

---

## - Bachelor/Master Thesis -

### Flow-based Spam Detection

---

<b>CASED</b>	<p>In CASED (Center for Advanced Security Research Darmstadt) collaborate the Technische Universität Darmstadt, Fraunhofer Institute for Secure Information Technology and the University of Applied Sciences Darmstadt in the fast developing field of IT Security. In a unique cooperation, which combines different areas of expertise from these renowned institutions, progressive IT security solutions are researched, developed and implemented into industrial economy: CASED brings together computer scientists, engineers, physicists, legal experts and business economists. Read more on <a href="http://www.cased.de">www.cased.de</a>.</p>						
<b>Motivation</b>	<p>Email spam is one of the central problems of the internet. Spam is defined as the use of electronic message systems to indiscriminately send unsolicited bulk messages. According to estimates, spam emails make 89% - 97% of the total amount of emails per year. Today's approaches to detect email spam usually rely on the analysis of a specific email as-is (i.e. email header and email body) and thus require processing of large data sets, which leads to scaling issues. Furthermore, analysis of email headers and bodies leads to privacy issues as this is highly sensitive data.</p>						
<b>Task</b>	<p>In this thesis, the possibility of using network flow data – i.e. 5-tuples consisting of source and destination IP addresses, source and destination TCP/UDP port numbers, the layer 4 protocol number as well as timing and intensity information – for email spam detection shall be investigated. Using network flow data for this task leads to highly reduced data sets. Furthermore, due to not containing any payload information (e.g. email header and body), using network flow data for spam detection increases the end user's privacy. During this thesis, the student shall work on the whole chain of flow data processing and analysis. For spam detection, appropriate pattern recognition techniques shall be applied.</p>						
<b>Requirements</b>	<ul style="list-style-type: none"><li>• High motivation, creativity and ability to work independently</li><li>• Good communication skills</li><li>• Good programming skills (e.g. C, Perl, Ruby)</li><li>• Good knowledge of internet protocols, especially SMTP</li><li>• Good knowledge of pattern recognition and ML techniques</li><li>• Knowledge of Linux operating system is a plus</li><li>• Very good knowledge of the German or English language</li></ul>						
<b>Start date</b>	Immediately						
<b>Contact</b>	<table><tr><td><b>Sebastian Abt</b></td><td>CASED</td></tr><tr><td><a href="mailto:sebastian.abt@h-da.de">sebastian.abt@h-da.de</a></td><td>Mornwegstraße 32</td></tr><tr><td>06151.16-8416</td><td>64293 Darmstadt</td></tr></table>	<b>Sebastian Abt</b>	CASED	<a href="mailto:sebastian.abt@h-da.de">sebastian.abt@h-da.de</a>	Mornwegstraße 32	06151.16-8416	64293 Darmstadt
<b>Sebastian Abt</b>	CASED						
<a href="mailto:sebastian.abt@h-da.de">sebastian.abt@h-da.de</a>	Mornwegstraße 32						
06151.16-8416	64293 Darmstadt						