Towards Open Data-Driven Evaluation of Access Control Policies

Benjamin Aziz

School of Computing
University of Portsmouth
Portsmouth, UK

Abstract

Modern approaches towards the understanding of the behaviour of systems and policies have recently been driven by the abundance of open and non-open data moving away from the classical model-based approaches, in which data were secondary to the solution. In this paper, we present a similar approach by suggesting that the analysis of the risk probability for access control and security policies can be based on an empirical data-driven study. We outline a constraint-based approach that allows organisations to examine policies in light of the probabilities of internal actors damaging organisational assets. Our approach is validated using Verizon's open community dataset for security incidents, known as VERIS/VCDB.

Keywords: Access Control, Data-driven Risk, Security Policies, Security Constraints, Security Datasets

1. Introduction

The use of the term *data-driven* has nowadays become synonymous with the application of various analysis techniques to large (open) datasets in order to understand better the behaviour and properties of computing and IT systems. This approach, with data at its heart, is slowly but surely overtaking classical pure model-driven approaches, where data are oftentimes considered secondary or even irrelevant to the problem at hand. Instead, a data-driven approach makes use of the information and knowledge embedded within the data to validate our understanding of the behaviour and characteristics of the systems we develop and utilise. For example, in recent years, Google has gradually replaced the heavyweight machine learning algorithms underlying its automatic language translation service, Google Translate, with purely statistical-based algorithms [1] applied to its large corpus of data.

In the computer and security research community, utilising data as a source of information and knowledge has been accomplished to varying degrees of success. There are still many factors that obstruct and hinder the adoption of open

data in this field. Some such factors are political, for example, the continuing diffidence on the part of commercial organisations to share data due to the fear that such data may eventually reveal sensitive information. Others are more technical, related to the consistency, quality and the lack of consensus on the nature of variables that should be monitored or indeed the metrics that should be used to quantify the data themselves [2]. There are also philosophical questions related to whether past data are in any way relevant to future events [3]. Nonetheless, in recent years, this trend has started to shift with the arrival of large open datasets backed by the reliability of reputable organisations (e.g. VCDB [4] and CERT's Vulnerability Notes Database at Carnegie Mellon University [5], SecRepo [6], CAIDA [7], LANL [8]), and we have started to notice an increasing use of such datasets.

The main goal motivating our work here is to demonstrate how open data can be used to evaluate and reason about security policies based on risk probabilities. We consider one example approach for evaluating the risk probability associated with access control policies defined in XACML [9] based on information recorded in the community-based open dataset, VERIS/VCDB [10, 4]. Thus, we combine both the data-driven approach in policy configuration with standards in defining access control policies. In doing so, we agree with the view by [2] that past data are still relevant to new security incidents and that despite the fact that *the road ahead may bend with human whim and technological advance, ... it does not appear to bend too sharply too often.* Our view therefore is that such past data can still be used to gain some approximate knowledge of how systems should be controlled and defended through the understanding of the level of risk probability that their access control policies imply, before such policies are deployed. Our focus on a standard such as XACML ensures that our approach has wide and general applicability and is not restricted to proprietary languages.

We adopt a frequency-based definition of risk probability [11]; that is, given n_{total} number of events, then the risk probability, P_{risk}, that a specific resource type is damaged by a specific type of user, is approximated by the relative frequency:

$$P_{risk} \approx \frac{n_{specific}}{n_{total}}$$

with the usual assumption that:

$$P_{risk} = \lim_{n_{total} \to \infty} \frac{n_{specific}}{n_{total}}$$

and where $n_{specific}$ is the number of times the resource has been recorded as damaged by some user type. In the most general form of this definition, n_{total} should represent the total number of times the damaged resource was accessed. However, due to the open nature of the VCDB dataset, which restricts the release of this kind of critical information (as well as other information, e.g. the correlation among the different events), we adopt a more limited and VCDB-constrained definition of n_{total}, which represents only the total number of resources damaged in Cyber security incidents.

We argue that the vast majority of existing works in literature follow instead a model-based approach paying little attention to such growing body of large security-related information available in the open domain. This is despite the fact that analysis of security datasets has been a central activity in some many areas of computer security-related research (e.g. [12, 13]) as well as other more general security areas (e.g. [14, 15]). Therefore, the main contribution of this work is to demonstrate a simple but general approach in which data-based evidence publicly available can be used to drive the decision making process when it comes to the evaluation of access control decisions. However, the current paper is aimed only at sketching up the approach to address an interesting and much needed idea, and demonstrate it with some simple examples. It is by no means an exhaustive study and so future work will extend this work both in terms of its techniques and metrics as well as its applicability to more complicated real world case studies that will validate better the proposed approach.

The rest of the paper is structured as follows. In Section 2 we give an overview of the current state-of-the-art literature in risk-based access control and compare the various open Cyber security datasets. In Section 3, we give some background on the main concepts included in this work, including VERIS and XACML. In Section 4 we summarise our approach. In Section 5, we show how we transform concepts from the XACML language to entities in the VERIS schema. This will be particularly focused on transforming subjects and resources. In Section 6, we give a VCDB-based data-driven definition of the risk probabilities of every type of actors damaging every type of asset. In Section 7, we define a set of risk probability constraints that we use as examples to demonstrate the applicability of our approach. We give an example in Section 8 and discuss in Section 9 our findings and contributions. Finally, we conclude the paper in Section 10 giving directions for future work.

2. Related Work

2.1. Risk-based Access Control

The idea of defining, analysing and enforcing security policies based on a model of risk has been widely researched in the security community and there is a plethora of works in literature that combine the two concepts of risk and access control under the same framework and at various levels of abstraction. We group this literature into two broad areas: risk-based decision making (e.g. [16, 17, 18, 19, 20]) and risk-based policy configuration (e.g. [21, 22, 23, 24]).

Risk-based decision making is concerned with the use of risk as a guiding factor when evaluating access control decisions. Dimmock et al. [16] proposed a model that incorporated risk analysis into the decision making process when granting access to resources in a distributed system. Their model is implemented based on the Open Architecture for Secure Interworking Services (OASIS), which included a language for expressing Role-Based Access Control (RBAC) policies [25]. The model also defines predicates for thresholding risk, which allow a decision to permit access to be deemed either risky or not. Another model that extends RBAC policy evaluation with risk is proposed by [17].

Their main contribution is based on computing risk at runtime when a policy is evaluated. The main difference from our proposed approach here is that their risk computation is not empirical, rather based on a model-driven approach.

Cheng et al. [18] proposed the extension of Multi Level Security (MLS) based access control policies with risk, in what they termed *fuzzy MLS* policies. Their model incorporates the probability of human misbehaviour as well as inadvertent information disclosures. Their approach follows a quantified risk-adaptive access control approach, where the band between deny and permit decisions is deemed to be a fuzzy area where access may be granted subject to mitigation controls.

Molloy et al. [20] propose a new model of decision making (e.g. the XACML PDP), which includes a risk assessment element. Unlike traditional decision making processes, which render a binary decision, this new model also include a decision defer outcome. The allow, deny or defer outcomes are coupled with a set of metadata that represent gains, costs and damages associated with a particular outcome. Again, such a model suffers from the complexity associated with determining such multi-dimensional risk values from real world data, and as such, would be impractical in reality.

XACML [26] itself has been considered as a medium for incorporating quantitative definitions of risk. In [19], the authors extend the XACML standard with primitives for expressing qualified risk adaptive access control. The new language, termed RXACML, considers both the risk of allowing access and the risk of denying access. As we outline in this paper, it is relatively straightforward to quantify the risk of granting access based on security incidents information, however, it is far less obvious calculating the risk of denying access due to the lack of any concrete data.

Risk-based policy configuration, on the other hand, is concerned with the use of risk as a factor when designing and configuring access control policies. Aziz et al. [21] consider the problem of reducing risk in RBAC policies as a reconfiguration problem. They define a risk-aware semantic model for such policies, and then introduce a reconfiguration analysis that reconfigures a *risky* policy into a less risky one. More recently, Bijon et al. proposed a framework in [22, 23] for defining risk-aware RBAC policies. The framework introduces a set of algorithms that can be used to define new RBAC policies (relations) taking into consideration risk, both as a quantitative and qualitative concept. In [24], Krautsevich et al. propose the extension of the Usage Control [27] model of security policies with quantitative risk methods. Their model of risk is integrated within an extension of XACML [26], termed U-XACML [28], for expressing usage control policies.

Aside from the decision-making and policy configuration problems, other works have used risk to improve the understanding of the quality of infrastructures and processes involved in the management of data and information. For example, Schläger and Nowey [29] propose the use of annual loss expectancy as a measure of risk when assessing and managing authentication and authorisation infrastructures. Their model is geared towards the analysis of infrastructures rather than access control policies, as is our aim here. In [30], another model,

called benefit and risk access control, is proposed, which combines the risk of information disclose with the benefit of information sharing, when evaluating transactions that involve the movement of data. In [31], the concept of risk is combined with that of trust in proposing a new access control mechanism that takes the two concepts into consideration when granting access to Grid-based resources, and in [32], risk is used to measure the consequences of non-conformity between a real execution of a business process and its specified description. Finally, some other works, such as [33], treat the notion of risk as a *quality*, therefore they are less relevant to our perception of risk as a *quantity*.

In all the above example works, and many others, risk is defined as an abstract quantity or a mathematical value that can be calculated from specific operations. This is even the case in more recent works such as [34], who proposed a risk-aware framework for enforcing access control policies in Grid systems and [35] who did the same for Cloud systems. None such works suggest a *data-driven empirical* approach to the problem of evaluating and configuring security policies based on concrete real world datasets, which is where our work here provides its major contribution. Most of the current literature adopts a *model-driven* approach to the definition of risk in the context of security policies.

Perhaps the closest work to ours is that proposed in [36], where the problem of missing attribute values in evaluating attribute-based access control policies, such as those written in XACML, is seen as a non-deterministic retrieval process that can be formulated in terms of a model of attribute probabilities. This can then be used to analyse the correctness probability of a policy decision evaluation. The main difference from our proposed method is that in our case, we do not doubt the correctness of an attribute value (for example, as the value may be missing or dubious), rather we estimate how dangerous the policy can be given the combination of particular attribute values (in our case, the combination of the subject and the resource it is allowed to access) based on empirical evidence existing in open data.

2.2. Comparison of Open Cyber Security Datasets

As we stated in the Introduction, recent years have noticed the opening of several datasets related to Cyber security, attack incidents and other relevant information in the domain to the public. Of these, the most notable ones are SecRepo [6], VCDB [4], CERT's Vulnerability Notes Database at Carnegie Mellon University [5], CAIDA [7] and the open datasets from the Los Alamos National Laboratory (LANL) [8]. Due to its suitability, our choice in this work has been VCDB, which we discuss in the next section. However, we provide here an overview and discussion of the characteristics of the other datasets.

2.2.1. SecRepo

"SecRepo.com - Samples of Security Related Data" [6] is a repository and directory of various datasets related to the security domain maintain by Mike Sconzo. The datasets are categorised by their relevance into categories such as network, system, malware, file, password, threat feeds and others. Despite being

a rich repository, the datasets referred to do not relate to incidents specifying types of users and resources, and are therefore outside the scope of our work in this paper. However, we consider this repository to be of value for any future studies that would be focused on malware, system and network security.

2.2.2. CERT's Vulnerability Notes Database

The CERT Vulnerability Notes Database [5] provides a list of software vulnerabilities discovered including summaries of the vulnerabilities, their technical detail, mitigation information and lists of affected vendors. The database is more geared towards describing vulnerabilities rather than actual incidents, and does not relate directly to types of resources nor does it mention what roles of attackers may be able to benefit from the vulnerabilities. Thus, the information included in the database does not map in any intuitive manner to XACML policy concepts. The database also includes metrics based on the Common Vulnerability Scoring System (CVSS) standard [37]. The CVSS metrics providing a score for the severity of vulnerabilities, and can be used in the future for any new impact-related study.

2.2.3. CAIDA

The Center for Applied Internet Data Analysis (CAIDA) set [7] is a collection of datasets, monitors and reports not all directly related to the security domain (some are related to network traffic and topology information, for example), but from which security information can be obtained nonetheless. In general, all the referenced datasets are more relevant to network security than to access control-based security, and therefore, we do not consider these to be of relevance to the approach proposed in this paper. Like SecRepo, in order to explore further the structure of these data, it is necessary to study the individual datasets and reports included in the collection.

2.2.4. LANL Open Datasets

The Los Alamos National Laboratory (LANL) open datasets [8] consist currently of two datasets. The first represents "Comprehensive, Multi-Source Cyber-Security Events" that include authentication events, process start and stop events, network flow events, DNS resolution events and events taken from the authentication data that present known compromise events. Of these, the last is probably most related to our work. This dataset has the form "time, user@domain, source computer, destination computer", which represents a compromise event at some given time. Since this dataset refers directly to anonymised user identities and anonymised resource identities without referencing their types, it would not be possible to directly map from subjects and resources in an XACML policy to elements of this dataset or vice versa. Therefore, we do not consider the dataset here. The second dataset represents user-computer authentication associations in time, which is of little relevance to our policy configuration problem particularly again that neither the type of the computer (resource) nor the role of the user are referenced in the dataset.

3. Background

3.1. VERIS: A Schema for Describing Cyber Security Incidents

The Vocabulary for Event Recording and Incident Sharing (VERIS) [10] is a dataset and schema defining a set of metadata and metrics for describing Cyber security incidents. It is currently considered a leading provider of open quality information in the IT security domain and provides a framework that organisations can use to collect and share information on security incidents in a responsible and anonymous manner, with the aim of constructing a ground on which researchers and experts in the IT security industry can cooperate to learn from their knowledge and experiences.

The VERIS schema itself consists of five general categories, containing descriptions of the security incidents in the VERIS dataset. These five categories can be summarised as follows:

- *Incident Tracking*: this category contains general information about the incidents, for example, the source identity, summary of the incident and whether the incident is related to other incidents.
- *Victim demographies*: this category contains information related to the organisation being affected by the incident, for example, its country of operation, number of employees, revenue and industry type.
- *Incident description*: this category contains information related to the question of "who did what to what (or whom) with what result". It is based on the A4 threat model developed by Verizon and contains descriptions related to the Actors, Assets, Actions and Attributes (A4) of an incident. Here, we focus only on the Actors and Assets metadata, as these can directly relate to concepts in XACML.
- *Discovery and response*: this category contains information related to the incident's timeline, its discovery method, root causes, corrective actions and so on.
- *Impact assessment*: this last category contains information on loss categorisation and estimation and impact rating. As we highlight later, information in this category is currently poor, therefore we do not consider it here.

The significance of the VERIS dataset lies in the fact that it is a *community-based* dataset. This means that its data are collected from a wide range of industries and varied over different types and sizes of organisations, therefore providing a rich ground for organisations to learn about the various risks and threats that could exist on a global level.

On the other hand, the VERIS dataset, known as VCDB [4], has currently over 6800 recorded incidents (latest update as of November 2016), with its schema metadata ranging over 2500 elements. In these, there were 8146 assets reported to have been affected. Of these, only 585 (approx. 8.5%) were reported

to be of type Unknown, therefore we consider this data to be of reasonably good quality. On the other hand, there were 3402 internal actors reported to have been involved in these incidents

The quality of the data on the impact of incidents as reported under the overall rating metadata is poor in the current version of VCDB. Out of the 6860, over 80% of incidents have an unknown impact and only in one incident (i.e. in 0.015% of cases) was such impact reported in detail. The rest of the incidents have unreported rating for their impact.

One of the drawbacks of the dataset in VCDB is that nothing is reported on the correlation between compromised assets, and whether attacks have complex vectors. Incidents are reported as independent of one another; i.e. there is no information suggesting that several incidents are part of the same attack. This is likely to be due to anonymity reasons, but it does prevent more complex probabilistic analyses to be applied, for example the likelihood that a specific type of actor can launch a complex attack over multiple resources.

Additionally, VERIS lacks information related to whether access control breaches were part of the Cyber security incidents. This means that inevitably, there would be cases where the incidents are not related to the problem of access control, and hence, our adopted analysis later is an approximation of the real number of cases. Nonetheless, we argue that this approximation (e.g. of the probability that a specific type of actors would damage a specific type of resources) is a safe one.

Generally, we chose the VERIS/VCDB dataset over other datasets (e.g. [5, 6, 7]) due to the straightforward nature of the mapping from XACML elements (subjects and resources) to VERIS schema elements (actors and asset varieties). Our approach is general enough to be able to incorporate probabilities created based on the analysis of any other dataset, which would be deemed suitable in terms of its content. Such probabilities would then be combined with the existing results (see Appendix A) to achieve more refined results. This could be carried out as an element of future research stemming from the current work.

3.2. XACML: An Industrial Standard for Access Control Policies

XACML stands for the eXtensible Access Control Markup Language, which is in its core a special XML schema for expressing access control policies and rules. XACML is maintained by OASIS (Organization for the Advancement of Structured Information Standards) and is currently in its 3rd version [9]. In addition to the schema, XACML also defines a reference architecture based on the concept of a policy enforcement point that implements and enforces XACML policies. The XACML policy language model is shown in Figure 1.

In the context of this paper, our main interest in XACML is related to its definition of *subjects* and *resources* as parts of a security policy's target. A subject is defined as an actor whose attributes can be referenced by predicates and who represents the entity requesting access to some resource. On the other hand, a resource represents any data, service or system component to which access is requested.

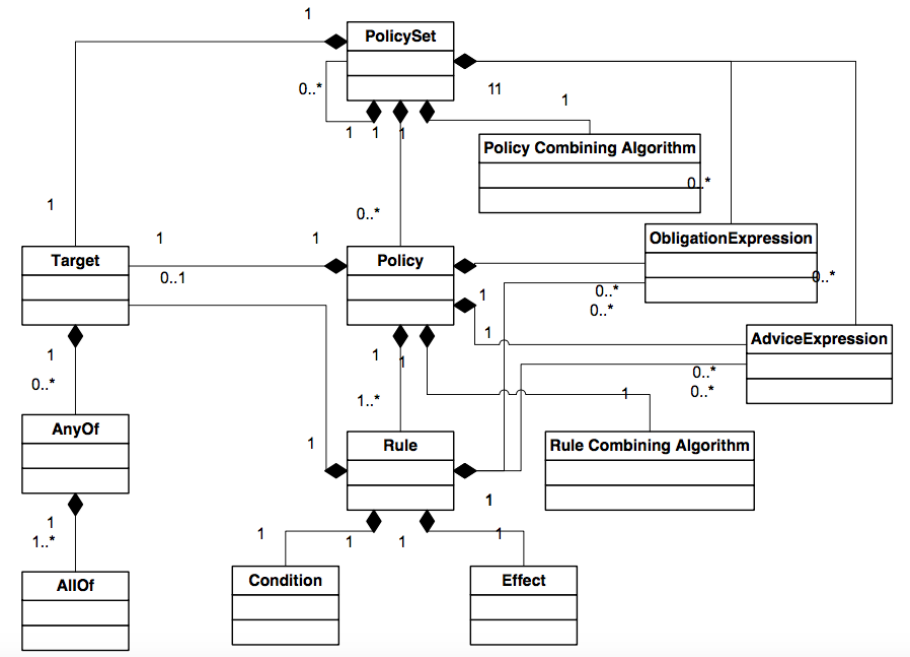

Figure 1: The XACML Policy Language Model [9]

We keep an open view of the specific attributes that identify a subject or a resource. For example, a subject could be identified through attributes related to its identity, IP address or DNS name:
`urn:oasis:names:tc:xacml:1.0:subject:subject-id`
`urn:oasis:names:tc:xacml:3.0:subject:authn-locality:ip-address`
or `urn:oasis:names:tc:xacml:1.0:subject:authn-locality:dns-name`

Similarly, for resources, they may be identified through attributes related to the identity or file name (in the case the resource is a considered a file):

`urn:oasis:names:tc:xacml:1.0:resource:resource-id`
`oasis:names:tc:xacml:1.0:resource:simple-file-name`

Our analysis of risk probabilities is not limited by the source of such attributes, only by how these attributes are mapped to VERIS concepts as we discuss in Section 5.

4. Our Approach

Our approach can be summarised in Figure 2. The dark-coloured boxes represent the various outcomes from this approach and the arrows represent the steps that need to be taken to generate those outcomes.

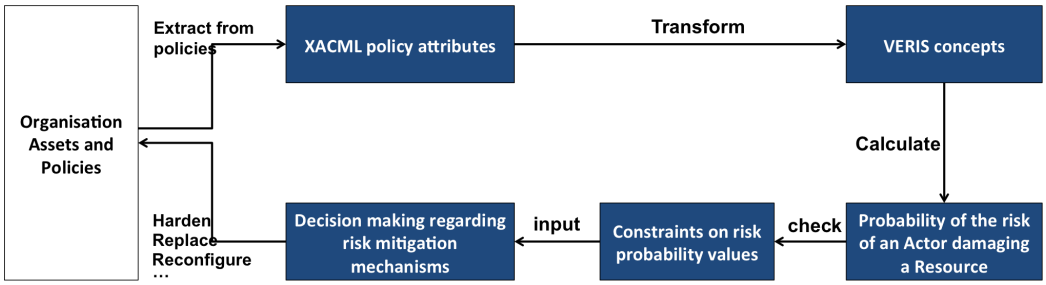

Figure 2: Our Proposed Approach

Our starting point are the assets owned by an organisation and the security policies protecting those assets set up by the administrators. The first step involves identifying and extracting policy attributes that refer to internal subjects within the organisation and the resources they can access. These attributes are then transformed to VERIS schema concepts, which in our case, will be the Actors and Assets metadata. Our focus in this work is on internal actors, as an example, however, the analysis can be expanded to also the external and partner actors. Once we are in the VERIS model, we are able to calculate the frequency-based probability approximate of a type of actor damaging a type of resource, based on the data recorded in VCDB. This gives us the probability of the risks involved by having a badly configured policy. The next step then is to check how bad such configuration is by validating whether any number of constraints on the risk probabilities calculated in the previous step are violated for the given policy. This information is then used as input to a decision-making process that aims to mitigate against the effects of such risks in the policy. Such mitigation mechanisms may include, for example, the security hardening of the assets, the reconfiguration of the system or the policy protecting the assets or the replacement of assets or policies altogether with more security ones. This is then fed back into the original organisation set up, and the approach is repeated until a fixed point is reached.

In the following sections, we focus on the transformation of concepts, the definition of risk probabilities and give examples of constraints on policies.

5. From XACML to VERIS Model Transformation

Our first step here is to transform elements in the XACML model to elements in the VERIS schema. In the following sections, we focus on transforming subject attributes to actors and resources to asset varieties.

5.1. Transforming Subjects to Actors

We consider in this paper only internal actors; i.e. employees or people in roles internal to some organisation. In future works, we shall also consider the

risk analysis of external actors and partners of organisations; the other two broad types of actors that the VERIS schema considers.

We define the set of all VERIS internal actors as $Actors = \{$*Auditor, Call Center, Cashier, End-user, Executive, Finance, Helpdesk, Human Resources, Maintenance, Manager, Guard, Developer, System administrator, Unknown, Other*$\}$. We shall use this set as the range for our transformation function.

On the other hand, we assume that XACML subjects are captured by the values of attributes such as `subject-id`, `ip-address` or `dns-name`, although these can be extended to include the values of any other attributes, as long as the attribute defines the identity of a subject. We refer to the set of all XACML subjects as *Subjects*, which represents the values that any subject attribute can assume. We shall write $Subjects_{policy} \subseteq Subjects$ to refer to the set of subjects that appear in a specific XACML policy or policy set. Similarly, we define $Actors_{policy} \subseteq Actors$ as the subset of actors that corresponds to the subjects referenced in a specific XACML policy.

We define the following transformation function to transform XACML subjects to VERIS actors:

$$\mathcal{T}_{su} : Subjects \rightarrow Actors$$

Which permits us to define more specifically $Actors_{policy}$ as the set $\{a : \ a \in Actors \ \wedge \ (\exists s \in Subjects_{policy}.\ a = \mathcal{T}_{su}(s))\}$. In reality, the definition of $\mathcal{T}_{su}$ will be specific to each organisation. It's instantiation will be carried out according to some internal role-based access control policy, which will be used to align the roles of the policy with VERIS actor values above. For example, if we assume that the organisation's policy considers a subject named *James Brown* to be in the role of a manager, then we will have that $\mathcal{T}_{su}(James\ Brown) = Manager$.

5.2. Transforming Resources to Asset Varieties

Next, we need to give a type for XACML resources in terms of VERIS concepts. This is done by mapping XACML resources to what is known as an asset in the VERIS schema. The VERIS terminology provides a type for each asset affected in an incident called the asset's *variety*.

The set of all VERIS asset varieties is defined as the set $AssetVariety =$ $\{$*Server - Authentication*, *Server - Backup*, *Server - Database*, *Server - DHCP*, *Server - Directory (LDAP, AD)*, *Server - Distributed Control System (DCS)*, *Server - DNS*, *Server - File*, *Server - Log or Event Management*, *Server - Mail*, *Server - Mainframe*, *Server - Payment Switch or Gateway*, *Server - POS Controller*, *Server - Print*, *Server - Proxy*, *Server - Remote Access*, *Server - SCADA System*, *Server - Web Application*, *Server - Code Repository*, *Server - Virtual Host*, *Server - Other*, *Server - Unknown*, *Network - Access Control Reader*, *Network - Camera or Surveillance System*, *Network - Firewall*, *Network - Hardware Security Module (HSM)*, *Network - IDS or IPs*, *Network - Mobile Broadband Network*,

Network - Private Branch Exchange (PBX), *Network - Private WAN*,
Network - Programmable Logic Controller (PLC), *Network - Public WAN*,
Network - Remote Terminal Unit (RTU), *Network - Router or Switch*,
Network - Storage Area Network (SAN), *Network - Telephone*,
Network - VoIP Adapter, *Network - Wired LAN*, *Network - Wireless LAN*,
Network - Other, *Network - Unknown*,
User Device - Authentication Token or Device,
User Device - Media Player or Recorder,
User Device - Desktop or Workstation, *User Device - Laptop*,
User Device - Mobile Phone or Smartphone, *User Device - Peripheral*,
User Device - POS terminal, *User Device - Tablet*, *User Device - Telephone*,
User Device - VoIP phone, *User Device - Other*, *User Device - Unknown*,
Public Terminal - Automated Teller Machine (ATM),
Public Terminal - Detached PIN Pad or Card Reader,
Public Terminal - Gas "Pay-at-the-Pump" Terminal,
Public Terminal - Self-Service Kiosk, *Public Terminal - Other*,
Public Terminal - Unknown,
Media - Backup Tapes, *Media - Disk Media*, *Media - Documents*,
Media - Flash Drive or Card, *Media - Hard Disk Drive*, *Media - Smart Card*,
Media - Payment Card, *Media - Other*, *Media - Unknown*,
People - System Administrator, *People - Auditor*, *People - Call Center*,
People - Cashier, *People - Customer*, *People - Developer*, *People - End-user*,
People - Executive, *People - Finance*, *People - Former Employee*,
People - Guard, *People - Helpdesk*, *People - HR*, *People - Maintenance*,
People - Manager, *People - Manager*, *People - Other*, *People - Unknown*,
Unknown}

On the other hand, we assume that resources are identified in an XACML policy by the values of resource-related attributes in XACML such as `resource-id` or `simple-file-name` and any other relevant attributes. We call the set of all resources *Resources* and the resources referenced in a specific XACML policy $Resources_{policy} \subseteq Resources$. We define the transformation function on XACML resources as follows:

$$\mathcal{T}_{re} : Resources \rightarrow AssetVariety$$

which maps a resource in XACML to one of the above asset varieties. This function allows us to define more precisely the set of asset varieties corresponding to some XACML policy as the set $AssetVariety_{policy} = \{v : \ v \in AssetVariety \ \wedge \ (\exists r \in Resources_{policy}.\ v = \mathcal{T}_{re}(r))\}$. The instantiation of $\mathcal{T}_{re}$ can be done at an organisational level with regards to some internal policy, which will identity the asset variety type of each resource in the organisation. For example, it could be considered that files are a type of documents, hence $\mathcal{T}_{re}(file) = Media$ - *Documents*.

6. A Data-driven Definition of Risk Probabilities

We define risk based on the standard formula that describes it as the product of probability by impact:

$$risk = probability \times impact$$

This definition of quantitative risk was first articulated in a formal manner within the domain of computing systems by IBM's Robert Courtney, Jr. [38]. This formula for risk was adopted earlier by the National Bureau of Standards in its Federal Information Processing Standard (FIPS) publication number 65, *Guideline for Automatic Data Processing Risk Analysis* [39], which quantified risk using a metric called *Annual Loss Expectancy (ALE)*.

Whilst calculating the probability of an event occurring in the future is by itself a predictive exercise and as such much work has been done in literature on predicting the probabilities of new security incidents (e.g. [40, 41]), the prediction of the impact of new incidents from the impact of older ones, particularly those that would have occurred in a different organisation, is futile, since the business value of the affected assets and the indirect cost of incidents is a local model that cannot be ported across organisations. Even notable works like [42], have stopped at the analysis of the impact of security incidents on the market without attempting to predict future impact. As a result, there has been a severe lack of insight in literature into how impact of new security incidents can be derived from the impact of past ones. Therefore, at this stage, we focus only on the definition of risk probability and defer the study of impact to future works.

Therefore, we give here a VCDB-based definition of the probability that expresses how dangerous a specific type of internal actors may be in relation to accessing a particular type of assets. We define a *risk probability* function, r, to express this probability:

$$r : Actors \times AssetVariety \to \mathbb{R}$$

This function defines the risk probability for each type of VERIS actors in relation to a particular type (variety) of assets, which the actor may affect either maliciously or unmaliciously. Its definition is empirical in nature; it is based on the data provided in the VCDB dataset, and therefore it provides a true real world reflection of the probability that some actor will misuse an asset. Its concept is general; one can easily exchange the data in VCDB with data from another source, as long as the information provided is compatible with the signature of r.

For the case of VCDB, we calculate this probability for each asset variety s as the ratio between the number of assets $N_{a,s}$ affected by a particular actor a over the total number of all assets affected in all incidents N:

$$r(a, s) : \frac{N_{a,s}}{N} \times 100\% \tag{1}$$

Naturally, when an asset s is not affected by the actions of an actor a, then $N_{a,s} = 0$ and consequently, this gives us an empirical evidence that $r(a, s) = 0\%$ in relation to the data provided by VCDB.

The full VCDB-based definition of r is provided in detail in Appendix A, where each Actor's probability of damaging an Asset Variety is calculated according to (1) above.

7. Risk Probability Constraints

After defining the VCDB-based r function, the next step would be to set up any number and type of risk probability constraints that an organisation may wish to check on its internal access control policies expressed as XACML policies. In terms of their nature, constraints are logical formulæ that evaluate to a Boolean value. We give here a few examples of such constraints.

7.1. Actor Risk Probability Constraint

The first example of constraints we define here is a constraint on the maximum risk probability k that an organisation is willing to tolerate (i.e. its appetite) regarding any type of actors, in the context of some *policy*:

$$\mathcal{C}_{actor}(k, policy) = \forall a \in Actors_{policy}, s \in AssetVariety_{policy} :$$
$$rule\_permitting\_access(a, s, policy) \Rightarrow r(a, s) \leq k$$

Where $rule\_permitting\_access(a, s, policy)$ is a predicate that returns True whenever there exists a rule in the policy permitting access to s by a. The constraint states that the probability level for any actor with respect to any asset that it can access, as evaluated by the r function, must not exceed k, where k is selected based on some internal decision-making process.

A more refined version of this constraint would specify for each different actor, a, their own appetite value k_a, as follows:

$$\mathcal{C}'_{actor}(k_a, a, policy) = \forall s \in AssetVariety_{policy} :$$
$$rule\_permitting\_access(a, s, policy) \Rightarrow r(a, s) \leq k_a$$

Again the selection of k_a must be based on some decision-making process internal to the organisation itself.

This type of constraints represents the appetite an organisation is willing to have with respect to the risk posed by a specific type of actors, either in a uniform or a nonuniform manner. In a practical setup, it would constitute part of a profile of a specific role in an access control policy, which would be consulted whenever new users are added to that role. It is easy to envisage how such risk profile could be used to enhance role assignment [43] in an XACML policy (set).

7.2. Asset Risk Probability Constraint

The second example of risk probability constraints is on the maximum risk probability, k_s, tolerated for some individual asset, s, over all the types of actors that can access s in some given policy:

$$\mathcal{C}_{asset}(k_s, s, policy) = \forall a \in Actors_{policy} : \\ rule\_permitting\_access(a, s, policy) \Rightarrow r(a, s) \leq k_s$$

It is also possible to define a different variant of this constraint that aggregates the probabilities of all the actors allowed to access the asset s:

$$\mathcal{C}'_{asset}(k_s, s, policy) = \\ (\sum_{i=1}^{|MSet|} r(a_i, s)) \leq k_s,\ where\ a_i \in MSet \\ and\ MSet = \{\!|a :\ a \in Actors_{policy} \wedge rule\_permitting\_access(a, s, policy)|\!\}$$

MSet is a multiset representing all the actors in the policy (preserving their multiplicity) where there is an access-permitting rule that allows them access to the resource s.

Similar to the previous pair of constraints on actors, these constraints can be used in practice as part of the conditions of deployment of specific types of resources in an organisation. For example, if the resource is provided as a service to external users, those users may request as part of the service level agreement that the resource have a maximum level of risk probability associated with it with respect to the actors that have access to it.

7.3. Overall Policy Risk Probability Constraint

The last example of such constraints represents an aggregation of individual policy risk probabilities for each of the assets mentioned in the policy:

$$\mathcal{C}_{policy}(k, policy) = (\sum_{j=1}^{|AssetVariety_{policy}|} \sum_{i=1}^{|MSet|} r(a_i, s_j)) \leq k$$

$$where\ a_i \in MSet,\ s_j \in AssetVariety_{policy} \\ and\ MSet = \{\!|a :\ a \in Actors_{policy} \wedge rule\_permitting\_access(a, s_j, policy)|\!\}$$

This constraint, which is a general form of $\mathcal{C}'_{asset}$, covers all the actor/asset combinations in a policy. Naturally, a_i and s_j are obtained from the model transformations of Section 5, as the actual entities in an XACML policy are subjects and resources.

This kind of constraints can be used in a real setup for liability insurance purposes, for which the risk profile of the whole policy committed between the organisation and the resources/users is required.

8. Example: Employee Records

We consider here the scenario from Dell EMC [44], which deals with policies regulating employee records' access and update in an organisation. The example

starts with the assumption that any user of a record management application can access and update their own organisational records in addition to any direct and indirect reports generated based on those records. Furthermore, other employees in the organisation with proper authorisation, e.g. employees in the human resources department, can also view all the other employees' records.

Sam is an employee who is an end-user. He requires access to his self-appraisal document, `self-appraisal-2009`. Additionally, both Michelle, who is Sam's manager, and Peter, who is Michelle's manager and head of division, require access to Sam's absence records. Finally, Diane, who is an officer in the human resources department requires only a view access to the records. A simple XACML policy expressing the above scenario as an access control list is depicted below [44]:

```
<Policy PolicyId="pol_self-appraisal-2009" RuleCombiningAlgId=
    "urn:oasis:names:tc:xacml:1.0:rule-combining-algorithm:permit-overrides">
  <Description>ACLs for Sam's 2009 self appraisal</Description>
  <Target>
    <Resources>
      <Resource>
        <ResourceMatch MatchId=
            "urn:oasis:names:tc:xacml:1.0:function:string-equal">
          <AttributeValue DataType=
              "http://www.w3.org/2001/XMLSchema#string"> self-appraisal-2009
          </AttributeValue>
          <ResourceAttributeDesignator>
            urn:oasis:names:tc:xacml:1.0:resource:resource-id
          </ResourceAttributeDesignator>
        </ResourceMatch>
      </Resource>
    </Resources>
  </Target>
<Rule RuleId="rul_self-appraisal-2009_sam" Effect="Permit">
  <Description>Sam can do anything with his 2009 self appraisal</Description>
  <Target>
    <Subjects>
      <Subject>
        <SubjectMatch MatchId=
            "urn:oasis:names:tc:xacml:2.0:function:string-equal">
          <AttributeValue DataType=
              "http://www.w3.org/2001/XMLSchema#string"> Sam
          </AttributeValue>
          <SubjectAttributeDesignator DataType=
              "http://www.w3.org/2001/XMLSchema#string">
            urn:oasis:names:tc:xacml:1.0:subject:subject-id
          </SubjectAttributeDesignator>
        </SubjectMatch>
      </Subject>
    </Subjects>
  </Target>
</Rule>
<Rule RuleId="rul_self-appraisal-2009_michelle+peter+diane" Effect="Permit">
  <Description>Sam's bosses and HR can only view the report</Description>
  <Target>
    <Actions>
      <Action>
        <ActionMatch MatchId=
            "urn:oasis:names:tc:xacml:1.0:function:string-equal">
          <AttributeValue DataType=
              "http://www.w3.org/2001/XMLSchema#string">
            urn:oasis:names:tc:xacml:1.0:action:action-id
          </AttributeValue>
          <ActionAttributeDesignator>
            view
          </ActionAttributeDesignator>
```

```
      </ActionMatch>
    </Action>
  </Actions>
 </Target>
 <Condition>
  <Apply FunctionId=
      "urn:oasis:names:tc:xacml:1.0:function:string-is-in">
    <SubjectAttributeDesignator DataType=
        "http://www.w3.org/2001/XMLSchema#string">
      urn:oasis:names:tc:xacml:1.0:subject:subject-id
    </SubjectAttributeDesignator>
    <Apply FunctionId=
        "urn:oasis:names:tc:xacml:1.0:function:string-bag">
      <AttributeValue DataType=
          "http://www.w3.org/2001/XMLSchema#string"> Michelle
      </AttributeValue>
      <AttributeValue DataType=
          "http://www.w3.org/2001/XMLSchema#string"> Peter
      </AttributeValue>
      <AttributeValue DataType=
          "http://www.w3.org/2001/XMLSchema#string"> Diane
      </AttributeValue>
    </Apply>
  </Apply>
 </Condition>
</Rule>
</Policy>
```

In the above example, subjects are identified using the attribute (part of the incoming request) `subject-id`, which is matched against the above names using the functions `string-equal` and `string-bag`. In order to obtain the type of actor and asset variety in terms of the VERIS schema, we give the following two transformation definitions, so that we can map the subjects and the resources to the VERIS schema:

$\mathcal{T}_{su} = \{(Sam, \textit{End-user}), (Michelle, Manager), (Peter, Executive),$
$(Diane, HR)\}$

$\mathcal{T}_{re} = \{(\textit{self-appraisal-2009}, Documents)\}$

These definitions are only for demonstration and are constructed based on the scenario itself. In a real case, they will need to be predefined in advance by the policy administration team in the organisation, in a more efficient and general manner.

The risk probability for each of the above users causing a security damage to `self-appraisal-2009` as taken from the definition of r, is as follows:

- For Sam, $r(\textit{End-user}, Documents) = 1.927\%$
- For Michelle, $r(Manager, Documents) = 0.172\%$
- For Peter, $r(Executive, Documents) = 0.196\%$
- For Diane, $r(HR, Documents) = 0.061\%$

Now we consider how each of the constraints we defined earlier is upheld under this scenario.

8.1. Evaluation of Risk Probability Constraints

The organisation may have various levels of appetite in relation to each of the constraints mentioned above. We consider the following examples:

- $\mathcal{C}_{actor}(2.00\%, \texttt{pol\_self-appraisal-2009})$

- $\mathcal{C}'_{actor}(2.00\%, \textit{End-user}, \texttt{pol\_self-appraisal-2009})$,
  $\mathcal{C}'_{actor}(0.2\%, \textit{Manager}, \texttt{pol\_self-appraisal-2009})$,
  $\mathcal{C}'_{actor}(0.2\%, \textit{Executive}, \texttt{pol\_self-appraisal-2009})$,
  $\mathcal{C}'_{actor}(0.1\%, \textit{HR}, \texttt{pol\_self-appraisal-2009})$

- $\mathcal{C}_{asset}(2.00\%, \textit{Documents}, \texttt{pol\_self-appraisal-2009})$

- $\mathcal{C}'_{asset}(2.50\%, \textit{Documents}, \texttt{pol\_self-appraisal-2009})$

- $\mathcal{C}_{policy}(2.50\%, \texttt{pol\_self-appraisal-2009})$

In this case, we see that all of these constraints evaluate to **True** as no individual type of actors or assets exceeds their risk probability limit and the overall policy risk probability also remains within its limit.

However, assume that, as part of a new event requiring the re-structuring of the organisation, new members of the board of executives Graham and Gail are to be allowed access to Sam's record. Despite the fact that both these members still maintain the validity of the rest of the constraints, they however cause the violation of the last two, $\mathcal{C}'_{asset}$ and $\mathcal{C}_{policy}$. This is because:

$\mathcal{C}'_{asset}(2.50\%, \textit{Documents}, \texttt{pol\_self-appraisal-2009}) =$
$(\sum_{i=1}^{|MSet|} r(a_i, \textit{Documents})) \leq 2.5\%$

where, $MSet = \{|\textit{End-user}, \textit{Manager}, \textit{Executive}, \textit{Executive}, \textit{Executive}, \textit{HR}|\}$

And therefore, $\mathcal{C}'_{asset}(2.50\%, \textit{Documents}, \texttt{pol\_self-appraisal-2009}) =$
$(1.927\% + 0.172\% + 0.196\% + 0.196\% + 0.196\% + 0.061\%) \leq 2.5\% =$ **False**

which violates the constraint. As a result, the organisation may re-consider the requirement that Graham and Gail should have access to Sam's record, or if not possible, it will consider how to include additional protection mechanisms for such a record.

9. Discussion

Our approach presented in this paper helps assign a risk meaning to access control policies in terms of understanding the danger that specific types of users can pose to organisational assets. Unlike most of the existing literature, it

is *data-driven* and therefore evidence-based. In many ways, this is important when configuring and deploying policies particularly at the bootstrapping stage, when little or no local evidence is present at the organisation or team levels, and therefore it is useful to draw from evidence present in open datasets such as VERIS/VCDB.

We find however that the nature of open data may not be ideal or suited to the kind of information an analyst would be looking for. For example, we had to modify some of our probability definitions (e.g. P_{risk}) to take into account the lack of any data on general accesses to organisational resources where no attacks take place. Another example is the lack of data on the impact of incidents, which restricts a more generalised definition of the risk of access control policies.

Nonetheless, we feel that the current approach provides benefits in terms of the quantitative optimisation of access control policies. For example, it is possible to conceive of the constraints presented in Section 7 as part of a *chance constraint* problem that can be solved through a program [45] with the aim of optimising the risk level of an access control policy. This optimisation then would lead to the reconfiguration of the policy in order to meet business and administrative requirements and conditions (see for example previous work by the author in [21]).

Finally, the current model as it stands is VERIS-focused. While not necessarily a negative aspect as it grounds the model on some concrete dataset (and thus fulfils the evidence-based claim) and it provides a standardised language for describing security events, it does limit the information that can be derived when reasoning about the risk level of a policy. In particular, the lack of correlation among the various events recorded in VCDB makes any frequency-based analysis difficult to make.

There are a few directions for future work that we discuss next in the concluding section.

10. Conclusion and Future Work

In this paper we presented a data-driven definition of the probabilities of subjects damaging assets in organisations, which is rooted in the large open data set of security incidents called VCDB.

There are several strands for future work that stem from the current work. In the current paper, the focus was on internal roles of actors, however, VERIS also provides definitions of roles for external and partner roles. This would promote the understanding of the risk probabilities for inter-organisational policies, for example, such as those deployed in virtual organisations or federations. Another direction would be to incorporate security attributes into the risk formula, which currently is limited to the probability of security compromises without considering their effects. These attributes in VERIS are known as the Parkerian Hexad [3], which include confidentiality, possession, integrity, authenticity, availability and utility.

We also plan to consider the impact of incidents into the risk formula to define a more complete risk equation which includes both probability of the

incident and its impact. VERIS includes information on the impact of incidents in terms of the type of loss and its value incurred as a result of the incident, however VCDB's data provided on such information is currently poor, but this could change in future editions of the VCDB.

Another area of improvement to the current approach is to consider the risk of *denying* legitimate access to resources. However, since our approach is entrenched in the data-driven philosophy, we will first need to discover publicly available or obtain private datasets that contain information on such type of risks, before we can extend our approach to include such risk.

Finally, in order for this approach to be usable, it is necessary as part of future work to develop new tools or extend existing ones incorporate the evidence produced from datasets and to use that evidence to evaluate any access control constraints required by the organisation in protecting its resources.

Bibliography

References

[1] P. Koehn, F. J. Och, D. Marcu, Statistical Phrase-based Translation, in: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1, NAACL '03, Association for Computational Linguistics, Stroudsburg, PA, USA, 2003, pp. 48–54.

[2] K. J. S. Hoo, How Much is Enough? A Risk-Management Approach to Computer Security (6 2000).

[3] D. B. Parker, Fighting Computer Crime: A New Framework for Protecting Information, John Wiley & Sons, Inc., New York, NY, USA, 1998.

[4] VERIZON, VERIS Community Database, last accessed: 21.11.2016.
URL `http://vcdb.org/`

[5] CERT Coordination Center, CERT Vulnerability Notes Database, last accessed: 14.08.2017.
URL `http://www.kb.cert.org/vuls`

[6] Mike Sconzo, SecRepo.com - Samples of Security Related Data, last accessed: 14.08.2017.
URL `http://www.secrepo.com`

[7] Center for Applied Internet Data Analysis, CAIDA Data, last accessed: 14.08.2017.
URL `http://www.caida.org/data/overview/`

[8] Los Alamos National Laboratory, Cyber Security Science Open Data Sets, last accessed: 14.08.2017.
URL `http://csr.lanl.gov/data/`

[9] E. Rissanen, eXtensible Access Control Markup Language (XACML) Version 3.0, OASIS Standard (2013).

[10] VERIZON, The Vocabulary for Event Recording and Incident Sharing (VERIS), last accessed: 21.11.2016.
URL http://veriscommunity.net/

[11] R. von Mises, Probability, Statistics and Truth, 1939.

[12] G. Davis, A. Garcia, W. Zhang, Empirical Analysis of the Effects of Cyber Security Incidents, Risk Analysis 29 (9) (2009) 1304–1316.

[13] M. A. Kuypers, T. Maillart, E. Paté-Cornell, An Empirical Analysis of Cyber Security Incidents at a Large Organization, working paper (2016).

[14] A. Sarabi, P. Naghizadeh, Y. Liu, M. Liu, Prioritizing Security Spending: A Quantitative Analysis of Risk Distributions for Different Business Profiles, in: the Annual Workshop on the Economics of Information Security (WEIS)(June 2015), 2015.

[15] Thomas Lenard and Scott Wallsten , An Economic Analysis of the FCC's Privacy Notice of Proposed Rulemaking, last accessed: 05.01.2017.
URL https://techpolicyinstitute.org/wp-content/uploads/2016/05/Lenard_Wallsten_FCCprivacycomments.pdf

[16] N. Dimmock, A. Belokosztolszki, D. Eyers, J. Bacon, K. Moody, Using Trust and Risk in Role-based Access Control Policies, in: Proceedings of the Ninth ACM Symposium on Access Control Models and Technologies, SACMAT '04, ACM, New York, NY, USA, 2004, pp. 156–162.

[17] J. Ma, K. Adi, M. Mejri, L. Logrippo, Risk analysis in access control systems, in: 2010 Eighth International Conference on Privacy, Security and Trust, 2010, pp. 160–166.

[18] P. C. Cheng, P. Rohatgi, C. Keser, P. A. Karger, G. M. Wagner, A. S. Reninger, Fuzzy multi-level security: An experiment on quantified risk-adaptive access control, in: 2007 IEEE Symposium on Security and Privacy (SP '07), 2007, pp. 222–230.

[19] C. Chen, W. Han, J. Yong, Specify and Enforce the Policies of Quantified Risk Adaptive Access Control, in: The 2010 14th International Conference on Computer Supported Cooperative Work in Design, 2010, pp. 110–115.

[20] I. Molloy, L. Dickens, C. Morisset, P.-C. Cheng, J. Lobo, A. Russo, Risk-based Security Decisions Under Uncertainty, in: Proceedings of the Second ACM Conference on Data and Application Security and Privacy, CODASPY '12, ACM, New York, NY, USA, 2012, pp. 157–168.

[21] B. Aziz, S. Foley, J. Herbert, G. Swart, Reconfiguring Role Based Access Control Policies using Risk Semantics, Journal of High Speed Networks 15 (3) (2006) 261–273.

[22] K. Z. Bijon, R. Krishnan, R. Sandhu, Risk-Aware RBAC Sessions, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 59–74.

[23] K. Z. Bijon, R. Krishnan, R. Sandhu, A Framework for Risk-aware Role Based Access Control, in: 2013 IEEE Conference on Communications and Network Security (CNS), 2013, pp. 462–469.

[24] L. Krautsevich, A. Lazouski, F. Martinelli, P. Mori, A. Yautsiukhin, Integration of Quantitative Methods for Risk Evaluation within Usage Control Policies, in: 2013 22nd International Conference on Computer Communication and Networks (ICCCN), 2013, pp. 1–8.

[25] D. Ferraiolo, R. Kuhn, Role-based Access Controls, in: Proceedings of 15th NIST-NSA National Computer Security Conference, Baltimore, MD, USA, 1992, pp. 554–563.

[26] T. Moses, eXtensible Access Control Markup Language (XACML) Version 2.0, OASIS Standard (2005).

[27] J. Park, R. Sandhu, The UCON_{abc} Usage Control Model, ACM Transactions on Information and System Security 7 (1) (2004) 128–174.

[28] M. Colombo, A. Lazouski, F. Martinelli, P. Mori, A Proposal on Enhancing XACML with Continuous Usage Control Features, Springer US, Boston, MA, 2010, pp. 133–146.

[29] C. Schläger, T. Nowey, Towards a Risk Management Perspective on AAIs, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 41–50.

[30] L. Zhang, A. Brodsky, S. Jajodia, Toward information sharing: benefit and risk access control (BARAC), in: Seventh IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY'06), 2006, pp. 9 pp.–53.

[31] Y. Li, J. Ren, H. Sun, H. Luo, Z. Chen, Trust-Risk-Game Based Access Control in Cross Domain Application, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 89–102.

[32] M. Alizadeh, N. Zannone, Risk-based Analysis of Business Process Executions, in: Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy, CODASPY '16, ACM, New York, NY, USA, 2016, pp. 130–132.

[33] B. Stepien, S. Matwin, A. P. Felty, Strategies for Reducing Risks of Inconsistencies in Access Control Policies, in: ARES 2010, Fifth International Conference on Availability, Reliability and Security, 15-18 February 2010, Krakow, Poland, IEEE Computer Society, 2010, pp. 140–147.

[34] S. D. Nogoorani, R. Jalili, TIRIAC: A Trust-driven Risk-aware Access Control Framework for Grid Environments, Future Generation Computer Systems 55 (2016) 238 – 254.

[35] D. R. dos Santos, R. Marinho, G. R. Schmitt, C. M. Westphall, C. B. Westphall, A Framework and Risk Assessment Approaches for Risk-based Access Control in the Cloud, Journal of Network and Computer Applications 74 (2016) 86 – 97.

[36] J. Crampton, C. Morisset, N. Zannone, On Missing Attributes in Access Control: Non-deterministic and Probabilistic Attribute Retrieval, in: Proceedings of the 20th ACM Symposium on Access Control Models and Technologies, SACMAT '15, ACM, New York, NY, USA, 2015, pp. 99–109.

[37] FIRST, Common Vulnerability Scoring System v3.0: Specification Document, last accessed: 14.08.2017.
URL https://www.first.org/cvss/specification-document

[38] J. Robert H. Courtney, Security Risk Assessment in Electronic Data Processing Systems, in: Proceedings of the June 13-16, 1977, National Computer Conference, AFIPS '77, ACM, New York, NY, USA, 1977, pp. 97–104.

[39] U. D. of Commerce, Guideline for Automatic Data Processing Risk Analysis, Tech. Rep. FIPS PUB 65, National Bureau of Standards, Washington, DC, USA (1975).

[40] C. Ishida, Y. Arakawa, I. Sasase, K. Takemori, Forecast Techniques for Predicting Increase or Decrease of Attacks using Bayesian Inference, in: PACRIM. 2005 IEEE Pacific Rim Conference on Communications, Computers and signal Processing, 2005., 2005, pp. 450–453.

[41] G. Magklaras, S. Furnell, Insider Threat Prediction Tool: Evaluating the Probability of {IT} Misuse, Computers & Security 21 (1) (2001) 62 – 73.

[42] K. Campbell, L. A. Gordon, M. P. Loeb, L. Zhou, The Economic Cost of Publicly Announced Information Security Breaches: Empirical Evidence from the Stock Market, Journal of Computer Security 11 (3) (2003) 431–448.

[43] A. Anderson, XACML Profile for Role Based Access Control (RBAC), OASIS Standard (2004).

[44] Remon Sinnema, XML and Security: Real World Examples of XACML Security Policies, last accessed: 02.01.2017.
URL https://community.emc.com/docs/DOC-7410

[45] A. Charnes, W. W. Cooper, Chance-Constrained Programming, Management Science 6 (1) (1959) 73–79.

Appendix A. VCDB-based Definition of the r Function

The diagrams below (Figures A.3-A.17) represent the percentage of incidents in which a given Actor will damage a particular Asset Variety, as recorded in the VCDB dataset. The horizontal values are therefore percentages. For example, in Figure A.3, the probability that an Auditor may damage a Database Server is about 0.037% of the VCDB incidents.

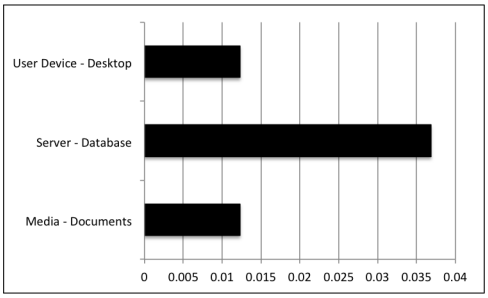

Figure A.3: Auditors Risk Probabilities

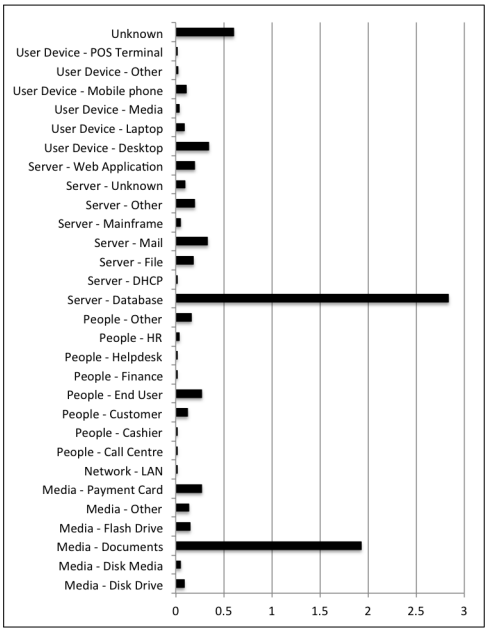

Figure A.4: End-users Risk Probabilities

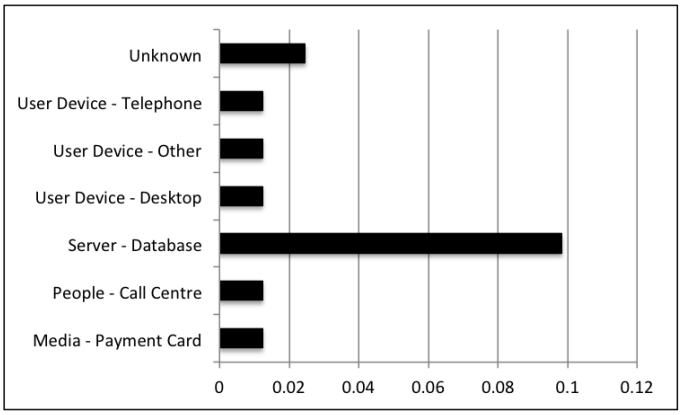

Figure A.5: Call Centre Risk Probabilities

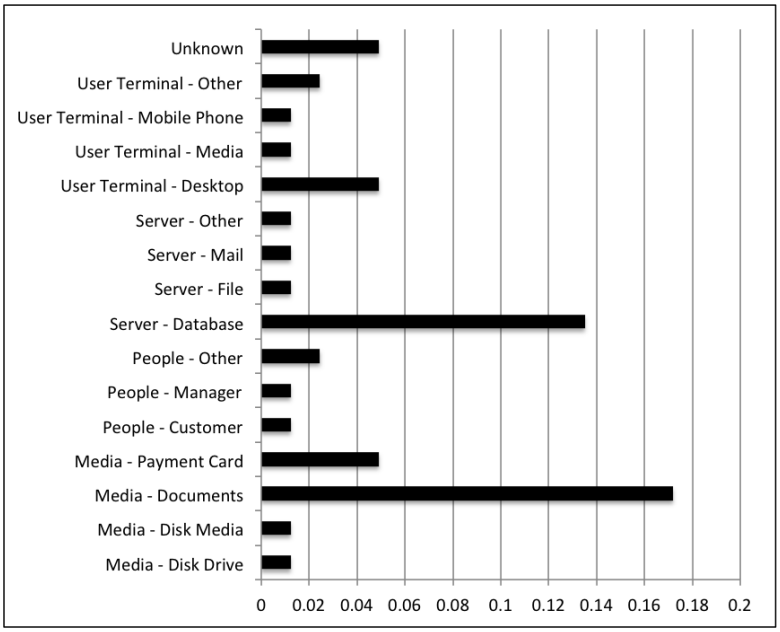

Figure A.6: Managers Risk Probabilities

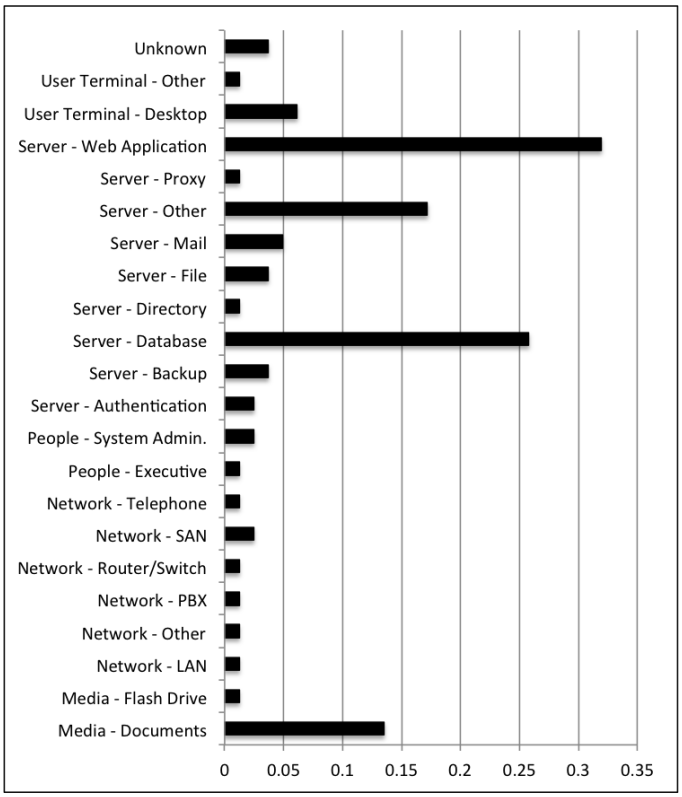

Figure A.7: System Administrators Risk Probabilities

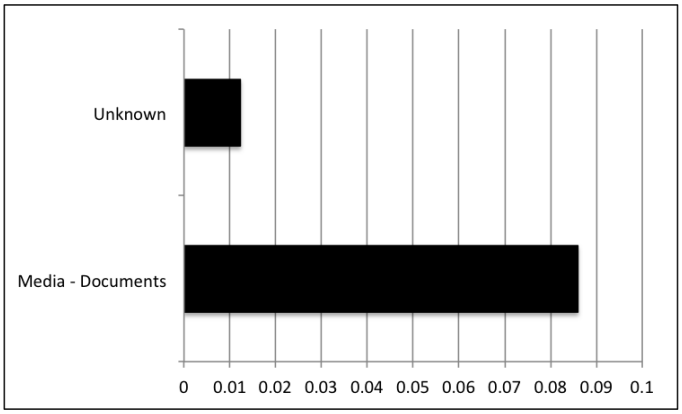

Figure A.8: Maintenance Risk Probabilities

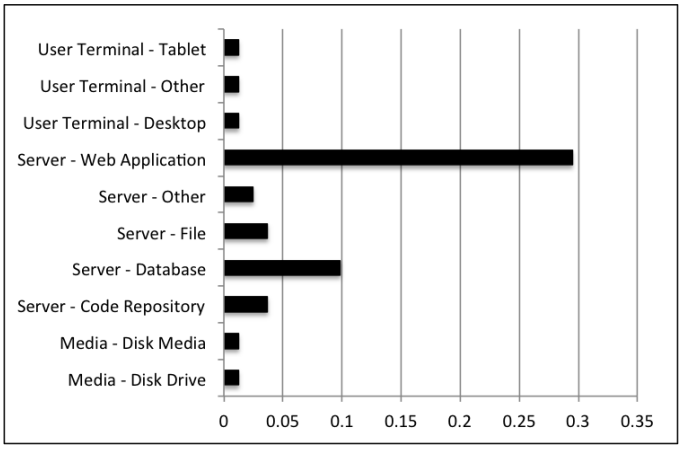

Figure A.9: Developers Risk Probabilities

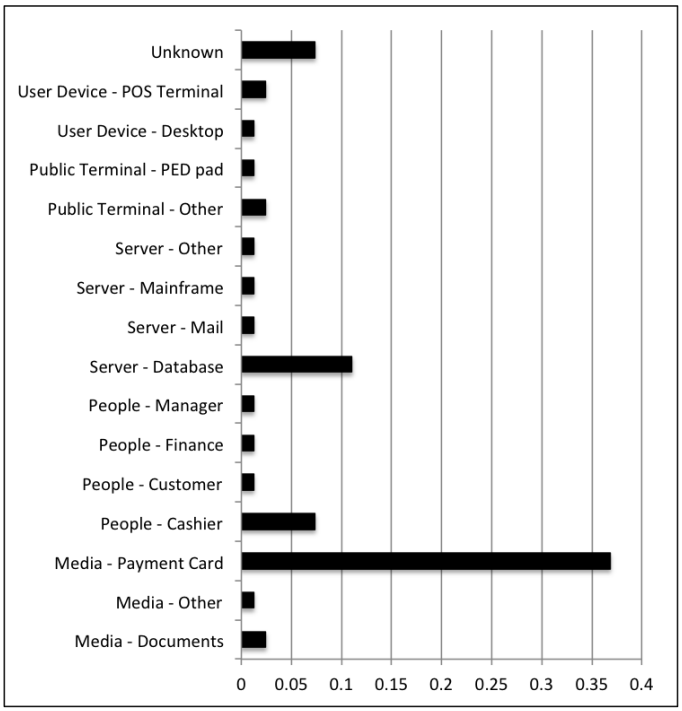

Figure A.10: Cashiers Risk Probabilities

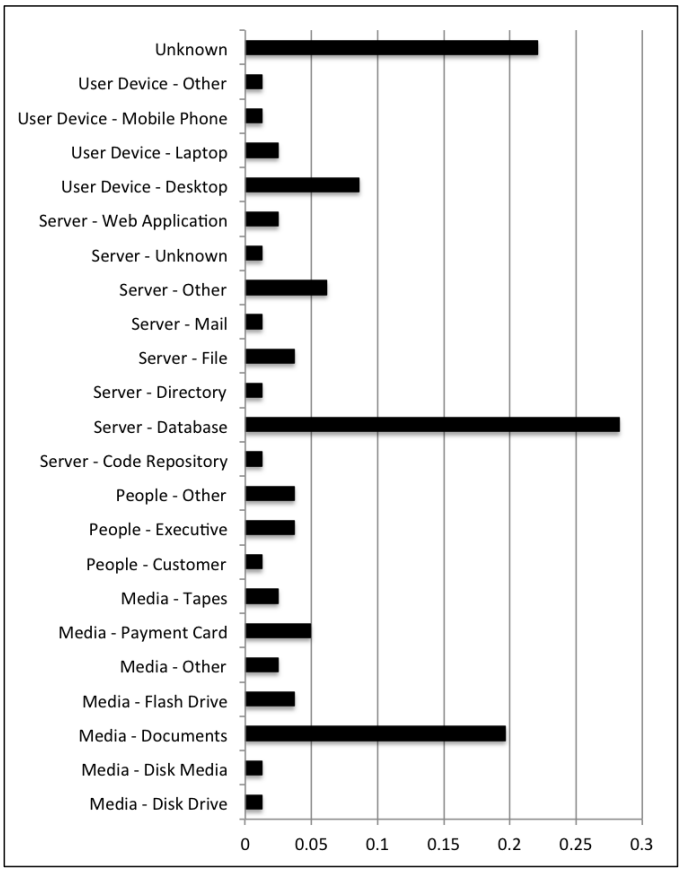

Figure A.11: Executives Risk Probabilities

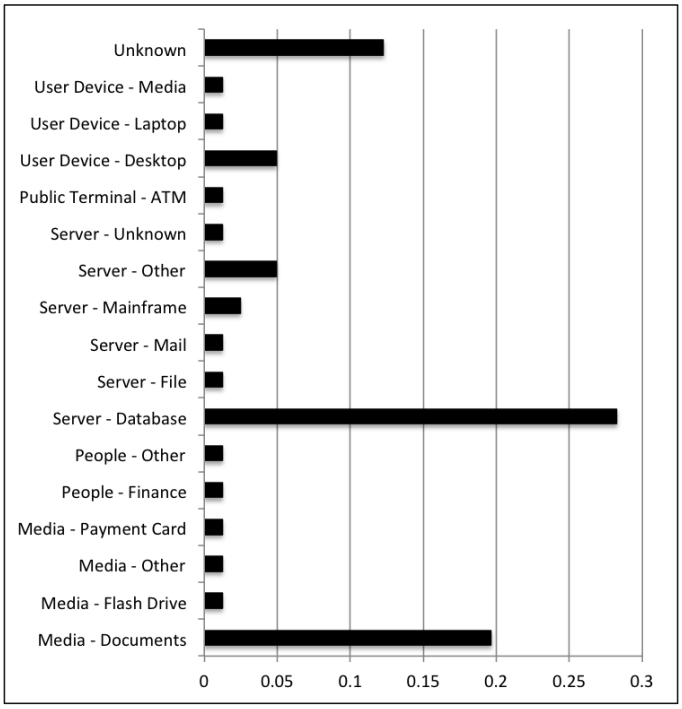

Figure A.12: Finance Risk Probabilities

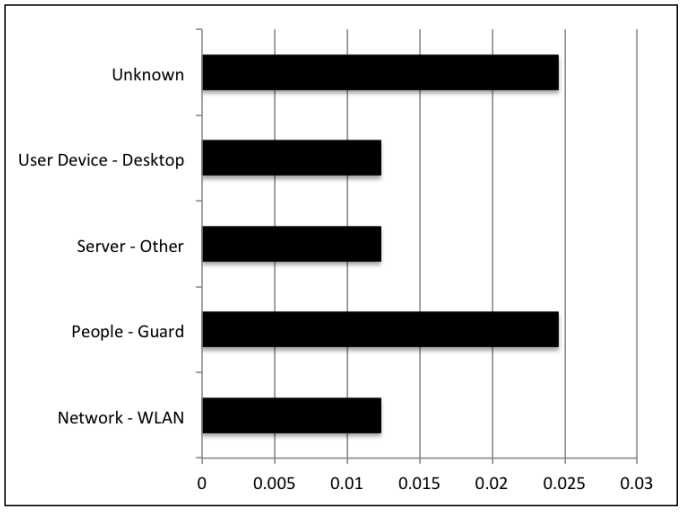

Figure A.13: Guard Risk Probabilities

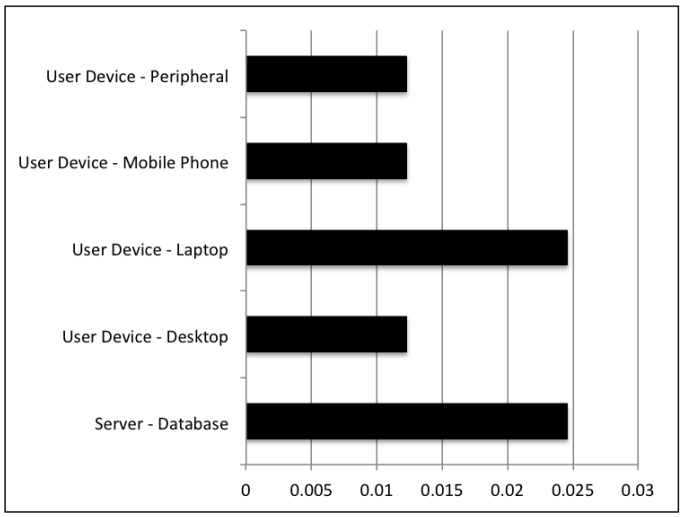

Figure A.14: Helpdesk Risk Probabilities

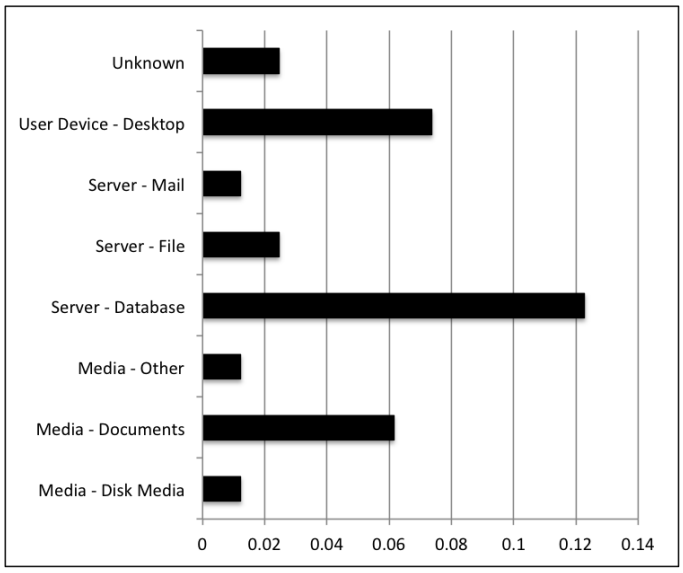

Figure A.15: HR Risk Probabilities

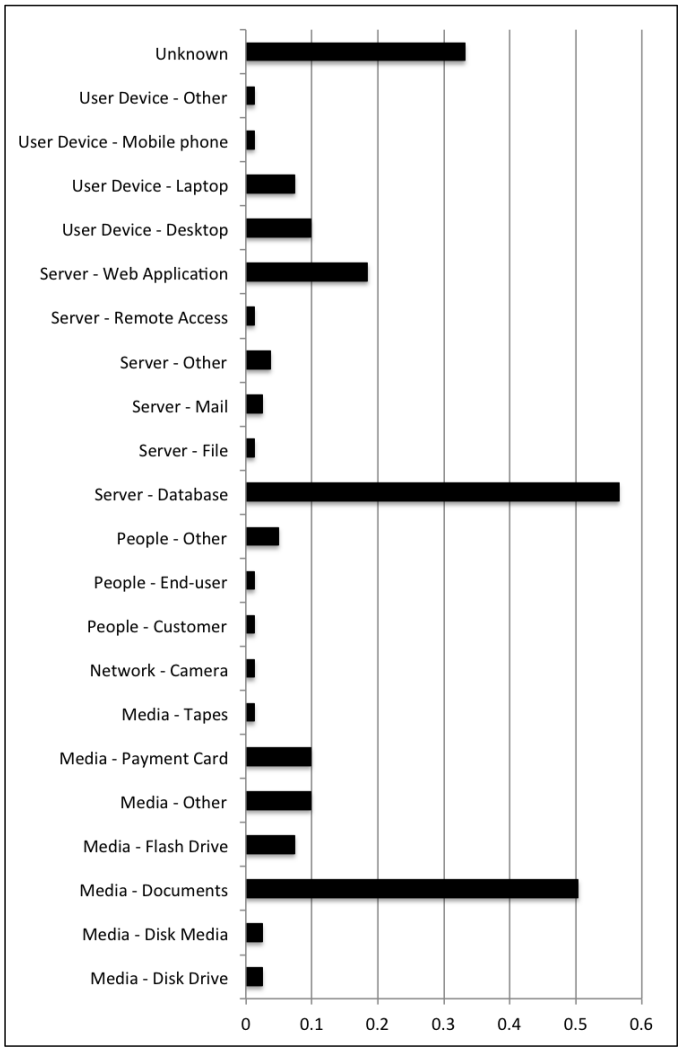

Figure A.16: Risk Probabilities for All the Other Roles

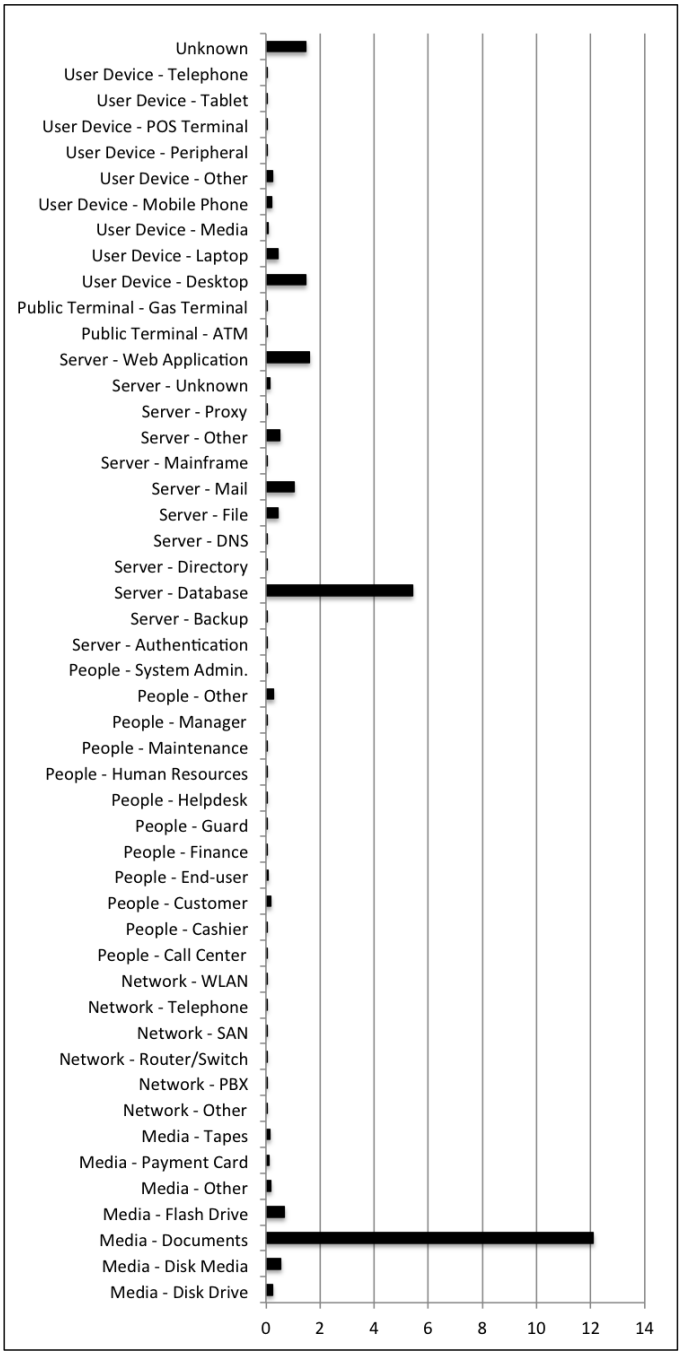

Figure A.17: Risk Probabilities for All the Unknown Roles