Blog Farm Detection Using Social Network Analysis and Social Cyber Forensics Informed Methodologies

Tuja Khaund, Billy Spann, Richard Young, Samer Al-khateeb, and Nitin Agarwal
{txkhaund, bxspann, rbyoung, nxagarwal}@ualr.edu , sameral-khateeb1@creighton.edu
Department of Information Science

Abstract

Blogs play a vital role on the Internet for retrieval of real time information, a place for users to gain insights of events around them and also find communities that share similar interests. However, over the years, blogs have been infused with spam content and also have fallen prey to link farming. Blog search engines are plagued with fake content that may or may not be automatically generated with the motive of increasing the search ranking and thereby affecting crowd manipulation. In this paper, we monitor 25 blog sites obtained from the Twitter timelines of around 50 Ukraine parliament members and detect relevant blogs based on tracking codes. Our results show preliminary observations of the formation of a blog farm or a network of blog bots. We conducted a ping test on all the suspicious blogs and we observed patterns in their activity levels that suggest a presence of a master controller for the blog farms.

Challenges

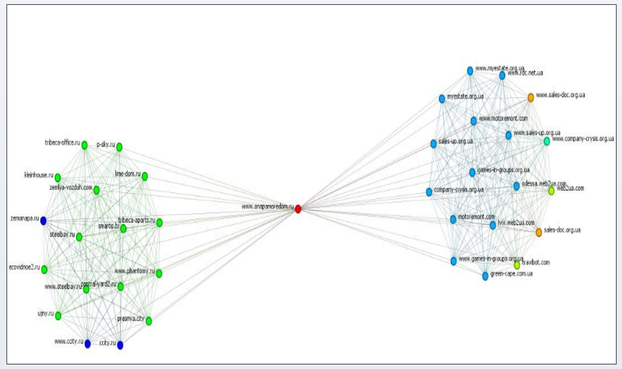
- **Blog sites names are randomized, sometimes harder to detect new blogs.**
- **Need more work to track Inactive sites over time and scale**

Methodology

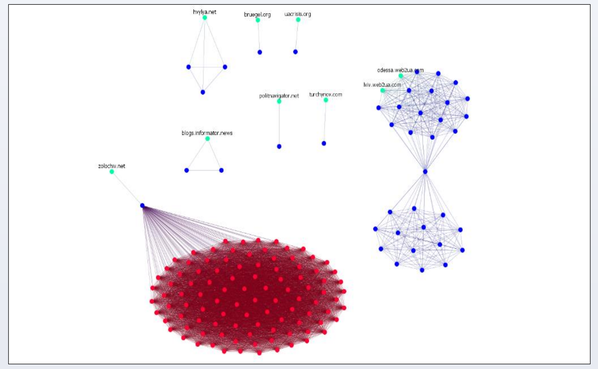
While studying Ukrainian parliamentary discourse we identified a set of 25 blog sites from Twitter timelines of the parliament members of Ukraine.

- **Using Maltego, a cyber-forensics tool, we first extracted tracking codes linked to these blogs. Out of 25 blogs, only 17 blogs contained a tracking code.**
- **Several blog sites containing arbitrary numbers and characters were detected. These arbitrary blogs also appeared in alphabetical order with similar starting characters as part of the name, further suggesting they were part of a blog farm.**
- **The source of these blog sites was taken down after sometime, which led to a reverse track of its tracking code to retrieve the lost data.**
- **ORA [14] was used to visualize the network of blogs in which two blogs are connected if they share a tracker code.**
- **We also ran a ping test on the set of newly identified blog sites, once a day, starting from April 23, 2018 to monitor their activity. We present the analysis up to May 30, 2018.**

Social Network Analysis and Findings



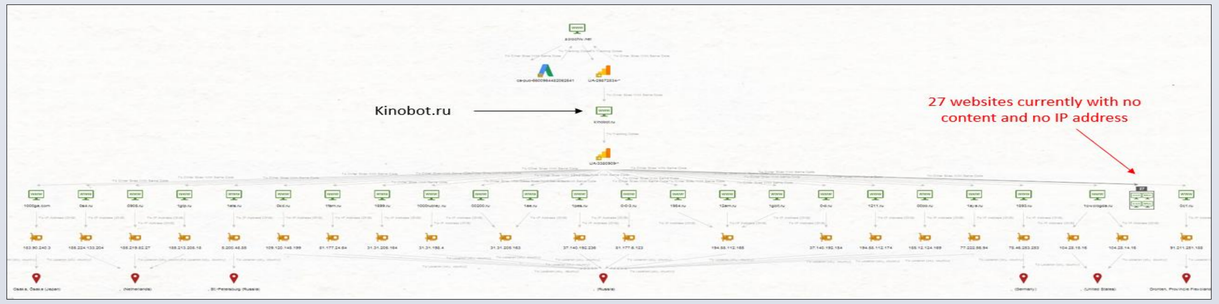
Blogs indicating polarization of domains originating from seed blogs 'lviv.web2ua.com' and 'odessa.web2ua.com'



Blog-Blog network sharing a tracking code

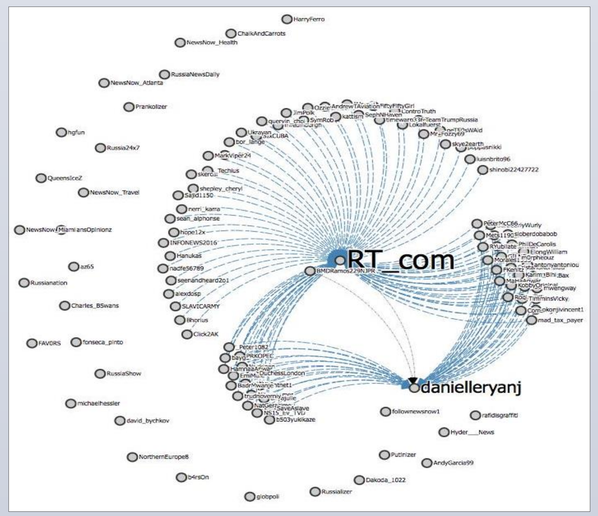
The network (right) shows nodes as blog sites and edges as tracker codes shared between the nodes. Nodes are colored based on their PageRank centrality where red indicates the highest number of in-links to the blog site and green indicates the lowest.

Blog Farm Detection using Social Cyber Forensics



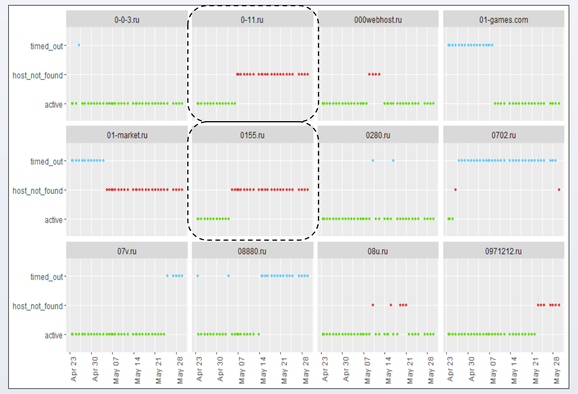
Kinobot network showing multiple arbitrary blogs using same tracking code as Kinobot

RT.com's tweet posting an RT.com article on Twitter to amplify its dissemination in order to delegitimize Reuters' article



Twitter network of the mentions of the RT.com's Swedish TV mast story and its author (@danielleryanj)

Activity Patterns using Ping Test



Ping test results for a sample of blogs that fluctuated over time.

To dig deeper into the activity trends of the suspicious websites (such as active, host-not-found, and timed-out), we analyzed the ping responses for the blog sites. Since these blog sites were inactive most of the time, we assume them to be sleeper cells that stay under the radar until they are activated by their controller.

Acknowledgments

This research is funded in part by the U.S. National Science Foundation (IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2605, N00014-17-1-2675), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock and the Arkansas Research Alliance. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.