The Ever-changing Labyrinth:
A Large-scale Analysis of Wildcard DNS Powered Blackhat SEO

Sites' search rankings is important for to the owners, because to most people, it is the entrance to all sorts of web sites on internet. Blackhat Search Engine Optimization (SEO) has been widely used unethically to promote a website's ranking to gain advantages in search results at low cost by gaming ranking algorithms.

In this paper, the authors carried out a comprehensive study to understand a new type of blackhat SEO infrastructure (called *"spider pool"*). They compare it with other SEO infrastructures: price of buying new or expiring domain is much cheaper nowadays. This is a big advantage over Private Blog Network (PBN) which asking for expensive expired domains with high PR(PageRank) value, and over link exchange service for expensive links. In the meantime, attackers can easily change the underlying link structure, which outperforms forum spam in flexibility. Moreover, since the sites in spider pool are not compromised, they are less likely to be detected and alarmed compared to sites recruited by SEO botnet.

"Spider pools" seeks a different operational model. First the owners of spider pools use cheap domains with low PR values to construct link networks and poison longtail keywords. Then the owners reduce the indexing latencies by search engines. They abuse *wildcard DNS* to create virtually infinite sites and construct complicated loop structure to force search-engine crawlers to visit them relentlessly.

The authors carried out their studies as the following: First they infiltrated a spider pool service and built a detection system to explore all the recruited SEO domains to learn how they were orchestrated. Next, they exploiting the unique features of the spider pool, we developed a scanner which examined over 13 million domains under 22 TLDs/SLDs and discovered over 458K SEO domains. Finally, they measured the spider-pool ecosystem on top of these domains and analyzed the crawling results from 21 spider pools.

The measurement result reveals their infrastructure features, customer categories and impact on search engines. They hope the study could inspire new mitigation methods and improve the ranking or indexing metrics from search engines.

In summary, the authors conducted the investigation on a new type blackhat SEO technique called "spider pool" which abuses wildcard DNS to tamper long-tail keywords of search engines. Based on the understanding through infiltrating a shared spider pool service, they developed a DNS prober which can identify the SEO domains with high accuracy and efficiency, together with a spider pool explorer which is able to excavate the domains used by individual spider pool through seed expansion. The results show that spider pool has become a big threat to registrars, search engines and their users, as more than 458K SEO domains have been discovered, popular sites like amazon.com are abused to promote illegal messages, and long-tail keywords can be easily polluted. The authors think this new threat should be mitigated and call for the attention from the security community.