Institution: Oxford Brookes University

Unit of Assessment: 11 - Computer Science and Informatics

Title of case study: Reducing fraud: Helping one of the world's leading internet registries to understand typosquatting and improve abuse detection

1. **Summary of the impact** (indicative maximum 100 words)

Database and URL hijacking is a very real and damaging threat for businesses and their brands. Professor David Duce and Dr Faye Mitchell successfully partnered with Nominet, a leading internet domain registry, to help detect abuse of their WHOIS system and develop tools to better understand and deal with typosquatting. Their approach enabled improvements to Nominet's information services and practices, whilst also influencing the wider technical community. These benefits included better policing of systems, securing brands, reducing fraud and starting to get people thinking about what can be done with data to gain insights and understanding of behaviours.

2. Underpinning research (indicative maximum 500 words)

With research expertise in web technology, visualisation, data mining and digital forensics, Professor David Duce and Dr Faye Mitchell of Oxford Brookes University, worked with Nominet (2007-2009) as part of a Knowledge Transfer Partnership (KTP)[1]. As the internet registry for .co.uk domain names, internet security is paramount for Nominet. The partnership aimed to develop technologies that would detect abuse of Nominet's information services - the WHOIS database. The scope of the project was then extended to include measurement of the incidents of "typosquatting". Two KTP associates were employed to work on the project; Oliver Buckley, supervised by Prof. David Duce concentrated on the data visualisation aspect and Dr Alessandro Linari, supervised by Dr Faye Mitchell, concentrating on the knowledge discovery aspect.

Typosquatting is the practice of registering domain names similar to well-known brands with intent to profit from the confusion, relying on errors being made by web users when typing web addresses. The work is based on the notion of a syntactic neighbourhood around a domain name. Other names falling within this neighbourhood are considered to be potential typosquats. This depends on finding an effective distance measure between domain names in order to characterise this neighbourhood. The original research in the KTP project explored a variety of distance measures from the literature and applied these to the .uk domain register.

Methods explored included;

- Edit distance with transpositions - the minimum number of insertions, deletions, substitutions of a single character and transpositions of a pair of characters needed to transform one string into another,
- Keyboard distance - derived from edit distance but giving higher importance to operations corresponding to frequent typing errors.

In addition a visual similarity measure was also used, an extension of a method previously published by Black. Experiments were carried on a snapshot of third level domains in the .co.uk registry. The snapshot, taken in March 2008 contained six million entries; the initial experiments were performed on the 1000 most popular domains[2]. The initial work extended previously published work by others in this area by considering other distance measures and having access to the full content of a real register. The work has now been enhanced for production use by developing an architecture and platform that enable searches to be carried out much more efficiently.

In order to develop ways to detect abuse of the WHOIS system, the work included an analysis of the data looking for past abuses and the means to characterise them. The project developed a methodology for anomaly detection and characterisation based on the analysis of information contained in application logs. Some bespoke visualizations were developed to aid detection of patterns in sources querying particular domains, the range of domains queried by particular sources and temporal variations in querying behaviour. These enabled various anomalies in overall volumes of queries to be analysed. A framework was produced that analysed the data; able to

generate on demand reports to indicate any suspicious behaviour in the WHOIS queries and could deal with the large number of queries received, approximately 500,000-750,000 queries per day[4,5,6].

3. References to the research (indicative maximum of six references)

The research within the KTP was informed, in part, by two EPSRC-funded grants:

- GR/R96224/01 (2003-2005) Visualization Middleware for e-Science. Award value: £56,275. Principal Investigator: Professor David Duce. http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=GR/R96224/01
- GR/S68514/01 (2004-2007) OPEN OVERLAYS: Component-based Communication Support for the Grid. Award value: £214,614. Principal Investigator: Professor David Duce http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=GR/S68514/01

1. KTP006502 Dr Faye Mitchell, Prof. David Duce, 'To create a system based on techniques of knowledge discovery, data mining and information visulisation, to support detection of fraudulent usage of the company's domain name registry' with Nominet UK August 2007 to August 2009 £124,222. http://info.ktponline.org.uk/action/details/partnership.aspx?id=6502

2. Linari, A. Mitchell, F., Duce, D. and Morris, S.*Typo-Squatting: The "Curse" of Popularity*. In Proc. of the WebSci'09: Society On-Line, 18-20 March 2009, Athens, Greece. Available at: http://journal.webscience.org/187/. (Poster presentation of 5 page paper.)

3. Linari, A. *The incidence of Typo-squatting in the .uk Registry*. 18th CENTR Technical Workshop, 4 May 2008. Available at http://www.centr.eu/main/4326-CTR/version/default/part/AttachmentData/data/Tech18%20-%20Linari%20-%20typosquatting-uk-registry3.pdf (CENTR is the Council of European National Top Level Domain Registries. This workshop was a specialist workshop attended by the domain registry community.)

4. Linari, A. *Abuse Detection Programme At Nominet*. 17th CENTR Technical Workshop, 21 October 2007, Amsterdam, the Netherlands. Available at http://www.centr.eu/main/3832-CTR/version/default/part/AttachmentData/data

5. Morris, S., Buckley, O., & Linari, A. (2009), *Nominet'sWhois Botnet Attack. 20th CENTR Technical Workshop*, 3 May 2009, Amsterdam, the Netherlands. Available at http://www.centr.eu/main/5028-CTR/version/default/part/AttachmentData/data/Tech20%20-%20Moris%20-%20Buckley%20-%20Linari%20-%20Botnet%20Presentation%20[Compatibility%20Mode].pdf

6. Buckley, O., Duce, D. and Morris, S. *The When, Where and Who of Visualising WHOIS Data*, poster presentation at EuroVis 2009 conference.

4. Details of the impact (indicative maximum 750 words)

The project, informed by the expertise of Duce and Mitchell, enabled Nominet to offer its clients improved performance, service and practice, through understanding the prevalence and characteristics of typosquatting and methods to detect abuse of the WHOIS system.

Furthermore, the research has enabled benefits that are not only immediate to Nominet but also to their industry through influencing their practice. Of further significance are the on-going benefits to Nominet's clients, who are better able to police their brands, and their customers through a reduction in losses due to fraud. In 2009, the number of .uk registered domain names passed 8 million increasing to 9.7 million in 2011 [7].

Importance of measuring typosquatting; improved policing and improved services

The typosquatting work made Nominet realise that identifying appropriate geometries to

characterise various aspects of the Internet is a potentially fruitful line of research. Based on this initial project, a new line of research is being explored by Nominet, looking at developing algorithms for analysing the registry that may lead to new and improved products and services for Nominet's customers.

One of the direct outputs of the partnership was a prototype of an extension to Nominet's Public Register Search Service (PRSS) query system[7], which can conduct a typosquatting check. The PRSS system is used, amongst other things, to assist in establishing or defending intellectual property rights and other similar matters. The extension allows companies to better police their brands, so reducing the amount lost to fraud.

The work showed the potential to resolve problems experienced not only by Nominet but also by other national domain name registries. It generated sufficient interest to warrant presentations at industry workshops under the auspices of the Council of European National Top-Level Registries (CENTR; May 2008)[3], and the DNS Operations and Analysis Research Centre (OARC; June 2008)[8].

Database protection; increasing security, confidence and efficiency on an international stage

The framework[4,5,6] developed for WHOIS[10] was turned into a production system to issue daily reports system to the Nominet operations team. This reduced the amount of work needed to detect abuse, increased the sensitivity of the detection of abuse and enabled the detection of distributed query systems which would have previously been undetected. The anti-abuse system has helped to raise the profile of Nominet in the international registry and internet community through presentations at the international CENTR Technical workshop[4].

Research carried out by Nominet showed that consumers see .uk as a trusted environment[11]. For businesses, the report highlighted that a .uk web address is an important asset when doing business in the UK with UK consumers and businesses. For consumers, a .uk address is an important consideration when making online purchases. Incidents of data loss can lead to a significant loss of confidence in the organisation concerned. In Nominet's case, the impact of a well-publicised loss of a large amount of data via abuse of its information systems could affect the trust in .uk domains names. This would affect not only Nominet but other businesses too. Such an impact is difficult to estimate, but from Nominet's perspective, should it lead to as little as a 0.5% drop in domain name registrations and renewals, turnover would be affected to the tune of £100,000 (0.5% of sales turnover).

Through the project, Nominet detected an attack at an early stage and managed to trace this to an organisation based in the US[5]. With the evidence gathered through the project, a legal counsel was rapidly able to obtain a promise for them to desist without the need to resort to very expensive legal procedures. Nominet were subsequently contacted by the Austrian registry who were having a similar problem and their legal team were helping them in their action against the same company. Nominet estimated that had they needed to go to court over this, they could have incurred expenditure in the region of £250,000[11,12]. The anticipated savings to Nominet come in the avoidance of costs that would be required if attacks on the WHOIS service and other information systems were to continue and grow.

5. Sources to corroborate the impact (indicative maximum of 10 references)

7. NominetAnnual Report and Accounts 2009 & 2011http://www.nominet.org.uk/sites/default/files/57812_report-and-accounts2009.pdf , http://www.nominet.org.uk/sites/default/files/nominet_report_and_accounts_2011.pdf

8. Public Register Search Service, Nominet.co.uk (http://www.nominet.org.uk/disputes/public-register-search-service)

9. DNS Operations and Analysis Research Centre 2008 Workshop agenda https://www.dns-oarc.net/dns-operations/workshop-2008/agenda 'A statistical approach to typosquatting detection'

10. WHOIS database http://www.nominet.org.uk/uk-domain-names/about-domain-names/domain-lookup-whois/whois-tool

11. '.co.uk – A great place to be' (http://www.agreatplacetobe.co.uk/)

12. KTP006502: Final Report (available from Oxford Brookes University Research and Business Development Office)

13. Corroborative contact 1: Senior Legal Counsel, Nominet