Resolving Dark Web Identities

By Babur Kohy
Cybersecurity Doctoral Student
Marymount University

Abstract

Criminals continue to pose a significant threat on the dark web; a threat that will likely continue for decades. As cybersecurity professionals, we often hear and read about many forms of breached data for sale on the dark web yet we seldom learn how to properly resolve the identities of these criminals. Awareness alone of data breaches, sales, and/or knowledge of an uncontrolled release of information is not sufficient to adequately diagnose, fix, and prevent these problems from happening again. Therefore, it is important to quickly identify and expose these bad actors using the same hidden networks they use for exploitation, extortion, and theft. My research will attempt to offer a solution to these problems by investigating several multi-faceted strategies to identify, analyze, and expose criminal dark web identities.

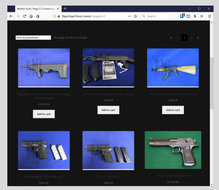
What is the dark web?

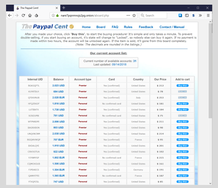
A managed attribution network that is layered on top of the internet. A decentralized network that uses advance / multi layer encryption for communication. Sometimes requires special software for access.

Background of The Problem

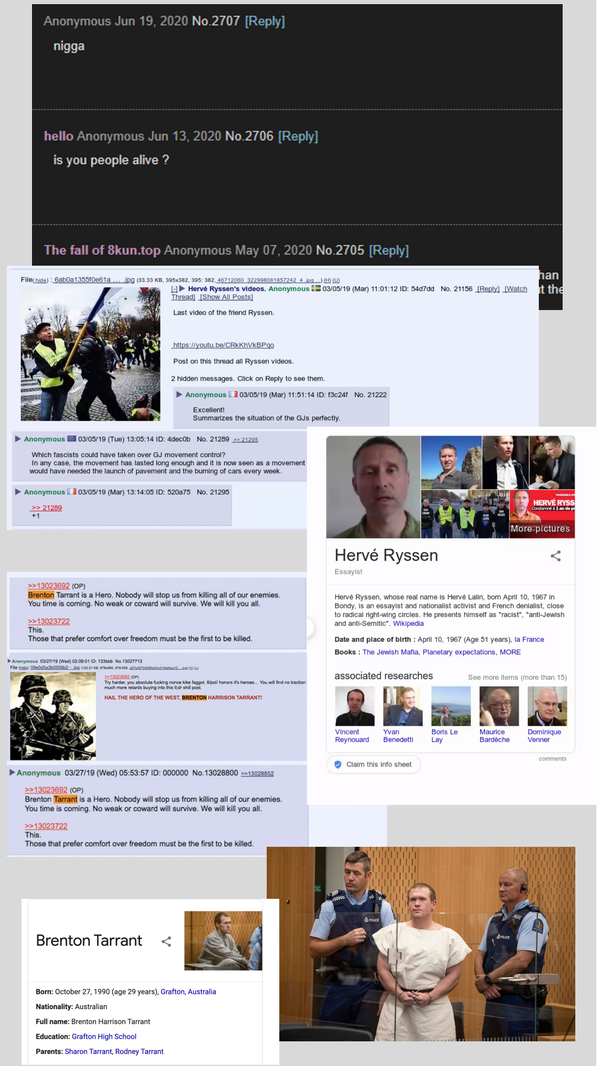
Dark web is notorious for illegal activity such as:

- Stolen financial data for sale
- Counterfeit currency trade
- Fake identities
- Assassination services
- Illegal guns
- Porn
- Cyber weapons
- Trolling, bullying, intimidation





Problem Examples



Significance of Research

- Empirical study of identity resolution techniques on the dark web
- Setup a process for researchers and industry to resolve anonymous identities
- Develop a nicely constructed dataset of dark web domains gathered during the research to be used as seed for other researchers

Dark Web Access

Research Questions?

Is it possible to resolve identities used on the dark web?

- Using language matching – pure linguistics (syntax (grammar), Semantics (meaning)
- Online patterns of life – being able to identify someone's digital pattern of life on one domain then match to another domain
- User handle resolution – take an anonymous alias from dark web to resolve to surface web (vise versa)
- Poster IDs and IP locations

Patterns of Life – How Are You Targeted?

Identify your patterns:

- Physical surveillance
- *Tags*
- *Observers*
- Electronic surveillance
  - Intercept
  - Implants
  - PAI collection

- Identify family, friends, and close associates
- Target weaknesses in your network
- Behavior over technology
- People fail because of poor behavior and not because of technology

Data Collection

- Data will be gathered from dark web hidden services and encrypted messaging apps
- Surface web sites gathered during the investigation will be searched for possible match

ML Models

- Language Matching: turning text into features using support vector machine, random trees, etc.
- Language Modeling for Anonymous Identities: Fuzzy Classifiers

Contact

Babur Kohy
Cybersecurity Doctoral Student
The Cyber Center
bkohy@Marymount.edu

Dr. Nathan Green
The Cyber Center
Committee Chair
ngreen@marymount.edu

MARYMOUNT UNIVERSITY