a given user-preference \( k = 3 \), \( s \) will not be able to distinguish \( u \)'s request from requests made by \( k-1 \) arbitrary other users \( u' \) and \( u'' \). Under reasonable assumptions about the scope of IIN, we show that \( s \) can know \( u \)'s location only within \( p_u \), even while \( u \) is moving. Particularly, the TVM algorithm operates in two phases as outlined below.

In Phase 1 of TVM, \( u \) computes a \( k \)-Anonymity Bloom (kAB) filter structure, which provides location privacy for snapshot localization tasks. When \( u \) needs continuous localisation (e.g., as \( u \) moves), the kAB of Phase 1 itself is not adequate to preserve the privacy of \( u \), since by issuing \( k \) independent requests, \( s \) can realise by exclusion that there are \( k - 1 \) invalid requests (as one of the requests will always relate to the real building A). This allows \( s \) to deterministically derive \( u \)'s real location.

To circumvent the above problem, in Phase 2 of TVM, \( u \) uses the bestNeighbors algorithm to issue a set of camouflaged localisation requests that follow a similar natural movement pattern to that of \( u \) (i.e., dotted circles in Figure 1). This provides the illusion to \( s \) that there are \( k \) other possible users moving in space, thus camouflaging \( u \) among \( k \) other users. Since our TVM algorithm transfers only a partial state of the database from \( s \) to \( u \), it requires less network traffic and smartphone-side energy than current privacy-aware approaches that transfer the complete database to \( u \) prior the localisation task. TVM is resilient to the (i) linking attack: the only uniquely identifying attribute is the fingerprint of a user's location. In fact, this is also the only attribute sent by the user to the server, therefore there are no other attributes that could link to the user's fingerprint value; and (ii) the homogeneity attack: there is an inherent diversity in the resulting \( k \)-anonymous set of TVM, since it uses hashing to generate a set of unique access point MAC values that has a uniform distribution over all values, and therefore, no information can leak due to lack of diversity in the sensitive attributes.

**References:**


**Please contact:**

Demetrios Zeinalipour-Yazti
University of Cyprus
+357 22 892755
dzeina@cs.ucy.ac.cy

---

**Social Fingerprinting - or the Truth About You**

*by Stefano Cresci, Marinella Petrocchi, Maurizio Tesconi (IIIT-CNR), Roberto Di Pietro (Nokia Bell Labs), and Angelo Spognardi, (DTU)*

_**Inspired by biological DNA, we model the behaviour of online users as “Digital DNA” sequences, introducing a strikingly novel, simple, and effective approach to discriminate between genuine and spambot online accounts.**_

Modelling the behaviour of online users, as well as analysing their properties, is of primary importance for a broad variety of applications - for example, to mine substantial information about events of public interest. Secondly, online behavioural analysis can be applied to make predictions: linking behaviours to some kind of ground truth in the past leads to predictions of what will likely happen in the future when similar behaviours take place.

Here, we consider online behavioural analysis as a means to detect fictitious and automated accounts, which distribute unsolicited spam, advertise events and products of doubtful legality, sponsor public characters and, ultimately, lead to a bias in public opinion and harm social relationships. Spambot detection is thus a must for the protection of cyberspace, in terms of both threats to users’ sensitive information and trolls that may want to cheat and damage them. Unfortunately, new waves of malicious accounts present advanced features, making their detection with existing systems extremely challenging [1].

Inspired by biological DNA, we propose to model online user behaviour with strings of characters representing the sequence of a user’s online actions [2]. Each kind of action (e.g., posting new content, following or replying to a user) can be encoded with a different character, in a similar manner to the bases of DNA sequences. According to this paradigm, online user actions represent the bases of their ‘digital DNA’.

Digital DNA is a flexible way of modeling the different kinds of user behaviour that are observed on the internet. Its flexibility lies in the ability to choose which actions will form the sequence. For example, digital DNA sequences on Facebook could include a different base for each user-to-user interaction type: comments, likes, shares and mentions.

Like its biological namesake, digital DNA is a compact representation of information. For example, the timeline of a Twitter user could be encoded as a single string of 3,200 characters (one character per tweet).

In contrast with the supervised spambot detection approaches largely used in recent years, we have devised an unsupervised way to detect spambots by comparing their behaviour with the aim of finding similarities between automated accounts. We model the behaviour of spambots via their digital DNA and we compare it to that of genuine accounts.
We exploit digital DNA to study the behaviour of groups of users following the intuition that, because of their automated nature, spambots are likely to share more similarities in their digital DNA than will a group of heterogeneous genuine users.

This process is called digital DNA fingerprinting and encompasses four main steps: (i) acquisition of behavioural data; (ii) extraction of DNA sequences; (iii) comparison of DNA sequences; (iv) evaluation. First, we create datasets of verified spambots and genuine Twitter accounts. Then, we extract the digital DNA of the accounts; that is, we associate each account to a string that encodes its behavioural information.

Successively, we study similarities among the DNA sequences of our accounts. We consider similarity as a proxy for automation and, thus, an exceptionally high level of similarity among a large group of accounts serves as a red flag for anomalous behaviours. In particular, we quantify similarity by looking at the Longest Common Substring (LCS) among digital DNA sequences. We show that the similarity, as measured by the LCS, between the DNA sequences of spambots is much higher than that of genuine accounts, and we leverage this distinctive feature to perform our spambot detection. Finally, we compare our spambot detection results with those of other state-of-the-art approaches.

Results show that our proposed technique outperforms best-of-breed algorithms that are commonly employed for spambot detection [2]. In addition, most of those state-of-the-art approaches require a large number of data-demanding features, as shown in [3]. Instead, our digital DNA fingerprinting technique on Twitter only exploits time-line data to perform spambot detection, thus being both effective and efficient. By relying on digital DNA, analysts can leverage a powerful set of tools that have been developed over decades for the analysis of biological DNA to validate their working hypotheses on online user behaviour.

Link: http://milb.projects.iit.cnr.it

References:

Please contact:
Marinella Petrocchi, IIT-CNR, Italy
+390503153432
marinella.petrocchi@iit.cnr.it

Flexible Decentralised Access Control using Invitation-Response Dialogue

by Arthur Melissen (Coblue Cybersecurity)

Distributed role-based access control (RBAC) has become a standard for decentralised systems to manage authorisation across networks. While this model is effective at providing authorization, it fails in providing the flexibility and authorisation accountability that organisations require today. We present an extension to standard distributed RBAC mechanisms by adding an invitation and response dialogue in the assignment of roles to entities for distributed resources, such as collections of shared files. This approach offers more flexibility for delegating roles across administrative domains and increases transparency and confidence in the authorisation structure of distributed resources.

Traditionally, decentralised systems have used distributed access control and simple public key infrastructure (SPKI) [1] mechanisms to manage which entities are provided access to a given resource. Authorisation for a resource is described using an access control list (ACL). Each entry in the ACL describes a set of entities and their assigned roles and authorisation characteristics, such as read and write permissions.

An entity that has a sufficiently high authorisation is able to modify the ACL itself and control the authorisations of other entities. Granting authorisations to other entities in the system is called delegation. We shall refer to entities that have the power to delegate authorisation to other entities as administrators of a resource.

Typically, existing SPKI systems delegate authorisation by letting administrators create a new entry in the ACL and signing it with an administrator’s cryptographic key. The authorisation describes the role granted and a set of entities by their public key. The updated ACL is then distributed across all relevant entities in the network and the entities are made aware of the new authorisation.

While this approach works well, it lacks the flexibility and accountability that is often desired in a modern distributed networking context. We found several areas for improvement:

First, it is possible that the entity desires no authorisation at all for a given resource. In a distributed file system, an