OSINT: Sammeln öffentlich verfügbarer Information

Das Internet erleichtert es Angreifern, Informationen über ihre Ziele und deren Schwachstellen zu sammeln und Angriffe darauf vorzubereiten. Verteidiger können jedoch dieselben Mittel und Werkzeuge nutzen, um Angriffswege zu versperren.

Von Frank Ully

Open Source Intelligence, kurz OSINT, beschreibt das Gewinnen von Erkenntnissen, indem man öffentlich verfügbare Informationen sammelt und auswertet. Der Begriff stammt aus dem Umfeld der US-Geheimdienste; früheste Ansätze dieser Praktik sind vom Ende des Zweiten Weltkriegs dokumentiert. Zunächst waren Printmedien, Radio und Fernsehen die öffentlichen Quellen – mit dem Aufkommen des Internets erlebte OSINT einen neuen Boom.

Neben Nachrichtendiensten und dem Militär bedienen sich heute Journalisten, Strafverfolger, Detektive, Ahnenforscher und Geschäftsleute dieser Art der Informationsbeschaffung.

Auch für Cyberkriminelle, die eine Organisation kompromittieren wollen, offenbart sich ein großer Datenschatz. Informationssammlung (Information Gathering oder Reconaissance) ist ein eigener, zeitaufwendiger Schritt am Beginn eines Angriffs [1]. Sie hilft wesentlich beim Social Engineering [2] und Hacker verwenden darauf viel Zeit in späteren Phasen, etwa beim Auslesen von Informationen aus dem Verzeichnisdienst Active Directory [3]. Je besser Angreifer ihr Ziel kennen und verstehen, desto einfacher können sie es attackieren.

Auf viele Arten nützlich

Admins und IT-Sicherheitsverantwortliche können selbst interessante Informationen über die eigene Organisation finden und dadurch die Angriffsfläche ermitteln, einschätzen und verringern. Verteidigern hilft OSINT zudem beim Untersuchen von Sicherheitsvorfällen, bei der Malware-Analyse und als sogenannte Threat Intelligence, die aktuelle Bedrohungslage zu beobachten, um sich zukünftig besser vor Schadsoftware zu schützen [4].

Informationssammlung findet passiv oder aktiv statt. Beim passiven Sammeln interagiert der echte oder simulierte Angreifer nicht direkt mit einer Zielperson oder dem System des Opfers. Bei aktiver Informationssammlung schon, indem er Ports scannt oder versucht, Domänennamen oder Verzeichnisse per Brute Force zu erraten. Die Scans in den vorherigen Artikeln dieses Tutorials haben aktiv Informationen gesammelt; im Kontext eines Pentests müssen die Systembetreiber dem zustimmen. Bei OSINT ermittelt man Informationen passiv aus öffentlichen Quellen oder, was die Definition etwas dehnt, aus prinzipiell jedermann nach Registrierung oder Zahlung zugänglichen Datenbanken. Eine strenge Auslegung zählt das Nutzen von Quellen, die nicht frei verfügbar sind, nicht zu OSINT.

Der erste Artikel dieser Reihe beschrieb, wie Angreifer aus Datenlecks zusammengesammelte Benutzerkennungen, meist E-Mail-Adressen, samt zugehörigem Klartext beim kostenpflichtigen Onlinedienst dehashed.com abfragen. Teil 2 zu Webanwendungen zeigte, wie die Browsererweiterung BuiltWith verwendete Webtechnologien ermittelt. BuiltWith bietet auch eine Webdatenbank. (Alle im Text genannten Artikel, Tools, Dienste, Webseiten et cetera sind über ix.de/zgp7 zu finden.) Ein ähnlicher Onlinedienst, der viele Informationen über eine Website zusammenfasst, ist der Netcraft-Site-Report (Abbildung 1). Neben Basisdaten wie hinterlegten Registrierungsinformationen und IP-Adressen zeigt er verwendete Zertifikate, genutzte Webtechnologien sowie eingebettete Tracker an und enthält eine Hosting-Historie.

Dieser Artikel beschreibt, wie Verteidiger dieselben Websites und Tools verwenden wie echte und simulierte Angrei-

iX-TRACT

- Öffentlich erreichbare Dienste muss man gut schützen – ein Konfigurationsfehler reicht aus, um Angreifern Tür und Tor zu öffnen. Sie suchen in öffentlich verfügbaren Quellen nach Angriffspunkten.
- Verteidiger können selbst interessante Informationen über die eigene Organisation finden und die dadurch offenbarte Angriffsfläche ermitteln, einschätzen und verringern.
- Verteidiger und Angreifer nutzen dieselben frei verfügbaren Websites und Tools, um verwendete IP-Adressen und Domänen aufzuspüren und darauf laufende Dienste auf mögliche Schwachstellen zu untersuchen.

fer, um unter anderem verwendete IP-Adressen und Domänen aufzuspüren und darauf laufende Dienste auf mögliche Schwachstellen zu untersuchen.

Erster Schritt: IP-Adressen

Einer der ersten Ansatzpunkte von Angreifern ist ein Blick in die RIPE-Datenbank. RIPE ist die Regional Internet Registry für Europa, den Nahen Osten sowie Zentralasien. Sie verwaltet einen Teil der weltweit vergebenen IPv4- und IPv6-Adressen. Auf der Seite „RIPE Database Text Search“ gibt man den Namen des eigenen Unternehmens ein, beispielsweise Heise, und wählt als Suchergebnis den Typ inetnum.

Die Suchergebnisse in Abbildung 2 zeigen alle IPv4-Adress-Registrierungen, in denen dieser Begriff vorkommt, auch in Adressbestandteilen wie der Straße. Für Heise gab es zum Zeitpunkt der Entstehung dieses Artikels 83 Einträge auf mehreren Seiten. Neben ignorierten, weil irrelevanten Ergebnissen notiert man sich beispielsweise die IP-Adressblöcke 193.99.145.0-255 und, folgt man dem dritten Link, 192.109.227.0-255 für Heise Medien sowie 193.99.144.0-255 für den Zeitschriftenverlag.

Es lohnt, bei allen Ergebnissen dem grün hinterlegten Link zu folgen und sich die Details anzusehen, um einzuschätzen, ob das Ergebnis zur gesuchten Organisation passt. Auf diese Weise stellt man selbst bei kleineren Betrieben mit wenigen Tochtergesellschaften rasch einige mutmaßlich verwendete IPv4-Adressen zusammen. Diese Liste ist ein guter Start, aber nicht unbedingt vollständig: Natürlich kann ein Unternehmen Systeme in einem Rechenzentrum betreiben oder Adressen verwenden, die auf den Namen eines Internetproviders registriert sind.

Nächste Anlaufstelle ist das BGP-Toolkit-Portal von Hurricane Electric.

Tutorialinhalt

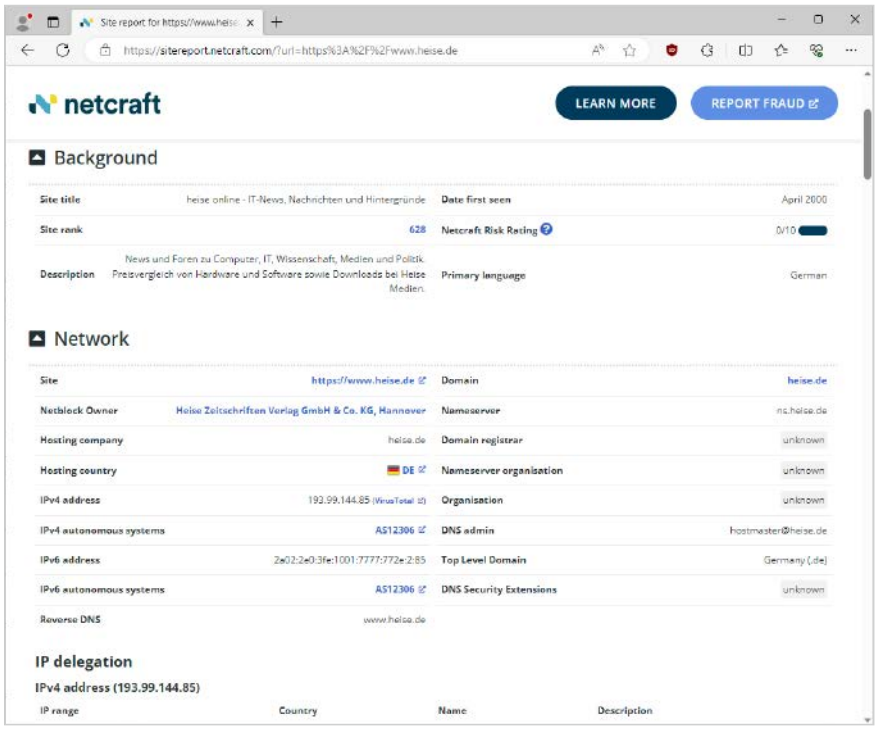
Teil 1: Scannen und Verifizieren der eigenen Systeme

Teil 2: Webapplikationen angreifen

Teil 3: Auditieren interner Netzwerke, Domänen und Systeme

Teil 4: Sammeln öffentlich verfügbarer Informationen und Analysieren der Angriffsfläche

Teil 5: Überprüfen von Cloud-Umgebungen



Der Netcraft-Site-Report zeigt beispielsweise für heise online Informationen über IP-Adressen, verwendete Adressblöcke, Zertifikate und eine Hosting-Historie an (Abb. 1).

BGP steht für Border Gateway Protocol, das Routingprotokoll des Internets. In das Suchfeld gibt man ebenfalls den Organisationsnamen ein. Das Ergebnis zeigt die schon aus der RIPE-Datenbank bekannten Adressblöcke. Bei größeren Organisationen stehen dort deren autonome Systeme (AS) mit ihrer Nummer (ASN). Diese Nummer ist eindeutig und öffentlich zugänglich, um Routinginformationen mit anderen Systemen auszutauschen. Autonome Systeme sind Teile des Internets, die große Organisationen verwalten. Findet man AS-Einträge im BGP-Portal, folgt man ihrem Link und sieht in der Detailansicht in den Registerkarten „Prefixes v4“ und „Prefixes v6“ jeweils darüber verteilte Adressblöcke.

Liefert keine der beiden Quellen Ergebnisse für den eigenen Betrieb, ist das kein Problem. Es gibt weitere Techniken.

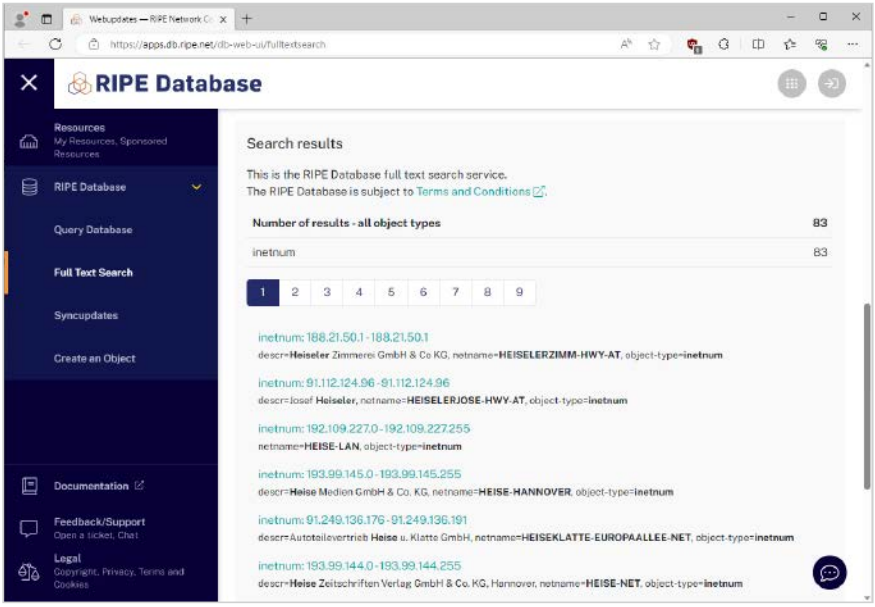
Eine Ebene darunter: Subdomänen

Als Nächstes gilt es, eine Liste von Subdomänen zusammenzustellen, auf denen das Unternehmen Dienste betreibt. Zum Beispiel ist www.heise.de eine Subdomäne der Domäne heise.de. Eine Subdomäne kann für jeden beliebigen Zweck erstellt werden; die Anzahl an Subdomänen für eine Domäne ist nicht begrenzt.

Erste Anlaufstelle ist DNSdumpster (Abbildung 3), das einen spannenden Einblick in das Domain Name System (DNS) erlaubt. Eine naheliegende Eingabe dort ist eine bekannte Domäne, auf der die Hauptwebsite der Organisation läuft. Tippt man etwa heise.de ein, steht zuoberst eine Liste verwendeter DNS- und E-Mail-Server (MX Records, das steht für Mail Exchange). Mit den MX-Datensätzen kann ein Angreifer die verwendeten Spamfilter anhand der IP-Adresse und der ASN ermitteln, die mit dem E-Mail-Host verbunden sind. Im Abschnitt TXT-Records ist im SPF-Eintrag (Sender Policy Framework) sichtbar, wenn eine Organisation Cloud-E-Mail-Dienste oder Mailinglistenanbieter nutzt. Als TXT-Einträge stehen ebenfalls Kennnummern für Cloud-Dienste wie Atlassian Online oder Microsoft Azure.

Der eigentlich interessante Teil der Seite ist der Abschnitt „Host Records“. DNSdumpster löst automatisch Domänennamen in IP-Adressen auf, zeigt an, von wem die IP-Adresse registriert ist – in den meisten Fällen ein Hosting- oder Internetprovider – und versucht, den geografischen Standort zu ermitteln. Außerdem verzeichnet DNSdumpster manchmal die auf dem Server laufenden Dienste, wenn es diese identifizieren kann.

Die Liste der Hosts ist eine gute Ausgangsbasis zum Bewerten der eigenen Angriffsfläche. Anhand der Hostnamen stechen Systeme beispielsweise zum Erraten von Passwörtern und für den internen Zugriff ins Auge: vpn, owa, adfs und Variationen von citrix, admin oder remote.



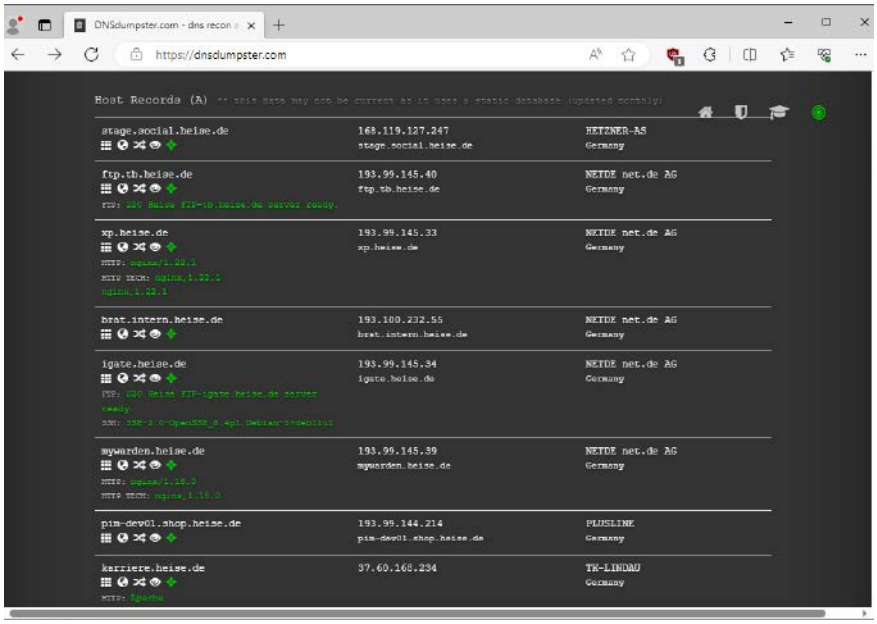
Die RIPE-Datenbank zeigt bei der Suche nach Objekten vom Typ inetnum IPv4-Adressblöcke an, in denen der gesuchte Begriff im Organisationsnamen oder an anderen Stellen vorkommt (Abb. 2).

Neben dem Durchführen von Passwort-Audits (siehe den ersten und dritten Tutorialteil) sollten Admins alle Anmeldeoberflächen mit einem zweiten Faktor schützen. Auch interne, Test- und Entwicklungssysteme (etwa intern, dev) sind interessant, weil dort womöglich veraltete Anwendungen laufen oder die Systeme schlechter gesichert sind als die Produktionsumgebung.

Beziehungen offenlegen

Unten auf der Seite lässt sich eine Karte herunterladen (zum Vergrößern anklicken), die die Beziehungen zwischen Domänen und IP-Adressen beziehungsweise Servern zeigt. Die Schaltfläche „View Graph“ ruft einen interaktiven Graphen auf, der nach dem autonomen System sortiert ist. Eine Excel-Tabelle der gefundenen Hosts kann man direkt exportieren, alternativ notiert man die gefundenen Subdomänen und IP-Adressen in eigene Textdateien.

Man sollte sich nicht nur auf eine einzelne Webquelle verlassen; ebenfalls zurate ziehen kann man den Subdomain Finder. Dort das Kontrollkästchen „Private scan“ aktivieren, um zu verhindern, dass die eigene Domäne in der Liste der letzten Scans auftaucht.



Der Webdienst DNSdumpster zeigt nach Eingabe einer Domäne unter anderem verwendete Namensserver, spezielle DNS-Einträge wie TXT, vorhandene Subdomänen und zugeordnete IP-Adressen an (Abb. 3).

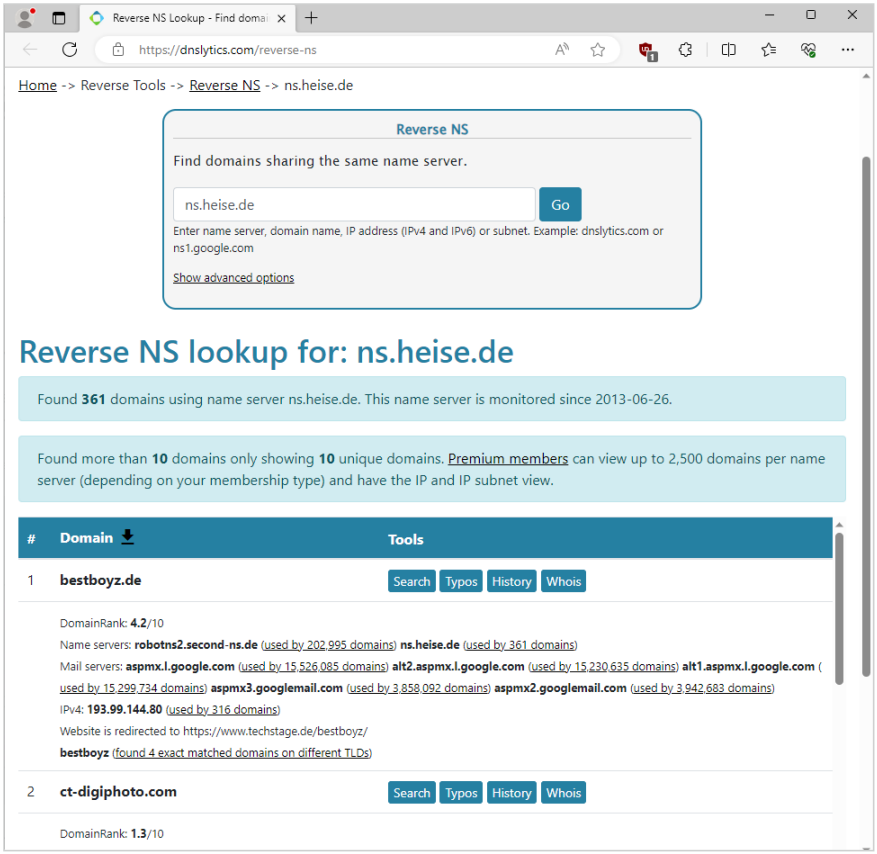
Weitere lohnende Dienste, die allerdings ein kostenfreies Konto erfordern, sind SecurityTrails und Phonebook.cz. Bei SecurityTrails gibt der Suchende direkt in das mittige Eingabefeld die Domäne ein. Auf der Ergebnisseite sieht er links die erfassten Subdomänen. Die ersten Ergebnisse sind ohne Anmeldung lesbar, die unscharfen Einträge werden erst danach aufgedeckt. Bei Phonebook.cz, betrieben vom Anbieter Intelligence X, muss man sich noch vor einer Suche nach Domains anmelden. Daneben kann man E-Mail-Adressen einer Organisation oder indexierte URLs nachschlagen. Nebenbei: Auf der Anbieter-Hauptseite intelx.io durchsucht man mit einem Konto den kompletten Datenbestand, der aus dem Darknet und öffentlich gewordenen Leaks gespeist wird. Die Ergebnisse dort sind allerdings geschwärzt und werden nur durch teure Lizenzen sichtbar.

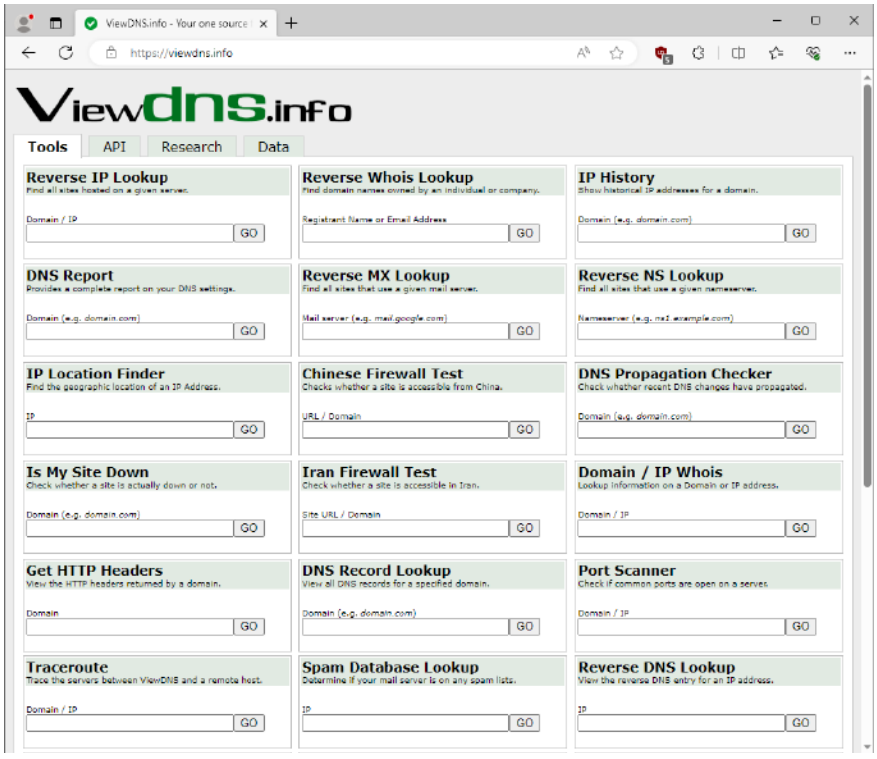
Woher wissen diese Datenbanken eigentlich, welche Subdomänen eine Organisation angelegt hat? Eine der Quellen sind Dienste zur Certificate Transparency. Dieser Standard sieht vor, alle von vertrauenswürdigen Zertifizierungsstellen ausgestellten öffentlichen Zertifikate revisionssicher zu protokollieren. So können fälschlicherweise oder mit böser Absicht ausgestellte Zertifikate entdeckt werden. Auf crt.sh lässt sich eine dieser Zertifikatsdatenbanken abfragen. Im Fall von Heise findet man dort Informationen über zahlreiche TLS-Zertifikate und die Subdomänen, für die sie ausgestellt wurden. Einen spannenden Ansatz, wie man mit Certificate Transparency über die Korrelation der Ausstellungszeit von Zertifikaten ähnliche Seiten findet, der den Rahmen dieses Artikels sprengen würde, beschreibt ein Blogbeitrag bei PT Security (siehe ix.de/zgp7).

Geteilte DNS-Server

Bereits jetzt hat man eine lange Liste von Subdomänen, die mutmaßlich zumindest in der Vergangenheit von der untersuchten Organisation verwendet wurden. Allerdings hat sie wahrscheinlich weitere Domänen wie ix.de registriert. Auch diese kann man ermitteln.

Die Rückwärtsabfrage von Namensservern über Reverse NS beim Dienst DNSlytics oder bei Hacker Target liefert Domänen, die vom selben Nameserver verwaltet werden und möglicherweise ebenfalls der untersuchten Organisation gehören (Abb. 4).





Das DNS-Multitool ViewDNS.info bietet ohne Anmeldung und kostenfrei zahlreiche Abfragen zu IP-Adressen und Domänen (Abb. 5).

Der erste Weg ist, gemeinsam verwendete DNS-Server aufzuspüren. Dazu dient beispielsweise die Funktion „Reverse NS“ des Anbieters DNSlytics. Gibt man dort heise.de ein und filtert die Suchergebnisse nach dem Nameserver ns.heise.de durch Klicken auf die Schaltfläche „View domains using this name server“, sieht man zehn weitere Domänen (Abbildung 4) von 361 Ergebnissen, darunter auch heise-shop.de. Nur die ersten zehn Ergebnisse sind kostenfrei einsehbar, für weitere Resultate benötigt man eine Premium-Mitgliedschaft, die mit 30 Euro für einen Monat recht preiswert ist. Im Beispiel wurde der Server ns.heise.de aus der Namensserver-Liste gewählt, weil er Heise zugeordnet werden kann. Server wie ns.plusline.de, die einem Hoster gehören, liefern wahrscheinlich viele Ergebnisse für andere Organisationen.

Ein alternativer Dienst für diese Rückwartsabfrage nach dem Namensserver, der alle erfassten Daten ohne Anmeldung ausspuckt, ist „Find Shared DNS Server“ von Hacker Target. Gibt man dort den zuvor ermittelten Nameserver ns.heise.de ein und löst das Captcha, erhält man eine lange Liste von Domänen, darunter ct.de und telepolis.de.

SecurityTrails, für das man in diesem Fall ein kostenfreies Konto benötigt, unterstützt die Funktion des „Reverse NS Lookup“, wenn man dort auf der Übersichtsseite für eine Domäne im Bereich „NS records“ auf einen der verlinkten Nameserver klickt.

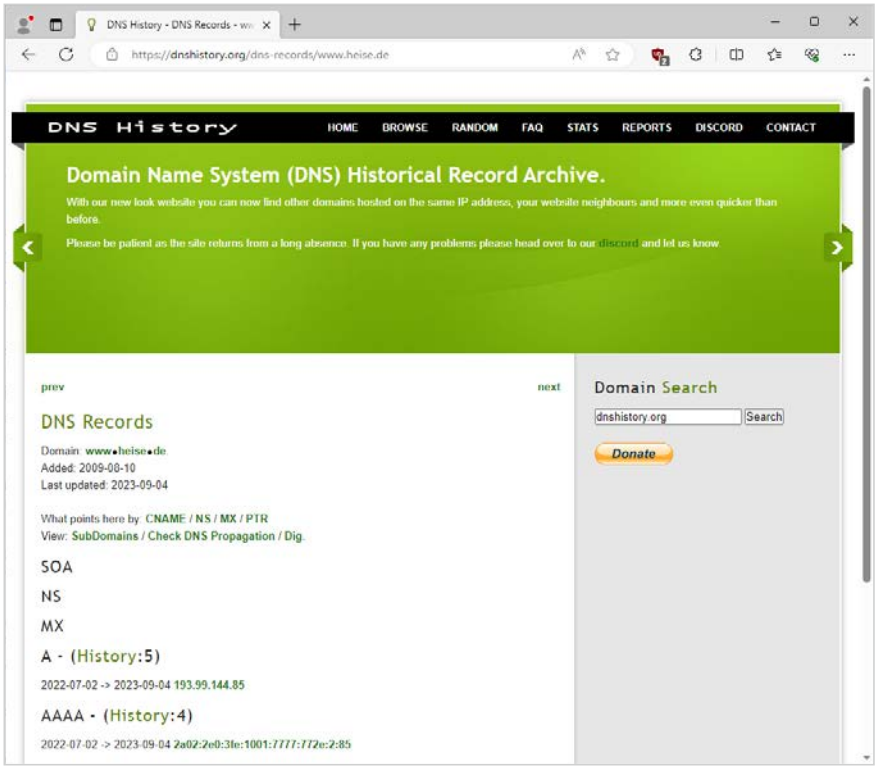
Zyklisches Vorgehen

Zu empfehlen ist, den gesamten OSINT-Prozess in Zyklen zu organisieren und die Suche mit den neuen Informationen für jeden Zyklus zu wiederholen. Für die ermittelten neuen Domänen sucht man wieder in Diensten wie DNSdumpster oder Phonebook.cz nach deren Subdomänen.

Dieser systematische Ansatz führt zu einem strukturierten Arbeitsablauf, der nachvollziehbare und recht vollständige Ergebnisse beschafft und auf Wunsch eine saubere Dokumentation ermöglicht.

Das bereits vorgestellte DNSlytics enthält weitere Reverse-Tools. Über die Funktion „Reverse IP“ findet man Domänen, die auf dieselbe IP-Adresse auflösen. Kostenfrei angezeigt werden die ersten hundert Ergebnisse, die sich stark mit der Nameserver-Rückwartsabfrage überschneiden. Dieselbe Funktion mit kostenfreier Anmeldung gibt es bei SecurityTrails, wenn man dort auf der Übersichtsseite für eine Domäne im Bereich „A records“ auf eine IP-Adresse klickt.

Ein Lesezeichen wert ist die Website ViewDNS.info (Abbildung 5). Dort lassen sich die vorgestellten und weitere Recherchen über DNS ohne Anmeldung



DNS History sammelt historische DNS-Daten, mit denen man IP-Adressen findet, die früher einer Domäne zugeordnet waren (Abb. 6).

durchführen. Alle Seiten auf einem Server wie gerade für andere Dienste beschrieben spürt die Funktion „Reverse IP Lookup“ auf.

IP-Adressen rückwärts und Whois-Daten durchsuchen

Whois ist ein Protokoll, das Datenbanken für Domänennamen, IP-Adressen oder autonome Systeme abfragt und unter anderem Informationen über eine Domäne wie Registrar und Eigentümer zurückliefert. Denn die ICANN (Internet Corporation of Assigned Names and Numbers) verlangt von zugelassenen Registrierstellen, dass sie unmittelbar nach Registrierung einer Domäne die Kontaktdaten des Inhabers, das Erstellungs- und Ablaufdatum der Domäne sowie weitere Informationen in die Whois-Datenbank eintragen. Einfach ausgedrückt ist die Whois-Datenbank eine durchsuchbare Liste aller weltweit registrierten Domänen. Im Zuge des gestiegenen Datenschutzes lassen sich in der Regel nicht mehr Personen, Anschrift oder Telefonnummer der Organisation auslesen.

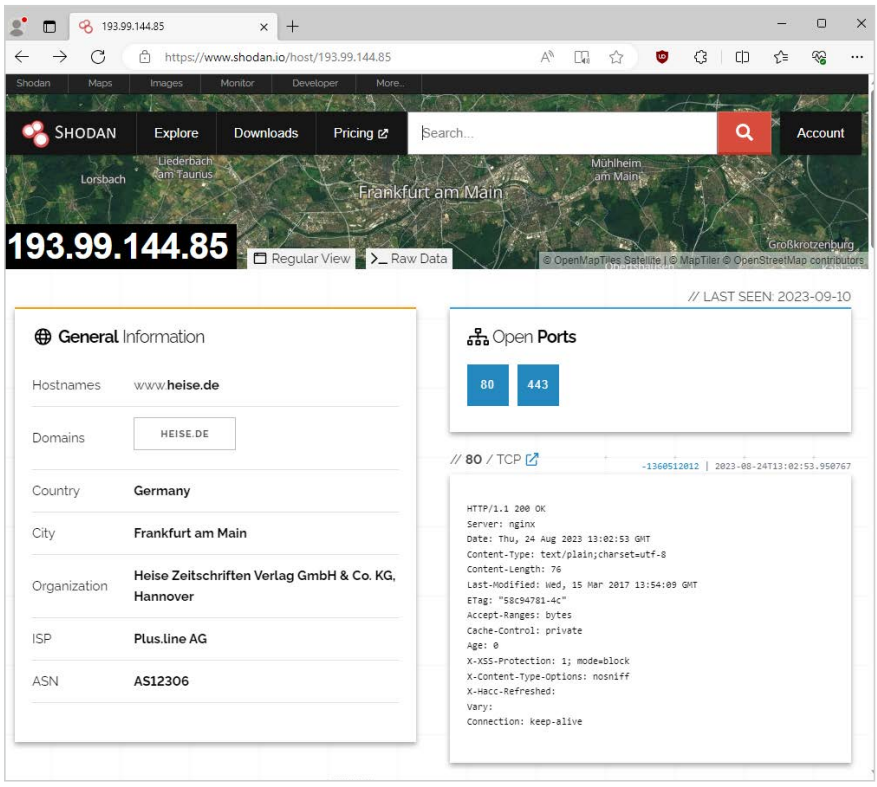
In der Vergangenheit haben zahlreiche Webdienste Whois-Daten indexiert. Dadurch kann man auf der Website Reverse Whois Lookup sowie auf ViewDNS im Eingabefeld „Reverse Whois Lookup“ den Namen einer Organisation eingeben und erhält eine Liste mit Domänen, die möglicherweise von ihr registriert wurden. Die Ergebnisse muss man mit Vorsicht genießen, da die Suche gleichlautende Namen berücksichtigt. Der wohl beste Dienst für Whois-Abfragen, Whoxy, unterstützt keine de-Domänen und zeigt kostenfrei nur begrenzte Ergebnisse. Gibt man auf Whoxy in das Eingabefeld „Whois Lookup“ microsoft.com ein und folgt auf der Ergebnisseite im Bereich „Registrant Contact“ dem Link neben dem Unternehmensnamen Microsoft Corporation, wird deutlich, wie mächtig dieser Suchansatz ist.

Historische DNS-Daten

Ein spannender Ansatz, weitere IP-Adressen zu finden, ist die Abfrage historischer DNS-Daten. Für Webpräsenzen wie heise.de ist meist die Suche nach Subdomänen wie www.heise.de sinnvoller. Tippt man das im kostenfreien Dienst DNS History in das Eingabefeld ein, sieht man auf der Ergebnisseite (Abbildung 6) neben dem A-Eintrag, dass das Archiv fünf IPv4-Adressen archiviert hat, und neben AAAA, dass vier IPv6-Adressen im Archiv stehen. Ein Klick auf den History-Link zeigt die Ergebnisse.

Ein anderer Dienst mit derselben Funktion ist PassiveDNS. Das bereits genannte DNS-Multitool ViewDNS bietet diese Suchfunktion unter IP History.

Auf diese Weise stellt man eine Liste mit IP-Adressen zusammen, unter de-



Die Internet-of-Things-Suchmaschine Shodan zeigt aus aktuellen, regelmäßigen Portscans offene Ports eines Systems und darauf bezogene Daten an (Abb. 7).

nen womöglich von der Organisation längst vergessene Dienste noch angeboten werden.

Will man schnell Allgemeines über eine IP-Adresse herausfinden, ist der Dienst IPinfo.io ideal. Er fasst Daten wie ungefähren Standort, ASN und Missbrauchsmeldekontakt übersichtlich zusammen. Außerdem verrät er, ob die Adresse zu einem VPN-Provider oder dem Tor-Netzwerk gehört.

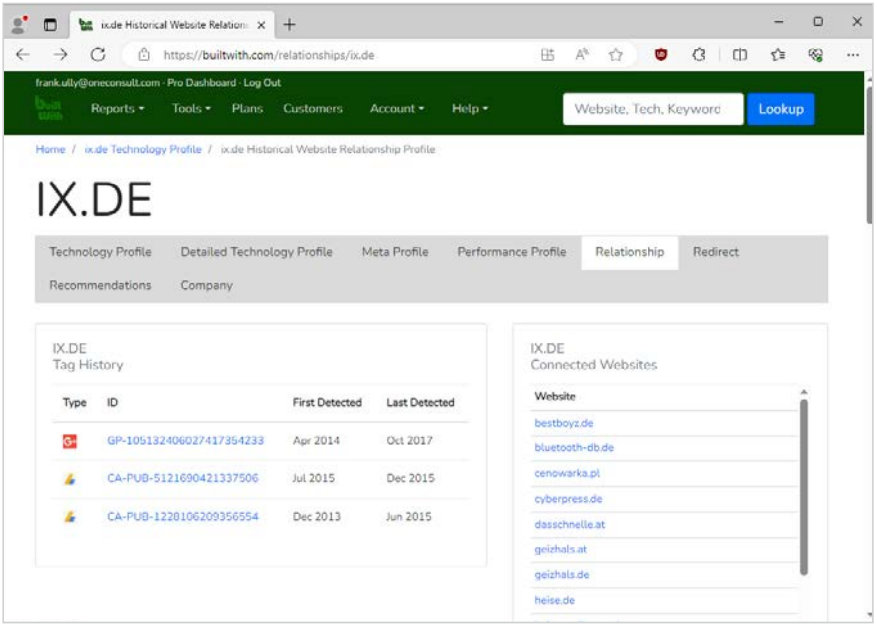
Die IoT-Suchmaschinen Shodan und Censys

Was fängt man ihm Rahmen von OSINT mit der Liste dieser Domänen, Subdomänen und IP-Adressen an? Hier kommen Internet-of-Things-Suchmaschinen wie Shodan und Censys Search ins Spiel. Diese spezialisierten „Dinge"-Suchmaschinen scannen regelmäßig den gesamten IPv4-Adressraum. Finden sie ein Gerät, folgt ein ausführlicher Portscan. Damit entdecken sie etwa Webdienste auf typischen Ports wie 80 oder untypischeren Ports wie 8888, verzeichnen weitere Informationen wie die Version einer Serversoftware und bereiten die Ergebnisse in ihren Weboberflächen durchsuchbar auf. Shodan ist ganz ohne Anmeldung nutzbar, dazu gibt man einfach die IP-Adresse in die Suchleiste ein (Abbildung 7).

Neben Shodan sollte man Censys Search konsultieren, das ähnlich funktioniert, teilweise aber andere Ergebnisse liefert, weil es eine eigene Scan-Engine verwendet und zu einem anderen Zeitpunkt zuletzt beim abgefragten System vorbeikam.

Für alle IP-Adressen sollte man sich die Einträge genau ansehen. Wahrscheinlich entdeckt man in einem der beiden Dienste ein unerwartetes System oder überraschenderweise offene Dienste wie eine ungeschützte Weboberfläche, die eigentlich nur aus dem Unternehmensnetz erreichbar sein sollte. Dann sollte man den Dienst abklemmen oder die Firewallregeln nachschärfen – und der Ursache nachgehen, wie es zu dieser Fehlkonfiguration kommen konnte, damit zukünftig die auf diese Art entstehende Angriffsfläche verkleinert wird.

Neben IP-Adressen kann man in beiden Webdiensten Domänen und Subdomänen eintragen, erhält dann aber auch irrelevante Ergebnisse. Shodan und Censys sind grundsätzlich ohne Kontozwang nutzbar. Ein kostenfreies Konto anlegen sollte man dennoch. Ein kostenpflichtiger Zugang für Shodan lohnt sich ebenso, weil er weitere Suchfilter und die Abfrage über eine API freischaltet. Neben einer lebenslang gültigen Einmalzahlung gibt es monatliche Pläne. Zahlende Benutzer können IP-Adressen überwachen lassen und werden benachrichtigt, sobald sich etwas ändert. Shodan bietet ein Kommandozeilenwerkzeug, für das man einen API-Schlüssel benötigt. Das Tool ist einfacher zu verwenden als die Website, wenn man nach vielen Hosts oder IP-Adressen sucht.



Das Relationship Profile im Website-Analysedienst BuiltWith zeigt verwandte Seiten, die dieselben Tags für Werbe- und Analysedienste enthalten oder enthielten (Abb. 8).

Kombinierte Abfragen

Mehrere IP-Adressbereiche oder einzelne Adressen kombiniert in einer Abfrage findet man auf Shodan mit der Syntax:

```
net:193.99.145.0-255,↵
    192.109.227.0-255,193.99.144.85
```

Dazu benötigt man ein Konto. Klickt man auf blau hinterlegte Organisationsnamen, sucht Shodan nach mutmaßlich zugehörigen Systemen; etwaige überflüssige IP-Adressen in der Abfrage entfernt man händisch. Bei Censys sucht man ohne Anmeldung mit folgender Syntax gleichzeitig nach mehreren IP-Bereichen und -Adressen:

```
ip:[193.99.145.0 to 193.99.145.↵
    255] OR ip:[192.109.227.0 to ↵
                192.109.227.255] OR ↵
                    ip:193.99.144.85
```

Eine hilfreiche Übersicht über diese Art von Maschinensuchmaschinen bietet die Seite Stuff Off Search der amerikanischen Cybersecurity-Agentur CISA. Bei Shodan sollte man unbedingt auf die Hilfe-Seite sehen und in die Datenbank fertiger Exploits. Weitere Ideen, was mit Shodan-Suchabfragen, sogenannten Dorks, möglich ist, liefern die GitHub-Projekte Shodan Dorks und Awesome Shodan Search Queries. Bei intensiverer Beschäftigung mit dieser Art von Suchmaschinen sind auch die Dienste BinaryEdge und Netlas einen Blick wert.

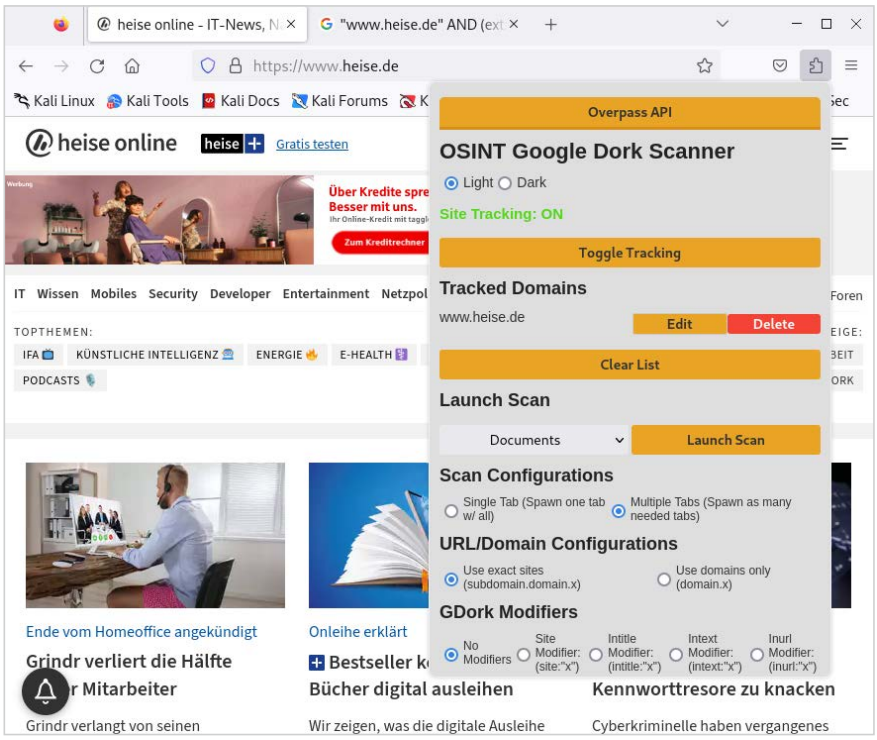
Spuckt eine dieser Internet-of-Things-Suchmaschinen in einem Banner den Namen oder sogar die Version einer Software aus, lohnt ein Blick auf den empfehlenswerten Known Exploited Vulnerabilities Catalog der CISA. Er verzeichnet inzwischen knapp tausend Schwachstellen samt ihrer CVE-Nummer, die bösartige Akteure aktiv ausnutzen, mit klaren Hinweisen, wie die Schwachstelle zu beheben ist.

Alternativen: Werbungs-IDs und Favicon

Ein innovativer Ansatz, verwandte Websites zu finden, die von derselben Person oder Organisation betrieben werden, ist die Suche nach Tags von Analyse- und Werbediensten wie Google Analytics oder AdSense. Ist ein Tag mit derselben ID in mehrere Websites eingebunden, liegt es nahe, dass sie einen gemeinsamen Betreiber haben. DNSlytics bietet in seinen Reverse-Tools die Recherche sowohl nach Analytics wie nach AdSense.

Dieselbe Funktion mit einer Suche nach mehreren Tag-IDs ermöglicht der Website-Analysedienst BuiltWith, klickt man in der grauen Registerleiste auf Relationship. Für diese Funktion benötigt man ein kostenloses Konto. Im Relationship Profile von ix.de in Abbildung 8 sieht man auf diese Art verbundene Domänen, basierend auf aktuellen und historischen Daten.

Das Favicon verkörpert die visuelle Identität einer Website in einem meist 16 × 16 Pixel großen Bild; Browser zeigen es in der Tableiste an. Für jedes dieser Symbole lässt sich ein eindeutiger Favicon-Hash berechnen, nach dem man in



Die Firefox-Erweiterung OSINT Google Dork Scanner überwacht die besuchten Websites und bietet vorgefertigte Google-Dork-Abfragen, beispielsweise nach Dokumenten oder Anmeldeseiten (Abb. 9).

Shodan suchen kann. Dafür ist ein Konto notwendig.

Auf GitHub gibt es eine Liste mit Shodan-Favicon-Hashes verbreiteter Dienste. Sucht man in Shodan zum Beispiel nach `http.favicon.hash:81586312`, findet man Zehntausende Systeme mit dem Automatisierungsserver Jenkins. Für nicht in der Liste verzeichnete Favicons berechnet ein Python-Skript den Hash, beispielsweise wie folgt für das Heise-Symbol, ausgeführt in Kali Linux, das man im ersten Tutorialteil eingerichtet hat:

```
python3 -m pip install mmh3
wget https://raw.githubusercontent.↵
  com/phor3nsic/favicon_hash_shodan/↵
                    master/favicon.py
python3 favicon.py ↵
     https://www.heise.de/favicon.ico
```

So entdeckt man weitere mit einer Organisation verbundene IP-Adressen und Dienste. Die Python-Werkzeuge fav-up und FavFreak helfen bei einer fortgeschrittenen Informationssammlung mit Favicons.

Hacken mit der Suchmaschine: Google Dorks

Kann man mit Google hacken? Klar, man kann. Die Crawler von Google und anderen Suchmaschinen durchkämmen so viele Teile des öffentlichen Webs, auch Surface Web oder Clear Web genannt, wie sie können. (Das Deep Web besteht im Gegensatz dazu aus Websites und Datenbanken, die zu Unternehmen, Behörden oder Universitäten gehören. Diese Inhalte sind mit einer Anmeldung geschützt oder zahlungspflichtig.) Sofern die Betreiber einer Website Crawler nicht mit einer robots.txt-Datei oder Metadaten auf den Seiten anweisen, Inhalte auszuschließen, indexieren die Krabbler alle Informationen, die auf einer Website vorhanden und öffentlich abrufbar sind – einschließlich womöglich vertraulicher Dateien oder verwundbarer Anwendungen.

Die meisten Menschen suchen nur mit Wortfolgen in Google, beispielsweise `Wetter in München`. Mit einer Reihe von Suchoperatoren lassen sich Ergebnisse jedoch eingrenzen: `ChatGPT site:heise.de` zeigt nur Inhalte von Heise zum KI-Chatbot; der Operator `site:` beschränkt die Suche auf eine bestimmte Domäne. Mit `filetype:pdf site:heise.de` findet man PDF-Dateien auf allen indexierten Subdomänen von Heise. Apropos Subdomänen: Mit einem Minus vor einem Wort schließt man es von der Suche aus. Dasselbe funktioniert mit Operatoren: `site:heise.de -site:www.heise.de` zeigt nur Inhalte von Subdomänen an, nicht vom Heise-Portal. Die Operatoren `inurl:` und `intext:` begrenzen die Suche auf die URL beziehungsweise den Seiteninhalt.

Einfach nutzen kann man Suchoperatoren mit der erweiterten Suche auf Google selbst, die in einem Formular wichtige Operatoren einbindet, mit denen man Ergebnisse auf Zahlenbereiche, die letzte Aktualisierung, die Domäne oder den Dateityp beschränkt. Schickt man das Formular ab, sieht man die Suchsyntax. Das einstündige YouTube-Video „Google Like a Pro“ stellt alle Suchoperatoren vor.

Verräterische Suchabfragen

Einen ersten Eindruck, welche Suchabfragen zum Hacken mit Google interessant sind, auch bekannt als Google Dorks, erhält man auf der Seite Google Hacking von Pentest Tools. Dort gibt man die Domäne ein und wählt aus einer Liste von Einträgen zum Beispiel, dass man Login-Seiten finden möchte.

Will man für die eigenen Websites schnell Dorks nutzen, hilft eine Firefox-Erweiterung. Dazu startet man in Kali Linux den Browser, klickt auf das Hamburger-Menü und wählt „Add-ons and themes“ aus. In das Suchfeld gibt man „OSINT Google Dork Scanner“ ein und findet eine Erweiterung von Bandit Pingu. Diese installiert man durch Klick auf die Schaltfläche „Add to Firefox“ und Bestätigen der Abfrage nach Berechtigungen für alle Websites. Nun steht im Puzzle-Menü, direkt links neben dem Hamburger-Hauptmenü, der OSINT Google Dork Scanner bereit (Abbildung 9).

Klickt man auf den Namen der Erweiterung und dann auf „Toggle Tracking“, zeichnet sie alle besuchten Domänen auf. Öffnet man anschließend wieder die Dork-Scanner-Erweiterung, steht die besuchte Seite unter „Tracked Domains“. Im Bereich „Launch Scan“ wählt man einen Eintrag, beispielsweise Documents, und startet die Suche mit Google Dorks, woraufhin sich eine oder mehrere neue Registerkarten mit der passenden Suchsyntax öffnen.

Artikel bei SecurityTrails und im OSINTTEAM-Blog weihen in tiefere Geheimnisse des Dorking ein (siehe ix.de/zgp7). Die größte Sammlung von Dorks ist die Google Hacking Database (GHDB), ein Index von Suchabfragen, mit denen Pentester und Sicherheitsforscher die für sie interessanten öffentlichen Informationen aufspüren können. Die Seite Dork Search macht die Abfragen aus der GHDB einfacher nutzbar.

Es gibt jedoch nicht nur Google. Nicht vergessen sollte man bei tiefgehenden Recherchen andere Suchmaschinen wie Bing, DuckDuckGo, Yahoo, Baidu und Yandex.

Fündig werden mit Codesuchmaschinen

Entwicklungs- und Codeaustauschplattformen wie GitHub sind ebenfalls gute Informationsquellen, weil man dort oft Quelltextschnipsel findet, die für Organisationen potenziell gefährliche Informationen liefern. Das reicht von IP-Adressen und Hostnamen über Konfigurationsdateien bis hin zu Anmeldedaten und Backups. Obwohl viele Entwickler wissen, welche Daten dort gespeichert sein könnten, unterschätzen sie das Risiko.

PublicWWW ist eine Suchmaschine auf Codebasis. Die Suche startet man mit dem Firmennamen oder dem Namen der Anwendung, die das eigene Unternehmen anbietet. Gute Einstiegspunkte sind ebenfalls bekannte Zeichenfolgen, IP-Adressen und Domänennamen. Auch nach JavaScript-Dateinamen, HTML-Fragmenten, HTTP-Headern oder Variablennamen kann man suchen; es lassen sich reguläre Ausdrücke verwenden. Unter Examples findet man in mehreren Registerkarten Beispielabfragen. Die Ergebnisse können als CSV exportiert werden. Zum Anzeigen der meisten Suchergebnisse genügt ein kostenfreies Konto, das die beliebtesten Ergebnisse freischaltet. Um Zugriff auf die gesamte Datenbank zu erhalten, benötigt man ein kostenpflichtiges Abo.

Grep.app ermöglicht die Suche über eine halbe Million Git-Repositorys und funktioniert ohne Anmeldung. Ein Blogartikel auf Include Security beschreibt, wie man damit sogar systematisch eine bestimmte Schwachstellenkategorie aufspürt. Ebenfalls ohne Log-in sucht man in Searchcode nach verschiedenartigen Codeschnipseln. Beispielhafte Abfragen sind direkt auf der Startseite verlinkt. Einen Blick wert ist die neue Codesuche von GitHub.

Listing: Installieren des pdtm und aller Werkzeuge

```
sudo apt update && sudo apt install -y golang libpcap-dev moreutils eyewitness
sudo apt remove -y python3-httpx
echo 'export PATH=$HOME/go/bin:$PATH' >> ~/.zshrc && source ~/.zshrc
go install -v github.com/projectdiscovery/pdtm/cmd/pdtm@latest
pdtm && source ~/.zshrc
pdtm -install-all
```

Arbeitserleichternde Werkzeuge

Diverse Open-Source-Werkzeuge dienen dazu, OSINT und aktives Sammeln von Informationen zu vereinfachen oder teilweise zu automatisieren.

Besonders hervorzuheben ist das Project Discovery, das viele Open-Source-Tools für aktive und passive Informationssammlung anbietet, darunter die Nmap-Portscanner-Alternative naabu oder der stark anpassbare Schwachstellenscanner nuclei. Das Projekt bietet einen Open-Source-Tool-Manager pdtm, der alle Werkzeuge des Projekts in einem Rutsch bereitstellt. Das Listing zeigt das Installieren unter Kali Linux.

Das Subdomänen-Suchwerkzeug subfinder findet nun schnell und einfach mit passiven Abfragen Subdomänen und schreibt sie in eine Datei:

```
subfinder -d heise.de -o ↵
                        subdomains.txt
```

Darunter befinden sich allerdings auch viele Subdomänen, die längst nicht mehr in Betrieb sind. Mit dem DNS-Abfragetool dnsx ermittelt man für alle Domänen aktuelle DNS-Einträge wie IP-Adressen:

```
cat subdomains.txt | dnsx -silent ↵
                -resp -a -aaaa -cname
```

Im Folgenden fragt der Portscanner naabu aus der Shodan-Datenbank offene Ports passiv ab, ohne selbst aktiv zu scannen. Zuvor wird die Liste der Subdomänen nach aktuellen IPv4-Adressen aufgelöst und sortiert:

```
cat subdomains.txt | dnsx -a ↵
       -silent -resp-only | sort -u ↵
             | sponge ipaddresses.txt
naabu -l ipaddresses.txt -passive
```

Scannt man viele IP-Adressen passiv mit naabu, können aufgrund von Timeouts Ergebnisse fehlen. Ein interessantes alternatives passives Scan-Tool ist Smap, das ebenfalls Shodan-Daten abfragt und gleiche Kommandozeilenparameter wie Nmap unterstützt.

Mit weiteren Programmen von Project Discovery wie dem TLS-Scanner tlsx und dem HTTP-Tool httpx sowie Werkzeugen anderer Entwickler sind aktive Abfragen möglich, die über den Rahmen von OSINT hinausgehen, weil sie direkt mit den Zielsystemen Verbindungen herstellen. Der folgende Befehl für das bereits im ersten Tutorialteil vorgestellte

Screenshot-Tool eyewitness erstellt automatisiert Bildschirmfotos, wozu es den Aufruf der jeweiligen Startseite mit einem Browser simuliert:

```
eyewitness --prepend-https -f ↵
                        subdomains.txt
```

Eine ausführliche Anleitung zu Subfinder findet sich im Blog von Project Discovery. Ebenfalls dort beschreibt eine fünfteilige Artikelreihe zu Reconaissance Vor- und Nachteile von aktiver und passiver Informationsbeschaffung. Sie zeigt außerdem, wie mit Tools von Project Discovery aktiv Informationen gesammelt werden.

wtfis bereitet passive Abfragen von Daten zu Domäne und IP-Adresse in einer textbasierten Oberfläche für Menschen auf. Eine gute deutschsprachige Anleitung, wie man das Tool samt notwendigen API-Schlüsseln einrichtet, findet sich in Ausgabe 23/2022 der iX-Schwesterzeitschrift c't [5] und auf heise+.

Daneben gibt es zahlreiche OSINT-Frameworks. Ein neuer spannender Vertreter ist der Paketmanager sn0int. Bereits etabliert sind Recon-ng, Spiderfoot sowie Maltego Community. Für alle Frameworks findet man im Web zahlreiche Anleitungen.

Deutlich über den Rahmen von OSINT hinausreichend, aber sehenswert ist reconftw, ein Toolset zur automatisierten Erkundung einer Zieldomäne, das die besten Werkzeuge zum Scannen und Auffinden von Schwachstellen kombiniert. Es führt auch aktive Scans durch. Zwei gute Anleitungen dazu stehen in Beiträgen des Subdomain Enumeration Guide und im Blog von Jason Haddix.

Weiterführende kostenpflichtige Quellen

OSINT ist eine eigene Spezialisierung innerhalb der Informationssicherheit. Dieser Artikel hat nur einige Aspekte angerissen, die Organisationen beim Einschätzen der eigenen Angriffsfläche helfen. Jedoch gibt es weitere Datenquellen wie soziale Medien, Fotos und Videos oder Kryptowährungen sowie andere Anwendungsfälle wie das Fahnden nach vermissten oder mutmaßlich kriminellen Personen.

Referenzwerk und unbedingt lesenswert ist das Fachbuch „Open Source Intelligence Techniques“, das der Autor Michael Bazzell jährlich auf den neuesten Stand bringt. Es beschreibt, wie man sich mit einer Ubuntu-VM eine OSINT-Umgebung einrichtet, Tools installiert und nutzt und unterschiedliche Daten aus sozialen Netzwerken von Bildern über Videos bis hin zu Livestreams auswertet. Außerdem stellt Bazzell auf seiner Website mit den „IntelTechniques Search Tools“ kostenlos eine umfangreiche und einfach nutzbare Sammlung öffentlicher Quellen bereit; die Suchfelder sind direkt in die Seite eingebettet.

Ein neueres Buch ist „Deep Dive: Exploring the Real-world Value of Open Source Intelligence“ von Rae Baker. Baker gibt einen Überblick über die Geschichte von OSINT und Grundlagen wie Operations Security (OPSEC) und wie man dabei einen kleinen Fußabdruck hinterlässt, um beim Sammeln der Daten nicht aufzufallen und identifiziert zu werden. Sie zeigt Strategien, wie man auf die Jagd nach Open-Source-Informationen über Personen, Organisationen, Kryptowährungen und Verkehrsmittel wie Schiffe oder Flugzeuge geht.

Nicht unbedingt relevant zum Sichern von Organisationen, aber spannend: Bei „Kase Scenarios“ lernt man in realistischen Szenarien, mithilfe öffentlicher Informationen Kriminalfälle zu lösen; das erste Modul ist kostenfrei.

Weiterbildung mit lehrreichen kostenfreien Quellen

Eine empfehlenswerte Lernplattform zu Informationssicherheit ist TryHackMe. Die Inhalte dort sind in sogenannten Räumen in gut konsumierbaren Häppchen aufbereitet, mit Übungen versehen und nach einer Registrierung häufig kostenfrei zugänglich. Für OSINT spannend sind die Räume Red Team Recon, Passive Reconaissance, WebOSINT und OhSINT sowie diejenigen zu Shodan und Google Dorking. Auf YouTube stellt der Cyber Mentor kostenfrei einen fünfstündigen Videokurs zu OSINT bereit.

Wer sich nicht wie im ersten Tutorialteil skizziert auf Grundlage von Kali Linux eine OSINT-VM einrichten will, kann bei Trace Labs eine betriebsfertige Maschine herunterladen, die auf Kali basiert. Trace Labs ist eine gemeinnützige Organisation, die hilft, mit öffentlichen Informationen vermisste Personen aufzuspüren. Browsererweiterungen für Chrome und Firefox, die bei der OSINT-Arbeit unterstützen, listet das GitHub-Projekt „Awesome Browser Extensions for OSINT“.

Eine der bekanntesten Quellen für Open Source Intelligence ist das OSINT-Framework, das in einer großen Mindmap-Struktur Webseiten und Werkzeuge sammelt. Das Framework wird allerdings kaum aktualisiert und überwältigt Einsteiger mit der Menge der Einträge. Ein aktueller Framework-Nachfolger ist die „Malfrat's OSINT Map“.

Viele thematisch sortierte OSINT-Quellen sammeln die GitHub-Repositorien Awesome OSINT, Awesome Intelligence sowie Web Recon. Wen die zahlreichen Quellen dort überfordern, der findet auf cylect.io eine durchsuchbare Sammlung. Schließlich bleibt man auf dem Laufenden mit dem Lesen des Newsletters Week in OSINT von Sector035 sowie dem Blog von OSINT TEAM.

Fazit

Mit OSINT nutzt man dieselben Informationsquellen und Werkzeuge wie Angreifer, um Schwachstellen in der eigenen IT zu finden. Der neudeutsche Begriff dafür ist Attack Surface Management oder Angriffsflächenmanagement oder -monitoring. In dieser Produktkategorie tummeln sich zahlreiche kommerzielle Anbieter wie Assetnote, RedHunt Labs, runZero oder SecurityTrails. Eine Übersicht bietet die GitHub-Seite „Awesome Attack Surface Monitoring“. Wenn man OSINT geschickt nutzt, zieht man denselben Nutzen aus öffentlichen Quellen. (ur@ix.de)

Quellen

[1] Sascha Herzog; Mit allen Mitteln; Sicherheitstests: Angriffe auf Technik und Mensch; iX 2/2018, S. 78

[2] Andreas Heideck; Layer 8 – Faktor Mensch; iX 12/2022, S. 84

[3] Frank Ully; Nach oben gehangelt; Informationsbeschaffung – was jeder Domänenbenutzer alles sieht; iX 10/2020, S. 58

[4] Nadia Meichtry, Fabian Murer, Tabea Nordieker; Malware-Analyse per OSINT und Sandbox; iX 3/2023, S. 122

[5] Wilhelm Drehling; Mit wtfis detaillierte Informationen über Domains und IP-Adressen herausfinden; c't 23/2022, S. 168

[6] Alle im Text genannten Artikel, Webseiten, Dienste und Tools sind über ix.de/zgp7 zu finden.

FRANK ULLY

ist Head of Research der Oneconsult Deutschland AG in München. Er beschäftigt sich mit aktuellen Themen der offensiven IT-Sicherheit.