Technische Hochschule Ingolstadt

Seminar zu Themen der Informatik

Sommersemester 2022

Seminararbeit

Analyse von PST-Dateien

von

Alexander Pilz

1 Einleitung

E-Mails sind im privaten wie auch im geschäftlichen Umfeld nicht mehr als Kommunikationsmedium wegzudenken. Dies ist klar erkennbar an den versandten und empfangenen E-Mails weltweit. So lag diese Zahl im Jahr 2021 bei 319,6 Milliarden E-Mails und steigt laut einer Prognose bis ins Jahr 2025 auf 376,4 Milliarden E-Mails [2].

E-Mails werden heutzutage aber nicht nur für die Kommunikation verwendet, sondern auch zum Versenden von Spamnachrichten. So beläuft sich der Anteil an Spam-Mails am weltweiten Anteil der versandten E-Mails auf 46 Prozent [1]. Unter Spam versteht man Nachrichten, die ohne Aufforderung und unterwünscht zugestellt werden. Meist haben diese Nachrichten den Zweck Werbung zu verbreiten. Jedoch gibt es auch weitaus bedenklichere Nachrichten. So wird mithilfe von Spam-Mails versucht, unvorsichtige Nutzer dazu zu bringen, persönliche Daten preiszugeben oder finanzielle Gewinne zu erzielen.

Im geschäftlichen Umfeld wird hier dem Nutzer meist die Arbeit abgenommen, weil professionelle Spam-Filter eingerichtet sind und so fast keine unerwünschten Nachrichten mehr ankommen. Im privaten Umfeld werden solche Spamfilter oft nicht bzw. nicht richtig angewandt. Hinzu kommt, dass sich viele Nutzer mit dieser Thematik gar nicht auseinandersetzen. Dies führt häufig dazu, dass die Postfächer der Nutzer hier regelrecht durch Spam-Mails überladen werden.

Im Rahmen dieser Seminararbeit sollen diese unerwünschten E-Mails mithilfe des .PST-Dateiformats analysiert und die Ergebnisse dokumentiert werden.

2 Allgemeines

3 PST-Datei

Das .pst-Dateiformat ist ein proprietäres Dateiformat, welches von Microsoft im Mailprogramm Öutlook"verwendet wird. PST steht dabei für Personal Storage Table. Microsoft nutzt das Dateiformat zum Speichern von Nachrichtenkopien, Kalendereinträgen und Kontakten. Nutzt man den Microsoft Exchange Server werden die Daten an den Server übermittelt und dort gespeichert. Im Gegensatz dazu speichert Microsoft Outlook ohne Exchange Server diese Elemente auf dem lokalen Computer. Dabei

werden .pst-Dateien meist zum Speichern archivierter Elemente verwendet.

4 Aufbau einer PST-Datei

5 Sicherheitsprobleme mit PST-Dateien

6 Schritte bei der Analyse der PST-Datei

In diesem Kapitel werden die durchgeführten Analyseschritte näher erläutert. Dabei wird auf eine Senderanalyse, eine zeitliche Analyse sowie eine Analyse bestimmter Schlagwörter eingegangen.

6.1 Datensatz

Zuerst wird näher auf den verwendeten Datensatz für die Analyse eingegangen. Dabei war die Auswahl eines geeigneten Datensatzes schwieriger als anfangs erwartet, da der Datensatz auch Kriterien für die Analyse erfüllen sollte. Der Datensatz sollte eine relativ hohe Anzahl an E-Mails enthalten, es sollte ein schwacher bzw. kein Spam-Filter vorhanden sein und der Datensatz sollte in ein .pst-Dateiformat überführt werden können. Eine Online-Suche nach geeigneten Datensätzen ergab eine Anzahl an Treffern, die E-Mail-Postfächer simulieren, diese eigneten sich aufgrund ihrer Formate jedoch nicht für die Weiterverarbeitung die in dieser Seminararbeit geplant war.

Aufgrund dessen wurde im privaten Umfeld nach verfügbaren Postfächern gesucht, die ich für die Analyse verwenden durfte. Dabei wurde mir ein E-Mail Konto vom Anbieter "Web.deßur Verfügung gestellt, welches 4102 E-Mails enthält. Dieses Konto wurde dann in Microsoft Outlook eingebunden um mithilfe des Mail-Programmes einen .pst-Export erstellen zu lassen, der dann als Grundlage für die Analyse verwendet wurde.

6.2 Parsing der PST-Datei

Für die Analyse wurde die Python Bibliothek "libpffin Version 20211114 verwendet. Diese Bibliothek wurde speziell für den Zugriff auf das Personal Folder File (PFF) und das Offline Folder File (OFF) entwickelt. Diese Formate werden von Microsoft Outlook verwendet, um E-Mail, Kontakte und andere Daten zu speichern. PFF und OFF werden dabei in mehreren Dateitypen verwendet, unter anderem auch bei PST.

6.3 Senderanalyse

Eine Analyse der Sender wurde durchgeführt, um festzustellen, wie viele verschiedene Absender einer E-Mail an das Analysierte Postfach gesendet haben und welche davon am häufigsten vorhanden waren.

6.4 Zeitliche Analyse

Sobald der Ransomware-Client erfolgreich im Fahrzeug installiert wurde, wird die eigentliche Geiselnahme der Ransomware durchgeführt. Unter Geisel ist hierbei eine gesperrte Komponente, die nicht leicht wiederhergestellt werden kann oder bei der eine lange Ausfalldauer nicht vertretbar ist, die Beschlagnahme oder angedrohte Veröffentlichung von internen Daten, durch die ein erheblicher Schaden verursacht

wird oder etwas anderes zu verstehen, um das Opfer zur Zahlung des Lösegeldes zu zwingen. Sobald ein Angreifer Zugriff auf das System hat, kann er nach Belieben vorgehen, um das Opfer zu schädigen und somit eine Lösegeldzahlung zu erreichen. Beispiele für Angriffe sind unter anderem das Blockieren einer Türverriegelung, Sperren wichtiger kryptografischer Zugangsdaten wie dem Fahrzeugschlüssel, die Verschlüsselung kritischer fahrzeuginterner Daten, Einschüchterung des Opfers mithilfe gefälschter kritischer technischer Zustände wie einer überhitzten Batterie, oder reale physische Manipulationen wie dem Auslösen aller Airbags oder Abschalten des Motors. Die Möglichkeiten sind praktisch unbegrenzt.

Nach erfolgter Geiselnahme wird dies auch für das Opfer sichtbar. Dabei wird klar und deutlich erkennbar, was passiert ist und was die notwendigen Schritte sind, um wieder Zugriff auf sein System zu erhalten. Zur Veranschaulichung wie es aussehen könnte, wenn ein System übernommen wurde, haben die Autoren des dieser Arbeit zugrundeliegenden Artikels einen Demonstrator erstellt, der auf Abbildung zu sehen ist.

6.5 Analyse von Spam-Wörtern

In diesem Kapitel wird auf die Analyse bestimmter Schlagwörter in den E-Mails eingegangen.

7 Ergebnisse

In diesem Kapitel werden die Ergebnisse der Analyse näher erläutert.

8 Fazit

In dieser Seminararbeit wurde eine forensische Analyse einer .pst-Datei durchgeführt. Hierfür wurden verschiedene Analysen durchgeführt. Die Analysen wurde mithilfe von eigens geschriebenen Python-Skripten durchgeführt, da Tools für die Analyse von .pst-Dateien nur kostenpflichtig zur Verfügung standen bzw. bei den kostenfreien Versionen nur sehr eingeschränkte Funktionalitäten besessen haben. Durch die Verwendung professioneller Tools wären mit großer Wahrscheinlichkeit noch viel detailliertere Ergebnisse zustande gekommen.

Zum Parsing der .pst-Datei wurde die Python Bibliothek ”libpff”verwendet, die sich speziell für den Zugriff auf Dateiformate von Microsoft Outlook eignet. Mithilfe dieser Bibliothek wurden dann die wichtigsten Eigenschaften der E-Mails extrahiert, um sie für die weitere Analyse aufzubereiten. Im Anschluss daran wurden dann eine Senderanalyse, eine zeitliche Analyse und eine Spamwortanalyse durchgeführt.

Die Senderanalyse hat gezeigt, dass Facebook mit 752 E-Mails auf Platz 1 der meisten gesendeten E-Mails liegt. Dies liegt höchstwahrscheinlich daran, dass der Besitzer des Postfaches seine Facebook Benachrichtigungen aktiviert hat. Auf dem zweiten Platz liegt jedoch ”Lidl Insider”mit 250 E-Mails. Dieses Ergebnis ist erschreckend, da ich nach einer Rücksprache mit dem Besitzer des E-Mail Kontos erfahren habe, dass genau eine Bestellung vor einigen Jahren durchgeführt wurde und seitdem zahlreiche E-Mails mit Werbung empfangen werden.

Die Analyse der Empfangszeiten der E-Mails hat gezeigt, dass deutliche Muster von Häufungen der empfangenen E-Mails zu bestimmten Zeiträumen erkennbar sind. In dem verwendeten Datensatz war klar erkennbar, dass sich eine Häufung zwischen 16 und 18 Uhr abzeichnet. Dies ist für viele Personen die Zeit um Feierabend zu machen. Das ist insofern eine gute Zeit, da viele Leute dann erschöpft und somit unvorsichtiger sind und leichter auf eine Spam-Mail hereinfallen.

Bei der Analyse der auftretenden Spam-Wörter in den E-Mails wurde eine mithilfe häufig verwendeter Wörter eine SSpam-Wort-Listeërstellt. Die Inhalte der E-Mails wurden dann mit der Liste überprüft um

zu sehen wie häufig bestimmte Wörter auftreten. Dabei war Ängebot"mit 5092 Treffern klar auf Platz 1 und "Date"mit 2991 Treffern auf Platz 2. Somit kam das Wort Angebot im Schnitt in 1,2 E-Mails vor.

Zu erwähnen ist jedoch, dass diese Ergebnisse nicht repräsentativ sind. Um bessere Ergebnisse zu erzielen bräuchte man zum einen mehr E-Mails und zum Anderen mehrere E-Mail Konten verschiedener Benutzer um andere Verhaltensmuster im zu analysierenden Datensatz vorzufinden.

Literaturverzeichnis

[1] A-SIT Zentrum für sichere Informationstechnologie – Austria. Onlinesicherheit - spam, 26.06.2022.

[2] Statista. Anzahl der e-mails pro tag weltweit 2025, 26.06.2022.