Wie Sender Tracking und Cookies einsetzen

HBBTV UND DIE DATENSAMMELWUT

Hybrid HbbTV hat sich zu einer beliebten Möglichkeit entwickelt, traditionelles Fernsehen mit internetbasierten Inhalten zu kombinieren. Die Technologie wirft jedoch auch Bedenken hinsichtlich des Schutzes der Privatsphäre der Nutzer auf. Eine Studie hat nun untersucht, wie Daten gesammelt und Nutzer getrackt werden – mit alarmierenden Ergebnissen.

Trotz der wachsenden Beliebtheit moderner Streaming-Dienste ist das Fernsehen nach wie vor ein wichtiges Kommunikations- und Unterhaltungsmedium. Seit der Einführung des HbbTV-Standards im Jahr 2006 hat sich das traditionelle lineare Fernsehen stark weiterentwickelt. Hybrid Broadcast Broadband TV (HbbTV) kombiniert klassisches Fernsehen mit On-Demand-HTML5-Inhalten. Die Technologie, die eine Internetverbindung voraussetzt, bietet zusätzliche Funktionen wie Hintergrundinformationen zu Sendungen und den Zugang zu Mediatheken. HbbTV hat sich vor allem in Europa und besonders in Deutschland etabliert.

HbbTV wirft jedoch auch Fragen der Sicherheit und des Datenschutzes auf. Die Nutzung dieser Technologie ermöglicht es den Kanalbetreibern, Informationen über ihre Zuschauer zu sammeln und zu verfolgen, was zur Erstellung detaillierter „Zuschauerprofile" führt. In der Europäischen Union sind die Anbieter verpflichtet, Datenschutzgesetze wie die Datenschutz-Grundverordnung (DSGVO) einzuhalten, um die Privatsphäre der Nutzerinnen und Nutzer zu schützen.

Obwohl sich frühere Studien mit den Sicherheits- und Datenschutzaspekten von HbbTV befasst haben, wurden das HbbTV-Tracking-Ökosystem und die damit verbundenen Datenschutzfragen bisher kaum umfassend untersucht. Eine groß angelegte Untersuchung von 391 europäischen HbbTV-Kanälen hat nun herausgefunden, wie Daten gesammelt und Nutzerinnen und Nutzer getrackt werden. Darüber hinaus wurden die Datenschutzrichtlinien bewertet und die Einwilligungserklärungen analysiert.

DAS MESS-FRAMEWORK

Die Studie erforderte eine präzise und sorgfältig konzipierte Messinfrastruktur – Abbildung 1 stellt einen Überblick über den verwendeten Versuchsaufbau dar. Für die Untersuchung wurde ein LG 43UK6300 LLB-Fernseher mit HbbTV-Unterstützung verwendet, der mit dem RootMyTV 2.0 Rootkit [8] gerootet wurde. Das ermöglichte die Installation eines Zertifikats im Zertifikatsspeicher des Fernsehers, um TLS-geschützten Netzwerkverkehr abzufangen und zu entschlüsseln. Mit der aktualisierten Version der LG-Firmware funktioniert dieses Rootkit allerdings nicht mehr.

Als Analysegerät diente ein Desktop-Computer, der den Fernseher über einen Wi-Fi-Hotspot mit dem Internet verband. Mit dem HTTP(S)-Proxy mitmproxy (Version 9.0) [1] ließ sich dann der HTTP(S)-Verkehr abfangen. Das Proxy-Zertifikat wurde auf dem gerooteten TV-Gerät installiert, sodass die Forscher den größten Teil des HbbTV-Datenverkehrs entschlüsseln konnten. Aufgrund der fehlenden Zertifikatsvalidierung durch die analysierten Sender war es möglich, den gesamten HTTP(S)-Verkehr zu erfassen. Zusätzlich wurde die webOS TV Developer API [4] verwendet, um Informationen über den aktuellen Sender zu sammeln und Screenshots der angezeigten Inhalte zu erstellen. Auch Daten aus dem Cookie-Speicher und dem lokalen Speicher des Fernsehers wurden erfasst.

Der Fernsehempfang erfolgte über eine Parabolantenne, die Signale von drei Satelliten empfing: Astra 1L (19,2°O), Hot Bird 13E (13,0°O) und Eutelsat 16,0°O. Diese ermöglichten den Empfang von Fernsehsendern aus verschiedenen europäischen Ländern, wobei der physische Standort der Studie in Deutschland war.

Das anfänglich empfangene Signal umfasste 3.575 Sender. Viele davon waren jedoch ungeeignet, da sie entweder keine Programme ausstrahlten, verschlüsselt waren oder Radiosender darstellten. Am Ende des Filterprozesses verblieben 391 Fernsehsender.

Das „Fernbedienungsskript" implementierte fünf Profile, die verschiedene Benutzerinteraktionen mit dem Fernseher simulierten, um unterschiedliche HbbTV-Anwendungen für jeden Kanal auszulösen.

- **Profil ohne simulierte Benutzerinteraktion (General):** Beobachtete jeden Kanal 900 Sekunden lang ohne weitere Interaktionen.
- **Farbtasten-Profile (Rot, Blau, Gelb und Grün):** Jedes Profil beinhaltete geskriptete Interaktionen mit der jeweiligen farbigen Taste. Nach dem Kanalwechsel wartete das Skript 10 Sekunden und drückte dann die jeweilige Farbtaste. Anschließend wurden zufällig Navigationsknöpfe gedrückt, um mit dem möglicherweise geladenen neuen Inhalt zu interagieren.

Diese detaillierte und methodische Herangehensweise ermöglichte eine umfassende Erfassung und Analyse des HbbTV-Datenverkehrs und der damit verbundenen Datenschutzpraktiken.

ERGEBNISSE: TRACKING UND 1700 COOKIES

Die Messungen ergaben, dass 1.705 verschiedene Cookies über HTTP(S) gesetzt wurden. Durchschnittlich wurden pro Kanal 4,1 Cookies gesetzt, wobei 166 unterschiedliche Drittanbieter beteiligt waren. Bemerkenswert ist, dass nur 20,5 Prozent der Cookies von Cookiepedia klassifiziert werden konnten, was darauf hindeutet, dass sich das HbbTV-Ökosystem stark vom Web-Ökosystem unterscheidet.

Verbreitung von Third-Party-Cookies

Cookies von Drittanbietern sind im HbbTV-Ökosystem mit durchschnittlich 3,1 Cookies pro Kanal weit verbreitet. Eine signifikante Anzahl dieser Cookies wird für Trackingzwecke verwendet. Die häufigste Third-Party-Domain war xiti.com, die auf 119 Kanälen beobachtet wurde. Insgesamt zeigen die Ergebnisse, dass HbbTV-Kanäle häufig auf Third-Party-Dienste und Cookies zurückgreifen, was auf umfangreiche Tracking-Praktiken hindeutet. Nur 25 Drittparteien wurden auf mehr als zehn Kanälen verwendet, was

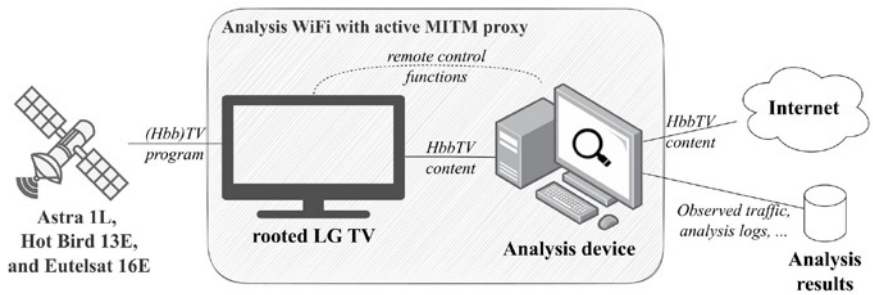

Abbildung 1: Messframework HbbTV (Bild: if(is))

Bild: Robert Daly/KOTO – stock.adobe.com

auf ein verstreutes Third-Party-Cookie-Ökosystem hinweist, im Gegensatz zum Web, das von wenigen großen Akteuren dominiert wird. In Tabelle 1 ist ein Überblick, über die Anzahl der Third-Parties und wie viele Cookies diese bei jedem Profil setzen.

PROFIL	# 3PS	# 3P COOKIES	MEAN	MIN	MAX	SD
GENERAL	36	167	2,31	1	8	1,74
ROT	107	560	3,59	1	22	5,82
GRÜN	77	287	3,69	1	21	4,27
BLAU	47	189	2,04	1	16	2,34
GELB	88	300	3,2	1	24	4,16

Tabelle 1 Cookies von Third Parties nach Profilen

Kommunikation zwischen Third-Parties: Cookie-Syncing

Während viele Drittanbieter direkt auf mehreren Kanälen eingebettet sind, bleibt unklar, ob sie untereinander Daten austauschen. Cookie-Syncing, ein zweistufiger Prozess, ermöglicht den Datenaustausch zwischen Drittparteien. Dabei lädt eine Website zunächst ein Drittanbieter-Skript, welches dann an einen Synchronisierungspartner weitergeleitet wird. Diese weitergeleitete Anfrage enthält beispielsweise eine Benutzer-ID.

Um Cookie-Syncing in den Daten zu identifizieren, analysierte man, ob eine Drittpartei eine HTTP-Anfrage mit einem (Cookie-)Identifikator an eine andere Partei sendete. Dazu adaptierten die Forscher eine Methode von Acar et al. zur Identifizierung von Cookies, die möglicherweise eine ID enthalten könnten. Dabei betrachtet man einen Cookie-Wert als einen Identifikator, wenn er: (1) zwischen 10 und 25 Zeichen lang war (d. h. genügend Entropie für eine ID aufwies) und (2) kein gültiger Unix-Zeitstempel innerhalb des Messzeitraums war. Viele Cookie-Werte enthalten solche Zeitstempel für unterschiedliche Zwecke, zum Beispiel zur Einholung von Einwilligungen oder beim Kanalwechsel.

Diese Methode identifizierte 14.236 Cookie-Werte, die potentiell eine ID sein könnten. Von diesen wurden 25 Werte in einer HTTP-Anfrage an eine andere Partei übertragen. Die meiste Synchronisierungsaktivität fand im Rot-Profil statt, während im Allgemeinen und im Gelben Profil keine einzige Instanz beobachtetet wurde. Nur zwei Domains (nach eTLD+1) verursachten Synchronisierungsaktivitäten, wobei die Cookies Benutzer-IDs enthielten. Diese Domains waren adform.net mit 22 (88 %) Synchronisierungsaktivitäten (Cookie-Name: uid) und arte.tv mit drei (12 %) Synchronisierungsaktivitäten (Cookie-Name: deviceId).

Insgesamt beobachteten die Autoren der Studie Cookie-Syncing-Aktivitäten auf 20 Kanälen. Im Vergleich zu Cookie-Syncing im Web, wo etwa 500 Synchronisierungsverbindungen identifiziert wurden [9], ist die Technologie in HbbTV-Anwendungen weniger verbreitet.

Tracking-Methoden und ihre Verbreitung

Tracking ist ein weit verbreitetes Phänomen in vielen digitalen Diensten. Um zu bewerten, inwieweit HbbTV-Anwendungen solche Techniken nutzen, wurden alle vollständigen URLs aus dem HTTP-Verkehr mit Filterlisten wie EasyList [2] und der Pi-hole-Blockliste [7] verglichen. Es stellte sich heraus, dass nur 0,5 Prozent der URLs von EasyList und 1,17 Prozent von Pi-hole markiert wurden (siehe Tabelle 2). Dies könnte darauf hindeuten, dass Tracking in HbbTV-Anwendungen entweder weniger verbreitet ist oder von anderen Parteien durchgeführt wird als im Web.

Einsatz von Tracking-Pixeln

Tracking-Pixel sind nahezu unsichtbare Bilder, die verwendet werden, um Nutzer zu verfolgen. Die Studienautoren identifizierten Tracking-Pixel anhand von drei Kriterien: (1) Der HTTP-Content-Typ zeigt ein Bild an, (2) die Datenmenge ist kleiner als 45 Byte und (3) der HTTP-Response-Code ist 200 (OK). Es konnten so insgesamt 277.574 Tracking-Anfragen identifiziert werden, von denen nur 0,2 Prozent von EasyList markiert wurden. Diese Anfragen stammten von 47 verschiedenen eTLD+1, wobei 8 (17 %) auf EasyList standen.

Fingerprinting-Techniken

Fingerprinting wird häufig verwendet, um Nutzende ohne Cookies zu verfolgen. Die Studienautoren untersuchten Antworten, die JavaScript-Code enthielten und Skripte, die APIs wie Canvas oder WebGL nutzen. Dieser Ansatz identifizierte 60 (15 %) Sender, die Fingerprinting-Techniken verwendeten, was mit dem Anteil der Websites vergleichbar ist, auf denen diese Technik genutzt wird. Von allen identifizierten Fingerprinting-Anfragen wurde nur eine eTLD+1 von EasyList markiert.

Unzureichende Blocklisten für Smart TVs und HbbTV

Spezielle Blocklisten für Smart-TVs und HbbTV, wie die von Perflyst [6] und Kamran [5], blockierten deutlich weniger Tracking-Anfragen als die reguläre Pi-hole-Filterliste. Dies zeigt, dass diese Listen Nutzer nicht ausreichend vor HbbTV-spezifischem Tracking schützen. Häufig verwendete Tracker wie tvping.com fehlen auf diesen Listen, was darauf hinweist, dass der Fokus dieser Listen auf Anwendungen wie Netflix liegt und nicht auf HbbTV.

PROFILE	PI-HOLE	EASY LIST	TRACKING PIXEL	FINGERPRINTS
GENERAL	203	6	80.960	51
ROT	2.120	1.375	90,199	536
GRÜN	1.051	463	14.593	161
BLAU	313	8	5.925	179
GELB	1.668	660	85.897	151

Tabelle 2: Tracking in HbbTV Profilen

Besonders kritisch ist das Tracking auf Kinderkanälen. Gemäß der DSGVO [3] müssen Kanäle, die Programme für Kinder anbieten, ihre Datenpraktiken anpassen. In der Untersuchung wiesen 12 Kinderkanäle 2.214 Tracking-Anfragen auf und setzten 97 Drittanbieter-„Targeting/Advertising"-Cookies. Das könnte gegen die DSGVO verstoßen. In der statistischen Analyse trat jedoch kein signifikanter Unterschied im Tracking-Verhalten zwischen Kinderkanälen und anderen Kanälen auf.

Das HbbTV-Tracking-Ökosystem

Das HbbTV-Tracking-Ökosystem unterscheidet sich signifikant vom Web-Ökosystem. Ein Netzwerkgraph basierend auf dem beobachteten HbbTV-Verkehr offenbarte ein gut vernetztes System mit 429 Knoten und 675 Kanten. Die am stärksten verbundenen Knoten waren ard.de, redbutton.de und rtl-hbbtv.de, die alle zu deutschen TV-Netzwerken gehören. Diese Ergebnisse werfen die Frage auf, ob HbbTV-Kanäle transparent über ihr Tracking informieren und den Nutzenden Wahlmöglichkeiten bieten. Abbildung 2 stellt das HbbTV-Tracking-Ökosystem dar, in dem blaue Knoten die Kanäle und rote Knoten die First- und Third-Parties repräsentieren. Die Knotengröße gibt die Anzahl der Verbindungen an.

Kanalbasierte Analysen

Die vorangegangene Analyse konzentrierte sich auf verschiedene Messprofile, die die Interaktion der Nutzer mit HbbTV-Anwendungen simulierten. Darüber hinaus wurden die Datenschutzpraktiken der einzelnen Sender untersucht, um

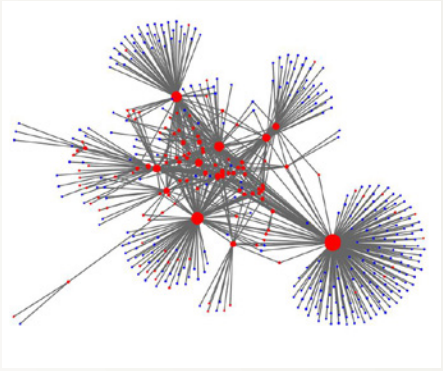

Abbildung 2: Das HbbTV-Tracking-Ökosystem. Blaue Knoten repräsentieren die Kanäle, rote Knoten die First- und Third-Parties, und die Größe eines Knotens zeigt die Anzahl seiner Kanten an. (Bild: if(is))

festzustellen, ob einige Sender die Privatsphäre der Zuschauer stärker beeinträchtigen, indem sie

HBBTV: REVOLUTION DES FERNSEHENS DURCH HYBRIDTECHNOLOGIE

Hybrid Broadcast Broadband TV (HbbTV) ist ein Industriestandard, entwickelt vom Europäischen Institut für Telekommunikationsnormen (ETSI), der darauf abzielt, internetbasierte Inhalte nahtlos mit linearen TV-Programmen zu verbinden. HbbTV ermöglicht es Geräten, die einen Decoder für digitales Fernsehen (Broadcast) und einen Internetzugang (Breitband) besitzen, interaktive Anwendungen zu nutzen. Diese Anwendungen bieten Zusatzelemente zum TV-Programm, wie Video-on-Demand-Dienste, elektronische Programmführer oder interaktive Werbung.

Beispielsweise können solche Inhalte als Overlay über das laufende TV-Programm eingeblendet oder das Programm vollständig ersetzt werden, sodass es nicht mehr sichtbar oder hörbar ist. Kanäle bieten oft einen Einstiegspunkt zu diesen zusätzlichen Inhalten, wenn der Nutzer den Kanal wechselt.

Ein wesentlicher Meilenstein für HbbTV war die Veröffentlichung der Hauptversion 2.0 im Jahr 2015, die bedeutenden Änderungen im HbbTV-Ökosystem einführte, darunter die Fähigkeit, HTML5-Inhalte auf Fernsehgeräten darzustellen. Diese Weiterentwicklung hat die Möglichkeiten für interaktive Inhalte erheblich erweitert und das Nutzerlebnis verbessert.

Um HbbTV-Inhalte anzubieten, wird die URL jeder Anwendung, die ein Kanal bereitstellt, im linearen Broadcast-Signal kodiert. Unterstützt ein Fernseher den HbbTV-Standard, kann er sich mit diesem Endpunkt verbinden und die entsprechende Anwendung laden. Die Übertragung der Anwendungen erfolgt üblicherweise über das HTTP-Protokoll. Für die Darstellung und Ausführung der Anwendungen muss jeder Fernseher eine geeignete Laufzeitumgebung implementieren, die browserähnlich ist und HTML5-Seiten anzeigen, JavaScript-Code ausführen und andere Elemente von Webanwendungen verarbeiten kann.

HbbTV eröffnet neue Möglichkeiten der Interaktion und Informationsbereitstellung und trägt dazu bei, das Fernsehen in das digitale Zeitalter zu führen. Abbildung 1 zeigt ein Beispiel für HbbTV-Inhalte, während Abbildung 2 einen allgemeinen Ablauf darstellt, wie HbbTV-Inhalte über dem linearen TV-Programm angezeigt und aus dem Internet geladen werden.

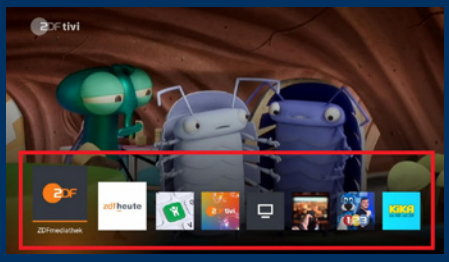

Abbildung 1: Beispiel für HbbTV Elemente (Bild: if(is))

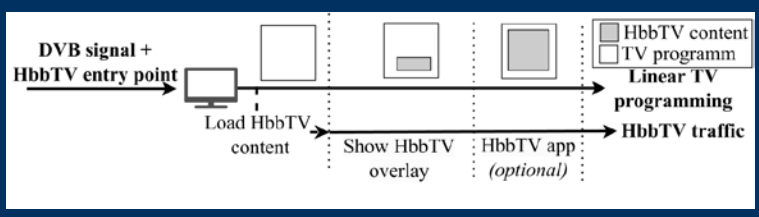

Abbildung 2: HbbTV Ablauf für das Laden von linearem TV und Interaktiven Elementen aus dem Internet (Bild: if(is))

beispielsweise mehr personenbezogene Daten sammeln. Dabei wurden nur Kanäle berücksichtigt, bei denen mindestens eine Tracking-Anfrage beobachtet wurde.

Durchschnittlich sendete ein Kanal 1.132 Tracking-Anfragen, wobei die Spannweite von 1 bis 59.499 reichte. Ein einzelner Kanal stellte dabei 59.499 Tracking-Anfragen, von denen 99,7 Prozent an tvping.com gingen – und das ausschließlich im Rot-Profil. Im Schnitt kontaktierten die Kanäle 7,25 Tracker. Die zehn Kanäle mit den meisten Trackern machten 6,34 Prozent der gesamten Tracking-Anfragen aus. Das verdeutlicht, dass nicht nur wenige Kanäle, sondern viele für die Tracking-Anfragen verantwortlich sind. Statistische Analysen ergaben, dass die Anzahl der verwendeten Tracker stark vom jeweiligen Kanal abhängt. Bestimmte Kanäle verfolgen Nutzer intensiver als andere, aber Tracking ist generell auf allen präsent. Zudem hatte das Benutzerprofil, also welcher Knopf gedrückt wurde, einen größeren Einfluss auf das Tracking-Verhalten als der Kanal selbst. Dies bedeutet, dass die Benutzerinteraktion das Ausmaß des Trackings stärker beeinflusst als der betrachtete Kanal.

EMPFEHLUNGEN: DATENSCHUTZ VERBESSERN

Die Studie hat den Datenschutz im europäischen HbbTV-Bereich unter die Lupe genommen. Anhand eines Messrahmens sammelten die Forscher den HTTP(S)-Datenverkehr von 391 HbbTV-Kanälen und bewerteten deren Datenschutzpraktiken in drei wesentlichen Bereichen: Nutzende-Tracking, Einwilligungskontrollen und Datenschutzrichtlinien. Die Ergebnisse sind alarmierend: Die überwiegende Mehrheit (95 %) der analysierten Kanäle verfolgt ihre Zuschauer mittels Tracking-Pixeln oder Geräte-Fingerprinting.

Dabei unterscheidet sich das HbbTV-Tracking-Ökosystem deutlich vom Web-Ökosystem und bezieht weitere Akteure ein. Trotz der verbreiteten Nutzung von Tracking-Technologien sind diese Praktiken oft schlecht dokumentiert oder nicht offengelegt. Besonders bedenklich ist das Tracking auf Kanälen, die sich an Kinder richten, da hier besondere Datenschutzanforderungen gemäß der DSGVO gelten.

Von den 57 untersuchten Datenschutzrichtlinien wiesen viele erhebliche Mängel auf, darunter nicht deklariertes Tracking und inkonsistente Offenlegungen. Die Zuschauer können sich derzeit nicht darauf verlassen, dass die Standardfilterlisten sie vor Tracking im HbbTV-Bereich schützen.

Angesichts der Ergebnisse sollten die Verantwortlichen maßgeschneiderte Filterlisten für HbbTV entwickeln, die spezifisch auf die Tracking-Praktiken in diesem Bereich abgestimmt sind, um so einen besseren Schutz der Privatsphäre der Zuschauer zu gewährleisten. Zudem ist es essenziell, dass HbbTV-Anbieter ihre Datenschutzpraktiken überarbeiten und für mehr Transparenz sorgen. Das umfasst klare und vollständige Datenschutzrichtlinien sowie effektive Einwilligungskontrollen, die den Zuschauern echte Wahlmöglichkeiten bieten.

Die Umsetzung dieser Maßnahmen würde dazu beitragen, den Datenschutz im Bereich HbbTV deutlich zu verbessern und das Vertrauen der Zuschauer in die Nutzung dieser Dienste zu stärken. ■

Literatur

[1] Cortesi, A., Hils, M., Kriechbaumer, T., and contributors. 2010. mitmproxy: A free and open source interactive HTTPS proxy. https://mitmproxy.org/.

[2] EasyList. 2023. EasyPrivacy.

[3] European Parliament and the Council of the European Union, The. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[4] Iyer, S. 2023. pywebostv 0.8.2.

[5] Kamran, H. 2023. Blocklists.

[6] Perflyst. 2018. PiHoleBlocklist.

[7] Pi-hole. 2023. StevenBlack/hosts.

[8] RootMyTV. 2023. RootMyTV 2.0.

[9] Urban, T., Tatang, D., Holz, T., and Pohlmann, N. 2018. Towards Understanding Privacy Implications of Adware and Potentially Unwanted Programs. In esorics. ESORICS, 449–469. DOI=10.1007/978-3-319-99073-6_22.

CHRISTIAN BÖTTGER
Doktorand im Themenschwerpunkt „Privatsphäre im Internet" im Institut für Internet-Sicherheit – if(is) an der Westfälischen Hochschule in Gelsenkirchen

NURULLAH DEMIR
Post-Doc mit dem Schwerpunkt „Web-Analyse und -Auswertung" im Institut für Internet-Sicherheit – if(is) an der Westfälischen Hochschule in Gelsenkirchen

NORBERT POHLMANN
ist Professor für Cybersicherheit und Leiter des Instituts für Internet-Sicherheit – if(is) an der Westfälischen Hochschule in Gelsenkirchen sowie Vorstandsvorsitzender des Bundesverbands IT-Sicherheit – TeleTrusT und im Vorstand des Internetverbandes – eco.

TOBIAS URBAN
Professor für Cyber-Sicherheit mit dem Forschungsschwerpunkt „Schutz von Online-Anwendungen und Verbesserung der Privatsphäre im Internet" im Institut für Internet-Sicherheit – if(is) an der Westfälischen Hochschule in Gelsenkirchen