

Internet Background Noise

Themenbereiche:	Security, Data Science
Studierende:	Jessica Schumacher
Betreuungsperson:	Peter Infanger
Experte:	Martin Burri
Auftraggebende:	SWITCH
Keywords:	IBN, Darknet, Network Security, Monitoring, DDoS Backscatter, Scans, Bots, Garbage

1. Aufgabenstellung

Die Stiftung SWITCH betreibt das National Research and Education Network (NREN) und vernetzt damit die Universitäten, Hochschulen und in den Diensten der Lehre und Forschung stehenden Stellen innerhalb der Schweiz. Zu diesem Zweck verwaltet und verwendet sie mehrere grössere IP-Netzwerke. Nicht alle dieser IP-Adressen werden verwendet. Ein bestimmter Bereich wird absichtlich bereits seit längerer Zeit nicht vergeben. In diesem IP-Adressen-Bereich wird jeglicher Netzwerkverkehr gesammelt und aufgezeichnet. Diese Daten, auch Internet Background Noise (IBN) genannt, werden jedoch aktuell nicht weiterverarbeitet. Das Ziel der Arbeit ist es, die gesammelten Daten aufzubereiten und zu analysieren, um Erkenntnisse zu erlangen. Folgende Frage soll nach Abschluss der Arbeit beantwortet werden: «Bietet die Sammlung und Analyse von IBN für SWITCH einen solchen Mehrwert, dass die Weiterentwicklung und Aufrechterhaltung dieser Tätigkeit in Zukunft als lohnenswert betrachtet werden kann?»

2. Lösungskonzept

Das Vorgehensmodell dieser Arbeit wurde an den CRISP-DM Prozess aus der Data Science angelehnt. Nach der Initialisierung und Planung der Arbeit wurde die aktuelle Situation bei SWITCH erfasst. Anschliessend musste die Datensammlung analysiert, kritisch hinterfragt sowie optimiert werden. Gleichzeitig zeigten Recherchen auf, welche Erkenntnisse die Forschung anhand der Daten gewinnen konnte und wie und ob andere Unternehmen IBN einsetzen. Anhand einer Nutzwertanalyse wurde die Plattform zur späteren Analyse der Daten bestimmt. Die Daten mussten aufbereitet und in die gewählte Plattform transportiert werden. Erste deskriptive Statistiken wurden erstellt um die Datenlage zu beschreiben und unter Umständen spannende Phänomene in den Daten zu erkennen. Einige Use Cases wurden für die weitere Analyse ausgewählt und ausgearbeitet. Vor allem beim letzten Schritt ermöglichte ein Abgleich mit Threat Intel das Detektieren von böartigem Verkehr.

3. Spezielle Herausforderungen

Die Datensammlung, auf welcher diese Arbeit aufbaute, war nicht homogen implementiert und zum Teil fehlerhaft. Zudem war die Dokumentation der Datensammlung eher spärlich. Deshalb musste ausserplanmässig zusätzliche Zeit in diesen Teil der Arbeit investiert werden. Die enorme Datenmenge im Allgemeinen war zudem eine spezielle Herausforderung, welche ein Umdenken erforderte, da teilweise nicht mit Standard-Tools oder Software gearbeitet werden kann. Eine weitere Herausforderung dieser Arbeit, im Vergleich zu herkömmlichen Data Science Projekten, ist die Kontinuität der Daten. Diese werden in Form

eines Streams fortlaufend konsumiert und die Analysen müssen dementsprechend kontinuierlich durchgeführt werden.

Als besonders anspruchsvoll wurde die Unterscheidung zwischen böartigem und gutwilligem Verkehr betrachtet. Eine Unterscheidung ist nicht ganz trivial und wurde stellenweise nur angedeutet oder nicht klassifiziert. Die eingesetzte Analyseplattform stiess vor allem bei den komplexeren Analysen und Statistiken an ihre Grenzen. Dies erforderte eine erhöhte Kreativität im Schlussteil der Arbeit.

4. Ergebnisse

Aufgrund der Datenanalyse konnten spannende Ereignisse beobachtet werden. Es folgt eine Auflistung einiger observierter Fakten:

- Bis zu 50'000 Pakete pro Minute wurden an die 256 überwachten IP-Adressen gesendet.
- TCP ist das meist verwendete Protokoll.
- Ein grosser Teil von IBN sind Scanner, welche das gesamte Internet scannen.
- Eine DNS Amplification Attacke wurde beobachtet.
- DDoS Backscatter ist reichlich vorhanden und konnte identifiziert werden.
- Mirai Bots kontaktierten einige IP-Adressen aus dem überwachten Netzwerk.

Die deskriptiven Statistiken beschreiben die einzelnen Protokolle, welche auf Splunk entgegengenommen und interpretiert werden konnten. Als Use Case wurde das Erkennen von Scannern, DDoS Backscatter Verkehr sowie infizierten Bots, welche sich ausbreiten, ausgewählt.

Es wurde aufgezeigt, dass das Internet von verschiedensten Instanzen abgescannt wird. Dabei handelt es sich um gutartige Scanner, wie auch böartige Scanner.

DDoS Backscatter Verkehr konnte als solches erkannt und klassifiziert werden. Die Identifikation des Backscatters Verkehrs erfolgte mithilfe von bereits definierten Patterns aus der Forschung. Während der Dauer der Arbeit wurde keine grössere DDoS Attacke festgestellt, welche in den Daten nachvollzogen werden konnte.

Das Detektieren von Bots wurde anhand von mehreren Methoden umgesetzt. Als erstes wurde ein Analysieren der Peaks von Anfragen von distinkten Source IP-Adressen realisiert. Peaks wurden detektiert, jedoch konnte der Verkehr nicht eindeutig zu infizierten Bots zugewiesen werden. Bei der zweiten Methode wurden IoT Botnetze genauer betrachtet und deren herkömmlichen Kommunikationskanäle untersucht. Aufgrund von mangelnden Paketinformationen auf der Analyseplattform, was ein Fingerprinting erschwert hat, konnten auch hier keine eindeutigen Ergebnisse gemacht werden. In einem späteren Schritt wurden IP-Adressen aus bekannten Botnetzen, wie beispielsweise Mirai, in den Daten gesucht. Diese Suche war erfolgreich. Das zeigt auf, dass gewissen Bots immer noch zufällige IP-Adressen verwenden, um weitere infizierbare Geräte zu finden.

In den Daten wurden zudem etliche unerklärliche Phänomene entdeckt, welche nicht genau zugeordnet werden konnten.

Die gestellte Fragestellung kann anhand der Ergebnisse bejaht werden. Vor allem für das Erkennen von Trends und neuartigem böartigem Verkehr bietet IBN viel Potenzial. Diese Trenderkennung wäre für die Vorbereitung auf mögliche Attacken ein Vorteil und könnte mit den Kunden von SWITCH geteilt werden.

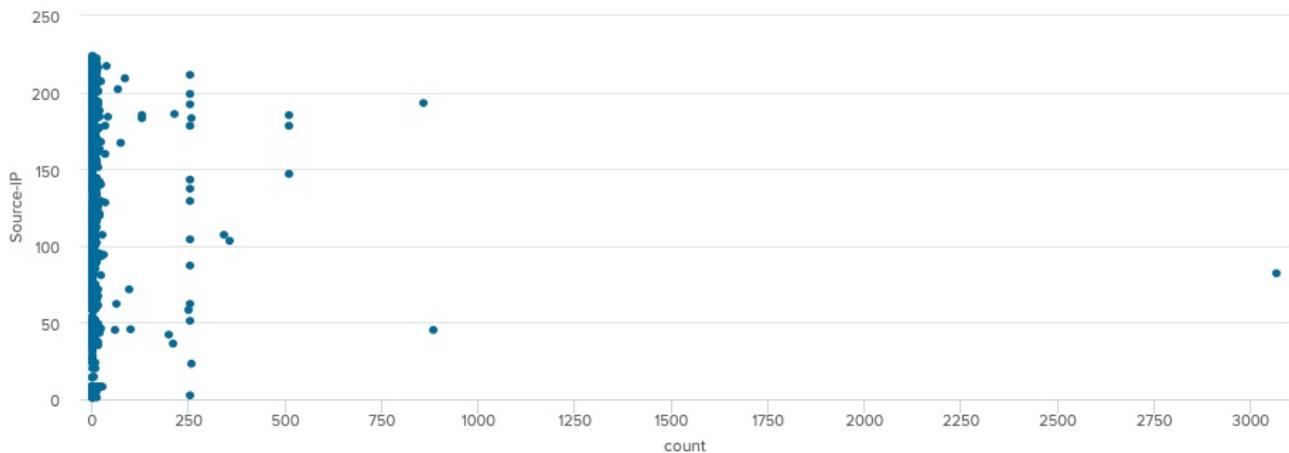


Abbildung 1: Verteilung Source IP-Adressen und Anzahl gesendete Pakete: Viele IP-Adressen senden eher wenige Pakete und nur wenige IP-Adressen senden viele Pakete (Output gekürzt).

5. Ausblick

Die geleistete Arbeit bietet eine gute Grundlage, um die Analyse von Internet Background Noise bei SWITCH einen Schritt weiterzuführen. Die deskriptive Beschreibung der Daten ist bereits vollumfänglich implementiert und das Potenzial der Daten konnte in der Arbeit aufgezeigt werden. Es sollen zudem noch weitere Use Cases erkannt und analysiert werden. Die implementierten Use Cases sind aktuell in einem Proof-of-Concept Status und können noch erweitert und ausgearbeitet werden. Hier muss später aber auch die Frage gestellt werden, ob die gewählte Analyseplattform für solch komplexe Klassifizierungen die geänderten Anforderungen immer noch erfüllt. Das Detektieren von Paketen, welche von infizierten Bots stammt, könnte mit Machine Learning besser umgesetzt werden als die aktuell vorgestellten Methoden. Dies wurde bei verschiedenen Forschungen bereits umgesetzt und beschrieben. Es müsste abgeklärt werden, ob dies vom Auftraggeber erwünscht ist und in welchem Umfang Machine Learning eingesetzt werden soll.

In Zukunft sollen Trends in den Daten besser und vollkommen automatisch detektiert werden können. Es sollen Benachrichtigung versandt werden, damit diese Trends schnellstmöglich verifiziert und weitere Schritte eingeleitet werden können.

Zudem ist das Bereitstellen der Daten an die Kunden eine Option, falls dies von Kundenseite gewünscht wird. Unter Umständen könnte man die Daten auch für Forschungszwecke an Universitäten zur Verfügung stellen, dies müsste aber vor allem aus Datenschutzgründen noch abgeklärt werden. Im Weiteren kann ein Teilen der Daten oder den erlangten Erkenntnissen an andere NRENs in Betracht bezogen werden.