BBND - BigBrotherNewsletterDetektor

Schlussbericht

Zuwendungsempfänger:

BBND Meder, Rutz, Rüther, Halbmann GbR

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS22S38 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Auch im Jahre 5 nach DSGVO enthalten noch viele Newsletter personalisiertes Tracking, d.h. es wird ein persönliches Profil angelegt, mit genauen Angaben, wer wann welche Mail geöffnet, wann welchen Link geklickt hat. Dieses persönliche Profil bleibt auf Dauer gespeichert und gelangt schlimmstenfalls mittels Google Analytics in die Hände von Google und NSA.

Selbst in der sonst recht sensibilisierten Nachhaltigkeits-Szene gibt es wenig Bewusstsein für das Thema, das aber durchaus kampagnenfähig ist. Viele NGOs nutzen Mailchimp mit der Standardeinstellung persönliches Öffnungs- und Link-Tracking. Nutzende sind diese Tracking-Links gewohnt, und selbst bei sonst konzernkritischen Menschen gibt es wenig Sensibilisierung. For-Profits - selbst einzelne, die mit "Datenschutz" werben, haben bei der geringen Sensibilisierung ihrer Kunden wenig Anreiz, auf personalisiertes Tracking zu verzichten.

Mit dem BigBrother-Newsletter-Detektor sollte die Sensibilisierung für personalisiertes Tracking sowohl bei Verbrauchern als auch bei Anbietern erhöht werden.

Ziel war, zunächst eine Analyse-Library zu entwickeln, die in der Lage ist, typische Muster der gängigsten Newsletter-Software zu identifizieren und kategorisieren. Ergebnisse sollten datenschutzkonform gespeichert werden können und weitergehender Forschung als Open Data zur Verfügung stehen. Ein nutzerfreundliches Frontend sollte die einfache Anwendung ermöglichen, und ergänzende Informationen zum Thema Tracking und Überwachung sollte in Zusammenarbeit mit einem Content-Partner erstellt und bereitgestellt werden.

Beitrag des Projektes zu den Zielen der Förderinitative „Software-Sprint“

Das Projekt ermächtigt Nutzende, selbst aktiv zu werden und der allgegenwärtigen Überwachung ein kleines persönliches Ausrufezeichen entgegen zu stellen und führt so zu einer Wahrnehmung der eigenen

Selbstwirksamkeit. Gleichzeitig klärt das Projekt allgemein zu den Themen Tracking und Datenschutz auf, und trägt damit zu einem bewussteren Umgang mit den eigenen personenbezogenen Daten bei.

Ausführliche Darstellung der Ergebnisse

Software Ergebnisse

Auf der Software Ebene alle Meilensteine erreicht und über-erfüllt. So wurde neben der angestrebten Newsletter-Analyse eine Ergebnis-Datenbank sowie ein kompletter Feedback-Ablauf für die Newsletter-Verantwortlichen umgesetzt. Folgende Ergebnisse wurden dabei erzielt:

Analyzer Komponente

Ein Prototyp des Analyzer Pakets wurde erstellt und auf gitlab veröffentlicht (https://gitlab.com/unspy.email/untrack-email-analyzer). Die Analyzer Komponente nimmt als Eingangsdaten eine E-Mail inklusive Header entgegen, und gibt eine vollständige Analyse aus.

Der Analyzer nimmt folgende Analysen vor:

- Prüfung der DKIM Signatur
- Extrahieren aller Header
- Extrahieren aller Link URLs
- Extrahieren aller Bild URLs
- Extrahieren mutmaßlicher Unsubscribe URLs (Whitelist + Heuristik)
- Online-Abfrage aller CName Domain Aliase der verwendeten URLs
- Check aller URLs auf bekannte Analytics Muster
- Check aller URLs auf gegen Muster einer Provider Datenbank
- Online-Abfrage aller Nicht-Unsubscribe URLs auf Weiterleitungen (Asynchron, um Zeit zu sparen, gleichzeitig gedrosselt, um keine Provider-Sperre auszulösen)
- Extrahieren der Mailinglisten-Identität (Heuristik)
- Extrahieren eines "Verdikts" aus allen bisherigen Analyse-Ergebnissen (Enthält die Mail eine bekannte oder mutmaßliche Spionage-Komponente)

Die Analyse-Ergebnisse werden in einer Datenstruktur übergeben, die einerseits allgemein genug bleibt, um als Ausgangspunkt weiterer Analysen zu dienen, andererseits so strukturiert ist, dass eine Bildschirm-Ausgabe mit überschaubarer Logik einfach möglich ist.

Die Muster Datenbank enthält folgende Provider:

- Mailchimp
- Inxmail
- Cleverreach

Diese Provider decken geschätzt 30-60% der Newsletter von Vereinen und KMU ab.

Storage Komponente

Ein Prototyp des Storage Pakets wurde erstellt und auf gitlab veröffentlicht (https://gitlab.com/unspy.email/untrack-email-analyzer). Die Storage Komponente ist für Speicherung und das komplette Web-Frontend zuständig:

- Hochladen des E-Mail Quelltexts in ein Formular
- Übergabe an den Analyzer
- Einholen des Einverständnisses zur Verarbeitung
- Speichern der Roh-Analysen (enthält alle URLs)
- Konvertieren und Speichern von Archiv-Analysen (alle URLS u.a. anonymisiert)
- Löschen der Roh-Analysen nach 90 Tagen
- Zuordnen der Analysen zu Newsletter-Listen und Domains
- Versand von Informations-Mails an die Newsletters-Verantwortlichen, mit einer geheimen Feedback-URL
- Entgegennahme von Feedback der Newsletters-Verantwortlichen in einem strukturierten Formular
- Darstellen einer Übersichtsliste aller Newsletter nach Verantwortlichen, sowie
    - Status der jeweils letzten Analyse
    - eventuelles Feedback der Newsletter-Verantwortlichen

Als Feedback wird abgefragt:

- How did your organization decide whether to leverage user tracking in your newsletter emails? (Wie hat ihre Organisation entschieden, ob sie User Tracking in ihren Newslettern einsetzen wollen?)
- Campaign tracking provides a privacy-friedly alternative to user tracking. Have you been aware of that option? (Kampagnen-Tracking bietet eine datensparsame Alternative zum User Tracking. War ihnen das bewusst?)
- How do you plan to proceed? WIll you de-activate user tracking? (Wie werden sie weiter vorgehen? Werden sie User Tracking deaktivieren?)
- If no: Will you utilize user tracking data? (Falls nein: werden sie User Tracking Daten auswerten?)
- If yes: Why is campaign tracking not enough? How exactly will you analyze user tracking data? (Falls ja: warum ist Kampagnentracking nicht ausreichend? Wie genau werten sie User Tracking Daten aus?)

Demo Installation

Als Arbeitsinstanz wurde eine Demo Installation erstellt (qa.unspy.email, u.a. wegen fehlender Rechtstexte noch hinter Passwortschutz).

Weitere Beiträge

Im Projekt wurde bei verwendeten Open Source Komponenten "upstream" und rund einem Dutzend Stellen Verbesserungen oder Problem-Meldungen rück-gespeist.

Zu folgenden Komponenten wurden wesentliche Weiterentwicklungen beigesteuert:

- DCache (https://www.drupal.org/project/dcache) - ein horizontal skalierbares Caching API
- DomainParser (https://www.drupal.org/project/domainparser) - extrahiert den registrierbaren Teil einer Domain.
- LiveFilter (https://www.drupal.org/project/livefilter) - ermöglicht schnelles Volltextfiltern beliebiger Tabellen.
- ViewsFieldEntityLabel (https://www.drupal.org/project/views_field_entity_label) - Entity Label als Drupal Views Field.
- WebformPreset (https://www.drupal.org/project/webform_preset) - stellt geheime URLs für authentifizierte Webform Eingaben bereit.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Open-Source Komponenten

Sicherheitsforscher können die o.g. Komponenten direkt nutzen, und über Gitlab zusätzliche Provider-Muster beisteuern. Relevante Teile der Analyzer und Storage Komponenten können direkt in andere Projekte kopiert werden, oder es kann aus der Art der Problemlösung gelernt werden.

Entwicklung des unspy.email Portals mit Partnern

Mit den Projektpartnern (DigitalCourage, Digitale Gesellschaft angefragt) wird in den kommenden Monaten aus dem Prototyp das unspy.email Portal feingeschliffen:

- Alle Abläufe werden praxiserprobt
- UI-Texte, Bildmotive, Metaphern und Framings werden feingeschliffen
- Rechtstexte werden geklärt und eingefügt.

Als Faktoren für eine gute Erfolgsprognose sehen wir, dass die DigitalCourage Web-Admins bereits Erfahrungen mit dem Drupal Framework mitbringen, "richtig Lust" darauf haben, und mit Jörg Heidrich starke anwaltliche Fachkompetenz mit im Boot sitzt.

Fachliche Take-Aways

Zu folgenden Technologien wurden das Verständnis wesentlich vertieft:

- PHP8 mit seinen "cheap immutable objects"
- Funktionale Collections (loophp/collection)
- DKIM Signaturen
- IMAP Protokoll
- Drupals experimentelle Single Directory Components (SDC)
- Async in PHP mit Promises, Fibers, und Swoole

Wie kann's weiter gehen?

In dem Projekt wurde nicht nur der Problemraum erkundet und verstanden, es entstanden auch potente Ansätze für Folge-Projekte:

E-Mail Einlieferung per Weiterleitung

Für die Analyse von E-Mails wird deren Quelltext inklusive Headern benötigt. Eine komfortable Einlieferung dieses Quelltexts war von Anfang an als Knackpunkt definiert. Die aktuelle Einlieferung per Upload der abgespeicherten E-Mail ist für den Prototyp pragmatisch, stellt aber u.e. für eine Live Version eine zu hohe Hürde dar.

Einlieferung per Browser und oder Mailclient Plugin war angedacht, hätte aber den Rahmen des Projekts gesprengt. Solche Plugins müssten allerdings für jeden Client entwickelt und gepflegt werden, was uns nach einer wirtschaftlicheren Lösung Ausschau halten hat lassen.

Eine Einlieferung von Mails mit Headern als .eml Anhang an eine definierte Mail-Adresse funktioniert mit den meisten E-Mail Clients. Das Umsetzen über das IMAP Protokoll ist nicht ganz unaufwendig, bietet aber eine universelle und niedrigschwellige Lösung, und hat immense Synergien mit anderen Ideen (s.u..). Diese Schnittstelle wollen wir in einem Folgeprojekt entwickeln.

"E-Mail Cleaner Service"

Bei der Arbeit am Weiterleitungs-Part des Analyzers hatten wir eine wesentlich weitreichendere Idee: Spionage-Links funktionieren über Weiterleitungen. Wenn wir mit einem Service diese Weiterleitungen auflösen, ist das Ergebnis ein vollständig spionagefreier Newsletter. Beim Tracking-Service erscheint es so, dass jeder Newsletter sofort geöffnet und alle Links geklickt werden.

Dabei sind ein paar Feinheiten zu beachten: Die Umsetzung in einem Mail-Client-Plugin würde als Öffnungstracker funktionieren, muss also verworfen werden. Eine Umsetzung muss also zwingend auf Ebene des Mail-Servers erfolgen. Den Cleaner auf dem Mail-Server installieren zu müssen, verspricht keinen Erfolg. Folgendes dagegen funktioniert für die meisten Mailkonto-Anbieter: "Wenn Newsletter X, leite weiter an cleaner@unspy.email". Der Service reinigt dann die E-Mail und sendet das Ergebnis wieder zurück.

Dieser Cleaner Service würde den "Detektor" auf gesellschaftlicher Ebene um ein Element des Empowerments ergänzen: Statt nur (ggf. mit gesellschaftlichem oder rechtlichem Nachdruck) um das Weglassen von Spionage-Links zu bitten, würde ein Werkzeug für alle geschaffen, das Spionage-Links direkt entfernt.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Für die Async Auflösung von Weiterleitungen wurden mehrere Ansätze umgesetzt, bis ein befriedigender gefunden wurde.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Auf der künftigen Projekt-Domain www.unspy.email werden die Quelltext Repositories bereits verlinkt.

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Die Stabilisierung des Backends verbrauchte mehr Zeit und Aufwand, in Folge wurde der Feinschliff des Frontends mit wesentlich geringerer Tiefe gemacht.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Für die Storage Komponente wurde die gerade in Entwicklung befindliche Drupal Single Directory Components (SDC) Technologie genutzt. In der anfänglichen Version hatte SDC eine unserer Anforderungen nicht erfüllt. Durch Upstream Kommunikation konnten wir unsere Anforderungen einbringen, SDC wurde entsprechend erweitert, so dass wir es nun gut nutzen können.