Pro-active Data Breach Detection: Examining Accuracy and Applicability on Personal Information Detected

Johnny Botha1, 2, M.M. Eloff1, Ignus Swart2
1Institute for Corporate Citizenship UNISA, Pretoria, South Africa
2CSIR, Pretoria, South Africa
1jbotha1@csir.co.za
1eloffmm@unisa.ac.za
2iswart@csir.co.za

Abstract:

Data breaches remain a common occurrence affecting both companies and individuals alike, despite promulgated data protection legislation worldwide. It is unlikely that factors causing data breaches such as incorrect device configuration or negligence will stop unless effective enforcement of relevant legislation is applied. While several information privacy regulators exist, the dominant norm is to respond reactively on reported incidents. Reactive response is useful for cleaning up detected breaches but does not provide a clear indication of the level of personal information available on the internet since only reported incidents are taken into account. The possibility of pro-active automated breach detection has previously been discussed as a capability augmentation for existing privacy regulators. By pro-actively detecting leaked information, detection times can potentially be reduced to limit the exposure time of Personal Identifiable Information (PII) on publicly accessible networks. At present the average time for data breach detection is in excess of three months internationally and breach discovery it most often not by the data owner but an external third party increasing exposure of leaked information. The duration of time that data is exposed on the internet has severe negative implications since a significant portion of information disclosed in data breaches have been proven to be used for cybercrime activities. It could then be argued that any reduction of data breach exposure time should directly reduce the opportunity for associated cyber-crime. While pro-active breach detection has been proven as potentially viable in previous work, numerous aspects of such a system remain in question. Aspects such as legality, detection accuracy and communication with affected parties and alignment with privacy regulator operating procedures are all unexplored. The research presented in this paper considers the results obtained from two iterations of such an experimental system that was conducted on the South African .co.za domain. The first iteration conducted in early 2014 was used as a baseline for the second iteration that was conducted one year later in 2015. While the experiment was conducted on the South African cyber domain, the concepts are applicable to the international environment.