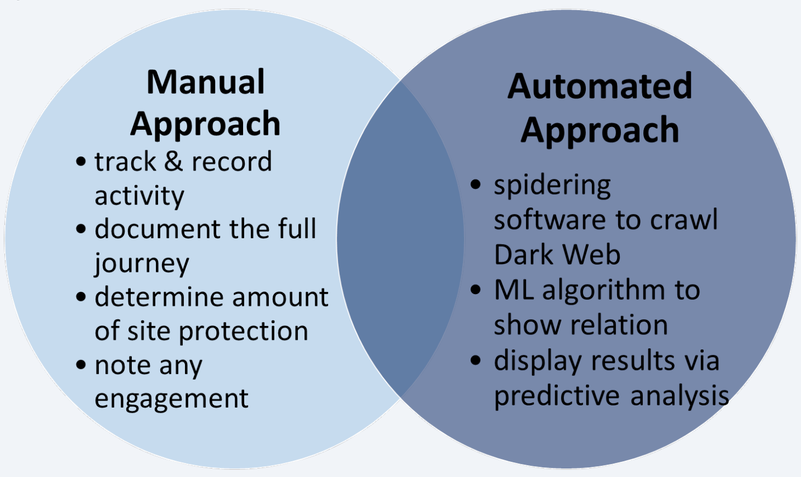
A Framework for Identifying Malware Threat Distribution on the Dark Web

Shelby A. Caldwell, J. Todd McDonald, Ryan G. Benton, David Bourrie

School of Computing, University of South Alabama

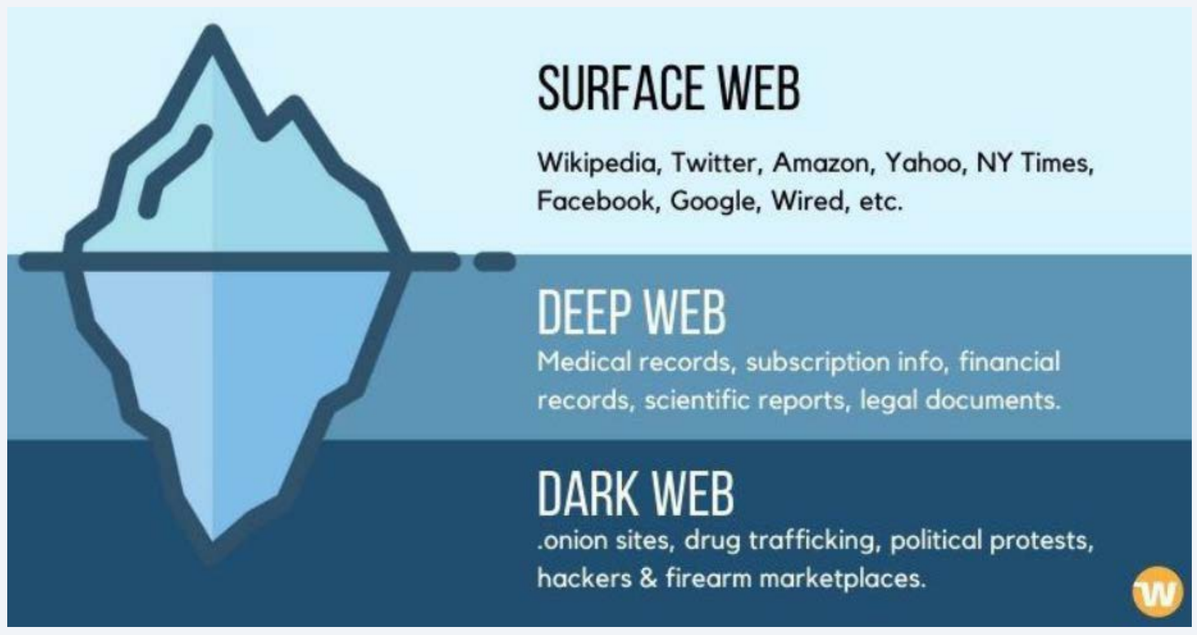
ABSTRACT

The **Dark Web** is an ever-growing phenomenon that has not been deeply explored. Contents of the dark web are primarily the buying and selling of **unauthorized goods, illegal activity**, and **trading services**. It is no secret that in recent years, malware has become a potent threat to technology users. Cybercriminals and hackers utilize the dark web for **malware distribution** by selling the code under aliases and via cryptocurrency. The Dark Web is known for supporting anonymity and secure connections for private interactions, thus making it a favored medium for people who engage in illegal activity, and thus a rich environment for discovering trends, details, and indicators of emerging malware threats. By examining this malware threat distribution, we gain useful information regarding this Dark Web activity. Through the application of **data science** and **open-source intelligence** techniques, trends in malware distribution can be studied such as the types of people selling the malware, the amount of malware being exchanged, the type of currency being used, and what malware is gaining popularity at a specific time. In this research, we aim to create a framework for helping identify malware threat distribution patterns.



We will examine this type of dark web activity using an **automated** and **manual approach** that collects data on malware exchanges. Furthermore, a comparative analysis is conducted to determine which approach is more effective and efficient. Our framework for identifying current or future malware threats that are distributed on the Dark Web will be refined by examining the weaknesses and strengths of each gathering approach.

MOTIVATION

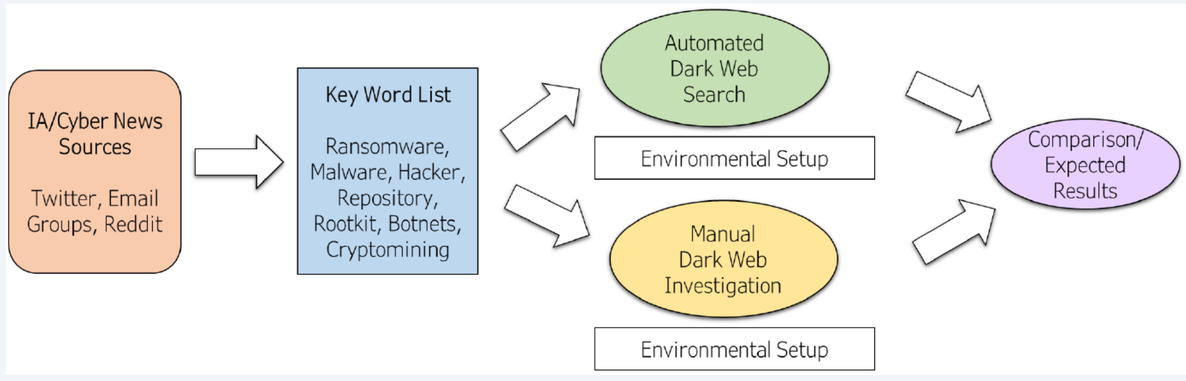


In recent years, malware has become an increasing threat to all cyber users. CISA, FBI, NSA, along with many other government agencies have warned about increasing malware attacks and steps users can take to prevent them. In this research, we will address malware threat distribution and the specific mechanisms used to distribute it. The goal of this research is to identify emerging malware threats such as rootkits, ransomware, and target specific code. By analyzing how they are advertised, distributed and purchased, we can aid law enforcement, researchers, and businesses that are in pursuit of mitigating the risk and containing the spread of malware.

RESEARCH QUESTIONS

- ❑ **How is malware advertised and purchased on the Dark Web?**
- ❑ **What are strengths and weaknesses of automated vs. manual approach for identifying malware distribution trends?**
- ❑ **Can we effectively use data science techniques to classify or link parts of the distribution process?**

METHODOLOGY



Goals:

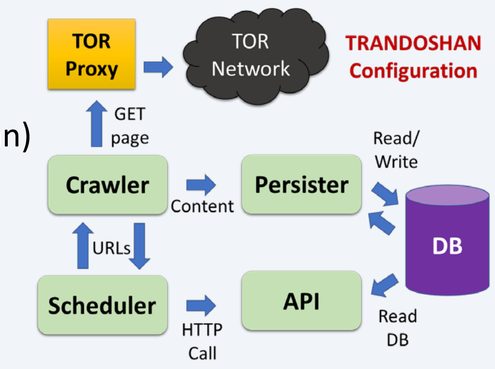
- ❑ Develop a framework for identifying **current** or **future** malware threats that are distributed on the Dark Web
- ❑ Develop and evaluate a **manual** process for documenting the distribution sources of prominent malware threats
- ❑ Develop and evaluate an **automated** machine learning approach for documenting the distribution sources of prominent malware threats
- ❑ Compare the **effectiveness** of manual vs. automated methods

Manual Approach:

- ❑ Examine 5-10 domains
- ❑ Track time spent on browsing, dialogue, search, etc.
- ❑ Document journey/links between starting point and locations
- ❑ Note password-protection, user account, captcha, puzzles, etc.
- ❑ Record engagements with vendors

Automated Approach:

- ❑ Setup/install spidering software (Trandoshan)
- ❑ Create/adapt scripts
- ❑ Expected data will be HTML, no photos
- ❑ Run ML algorithm to trace patterns/trends



Case Study Details:

- ❑ Corpus data will be gathered in the form of HTML
- ❑ Due to the nature of the Dark Web, no images will be collected throughout this process
- ❑ Machine learning algorithm applied to find relationship between the total sites found and the sites containing keywords that were matched
- ❑ Algorithmically examine .onion domains and perform information extraction of any keywords found based on site name, repeated seller names, and cryptocurrency used

Comparative Analysis:

- ❑ Assess weaknesses and strengths of both automated and manual methods
- ❑ Take note of successes (presence of malware) and identification of source or potential sources for where malicious software originated or were sold on the Dark Web
- ❑ For each original word that was searched for, we identify a market where that malware or organization communicates
- ❑ Identify communications between specific people looking for or selling that malware by checking forums and comment sections on the Dark Web.
- ❑ Identify means to download or obtain a copy of the malware
- ❑ Identify goals, outcomes, and prices for the malware
- ❑ Profile the supply and demand of malware and how frequent these transactions are occurring

ACKNOWLDGEMENTS

This research was supported in part by the National Science Foundation under grant DGE-1564518.

QR Code