ON OPTIMAL DECISIONS IN AN INTRODUCTION-BASED REPUTATION PROTOCOL

Richard Al-Bayaty and O. Patrick Kreidl

University of North Florida
School of Engineering
Jacksonville, Florida 32224

ABSTRACT

Consider a network environment with no central authority in which each node gains value when transacting with behaving nodes but risks losing value when transacting with misbehaving nodes. One recently proposed mechanism for curbing the harm by misbehaving nodes is that of an introduction-based reputation protocol [1]: transactions are permitted only between two nodes who consent to being connected through introduction via a third node. This paper models the main decision process in this protocol, namely that of continuing/closing an active connection, as a sequential detection problem in which each stage corresponds to a transaction that is (perhaps erroneously) classified as either benign or harmful. It is shown that the optimal decision takes the form of a reputation threshold policy, the exact threshold determined by a Bellman equation that admits a tractable iterative solution.

Index Terms— Reputation Systems, Trust, Ratings, Sequential Detection, Dynamic Programming

1. INTRODUCTION

A growing portion of daily Internet commerce incentivises parties involved in any particular transaction to rate one another, so that ratings of a given party in previous transactions can be leveraged to decide whether or not to transact with that party in the future. Well-known examples include the feedback forum on the auction website "Ebay," the user-based moderation system on the discussion forum "Slashdot" and the partly crowdsourced website rating tool within the browser plug-in "Web-of-Trust." While these systems may aggregate ratings and transform them into reputations differently [2], they share an intent to use the reputation-based signaling of trustworthiness for discouraging interactions with parties that repeatedly misbehave and, in the long run, have a positive effect on quality-of-service for the behaving parties.

In peer-to-peer networks for which a central reputation authority becomes infeasible, a recently proposed scheme (described in [1] for secure Internet packet routing) is a so-called introduction-based approach. Fundamental to such an approach is that transactions are allowed only between two parties that are *connected*, where both parties consent to the connection through an introduction sequence involving a third party. In other words, as illustrated in Fig. 1, there are three parties, or nodes, involved in every introduction sequence: the *requester* is the node who initiates the sequence, the *target* is the node to which the requester wishes to be introduced and the *introducer* is the node connected to both requester and target who is asked to make the introduction. The introducer may or may not offer the introduction (based on its reputations of requester and target) and the target may or may not accept the introduction (based on its reputation of the introducer), but if offered and accepted then the connection between requester and target is established and the two nodes can transact and/or request introductions to others. The connection exists indefinitely until the requester or target elects to close it, and each provides feedback to the introducer over the lifetime of the connection. Note that, depending on the state of all nodes' connections, forming a new connection may require multiple consecutive introductions; moreover, it is also assumed that the network initializes with every node having at least one a-priori connection in place.

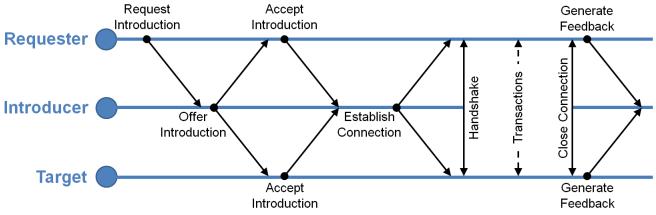

Fig. 1. Sequence Diagram of a Successful Introduction

The described introduction-based protocol assumes a constructive interplay among several different decision processes within the different roles e.g., whether a requester or target closes a connection, whether a target accepts an offered introduction, whether an introducer offers a requested introduction, how feedback is generated by requester/target as well as interpreted by the introducer. These decisions for each

Work supported by the Air Force Research Laboratory (AFRL) under contract FA8750-10-C-0178. The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

node are prescribed by its so-called *policy*, which for the experiments in [1] was selected somewhat ad-hoc i.e., a set of heuristic rules with parameters tuned via a time-consuming simulation-based search. This paper presents an initial step towards a more model-based optimization approach to policy selection. Specifically, it is shown (Section 2) that the optimal continue vs. close decision within any established connection is given by a variant of the Sequential-Probability-Ratio-Test (SPRT) solution to the famous Wald problem [3]. A summary of related work (Section 3) and ideas for future work (Section 4) close the paper.

2. PROBLEM FORMULATION AND MAIN RESULTS

The model considered in this paper is from the perspective of one node over the lifetime of a connection to another node i.e., in the sequence diagram of Fig. 1, it focuses on the decision process by either introducee (requester or target) starting from when the connection is established to when it is closed. The implementation of this decision process in [1] is illustrated in Fig. 2. For each incoming transaction, an (imperfect) detector reports a "yes" or a "no" to the question of whether that transaction is a harmful attack and, in turn, the reputation is decremented or incremented, respectively. The updated reputation is then compared to a threshold to answer the question of whether the remote node is misbehaving, closing or continuing the connection if the answer is "yes" or "no," respectively. Mathematically, upon receiving the nth binary-valued report z_n for $n = 1, 2, \ldots$, the real-valued reputation R_n is updated according to a recursion

$$R_n = R_{n-1} + \rho(z_n) \tag{1}$$

with initial reputation R_0 determined when establishing the connection and

$$\rho(z_n) = \begin{cases} R_{\text{INC}} & , \text{ if } z_n \text{ reports a benign transaction} \\ R_{\text{DEC}} & , \text{ if } z_n \text{ reports a harmful transaction} \end{cases}, \tag{2}$$

feeding the associated continue vs. close decision according to a threshold policy

$$\mu(R_n) = \begin{cases} \text{continue} & , \text{ if } R_n \geq R_{\text{THR}} \\ \text{close} & , \text{ if } R_n < R_{\text{THR}} \end{cases}. \tag{3}$$

Subject to design are the (real) values for parameters R_{INC}, R_{DEC} and R_{THR}, referred to as the policy's *increment*, *decrement* and *threshold*, respectively.

The following sections show that the above parameterization of the continue vs. close policy is provably optimal when (i) the total utility of the connection is the expected net-sum-reward of all transactions over the lifetime of that connection and (ii) the detector's error probabilities are independent and identically distributed across all transactions. The associated increment and decrement are given in closed form, while the threshold is given as the solution to a specific iterative algorithm. The proof rests upon casting the decision process as a variant of the well-studied Wald problem [3], also leveraging results for optimal stopping problems in general [4].

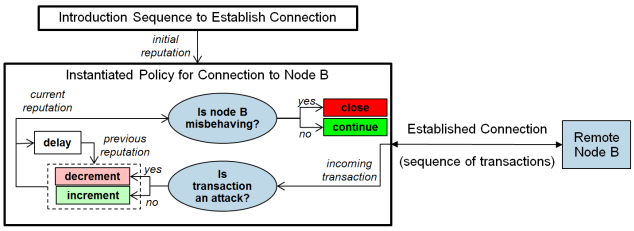

Fig. 2. The Continue vs. Close Decision Process in [1]

2.1. Main Model

Let a binary random variable X denote the (hidden) state of the remote node, which is either well-behaved ($X = 0$) or mis-behaved ($X = 1$). The overall probabilistic model makes the following assumptions:

- A misbehaving node attacks in any particular transaction with probability q (and does not attack with probability $1 - q$), independently of the attack sequence over all previous transactions.
- The detector classifies each transaction as "benign" ($Z = 0$) or "harmful" ($Z = 1$), misclassifying a non-attack transaction as harmful with false-positive rate q_{FP} and an attack transaction as benign with a false-negative rate q_{FN}, independently of the error sequence over all previous transactions.

Observe that the detector's overall error probabilities reflect not just the per-transaction misclassification rates q_{FP} and q_{FN} but also the misbehaver's attack rate q i.e.,

$$\mathbf{P}\left[Z = z | X = x\right] = \begin{cases} 1 - q_{\text{FP}} & , \text{ if } x = 0 \text{ and } z = 0 \\ q_{\text{FP}} & , \text{ if } x = 0 \text{ and } z = 1 \\ (1-q)(1-q_{\text{FP}}) + q q_{\text{FN}} & , \text{ if } x = 1 \text{ and } z = 0 \\ (1-q) q_{\text{FP}} + q(1-q_{\text{FN}}) & , \text{ if } x = 1 \text{ and } z = 1 \end{cases}. \tag{4}$$

The utility of the connection is expressed as an expected total discounted reward with discount factor $\delta \in (0, 1)$, where every transaction with a behaving node incurs a positive reward v, while with a mis-behaving node every non-attack transaction incurs zero reward and every attack transaction incurs a positive cost c. If the number of transactions were fixed at some positive integer n, then it is straightforward to express the utility i.e.,

$$\mathbf{P}\left[X = 0\right] U_0(n) + \mathbf{P}\left[X = 1\right] U_1(n)$$

with

$$U_x(n) = \begin{cases} \sum_{k=1}^{n} \delta^{k-1} v & , \quad \text{if } x = 0 \\ -\sum_{k=1}^{n} \delta^{k-1} qc & , \quad \text{if } x = 1 \end{cases}.$$

In our formulation, however, the number of transactions N is itself a random variable with distribution depending upon the policy. For a fixed policy, however, if we consider any positive integer n such that $\mathbf{P}[N = n]$ is nonzero, then conditioned on the event that $N = n$ the n-stage utility is

$$U(n) = \mathbf{P}[X = 0|N = n]\, U_0(n) + \mathbf{P}[X = 1|N = n]\, U_1(n).$$

In turn, introducing a final expectation with respect to stopping time N, the overall utility is $U = \sum_{n=1}^{\infty} \mathbf{P}[N = n]\, U(n)$ or, equivalently,

$$\begin{aligned} U \quad = \quad & \mathbf{P}[X = 0] \left(\sum_{n=1}^{\infty} \mathbf{P}[N = n|X = 0]\, U_0(n) \right) + \\ & \mathbf{P}[X = 1] \left(\sum_{n=1}^{\infty} \mathbf{P}[N = n|X = 1]\, U_1(n) \right). \end{aligned} \tag{5}$$

The first term quantifies the utility when connected to a well-behaving node, which is maximized by policies that result in longer stopping times, while the second term quantifies the utility when connected to a mis-behaving node, which is maximized by policies that result in shorter stopping times. Of course, the heart of the problem is that the true state of the remote node is never known with certainty but must rather be inferred from the report sequence of past transactions.

2.2. Summary of Results

The problem formulated in the preceding subsection is similar to the famous Wald problem [3] in most ways that general results in the field of stochastic dynamic programming have been organized i.e., both are infinite-horizon optimal stopping problems involving a partially-observable two-state system with bounded cost per stage [4]. One difference is that Wald's problem uses an expected total cost criterion without discounting. Another difference is that our single-stage cost of continuing another transaction also depends on the (hidden) state of the remote node. Even so, the form of the optimal policy continues to be the celebrated Sequential-Probability-Ratio-Test (SPRT) solution first derived by Wald. In an SPRT, the decision on whether to stop after the nth transaction amounts to comparing a pair of thresholds to the a-posteriori state distribution, which is Bernoulli with parameter p_n depending on the observed sequence $y_n = (z_1, z_2, \ldots, z_n)$. More specifically, initialize probability $p_0 = \mathbf{P}[X = 0]$ and, in each period $n = 1, 2, \ldots$, first apply the probabilistic state recursion

$$p_n \equiv \mathbf{P}[X = 0|y_n] = \begin{cases} \dfrac{p_{n-1}(1 - q_{\text{FP}})}{f(p_n, 0)} & , \quad \text{if } z_n = 0 \\ \dfrac{p_{n-1} q_{\text{FP}}}{f(p_n, 1)} & , \quad \text{if } z_n = 1 \end{cases} \tag{6}$$

with the denominator in each case using (4) via operator

$$f(p, z) = p\mathbf{P}[Z = z|X = 0] + (1 - p)\,\mathbf{P}[Z = z|X = 1],$$

and then choose to stop only if probability p_n is outside of the closed interval between the thresholds, classifying the 0-state if its above the upper threshold and the 1-state if its below the lower threshold. Our solution is in fact a special case of Wald's SPRT: in our setup, the decision to stop and the decision to classify the 1-state will be seen to be one in the same (i.e., in our setup, the upper SPRT threshold is always unity).

Definition 1 (Remote Node's Reputation) *Consider a connection in which there have been $n = 0, 1, 2, \ldots$ transactions resulting in the length-n sequence of reports $y_n = \{0, 1\}^n$ described by the sensor model in (4). The* reputation R_n *is in one-to-one correspondence with the probabilistic state p_n through the identity*

$$R_n = \log\left(\frac{p_n}{1 - p_n}\right) \quad \Longleftrightarrow \quad p_n = \frac{\exp(R_n)}{1 + \exp(R_n)}.$$

Note that as probability p_n approaches unity (zero), reputation R_n approaches positive (negative) infinity. In SPRT terms, reputation is exactly the log-likelihood that the remote node is behaving based on the history of sensor reports y_n.

Lemma 1 (Predicted Sensor Report) *Consider a connection described by the sensor model in (4), assuming that the current probabilistic state is p. Then, the sensor report to be generated by the next transaction, if the decision is to continue for another time period, is described by a Bernoulli random variable Z with parameter*

$$\mathbf{P}[Z = 1; p] = pq_{\text{FP}} + (1 - p)\left[(1 - q)q_{\text{FP}} + q(1 - q_{\text{FN}})\right].$$

Proof: Consider any sequence y_n in time period n, so we are assuming that $p = \mathbf{P}[X = 0|y_n]$ and we need to determine $\mathbf{P}[Z = 1; p] = \mathbf{P}[Z_{n+1} = 1|y_n]$. Using (4),

$$\begin{aligned} &\mathbf{P}[Z_{n+1} = 1|y_n] = \\ &\qquad \mathbf{P}[Z_{n+1} = 1, X = 0|y_n] + \mathbf{P}[Z_{n+1} = 1, X = 1|y_n] \end{aligned}$$

with

$$\begin{aligned} &\mathbf{P}[Z_{n+1} = 1, X = x|y_n] = \\ &\qquad \mathbf{P}[Z_{n+1} = 1|X = x]\,\mathbf{P}[X = x|y_n] \end{aligned}$$

for each $x \in \{0, 1\}$. □

Proposition 1 (Continue vs. Close Policy) *Given the connection model of Subsection 2.1, with sensor parameters q, q_{FP} and q_{FN} in (4) as well as utility parameters v, c and δ in (5), the utility-maximizing continue vs. close policy takes the form of (1)-(3) with* increment

$$R_{\text{INC}} = \log\left(\frac{1-q_{\text{FP}}}{(1-q)\,(1-q_{\text{FP}})+qq_{\text{FN}}}\right),$$

decrement

$$R_{\text{DEC}} = \log\left(\frac{q_{\text{FP}}}{(1-q)q_{\text{FP}}+q\,(1-q_{\text{FN}})}\right),$$

and threshold

$$R_{\text{THR}} = \log\left[p^*/(1-p^*)\right].$$

Here, $p^* \in (0,1)$ denotes the probability threshold implied by the optimal utility function $U^* : [0,1] \to \mathbb{R}$ governed by the Bellman equation

$$U^*(p) = \max\left\{0, pv-(1-p)qc+\delta\mathbf{E}\left[U^*\left(f(p,Z)\right)\right]\right\}, \tag{7}$$

where expectation $\mathbf{E}$ is taken with respect to the predicted sensor report Z of Lemma 1.

Proof (outline): From Sect. 5.4 in Vol. 1 of [4], the imperfect state information problem involving a two-state system can be reduced to a perfect state information problem involving the probabilistic state recursion in (6). Like the partially-observable problem, the reformulated problem is an infinite-horizon discounted problem with bounded cost per stage, so from Sect. 1.2 in Vol. 2 of [4] the Bellman equation for all $p \in [0,1]$ specializes to (7). By arguments analogous to those for the Wald problem (see Sect. 3.4 in Vol. 2), the optimal utility function U^* of this Bellman equation is ensured to be convex over $[0,1]$ with boundary conditions $U^*(0) = -qc$ and $U^*(1) = v/(1-\delta)$. In turn, the reasoning to optimality of a stationary threshold rule also still holds; see Fig. 3. The last step is to translate this probabilistic state solution in terms of reputation via Definition 1. □

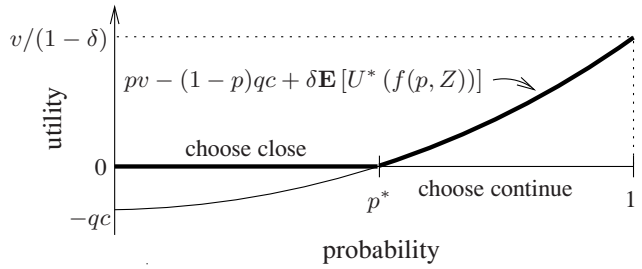

Fig. 3. The Optimal Utility $U^*(p)$ vs. Behaving Probability p

3. RELATED WORK

The fundamental concepts underlying Internet-based reputation systems are essentially no different from those underlying the "word-of-mouth" mechanism within human society, in general, about which much has been written e.g., [2, 5–7]. The lack of a central authority in a reputation system has also been considered in other work [8] e.g., the eigentrust scheme proposed in [9] has a fully distributed version, in which reputations are iteratively communicated. However, an introduction-based approach explicitly requires that every node keep its reputations about others nodes private [1] i.e., its architecture is not only computationally distributed but it is informationally decentralized, which makes its analysis both challenging and interesting. The application of probabilistic models for interpreting trust and reputations is also not new [10], nor is the application of a utility-based incentive structure for peer-to-peer networks [11, 12], but incorporating both within a common formulation for reputation systems has (to our knowledge) not been previously suggested. Finally, our development of Wald's SPRT solution for a reputation-based system is similar to its development in [13] for a network security classification game, but the link between probabilistic state and reputation was not made there.

4. CONCLUSION

A core decision process of a recently proposed introduction-based reputation protocol [1], aiming to retain the attractive properties of trust systems but without the assumption of a centralized reputation server, has been modeled as a sequential detection problem. It is shown that the connection policy studied through simulation in that initial proposal is optimal in form, characterizing the three policy parameters exactly as a function of the misbehavior model, the sensor model and the utility model. This characterization alleviates some of the tax associated with the simulation-based policy optimization approach employed in [1].

An analogous mathematical characterization of the numerous other decision processes among the different roles within the protocol would be similarly helpful. These include decisions by requester or target on whether to accept an offered introduction, decisions by requester or target on how feedback to the introducer is generated, decisions by introducer on how feedback from the requester and target is interpreted as well as decisions by introducer on whether to offer a requested introduction. Understanding the interplay of all of these different processes with the connection policy derived here is an open problem. Also of interest is the impact of richer adversary models (e.g., on-off misbehaviors), ultimately including strategic misbehaviors by adversaries with deep knowledge of the reputation protocol as well as the possibility of collusion among multiple nodes and across roles.

5. REFERENCES

[1] Gregory Frazier, Quang Dong, Michael P. Wellman, and Edward Petersen, "Incentivising responsible networking via introduction-based routing," in *Proceedings of the 4th International Conference on Trust and Trustworthy Computing*. 2011, Trust'11, pp. 277–293, Springer-Verlag.

[2] Audun Josang, Roslan Ismail, and Colin Boyd, "A survey of trust and reputation systems for online service provision," *Decision Support Systems*, vol. 43, no. 2, pp. 618–644, 2007.

[3] Abraham Wald, *Sequential Analysis*, John Wiley and Sons, New York, NY, 1947.

[4] Dimitri P. Bertsekas, *Dynamic Programming and Optimal Control (Vols. 1 and 2)*, Athena Scientific, Belmont, MA, 1995.

[5] Yafei Yang, Yan Lindsay Sun, Steven Kay, and Qing Yang, "Defending online reputation systems against collaborative unfair raters through signal modeling and trust," in *Proceedings of the 2009 ACM symposium on Applied Computing*, New York, NY, USA, 2009, pp. 1308–1315, ACM.

[6] Kevin Hoffman, David Zage, and Cristina Nita-Rotaru, "A survey of attack and defense techniques for reputation systems," *ACM Comput. Surv.*, vol. 42, no. 1, pp. 1:1–1:31, Dec. 2009.

[7] Yan Sun and Yuhong Liu, "Security of online reputation systems: The evolution of attacks and defenses," *IEEE Signal Processing Magazine*, vol. 29, no. 2, pp. 87–97, 2012.

[8] Yan Lindsay Sun and Yafei Yang, "Trust establishment in distributed networks: Analysis and modeling.," in *Proceedings of the IEEE International Conference on Communications*, June 2007, pp. 1266–1273.

[9] S. Kamvar, M. Schlosser, and H. Garcia-Molina, "The eigentrust algorithm for reputation management in p2p networks," in *Proceedings of the IEEE International World Wide Web Conference*, May 2003.

[10] Y. Sun, W. Yu, Z. Han, and K. J. Ray Liu, "Information theoretic framework of trust modeling and evaluation for ad hoc networks," *IEEE Journal on Selected Areas of Communicatins (Special Issue on Security in Wireless Ad Hoc Networks*, vol. 24, no. 2, pp. 305–317, Feb 2006.

[11] Qinyuan Feng, Yu Wu, Yan Sun, Jing Jiang, and Yafei Dai, "User behavior modeling in peer-to-peer file sharing networks: Dissecting download and removal actions," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, Washington, DC, USA, 2009, pp. 3477–3480.

[12] Yu Zhang, Jaeok Park, and M. van der Schaar, "Social norm based incentive mechanisms for peer-to-peer networks," in *Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, 2011, pp. 3116–3119.

[13] Ning Bao, O. Patrick Kreidl, and John Musacchio, "A network security classification game," in *GAMENETS*, 2011, pp. 265–280.