Dark Web掲示板投稿者の特定手法の提案

Proposal of a Method for Identifying Bulletin Posters on Dark Web

青木卓矢・システム分科会・情報セキュリティ大学院大学

In recent years, there is a Web service that can not be reached from the normal network infrastructure. Such Web service is growing rapidly as a service called Hidden Service using a network anonymous communications system Tor. This Hidden Service is presented by various servers without revealing their identities and is used as a foundation of black business. Dark Web services also exist in Japan, and they primarily provide some confidential bulletin boards. So this paper studies a method to identify criminals contributing to the bulletin boards, by finding and correlating similar posters on the both boards. The study was carried out by collecting and analyzing threads information from both bulletin boards. As a result, it can identify the person who carried out the illegal trade by using the same author name from both bulletin boards.

【Dark Webとは】

- サービス提供者及びユーザが互いに身元を明かさずに各種サービスを提供することができる.
- 独自の擬似アドレスを持たせることで，特定のIPアドレスと結びつける必要がない.
- 通常のTor経由による一般サイトへのアクセスがExitノードを介すのに対し，Dark WebはTorネットワーク内で完結する点と点への接続が提供される.

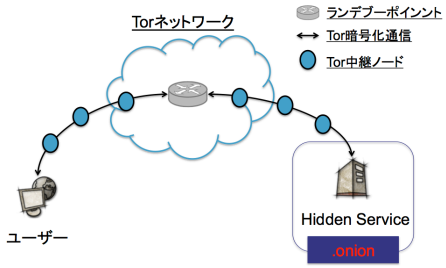


図1 Dark Webの構成

【提案手法】

- Dark Web掲示板及び通常のネットワーク上に存在する掲示板の全スレッド情報を取得し，データベース化
- Dark Web掲示板に投稿された違法取引キーワードを目視で抽出し，得られたキーワード基に，データベースから両掲示板より，類似する投稿を取得.
- 取得した投稿データを基に，投稿者名，投稿時間，投稿文の特徴から同一性を分析する.

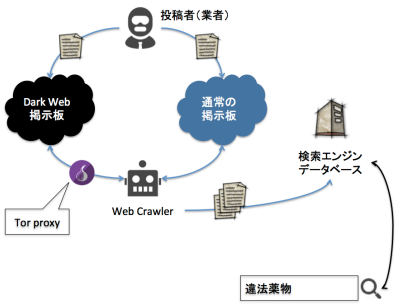


図2 両掲示板より情報収集

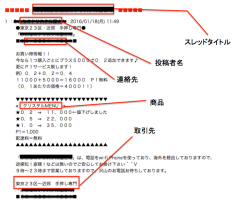


図3 キーワード抽出例

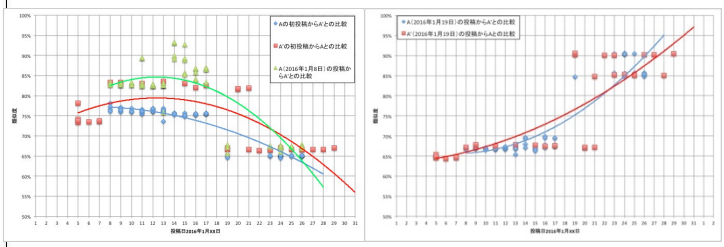
【投稿者の同一性分析】

- 両掲示板より得られた情報から同一名で投稿するものを一人選択
- 両掲示板で投稿日時と投稿本文に類似度が出ると想定. 投稿文の類似度については，Jaro-Winkler距離アルゴリズムを用いて分析を行う.

【分析対象】

- Dark Web掲示板投稿者 → A
  - 2016年1月5日～2016年1月29日の投稿
- 通常のネットワーク上に存在する掲示板投稿者 → A'
  - 2016年1月8日～2016年1月26日の投稿

【分析結果】



- Aの初投稿からA'の比較
  - A'には1月5日～7日の間投稿がなかったため，類似度は高くても70%～80%の結果が出た
- A'の初投稿からAの比較
  - 上記の理由により1月7日以前の投稿では，80%未満の類似度しか得られなかったが，1月8日以降に80%以上に増加した.
- A（2016年1月8日）の投稿からA'との比較
  - Aの比較基準をA'の初登校日と合わせた結果，1月8日～1月19日までの間85%～95%の高い類似度が得られた.

- 3つの類似比較を行った結果，同様に65%～70%の結果が出たものについては，1月19日以降に投稿の広告内容が大幅に変わったためである.