Identity Analytics: Using Modeling and Simulation to Improve Data Security Decision Making

Marco Casassa Mont, Adrian Baldwin, Jonathan Griffin, Simon Shiu, Yolanta Beres

HP Laboratories
HPL-2008-188

Keyword(s):

Identity Analytics, Identity Management, Security, Security Analytics, Risk Assessment, Modelling Simulation, Economics, Trade-offs, Policies, Threats , Data Protection

Abstract:

This short paper describes research in the space of Identity Analytics that is helping decision makers (e.g. CIOs/CISOs) reason about choices of identity controls. We present an approach of using modeling and simulation to explore the effects of different identity controls on security and business outcomes. More specifically we present a case study on "collaborative data sharing" showing how the methodology can be applied to understand the trade-offs around centralizing identity management and educating users. This case study shows how simulation and modeling enhance decision making around identity management, along with its impact on data security.

External Posting Date: November 6, 2008 [Fulltext] Approved for External Publication
Internal Posting Date: November 6, 2008 [Fulltext]

Submitted to Financial Cryptography and Data Security 2009, February 23-26, 2009, Barbados

Identity Analytics: Using Modeling and Simulation to Improve Data Security Decision Making

Marco Casassa Mont, Adrian Baldwin, Jonathan Griffin,
Simon Shiu, Yolanta Beres

Hewlett-Packard Labs, Systems Security Lab, Filton Road, Stoke Gifford, Bristol, UK

marco.casassa-mont@hp.com, adrian.baldwin@hp.com, jonathan.griffin@hp.com,
simon.shiu@hp.com, yolanta.beres@hp.com

Abstract. This short paper describes research in the space of Identity Analytics that is helping decision makers (e.g. CIOs/CISOs) reason about choices of identity controls. We present an approach of using modeling and simulation to explore the effects of different identity controls on security and business outcomes. More specifically we present a case study on "collaborative data sharing" showing how the methodology can be applied to understand the trade-offs around centralizing identity management and educating users. This case study shows how simulation and modeling enhance decision making around identity management, along with its impact on data security.

Keywords: Identity Analytics, Identity Management, Security Analytics, Risk Assessment, Economics, Trade-offs, Security, Threats, Data Protection

1 Introduction

Business, confidential and personal data are valuable assets for financial and commercial organisations. Strategic decisions need to be made on how to protect this information, based on the context, people and applications that might access it, the threat landscape and involved risks. These decisions must also ensure the smooth running of enterprise's business operations. Strategic decision makers (e.g. CIOs/CISOs) develop policies and risk postures based on their risk appetite and approach to security. Identity Management (IdM) solutions [7,8] play a key role in this space, to mitigate risks: they are part of security strategies of any enterprise.

Important questions that decision makers might be interested in getting an answer to include: What is the best approach to protect valuable data? What are the effects of different authentication and access control strategies? What level of identity management automation is actually required? What is the trade-off between reducing risks by tightening access control vs. the loss of productivity that this might introduce? What is the impact of emerging enterprise web 2.0 technologies, such data sharing tools and social networking? How to more effectively curb identity thefts? Attitudes and motivations of people (end-users, employees, etc.) need to be taken into account in the decision making process as they affect security and can fuel threats.

Most of current solutions and R&D in the space of identity management aims at improving the technical IT controls and automating compliance monitoring. These solutions do not provide strategic, predictive capabilities based on analysis of trade-offs in investments and hence they do not aid the Chief Information Security Officer (CISO) in becoming more risk driven or prioritize their limited budget.

In this paper, we introduce an approach to Identity Analytics, as part of a wider Security Analytics project [1,2,10], based on rigorous modeling and simulation techniques. We work with the security decision makers to understand the risk questions that are concerning them. We refine the questions into models and simulations that can animate different scenarios ("what-if" analysis) thereby helping the decision makers explore the implications and consequences of different choices.

Our vision and methodology are introduced. As an example of our approach, a case study is discussed, about the enterprise trends towards using "unstructured data" and the increased adoption of "web 2.0 collaborative data sharing tools".

2 Identity Analytics: Vision and Methodology

In our vision Identity Analytics [6] consists of a set of approaches, techniques and methodologies to explain and predict the impact of identity, identity management and people's behaviour on aspects of relevance to decision makers such as security exposure and effect on productivity (see Figure 1).

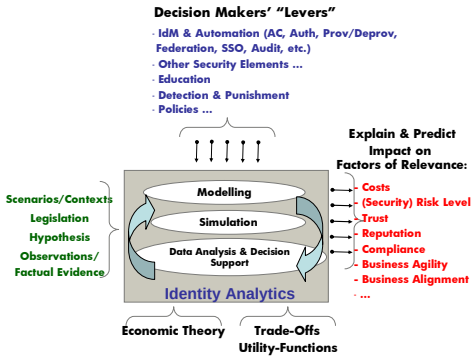


Fig. 1. Aspects of Identity Analytics

Identity Analytics aims at providing decision makers with decision support tools and services describing the controls or "levers" that they can influence and reflecting the likely consequences on impact metrics (exposure/productivity). Based on assumptions of the current state such tool allow for an exploration of the trade-offs of different decisions. Identity management is an area where even the experts have little intuition as to how to invest for the best (security) outcomes. As such, it is likely to be a high profile and rich problem area for Identity Analytics.

Modelling, simulation and analysis techniques are at the very core of Identity Analytics to support predictive capabilities and "what-if" analysis. This is coupled with social and economic theories to keep into account users' behavious and

motivations and economic drivers in analyzing trade-offs. To achieve this we rely on the mathematical foundation being provided by our Security Analytics project [2]. This foundation helps the decision makers explore their challenges and the trade-offs of various solutions; this approach also helps in discussions with other stakeholders.

The methodology we adopted to make progress in this space is based on the Scientific Method [9], tailored to the specific domain of security and identity management. Given a specific scenario/context of interest, this requires building a theory of a specific phenomenon we are interested in (e.g. involved threats and related risks, impact of control points, trade-offs, etc.), making related hypothesis, design experiments to prove/disprove these hypothesis (and hence the theory) and, based on the outcomes, potentially refine the initial theory. This involves gathering observational facts, using them to create models via an inductive process, using simulation techniques on top of these models to draw conclusions and validating these conclusions against the real world. The aim is to build models that covers the current and potential alternatives and/or future situations, by exploring security threats and risks, the impact of controls and trade-offs between different options. The model is verified against the 'known situations' with the hope that it generalises appropriately to the unknown ones.

Our current preliminary work and exploration of this space has been based on a "simulation-based predictive modelling" approach. Our initial investigation of this approach suggests advantages over the analytical approach as it allows exploration (in a more natural way), via experimental results, of the dependencies among different involved entities, asynchronous and competing processes and decisions. This is particularly the case for scenarios involving modelling business process aspects, interdependencies with identity management solutions and probabilistic users' actions and behaviours. Specifically, we have used a specialised simulation language Demos2k [3, 4, 5], which is based on the mathematical foundations of a synchronous calculus of resources and processes, together with an associated modal logic. Models are executed via repeated experimental simulations in the specially developed experimental environment, where statistically significant information is gathered.

3 Case Study on Data Repositories and Unstructured Data

To illustrate the concepts involved in Identity Analytics, we present a case study and a related scenario, that we believe is of relevance for financial and commercial environments. For the purpose of this paper we just aim at illustrating the principles and some of the involved steps. More details are available [6].

We consider an enterprise scenario focusing on the emerging trend towards the adoption of (web 2.0) collaborative data sharing tools, such as Wiki, Twiki and Microsoft Sharepoint tools. On one hand, people within organisations might be encouraged to share data and information, hence improving communication and projects effectiveness. On the other hand this presents security risks, e.g. that data may be accessed and shared inappropriately. Traditional authentication and access control/authorization solutions show their limitations, due to the type of flexibility provided by these data sharing tools in dynamically managing content, handling

access credentials and enabling people to collaborate and share content. This aspect is of particular concern to enterprises, including financial and commercial organisations, given the kind of sensitive data they need to protect.

This introduces dilemmas for decision makers. What are the most appropriate security policies? Should data sharing tools be forbidden? If so, this decision could improve security but have a negative impact on business effectiveness. If, instead, collaborative data sharing tools are allowed, which constraints should apply? For example, could users install and use their own preferred solutions e.g. Ad-Hoc/Self IT (SIT) solutions? Or should solutions only be provided by "Central IT (CIT) enterprise services"? In these two cases, different IT security profiles and threats apply depending on how the data sharing sites are designed and managed. How vulnerable is an enterprise to the correct behaviour of the IT support staff?

In general, data confidentiality, integrity and availability threats need to be considered and analysed. Data can have different levels of confidentiality and value. Data could be accidentally or maliciously leaked by employees or IT Staff. Data could be tampered with. Risks include financial losses and negative impact on reputation due to data leakage or unauthorized disclosure of confidential data.

User's profiles and their motivations are important. People might have different attitudes, including being compliant to security policies, be loose (slight effort compliance), non-compliant, traitor or disgruntled. Depending on their behaviours, people might or might not choose to use Central IT data sharing sites. Their attitude at managing their credentials (e.g. UserId and passwords) might be different, including in writing them down or sharing them with colleagues.

Automating Identity Management (single-sign-on, authentications, user account management, credential expiration, etc.) and providing them as part of the Central IT enterprise services could in some way mitigate these risks and protect data; for example, by expiring credentials over time. However, would this realistically curb the problem, if employees decide to use their own data sharing tools and ad-hoc identity management? Would an investment in educating people be more effective?

Our approach involves the analysis of all these aspects, including identifying entities, their interactions, involved processes, threat & risks and outcomes of interest that can be measured. Figure 2 provides an overview of these aspects as analysed in this case study along with a way to relate threats to information flow and interactions.

Based on this analysis, we can build a model by using Demos2k [36]. We represent a variety of aspects, ranging from human behaviours to security aspects and technological components. We can the start to tackle the analysis of their aggregated effects, consequences and explore trade-offs. "Factors of interest" to decision makers can be identified as potential outcomes to measure and analysed, e.g. the overall amount and value of leaked data, the number of written down or shared credentials and the number of these credentials that have been misused. We repeatably run the model with assumed probability distributions, using the Monte Carlo method, to get experimental results for the distributions of the potential outcomes [6].

"What-if" scenarios can then be analysed. For example, Figure 3 shows the impact on data leakage, obtained by changing people's behaviours and attitudes, in case investments are made in education and/or compliance monitoring and punishment. Based on our initial assumptions, Figure 3 shows that, by ensuring that 55% of the population is compliant, the leakage of data is drastically reduced. The marginal

return of additional investments in IdM is minimal. These results could encourage decision makers to further invest in education rather than in IdM solutions.

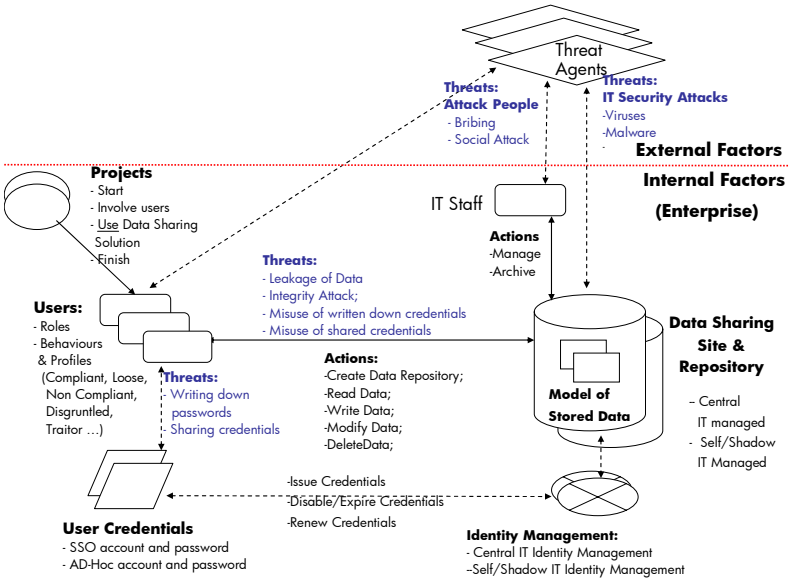
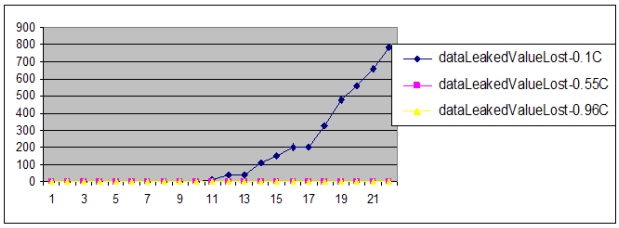


Fig. 2. Enterprise Data Sharing Scenario: involved entities, actions, interactions and threats

(A)

	Compliant	Loose	Non-compliant	Traitors	Disgruntled
Ref. Model	0.1	0.58	0.3	0.01	0.01
Alternative #1	0.55	0.28	0.15	0.01	0.01
Alternative #2	0.96	0.01	0.01	0.01	0.01

(B)



(C)

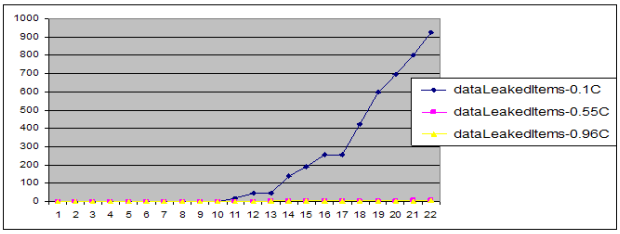


Fig. 3. Outcomes of Simulation

4 Related Work and Discussion

Decision support systems based on modelling and simulation techniques are not new and have been successfully used in many disciplines to explain and predict various trends and phenomena. Our contribution consists in coupling these techniques, along with economic and probability theory, in the context of security, identity and identity management and aiming at providing a rigorous mathematical foundation to observations and the decision-making process. Some important related work in this space include [11,12,13, 14], but their focus is limited on specific areas.

"Identity Analytics" is an overloaded world. Most of current solutions and R&D work focus on identity controls and compliance solutions. These are "bottom-up" solutions, starting from a low-level technical perspective. We aim at providing "top-down" solutions to key decision makers to aid their decisions, by explaining and predicting the impact of their choices (both technological and policy investments). We aim at leveraging current standards such as ITIL, CoBIT and ISO 27001 and use them as drivers and references but our work in the space of Identity Analytics will add the value of grounding the reasoning to specific contexts.

The modeling of the impact of identity controls can be complex in that the models need to include aspects of human behavior; how they interact with information, systems and policies. This poses key challenges in getting sufficient observations to create a realistic model. Such data may not be available, too expensive to extract and hence we also rely on expert validation allowing us to qualitatively explore the space of the outcomes. In this way we can provide models that explore the consequences of possible decisions; showing the shape of trade off relationships but where we cannot trust the exact figures from the model. More details about related work are available here [6].

5 Conclusions

This paper introduced the concept of "Identity Analytics" as an approach to explore and predict the impact of identity, identity management and user behaviour on enterprise systems. We presented our vision and the adopted methodology, based on an adaptation of the Scientific Method, along with leveraging modeling and simulation techniques, social and economic sciences. We have described the approach and its possible impacts on decision-making through a case study.

We believe that this approach is of key relevance to the financial and commercial context to steer the directions of decision makers toward more effective risk management and (IT security) policy choices. This area provides plenty of research opportunities as well as challenges to overcome.

References

1. HP Labs, Systems Security Laboratory (SSL), HP Labs, http://www.hpl.hp.com/research/systems_security.html, 2008

2. Security Analytics, HP Labs, Systems Security Laboratory, 2008
3. Demos2k, Demos 2k, http://www.demos2k.org/, 2000
4. Birtwistle, G., Demos,discrete event modelling on Simula. Macmillian, 1979
5. Pym, D., Monahan, B., A Structural and Stochastic Modelling Philosophy for Systems Integrity. HP Labs Technical Report Series, HPL-2006-35, Feb 2006
6. Casassa Mont, M, Baldwin, A., Shiu, S., On Identity Analytics: Setting the Context, HPL Technical Report, HPL-2008-84, 2008
7. Birch, D., Digital Identity Management: Technological, Business and Social Implications, Book, 2007
8. Windley, P., Digital Identity, O' Reilly, 2005
9. Wilson, E. B., An Introduction to Scientific Research McGraw-Hill, 1952
10. Trust Economics, UK DTI grant P0007, Trust Economics Project, 2008
11. Adams, A, Sasse, M.A., Users are not the enemies, Communications of the ACM, Volume 42, Issue 12, pages 40-46, 1999
12. Moore, T., Clayton, R., The Consequence of Non-Cooperation in the Fight Against Phishing, 3rd APWG eCrime Researchers Summit, 2008
13. Shay, R., Bhargav-Spantzel, A., Bertino, B., password policy simulation and analysis, DIM 2007, 2007
14. Anderson, R., Moore, T., The economics of information security. Science, 314:610–613, http://www.cl.cam.ac.uk/˜rja14/Papers/toulouse-summary.pdf, 2006