A Behavioural Model for Consumer Reputation

Anirban Basu, Ian Wakeman and Dan Chalmers

Department of Informatics, University of Sussex

{A.Basu, ianw and D.Chalmers}@sussex.ac.uk

Introduction

Future interactions between a client (service consumer) and a server (service provider) in a networked environment are often unaffected by past observations or any form of local or global behavioural history. Instead, filtering of network connections or variations of service levels are dependent on other techniques such as content filtering in case of messaging systems (e.g., email). In an effort to inform policy decisions for future interactions, some work has been done towards developing an architecture for large-scale sharing of behavioural history [ABP05]. Others [GKF+06, GH04] have proposed use of reputation mechanisms to combat the problem of email spam.

Research question: We explore the research question: *can a local and a global reputation scheme based on behavioural history of long-lived network identities be used to implement policies for future network interactions?*

Assuming that a long-lived identity infrastructure is in place, we propose a reputation scheme based on behavioural history of such identities. In many scenarios, network identities are either short-lived or anonymous. A proposal for developing long-lived identities using group memberships is discussed in [WCF07]. When identity of the client is anonymous or short-lived, a fall-back option, such as best effort service, will be used.

Acceptable behaviour model

There needs to be means for defining "acceptable behaviour" for network actors. Service contracts, which include Acceptable Use Policies and Service Level Agreements, provide good starting points but these are usually legal agreements with vague technical terms. We are exploring the possibilities of identifying technical terms. We are developing an *acceptable behaviour model*, which is a mapping of technical terms to notions of good or bad behaviour through a logic-based formalism, such as Event Calculus [KS86]. This formalism will help quantising good or bad behaviour in accordance with service contracts.

Local reputation

We have defined a local reputation response to change in quantised behaviour. We have experimented with some mathematical models to best represent the expected reputation response. We will use the terms *score* and *rank* to denote reputation of a consumer and of a provider respectively. Let us denote score variable with r; consumer behaviour variable with b; positive score saturation with r_{psat}; negative saturation with r_{nsat}; and two adjustable response parameters λ and μ. Also, for any event (v) for which a change of behaviour is noted, the corresponding cumulative behaviour is b_v and the corresponding reputation is r_v. Further p and n suffixes will signify positive and negative respectively.

The equation for good reputation getting better with good behaviour is:

$$r = r_{psat}\left(1 - e^{-\lambda b}\right) \quad \text{for} \quad \Delta b > 0,\, b > 0,\, r_{v-1} \geq 0 \tag{1}$$

and the equation for bad reputation getting worse with bad behaviour is:

$$r = r_{nsat}\left(1 - e^{\lambda b}\right) \quad \text{for} \quad \Delta b < 0,\, b < 0,\, r_{v-1} \leq 0 \tag{2}$$

and the equation for good reputation (r_{v_p}) getting worse with bad behaviour is:

$$\begin{aligned} r = \frac{r_{v_p}}{b_{v_p}} b \quad &\text{for} \quad \Delta b < 0,\, b > 0,\, r_{v-1} > r_v \geq 0 \\ &\text{and} \quad r_{v_p} = r_{psat}\left(1 - e^{-\lambda b_{v_p}}\right) \end{aligned} \tag{3}$$

and the equation for bad reputation (r_{v_n}) getting better with good behaviour is:

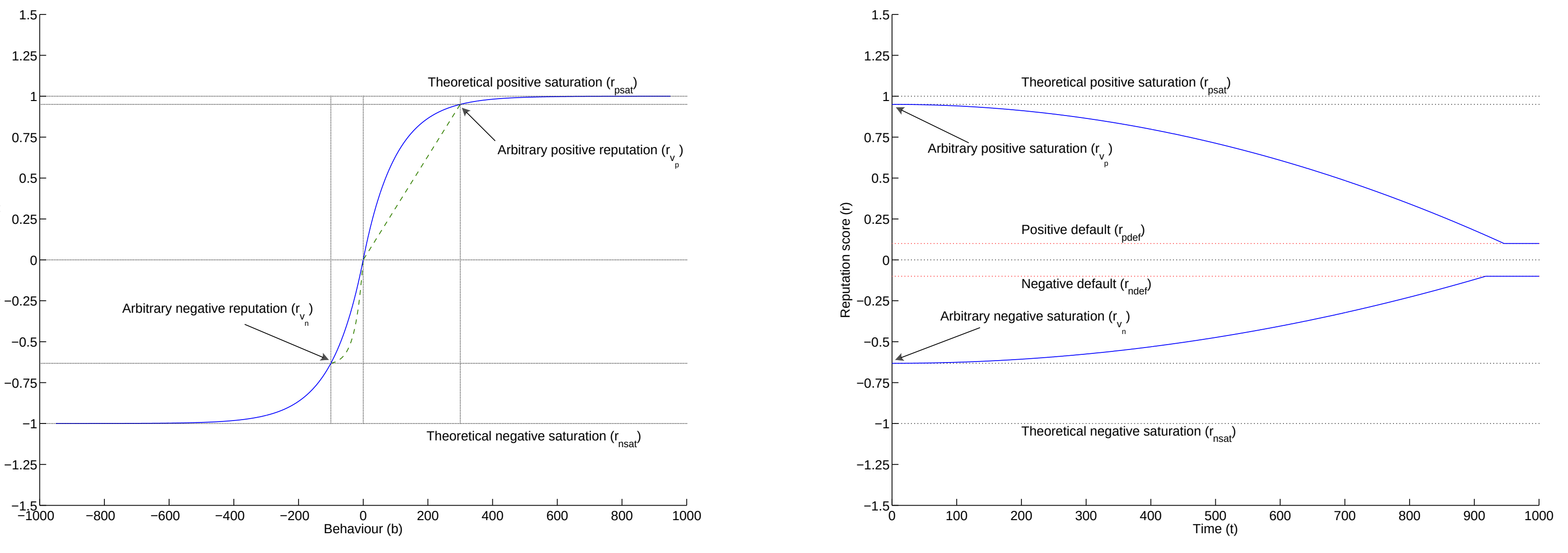
$$\begin{aligned} r = \frac{r_{v_n}}{\left(1 - e^{\mu b_{v_n}}\right)}\left(1 - e^{\mu b}\right) \quad &\text{for} \quad \Delta b > 0,\, b < 0,\, r_{v-1} < r_v \leq 0 \\ &\text{and} \quad r_{v_n} = r_{nsat}\left(1 - e^{\lambda b_{v_n}}\right) \end{aligned} \tag{4}$$

Time decay: Saturated reputation indicates either "too good" or "too bad" values. Therefore, a decay with no activity over time helps a saturated bad reputation to recover; and it also questions a saturated good reputation. A neutral zone $[r_{ndef} \quad r_{pdef}]$ (positive and negative default) is defined for this purpose. Positive reputation higher than r_{pdef} decays to positive default, while negative reputation value lower than r_{ndef} increases to negative default. An adjustable decay rate parameter ϵ is introduced in this context. The equation for positive reputation decaying over time is given as:

$$r = \begin{cases} r_{v_p}\left(1 - \epsilon t^2\right) & \text{for } r \geq r_{pdef} \\ r_{pdef} & \text{for } r < r_{pdef} \end{cases} \tag{5}$$

and the equation for negative reputation increasing over time is given as:

$$r = \begin{cases} r_{v_n}\left(1 - \epsilon t^2\right) & \text{for } r \leq r_{ndef} \\ r_{ndef} & \text{for } r > r_{ndef} \end{cases} \tag{6}$$



(a) (b)

Figure 1: *Part (a): local reputation response to behaviour; part (b): the time decay of local reputation.*

Global reputation

The local scores are aggregated to develop a global score. The global score can be stored in an infrastructure with single administrative control. The global score is calculated, at query time, over a window of previously submitted local scores. If local reputation score from provider i is denoted by r_{local_i}; rank of the provider is $rank_i$; reputation to behaviour response parameters are λ_i and μ_i; time at which the reputation score is reported is t_{report_i}; and time at which global score is calculated is t. Calculated over a window of n submitted local scores, the i^{th} component of the global score is given as:

$$r_{global_i} = \begin{cases} r_{local_i} rank_i \left(1 - \lambda_i \epsilon_i (t - t_{report_i})^2\right) & \text{for } 0 \leq r_{local_i} \leq 1 \\ r_{local_i} rank_i \left(1 - \mu_i \epsilon_i (t - t_{report_i})^2\right) & \text{for } -1 \leq r_{local_i} \leq 0 \end{cases} \tag{7}$$

and thus, the global score is given as:

$$r_{global} = \frac{\sum_{i=1}^{n} r_{global_i}}{n} \tag{8}$$

If the requesting provider j has already submitted its own component of global score in the past then the j^{th} component is ignored; hence:

$$r_{global} = \frac{\sum_{i=1}^{n} r_{global_i}}{n-1} \quad \text{where} \quad i \neq j \tag{9}$$

We are investigating use of other statistical measures, such as standard deviation or distributions, along with the weighted average to detect inconsistencies in the global components of the score. This, in turn, forms a defense mechanism for certain attacks on the model.

Global score aggregation: The global score is reported to a *score aggregation system* (SAS) is described by a six step process as follows.

1. At the start of service provision, provider (P) requests authorisation from consumer (C) to look up C's global score (r_{global_C}) stored in the SAS
2. C sends authorisation token (AT) to P, which also contains the permit to report a score for C. In addition, C notifies SAS that AT has been created.
3. P provides service to C and makes local observations. If a service is continuous, P can submit scores several times but for each submission a new AT is required.
4. At any time, P can send the local score to SAS for aggregation.
5. SAS contacts C (or its agent) with an optional requirement to submit its assessment for P (rank of P). If C declines to comment or is unavailable, SAS will assume a value 1 (highest) for P's rank.
6. SAS updates the rank for P only if C's global score at that point (prior to the current aggregation) is positive. SAS aggregates C's score and invalidates AT.

Provider ranking is intuitive at the moment, such as "did I like (range: $(0 \quad 1]$) the service I was provided?". We are investigating if this can be formalised. Storage and calculation of provider ranks can also be done over a resizable window of submitted ranks.

Implementation and simulation (future work)

Mathematical validation of the model is being achieved through the use of differential calculus. This will be followed by the implementation. The simulation of the model will be done based on any available real world input data (e.g., The Internet Traffic Archive[1]) as well as synthetic data that represent the full spectrum of users with behaviour between fully malicious and fully non-malicious. The results will illustrate how well the proposed model can act as a security measure augmented with existing policies to protect unsolicited transactions over a network. We expect that consumers having accidental and occasional short spells of bad behaviour but generally good behaviour otherwise should not have their reputation badly affected. However, consumers consistently behaving bad will have their service levels drop to minimum or be cut off.

We are also interested in an experiment to use such consumers scores as an incentive mechanism in a peer-to-peer content distribution system. In addition to this, we will simulate a variety of attacks and check our model for defense against such attacks.

References

[ABP05] M. Allman, E. Blanton, and V. Paxson. An Architecture for Developing Behavioral History. *Proc. Workshop on Steps to Reducing Unwanted Traffic on the Internet*, 2005.

[GH04] J. Golbeck and J. Hendler. Reputation Network Analysis for Email Filtering. *Proceedings of Conference on Email and Anti-Spam (CEAS)*, 2004.

[GKF+06] S. Garriss, M. Kaminsky, M.J. Freedman, B. Karp, D. Mazières, and H. Yu. Re: Reliable Email. *Proceedings of the 3rd Symposium of Networked Systems Design and Implementation (NSDI '06)*, 2006.

[KS86] R.A. Kowalski and M.J. Sergot. A Logic-based Calculus of Events. *New Generation Computing*, 4(1):67–95, 1986.

[WCF07] Ian Wakeman, Dan Chalmers, and Michael Fry. Reconciling privacy and security in pervasive computing: The case for pseudonymous group membership. *Submitted for publication*, June 2007.

[1] http://ita.ee.lbl.gov/

IWSOS 2007, International Workshop on Self-Organizing Systems, 11-13 September 2007, The Lake District, UK