Show me how you play and I will tell you who you are

Cédric Lauradoux

Université de Lyon, INRIA,
INSA-Lyon, CITI,
F-69621, Villeurbanne, France
cedric.lauradoux@inria.fr

Abstract. Can a chess player play anonymously in the Internet? This question is critical for all tournament players: games are collected and gathered on online database. Identifying a player is "only" a problem of data mining, *i.e.* players characterization, pattern matching, and hypothesis testing. This abstract presents the general issues in the analysis and identification of a chess player. The problem depicted has consequences for many other domains in which privacy can be jeopardized. Chess is certainly the *Drosophila* of privacy for online games.

Keywords. Privacy, chess, data mining.

1 Introduction

John McCarthy [1] has referred to chess as the *Drosophila* of AI, speaking metaphorically about the pioneer works on genetics with fruit flies. The author's opinion is that chess is also the *Drosophila* of online game privacy. The main question of this paper is about the privacy of players in online games. Can my chess moves reveal my identity? The main issue is that chess players cannot control the access to the most sensitive information: the record of their own games! Chess players privacy looks harmless at a first sight and we have been confronted to such question numerous time in computer science. So nothing seems new under the sun.

Why studying chess fantasy while there are far more worrying applications like geo-localization [2] or the Facebook [3, 4]? Apart from being fun, there are four reasons:

- Chess is a finite game. It simplifies significantly the analysis task over modern real-time-strategy or massively multiplayer online role-playing games in which the players actions are difficult to model.
- It is easy to access large database of games. Therefore, statistical tools can be used and tested to verify hypothesis.
- Understanding the data mining problem found in chess might be very useful for data mining over the Facebook.

- The author was rated FIDE 2131 in 2006. Several of his own games appear in these databases...

The project envisioned by the author is particularly time consuming since large games database (more than four billions of games) must be explored to construct players profiles and many characteristics must be computed. This abstract details the goal and expectation of the author.

2 Player characteristics

To start our study, we need to secure a good source of information. Most of the organizers digitize all the tournament games and send them to free online databases such as TWIC [5]. Commercial products like the Big Database of Chessbase [6] can also be used. The primary purpose of these databases is to increase/share the knowledge on chess games and positions.

We do not remember the conduct of a chess game, the readers who have never approached a chessboard can consult [7] for the strict rules or [8] for a more entertaining way to learn chess.

A chess player can be characterized in many ways by his games. All chess players have their pet openings and some move orders they favor in specific positions. The Fig. 1 show the profile of five top grandmasters (Anand, Carlsen, Aronian, Kramnik and Ivanchuk) when they play their first move with the white. The number of games analyzed for each player is between 700 and 1500. It corresponds to all their known recorded games. The Fig. 1 shows the statistics of each grandmaster.

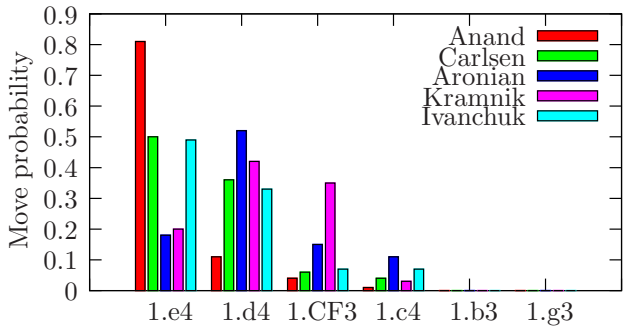

Fig. 1. Opening profile for the first move of five top grandmasters.

Let define $\chi = (\chi_1, \chi_2, \cdots \chi_n)$ a vector of n characteristics $\chi_i,\ i \in [1, n]$ obtained from the database exploration for a given player. A characteristic can correspond to the first move occurrence and so on... Everything can be considered. Let $\chi' = (\chi'_1, \chi'_2, \cdots \chi'_n)$ be the observed characteristics collected from

several online games. The challenge is to apply hypothesis testing [9] on χ and χ' to determine whether or not the anonymous player matched a given profile. We need also to determine the average minimal number of games needed to identify a player.

3 Conclusion

The contribution of chess to privacy looks marginal at a first sight: only the 200 000 players found in the database are concerned. However, chess is the unique opportunity to understand precisely all the difficulties encountered by a malicious user who wants to track or identify anonymous gamers or users based on their chess habits. Compared to other games, chess is finite and games database are easy to find offering a primary source of information to the malicious player. The results of this exploration look very promising to establish a theoretical foundation for more sensitive applications like the Facebook.

References

1. McCarthy, J.: Chess as the Drosophila of AI. Computers, Chess, and Cognition (1990)
2. Gambs, S., Killijian, M.O., del Prado Cortez, M.N.: Show me how you move and i will tell you who you are. In: 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS - SPRINGL 2010, San Jose, CA, USA, ACM (2010) 34–41
3. Gross, R., Acquisti, A., III, H.J.H.: Information revelation and privacy in online social networks. In: ACM Workshop on Privacy in the Electronic Society - WPES 2005, Alexandria, VA, USA, ACM (2005) 71–80
4. Acquisti, A., Gross, R.: Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. In: Privacy Enhancing Technologies - PET 2006. Lecture Notes in Computer Science 4258, Cambridge, UK, Springer (2006) 36–58
5. Crowther, M.: (2011) http://www.chess.co.uk/twic/twic.html.
6. ChessBase: Big Database 2011 (2011) http://www.chessbase-shop.com/en/products/5852.
7. Fédération Internationale des Échecs: FIDE Laws of Chess (2009) www.fide.com/FIDE/handbook/LawsOfChess.pdf.
8. Chandler, M.: How to Beat Your Dad at Chess. Gambit Chess (1998)
9. Blahut, R.E.: Principles and practice of information theory. Addison-Wesley Longman Publishing Co., Inc. (1987)
10. Coles, S.: Computer Chess: The Drosophila of AI. Dr. Dobb's Journal (2002) http://drdobbs.com/high-performance-computing/184405171.