 1–13 (2015)
8. Roßnagel, A.: Der Datenschutz von Kindern in der Datenschutz-Grundverordnung - Vorschläge für die Evaluierung und Fortentwicklung. Zeitschrift für Datenschutz **9**(2), 88–92 (2020)
9. Seneviratne, S., Seneviratne, A., Mohapatra, P., Mahanti, A.: Predicting user traits from a snapshot of apps installed on a smartphone. ACM SIGMOBILE Mobile Comput. Commun. Rev. **18**(2), 1–8 (2014)
10. Simitis, S.; Hornung, G.; Spieker genannt Döhmann, I. (Hrsg.): Datenschutzrecht (DSGVO mit BDSG). Kommentar. Nomos, Baden-Baden (2019)
11. Stapf, I., Meinert, J., Heesen, J., Krämer, N., Ammicht Quinn, R., Bieker, F., Friedewald, M., Geminn, C., Martin, N., Nebel, M., Ochs, C.: Privatheit und kinderrechte. Schriftenreihe Forum Privatheit und selbstbestimmtes Leben in der digitalen Welt, Karlsruhe S. 24 (2020)
12. Tengler, K., Schrammel, N., Brandhofer, G., Sabitzer, B.: Lernen auf distanz während der corona-krise. chancen und herausforderungen des distance learning für die primarstufe. In: Bildung und Digitalisierung, S. 195–216. Nomos Verlagsgesellschaft mbH & Co. KG (2020)