Towards Identity Analytics in Enterprises

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

HP Laboratories
HPL-2008-186

Keyword(s):
Identity Analytics, Identity Management, Security, Security Analytics, Modelling, Simulation, Economics, Trade-offs, Policies

Abstract:

This paper aims at setting the context for "Identity Analytics" within enterprises. In our vision, Identity Analytics is about helping decision makers (e.g. CIOs, CISOs) to explain and predict the impact of identity and identity management (along with other related aspects, such as users' behaviours) on key factors of relevance to them (e.g. costs, risk exposure, reputation, trust, etc.) - based on their initial assumptions and investment decisions, in complex enterprise scenarios. Decision makers are increasingly asked to justify their decisions and provide evidence about returns of IT and security investments. Our goal is to provide them with rigorous techniques to gain a better understanding of the involved threats & risks and investment trade-offs within the identity space (e.g. investing in technologies vs. changing processes vs. investing in users' education). This means providing "decision support" and "what-if analysis" capabilities to explore options, formulate new policies and/or justify existing ones. Our vision is introduced and discussed, along with the methodology that we adopted. There are many opportunities and challenges in this space: a scientific approach is required, involving the use of modelling and simulation techniques, coupled with the understanding of involved technologies and processes, human behaviours and economic aspects. As a significant example, we describe a case study focusing on emerging "web 2.0 enterprise collaborative data sharing tools", where unstructured information is created, stored and shared by people in collaborative contexts, within and across organizations. We discuss related threats and risks and demonstrate how trade-offs can be explored using the modelling approach hence allowing decision makers to investigate the different impacts of policy choices.

External Posting Date: October 29, 2008 [Fulltext] Approved for External Publication
Internal Posting Date: October 29, 2008 [Fulltext]

Submitted to 24th IFIP SEC 2009, 18-20 May 2009, Papos, Cyprus

Towards Identity Analytics in Enterprises

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

Hewlett-Packard Laboratories, Systems Security Lab (SSL), Bristol, UK

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

Abstract *This paper aims at setting the context for "Identity Analytics" within enterprises. In our vision, Identity Analytics is about helping decision makers (e.g. CIOs, CISOs) to explain and predict the impact of identity and identity management (along with other related aspects, such as users' behaviours) on key factors of relevance to them (e.g. costs, risk exposure, reputation, trust, etc.)—based on their initial assumptions and investment decisions, in complex enterprise scenarios. Decision makers are increasingly asked to justify their decisions and provide evidence about returns of IT and security investments. Our goal is to provide them with rigorous techniques to gain a better understanding of the involved threats & risks and investment trade-offs within the identity space (e.g. investing in technologies vs. changing processes vs. investing in users' education). This means providing "decision support" and "what-if analysis" capabilities to explore options, formulate new policies and/or justify existing ones. Our vision is introduced and discussed, along with the methodology that we adopted. There are many opportunities and challenges in this space: a scientific approach is required, involving the use of modelling and simulation techniques, coupled with the understanding of involved technologies and processes, human behaviours and economic aspects. As a significant example, we describe a case study focusing on emerging "web 2.0 enterprise collaborative data sharing tools", where unstructured information is created, stored and shared by people in collaborative contexts, within and across organizations. We discuss related threats and risks and demonstrate how trade-offs can be explored using the modelling approach hence allowing decision makers to investigate the different impacts of policy choices.*

1. Introduction

Identity and access management [3,4,5,6] are key aspects of the security strategy for any enterprise and one that is critical for ensuring people gain timely access only to the business applications necessary for their job. The *Chief Information Security Officer* (CISO) will develop risk postures and policies that reflect the enterprises risk appetite and approach to managing identity and access. Such policies are key to understanding how the enterprise defines who can access a given data set or perform a particular transaction. These policies are then refined into a number of technical and procedural controls that help achieve the appropriate risk posture. Within the identity space, such controls include the management of employee lists though active directories and single-sign-on (SSO); user account lifecycle management; the management of access control to information and systems; authentication mechanisms; compliance and reporting tools. As the enterprise operates, the CISO and company officers need to know how well they are meeting the risk postures they have defined and hence monitoring key metrics and auditing the controls is an important part of the risk life-cycle.

Currently there is much research into improving the technical controls and automating compliance monitoring for security in general and more specifically in the area of Identity Management (IdM). In this paper we are concerned with Identity Analytics, that is tools that help the CISO and other security decision makers rigorously explore the consequences of their policy choices, demonstrate due diligence and more accurately explain to senior management why they should invest in the recommended identity or access solution. Moreover, by exploring the relationships between technical and procedural controls, Identity Analytics provides a better understanding of which metrics should be monitored and ways in which they indicate risk. Typical questions addressed by Identity Analytics would be:

- What is the trade-off between reducing risk in tightening the access to critical applications vs. the loss in productivity as access rights are more limited and time taken to gain these access rights will increase?
- Is it better to spend a limited budget on user education or implementing a given technical control, such as automating user provisioning/deprovisioning or providing two-factor authentication?
- Should users and business units be allowed to run their own IT solutions or is it better to have centrally managed services?
- What is the impact of emerging collaboration technologies such as blogging, Wikis and second life?
- Do changes to working patterns such as greater mobility lead to additional risks?

Additionally, Identity Analytics solutions that help answer and animate the trade-offs within these questions can help the CISO influence other parts of the organization in understanding why particular security policies or controls are necessary, hence aiding the CISO reaching out to the rest of the enterprise.

In the security and identity management realms, current solutions providing "compliance assessment" capabilities (for specific laws and policies, against potential exposures to risks) can only partially address these issues. Many solutions in this space are reactive and driven by a bottom-up approach. Again, their decision support capabilities are limited to reporting compliance violations and highlighting potential risks, based on existing policies, common security criteria and current IT operational environments. These solutions do not provide strategic, predictive capabilities based on analysis of trade-offs in investments and they do not take into account the current strategic transition from a pure compliance-based approach to a risk-based approach, driven by the CIOs/CISOs' needs to cope with limited budget and resource issues and prioritize their investments.

In this paper, we are proposing an approach to identity analytics, as part of a wider HP Labs security analytics project [1,2], where we work with the security decision makers to understand the risk questions that are concerning them. We then refine these questions into models that can animate and explore different scenarios thereby helping the decision makers explore the implications and consequences of different decisions. Within this context, we are using modelling languages [22] with strong mathematical foundations [23,24] to simulate processes, people and systems in the enterprise and potential changes thus gaining a better understanding of the possible impact of decisions.

The remaining part of this paper is structured as follows: Section 2 describes in more detail our vision for "Identity Analytics"; Section 3 illustrates the suggested methodology to make progress towards Identity Analytics; Section 4 walks through a significant example of Identity Analytics, focusing on the management of unstructured data by employees by using collaborative "enterprise web 2.0 data sharing" tools; Section 5 compares this against related work in this space; Sections 6 discusses current results and next steps. Section 7 draws a few conclusions.

2. Our Vision of Identity Analytics

Current identity management decision making relies on a mixture of the deployment of best practice solutions along with a reliance on experts' judgement and intuitions. We believe, however, that it can be hard for an expert to gain good intuitions and confidence in those intuitions given the complexity of the identity management space and its tight relationship with business processes. Decision makers need to ensure that investments in identity management processes are made to achieve an appropriate level of risk to the business whilst allowing for productive working. Identity analytics is concerned with providing a set of approaches, techniques and methodologies that help explain and predict the impact of changes to identity controls on the business.

Towards this vision we are exploring a methodology based on modelling, simulation and analysis techniques that can be used to explore the impacts of a range of identity controls within a given business scenario. Figure 1 summarizes the approach showing the range of controls or levers the decision makers can affect with the use of modelling and simulation to explore the effect on a number of significant factors. The decision makers also need to understand the trade-offs they are trying to explore and hence formulate questions to be answered by the analysis.

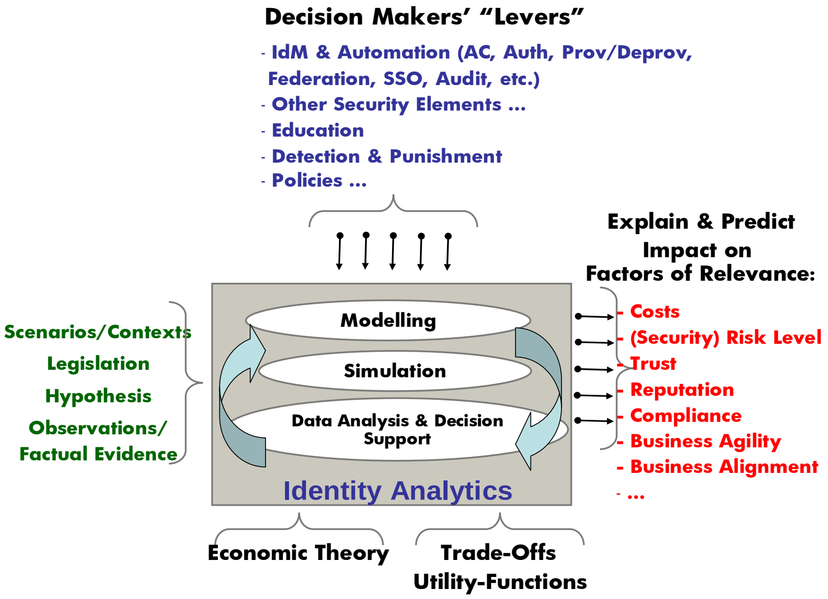


Fig. 1 Aspects characterizing Identity Analytics

Our approach starts by understanding a scenario where identity control changes are being explored and then abstracting this scenario into a model that captures the important entities, assets, relationships, related threats and the dynamic behaviour of the system. The model also includes various metrics or measures representing the impacts of these controls. We then animate the model using simulation techniques to understand how these elements interact, collecting the metrics so that we can understand the various impacts. This model captures our understanding of the system that we can start to validate. We can then create alternative versions of the model to explore different investment options, i.e. different selections of the identity controls or levers.

Once we have models that explore different investment options and simulation results showing the different effects we perform some data analysis to help understand the different impacts. The comparison between different investment options does not result in a simple single change. For example, investing in an improved identity provisioning process can reduce the risk that people have inappropriate access to systems; but at the same time there is a productivity hit in the time taken to gain an appropriate account and risks as accounts are more likely to be shared. The relationship between these metrics that emerges from the models will not necessarily be linear and hence the results require more analysis in helping the decision maker understand the consequences.

Decision makers usually have an informal utility function reflecting their experience and that helps them make judgements as to the importance of the different predicted outcomes. Economic theory provides us with rigorous techniques to explore these trade-offs and preferences and following other research on the economics of information security [8,9,10,11,12,13,14,15,16], we aim to include these techniques within our approach. In particular, we will use the idea of utility functions to help capture the relative importance of the different metrics thereby helping the decision maker in understanding and value the different aspects of the trade-offs.

Decision-making based on modelling and simulation techniques is not new and has been successfully used in many disciplines to explain and predict various trends and phenomena. The contribution of this approach to security analytics [2] and here particularly to identity analytics is two fold. Firstly, we aim to couple techniques from cognitive science and economics with the modelling and simulation to help the models reflect human behavior in exploring the trade-offs. Secondly, we are applying these techniques to bring a rigorous approach to decision making around identity controls—hence reducing the reliance of expert intuition.

3. Moving Towards Identity Analytics: Methodology and Modelling Tools

The basic methodology that we have adopted, based on the scientific method [17] and tailored to security and identity management, involves hypothesizing a theory or model that explains the current situation. We iterate with this model starting with some observational facts about the current scenario and validating results against experts' opinions and other observed facts. We can gradually add detail to the model based on further observations, e.g. through user interviews, in order to ensure the output of the model sufficiently reflects the current scenario. We can then use the model to explore specific phenomena by varying the assumptions, or adding additional facets representing controls. This overall process is illustrated in figure 2.

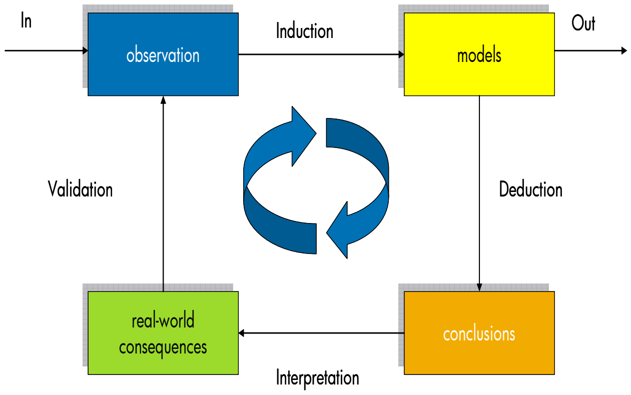


Fig. 2 Modelling Methodology

For example, theories or models can be built, in an enterprise scenario, about the impact and effects of policies such as password policies (based on length, complexity and expiration times) on risks based on observations about users' behaviour. We animate the model using it to drive simulations recording the outcomes, which can be compared with observations (such as password resets, or the results of password cracking programs). We can then hypothesize the effects of changing policies (e.g. increasing password length) by changing the assumptions in the model around the likelihood of people writing down passwords and that the password is guessable. Ideally, such assumptions are based on further observations, for example, understanding how people react as password complexity increases. Simulations then allow us to explore likely and possible effects of imposing a new password policy.

Given a specific scenario and context, empirical studies can be conducted to gather observational facts and evidence. Observational information and relevant events can be collected from the field (by means of surveys, analysis, etc.) to describe, in probabilistic terms [35], some of the involved interactions (such as the likelihood of users performing specific actions on enterprise resources) and processes where the above entities are involved. Two kinds of events (and related observations) can be analysed: (1) **external events**: these are events that just happen, where there is little degree of control, at least in the short term.; (2) **internal events**: these are events that can be influenced and/or for which there is a degree of control. This information is used in an inductive step, to produce (semi-) formal modelling components (such as probability distributions and statistics) and iteratively build one or more mathematical models. Simulations, based on these models, are used to generate experimental results.

Experiments are run multiple times, with Monte Carlo simulation methods, to derive the probability distributions of the potential outcomes. A subsequent analysis of these results drives the deductive process leading to conclusions on factors of interests (e.g. involved costs, impact on reputation, impact on trust and compliance, etc.) and presenting the impact of different trade-offs (e.g. investing in a technological approach to identity management vs. investing in an educational approach). Experts in the field, including decision makers, are likely to be actively involved in this process. Multiple iterations of the entire process might be required, before a model matches expectations and can provide meaningful predictions to decision makers about non-intuitive situations and aspects. This methodology has already been successfully explored at HP Labs and applied both internally and in the service consulting context by the Open Analytics project [19,20,21]. It is also consistent with the methodology that is used in the context of the UK TSB Trust Economics Project [18], in which HP Labs is also involved

The modelling of the impact of identity controls can be complex in that the models need to include aspects of human behavior and 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 or too expensive to extract, and in this case, we can still construct models validated by experts that allow 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. For example, we may not know the number of “traitors” within an organization but the model may help us understand that the effects of increasing numbers quickly tails off.

Section 4 walks through an illustrative case study, focusing on an “enterprise data sharing” scenario with the objective of exploring the shape of the outcome space. We also aim to follow a rigorous, scientific approach in analysing complex contexts where the outcomes of the modelling and simulation steps might not be trivial and intuitive, given the complexity and non-deterministic aspects of the involved interactions. This is where we see the greatest value of Identity Analytics, i.e. in providing insights and analysis in complex contexts where intuitions and expertise can only help up to a point. This is particularly true when trying to provide indications to decision makers about the outcomes of trade-offs.

3.1 Modelling and Simulation Tools

Analytical and predictive mathematical modelling approaches are potentially suitable for carrying out modelling and simulation activities in the Identity Analytics area. Our current preliminary work and exploration of this space has been based on a “simulation-based predictive modelling” approach. Based on our initial investigation, the predictive modelling approach provides advantages over the analytical approach as it allows us to explore (in a more natural way), via experimental results, the dependencies among different involved entities without over-simplifying probability distributions. It also appears easier for security experts to understand the models behind a simulation rather than the complex equations behind a more analytical approach.

Specifically, we have used a specialized simulation-oriented language, Demos2k [22,23], which implements a modelling framework based on the mathematical foundations of a synchronous calculus of resources and processes, together with an associated modal logic [24]. Because of its strong mathematical foundations and sound semantics, we have assurance that simulations based on the models developed in the Demos2k language are robust and reliable—thus, meaningful observations can be taken. The code is executed via repeated experimental simulations in the specially developed experimental environment [25], where statistically significant information is gathered. The mathematical framework behind the Demos2K programming language revolves around four key concepts: (1) **resources,** capturing the essentially static components of the system; (2) **processes,** capturing the dynamic components of the system; (3) **location,** capturing the spatial distribution and connectivity of the system; (4) **environment** within which a system functions. A full description of this mathematical framework can be found in [24].

In the domain of Identity Analytics, “resources” could be any valuable asset or element we might want to model. For example, this could include confidential and personal information, user accounts and related passwords/credentials, identity and authentication tokens, etc. Modelled “processes” could include, among other things, identity management processes and systems, data lifecycle processes and data flows, enterprise business processes, and human activities and behaviours. “Location” modelling aspects are also of particular importance in Identity Analytics: they are required to represent spatial distribution aspects of identity management systems, data repositories and people’s locations and interdependencies. Finally, the “environment” aspect is used to model additional characteristics of the scenario under observation that are of relevance for the simulation steps.

In general, the model of a specific system in a scenario usually has multiple processes, either consuming resources, or taking a certain time to finish. As a result of one process finishing, another might be triggered. Alternatively, they might start concurrently depending on the structure and complexity of the system to be modelled. Some of the processes could be triggered by events from the environment. Demos2k efficiently handles concurrency, queuing, and prioritization among processes. Based on our current investigation, Demos2k seems to be suitable to address most of the modelling and simulation needs for Identity Analytics. A critical aspect for Identity Analytics is the capability of modelling and simulating the behaviour of large populations. This is where we are still testing how to better use Demos2k. Section 4 illustrates an example of our current approach.

4. Case Study on Enterprise Data Sharing Tools and Unstructured Data

This section aims at grounding the Identity Analytics concepts described in this paper by walking through a case study of how we deal with modelling, simulation and data analysis aspects. The goal is to show how to help decision makers to understand and explore the involved threats and risks; explain and predict the impact of identity management solutions in mitigating involved risks; and explore trade-offs on alternative options. We followed the methodology illustrated in Section 3: at the current stage, the results illustrated in this section are the outcomes of an experiment that is evolving over time. It shows the range of outcomes and brittleness to different assumptions (e.g. about employee population types). Within the overall methodology this work sits in the early experimentation stage of trying to model based on experts' intuitions and analysis of scenarios. This forms the basis for a more rigorous experimental design including user studies to better understand the details of how people act in a given situation.

The scenario we considered is about an emerging enterprise trend, consisting in the adoption and usage by employees of collaborative, customizable data sharing tools, such as Wiki, Twiki and Microsoft Sharepoint tools. These collaborative tools provide an unprecedented level of flexibility and simplicity of usage, in terms of creation of related data sharing sites, creation, posting and retrieval of "unstructured data" [38] and information, collaborative generation of content and wide options for sharing unstructured material and information. Additional details are available [7].

Decision makers (such as CIOs/CISOs) face a dilemma in these situations. On one hand, people within organizations might be encouraged to share data and information, improving communication among parties involved in projects and increasing effectiveness. On the other hand this presents security risks, e.g. that data may be accessed and shared inappropriately. So how should the decision makers act? By completely forbidding the usage of data sharing tools, they could undermine collaboration, creativity and innovation. Alternatively they could decide to directly supply these data sharing services by means of centralized enterprise IT services, hence meeting the basic security requirements and being compliant with security policy but undermining the level of flexibility and customization of these tools. Further effort could be made in educating the workforce about good practice and security risks. What are the implications of each of these options, for example in terms of data leakages? Which investments need to be done on "control points" in the IT infrastructure, in particular in terms of Identity Management?

4.1 Assumptions and Analysis

Our analysis covered multiple aspects, including: (1) processes, people, identity management solutions, etc.; (2) the profile and behaviour of users/employees; (3) various interactions and information flows; (4) analysis of related threats and risks; (5) desired outcomes and metrics. Figure 3 provides a high level overview of the key aspects we believe need to be taken into account.

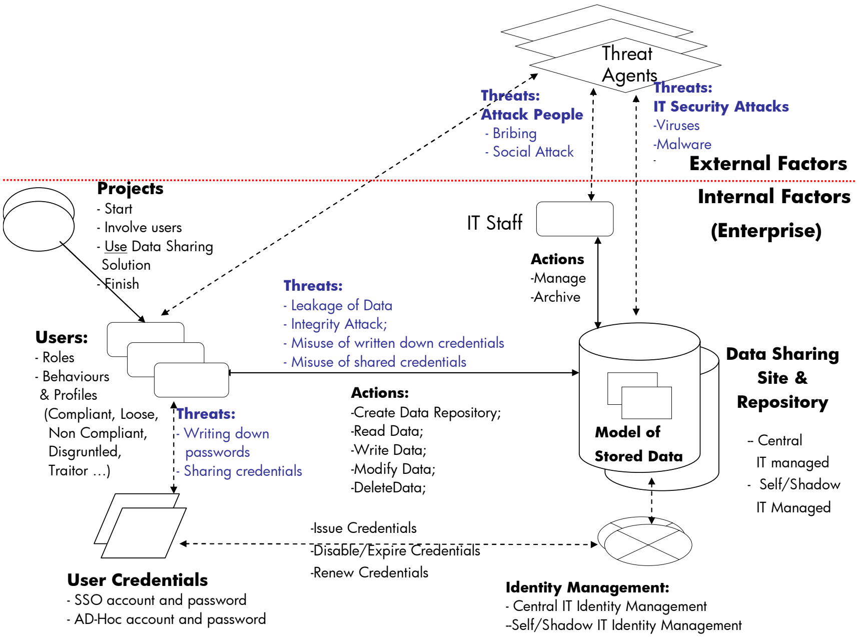


Fig. 3 Enterprise Data Sharing Scenario: involved entities, actions, interactions and threats

Different types of security threats need to be considered, which can be roughly categorized as follows:

- **Confidentiality Threats**. Most of these threats are related to data leakage, e.g., sending or sharing info with the wrong audience, or leaving data in unprotected areas (e.g. laptop in a taxi). Some of the leakages could be of a malicious nature, whether originating with external attackers or untrustworthy employees.
- **Integrity Threats**. They include sabotage by modifying/tampering with data, vandalism/graffiti attacks, etc.
- **Availability Threats**. Examples include denial of service (DoS) on data repositories, and loss of encryption keys.

Potential impacts include financial losses, loss of reputation and trust, and public embarrassment.

We considered two key categories of enterprise data sharing tools—and related data repositories supporting them—that could be adopted and used by employees to support their collaborative needs:

- **Central IT (CIT) Data Sharing Sites**: these sharing sites (e.g. Twiki, MS Sharepoint, shared file systems, etc.) are hosted and run on enterprise approved IT infrastructures;
- **Shadow/Self-IT (SIT) Data Sharing Sites**: these sharing sites are run on “shadow/self IT”, i.e. not officially approved enterprise IT infrastructures, such as personal servers or individuals’ PCs.

Each type of data sharing site (and related data repositories) has a different IT security profile and exposes organizations to different risks. We might expect that Central IT data sharing sites are patched and updated on regular basis, and are linked to well run identity management solutions, whilst Shadow/Self-IT sites may lack good security practices, hence exposing their content to leakage and external attacks.

Identity management [3,4,5,6] plays a key role, by providing basic mechanisms for authentication, authorization and protection of data. User provisioning and deprovisioning solutions can be used to automate the management of the lifecycle of user accounts and their credentials. Federated identity management and Single-Sign-On (SSO) solutions can simplify users’ access to various systems and sites by reducing the number of required user accounts and credentials. We considered two categories of identity management solutions that could be used in this scenario:

- **Central IT Identity Management Solutions**: These IdM solutions are provided by the organization’s central IT services and are expected to be compliant with the organization’s security policy. They might include automated provisioning and deprovisioning of user accounts; management of single-sign-on credentials; lifecycle management of passwords and credentials. If a data sharing site is managed by the enterprise central IT, we assumed that Central IT Identity Management solutions are used.
- **Ad-Hoc/Shadow/Self-IT Identity Management solutions**: These identity management solutions are provided on ad-hoc basis, by Shadow/Self-IT sites, for example in terms of ad-hoc management and setting of user accounts and passwords. These identity management solutions might not be compliant with security requirements imposed by the organization. For the purposes of this model, we assumed that Shadow IT data sharing sites use these IdM solutions.

4.2 Modelling

Based on our analysis, an initial, full working model has been built by using Demos2k [22], a stochastic, discrete-event modelling tool and framework. A fully copy of this model is available in Appendix A. By representing in our model a variety of aspects, ranging from human behaviours to security aspects and technological components, we can start tackling the analysis of their aggregated effects and consequences and explore trade-offs. Specifically this model represents the involved systems (i.e. different types of data sharing sites, identity management solutions, etc.) the involved categories/classes of users, their interactions and behaviours when dealing with these data sharing sites and stored information.

The model starts with the assumption that “collaborative projects”, groups of people pursuing similar objectives, are created; employees, with various profiles, join these projects and collaborate by sharing information. To satisfy collaborative needs, a data sharing repository is created for each new project, either a central IT managed or a self-managed solution. Users interact with the data sharing tools (and associated data repositories) by <u>creating</u> new documents, <u>reading, modifying</u> and <u>deleting</u> existing documents. Documents stored in various data sharing sites have different “levels of confidentiality and value”, ranging from little or no value to confidential and private data. The impact of data leakages depends on the value of this data.

To access the data repositories we assumed that users need “authentication credentials”. If a Central-IT data sharing site is used, single-sign-on (SSO) credentials are provided by *Central IT Identity Management.* This IdM solution will take care of the lifecycle of these SSO credentials, including expiring them and asking users to renew them. The same credential can potentially be shared across multiple Central IT data repositories. If a Shadow/Self-IT data sharing site is used, new Ad-Hoc credentials need to be obtained from the local access control system. New credentials need to be created for each different system. No support is provided by Central IT Identity Management, so credentials might not expire and accounts might be retained indefinitely on the sharing site, etc.

Employees (users) play a key role in the model. Ultimately they are the entities that influence and drive the overall data sharing process. Important aspects that need to be represented are their motivations and behaviours. Based on these, users might choose to use different types of data sharing tools and have different postures to policy compliance and risk. This includes their attitudes to use and misuse of their authentication credentials, e.g. by writing them down or sharing them, behaviours which expose the organization to the risks described above.

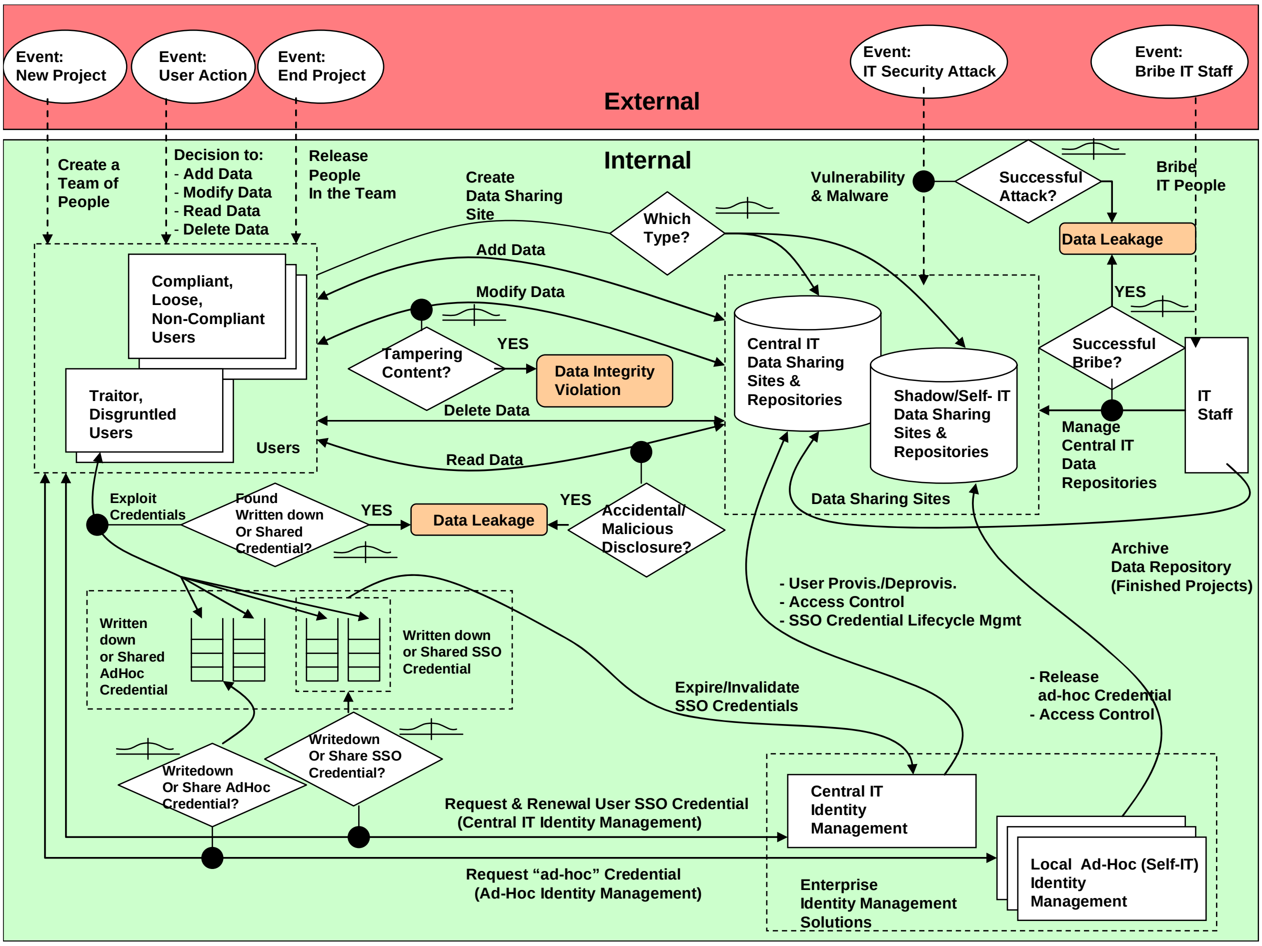
Taking a “common sense” approach, we identified different categories of users, as follows:

- **Compliant Users,** who are policy aware and act to the best of their knowledge to comply with prescribed policies and guidelines;
- **Loose Users,** who are not fully aware of policies or guidelines, and might take actions without fully understanding the implications;
- **Non-Compliant Users,** who might be well aware of policies and guidelines but deliberately act against them, not necessarily because of bad intentions but because they perceive that some of the existing policies can undermine their work and business objectives;
- **Traitors and Disgruntled People,** who deliberately act against the interests of an organization and create potential losses, for whatever reason that motivates them

Depending on the actual “distribution” of these behaviours within the population of an organization, different dynamics and outcomes are likely to happen. We assumed that users are going to have different approaches to how they handle their credentials, according to their profiles. In particular we focused on misbehaviours, e.g., that users might decide to write down their credentials or decide to share them with colleagues. We made the hypothesis that people that have “loose”, “non-compliant” or “disgruntled” profiles will be more inclined to misuse their credentials. We also assumed that traitors or disgruntled users might actually leverage these credentials for criminal intent. In this case it is likely that data will be leaked i.e. credentials misused to disclose valuable stored data to a third party.

We considered the impact that identity management solutions could have in mitigating these risks. In particular, if credentials are managed by Central IT Identity Management solutions, they will automatically expire at the due time, reducing in this way the exposure risk. In the case of Ad-Hoc credentials no risk mitigation is in place.

Figure 4 shows a comprehensive view of the various aspects that have been included in the model, with particular emphasis on entities (users, data repositories, identity management solutions, etc.), their interactions, and threats. Additional threats we considered and factored in are: (1) users accidentally leaking data, for example sharing it by email with unauthorized people; (2) users maliciously tampering with the integrity of stored documents; (3) IT Staff could be bribed, and data could be leaked as a consequence of this.



Event: New Project
Event: User Action
Event: End Project
Event: IT Security Attack
Event: Bribe IT Staff
External
Internal
Create a Team of People
Decision to:
- Add Data
- Modify Data
- Read Data
- Delete Data
Release People In the Team
Create Data Sharing Site
Which Type?
Vulnerability & Malware
Successful Attack?
Bribe IT People
Data Leakage
Add Data
Modify Data
Compliant, Loose, Non-Compliant Users
Traitor, Disgruntled Users
Users
Tampering Content?
YES
Data Integrity Violation
Delete Data
Read Data
Central IT Data Sharing Sites & Repositories
Shadow/Self- IT Data Sharing Sites & Repositories
Data Sharing Sites
YES
Successful Bribe?
IT Staff
Manage Central IT Data Repositories
Exploit Credentials
Found Written down Or Shared Credential?
YES
Data Leakage
YES
Accidental/ Malicious Disclosure?
Archive Data Repository (Finished Projects)
Written down or Shared AdHoc Credential
Written down or Shared SSO Credential
- User Provis./Deprovis.
- Access Control
- SSO Credential Lifecycle Mgmt
Expire/Invalidate SSO Credentials
- Release ad-hoc Credential
- Access Control
Writedown Or Share AdHoc Credential?
Writedown Or Share SSO Credential?
Request & Renewal User SSO Credential (Central IT Identity Management)
Request “ad-hoc” Credential (Ad-Hoc Identity Management)
Central IT Identity Management
Local Ad-Hoc (Self-IT) Identity Management
Enterprise Identity Management Solutions

Fig. 4 Diagram Showing the Breadth of Complexity of the Model

A few factors of interest to decision makers have been initially identified as potential outcomes to measure and analyse. All these "coarse-grain" indicators are calculated during the simulation process by keeping account of the combined effects of all the involved entities, human behaviours and processes. In this example we focused on the following indicators both for Central-IT and Shadow/Self-IT data sharing sites:

- **Number and Value of "Leaked Documents"**: number and value of leaked documents as an effect of having written down passwords, shared accounts, having security weaknesses on sharing sites, etc. In other words, this give a very raw indication of the "exposure surface";
- **Number of Written Down SSO and Ad-Hoc credentials**: number of credentials (both SSO and Ad-Hoc) that users have written down;
- **Number of Shared SSO and Ad-Hoc credentials**: SSO and Ad-Hoc credentials that have been shared between people;
- **Number of exploited SSO and Ad-Hoc credentials**: number of credentials that have been actually misused with consequent damage for the organization (e.g. in terms of data leakage);
- **Number of attacked Central IT and Shadow/Self IT sites**: number of sites that have been successfully attacked (e.g. by external hackers) due to their security vulnerabilities or lack of security compliance.

4.3 Simulation and Experimental Results

Experiments have been carried out by running simulations of the described model, using Demos2k and giving initial values to various parameters, including the above mentioned probability distributions. The model has been designed to allow us to simulate the actions carried out by each individual on a daily basis. Snapshots of all the involved indicators have been captured on a monthly basis with an overall simulation time period of two years.

Here we provide an overview of some of the potential outcomes and analysis that could be carried out in the space of Identity Analytics. As this is an initial exploratory study, we make some educated guesses for the probability distributions based on our and other experts' intuitions about the situations. These settings can be changed as we refine the experiment and understand the details of a specific scenario. This allows the area to be explored, with the next step being to carry out empirical studies that better ground these distributions. Parameters in the model for which values were chosen include:

- Relative proportions of each category of user (compliant, loose, non-compliant, etc.) in the population of employees;
- Rates and relative frequencies of user actions (i.e. read, write, delete, modify data);
- Likelihoods of different misbehaviours, varied by category of user;
- Rates of other attempted attacks, such as by IT staff, and likelihoods of success;
- Various lifecycle parameters for projects and for credentials provided by central IT and Shadow/Self IT Identity management.

For example, the initial population of users was set to 10% compliant, 58% loose, 30% non-compliant, 1% disgruntled and 1% traitors, with each project member carrying out an average of 1 action per day, with a probability of accidentally revealing shared data of 0.001.

Based on these settings, simulations have been used to generate experimental values for the indicators of interest described above. Figure 5 provides an example of results obtained from repeated Monte Carlo based experiments, measured from our model.

Of course, these outcomes depend on our initial settings and assumptions, so the following discussion is not a detailed analysis of the specific quantitative results, rather an illustration of possible types of outcomes and analysis that can be carried out with Identity Analytics, and which would be interest to decision makers.

Figure 5.A illustrates the number of overall data leakages and the cumulative value of disclosed documents. These numbers increase rapidly in the second year of the simulated period. This can be related to an increase in the number of written down (WD) and shared SSO and Ad-Hoc (AH) credentials (Figure 5.B) and an increase in the number of active projects (Figure 5.D), each of them associated with data sharing sites: in particular Shadow/Self-IT (SIT) data sharing sites persist over time, as no active management or archival is provided. Finally, Figure 5.C shows how many written down or shared credentials have been actually exploited. We made the assumption that these credentials could be exploited multiple times by individuals, as long as they are available.

From a different perspective, Figures 5.B and 5.C show the impact of Central IT Identity Management on the management of SSO credentials. Specifically the automatic expiry of credentials reduces the number of exploitable (written down or shared) SSO credentials. However, in this specific case, its effect is minimal due to (1) the users' "bad behaviours" and their increasing number of written down and shared Ad-Hoc credentials, and (2) lack of adoption of Central-IT (CIT) data sharing sites.

We designed the model to support three main levers that decision makers can act on to explore the implications of their investment decisions: investments in automation and security, including further centralization of identity management solutions; investments in user education, via training courses, awareness campaigns, etc.; and investments in detecting misbehaviours and punishment via HR. In this example, decision makers looking at the outcomes depicted in Figure 5 might deduce that investments in automation and Identity Management solutions might have little value if no action is taken to (1) change the way people behave, which choices they make and how securely they handle their credentials, and (2) improve the compliance with security policies for data sharing sites, in particular for Shadow/Self IT data sharing.

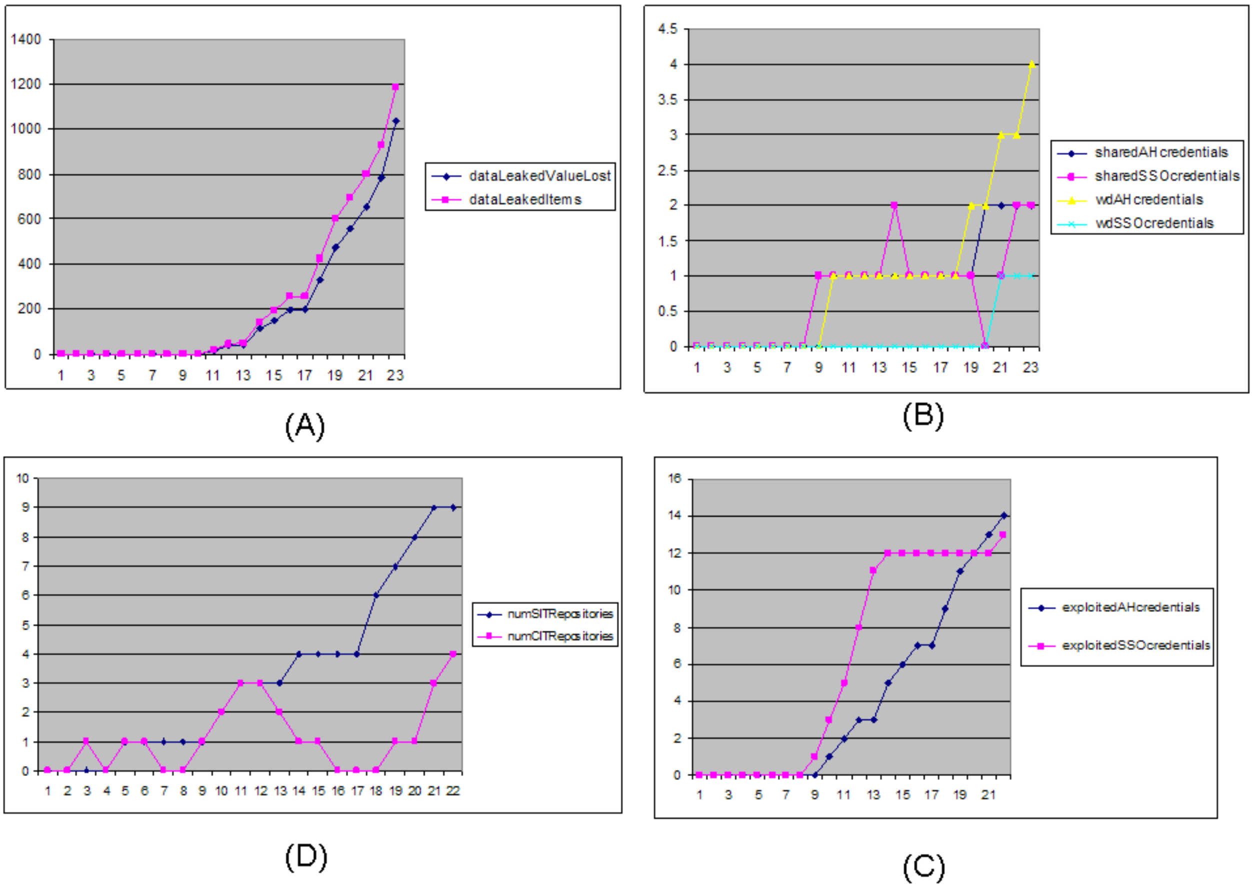


Fig. 5 Some Outcomes of Simulations

The power and strength of this Identity Analytics approach, driven by modelling and simulation, is that it can help decision makers to explore trade-offs and carry out what-if analysis on aspects that might not necessarily be so intuitive. Given a model that reflects the current situation, a decision maker could play out the potential effects of a decision to explore its value and impact. Variation in the outcomes of the model, given different possible effects of the decision, may help design metrics that allow the decision maker to ensure that the new policies are working as required.

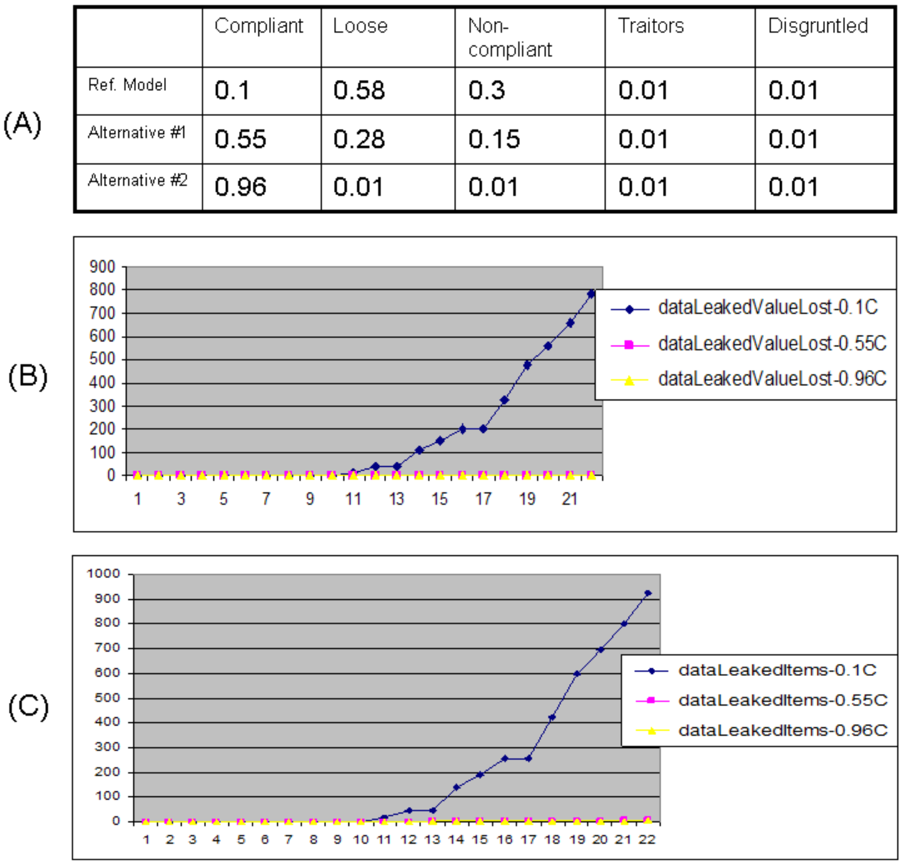
As an illustration, we considered the case where decision makers are interested in investing in actions that cause the distribution of the employee population to be shifted towards more policy-compliant users. Changing this distribution might be achieved by acting on one or more levers, including education, training, monitoring & detection and punishment. The outcomes of related experiments are shown in Figure 6. It is beyond the scope of this specific analysis to investigate how to achieve this: we aim at providing an indication of what would be the outcomes if this happens.

Specifically, Figure 6.A shows a possible what-if analysis, starting from the initial situation described at the beginning of this section, and considering two cases for shifting the distribution of the population towards greater levels of compliance—alternative #1: 55% population is compliant, alternative #2: 96% population is compliant—all other assumptions and parameters being the same. Figure 6.B and 6.C respectively show the impact in terms of amounts of leaked data and involved losses.

The outcomes shown in Figure 6 confirm that the amount of leaked data can be reduced substantially, by ensuring that there are a greater number of compliant users. The model is very sensitive to the distribution of users' behaviours. Specifically, based on the assumptions made, the data leakage threat can be drastically minimized by ensuring the compliance of at least 55% of the users. The results also suggest that achieving alternative #2, with 96% compliant users, would produce a relatively small further reduction in data leakage.

Decision makers might conclude that the most effective strategy would be to keep constant their current investments in Identity Management, and that new investments need to be made in education, detection & punishment to effectively change user behaviour and thus mitigate risks related to data leakage. As a consequence, they might decide to define security policies mandating this.

The next stage in the modelling process would be to validate and refine these assumptions with experts and carry out experiments to validate or refine the basic probability distributions used within the simulation. However, this initial model demonstrates the power of this approach, in providing qualitative predictions in complex situations.



(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

Fig. 6 Some Simulation Outcomes

5. Related Work

Identity Analytics is currently an "overloaded" term, with multiple meanings. It is used to refer to: (1) analysis of personal data and profiles, in order to extract meaningful "identity patterns", characterizing individuals or classes of individuals; (2) analysis and processing of organizational systems' log files, events and configuration information to assess its compliance to guidelines, policies and legislation, in the space of identity management, privacy and security, and report violations; (3) provision of indications to the management team about potential risks and security exposures an organization might incur, as a further processing step of information gathered by reporting solutions described in the previous point. There is no agreed, common definition of Identity Analytics.

Most current commercial work, solutions and approaches that make claims in the Identity Analytics area are "bottom-up"-driven solutions, dealing with compliance and governance issues. Their main functionalities are around analysis of log files, events to report on compliance and violations based on current processes, policies and guidelines. Solutions in this space provide indications of risk levels and exposures, based on predefined priorities and processes, e.g. [26,27]. Some other initiatives mentioning Identity Analytics capabilities are pretty much about business intelligence and data mining of identity information, for profiling purposes, e.g. [28,29,30]. This work is complementary to what we are aiming to do. They can provide observational and factual data in specific contexts, by processing and analysing identity and other information collected within the organization. Our approach to Identity Analytics is driven by decision makers' needs and aims at exploring and predicting the impact of their decisions along with possible trade-offs in making investment choices. It is a top-down approach, driven by models of scenarios and contexts under examination, based not only on current situations but also hypothetical ones (what-if analysis), along with related simulations and analysis of results. Existing solutions in the Identity Analytics space focusing on compliance management only help decision makers to assess decisions and policies that have already been made in an organization. Instead, our work focuses on the current shift from compliance to risk management and can provide upfront support to decision makers, at the decision making time. Decision makers using our approach based on modelling and simulation will be able to understand the implications of their possible decisions (before actually making them), choose the most suitable trade-offs, shape policies and/or justify current ones.

There is related work in Identity Management and Privacy using modelling and simulation, but only in narrow and specific IdM areas, such as on formulating password policies [31,36] and the role of cooperation on addressing identity phishing [37]. This is important work and provides valuable analysis and experimental data that can be leveraged in our work. However, we are not aware of any current research or commercial work that aims at modelling and simulating the overall complexity and different identity management dimensions (i.e. technologies, human behaviour, various interactions between involved entities, enterprise processes, legislation, etc.) that combine to influence an organization and that need to be taken into account when making strategic investment decisions. We are also not aware of related analysis of trade-offs (by factoring in economics aspects) involving identity management, taking into account this underlying complexity. Our work focuses exactly on these aspects and aims at using modelling and simulation techniques to cope with this complexity and provide useful decision support capabilities in this space.

Standards such as ISO 27001 [32], CoBit [33], ITIL [34] describe best practices and methodologies respectively in terms of information security management, IT governance and service management. These standards define valuable common methodologies and guidelines on how to address these management aspects, including aspects of Identity Management. Decision makers still need to understand, interpret and instantiate them in their specific operational environments. We can 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 and related needs and predicting the impacts. Further they represents a one size fits all approach to security and companies wanting to move from a compliance driven to a risk driven mentality need tools to understand the impacts of deviating from best practice.

Our work in Identity Analytics relies on mathematical models and related simulations. The use of mathematical models in engineering has a long and distinguished record of success. From earthworks to suspension bridges, from bicycles to spacecraft, mathematical models are used to predict behaviour and give confidence that necessary properties of the constructions—such as capacity, resilience, and cost—obtain. Such applications of applied mathematics in engineering are useful, and usable, by virtue of the scientifically rigorous modelling methodology, where observations about the external environment and the parameters that the system depends upon are interpreted and a range of properties of the mathematical model are deduced. In the worlds of traditional engineering, ranging over mechanical, civil, environmental and electrical/electronic engineering, the mathematical methods used are mainly concerned with continuous phenomena and typically use techniques from calculus such as differential equations.

For modelling security and identity management operations the appropriate mathematical methods are more discrete, being drawn from algebra, logic, theoretical computer science, and probability theory. In order to apply these methods, we require a conceptual analysis of the relevant aspects of the systems of interest. In our work, we leverage the seminal work done by HP Labs in the Open Analytics project [19,20,21], that we will consider as a reference. We are using Demos2k [22,23] as the reference tool for our modelling and simulation activities. Finally, an important aspect of Identity Analytics is the studies in the space of Social Science. We aim to leverage work done in this space, such as [18], in order to build mathematical models that realistically reflect users' behaviours and the associated impact.

6. Discussion and Next Steps

Identity Analytics is a greenfield area, open to innovation and contributions. There is the unique opportunity to explore the space of identity and identity management from a different perspective, i.e. not only technology-driven but also taking into account social, behavioural and economical aspects. Whilst the area of identity management is quickly maturing and commoditizing from a technological and solution perspective, little has been done so far to understand implications of these solutions in complex enterprise contexts by factoring in human motivations, policies, social aspects and legislation.

The main research opportunity we see in this space is in "turning the table around" and focusing on the decision makers' perspective (rather than on the usual IT perspective), by providing decision support tools and solutions that allow them to explore and predict the impact and consequences of their decisions, by taking into account all the above aspects—on factors that are of relevance to them, such as costs, security risks and exposures, financial losses and impacts on trust and reputation. We also observe that this kind of approach can bring shared stakeholder understanding and justification for identity investments. This suggests it is a promising area that will support the current enterprise trends in the strategic/executive decision-making area from a compliance-driven approach to a risk-driven approach.

Work in this space can potentially be very challenging, as the predictive capabilities of models and simulations depend on the availability of observational data, expertise in this space, access to CIOs/CISOs and their perspectives and validation of the outcomes on the field. The collection of observational data could be particularly challenging too, due to the potential lack of this information and related statistics, especially in emerging scenarios.

Related challenges include: (1) how in the absence of data, to construct models that can provide structural insights about the effects of certain factors (e.g. where the outcome is sensitive to small changes in assumptions and where it is not); (2) how to properly model and simulate complex human and social behaviours, particularly in large populations where different categories of behaviours could apply and where these behaviours could change over time. Equally, bringing economic ideas of understanding the incentives behind individuals and organizational behaviours can help build more realistic models.

We are planning to carry on our explorations in this space and identify suitable modelling and simulation approaches and tools to deal with the involved complexity. We are also planning to investigate other potential approaches to (mathematical) modelling and simulation, compare them and understand their pros and cons, driven by our needs and objectives in this space. Alternative/complementary approaches we are planning to investigate include: probabilistic rule-based modelling: modelling based on swarm theory and swarm intelligence: modelling based on chaos theory, in particular exploring this theory from an organizational perspective; probabilistic agent-based modelling and simulations: game theory, game modelling and simulations.

At the current stage, we do not exclude the fact that hybrid approaches, using two or more different modelling and simulation approaches might be required in Identity Analytics.

7. Conclusions

This paper introduced the concept of "Identity Analytics" as a mechanism and tool to explain and predict the impact of identity, identity management and other related aspects (such as users' behaviours, security, legislation and social aspects) on key factors of relevance to decision makers (e.g. CIOs, CISOs), in complex enterprise scenarios, based on their initial assumptions and investment decisions. The goal is to provide decision support and "what-if" analysis to decision makers, to explore possible trade-offs (e.g. using technologies vs. changing processes vs. investing in education of users, to change their behaviours) driven by an economic perspective and formulate new policies or justify existing ones.

We discussed our vision and the adopted methodology, based on the adaptation of the "scientific method" to this domain. It is a "top-down" approach, leveraging and applying modelling and simulation techniques to the identity management (and security) space, jointly with aspects of social science and economics. To ground some of the Identity Analytics concepts illustrated in this paper, we discussed a significant case study, focusing on emerging "web 2.0" enterprise data sharing scenarios, where "unstructured" information is created, stored and shared by people in collaborative contexts, within and across organizations. We discussed (as a significant example) some qualitative outcomes and what-if analysis that can be provided by Identity Analytics. We believe this area will provide plenty of research opportunities as well as challenges to overcome, in terms of identifying suitable scenarios, gathering relevant observational data, being able to access the expertise and judgment of CIOs/CISOs and validate work in this space in real-world scenarios, by means of trials. We described some existing challenges, our plans along with our next steps.

8. 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, SSL, http://www.hpl.hp.com/research/systems_security.html, 2008

[3] Birch, D., Digital Identity Management: Technological, Business and Social Implications, Book, 2007

[4] Windley, P., Digital Identity, O' Reilly, 2005

[5] Pato, J., Identity Management: Setting the Context, HPL Technical Report, HPL-2003-72, 2003

[6] Casassa Mont, M, Bramhall, P., Pato, J., On Adaptive Identity Management: The Next Generation of Identity Management Technologies, HPL Technical Reports, HPL-2003-149, 2003

[7] Casassa Mont, M, Baldwin, A., Shiu, S., On Identity Analytics: Setting the Context, HPL Technical Report, HPL-2008-84, 2008

[8] Aliprandis, C.D. (Editor), Economic Theory Journal, http://www.springer.com/economics/economic+theory/journal/199, 2008

[9] Nobay, R.A., Peel, D.A., Optimal Monetary Policy in a Model of Asymmetric Bank Preferences. London School of Economics, Mimeo.

[10]Varian, H., A bayesian approach to real estate management. In S.E. Feinberg and A. Zellner, editors, Studies in Bayesian Economics in Honour of L.J. Savage, pages 195–208. North Holland, 1974

[11]Anderson, R., Bohme, R., Clayton, R., Moore, T., Security economics and the internal market, European Network and Information Security Agency, 2007

[12]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

[13]Anderson, R., Why information security is hard: An economic perspective. In Proc. 17th Annual Computer Security Applications Conference, 2001

[14]Gordon, L.A., Loeb, M.P., The Economics of Information Security on Information and Systems Security, 5(4):438–457, 2002

[15]Yearworth, M., Monahan, B., Pym, D., Predictive modelling for security operations economics (extended abstract). In Proc. I3P Workshop on the Economics of Securing the Information Infrastructure, Proceedings at http://wesii.econinfosec.org/workshop/, 2006

[16]Beautement, A., Coles, R., Griffin, J., Ioannidis, C., Monahan, B., Pym, D., Sasse, A., Wonham, M., Modelling the human and technological costs and benefits of USB memory stick security, to appear in Proceedings of WEIS 2008, 2008

[17]Wilson, E. B., An Introduction to Scientific Research McGraw-Hill, 1952

[18]Trust Economics, UK DTI grant P0007, Trust Economics Project, 2008

[19]Pym, D., Taylor, R., Tofts, C., Yearworth, M., Monahan, B., Gittler, F., Systems and services sciences: a rationale and a research agenda (Open Analytics Project, HP Labs, Bristol, UK), http://www.hpl.hp.com/techreports/2006/HPL-2006-112.html, 2006

[20]Taylor, R., Tofts, C., Model Based Services Discovery and Management, PICMET 2008, 2008

[21]Taylor, R., Tofts, C. Taking a RaSP to Enterprise Stakeholder Dissonance, accepted EDOC 2008, 2008

[22]Demos2k, Demos 2k, http://www.demos2k.org/, 2000

[23]Birtwistle, G., Demos,discrete event modelling on Simula. Macmillian, 1979

[24]Pym, D., Monahan, B., A Structural and Stochastic Modelling Philosophy for Systems Integrity. HP Labs Technical Report Series, HPL-2006-35, Feb 2006

[25]Brian Monahan, DXM - The Demos eXperiments Manager, HP Labs Technical Report, 2008

[26]SailPoint, SailPoint – Identity Risk Management, http://www.sailpoint.com/product/reporting.php, 2008

[27]Beres, Y., Baldwin, A., Shiu, S., Model-based Assurance of Security Controls, HPL Technical Report, HPL-2008-7, 2008
[28]Oracle, Reporting and Auditing Solutions Roadmap, ftp://ftp.oracle.com/sales/outgoing/oam/roadmap.pdf, 2008
[29]IBM, Identity Analytics, http://www-935.ibm.com/services/us/gbs/bus/pdf/g510-6527-ibm-identity-risk.pdf, 2008
[30]IdAnalytics, IdAnalytics, http://www.idanalytics.com/, 2008
[31]Shay, R., Bhargav-Spantzel, A., Bertino, B., password policy simulation and analysis, DIM 2007, 2007
[32]ISO, ISO 27001, Information Security Management, http://www.iso.org/iso/catalogue_detail?csnumber=42103, 2005
[33]ISACA, Cobit, IT Governance, http://www.isaca.org/, 2008
[34]ITIL, ITIL IT Infrastructure Library for Service Management, http://www.itil-officialsite.com/home/home.asp, 2008
[35]Grimmett, G. and Stirzaker, D. "Probability and Random Processes", 3rd ed., Oxford UP, 2001
[36]Adams, A, Sasse, M.A., Users are not the enemies, Communications of the ACM, Volume 42, Issue 12, pages 40-46, 1999
[37]Moore, T., Clayton, R., The Consequence of Non-Cooperation in the Fight Against Phishing, 3rd APWG eCrime Researchers Summit. 2008,
[38]Wikipedia, Unstructured data, http://en.wikipedia.org/wiki/Unstructured_data, 2008

Appendix A: Demos2k Model – Case Study: Collaborative Data Sharing

This section provides additional details about the Demos2K model [22,23,24] we created to carry out the simulations and experiments illustrated in Section 4. The complete model (executable in a Demos2k simulation shell) follows.

```
(*  Demos 2k Model - Generic "Collaborative Data Sharing" involving People, Data Repositories, Identity Management Controls:

   - Users/Researchers. Users have different profiles and motivations
   - Project gouping a set of users/reserchers. A project has a lifecycle. It has a data repository associated
   - Data Repository stores project's data. It can be managed by Central IT Staff or managed ad-hoc by a user
   - ThreatAgents (entities external or itnernal to the project, e.g. hackers, bribed IT Staff, traitors and disgruntled people, etc.)
   - Identity Management solutions (e.g. User Provisioning/Deprovisioning and credential management) mitigating security risks

 *)

// Demos2k Directives
//* LIVELOCK-STEPS : 100000
//* SPAWN-LIMIT : 100000

// Timescaling constants
cons days       = 1;        // time unit = days
cons hrs        = days/24;
cons hours      = days/24;  // alternative spelling
cons mins       = hrs/60;
cons secs       = mins/60;
cons msecs      = secs/1000;

cons weeks      = 7 * days;
cons months     = 4 * weeks;
cons years      = 365 * days;
cons centuries  = 100 * years;


//simulation time constants
cons simulationAdvancementTimePeriod = days;
cons observedTimePeriod = months;
cons simulationTimeframe = 2*years;

//Experiments: dxm tool - parameters
cons noparam = 0;

//General constants
cons u_num = 100;                              // number of involved users
cons p_researcher = puni (1, u_num);           // select a random user (researcher)
cons interActionInterval = negexp (days);      // one action per day on average
cons interProjectInterval = negexp (months);   // time elapsed between creation of two projects
cons minProjectSize =2;                          // min size of a Project
cons maxProjectSize =12;                         // max size of a Project
cons projectSize = puni (minProjectSize, maxProjectSize);    // size of a project
cons projectDuration = negexp (3*months);       // duration of a project
cons interITActionInterval = negexp (weeks);    // time elapsed between actions carried out by IT Staff
cons archiveProjectWait = negexp (months);      // time elapsed before a project is archived by IT Staff


// General Properties of Data Repositories (underpinning Clllaborative Data Sharing Sites)

//Types of data repositories
cons DATAREPOSITORY_TYPE_CIT = 1; // Central IT (CIT) Data Repository
cons DATAREPOSITORY_TYPE_SIT = 2; // Self/Shadow/Ad-Hoc IT (SIT) Data Repository

cons secITComplianceCIT = 0.9999;  // average IT security compliance (patching, AV, etc.) for CIT data repositories
cons secITComplianceSIT = 0.2;     // average IT security compliance (patching, AV, etc.) for SIT data repositories

cons p_secITattackCIT = binom (1, 1-secITComplianceCIT);  // probability of success of CIT security attack
cons p_secITattackSIT = binom (1, 1-secITComplianceSIT);  // probability of success of SIT security attack


// User credential management

cons SSOCredExpirationTime = 6*months; // Expiration Time defined by Central IT for managed SSO credentials/passwords


// Threat likelihoods

cons testAccidentalReveal = binom (1, 0.001);
cons testIntegrityBreach = binom (1, 0.001);
cons interBribeInterval = negexp (10 * years / u_num);  // each employee can expect to be approached once every 10 years on average
cons testITBribable = binom (1, 0.1);                   // This represents the chance that _any_ member of IT can be found to bribe
cons interITBribeInterval = negexp (1 * years);         // once a year on average somebody tries to bribe IT

cons credentialBasedUserAttackInterval = negexp(weeks);   // frequency by which a traitor/disgruntled employee tries to
                                                          // exploit written-down or shared credentials

cons interITAttackInterval = negexp (3 * months);  // frequency of external attack attempts on Data Repositories
cons ATTACK_BRIBE = 1;

cons testFindingWrittenDownSSOCredential = binom (1, 0.01); // probability of finding a written down SSO Credential
cons testFindingWrittenDownAHCredential = binom (1, 0.01);  // probability of finding a written down AH Credential

cons testReceivedSharedSSOCredential = binom (1, 1/((minProjectSize+maxProjectSize)/2));  // probability of having received a shared SSO Credential
cons testReceivedSharedAHCredential = binom (1, 1/((minProjectSize+maxProjectSize)/2)); // probability of having received a shared AH Credential


//User Definition - Constants

// Different types of user behaviours, based on a distribution

// type 0: compliant (to policies, guidelines, etc.)
// type 1: loose
// type 2: non-compliant
// type 3: traitor
// type 4: disgruntled

cons p_userBehaviour= pud[(0.1,0),(0.58,1),(0.3,2), (0.01,3), (0.01,4)]; //probability distribution defining the distribution of user behaviours
cons userTypeBehaviour_Compliant = 0;       // type 0: compliant (to policies, guidelines, etc.)
cons userTypeBehaviour_Loose= 1;            // type 1: loose
cons userTypeBehaviour_NonCompliant = 2;    // type 2: non-compliant
cons userTypeBehaviour_Traitor = 3;         // type 3: traitor
cons userTypeBehaviour_Disgruntled = 4;     // type 4: disgruntled


// Possible way to define different type of users' roles
// Current simple assumptions: (1) a user has a main role, during the entire simulation time (to be double-checked ...)
//                             (2) this role is statistically determined, e.g. based on project population???
// NOTE: For the sake of simplicity, in this model we assume all users are "researchers"

cons p_userRole= pud[(1.0,0),(0.0,1),(0.0,2), (0.0,3), (0.0, 4) ]; //probability distribution defining users' roles
cons userRole[0]= 0;        // type 0: researcher
cons userRole[1]= 1;        // type 1: clerck/admin
cons userRole[2]= 2;        // type 2: manager
cons userRole[3]= 3;        // type 3: IT staff
cons userRole[4]= 4;        // type 4: other

// User Activities - constants
// User's actions (on data repositories) are more primitive than user's activities
// User actions involved data

cons ACT_CREATE = 1;
cons ACT_READ = 2;
cons ACT_WRITE = 3;
cons ACT_DELETE = 4;

cons p_researcherAction = pud [ (0.04, ACT_CREATE), (0.01, ACT_DELETE), (0.25, ACT_WRITE), (0.70, ACT_READ) ]; //probability of each action


// User/researcher actions involved in handling/managing SSO and Ad-Hoc credentials

cons ACT_IDM_CREATE_CRED = 1;
cons ACT_IDM_RENEW_CRED = 2;


// Probability of choosing a type of data repository based on user's profile/behaviour
// User profiles

// type 0: compliant (to policies, guidelines, etc.)
// type 1: loose
// type 2: non-compliant
// type 3: traitor
// type 4: disgruntled

cons p_chooseDataRepositoryType[0] = pud [ (0.99, DATAREPOSITORY_TYPE_CIT), (0.01, DATAREPOSITORY_TYPE_SIT)];
cons p_chooseDataRepositoryType[1] = pud [ (0.50, DATAREPOSITORY_TYPE_CIT), (0.50, DATAREPOSITORY_TYPE_SIT)];
cons p_chooseDataRepositoryType[2] = pud [ (0.20, DATAREPOSITORY_TYPE_CIT), (0.80, DATAREPOSITORY_TYPE_SIT)];
cons p_chooseDataRepositoryType[3] = pud [ (0.50, DATAREPOSITORY_TYPE_CIT), (0.50, DATAREPOSITORY_TYPE_SIT)];
cons p_chooseDataRepositoryType[4] = pud [ (0.10, DATAREPOSITORY_TYPE_CIT), (0.90, DATAREPOSITORY_TYPE_SIT)];

// Probability of user of writing down a credentail/password based on their profile

// SSO Credentials

// Probability of writing down credentials
cons p_writeDownCredential[0] = binom (1, 0.001);
cons p_writeDownCredential[1] = binom (1, 0.01);
cons p_writeDownCredential[2] = binom (1, 0.05);
cons p_writeDownCredential[3] = binom (1, 0.01);
cons p_writeDownCredential[4] = binom (1, 0.1);
// Probability of sharing credentials
cons p_shareCredential[0] = binom (1, 0.001);
cons p_shareCredential[1] = binom (1, 0.01);
cons p_shareCredential[2] = binom (1, 0.05);
cons p_shareCredential[3] = binom (1, 0.01);
cons p_shareCredential[4] = binom (1, 0.1);


// Ad-hoc Credentials

// Probability of writing down credentials
cons p_writeDownAdHocCredential[0] = binom (1, 0.001);
cons p_writeDownAdHocCredential[1] = binom (1, 0.02);
cons p_writeDownAdHocCredential[2] = binom (1, 0.07);
cons p_writeDownAdHocCredential[3] = binom (1, 0.01);
cons p_writeDownAdHocCredential[4] = binom (1, 0.1);
// Probability of sharing credentials
cons p_shareAdHocCredential[0] = binom (1, 0.001);
cons p_shareAdHocCredential[1] = binom (1, 0.02);
cons p_shareAdHocCredential[2] = binom (1, 0.07);
cons p_shareAdHocCredential[3] = binom (1, 0.01);
cons p_shareAdHocCredential[4] = binom (1, 0.1);


//Data - constants

// Different types of data and their respective (relative) value in a [0,10] range of relevance

cons p_dataType= pud[(0.7,0),(0.249,1),(0.051,2)]; //probability distribution defining users' roles
cons dataType[0] = 0;  // no value/low value
cons dataType[1] = 1;  // confidential
cons dataType[2] = 2;  // private

// Value of dataTypes [0-10] range.
cons dataValue[0] = 0;
cons dataValue[1] = 3;
cons dataValue[2] = 10;

// run control
```

```
var demos_sample_tick   = 0;
var done = 0;


// Variables

var userIndex = 0;

// Security Metrics
// Metrics based on Confidentiality, Integrity and Availability aspects

var dataLeakedItems = 0;
var dataLeakedValueLost = 0;
var dataIntegrityBreaches = 0;
var dataIntegrityValueLost = 0;
var dataAvailabilityBreaches = 0;
var dataAvailabilityValueLost = 0;
var successfulBribes = 0;
var wdSSOcredentials = 0;
var wdAHcredentials = 0;
var sharedSSOcredentials = 0;
var sharedAHcredentials = 0;
var exploitedSSOcredentials = 0;
var exploitedAHcredentials = 0;
var successfulCITexternalAttacks = 0;
var successfulSITexternalAttacks = 0;
var numCITRepositories = 0;
var numSITRepositories = 0;

// Measures of activity and productivity

var liveProjects = 0;
var busyResearchers = 0;
var totalActions = 0;

bin (freeResearchers, 0);
bin (projectsTrackedByIT, 0);
bin (projectsAdHoc, 0);
bin (CentralITSSOUserCredentials, 0);
bin (AdHocUserCredentials,0);

bin (WrittenDownSSOUserCredentials, 0);
bin (WrittenDownAdHocUserCredentials,0);

bin (SharedSSOUserCredentials, 0);
bin (SharedAdHocUserCredentials,0);


// Class initialising all other classes

class initialise =
{
   local var userRol = -1;
```

```
    do u_num {
        userRol := p_userRole;
        entity(researcher, researcher(#userRol), 0);
     }
    entity (projectGenerator, projectGenerator, 0);
    bin (fileShareBin, 0);
    entity (fileShare, fileShare (fileShareBin), 0);
    entity (ITStaff, ITStaff, 0);

    entity (IdMCentralITUserCredentialProvisioning, IdMCentralITUserCredentialProvisioning, 0);
    entity (IdMCentralITCredentialManagement, IdMCentralITCredentialManagement, 0);

    entity (IdMAdHocUserCredentialProvisioning, IdMAdHocUserCredentialProvisioning, 0);

    entity (generateExternalThreats, generateExternalThreats, 0);

    entity (measurement, measurement, 0);
    hold(simulationTimeframe);
    done :=1;
}

// class keeping track of variables of relevance of this model
class measurement =
{
    repeat {
        trace ("liveProjects=%v", liveProjects);
        trace ("busyResearchers=%v", busyResearchers);
        trace ("totalActions=%v", totalActions);
        trace ("dataLeakedItems=%v", dataLeakedItems);
        trace ("dataLeakedValueLost=%v", dataLeakedValueLost);
        trace ("dataIntegrityBreaches=%v", dataIntegrityBreaches);
        trace ("dataIntegrityValueLost=%v", dataIntegrityValueLost);
        trace ("dataAvailabilityBreaches=%v", dataAvailabilityBreaches);
        trace ("dataAvailabilityValueLost=%v", dataAvailabilityValueLost);
        trace ("successfulBribes=%v", successfulBribes);
        trace ("wdSSOcredentials=%v", wdSSOcredentials);
        trace ("wdAHcredentials=%v", wdAHcredentials);
        trace ("sharedSSOcredentials=%v", sharedSSOcredentials);
        trace ("sharedAHcredentials=%v", sharedAHcredentials);
        trace ("exploitedSSOcredentials=%v", exploitedSSOcredentials);
        trace ("exploitedAHcredentials=%v", exploitedAHcredentials);
        trace ("successfulCITexternalAttacks=%v", successfulCITexternalAttacks);
        trace ("successfulSITexternalAttacks=%v", successfulSITexternalAttacks);
        trace ("numCITRepositories=%v", numCITRepositories);
        trace ("numSITRepositories=%v", numSITRepositories);


        demos_sample_tick :=  demos_sample_tick + 1;
        trace ("demos_sample_tick=%v", demos_sample_tick);
        hold(observedTimePeriod);
    }
}
```

```
// Entity representing an abstraction of Data Repositories (underpinning a Collaborative Data Sharing Site)
// Data repositories are differentiated by type (Central IT vs Self/Shadow/Ad-Hoc managed repositories)
// Entries in the associated "value bin" contain the attributes of the document: <project, type, creation time>

class fileShare (repBin) =
{
    local var cmd = 0;
    local var proj = 0;
    local var proj2 = 0;
    local var createTime = 0;
    local var type = 0;
    local var repositoryType = 0;
    local var repTemp = 0;

    repeat {
    // NB: If any of these actions take (simulation) time and we want to allow concurrent actions, we should change this to recursive style.

        getSV (fileShare, [cmd, proj, repositoryType], true);
        try [cmd == ACT_CREATE] then {
            createTime := DEMOS_TIME;
            type := p_dataType;
            putVB (repBin, [proj, repositoryType, type, createTime]);
        }
        etry [cmd == ACT_READ] then {
            // How do we choose a random document from the repository?
            // Do we have to?  Surely we do at some stage?
            // For the moment we simply choose the element on top of the queue (and put it back!)
            try [getVB (repBin, [proj2, repTemp, type, createTime], proj2 == proj)] then {
                putVB (repBin, [proj2, repTemp, type, createTime]);
            }
            etry [] then {
                //  failure is allowed
                proj := -1;  type := -1;  createTime := -1;
            }
        }
        etry [cmd == ACT_WRITE] then {
            // For the moment, write doesn't change any document properties - easily changed if we want
            try [getVB (repBin, [proj2, repTemp, type, createTime], proj2 == proj)] then {
                putVB (repBin, [proj2, repTemp, type, createTime]);
            }
            etry [] then {
                proj := -1;  type := -1;  createTime := -1;
            }
        }
        etry [cmd == ACT_DELETE] then {
            try [getVB (repBin, [proj2, repTemp, type, createTime], proj2 == proj)] then {
                skip;  // Don't put the document back if we're deleting it
            }
            etry [] then {
                proj := -1;  type := -1;  createTime := -1;
            }
        }
    etry [] then {
      trace ("Error: illegal fileShare command - %v", cmd);
      close;
    }
    putSV (fileShare, [proj, type, createTime]);
  }
}


// Generator of Projects, requiring that a set of people will collaborate together to achieve common goals
// Each project has an associated collaborative data sharing tool +data repository (either managed by Central IT or ad-hoc managed),
// storing collaborative data

class projectGenerator = {
  local var id = 0;
  repeat {
    hold (interProjectInterval);
    id := id + 1;
    entity (project, project (#id), 0);
  }
}

// Entity Handling the lifecycle of a project

class project (id) = {
  local var researcherNumbers = projectSize;
  local var s = AV_freeResearchers;
  local var s1 = s;
  local var sid = 0;
  local var lead = 1;
  local var repositoryType = -1;
  local var repTemp = -1;

  liveProjects := liveProjects + 1;


  // Wait as necessary until there are enough researchers, and grab the first ones that come available
  // Tell all researchers and IT about the start of the new project

  do researcherNumbers {
    // The first member of the project is the lead. He/she will choose the type of data repository
    getVB (freeResearchers, [sid], true);
    syncV (projectStart, [id, sid,lead, repositoryType], [repTemp]);

    // Project lead identified the first time, along with the type of data repository
    try [lead ==1] then
     {lead :=0;
      repositoryType := repTemp;
      trace ("Project %v - repository type: %v number of people: %v", id, repositoryType, researcherNumbers);
      }
    etry [] then {}

  }
```

```
    // Project can be tracked by IT
    try [repositoryType == DATAREPOSITORY_TYPE_CIT] then
     {
       // Only Centrally IT Managed Data Repositories are supported by IT Stuff
       syncV (projectStart, [id, -1, lead, repositoryType], [repTemp]);
       //trace ("Project %v - repository type: %v is Managed by IT Staff", id, repositoryType);
       numCITRepositories := numCITRepositories +1;

     }
    etry [] then
     {
        // Tracking Ad Hoc Data Repositories
        putVB (projectsAdHoc, [id, repositoryType]);
       //trace ("Project %v - repository type: %v is SELF Managed", id, repositoryType);
       numSITRepositories := numSITRepositories +1;

     }

    hold (projectDuration);
    do researcherNumbers + 1
      {
        syncV (projectFinish, [id, repositoryType], []);
      }
    lead :=1;
    liveProjects := liveProjects - 1;
}


// No general class user for the moment
class researcher (userRole) = {
    userIndex := userIndex + 1;
    local var userID = userIndex;
    var userBehaviour[userID] = p_userBehaviour;
    var pid[userID] = -1;

    var dataRepositoryType[userID] = 1;     // data repository type
    var userSSOCredential[userID] = -1;     // SSO credential owned by user, if any
    var userAdHocCredential[userID] = -1;   // ad-hoc credential owned by user, if any


    entity (researcherProject, researcherProject (#userID, #userRole, userBehaviour[userID], pid[userID], dataRepositoryType[userID], userSSOCredential[userID], userAdHocCredential[userID]), 0);
    entity (researcherCredentialHandling, researcherCredentialHandling (#userID, #userRole, userBehaviour[userID], pid[userID], dataRepositoryType[userID], userSSOCredential[userID], userAdHocCredential[userID]), 0);
    entity (researcherAction, researcherAction (#userID, #userRole, userBehaviour[userID], pid[userID], dataRepositoryType[userID], userSSOCredential[userID], userAdHocCredential[userID]), 0);
    entity (researcherInteraction, researcherInteraction (#userID, #userRole, userBehaviour[userID], pid[userID], dataRepositoryType[userID], userSSOCredential[userID], userAdHocCredential[userID]), 0);

     try [userBehaviour[userID] == userTypeBehaviour_Traitor || userBehaviour[userID]  == userTypeBehaviour_Disgruntled] then
       {
         entity (researcherAttacks, researcherAttacks (#userID, #userRole, userBehaviour[userID], pid[userID], dataRepositoryType[userID], userSSOCredential[userID], userAdHocCredential[userID]), 0);
       }
     etry[] then {}
```

```
}


// Track whether the researcher is a member of a project or "resting"
class researcherProject (userID, userRole, userBehaviour, pid, dataRepositoryType, userSSOCredential, userAdHocCredential) = {
    local var id2 = 0;
    local var pid2 = 0;
    local var lead = 0;
    local var repositoryType = 0;
    local var repTemp = 0;
    local var repTemp2 = -1;
    repeat {
        putVB (freeResearchers, [userID]);
        getSV (projectStart, [pid, id2, lead, repTemp], id2 == userID);

        // The first member of the project is the lead. He/she will choose a data repository type, based on profile
        try [lead ==1] then
         {
           repositoryType := p_chooseDataRepositoryType[userBehaviour];

           //trace ("Researcher %v (type: %v) is the lead of project: %v - repositoryType: %v", userID, userBehaviour, pid, repositoryType);

         }
        etry[] then
         {
           repositoryType := repTemp;
           //trace ("Researcher %v (type: %v) is a member of project: %v - repositoryType: %v", userID, userBehaviour, pid, repositoryType);
          }
        dataRepositoryType := repositoryType;
        busyResearchers := busyResearchers + 1;
        putSV (projectStart, [repositoryType]);

        // Handling User credential

        // Case of Data Repository managed by Central IT?
        try [dataRepositoryType == DATAREPOSITORY_TYPE_CIT] then
          {
            // check if user has a SSO credential
            try [userSSOCredential == -1] then
             {
               // need to get a SSO Credential
                  syncV(UserSSOCredentialMgmt,[userID, ACT_IDM_CREATE_CRED],[]);
              }
            etry[] then {}
          }
        etry[dataRepositoryType == DATAREPOSITORY_TYPE_SIT] then
          {
               // No matter what ad hoc credential the user had before (if any).
               // need to get a new, ad hoc Credential

              syncV(UserAdHocCredentialMgmt,[userID, ACT_IDM_CREATE_CRED],[]);
          }
        etry[] then {}
        getSV (projectFinish, [pid2,repTemp2], pid2 == pid);

        // The project is over.
        pid := -1;
        busyResearchers := busyResearchers - 1;
        //userAdHocCredential := -1; // despite the project is over, the data repository might still be
                              // running for a while and the ad-hoc credential be misused ...
        lead := 0;
        repositoryType := 0;
        putSV (projectFinish, []);
    }
}

//Actions carried out by users when handling credentials (login, password) related to data repositories
class researcherCredentialHandling (userID, userRole, userBehaviour, pid, dataRepositoryType, userSSOCredential, userAdHocCredential) = {

 local var uid2 = 0;
 local var request = 0;
 local var cred = 0;
 local var currentTime =0;

 repeat{
   try [getSV (UserSSOCredentialMgmt, [uid2, request], userID ==uid2)] then
    {
        // creation of SSO credentails
        try[request == ACT_IDM_CREATE_CRED] then
          {
            syncV(IdMSSOCredentialMgmt,[userID,request],[cred]);
            userSSOCredential := cred;
            //trace ("CREATED SSO Credential - Researcher %v : %v", userID, userSSOCredential);

            try [p_writeDownCredential[userBehaviour] ==1 ] then
            {
             // case where a user writes down password due to their profile/attitudes
             currentTime :=DEMOS_TIME;
             putVB(WrittenDownSSOUserCredentials, [userID, pid, userSSOCredential]);
             trace ("WRITTEN DOWN SSO Credential - Researcher %v : %v", userID, userSSOCredential);
             wdSSOcredentials := wdSSOcredentials +1;
            }
            etry [] then {}

            try [p_shareCredential[userBehaviour] ==1 ] then
            {
             // case where a user shares password due to their profile/attitudes
             currentTime :=DEMOS_TIME;
             putVB(SharedSSOUserCredentials, [userID, pid, userSSOCredential]);
             trace ("SHARED SSO Credential - Researcher %v : %v", userID, userSSOCredential);
             sharedSSOcredentials := sharedSSOcredentials +1;
            }
            etry [] then {}
          }
        // renewal of SSO credentials
        etry[request == ACT_IDM_RENEW_CRED] then
          {
            syncV(IdMSSOCredentialMgmt,[userID,request],[cred]);
            userSSOCredential := cred;
            //trace ("RENEWED SSO Credential - Researcher %v : %v", userID, userSSOCredential);

           try [p_writeDownCredential[userBehaviour] ==1 ] then
             {
             // case where a user writes down password due to their profile/attitudes
             currentTime :=DEMOS_TIME;
             putVB(WrittenDownSSOUserCredentials, [userID, pid, userSSOCredential]);
             trace ("WRITTEN DOWN SSO Credential - Researcher %v : %v", userID, userSSOCredential);
             wdSSOcredentials := wdSSOcredentials +1;
             }
            etry[] then {}

            try [p_shareCredential[userBehaviour] ==1 ] then
             {
             // case where a user shares password (credential) due to their profile/attitudes
             currentTime :=DEMOS_TIME;
             putVB(SharedSSOUserCredentials, [userID, pid, userSSOCredential]);
             trace ("SHARED SSO Credential - Researcher %v : %v", userID, userSSOCredential);
             sharedSSOcredentials := sharedSSOcredentials +1;
             }
            etry [] then {}

          }
        etry[] then {}

       putSV (UserSSOCredentialMgmt, []);
     }

    etry[getSV (UserAdHocCredentialMgmt, [uid2, request], userID ==uid2)] then
    {
        // creation of ad hoc credential
        try[request == ACT_IDM_CREATE_CRED] then
          {
            syncV(IdMAdHocCredentialMgmt,[userID,request],[cred]);
            userAdHocCredential := cred;
            //trace ("CREATED Ad-Hoc Credential - Researcher %v : %v", userID, userAdHocCredential);

            try [p_writeDownAdHocCredential[userBehaviour] ==1 ] then
             {
             // case where a user writes down password (credential) due to their profile/attitudes
             currentTime :=DEMOS_TIME;
             putVB(WrittenDownAdHocUserCredentials, [userID, pid, userAdHocCredential]);
             trace ("WRITTEN DOWN Ad Hoc Credential - Researcher %v : %v", userID, userAdHocCredential);
             wdAHcredentials := wdAHcredentials + 1;
             }
            etry [] then {}
       try [p_shareAdHocCredential[userBehaviour] ==1 ] then
       {
        // case where a user shares ad hoc password (credential) due to their profile/attitudes
        currentTime :=DEMOS_TIME;
        putVB(SharedAdHocUserCredentials, [userID, pid, userAdHocCredential]);
        trace ("SHARED AdHoc Credential - Researcher %v : %v", userID, userAdHocCredential);
        sharedAHcredentials := sharedAHcredentials + 1;
       }
       etry [] then {}
      }
  }
 }
}

// Actions carried out by the user/researcher on data repositories
class researcherAction (userID, userRole, userBehaviour, pid, dataRepositoryType, userSSOCredential, userAdHocCredential) = {
  local var pid2 = 0;
  local var action = 0;
  local var docType = 0;
  local var createTime = 0;
  repeat {
    hold (interActionInterval);
    try [pid != -1] then {
      action := p_researcherAction;
      syncV (fileShare, [action, pid, dataRepositoryType], [pid2, docType, createTime]);
      try [pid2 != -1] then {
        totalActions := totalActions + 1;
        // threats associated with user action
        try [action == ACT_READ] then {
          try [testAccidentalReveal == 1] then {
            trace ("Accidental reveal, type: %v", docType);
            dataLeakedItems := dataLeakedItems + 1;
            dataLeakedValueLost := dataLeakedValueLost + dataValue[docType];
          }
          etry [] then {}
        }
        etry [action == ACT_WRITE || action == ACT_DELETE] then {
          try [userBehaviour == userTypeBehaviour_Traitor || userBehaviour == userTypeBehaviour_Disgruntled] then {
            try [testIntegrityBreach == 1] then {
              trace ("Integrity breach, type: %v", docType);
              dataIntegrityBreaches := dataIntegrityBreaches + 1;
              dataIntegrityValueLost := dataIntegrityValueLost + dataValue[docType];
            }
            etry [] then {}
          }
          etry [] then {}
        }
        etry [] then {}
      }
      etry [] then {}  // failed action - do nothing
    }
    etry [] then {}  // No actions while resting
  }
```

```
}

// Interactions with the user/researcher, initiated by other entities - mostly attempted attacks

class researcherInteraction (userID, userRole, userBehaviour, pid, dataRepositoryType, userSSOCredential, userAdHocCredential) = {
   local var id2 = 0;
   local var pid2 = 0;
   local var request = 0;
   local var docType = 0;
   local var createTime = 0;
   local var cTime1 = 0;
   repeat {
      getSV (researcherInteraction, [id2, request], id2 == userID);
      try [pid != -1] then {
         try [request == ATTACK_BRIBE] then {
            try [userBehaviour == userTypeBehaviour_Traitor] then
            {  // Bribery always succeeds on a traitor, never otherwise
              // reveal all project documents

              syncV (fileShare, [ACT_READ, pid, dataRepositoryType], [pid2, docType, cTime1]);
               createTime := -1;
               while [pid2 != -1 && createTime != cTime1] {  // depending on how we implement the repository, we may get -1 (no more documents) or repeat the first document we saw (same createTime)
                  trace ("Deliberate reveal, type: %v", docType);
                  dataLeakedItems := dataLeakedItems + 1;
                  dataLeakedValueLost := dataLeakedValueLost + dataValue[docType];
                  syncV (fileShare, [ACT_READ, pid, dataRepositoryType], [pid2, docType, createTime]);
               }
               putSV (researcherInteraction, [1]);
            }
            etry [] then {
               putSV (researcherInteraction, [0]);
            }
         }
         etry [] then {  // no other requests acknowledged
            putSV (researcherInteraction, [0]);
         }
      }
      etry [] then {  // No interactions while resting
         putSV (researcherInteraction, [0]);
      }
   }
}



// Attacks directly carried out by users
// Profiles currently considered: Traitors, Disgruntled people

class researcherAttacks (userID, userRole, userBehaviour, pid, dataRepositoryType, userSSOCredential, userAdHocCredential) =
  {
   local var uid1 = 0;
   local var SSOCredId1 = 0;
   local var adhocCredId1 = 0;
local var pid1 = 0;
local var repTemp1 = -1;
local var pid2 = 0;
local var docType  = -1;
local var createTime = -1;
local var cTime1 = -1;

  trace (" Entity carrying out User Attacks, for user: %v  behaviour: %v", userID, userBehaviour);

  repeat
   {
      hold (credentialBasedUserAttackInterval);

       // User tring to leverage any written down credential
       // Current hypotesis: we keep into account the impact of this data leakage multiple times, as the same credential
       //                   can be misused multiple times

       try[getVB(WrittenDownSSOUserCredentials, [uid1, pid1, SSOCredId1], uid1 != userID) ] then
        {
         // Need to actually find this written down SSO credentials ...
        try [(testFindingWrittenDownSSOCredential ==1)] then
         {
            trace (" User Attacks, leveraging writtendown SSO credential: %v for project: %v", SSOCredId1, pid1);
            // Information is leaked.

            // reveal all project documents
            syncV (fileShare, [ACT_READ, pid, repTemp1], [pid2, docType, cTime1]);
            createTime := -1;
            while [pid2 != -1 && createTime != cTime1]
                {  // depending on how we implement the repository, we may get -1 (no more documents) or repeat the first document we saw (same createTime)
                   trace ("Deliberate reveal - leveraging writtendown SSO credential, type: %v", docType);
                   dataLeakedItems := dataLeakedItems + 1;
                   dataLeakedValueLost := dataLeakedValueLost + dataValue[docType];
                   syncV (fileShare, [ACT_READ, pid, repTemp1], [pid2, docType, createTime]);
                }

             //wdSSOcredentials := wdSSOcredentials -1;
             exploitedSSOcredentials := exploitedSSOcredentials +1;

          }
         etry [] then {}
        putVB(WrittenDownSSOUserCredentials, [uid1, pid1, SSOCredId1]);

        }
       etry [] then {}


       try [getVB(WrittenDownAdHocUserCredentials, [uid1, pid1, adhocCredId1], uid1 != userID)] then
        {
          // Need to actually find this written down AH credentials
          try[(testFindingWrittenDownAHCredential ==1)] then
           {
```

```
     trace (" User Attacks, leveraging writtendown Ad-Hoc credential: %v for project: %v", adhocCredId1, pid1);

     // reveal all project documents
     syncV (fileShare, [ACT_READ, pid, repTemp1], [pid2, docType, cTime1]);
     createTime := -1;
     while [pid2 != -1 && createTime != cTime1]
          {  // depending on how we implement the repository, we may get -1 (no more documents) or repeat the first document we saw (same createTime)
             trace ("Deliberate reveal leveraging writtendown AD-HOC credential, type: %v", docType);
             dataLeakedItems := dataLeakedItems + 1;
             dataLeakedValueLost := dataLeakedValueLost + dataValue[docType];
             syncV (fileShare, [ACT_READ, pid, repTemp1], [pid2, docType, createTime]);
          }

     //wdAHcredentials := wdAHcredentials -1;
     exploitedAHcredentials := exploitedAHcredentials +1;
    }
   etry [] then {}

  putVB(WrittenDownAdHocUserCredentials, [uid1, pid1, adhocCredId1]);
 }
 etry [] then {}



// User trying to leverage shared credentials with the project
// Current hypotesis: we keep into account multiple impacts of this data leakage

try[getVB(SharedSSOUserCredentials, [uid1, pid1, SSOCredId1], uid1 != userID && pid == pid1) ] then
 {
   // Testing if this user might actually have received this credential
   try [testReceivedSharedSSOCredential == 1] then
    {
     trace (" User Attacks, leveraging shared SSO credential: %v for project: %v", SSOCredId1, pid1);
     // Information is leaked.

     // reveal all project documents
     syncV (fileShare, [ACT_READ, pid, repTemp1], [pid2, docType, cTime1]);
     createTime := -1;
     while [pid2 != -1 && createTime != cTime1]
          {  // depending on how we implement the repository, we may get -1 (no more documents) or repeat the first document we saw (same createTime)
             trace ("Deliberate reveal - leveraging shared SSO credential, type: %v", docType);
             dataLeakedItems := dataLeakedItems + 1;
             dataLeakedValueLost := dataLeakedValueLost + dataValue[docType];
             syncV (fileShare, [ACT_READ, pid, repTemp1], [pid2, docType, createTime]);
          }

      //sharedSSOcredentials := sharedSSOcredentials -1;
      exploitedSSOcredentials := exploitedSSOcredentials +1;
      }

    etry [] then {}
    putVB(SharedSSOUserCredentials, [uid1, pid1, SSOCredId1]);
```

```
              }
             etry [] then {}


             try [getVB(SharedAdHocUserCredentials, [uid1, pid1, adhocCredId1], uid1 != userID && pid == pid1)] then
              {

                // Testing if this user might actually have received this credential
                try [testReceivedSharedAHCredential == 1] then
                 {
                  trace (" User Attacks, leveraging shared Ad-Hoc credential: %v for project: %v", adhocCredId1, pid1);

                  // reveal all project documents
                  syncV (fileShare, [ACT_READ, pid, repTemp1], [pid2, docType, cTime1]);
                  createTime := -1;
                  while [pid2 != -1 && createTime != cTime1]
                       {  // depending on how we implement the repository, we may get -1 (no more documents) or repeat the first document we saw (same createTime)
                          trace ("Deliberate reveal leveraging shared AD-HOC credential, type: %v", docType);
                          dataLeakedItems := dataLeakedItems + 1;
                          dataLeakedValueLost := dataLeakedValueLost + dataValue[docType];
                          syncV (fileShare, [ACT_READ, pid, repTemp1], [pid2, docType, createTime]);
                       }

                  //sharedAHcredentials := sharedAHcredentials -1;
                  exploitedAHcredentials := exploitedAHcredentials +1;
                 }
                 etry [] then {}
                putVB(SharedAdHocUserCredentials, [uid1, pid1, adhocCredId1]);

              }

             etry [] then {}
         }
   }

// One class to represent the whole body of IT staff, rather than trying for one per individual

class ITStaff = {
    local var pid = 0;
    local var uid = 0;
    local var pid2 = 0;
    local var docType = 0;
    local var cTime1 = 0;
    local var createTime = 0;
    local var request = 0;
    local var lead =0;
    local var repTemp = 0;
    local var repTemp1 = 0;
    local var repTemp2 = 0;

    entity (ITAction, ITAction, 0);
    repeat {
        try [getSV (ITAction, [], true)] then {
            // For the moment, in our model, routine IT actions don't impinge on the security of the data repository, so we don't model them.
            putSV (ITAction, []);
        }
        etry [getSV (ITInteraction, [request], true)] then {
            try [request == ATTACK_BRIBE] then {
                try [testITBribable == 1] then {
                    while [getVB (projectsTrackedByIT, [pid, repTemp1], true)] {
                        // reveal all project documents
                        syncV (fileShare, [ACT_READ, pid, repTemp1], [pid2, docType, cTime1]);
                        createTime := -1;
                        while [pid2 != -1 && createTime != cTime1] {  // depending on how we implement the repository, we may get -1 (no more documents) or repeat the first document we saw (same createTime)
                            trace ("Deliberate reveal, type: %v", docType);
                            dataLeakedItems := dataLeakedItems + 1;
                            dataLeakedValueLost := dataLeakedValueLost + dataValue[docType];
                            syncV (fileShare, [ACT_READ, pid, repTemp1], [pid2, docType, createTime]);
                        }
                        putVB (projectsTrackedByIT, [pid, repTemp1]);
                    }
                    putSV (ITInteraction, [1]);
                }
                etry [] then {
                    putSV (ITInteraction, [0]);
                }
            }
            etry [] then {  // no other requests acknowledged
                putSV (ITInteraction, [0]);
            }
        }
        etry [getSV (projectStart, [pid, uid,lead, repTemp], uid == -1)] then {
            putVB (projectsTrackedByIT, [pid, repTemp]);
            putSV (projectStart, [-1]);
        }
        etry [getSV (projectFinish, [pid, repTemp2], true)] then {

             try [ repTemp2 == DATAREPOSITORY_TYPE_CIT] then
             {
               // Only Centrally IT Managed Repositories are archived
               entity (archiveProject, archiveProject (#pid), 0);
             }
             etry[] then {}

            putSV (projectFinish, []);
        }
    }
}

class ITAction = {
    repeat {
        hold (interITActionInterval);
        syncV (ITAction, [], []);
    }
}
```

```
// After a while, old projects' documents are deleted from the data repository so they don't build up indefinitely
// This happens only for data repositories managed by IT Staff, i.e. Central IT data repositories

class archiveProject (pid) = {
   local var pid2 = 0;
   local var docType = 0;
   local var createTime = 0;
   local var repTemp = 0;

   hold (archiveProjectWait);
   syncV (fileShare, [ACT_DELETE, pid, -1], [pid2, docType, createTime]);
   while [pid2 != -1] {
      syncV (fileShare, [ACT_DELETE, pid, -1], [pid2, docType, createTime]);
   }
   try [getVB (projectsTrackedByIT, [pid2, repTemp], pid2 == pid)] then {
      trace ("Project %v (type: %v) has been archived", pid, repTemp);
      numCITRepositories := numCITRepositories -1;
   }
   etry [] then {
      trace ("Error: Can't find project %v to archive", pid);
      close;
   }
}


// Central IT Management - IdM
// handling single-sign-on (SSO) user credentials & account management

class   IdMCentralITUserCredentialProvisioning = {
local var uid =0;
local var request =0;
local var SSOCredId =0;
local var createTime = 0;

 repeat{
   try [getSV (IdMSSOCredentialMgmt, [uid, request], true)] then
    {
       //Issuing a new SSO credentials to the user
       try[request == ACT_IDM_CREATE_CRED] then
        {
          // generate new credential
          SSOCredId := SSOCredId + 1;
          createTime := DEMOS_TIME;
          // keep track of it for future lifecycle management
          putVB (CentralITSSOUserCredentials, [uid,SSOCredId, createTime]);
        }
       etry[request == ACT_IDM_RENEW_CRED] then
        {
          // generate new credential
          SSOCredId := SSOCredId + 1;
          createTime := DEMOS_TIME;
          // keep track of it for future lifecycle management
          putVB (CentralITSSOUserCredentials, [uid,SSOCredId, createTime]);
```

```
         }
        etry[] then {}

       // return credential
       putSV (IdMSSOCredentialMgmt, [SSOCredId]);
     }
    //etry[] then {}

  }

}


// Central IT Management - IdM
// handling renewal of user credentials

class   IdMCentralITCredentialManagement = {
local var uid =0;
local var SSOCredId =0;
local var createTime =0;
local var currentTime = 0;
local var uid2 =0;
local var SSOCredId2 = 0;
local var pid =0;

    repeat {
           // checking for expired passwords and asking users for renewal
           currentTime  := DEMOS_TIME;
           while [getVB (CentralITSSOUserCredentials, [uid, SSOCredId, createTime], (currentTime - createTime > SSOCredExpirationTime))]
            {
              trace ("EXPIRED SSO Credential %v for user %v",SSOCredId, uid );

              // the expiration of this credential makes any written down copy worthless. Remove it.
              try[getVB(WrittenDownSSOUserCredentials, [uid2, pid, SSOCredId2], uid2 == uid && SSOCredId2 == SSOCredId)] then
               {
                 wdSSOcredentials := wdSSOcredentials -1;
               }
              etry [] then {};

              // the expiration of this credential also makes any shared copy worthless. Remove it.
              try[getVB(SharedSSOUserCredentials, [uid2, pid, SSOCredId2], uid2 == uid && SSOCredId2 == SSOCredId)] then
               {
                 sharedSSOcredentials := sharedSSOcredentials -1;
               }
              etry [] then {};


              // asking user to renew credential
              syncV(UserSSOCredentialMgmt,[uid, ACT_IDM_RENEW_CRED],[]);
            }
            hold(days);
        }
```

```
}


// AdHoc IT Management - IdM - handling ad-hoc user credentials & account management
// In the real world, each ad-hoc sharing site would have its own creedential management solution
// For semplicity we model this with an entity issuing ad-hoc credentiald and deal with account management for all ad-hoc sites

class   IdMAdHocUserCredentialProvisioning = {
local var uid =0;
local var request =0;
local var AdHocCredId =0;
local var createTime = 0;


 repeat{
    try [getSV (IdMAdHocCredentialMgmt, [uid, request], true)] then
     {
         //Issuing an ad hoc credentials to the user
         try[request == ACT_IDM_CREATE_CRED] then
           {
             // generate new credential
             AdHocCredId := AdHocCredId + 1;
             createTime := DEMOS_TIME;
             // keep track of it - but no future lifecycle management is provided by teh system
             putVB (AdHocUserCredentials, [uid,AdHocCredId, createTime]);
           }
         etry[] then
           {
             // in this model we assume there is no automatic renewal of ad hoc credentials
           }

        // return credential
        putSV (IdMAdHocCredentialMgmt, [AdHocCredId]);
     }
    //etry[] then {}

  }

}

// Entity in charge of generating external Threats, such as bribing researchers and IT staff

class generateExternalThreats = {
    // Generate a variety of threats to represent the external "threat environment"
    // doing things (based on prob distributions) to harm the project, based on data, project phase and other contextual information
    entity (bribeResearcher, bribeResearcher, 0);
    entity (bribeIT, bribeIT, 0);
    entity (securityITAttack, securityITAttack, 0);

}

class bribeResearcher = {
   local var result = 0;
   repeat {
      hold (interBribeInterval);
      syncV (researcherInteraction, [p_researcher, ATTACK_BRIBE], [result]);
      try [result == 1] then {
         trace ("Successful bribery of researcher");
         successfulBribes := successfulBribes + 1;
      }
      etry [] then {}
   }
}

class bribeIT = {
   local var result = 0;
   repeat {
      hold (interITBribeInterval);
      syncV (ITInteraction, [ATTACK_BRIBE], [result]);
      try [result == 1] then {
         trace ("Successful bribery of IT");
         successfulBribes := successfulBribes + 1;
      }
      etry [] then {}
   }
}


// Class modelling IT attacks (e.g. by external hackers) on collaborative data sharing sites
// Impact will depend on the class (CIT, SIT) of the underlying data repository

class securityITAttack =
{
 local var numCITsites = 0;
 local var numSITsites = 0;
 local var numSites = 0;
 local var s = 0;
 local var s1 = 0;
 local var pid = 0;
 local var repTemp = 0;
 local var pid2 = 0;
 local var docType = 0;
 local var cTime1 = 0;
 local var createTime = 0;


    repeat {
      hold (interITAttackInterval);
      numCITsites := AV_projectsTrackedByIT;
      numSITsites := AV_projectsAdHoc;

      numSites := numCITsites + numSITsites;

      // finding a data repository site that is in place
      try [numSites != 0] then
```

```
{
  //targeting CIT sites
  try[binom(1, (numCITsites/numSites)) == 1] then
    {
      trace ("CIT site: EXTERNAL ATTACK ATTEMPT");
      // randomly choose a CIT site
      s := numCITsites;
      s := s - 1;
      s1 := puni (0, s);
      do s1 {
             getVB (projectsTrackedByIT, [pid, repTemp], true);
             putVB (projectsTrackedByIT, [pid, repTemp]);
           }

      getVB (projectsTrackedByIT, [pid, repTemp], true);
      putVB (projectsTrackedByIT, [pid, repTemp]);

      // check if the attach is successful
      try [p_secITattackCIT == 1] then
      {
        trace ("CIT site %v: SUCCESSFUL EXTERNAL ATTACK", pid);
        successfulCITexternalAttacks := successfulCITexternalAttacks +1;

        // reveal all project documents
        syncV (fileShare, [ACT_READ, pid, repTemp], [pid2, docType, cTime1]);
        createTime := -1;
        while [pid2 != -1 && createTime != cTime1]
            {   trace ("Deliberate reveal leveraging attacked CIT site, type: %v", docType);
               dataLeakedItems := dataLeakedItems + 1;
               dataLeakedValueLost := dataLeakedValueLost + dataValue[docType];
               syncV (fileShare, [ACT_READ, pid, repTemp], [pid2, docType, createTime]);
            }
       }
      etry[] then {}

    }
  // targeting SIT sites (if any)
  etry[numSITsites>0] then
   {
      trace ("SIT site: EXTERNAL ATTACK ATTEMPT");
       // randomly choose a SIT site
      s := numSITsites;
      s := s - 1;
      s1 := puni (0, s);
      do s1 {
             getVB (projectsAdHoc, [pid, repTemp], true);
             putVB (projectsAdHoc, [pid, repTemp]);
           }

      getVB (projectsAdHoc, [pid, repTemp], true);
      putVB (projectsAdHoc, [pid, repTemp]);

      // check if the attach is successful
```

```
         try [p_secITattackSIT == 1] then
         {
          trace ("SIT site %v: SUCCESSFUL EXTERNAL ATTACK", pid);
          successfulSITexternalAttacks := successfulSITexternalAttacks +1;
          // reveal all project documents
          syncV (fileShare, [ACT_READ, pid, repTemp], [pid2, docType, cTime1]);
          createTime := -1;
          while [pid2 != -1 && createTime != cTime1]
             {   trace ("Deliberate reveal leveraging attacked SIT site, type: %v", docType);
               dataLeakedItems := dataLeakedItems + 1;
               dataLeakedValueLost := dataLeakedValueLost + dataValue[docType];
               syncV (fileShare, [ACT_READ, pid, repTemp], [pid2, docType, createTime]);
             }
          }
         etry[] then {}


       }
     }
    etry [] then {}

  }

}

// initialising simulation
entity(initialise, initialise,0);


// holding for the entire simulation timeframe
req [done ==1];

close;
```