University of Pavia

Ph.D. School of Electrical and Electronics Engineering and Computer Science

SEMINAR

Protecting Web Servers from Web Robot Traffic
Prof. Derek Doran
Wright State University, USA

November 14th 2014 – h. 10:00
Dept. of Electrical, Computer and Biomedical Engineering
via Ferrata 5, Pavia
Aula Seminari (ex Dip Elettronica) D floor

Abstract

The proportion of traffic that can be attributed to Web robots, which are software agents that automatically submit HTTP requests to Web systems around the world without any human intervention, has risen to staggering levels - from only 20% of all requests just a decade ago to over 60% of traffic across a swath of public Web systems. This sharp rise is precipitated by the proliferation of Web 3.0 technologies that encourage users to share in-the-moment thoughts, observations, and social information, which has virtually only instantaneous value to many organizations. Web servers, however, are not prepared to handle this large (and unlikely to ever decrease) volume of robot traffic. This is because the performance optimizations that Web servers critically rely on to offer high QoS are based on behavioral and statistical expectations about human-generated Web traffic, which robots do not exhibit. Robots may thus be quietly dragging down the performance and scalability of Web servers, clusters, and clouds across the world.

This talk will present our recent work on curtailing the negative effects of Web robot traffic by understanding, preparing, and protecting Web servers from their requests. It will present: (i) an analysis of behavioral and statistical similarities and differences between robots and humans on Web servers across many domains (academic, e-business, digital archive); (ii) a novel approach for detecting Web robot traffic, both offline and in real-time, with high accuracy and low false positive rates; and (iii) preliminary results from an experimental, promising predictive Web caching algorithm to service both human and robot traffic.

Biography

Derek Doran is an Assistant Professor in the Department of Computer Science & Engineering at Wright State University and is affiliated with Kno.e.sis, a state-wide Center of Excellence on Knowledge-Enabled Computing. His research interests are in the analysis of large-scale network and relational datasets, Web mining (focusing on traffic characterization, behavior modeling, and social media analytics), and in the performance modeling of distributed systems. His work across each of these interests has received best conference paper awards/nominations in many venues. Dr. Doran is a National Science Foundation EAPSI Fellow, a past Research Awardee from the Transportation Research Board of The National Academies, a twice summer alumni of Bell Laboratories, and prior to joining Wright State University, was a researcher in the (Big) Data Analytics team at ABB Corporate Research.

Ph.D. Coordinator
Prof. M. CALZAROSSA

Seminar in English.