Minimising Collateral Damage:
Privacy-Preserving Investigative Data Acquisition Platform

Zbigniew Kwecka, Edinburgh Napier University, UK

William J. Buchanan, Edinburgh Napier University, UK

ABSTRACT

Investigators often define invasion of privacy as collateral damage. Inquiries that require gathering data from third parties, such as banks, Internet Service Providers (ISPs) or employers are likely to impact the relationship between the data subject and the data controller. In this research a novel privacy-preserving approach to mitigate collateral damage during the acquisition process is presented. This approach is based on existing Private Information Retrieval (PIR) protocols, which cannot be employed in an investigative context. This paper provides analysis of the investigative data acquisition process and proposes three modifications that can enable existing PIR protocols to perform investigative enquiries on large databases, including communication traffic databases maintained by ISPs. IDAP is an efficient Symmetric PIR (SPIR) protocol optimised for the purpose of facilitating public authorities' enquiries for evidence. It introduces a semi-trusted proxy into the PIR process in order to gain the acceptance of the general public. In addition, the dilution factor is defined as the level of anonymity required in a given investigation. This factor allows investigators to restrict the number of records processed, and therefore, minimise the processing time, while maintaining an appropriate level of privacy.

Keywords: *Data Mining, Data Retrieval, Investigative Data Acquisition Platform, Privacy Enhancing Technology, Private Information Retrieval Protocols*

INTRODUCTION

Those who would give up essential Liberty, to purchase a little temporary safety, deserve neither Liberty nor Safety (Benjamin Franklin, 11 Nov 1755).

Since the September 11, 2001 many western governments have passed laws empowering public authorities with wider rights to gather operational data (Home Office, 2009; Swire & Steinfeld, 2002; Young, Kathleen, Joshua, & Meredith, 2006). For many years public opinion accepted the invasion of personal privacy rights as the sacrifice needed to *fight the terror* (Rasmussen Reports, 2008). However, slowly, public opinion is shifting back to a state where such measures are considered unacceptable. This is shown by public opinion surveys, such as the one conducted by Washington Post in 2006 (Balz & Deane, 2006), where 32% of respondents agreed that they would prefer the

DOI: 10.4018/jitsa.2011070102

federal government to ensure that privacy rights are respected rather than to investigate possible terrorist threats. This was an 11% increase from the similar survey conducted in 2003.

In the UK, the public authorities, including Police, request investigative data from third-parties on regular basis (Information Commissioner, 2008) and the data protection legislation allows for such requests, even without warrants (European Parliament, 1995; Home Office, 2007). Depending on the way these requests are performed, human and natural rights of the data-subject can be breached, and/or the investigation can be jeopardized (Kwecka, Buchanan, Spiers, & Saliou, 2008). A recent proposal by the UK government went further and recommended allowing the public authorities direct access to data held by Content Service Providers (CSPs), such as mobile telephony providers and Internet Service Providers (ISPs) (Home Office, 2009). According to the public consultation document, there were a few major motivating factors behind this proposal, these included: increasing access speeds to records; allowing for covert enquiries by anti-terror and national security agencies; lowering collateral damage to potential suspects under investigation; and enabling the analysis of data to facilitate the profiling of terrorists activities. In response, concerns were raised that if the proposal was implemented, it would thwart the privacy of Internet users around the globe, in order to increase the security of one nation. This research shows that most of the objectives set out in the proposal can still be achieved while maintaining high level of privacy. It is shown that an investigative system can maintain the privacy of the data subjects and also preserve the confidentiality of investigations. However, both security and privacy must be built into the system at the design stage in order to achieve this (Swire & Steinfeld, 2002).

This paper gives an insight into use of Privacy Enhancing Technologies (PETs) in improving the current investigative data acquisition practices. The structure of this manuscript closely follows the methodology used to draw the final conclusions. First we provide a background to investigative data acquisition and to various privacy-preserving approaches to information retrieval. The analysis of the related research is presented and identifies an existing protocol, Private Equi-join (PE) that can facilitate efficient private database searching and information retrieval. It is also shown that the complexity of this protocol is lower that complexity of similar approaches, and for these reason the PE protocol is chosen as the base for the investigative data acquisition solution. This protocol is described and other commonly used privacy-preserving primitives that can be reused in order to build a platform suitable for investigative enquiries and the design considerations are discussed. The PE protocol is evaluated against the requirements derived from the literature described earlier. Finally, we describe the novelty of this paper – which are the three improvements needed to form an Investigative Data Acquisition Platform (IDAP) based on the PE protocol and the evaluation of IDAP is provided. IDAP is an efficient approach to maintain the secrecy, preserving the suspect's privacy and gaining the public's support for the PET technologies in digitalised investigative enquiries. The improvements include the introduction of the *dilution factor*, which is a numeric value that specifies the level of anonymity required for an investigation. Resultantly, the identity of a potential suspect can be hidden in a group smaller than the population, which permits the use of PIR systems with large databases. The second improvement is the technique for building complex privacy-preserving queries without affecting the complexity of the protocol. Finally, a semi-trusted proxy is introduced to the PE protocol, in order to ensure information theoretic privacy of data-subjects that are not the potential suspects.

BACKGROUND AND RELATED WORK

The background section is split into two parts. First the nature of public authorities' investi-