Science-Tracker Fingerprinting with Uncertainty: Selected Common Characteristics of Publishers from Network to Application Trackers on the Example of Web, App and Email

Stefan Kiltz, Robert Altschaffel, Jana Dittmann
Faculty of Computer Science.
Otto-von-Guericke University
Magdeburg, Germany
e-mail: sec-by-design <at> iti.cs.uni-magdeburg.de

***Abstract*—Science tracking in information systems provided by publishers is a common pattern in today's scientific world. Besides impacting on the privacy of researchers, science tracking can have potentially adverse or even grave consequences for researchers. The comparison of tracking mechanisms employed by publishers can lead to interesting findings about their popularity over time and to better fine-tuned countermeasures. Publisher could also be identified according to the tracking mechanisms employed. Studies concerning user tracking in general and in particular science tracking exist. The approaches considered therein are concerned with detecting whether tracking is employed on the example of Web, App and email. Our goals are to allow a better comparison of publisher tracking as a privacy measurement by also including results from different forensic tools and to get hints/leads to individualize (attribute) science trackers. Towards our goals we introduce a BNF-style expression based Science-Tracking Fingerprint (STF) as semantics to support individualizing a publisher and for privacy measurements generally. Syntactically, a vector consisting of element-value pairs is created. We propose a certainty category as a metric in support of privacy measurement to rank privacy measurements as plausible, uncertain, or non-match. We investigate intra- and inter-application matches per publisher on URL and tracker information of existing and adapted static tools (of-/on-premises). We analyze 4 exemplary chosen science publishers in 3 selected application areas (web, app, email) with 4 known tools for web, 3 known tools for apps and 2 known tools for email. For the latter, we also introduce a self-implemented tool. We show that the STF is, according to our first tests, fit for the purpose of comparing of publisher tracking and to provide hints/leads for individualization of publishers employing science tracking.**

Keywords-Security, trust and privacy metrics; IT forensics; Attribution.

I. INTRODUCTION

Science tracking is a common practice [1], [2] impacting the privacy and transparency of data processing of users when working with literature information systems. Apart from the obvious privacy violations, as stated in [3], tracking can put scientists at concrete and grave risk. At least it potentially encourages data misuse and academic espionage and can result in personal discrimination against researchers [1]. To have reliable information on the extent of this impact, techniques from IT-forensics are used [4], mainly because of its carefully designed, measured and systematic approaches. Important for IT-forensics is to have at least an estimate on the levels of error, loss and in particular uncertainty [5], [6]. One of the goals of forensics is to individualize the traces and being able to attribute an entity with a given course of action [7] and thus to extract a characteristic fingerprint (quite literally in the case of crime scene forensics). In related work, other user data tracking studies are concerned with detecting science tracking, in this article, however, the individualization of the tracker (attribution) is addressed to allow also a better comparison of publisher tracking. Also, as a result of the individualization, specific and more effective countermeasures against the specific form of science tracking employed by a publisher can be devised.

Our main contributions in this article are:

- the proposition of a Science-Tracking Fingerprint (STF) as semantics based on a BNF expression with a element-value vector description as syntax to individualize science-trackers in support of attribution and improve comparison of publisher tracking as well as the forensic tools results and generally for privacy measurements,
- the certainty category as a metric in support of privacy measurements describing matches between tools as being plausible, uncertain or non-existent.
- intra- and inter-application area (web, app, email) matching of results of existing tools based for URL and tracker detection,
- inclusion of DNS response based A-Record and CNAME evaluation as part of a dynamic network data stream examination involving 1st and 3rd parties,
- implementation of RA_email_forensic toolbox for a semi-automated email forensic examinations,
- practical privacy measurements on 4 selected publishers on all 3 application areas (web, app, email) totaling 191 comparisons
- both our self implemented open source software RA_email_forensics and our material, which is available as open data, can be requested via email-request towards sec-by-design <at> iti.cs.uni-magdeburg.de.

This article is structured as follows: In Section II the state of the state regarding previous work is discussed. In Section III, the fundamentals regarding the employed

 ISBN: 978-1-68558-092-6

forensic model, the existing tools and data sources, and uncertainty in forensic examinations is described.

In Section IV we describe our concept of the semantics and syntax of the Science-Tracking Fingerprint (STF) followed by a tool property comparison. Next we discuss the intra- and inter-application area tool comparison and provide a system landscape analysis. In Section V we describe the implementation of the Science-Tracking Fingerprint (STF) followed by a selected intra-application area matching (web) on the example of one publisher. We provide a Science-Tracking Fingerprint (STF) for 4 selected publishers. In Section VI we perform an evaluation of the findings. The article closes with an conclusion and an outlook on future work in Section VII.

II. STATE OF THE ART

There a number of studies looking into data tracking in general (e.g., [8], [9]). The study done by Hanson [10] looks into the extent of science tracking. These existing studies share the fact that they try to determine to what extent tracking exist on various application fields and elaborate on the consequences of user tracking. One study [4] already employs forensic techniques for the detection of tracking. However, according to our knowledge, no study devised forensically motivated systematic means to give hints/leads to individualize (attribute) tracking to identify an originator. The approach outlined in Section IV is a first attempt at fingerprinting originators of science tracking also for the task of comparing different originators. The authors are fully aware that the suggested approach alone will not suffice for individualization and thus attribution but believe that it can give hints/leads towards further investigation.

III. FUNDAMENTALS

In this section, we provide the prerequisites to comprehend our approach and discuss the need for their advancement. We start with a selection of a specific model of the forensic process, which does not only order steps to be taken during an investigation as many others do, but also locates and describes the data within the examined system and its transformation process.

A. DCEA forensic process model

To examine science tracking in a systematic way, an accepted course of action is to follow a model-based approach that describes a forensic examination. The model from [6] provides the notion of data streams and forensic data types, which together with forensic methods (represented by capabilities of forensic tools) supports a detailed description of the provenance of the data from the beginning of the examination to its end. This is a seen by the authors as an aid to attribution. The model from [6] distinguishes 3 data streams:

- Mass storage data stream DS_T (time-discrete, low volatility, long-term data retention),
- Main memory data stream DS_M (time-discrete, high volatility, short-term data retention),
- Network data stream DS_T (time-continuous, high volatility, short-term data retention).

Throughout article we will use DS_T and DS_N during our examinations. Those data streams can be further divided into 8 forensic data types with the assumption that data of a specific data type is created, processed, stored and used similarly by a given IT system and thus can be acquired, investigated, analyzed and documented similarly in a forensic examination [6]. For our article we use DT_3 (details about data) and DT_5 (communication protocol data) in the context of the network data stream and its representation in mass storage.

Of importance is the system landscape analysis as part of a forensic examination [6]. The spatial and temporal intricacies of tool placement and operation define what can be obtained and analyzed. As stated in [4], the usage of on-premises tools allows for finer control over the tool operation and external data (e.g., lists used for comparison against known tracker URLs) and better data access (e.g., regarding intermediate results). In our research we will use both on- and off-premises tools for two way corroboration of the tool results. In Section IV.E we discuss the properties of both approaches with our system landscape analysis.

The existing model-based approach of the forensic examination as described in [6] alone is not sufficient for the individualization (attribution). However, provides us with the elementary building blocks for the fingerprint (e.g., data streams, forensic data types).

B. Selected tools and data sources for URL and Tracker examination

We select mostly existing tools for the examination of science tracking based on URL and tracker information. These are forensic tools in the sense that they provide relevant forensic information during their operation (similar to the class of methods of the IT-Application, see [6]) but leave the comprehensive documentation and the integrity preservation of the results to the examiner. One exception is the self-implemented *RA_email_forensics* toolbox, which covers the integrity checksum creation and an extensive logging for forensic documentation. This software is available as Open Source per email-request towards sec-by-design <at> iti.cs.uni-magdeburg.de.

Tools can be hosted by a 3rd party (*off-premises*) or operated by the examiners (*on-premises*) and tools can capture a discrete snapshot (*static operation S*) or allow for continuous examinations (*dynamic operation D)*. The latter is used to gather DNS information (A-Record, CNAMEs) based on the requests from the local name services and its responses.

For website-based 3rd party URL information (DT_5 according to [6]) on the network data stream DS_N we select the *static, off-premises* tools privacyscore [11] and webbkoll [12] and the *static, on-premises* tool website-evidence-collector [13]. Additionally, we use wireshark [14] for the dynamic part of the examination.

For website-based *static* tracker information (DT_3 according to [6]) on the network data stream DS_N we use the information also offered by privacyscore [11] and webbkoll [12]. Their results are based on *external data* typically available as lists (e.g., disconnect [15] for webbkoll) and

their accuracy thus depends on the accuracy of that list. Naturally, those lists are dynamic in nature (new tracker sites, sites disappearing, etc.). Thus, for repeatability, the date of the respective list and the list itself need to be conserved, otherwise e.g., a tracker might get reported at t_{i+1} but not at t_i (see also [6]).

For Android app-based 3rd party static URL information (DT_5 according to [6]) we rely on the mass storage data stream DS_T since as the chosen *on-premises* tools of exodus-standalone [16] and appchecker [17] operate on the apk file based representation of an Android app. Those tools provide only a the subset of URL information that also contains a detected tracker. For appchecker this information is readily available, for exodus we perform a manual lookup using the tracker name at the online list provided by the developers [18] For the *dynamic*, *on-premises* investigation we rely on the network data stream DS_N. We use an environment where Android x86 [19] is executed as guest OS on VirtualBox [20] which is running a bridged network with Debian Linux [21] as the host OS. We download the app into the virtualized Android x86 and monitor the connection attempts of the app during startup and idling for 5 minutes using wireshark running on the host and capturing the bridged network. For app-based *static* tracker information (DT_3 according to [6]) we again rely on the mass storage data stream DS_T since this information is made automatically available alongside the URL information by exodus-standalone and appchecker.

For email-based *static* 3rd party URL information (DT_5 according to [6]) on the mass storage data stream DS_T we examine the emails *on-premises* as a file using the existing emlAnalyzer [22] and the self-implemented RA_email_forensics (see also Section IV.C). For a *dynamic*, *on-premises* URL examination we use wireshark whilst starting the ungoogled chromium browser [23] on a Debian system, both of which are configured for being passive on the network. Simulating an email-client's reaction, we paste the extracted URLs from the static examination and record DNS communication and potential redirections.

For email-based *static* tracker information (DT_3 according to [6]) on the storage data stream DS_T we use the RA_email_forensics tool on-premises, which automatically compares the URLs contained in an email against the disconnect list [15]. For email-based *dynamic* tracker information we feed the URLs identified by emlAnalyzer and RA_email_forensics into the ungoogled chromium browser and manually look for phenomena typically employed by trackers (e.g., 1x1 white tracking pixel, mismatch of downloading a gif but returning a webpage, etc.).

The following Table I shows our data sources together with the date and additional information for our 4 selected publishers and 3 application areas (web, app, email). This material is available as Open Data per email-request towards sec-by-design <at> iti.cs.uni-magdeburg.de.

TABLE I. OUR DATA SOURCES WITH DATE AND ADDITIONAL INFORMATION FOR 4 SELECTED PUBLISHERS AND 3 APPLICATION AREAS (WEB, APP, EMAIL)

Application area	Publisher	Data source	Date	Additional information
Web	ACM	https://dl.acm.org	06/06/23	due to unavailable 3rd party and tracker detection re-run for privacyscore on 16/06/2023
	Elsevier	https://www.elsevier.com	06/06/23	due to unavailable 3rd party and tracker detection re-run for privacyscore on 16/06/2023
	IEEE	https://www.ieee.org	06/06/23	due to unavailable 3rd party and tracker detection re-run for privacyscore on 16/06/2023
	SN-MME	https://www.springernature.com/de/macmillaneducation	06/06/23	due to unavailable 3rd party and tracker detection re-run for privacyscore on 16/06/2023
App	ACM	ACM TechNews V1.4.6	31/05/23 12:53	Store URL: https://play.google.com/store/search?q=ACM%20Technews&c=apps
	Elsevier	eReader V8.0.0	31/05/23 11:08	Store URL: https://play.google.com/store/apps/details?id=com.impelsys.elsapac.android.ebookstore
	IEEE	MyXplore V4.0.4	31/05/23 15:21	Store URL: https://play.google.com/store/apps/details?id=org.ieee.mobile.pubs.myxplore
	SN-MME	Macmillan Education eReader V1.0.8.18	01/06/23 09:22	Store URL: https://play.google.com/store/apps/details?id=com.macmillaneducation.ereader
Email	ACM	Subject: Publish Your Work in the ACM Journal on Responsible Computing (JRC)	29/09/22 16:00	Sender: "Kenneth R. Fleischmann, ACM JRC Editor-in-Chief (do not reply)" <call-for-papers@hq.acm.org>
	Elsevier	Subject: Programme announced | Register now to join our expert speakers	27/10/22 18:42	Sender: AI in Aging and Age-related Diseases 2022 <conferences@author.email.elsevier.com>
	IEEE	Subject: July 2022 Issue of IEEE Signal Processing w/Content Gazette Is Now Available	09/08/22 17:00	Sender: IEEE Signal Processing Magazine <ieee-pubs@deliver.ieee.org>
	SN-MME	Subject: Springer Nature Editorial Newsletter May 2022	12/05/22 13:11	Sender: Springer Nature <springernature@newsletter.springernature.com>

C. Uncertainty in forensic examinations

According to [5], all digital evidence comes with some degree of uncertainty. A forensic expert should be capable of determining the level of certainty (and thus having uncertainty to certain extent) given to a piece of evidence. Corroboration is a common mechanism used to raise the level of certainty, where multiple sources of evidence are taken into account. However, if inconsistencies exist, the level of certainty is reduced [5]. A source of inconsistencies is, if one or more forensic tools return different results in repeated runs [6]. In the absence of a ground truth to weight the results against, this results in uncertainty regarding the returned results. In our research presented in this article, we will encounter many instances of uncertainty especially regarding URL information and the tracker detection capabilities of tools.

D. Variations of URL information based on A-Records and CNAMEs

The domain name system (DNS) is used to provide a consistent name space to map resources [24]. Domain names are used to identify a node. Resource information associated with a (domain) name form a resource record (RR). For usage in our paper, two distinct records can be distinguished: A-Record and CNAME. An A-Record describes a host address whereas the CNAME-Record describes a canonical name of an alias to this host [24]. As we will encounter in our research, IP addresses can point towards multiple A-Records, multiple CNAMEs can point towards the same A-Record and possibly further combinations of those three items exist. This adds uncertainty in interpreting tool results, e.g., when trying to associate URLs to entities that could act as 1st or 3rd parties in science tracking research.

CNAMEs can be used for benign purposes, e.g., not having to alter A-Records in a changing environment. But

 ISBN: 978-1-68558-092-6

they can also used to disguise a 3rd party using this DNS mechanism (CNAME cloaking, see e.g., [25]).

IV. Concept of Science-Tracking-Fingerprint (STF) with an intra- and inter-application assessment using URL and tracking Similarities

In this section the semantics and syntax of the concept of a Science-Tracking Fingerprint (STF) is introduced. We look and tools properties and describe intra- and inter-application area comparisons and provide a system landscape analysis.

A. Semantics of Science-Tracking Fingerprint (STF)

The individualization [7] and attribution of science tracking on the application areas of web, app and email is the result of an orchestrated usage of tools following the model-based approach from [6]. The result is a (according to our first, preliminary results) distinct Science-Tracking Fingerprint (STF) that can provide hints/leads towards the publisher's system that is conducting the science tracking whilst also generally supporting privacy measurements. We assume that tool results characteristic and the mode of operation is stable for a given amount of time.

The fingerprint semantics are formed as an evaluation of the tool results when:

- accessing the publisher's website,
- opening an app available from the publisher,
- processing an email from the publisher.

The main challenge is define metrics for ordering the tool results to form the semantics of the Science-Tracking Fingerprint. Since we do not have a ground truth, generally, uncertainty (see also Section III.C) will remain. We propose the notion of Certainty as a metric (also in support of privacy measurements), which uses 3 result categories when matching the set of the individual tool results as being:

- plausible (pl): all tools return the same or comparable result,
- uncertain (unc): at least one tool returns a diverging result,
- none (-): no tool returns a meaningful result.

For URL information (DT_5, see Section III.A), the last category of none (-) is omitted since in our scenario there is no tracker detection (DT_3, see Section III.A) without an URL. URLs from 1st or 3rd party, however, can come with no associated tracker detection.

For the Science-Tracking Fingerprint (STF) we created a notation in the style of the Backus-Nauer Form (BNF) as depicted in Figure 1. A matrix is formed consisting of cells. Each cells carry the semantics of:

- Counter: Number of occurrences,
- Certainty: plausible, uncertain or none,
- Data stream: Mass storage (T) or Network (N),
- Data type: DT_5 (URL) or DT_3 (Tracker),
- Discovery mode: list-based (L) and/or manual (M).

<MATRIX> ::= <ROW> && <MATRIX>

<ROW> ::= <CELL> | /0/ <CELL> | /0/ <CELL> | /0/ <CELL> | /0/

<CELL> ::= <Counter> <EXPR>

<EXPR> ::= <EXPR1> | <EXPR>; <EXPR1>

<EXPR1> ::= <CERTAINTY>,<DATASTREAM>,<DATATYPE> |
<CERTAINTY>,<DATASTREAM>,<DATATYPE>,<DISCOVERYMODE>

Figure 1: BNF-style representation of the Science-Tracking Fingerprint (STF)

This does not only help avoiding uncertainties when comparing tool results. This information is also vital for forming the Science-Tracking Fingerprint (STF). The presence of a particular arrangement of CNAME usage can be characteristic for a given publisher and its embedded 3rd party content.

With those elements we can describe quantifiable and qualitative differences between the science-tracking employed by the publishers. The Science-Tracking Fingerprint (STF) obviously changes if the provision of the application area (web, app, email) by a publisher is altered (e.g., embedded 3rd party content and tracker in the web/app application, number of included items in campaign emails, etc.). We treat the Science-Tracking Fingerprint (STF) as a similarity measure.

B. Syntax of Science-Tracking Fingerprint (STF)

With the BNF-style description of the Science-Tracking fingerprint (STF) we capture its semantics. Syntactically, we create a vector consisting of an element and a value. The concatenation of the vector/element pairs leads to the matrix described in Section III.A as depicted in Figure 2.

	A-Record 1st Party	**CNAME 1st Party**	**A-Record 3rd Party**	**CNAME 3rd Party**
Web	BNF cell	BNF cell	BNF cell	BNF cell
	BNF cell	BNF cell	BNF cell	BNF cell
App	BNF cell	BNF cell	BNF cell	BNF cell
	BNF cell	BNF cell	BNF cell	BNF cell
Email	BNF cell	BNF cell	BNF cell	BNF cell
	BNF cell	BNF cell	BNF cell	BNF cell

BNF Cell: 0 | Counter<pl | unc | none>, <N | T>, <DT5 | DT3>, <M | L>

Figure 2: Syntactical matrix representation of the Science-Tracking Fingerprint (STF)

Each row contains one or more BNF-style Cells where the counter records the number of occurrences if the following conditions are met:

- matching certainty per cell,
- tracker certainty is either plausible or uncertain.

 ISBN: 978-1-68558-092-6

If a row contains entries where the DNS response provided URL information containing CNAMEs for the 1st and/or 3rd party, we duplicate the cell entries from the A-Record to the CNAME. The value of the counter, however, remains unchanged and will only be evaluated once per row since technically this row describes the same examination step.

C. Tool property comparison

The tools used for our experiments represent methods of the forensic process (see Section III.A). In the following Table II we compare the properties of the tools.

TABLE II. TOOL COMPARISON WITH THE PROPERTIES OF LOCATION, INPUT DATA STREAM, INPUT FORENSIC DATA TYPE, OUTPUT DATA STREAM, OUTPUT DATA TYPE, URL OUTPUT, IP OUTPUT, TRACKER DETECTION AND RELIANCE ON EXTERNAL DATA

Application	Tool	Location	Input data stream	Input data type	Output data stream	URL output (DT_5)	IP output (DT_3)	Tracker Output (DT_3) based on external data	external data
Web	Privacyscore	Off-premises	DS_N	(DT_3)	DS_T	x	N/A	x	easilist.to
	Webbkoll	Off-premises	DS_N	(DT_3)	DS_T	x	x	x	disconect
	Website Evidence Collector	On-premises	DS_N	DT_1	DS_T	x	N/A	N/A	N/A
	Wireshark	On-premises	DS_N	DT_1	DS_T	x	x	N/A	N/A
App	Exodus-Standalone	On-premises	DS_T	DT_1	DS_T	x	N/A	x	exodus
	AppChecker	On-premises	DS_T	DT_1	DS_T	x	N/A	x	AppAuthor's list
	Wireshark	On-premises	DS_N	DT_1	DS_T	x	x	N/A	N/A
Email	emlAnalyze	On-premises	DS_T	DT_1	DS_T	x	N/A	x	N/A
	RA_email_forensics	On-premises	DS_T	DT_1	DS_T	x	N/A	x	N/A
	Wireshark	On-premises	DS_N	DT_1	DS_T	x	x	N/A	N/A

All tools internally operate on raw data (DT_1 according to [6]). This also applies to our self-implemented RA_email_forensics (see also Section III.B), which is available per email-request towards sec-by-design <at> iti.cs.uni-magdeburg.de. However, in the case of *off-premises* tools (privacyscore [11] and webbkoll [12]) this data is inaccessible to us. Some tools operate on the network data stream DS_N, however, in our examination we require all tools to produce output data on the mass storage stream DS_T. In the case of the *off-premises* tools this is achieved by saving the results as the html archive using the web browser's export function. CNAME information as part of URL data DT_5 can only be obtained from the *dynamic* examination using wireshark [14]. Wireshark and webbkoll also acquire the IP, which can help finding matches but in times of load balancing and other mechanisms, it is not reliable in case of a mismatch. List-based tracker detection requires external data. Since those lists change over time, *off-premises* tools have a lower repeatability since old lists cannot be supplied easily.

D. Intra- / Inter-Application (Web, App, Mail) comparison to gather semantics

During the tool result comparison as part of our examinations we evaluate intra-application and inter-application matches per publisher. We look for URL data DT_5 and tracker detection data DT_3 intra-application matches between the tools used for examining the application area and maintain the two separately. To detect intra-application matches we aggregate the URL data results of the dynamic examination (e.g., multiple A-Record or CNAME data) using a best-fit approach. Of course the detailed information is kept for further use. Table III shows an exemplary table header for the web intra-application matching.

TABLE III. EXEMPLARY TABLE HEADER FOR INTRA-APPLICATION MATCHING (WEB)

Static examination						Dynamic examination			
Off-Premesis					On-Premesis				
Privacyscore Web_S: Off Premesis {DS_N; DT_5, DT_3}		*Webbkoll Web_S: Off Premesis {DS_N; DT_5, DT_3}*			Website Evidence Collector_S: On Premesis {DS_N; DT_3}	*Wireshark Web_D: On Premesis {DS_N; DT_5, DT_3}*		Detailled Intra-Application Test result (Comparison)	
3rd Parties	Tracker Requests	Domain/ Host	IP	Detected as Tracker	Third party hosts	IP	Address (based on A-Record [A] Or CNAME [C])	Intra-Application DT_5 match	Intra-Application DT_3 Known tracker match

Here, all the properties already discussed such as static/dynamic examination, off-/on-premises operation, forensic data types and data streams captured is maintained and the URL/tracker data and the matching certainty are recorded. A complete examination for a given publisher covers 3 separate tables and forms the semantics for the fingerprint formation.

Further, we are interested in inter-application matches for a given publisher as this could provide hints for cross-application tracking. The idea is to be able to speculate about cross application tracking by identifying shared URL and tracker channels. For evaluating inter-application matches by also applying the certainty categories we evaluate the detailed data returned from the tools (including A-Record and CNAME data).

E. System landscape analysis for Science-Tracking Fingerprint examination

To get an understanding about the opportunities and limitations proposed by the infrastructure, in [6] a system landscape analysis is proposed. Here the connections between the components of interest and the data flows can be visualized and conclusions can be drawn. Figure 3 shows a simplified system landscape analysis for our examinations.

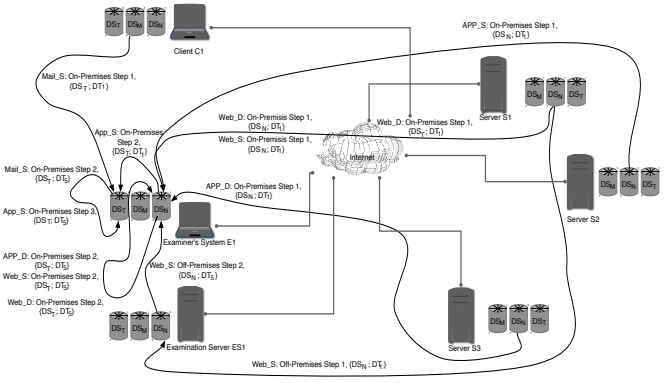

Figure 3: Simplified system landscape analysis for Science-Tracking Fingerprint (STF) examinations visualizing components connections and data flows during the forensic examination

All off-premises examination is performed by the Examination Server ES1. The examiner is located on the Examiner's System E1. Direct access to intermediate results, configurations and external data is only possible for tools running on E1. The data streams are represented as barrels and here the transformation process between data streams and data types during tool usage is visible. Data as input for the examination for the network data stream DS_N is originating from the servers S1...3 representing the

publisher's system. The email based examination is depicted by Client C1 retrieving the emails, which are then transferred using mass storage (USB thumb drive) to the Examiner's System E1. The system landscape is simplified, for an even more thorough examination, the data from the accessible network infrastructure elements such as routers, switches, etc. need to be considered.

V. EXEMPLARY IMPLEMENTATION OF THE SCIENCE-TRACKING FINGERPRINT (STF)

We test our approach of the Science-Tracking Fingerprint (STF) and the intra- and inter-application matching on a laptop that is acting as the Examiner's System E1 (see Section III.E). It is a Lenovo E15 employing an Intel Core i7-1255U CPU, 16GB RAM and a 256GB SSD drive for mass storage. We use a Debian 11 Operating System [21] and the ungoogled chromium Version 95.0.4638.54 [23], both of which are configured to be reasonably passive on the network to avoid contamination of our dynamic recordings. To test the apps we setup Android x86 Version 7.1 inside the virtualization provided by Virtual Box Version 6.1 [20]. For static web investigations we use privacy score [12] public beta (unversioned) and webbkoll [12] in version ec39808 through their respective web interfaces. The website evidence collector is used in Version 1.0.0. For dynamic investigations we use wireshark [14] in Version 3.4.10. For static app examination we use exodus-standalone [16] in Version 1.3.1 and appchecker [17] in Version 2020.05 We use emlAnalyzer [22] (unversioned) and Version 0.5 of the self-implemented RA_email_forensics software. The latter is available as Open Source per email-request towards sec-by-design <at> iti.cs.uni-magdeburg.de.

A. Results of intra-/inter application comparison (URL/Tracker) on the example website of the IEEE

The Table IX shows the result from our website visits of the IEEE (https://www.ieee.org) which are conducted on on June 6th 2023 with the exception of the privacyscore results, which date from June 16th, 2023. This is due to the off-premises scanning engine malfunctioned during the initial test and returned no results for 3rd party connections and known trackers.

We decide for a best-fit attempt when aggregating results for comparison of the static results of privacyscore, webbkoll and website evidence collector with the dynamic results of wireshark regarding the selection of A-Record or CNAME for the table. Greyed-out regions mark unavailable data, highlighting uncertainty through absence of data. Generally we apply the certainty result categories from Section IV.A. As can be seen in Table, we record the following occurrences of matches regarding URL information (DT_5):

- 23 plausible matches,
- 8 uncertain matches,

and regarding the tracker detection information (DT_3):

- 15 plausible matches,
- 6 uncertain matches,
- 10 mismatches (none).

Access to all tables (detailed and aggregated view) of all 4 publishers is available per email-request towards sec-by-design <at> iti.cs.uni-magdeburg.de.

B. Science-Tracking fingerprint of 4 selected publishers

The complete aggregated view on our data regarding all 4 publishers is shown in Table IV.

TABLE IV. AGGREGATED SUMMARY OF ALL INTRA- AND INTER-APPLICATION RESULTS FOR URL (DT5) AND TRACKER DETECTION (DT3) DATA

	aggregated DT_5 matches			aggregated DT_3 matches		
	pl	unc	none	pl	unc	none
Intra-Web (ACM)	19	5	0	5	9	10
Intra-App (ACM)	0	8	0	2	0	6
Intra-Emai (ACM)	3	1	0	1	0	3
Inter-Application (ACM)	0	1	35	1	0	35
Intra-Web (Elsevier)	5	12	0	2	4	11
Intra-App (Elsevier)	0	18	0	4	2	12
Intra-Email (Elsevier)	9	0	0	9	0	0
Inter-Application (Elsevier)	0	1	43	0	1	43
Intra-Web (IEEE)	23	8	0	15	6	10
Intra-App (IEEE)	0	15	0	1	3	11
Intra-Email (IEEE)	5	0	0	5	0	0
Inter-Application (IEEE)	2	3	46	1	3	47
Intra-Web (SN-MME)	11	11	0	2	6	14
Intra-App (SN-MME)	0	7	0	1	1	5
Intra-Email (SN-MME)	31	0	0	1	0	30
Inter-Application (SN-MME)	0	1	59	0	1	59

It represents all results of the intra-and inter-application area examination per publisher. Inter-application matches could hint towards cross-application tracking. The intra-application area results also form the basis of the Science-Tracking Fingerprint (STF) containing the BNF-style semantic cell description from Section IV.A and the syntactical vector from Section IV.B.

Table V shows the resulting Science-Tracking Fingerprint (STF) for the ACM publisher based on the data from Section III.B and tools from Section III.C.

TABLE V. SCIENCE-TRACKING FINGERPRINT (STF) OF THE ACM PUBLISHER USING THE SEMANTIC BNF-STYLE DESCRIPTION AND THE SYNTACTICAL VECTOR FORMED BY ELEMENT-VALUE PAIRS

	A-Record 1st party	CNAME 1st Party	A-Record 3rd Party	CNAME 3rd Party
Web	0	0	$3_{PL\_N,DT5;PL\_N,DT3,L}$	0
	0	0	$2_{PL\_N,DT5;PL\_N,DT3,L}$	$2_{PL\_N,DT5;PL\_N,DT3,L}$
	0	0	$6_{PL\_N,DT5;UNC\_N,DT3,L}$	0
	0	0	$2_{PL\_N,DT5;UNC\_N,DT3,L}$	$2_{PL\_N,DT5;UNC\_N,DT3,L}$
	0	0	$1_{UNC\_N,DT5;UNC\_N,DT3,L}$	$1_{UNC\_N,DT5;UNC\_N,DT3,L}$
App	0	0	$1_{UNC\_T,DT5;PL\_T,DT3,L;\ UNC\_N,\ DT5;\ PL,\ S,\ DT3,\ L}$	$1_{UNC\_T,DT5;PL\_T,DT3,L;\ UNC\_N,\ DT5;\ PL,\ S,\ DT3,\ L}$
	0	0	$1_{UNC\_T,DT5;PL\_T,DT3,L;\ UNC\_N,\ DT5;\ PL,\ S,\ DT3,\ L}$	0
Email	0	0	$1_{PL\_T,DT5;PL\_T,DT3,M,\ 1PL\_N,DT5;PL\_N,DT3,M}$	0

Table VI shows the resulting Science-Tracking Fingerprint (STF) for the Elsevier Publisher based on the data from Section III.B and tools from Section III.C.

TABLE VI. SCIENCE-TRACKING FINGERPRINT (STF) OF THE ELSEVIER PUBLISHER USING THE SEMANTIC BNF-STYLE DESCRIPTION AND THE SYNTACTICAL VECTOR FORMED BY ELEMENT-VALUE PAIRS

	A-Record 1st party	CNAME 1st Party	A-Record 3rd Party	CNAME 3rd Party
Web	0	0	$2_{PL\_N,DT5;PL\_N,DT3,L}$	$2_{PL\_N,DT5;PL\_N,DT3,L}$
	0	0	$2_{PL\_N,DT5;UNC\_N,DT3,L}$	0
	0	0	$1_{PL\_N,DT5;UNC\_N,DT3,L}$	$1_{PL\_N,DT5;UNC\_N,DT3,L}$
	0	0	$1_{UNC\_N,DT5;UNC\_N,DT3,L}$	0
App	$4_{UNC\_T,DT5;PL\_T,DT3,L;\ UNC\_N,\ DT5;\ PL,\ S,\ DT3,\ L}$	0	0	0
	$2_{UNC\_T,DT5;UNC\_N,DT3,L;\ UNC\_N,\ DT5;\ UNC,\ S,\ DT3,\ L}$	0	0	0
Email	$1_{PL\_T,DT5;PL\_T,DT3,M,\ 1PL\_N,DT5;PL\_N,DT3,M}$	0	0	$1_{PL\_T,DT5;PL\_T,DT3,M,\ 1PL\_N,DT5;PL\_N,DT3,M}$
	0	0	$8_{PL\_T,DT5;PL\_T,DT3,L,\ 8PM\_N,DT5;PL\_N,DT3,M}$	$8_{PL\_T,DT5;PL\_T,DT3,M,\ 8PL\_N,DT5;PL\_N,DT3,M}$

 ISBN: 978-1-68558-092-6

Table VII shows the resulting Science-Tracking Fingerprint (STF) for the IEEE Publisher based on the data from Section III.B and tools from Section III.C.

TABLE VII. SCIENCE-TRACKING FINGERPRINT (STF) OF THE IEEE PUBLISHER USING THE SEMANTIC BNF-STYLE DESCRIPTION AND THE SYNTACTICAL VECTOR FORMED BY ELEMENT-VALUE PAIRS

	A-Record 1st party	CNAME 1st Party	A-Record 3rd Party	CNAME 3rd Party
Web	0	0	$8_{PL,N,DT5;PL,N,DT3,L}$	0
	0	0	$5_{PL,N,DT5;PL,N,DT3,L}$	$5_{PL,N,DT5;PL,N,DT3,L}$
	0	0	$3_{PL,N,DT5;UNC,N,DT3,L}$	0
	0	0	$2_{PL,N,DT5;UNC,N,DT3,L}$	$2_{PL,N,DT5;UNC,N,DT3,L}$
	0	0	$2_{UNC,N,DT5;PL,N,DT3,L}$	0
	0	0	$2_{UNC,N,DT5;UNC,N,DT3,L}$	0
App	0	0	$1_{UNC,N,DT5;PL,N,DT3,L;\ UNC,T,\ DT5;\ PL,\ S,\ DT3,\ L}$	0
	0	0	$3_{UNC,N,DT5;UNC,N,DT3,L;\ UNC,T,\ DT5;\ UNC,\ S,\ DT3,\ L}$	0
Email	$5_{PL,T,DT5;PL,T,DT3,M}$, $5_{PL,N,DT5;PL,N,DT3,M}$	0	0	$5_{PL,T,DT5;PL,T,DT3,M}$, $5_{PL,N,DT5;PL,N,DT3,M}$

Table VIII shows the resulting Science-Tracking Fingerprint (STF) for the Springernature-Macmillan Education Publisher based on the data from Section III.B and tools from Section III.C.

TABLE VIII. SCIENCE-TRACKING FINGERPRINT (STF) OF THE SPRINGERNATURE-MACMILLANEDUCATION PUBLISHER USING THE SEMANTIC BNF-STYLE DESCRIPTION AND THE SYNTACTICAL VECTOR FORMED BY ELEMENT-VALUE PAIRS

	A-Record 1st party	CNAME 1st Party	A-Record 3rd Party	CNAME 3rd Party
Web	0	0	$1_{PL,N,DT5;PL,N,DT3,L}$	0
	0	0	$1_{PL,N,DT5;PL,N,DT3,L}$	$1_{PL,N,DT5;PL,N,DT3,L}$
	0	0	$3_{PL,N,DT5;UNC,N,DT3,L}$	0
	0	0	$3_{PL,N,DT5;UNC,N,DT3,L}$	$3_{PL,N,DT5;UNC,N,DT3,L}$
App	0	0	$1_{UNC,T,DT5;PL,T,DT3,L;\ UNC,N,\ DT5;\ PL,\ N,\ DT3,\ L}$	0
	0	0	$1_{UNC,T,DT5;UNC,T,DT3,L;\ UNC,N,\ DT5;\ UNC,\ N,\ DT3,\ L}$	0
Email	0	0	$1_{PL,T,DT5;PL,T,DT3,ML}$, $1_{PL,N,DT5;PL,N,DT3,ML}$	$1_{PL,T,DT5;PL,T,DT3,ML}$, $1_{PL,N,DT5;PL,N,DT3,ML}$

As can be seen from Tables 5-7, the Science-Tracking Fingerprints are significantly distinct. The distinctiveness would be maintained if small changes appeared such as slightly changing the number of 3rd parties involved or changing characteristics thereof.

VI. EVALUATION

Aggregated for the 4 publishers in total we identify 106 plausible, 85 uncertain and 0 mismatches for intra-application area URL matches per publisher out of 191 comparisons. For the tracker detection, we identify 48 plausible, 31 uncertain and 112 mismatches out of 191 comparisons. Additionally, covering all 4 publishers, we identify 2 plausible, 6 uncertain and 124 mismatches for intra-application area URL matches per publisher. For the tracker detection we identify 2 plausible, 6 uncertain and 125 mismatches.

The evaluation of the results leads to interesting findings regarding different name selections by the same domain owner. We propose to use whois [26] queries if some URL results seam similar, e.g.,www.google-analytics.com | firebase.google.com. This is partly backed by Google Inc. itself [27]. Such similarities result in uncertain intra- and inter-application matches Web/App for the publisher IEEE). Here we suggest to look for the entries of:

- Registrant,
- Admin,
- Tech.

If those entries match with regards to domain names (sometimes name server information differs), we would suggest this to be a manual match resulting to uncertain. Interestingly, such matches can also be found for URLs that at first look seem to be a sure mismatch (e.g., tiqcdn.com | tealium.com as intra- and inter-application match Web/App for the publisher IEEE). Again, we would opt for a manual match in that case. Also, the Adobe Ad cloud constituents from Marketo often have very dissimilar URLs (e.g., marketo.com | omtrdc.net | mktossl.com | mktoweb.com but all pointing towards Adobe Inc.). Those claims are partly backed by Adobe Inc. itself [28], [29].

Reliance on live DNS lookups in the case of wireshark is not recommended. In our experience in the case of multiple A-Records or multiple CNAMEs or a combination thereof often one hostname is chosen for display (with no easy detectable logic as to which one is chosen). Here we strongly suggest to look into the responses of the DNS server within the network data stream recording.

As already stated in Section II.B, automated tracker detection often relies on external data typically provided lists. Since the content is dynamic and thus differs between time t_i and t_{i+1}, for the Science-Tracking Fingerprint to deliver comparable results, dates and contents of the respective lists need to be conserved. This calls for local installments for privacyscore and webbkoll, which is possible due to both tools being released as OpenSource. We leave this to future work due to time constraints.

The results from Section IV.B show, that each publisher out of the tested 4 can easily be individualized. This provides leads/hints towards an attribution by providing a similarity measure. However, it must be stated that the Science-Tracing Fingerprint alone is not deemed by the authors to be sufficient for attribution beyond reasonable doubt.

Also, in a quick exemplary evaluation using the IEEE publisher, 3 calls-for-paper emails were evaluated (intra-application only). They cover a time span from August, 9th, 2022 to August, 28th, 2023, and show similarities and differences compared to the email investigated in the main body of work from August, 9th, 2022 which announces the availability of a journal as content. Their STF is shown in Table X. Similarities include the usage of CNAMEs pointing towards usage of 3rd parties whilst the A-Record suggests 1st party (except for Table 10.b, where only 3rd parties are involved). Also, in Table 10.b, an image resource is used that is hosted by a server identified by the disconnect list [15], Arguably, this stems from the conference organization and not from the publisher.

Generally, from the viewpoint of the user, unnecessary tracking is also impacting resource use, so the proposed Science-Tracking Fingerprint (STF) can also be used to look at energy efficiency for sustainability purposes.

VII. CONCLUSION AND FUTURE WORK

In this article we presented the Science-Tracking Fingerprint (STF) that provides a similarity measure to individualize and compare science publishers with qualitative and quantitative elements based on an existing model of the forensic process. The idea is to deliver

 ISBN: 978-1-68558-092-6

hints/leads towards attribution of a specific science literature provider. Its semantics are based on a BNF-style representation of the Science-Tracking Fingerprint (STF).

Syntactically the elements are based on a vector representation of element-value pairs forming a matrix that contains a number of occurrences on the quantitative part also usable for general privacy measurements. URL information is based on DNS responses of the local DNS server and split into A-Record and CNAME information as part of the dynamic examination. This offers a considerably larger chance of finding matchs. A certainty result category as a metric in support of privacy measurements based on URL and tracker detection information forms the qualitative part. The Science-Tracking Fingerprint (STF) is tool-independent. We tested it during privacy measurements on the application areas of web, app and email of 4 publishers with existing on- and off-premises tools and implement the dedicated email tool of RA_email_forensics.

Part of the STF is the intra-application matching, where URL and tracker detection results are compared using the certainty categories. Our results using the Science-Tracking Fingerprint show promising pointers towards its discriminatory power. We also calculated inter-application area matches per provider, which could provide indicators for cross-application tracking. The proposed Science-Tracking Fingerprint (STF) can also be used to look at energy efficiency for sustainability purposes through analyzing the data packets used for science tracking by the publishers.

Future work includes the STF broad scale testing based on automation, which is supported by the BNF-style formalization. Further, a local installation of the section of the tools that is yet off-premises is necessary to remove the reliance on 3rd party administered updates of external data. This is because external data influences the tracker detection.

ACKNOWLEDGEMENTS

The research from Robert Altschaffel is partly funded by the research project "CyberSec LSA_OVGU-AMSL, Security-by-Design-Orchestration_Booster" under the Grant Number ZS/2015/12/96222.

REFERENCES

[1] Deutsche Forschungsgemeinschaft, "Data tracking in research: aggregation and use or sale of usage data by academic publishers" [Online] https://www.dfg.de/download/pdf/foerderung/programme/lis/datentracking_papier_en.pdf (2023.09.05).

[2] E. Bettinger, and M. Bursic, and A. Chandler, "Disrupting the Digital Status Quo: Why and How to Staff for Privacy in Academic Libraries" [Online] https://publish.illinois.edu/licensingprivacy/files/2023/06/Whitepaper-on-Privacy-Staffing-Licensing-Privacy.pdf (2023.09.05).

[3] R. Siems, "When your journal reads you – user tracking on science publisher platforms", Elephant in the Lab. https://doi.org/10.5281/zenodo.4683778, 2021.

[4] R. Altschaffel, and S. Kiltz, and T. Lucke, and J. Dittmann, "Introduction to Being a Privacy Detective: Investigating and Comparing Potential Privacy Violations in Mobile Apps Using Forensic Methods", in Proceedings of the Fourteenth International Conference on Emerging Security Information, Systems and Technologies (Securware), Valencia, Spain, 21-25/09/2020, ISBN 978-1-61208-821-1, pp 60-68, 2020.

[5] E. Casey, “Error, Uncertainty and Loss in Digital Evidence”, In International Journal of Digital Evidence, Volume 1, Issue 2, pp. 1-45, 2002.

[6] S. Kiltz, "Data-Centric Examination Approach (DCEA) for a qualitative determination of error, loss and uncertainty in digital and digitised forensics", PhD Thesis, Faculty of Computer Science, Otto-von-Guericke-University Magdeburg, Germany, September, 2020.

[7] K. Inman and N. Rudin, “Principles and Practises of Criminalistics: The Profession of Forensic Science”, CRC Press LLC Boca Raton Florida, USA, ISBN 0-8493-8127-4, 2001.

[8] W. Christl, "Corporate Surveillance in Everyday Life" [Online] https://crackedlabs.org/dl/CrackedLabs_Christl_CorporateSurveillance.pdf (2023.09.05).

[9] H. Mildebrath "Unpacking 'commercial surveillance': The state of tracking" [Online] https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/739266/EPRS_BRI(2022)739266_EN.pdf (2023.09.05).

[10] C. Hanson, "User Tracking on Academic Publisher Platforms" [Online] https://www.codyh.com/writing/tracking.html (2023.09.05).

[11] D. Herrmann, "Welcome - PrivacyScore"[Online] https://privacyscore.org/ (2023.09.05).

[12] Dataskydd.net Sverige, "Analyze | Webbkoll - dataskydd.net" [Online] https://webbkoll.dataskydd.net/en (2023.09.05).

[13] European Data Protection Supervisor, "EDPS Inspection Software | European Data Protection Supervisor" [Online] https://edps.europa.eu/edps-inspection-software_en (2023.09.05).

[14] Wireshark Foundation, "Wireshark · Go Deep" [Online] https://www.wireshark.org/ (2023.09.05).

[15] Disconnect Inc., "GitHub - disconnectme/disconnect-tracking-protection: Canonical repository for the Disconnect services file" [Online] https://github.com/disconnectme/disconnect-tracking-protection (2023.09.05).

[16] Exodus Privacy, "GitHub - Exodus-Privacy/exodus-standalone: εxodus CLI client for local analysis" [Online] https://github.com/Exodus-Privacy/exodus-standalone (2023.09.05).

[17] J. Alemann, and N. Baier, and M. Streuber, and T. Nam, and L. Peters, "GitHub - Tienisto/AppChecker" [Online] https://github.com/Tienisto/AppChecker (2023.09.05).

[18] Exodus Privacy, "εxodus" [Online] https://reports.exodus-privacy.eu.org/en/trackers/ (2023.09.05).

[19] cwhung, "Android-x86 - Porting Android to x86" [Online] https://www.android-x86.org/ (2023.09.05).

[20] Oracle Inc., "Oracle VM VirtualBox" [Online] https://www.virtualbox.org/ (2023.09.05).

[21] Software in the Public Interest, Inc. , "Debian -- News -- Debian 11 "bullseye" released" [Online] https://www.debian.org/News/2021/20210814 (2023.09.05).

[22] F. Wahl, "GitHub - wahlflo/eml_analyzer: A cli script to analyze an E-Mail in the EML format for viewing the header, extracting attachments, etc." [Online] https://github.com/wahlflo/eml_analyzer (2023.09.05).

[23] ungoogled-chromium Authors, "GitHub - ungoogled-software/ungoogled-chromium: Google Chromium, sans integration with Google" [Online] https://github.com/ungoogled-software/ungoogled-chromium (2023.09.05).

[24] P. Mockapetris, "Domain names - concepts and facilites" [Online] https://datatracker.ietf.org/doc/pdf/rfc1034 (2023.09.05).

[25] Palo Alto Networks, Inc., "CNAME Cloaking: Disguising Third Parties Through the DNS" [Online] https://unit42.paloaltonetworks.com/cname-cloaking/ (2023.09.05).

[26] M. d'Itri, "whois(1) — whois — Debian bullseye — Debian Manpages" [Online]

https://manpages.debian.org/bullseye/whois/whois.1.en.html (2023.09.05).

[27] Google Inc., "What is Google Analytics for Firebase? - Firebase Help" [Online] https://support.google.com/firebase/answer/7388022?hl=EN (2023.09.05).

[28] Adobe Inc., "trackingServer | Adobe Analytics" [Online] https://experienceleague.adobe.com/docs/analytics/implementation/vars/config-vars/trackingserver.html?lang=en-US (2023.09.05).

[29] Adobe Inc., "Get started with tracking | Adobe Campaign" [Online] https://experienceleague.adobe.com/docs/campaign-classic/using/sending-messages/tracking-messages/about-message-tracking.html?lang=en (2023.09.05)

 ISBN: 978-1-68558-092-6

TABLE IX. AGGREGATED INTRA-APPLICATION COMPARISON FOR THE WEB APPLICATION ARE FOR THE PUBLISHER IEEE (HTTPS://WWW.IEEE.ORG, 2023.09.05)

Application Web									
Static examination						**Dynamic examination**			
Off-Premesis					**On-Premesis**				
Privacyscore Web_S:Off Premesis {DS_N; DT_5, DT_3}		*Webbkoll Web_S: Off Premesis {DS_N; DT_5, DT_3}*			Website Evidence Collector_S: On Premesis *{DS_N; DT_5}*	*Wireshark Web_D: On Premesis {DS_N; DT_5, DT_3}*		Detailled Intra-Application Test result (Comparison)	
3rd Parties	Tracker Requests	Domain/Host	IP	Detected as Tracker	3rd party hosts	IP	Address (based on A-Record [A] or CNAME [C])	Intra-Application DT_5 match	Intra-Application DT_3 known tracker match
insight.adsrvr.org	x	insight.adsrvr.org	15.197.193.217	x	insight.adsrvr.org	52.223.40.198	insight.adsrvr.org [A]	pl	pl
js.adsrvr.org	x	js.adsrvr.org	18.165.129.129	x	js.adsrvr.org	18.64.82.184	js.adsrvr.org [A]	pl	pl
						23.54.103.160	www.ieeee.org [A]	unc	-
s3.amazonaws.com		s3.amazonaws.com	52.216.38.0		s3.amazonaws.com	52.216.44.24	s3.amazonaws.com [A]	pl	-
s3-us-west-2.amazonaws.com		s3-us-west-2.amazonaws.com	52.218.217.72		s3-us-west-2.amazonaws.com	52.218.245.56	s3-us-west-2.amazonaws.com [A]	pl	-
cdnjs.cloudflare.com	x	cdnjs.cloudflare.com	2a06:98c1:3123:e000::		cdnjs.cloudflare.com	104.17.25.14	cdnjs.cloudflare.com [A]	pl	unc
4490791.fls.doubleclick.net	x	4490791.fls.doubleclick.net	142.250.74.166	x	4490791.fls.doubleclick.net	142.250.181.198	4490791.fls.doubleclick.net [A]	pl	pl
googleads.g.doubleclick.net	x	googleads.g.doubleclick.net	2a00:1450:400f:80b::2002	x	googleads.g.doubleclick.net	142.251.209.130	googleads.g.doubleclick.net [A]	pl	pl
stats.g.doubleclick.net	x	stats.g.doubleclick.net	2a00:1450:4010:c0d::9c	x	stats.g.doubleclick.net	142.250.147.155	stats.g.doubleclick.net [A]	pl	pl
						142.250.181.194	adservice.google.de [A]	unc	-
connect.facebook.net	x	connect.facebook.net	2a03:2880:f013:d:face:b00c:0:3	x	connect.facebook.net	157.240.223.15	connect.facebook.net [A]	pl	pl
www.facebook.com	x	www.facebook.com	2a03:2880:f113:81:face:b00c:0:25de	x	www.facebook.com	157.240.223.35	www.facebook.com [A]	pl	pl
kit.fontawesome.com		kit.fontawesome.com	2606:4700::6812:1734		kit.fontawesome.com	104.18.22.52	kit.fontawesome.com [A]	pl	-
adservice.google.com	x	adservice.google.com	2a00:1450:400f:801::2002	x	adservice.google.com	172.217.19.66	adservice.google.com [A]	pl	pl
www.google.com	x	www.google.com	2a00:1450:400f:802::2004	x	www.google.com	142.250.181.196	www.google.com [A]	pl	pl
www.google.de	x	www.google.de	2a00:1450:400f:801::2003	x	www.google.de	142.251.209.131	www.google.de [A]	pl	pl
region1.google-analytics.com		region1.google-analytics.com	2001:4860:4802:32::36	x	region1.google-analytics.com	216.239.32.36	region1.analytics.google.com [A]	pl	unc
www.google-analytics.com	x	www.google-analytics.com	2a00:1450:400f:802::200e	x	www.google-analytics.com			unc	pl
www.googletagmanager.com	x	www.googletagmanager.com	2a00:1450:400f:803::2008		www.googletagmanager.com	142.250.181.200	www.googletagmanager.com [A]	pl	unc
					securesso.ieee.org	140.98.193.42	securesso.ieee.org [A]	unc	-
code.jquery.com		code.jquery.com	2001:4de0:ac18::1:a:2b		code.jquery.com	69.16.175.42	code.jquery.com [A]	pl	-
app-ab24.marketo.com	x	app-ab24.marketo.com	104.16.96.80	x	app-ab24.marketo.com	104.16.96.80	app-ab24.marketo.com [A]	pl	pl
munchkin.marketo.net	x	munchkin.marketo.net	23.61.220.209	x	munchkin.marketo.net			unc	pl
756-gph-899.mktoresp.com	x	756-gph-899.mktoresp.com	192.28.144.124		756-gph-899.mktoresp.com			unc	unc
up.pixel.ad	x	up.pixel.ad	95.140.228.46		up.pixel.ad	178.79.242.16	up.pixel.ad [A]	pl	unc
di.rlcdn.com	x	di.rlcdn.com	35.244.174.68	x	di.rlcdn.com	35.244.174.68	di.rlcdn.com [A]	pl	pl
		6045067.global.siteimproveanalytics.io	18.185.183.56		6045067.global.siteimproveanalytics.io	18.185.183.56	6045067.global.siteimproveanalytics.io [A]	unc	-
		siteimproveanalytics.com	2606:4700:e4::ac40:ad0c		siteimproveanalytics.com	172.64.173.12	siteimproveanalytics.com [A]	unc	-
pixel.sitescout.com	x	pixel.sitescout.com	98.98.134.241	x	pixel.sitescout.com	98.98.134.243	pixel.sitescout.com [A]	pl	pl
tags.tiqcdn.com		tags.tiqcdn.com	2600:9000:2375:8a00:7:2bfb:7c00:93a1		tags.tiqcdn.com	18.64.79.94	tags.tiqcdn.com [A]	pl	-
www.youtube.com		www.youtube.com	2a00:1450:400f:80a::200e	x	www.youtube.com	142.251.209.142	www.youtube.com [A]	pl	unc

TABLE X. SCIENCE-TRACKING FINGERPRINT (STF) FOR CALL FOR PAPERS EMAILS FROM THE IEEE PUBLISHER DATING FROM A) AUGUST, 9TH, 2022, B) SEPTEMBER,13TH, 2022, C) AUGUST, 28TH, 2023

Fingerprint a) Mail from August, 9th, 2022

	A-Record First party	CNAME First Party	A-Record Third Party	CNAME Third Party
Email	$3_{PL,T,DT5;PL,T,DT3,M}$, $3_{PL,N,DT5;PL,N,DT3,M}$	0	0	$3_{PL,T,DT5;PL,T,DT3,M}$, $3_{PL,N,DT5;PL,N,DT3,M}$

Fingerprint b) Mail from September,13th, 2022

	A-Record First party	CNAME First Party	A-Record Third Party	CNAME Third Party
Email	0	0	$3_{PL,T,DT5;PL,T,DT3,M}$, $3_{PL,N,DT5;PL,N}$	$3_{PL,T,DT5;PL,T,DT3,M}$, $3_{PL,N,DT5;PL,N,DT3,M}$

Fingerprint c) Mail from August, 28th, 2023

Email	$5_{PL,T,DT5;PL,T,DT3,M}$, $5_{PL,N,DT5;PL,N,DT3,M}$	0	0	$5_{PL,T,DT5;PL,T,DT3,M}$, $5_{PL,N,DT5;PL,N,DT3,M}$
	$1_{PL,T,DT5;PL,T,DT3,M}$, $1_{PL,N,DT5;PL,N,DT3,M}$	0	0	$1_{PL,T,DT5;UNC,T,DT3,M}$, $1_{PL,N,DT5;UNC,N,DT3,M}$

 ISBN: 978-1-68558-092-6