ZeTA – Zero-Trust Authentication: Relying on Innate Human Ability, not Technology

Andreas Gutmann*[||], Karen Renaud[†], Joseph Maguire[†], Peter Mayer*,
Melanie Volkamer*[‡], Kanta Matsuura[§], and Jörn Müller-Quade[¶]

*Technische Universität Darmstadt
name.surname@secuso.org

[†]University of Glasgow
name.surname@glasgow.ac.uk

[‡]Karlstad University

[§]The University of Tokyo
kanta@iis.u-tokyo.ac.jp

[¶]Karlsruhe Institute of Technology
joern.mueller-quade@kit.edu

***Abstract*—Reliable authentication requires the devices and channels involved in the process to be trustworthy; otherwise authentication secrets can easily be compromised. Given the unceasing efforts of attackers worldwide such trustworthiness is increasingly not a given. A variety of technical solutions, such as utilising multiple devices/channels and verification protocols, has the potential to mitigate the threat of untrusted communications to a certain extent. Yet such technical solutions make two assumptions: (1) users have access to multiple devices and (2) attackers will not resort to hacking the human, using social engineering techniques. In this paper, we propose and explore the potential of using human-based computation instead of solely technical solutions to mitigate the threat of untrusted devices and channels. ZeTA (*Zero Trust Authentication on untrusted channels*) has the potential to allow people to authenticate *despite* compromised channels or communications and easily observed usage. Our contributions are threefold: (1) We propose the ZeTA protocol with a formal definition and security analysis that utilises semantics and human-based computation to ameliorate the problem of untrusted devices and channels. (2) We outline a security analysis to assess the envisaged performance of the proposed authentication protocol. (3) We report on a usability study that explores the viability of relying on human computation in this context.**

1. Introduction

Knowledge-based authentication, usually the alphanumeric password, is the dominant means of authentication, and has been for decades. Quite apart from the password's well-known usability flaws it is also easily compromised. Observation, by technical or human means, is a distinct possibility in our modern world with people authenticating in uncontrolled environments using multiple devices transmitting secrets over questionable networks.

This reality requires us to rethink the standard approach which relies firstly on the user and server establishing a secret during enrolment and then has the user divulge the full secret at each authentication attempt. In this scenario, observation (leakage of the secret) can occur if: (1) the device is compromised; (2) the transmission channel is intercepted; (3) the user's entry of the secret is observed or overheard.

A number of technical solutions have been proposed to mitigate the threat of untrusted channels or devices. For example, a number of researchers propose utilising a separate independent channel to deliver a one-time key [1], [2], [3] or ensuring that secure cryptography is used for communications [4]. Other researchers propose securing the device more effectively, so as to minimise opportunities for compromise [5], [6]. It is furthermore argued that decision makers have too little knowledge about strengths and weaknesses of different authentication schemes [7] and that supporting them in making context-dependent decisions [8] would improve security.

While these are viable and worthwhile solutions they rely on users having extra channels at their disposal or having the requisite expertise to be able to install and use secure software on their mobile devices. Neither of these is a given. Moreover, they do not address the human-based observation threat.

Thus, the question arises whether the problem can be addressed at a more basic level. It might be possible for the requirement for users to disclose their full secret during an authentication attempt to be relaxed or lifted altogether. Some authentication mechanisms, so-called *limited-disclosure* systems, have already relaxed this requirement by requiring only partial disclosure. Users are challenged to provide specific parts of the secret in a challenge-response

[||] *This work was supported by CASED (www.cased.de).*

fashion. This ensures that a single observation does not betray the full secret [9]. This technique is often used by telephone banking systems to prevent call centre staff from gaining knowledge of the full secret [10]. Multiple observations, however, still reveal the full secret.

A better way to prevent observation would be to find a way for people to confirm knowledge of the secret without providing the secret itself, or even a part thereof (ideally in a zero-knowledge fashion). We propose to exploit users' *semantic knowledge* of concepts related to their authentication secret, in order to prove that they have knowledge of the underlying secret.

The ZeTA concept is that people would have to memorise an authentication secret consisting of two or more words, together with logical connections between them. They would then be challenged by the provision of an attribute, and they would have to confirm or deny whether this particular attribute is, according to the logical connections, semantically related to their secret. Thereby, our proposal does not require disclosure of the user secret and makes (even repeated) person-to-person or technical observation unfruitful. A high level view of ZeTA is provided in Figure 1.

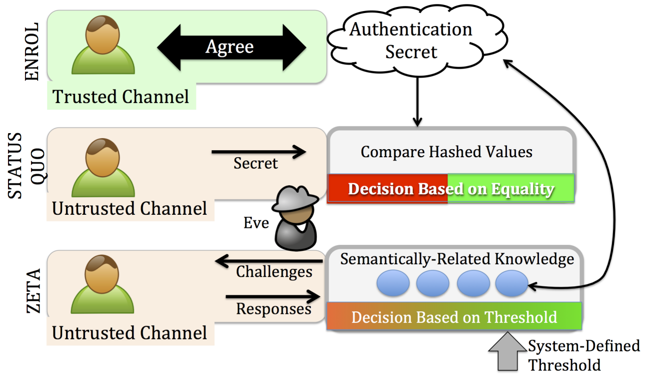

Figure 1. ZeTA: Zero-Trust Authentication

Why do we believe ZeTA to be a viable proposal? Humans seem to have an innate ability to build up semantic networks of related concepts of words and objects [11]. In fact, a disruption in the ability to mine this network is often one of the first signs of brain decline such as Alzheimer's disease [12]. So, if we can expect people to have this knowledge, does the proposed way of tapping into it seem feasible? Lee *et al.* [13] argue that the human brain stores semantic knowledge primarily in an attribute-based fashion. Hence, if people are prompted to confirm or deny the link between an attribute and a particular memorised secret, this should not be very demanding since that is the way the information is stored anyway. Moreover, by requiring people to confirm (recognise) rather than come up with these links themselves (recall), the cognitive load associated with these responses is as lightweight as possible [14].

The **main contributions** of this paper are:

- The formal definition of ZeTA, an authentication approach that relies on innate human-based computation.
- A security analysis of the proposed authentication approach.
- Empirical usability studies exploring the feasibility of the proposed authentication approach.

As a **minor contribution**, we challenge a widely held belief established by the work of Coskun and Herley [15]. They question the viability of secure challenge-response authentication approaches, and come to the daunting conclusion that they cannot be secure. However, their work concentrates on the required number of mathematical operations and the limited ability of humans to remember long secrets. ZeTA circumvents these constraints through the exploitation of readily available semantic knowledge.

The remainder of this paper is structured as follows. The next section introduces relevant background and related approaches. Section 3 describes our proposed authentication system ZeTA in detail. In section 4 we present a security analysis of ZeTA. Sections 5 and 6 describe the pilot and primary usability studies we conducted to investigate the viability of our approach. Section 7 discusses our findings and Section 8 enumerates the limitations of our approach and the study methodologies. Finally, Section 9 summarises, concludes and points out areas for future work.

2. Background

The problems of integrating authentication into systems and services has long been established [16]. There are numerous challenges, both technical and human, but a longstanding concern is securing authentication between users and verifying systems [17]. Automated Teller Machines or ATMs are a classic example of the challenges in securing the authentication process. System providers, such as banks, need to secure hardware and channels, as well as the surrounding environment, to have confidence in the authentication process. Furthermore, system providers need to ensure they are not placing users at undue risk when providing such services.

ATMs serve as an interesting consumer example as almost all aspects of the authentication process can be compromised, not only by the sophisticated and organised attacker but the opportunistic one as well. The user endpoint, i.e. the physical ATM, is susceptible to social engineering, as well as eavesdropping from opportunistic attackers determined to gain access to an account. A slightly more organised attacker with appropriate skills and resources could compromise hardware and communication channels to compromise several accounts [18].

The system-side end-point, while arguably the most challenging to attack, could be undermined by sophisticated attackers with sufficient resources [19]. The reality is that if an attacker can compromise any element of the authentication chain, they gain access to an individual's password. Consequently, secret knowledge is no longer secret and, as a result, an attacker can masquerade as the legitimate user.

Lamport argues that the difficulties of securing the authentication process could be addressed, in part, with a disposable or one-time authentication secret [17]. The verifying system does not store the user's password, x, but rather y = *F(x)*, i.e. an encrypted authentication secret where x is the key. Lamport argues that while this may prevent an attacker compromising the password at the verifying end-point, the authentication secret is still vulnerable during transmission and at the user's end-point. Consequently, Lamport proposes the notion of a sequence of passwords, x_1,x_2,. . . ,x_{1000}, where each password in the sequence can only be used once to authenticate. The sequence is generated by repeatedly applying the function, i.e. using the previous function output as input for next in sequence. The verifying system only maintains the last element in the sequence while the user starts from the penultimate element. The user submitted one-time password acts as the input for the function, authentication is successful if the output matches that of the list element maintain by the server. Upon successful authentication, the server discards the current list element and stores the element submitted by the user.

Haller *et al.* [20] implemented the aforementioned scheme as S/KEY authentication. The implementation relies on 64-bit numbers, but these are mapped to words of varying length to reduce the burden on the user. Nevertheless, while an interesting scheme, the approach is still vulnerable to man-in-the-middle attacks. Furthermore, the approach typically relies on users to either manage lengthy lists of disposable passwords or to carry a specialised device capable of generating a disposable password when needed.

The expense and inconvenience of using the scheme and its variants should not be discounted. This has fuelled research into developing a knowledge-based approach for a single untrusted channel. These schemes are often challenge-based, whereby the user and verifier share a secret, the verifier challenges the user to demonstrate knowledge of the secret, ideally in a way that does not divulge the shared secret [21].

Matsumoto and Imai [22] propose such a scheme, where the shared knowledge is a secret string and sequenced symbol set. The verifying system challenges the user with a random string, i.e. *str* = c_1,. . . ,c_n, the user is expected to substitute specific characters in the string and subsequently return it. The alternative string is based on the following rule: when a character within a string matches that of one in the sequenced symbol set — the user is to substitute it. The substitution is drawn from the secret string, using the position of the matching symbol in the secret sequence to select the correct character.

Similarly, Hopper and Blum [23] propose a scheme where the shared secret is a vector. In order to authenticate, the user is presented with a series of challenge vectors. The user has to calculate and return the outcome of the dot product of a each challenge vector with the secret vector, but intentionally give a wrong answer with some pre-agreed low probability. While the scheme is interesting and addresses the challenges of authenticating over insecure channels, it is undeniably difficult for users to perform [21], [24].

Nevertheless, there have been many proposals to address the problems of untrusted communication channels and devices. Mannan and van Oorschot [25] outline a protocol that relies on trusted devices while still utilising untrusted devices. Thorpe *et al.* [26] theorise that advances in technology may solve the problem of being observed by reading a user's thoughts. McCune [27] proposes the use of a visual channel to secure authentication, users scan barcodes with their smartphones. Roth *et al.* [28] propose a challenge-based approach to PIN-entry that minimises the threat of user's being observed entering their authentication secret . Nevertheless, while the proposed approach is not particularly burdensome, especially so when contrasted with the aforementioned approaches, Roth *et al.* reported that users simply did not perceive the benefits as worth the cost of entry.

Many of the schemes designed to support authentication in an untrusted environment do indeed resist observation but are effectively unusable. In the next section we outline ZeTA, a more usable scheme. We will also show that it resists observation in untrusted environments.

3. Authentication Approach

In this section we present a high-level description of our approach and identify constraints that arise due to the intended technical features of ZeTA.

3.1. Pre-Considerations and High-Level Description

The two main features the proposed protocol should fulfil are resistance against (1) untrusted devices and communications and (2) shoulder-surfers. To achieve this, a challenge-response style of interaction is proposed. The user's part needs to be executed mentally, without any external artifacts.

The ZeTA secret consists of a number of words. The protocol exploits users' *semantic knowledge* of the meanings of the secret words, and related attributes. Users are asked personalised questions with general answers, i.e. answers that constitute general knowledge about the secret. One of the main advantages in using human processing is that much of the knowledge stored in semantic memory has a high inter-human correlation, i.e. a large fraction of this knowledge is consistent among most people [29], [30], [31], [32], [33], [34].

For example, consider the ZeTA secrets *'Blue OR Tree'* and *'Red OR Bike'*. The basic idea is that the user receives challenges from the server such as *'Can any of your terms be used as a means of transportation?'*. The answer for the first secret would be *'No'* and for the second one would be *'Yes'* because you can ride the bike. If the owner of the first secret was presented with the challenge *'Do any of the terms in your secret have something in common with wine?'* he or she would answer *'No'*. If the secret were *'Red OR Mouse'* he or she would answer *'Yes'*, because wine is likely to be red.

The response to each challenge queries the user's pre-existing semantic knowledge. The user only memorises his/her secret and understands the task: confirm or deny a semantic association. Thus a large set of different challenges is possible without being constrained by the user's capabilities of memorising new content.

In order to facilitate the procedure outlined above, access to a reliable knowledge base comprising semantic relationships between its entries is required. It is needed to issue the challenges and to verify the user's responses. This requirement is discussed in greater detail in the next section.

3.2. Knowledge Base

In this section we introduce relevant concepts and terminologies from psychology and computational linguistics. We first define our terminology and provide a simplified description of semantic memory. This is followed by an explanation of semantic relations and relatedness. The section concludes with information on the measurement of semantic relatedness and how it supports derivation of a knowledge base.

3.2.1. Terminology. Each word or symbol can have more than one sense, e.g. orange can be a fruit or a colour, each with a broad meaning. The sum of the meanings of all senses is called the *semantic concept* of this expression. Semantic concepts can be simple and complex. We are only interested in simple concepts, e.g. *large* and *house*, rather than complex ones such as *large house*. Throughout this explanation, the term 'word' is frequently used interchangeably to represent of a semantic concept.

3.2.2. Semantic memory. The knowledge base for our protocol is supposed to be largely consistent with certain parts of the user's *semantic memory*, of which we give a basic description next.

Declarative memory (part of the long term memory) consists of episodic and semantic parts. Episodic memory stores information related to personal memories and experiences. Semantic memory, on the other hand, is the mental storage of our knowledge about the world in general. It contains "*organised knowledge [...] about words and other verbal symbols, their meaning and referents, [and] about relations among them*" [35] (p. 386). This implies that semantic memory is, among other things, responsible for the use of language and serves as a mental dictionary and thesaurus [36], [37].

3.2.3. Semantic relations and relatedness. When people read a text or listen to a speaker, they subconsciously recognise relations between words. Examples of semantic relations are *synonymy* (same meaning, e.g. 'to suggest' is a synonym for 'to propose') and *hyponyms* (superordinate, e.g. 'animal' is a hyponym for 'dog'). Dissimilar expressions can be semantically related too, e.g. 'sun' and 'day' are *functionally related*. These relations help give sentences meaning, extend over sentence boundaries and contribute to understanding the text.

Semantic relatedness is a special form of linguistic distance between words. It is designed to quantify semantic closeness of two semantic concepts; a unit of measurement thereof. This measurement yields high values for pairs in a semantic relation (synonyms, hyponyms, free associations, etc.) or any other kind of lexical, functional or logical association that may exist between the semantic concepts, and low values for unrelated pairs.

3.2.4. Measurement processes. We will sketch three general directions of ongoing research into multiple ways of measuring semantic relatedness.

The first approach arrives at scores by means of labour-intensive experiments. Human annotators are presented with selected word pairs and asked to rate their relatedness. Such ratings can either be binary decisions or a value on a specified scale. The overall relatedness is computed from the overall average rating for each word pair. Such manual-rating experiments have been conducted several times [38], [39]. More recently such experiments have frequently been crowd sourced, e.g. Bruni *et al.* [40] and Lofi [41].

WordNet [42], [43] is slightly different since the annotators in a first step were presented with sentences and asked to tag the semantic meaning of certain words therein. Then, in a second step, the resulting semantic concepts were evaluated on whether certain semantic relations between them exist.

Another procedure for acquiring relatedness measurements is word association experiments. Participants are presented with one word, the cue, and asked to respond with the first associated word that comes to mind. Answers that appear most frequently are accorded a higher relatedness. Examples of word association experiments are [44], [45], [46].

The third approach uses automated systems which foster large corpora of text, e.g. Wikipedia, and analyse them based on different criteria. Examples of such approaches are manifold [47], [48], [49], [50], [51], [52], [53].

The first approach is, for obvious reasons, the most accurate way to construct a database that needs to subsequently align with human judgment. Unfortunately, such approaches are also costly and time consuming. Word association experiments are promising but lack the ability to prompt a measurement of the relatedness of chosen word pairs. Lastly, the automated fostering of large corpora of text is fast and cheap, but the correlation with human judgement is unpredictable [54]. This arguably results less from wrong assumptions of related word pairs, but more from missing connections and relationships.

3.2.5. Generating a knowledge base. A combination of the three approaches seems to be better than one of them alone. Because automated systems are fast and cheap, it is reasonable to use the resulting generated data for either of the following:

- As a first step, the data from automated systems could be used as foundation. In a second step the correctness of this data can be improved by augmenting it with the results of human judgements or word association

tasks. Boyd-Graber *et al.* [55] followed this procedure when extending WordNet with evocation relations (a measurement of how much the first concept brings the second one to mind). Their augmentation was conducted by a human judgement task involving 20 participants.
- The data from automated systems, and from word associations tasks, could be used to accelerate human judgement tasks. In this case fewer human annotations per word-pair would be required as long as their judgements are in line with data collected elsewhere.

A complication during this process is that knowledge can differ significantly depending on the user's environment, e.g. it is influenced by culture and educational background. As stated previously, prior research has shown a significant cross-cultural overlap [29], [30], [31], [32], [33], [34] on the relatedness ratings of certain kinds of semantic knowledge. Relations between most natural, physical and biological concepts, e.g. *sun* and *warm*, are expected to be similar for the vast majority of humans. A knowledge base could contain only such cross-culturally recognised concepts, but these are limited by their very nature. Another possibility would be to have multiple knowledge bases, each specific to a certain subset of users, e.g. Western First World or Third World South American. Such knowledge bases would be constrained to include only those word-pairs for which the relatedness measurement overlaps culturally and socially for the target subset of users.

Moreover, there will always be word pairs upon which users have significantly different opinions about whether they are related or not. It would be useful for the knowledge base to include anticipated agreement information, e.g. the mean and standard deviation in human judgment tasks. The server could infer therefrom which challenges not to use in combination with any given secret. The exclusion of those challenges could reduce the user's error rate.

3.3. Definition

3.3.1. The Secret. ZeTA supports simple and complex secrets: an ordered list of words concatenated with logical operators from the set *AND*, *OR* and *NOT*. For simplicity sake, we can assume that the secret is in disjunctive normal form (DNF). This assumption is without loss of generality, as any ZeTA secret can be rewritten as a DNF.

Definition 1 (Secret)**.**
Let $\mathcal{D}$ be a dictionary of words. A secret ξ is w.l.o.g. a disjunctive normal formula and each literal in ξ consists of a word $w_i \in \mathcal{D}$.

The meaning of semantic relatedness between such a secret and another word is according to sentential logic. This means that any word is related with a secret if and only if it is related to at least one of its terms. A word is related with a term if and only if it is related with all of its literals.

Definition 2 (Semantic relations with the secret)**.**
Let $\mathcal{D}$ be a dictionary of words.
A word $w_i \in \mathcal{D}$ is related to a secret ξ if and only if it is related to at least one term of ξ.
A word $w_i \in \mathcal{D}$ and a term of ξ are related if and only if there is a relation between w_i and every non-negated literal, but no relation between w_i and any negated literal in that term.

3.3.2. Protocol Description. For the sake of readability, we refer to the genuinely interacting human and the server by $\mathcal{H}$ and $\mathcal{S}$, respectively.

Enrollment. $\mathcal{H}$ claims an identity, e.g. a user name, and either chooses a secret or is assigned one by $\mathcal{S}$. By doing so $\mathcal{S}$ ensures that only words in the knowledge base can comprise part of the secret. Subsequently $\mathcal{H}$ has to remember this secret. Then $\mathcal{S}$ chooses a distribution ϕ over the dictionary $\mathcal{D}$, consisting of all words within its knowledge base. Based on this distribution, $\mathcal{S}$ will draw the challenge words for this user in the future. Finally, $\mathcal{S}$ saves the triple of claimed identity, distribution and secret.

Authentication. The protocol requires two parameters and two variables controlled by the system. The first parameter $\mathtt{n}$ is a counter on the number of challenges used by the server for authentication. The second parameter $\mathtt{t}$ is a threshold on the required percentage of correct answers. This allows the legitimate user to make few errors in their responses, to be expected due to different people's idiosyncratic understandings of the world. The variables $\mathtt{a}$ and $\mathtt{c}$ count the number of correct answers and sent challenges, respectively.

Server $\mathcal{S}$ checks before any authentication whether the claimed identity is valid for authentication. Typical system security measures can be used here, e.g. the claimed identity must have been enrolled, must not be suspended and must not be suspected of being brute-forced.

1. $\mathcal{H}$ contacts $\mathcal{S}$ and informs it about his claimed identity.
2. $\mathcal{S}$ checks whether this identity is valid for authentication.
   If yes, $\mathcal{S}$ loads parameters $\mathtt{n}$ and $\mathtt{t}$, distribution ϕ, the user's secret, and initialises variables $\mathtt{a} = \mathtt{c} = 0$.
   a) $\mathcal{S}$ draws a challenge from the dictionary $\mathcal{D}$ according to the distribution ϕ and sends it to $\mathcal{H}$.
      $\mathcal{S}$ further sets $\mathtt{c} = \mathtt{c} + 1$.
   b) $\mathcal{H}$ responds whether the challenge is related to his secret or not.
   c) $\mathcal{S}$ validates the response based on its knowledge base.
      If and only if the answer is consistent to the knowledge base, $\mathcal{S}$ sets $\mathtt{a} = \mathtt{a} + 1$.
3. If $\mathtt{c}$ is smaller than the predefined counter $\mathtt{n}$, the protocol returns to step 2.a).
4. $\mathcal{S}$ checks whether $\mathtt{a}$ divided by $\mathtt{c}$ is greater than or equal to the predefined threshold $\mathtt{t}$.
   If this check is successful, $\mathcal{H}$ is accepted. Otherwise $\mathcal{H}$ is declined.

4. Security Analysis

The primary purpose of every authentication system is protecting access to restricted resources and services. Consequently, for any newly proposed mechanism, security is of high concern. The following three aspects are addressed:

1) Online trawling attacks;
2) Untrusted devices and shoulder surfers;
3) Authentication secret storage.

4.1. Online Trawling Attacks

The first serious concern in securing online authentication are guessing attacks. Usually servers block user accounts after a few unsuccessful log-in attempts in order to avert brute-force attacks. During an online trawling attack, an adversary makes only a few guesses for many user accounts. Those guesses usually follow heuristic patterns (e.g. password dictionaries) and are constrained by lock-out policies. The adversary's ambition is to corrupt at least a small percentage of random user accounts and its success chance is closely related to the schemes effective password space [56].

Our proposed protocol is based on multiple binary responses. The legitimate user must achieve a certain threshold of correct answers to be authenticated. A successful online trawling adversary would have to analogously guess the required number of correct responses.

Evenly distributed responses (i.e. the same number of "yes" and "no" answers) minimise this threat, which is easy to achieve. The distribution of responses (as given by the legitimate user) is directly related to the distribution of challenges posed by the system. ZeTA does not require a distinct challenge distribution. The distribution can, in fact, vary for each user if so desired. Therefore, without loss of generality, we can assume an even distribution of responses for all users. The protocol's vulnerability to online trawling attacks, or guessing attacks in general, is thus minimal.

4.2. Untrusted Devices and Shoulder Surfers

Untrusted channels could leak the credentials to an adversary. With this knowledge, the adversary could gain access to the user's account. This attack is thus a targeted version of the aforementioned online trawling attack.

With ZeTA we are concerned about the adversary's capabilities in terms of predicting the user's responses based on observed challenge-response pairs. In our analysis, we will first formalise the threat model and the subsequent impersonation attack. Then, we describe the probably approximately correct (PAC) learning model which is used to model the adversary's endeavour to learn from observed authentications. Thereafter, we introduce the Vapnik-Chervonenkis dimension which is used to measure the complexity of approximating an unknown function. Finally we present a security analysis based on an existing lower bound in the PAC model.

4.2.1. Threat Model and Impersonation Attack. The threat model, as introduced by Matsumoto & Imai [22], has the following properties:

- The user has access to a terminal and wants to prove his/her identity to a remote server.
- The adversary is a computer-assisted human and has access to every computational device, as well as the communication channels between those devices, but not to the server itself. This capability describes the context of untrusted devices.
- The adversary has visual access to the user, i.e. the adversary sees the human himself. This capability describes an adversarial shoulder surfer.

The adversary wants to impersonate a targeted user. The impersonation attack could run in two phases:

Phase 1: The adversary observes several authentication sessions of a particular targeted human. Observations can be internal, i.e. on the untrusted device, as well as external, i.e. as a shoulder surfer.

Phase 2: The adversary tries to use the observed credentials to convince the server and masquerade as the legitimate user.

4.2.2. Probably Approximately Correct Learning. The attempt to learn from observed challenge-response pairs is an instance of machine learning. The probably approximately correct (PAC) learning model [57] is a framework for mathematical analysis in machine learning.

In an instance of PAC learning, an algorithm has to produce a close approximation of an unknown binary function based on labelled examples. The unknown function is an element of a known class of functions and the labelled examples are drawn according to an unknown distribution. An algorithm is said to learn a class $\mathcal{C}$ of functions c: X $\rightarrow \{0, 1\}$ (where X denotes any set of arbitrary elements) if, with high probability, it can find a close approximation for any such function.

In the PAC learning model with classification noise, the label of each example is corrupted with error probability η.

Definition 3.
An algorithm $\mathcal{A}$ is said to learn a binary class $\mathcal{C}$ with noise rate η, accuracy parameter $\epsilon > 0$ and confidence parameter $\delta < 1$ if, given a random set of data points $\big(x, c(x)\big) \in (X, \{0, 1\})$ (sampled according to any distribution ϕ) from any function $c \in \mathcal{C}$, it outputs a hypothesis h: $X \rightarrow \{0, 1\}$ such that, with probability at least $1-\delta$

$$\Pr_{\phi}[h(x) \neq c(x)] < \epsilon$$

4.2.3. Vapnik-Chervonenkis Dimension. In determining the difficulty of approximation in the PAC model, the complexity of the considered class of functions $\mathcal{C}$ is an important element. The complexer this class, the more difficult an approximation. The Vapnik-Chervonenkis dimension [58] is a measurement of this complexity.

Any set X$'$ $\subseteq$ X is shattered by $\mathcal{C}$ if, for each of the $2^{|X'|}$ possible labellings of the points in X$'$, there exists some function in $\mathcal{C}$ consistent with that labeling. The Vapnik-Chervonenkis dimension of $\mathcal{C}$, denoted VC($\mathcal{C}$), is the maximal integer d such that there exists a set $X' \subseteq X$ of cardinality d that is shattered by $\mathcal{C}$.

4.2.4. Security. Theorem 1 by Aslam and Decatur [59, Theorem 6] provides a lower bound on the required sample size to learn any (non-trivial) class $\mathcal{C}$. It holds for all algorithms and independent of computational resources. The lower bound is parametrised by the desired quality of the approximation ϵ, the success probability of the algorithm δ, as well as the Vapnik-Chervonenkis dimension of $\mathcal{C}$. ϵ and δ depend on the algorithm, as described in the PAC model. The Vapnik-Chervonenkis dimension depends on the complexity of the class $\mathcal{C}$.

Theorem 1 ([59]).
PAC learning a function class $\mathcal{C}$ in the presence of classification noise η, and with respect to ϵ and δ, requires a sample of size

$$\Omega\left(\frac{VC(\mathcal{C})}{\epsilon(1-2\eta)^2} + \frac{log(1/\delta)}{\epsilon(1-2\eta)^2}\right).$$

Corollary 1.
Any algorithm which learns about a ZeTA secret based on the observations of corresponding challenge-response pairs is restricted by the lower bound of theorem 1.

Proof. We show that any knowledge about (and from) the ZeTA protocol, that is accessible to the adversary, is considered in theorem 1. Without loss of generality, we assume that the adversary observes successful authentications of the legitimate user only.

The possible knowledge of the adversary can be summarized as the following:

1) A copy of the server's knowledge base.
2) An estimation of a limit on the size of the secret.
3) A certain amount of available challenge-response pairs from observed authentications by the legitimate user.
4) Knowledge of the maximum possible error rate on the observed challenge-response pairs.

For any secret, the ZeTA protocol represents a binary function c: $X \rightarrow \{0,1\}$, whereby any $x \in X$ represents a challenge and $\{0,1\}$ represents the binary response. The set of all unique such functions is a binary class $\mathcal{C}$.

In the following, we will treat the ordered list of knowledge accessible to the adversary and describe how they are depicted in theorem 1:

1) Access to the server's knowledge base does help the adversary. Then, without loss of generality, the user's secret is the minimum necessary and sufficient information required to reconstruct any $c \in \mathcal{C}$. Thus, if the adversary has access to the knowledge base, the complexity of approximating any $c \in \mathcal{C}$ is reduced and equivalent to the complexity of approximating the corresponding secret.
2) The secrets in ZeTA are either in disjunctive normalized form (DNF) or can be rephrased as such. The set of all secrets in the ZeTA protocol is limited in its maximum possible size by the user. Therefore, the set of all possible user secrets $\mathcal{C}$ is the set of all l-term k-DNF (for some l and k).
It follows that the VC dimension of $\mathcal{C}$ can be defined by the VC dimension of the class of l-term k-DNF. This VC dimension's lower bound is the VC dimension of l-term monotone k-DNF, i.e. l-term k-DNF without negations. Littlestone stated such a lower bound:
Lemma 1 ([60]).
For $1 \leq k \leq n$ and $1 \leq l \leq \binom{n}{k}$, let $\mathcal{C}^*$ be the class of concepts expressible as l-term monotone k-DNF over domain $\{0,1\}^n$ and let m be any integer, $k \leq m \leq n$ such that $\binom{m}{k} \geq l$. Then $VC(\mathcal{C}^*) \geq k \cdot l\lfloor log_2 \frac{n}{m} \rfloor$.
3) Every observed authentication leaks a fixed number of labelled examples c(x) = $\{0,1\}$. At any given time, the amount of all previously leaked examples is the sample size.
4) If users are allowed to answer η percent of challenges wrong, the adversary has to expect an equal noise rate.

It follows that the adversary, independent of it's algorithm and computational resources, is limited by the above bound on the required sample size. □

If legitimate users were not allowed to make errors, the following theorem by Ehrenfeucht *et al.* [61] would be applicable instead of theorem 1:

Theorem 2 ([61]).
Let $\mathcal{C}$ be a non-trivial concept class. Then any (ϵ, δ)-learning algorithm $\mathcal{A}$ for $\mathcal{C}$ must use sample size

$$\Omega\left(\frac{1}{\epsilon}log\frac{1}{\delta} + \frac{VC(\mathcal{C})}{\epsilon}\right).$$

4.3. Authentication Secret Storage

There are different approaches to the problem of storing authentication secrets on the server. The challenge is to store the secret in such a way that it is usable for the purposes of authentication *and* remains resilient to internal and external threats, e.g. plain-text observation or brute-force dictionary attacks. The practical approach to address such a challenge, in many scenarios, is to store salted hashes rather than the authentication secrets themselves [62].

However, utilising salted hashes for the proposed protocol is unrealistic since it relies on the logical evaluation of arbitrary combinations of elements, drawn from a knowledge base, as well as elements known only to the user. Nevertheless, a realistic deployment of the proposed protocol needs to address the challenge of internal and external threats. Consequently, an alternative approach could be to utilise secret sharing or splitting to address the deployment challenge of secret storage.

Mayer and Volkamer propose such an approach [63], a (t, n)-threshold verification scheme that relies on Blakley's secret sharing [64] and key derivation functions to facilitate derivation of a common secret from all authorised subsets. Here n denotes the size of the authentication secret and

t denotes the size of the authorised subsets. The scheme can easily be adapted for ZeTA as the only conceptual requirement imposed by this (t, n)-threshold verification is that each authentication secret must be representable as a set of elements, which ZeTA secrets fulfil due to their sentential logic structure.

Therefore, ZeTA authentication can be considered as users evaluating and entering the logic of arbitrary combinations of elements drawn from the knowledge base and elements known only to users. An appropriate key derivation function (KDF) along with elements drawn from the knowledge base and responses are used to determine the shares. Then, as outlined in [63], a system of equations $Mx = y$ is solved for x and the corresponding server side stored s is compared to $s' = KDF(x_1)$.

4.4. Example Parameters

In Table 1 we compare the security provided by several ZeTA parameters with alphanumeric passwords and four digit PINs. To investigate the feasibility of ZeTA we present example parameters in Tables 2 and 3. We chose parameters such that several combinations emerged which we considered interesting from both security and usability perspectives.

4.4.1. Comparison of example parameters. We present a comparison of the security properties against guessing adversaries of several ZeTA parameters with passwords and PINs in Table 1: a listing of the bit strength of alphanumeric passwords in practice, as reported by Bonneau [65], and 4 digit PINs both in theory and in practice, as reported by Bonneau *et al.* [66], as well as several theoretical values for ZeTA as presented in Table 2.

Comparing values of bit strength from theory with those from practice doesn't provide optimal insight and one has to be careful with conclusions, as it has to be expected that user-chosen ZeTA secrets contain some kind of bias, too, and this will negatively influence the security. Nonetheless, as can be seen in Table 1, the security of ZeTA against guessing adversaries looks promising.

TABLE 1. SECURITY AGAINST GUESSING ADVERSARIES FOR ALPHANUMERIC PASSWORDS, 4 DIGIT PINS AND ZETA.

Authentication method	Security provided
Alphanumeric passwords in practice [65]	10 bit
4 digit PINs in practice [66]	12.9 bit
4 digit PINs in theory	13.29 bit
ZeTA with 25 challenges & 22 correct responses	13.64 bit
ZeTA with 14 challenges & 14 correct responses	14 bit
ZeTA with 21 challenges & 20 correct responses	16.54 bit
ZeTA with 25 challenges & 24 correct responses	20.3 bit

4.4.2. Parameters Regarding Guessing Adversaries. Table 2 depicts the probability of a false positive based on guessing. The data are valid under the assumption that correct responses are evenly distributed between 'yes' and 'no'. This is a realistic assumption, as argued in Section 4.1.

The table highlights the importance of determining how well people can cope with the ZeTA protocol. The fewer mistakes ZeTA needs to tolerate, the fewer challenges need to be used in order to achieve a required security level. Table 2 further demonstrates that a security level high enough for most settings can easily be achieved if the error rate is sufficiently small.

4.4.3. Parameters Regarding Untrusted Devices. Table 3 depicts the required sample size for an adversary that attempts to learn from observed authentications to raise its success rate. The data is derived according to theorem 1 and lemma 1.

It presents the minimum number of challenge-response pairs required by the adversary to improve its probability to guess correct responses for the sketched scenario. It contains three constant and two variable parameters:

For the constant parameters, we chose the knowledge base to consist of 65,536 words. We assume this to be reasonable and not to difficult to accomplish. This assumption is reasonable when considering the size of other knowledge bases available. The *Never-Ending Language Learner* (NELL) [67], for example, contains a knowledge base of more than 80 million *so-called* beliefs. It is partly constructed from human judgement, which was (mainly) collected by a specially purposed web-page, and mostly via the World Wide Web. NELL can not be directly used for ZeTA, unfortunately, since the authors admit that this ongoing research project contains many incorrect beliefs. The noise rate (i.e. percentage of mistakes made by the user) of 10% was chosen as a realistic but slightly pessimistic assumption. The probability $\delta = 0.25$ denotes that the adversary outputs a hypothesis of error rate lower than ϵ with probability at least $1 - \delta = 75\%$.

The error of approximation as first variable parameter denotes that a successful adversary would make fewer than ϵ percent mistakes on answering future challenges for the targeted user. The other variable parameter depicts three upper limits on the secret size – all monotonous (i.e. the secrets do not contain negations). Recall two important considerations:

- The VC dimension of general l-term k-DNF is unknown and can thus only be described by the inferior lower bound on l-term monotonous k-DNF. Nonetheless it is assumed that they are not equal and therefore the security against untrusted devices would considerably raise from the inclusion of negations.
- The secret sizes do not represent the actual size of a targeted user's secret, but rather what the adversary expects to be possible. If the adversary underestimates this size, his chances of success worsen considerably.

Especially when considering the second point, it becomes clear that these were moderate choices.

TABLE 2. PROBABILITY TO GET AUTHENTICATED BASED ON GUESSES WHEN 'YES' AND 'NO' RESPONSES ARE EQUALLY LIKELY.

		Number of challenges posed by ZeTA									
		14		18		21		25		30	
		%	$\mathcal{A}$	%	$\mathcal{A}$	%	$\mathcal{A}$	%	$\mathcal{A}$	%	$\mathcal{A}$
Number of errors tolerated by ZeTA	0	100	$1/16384$	100	3.81×10^{-6}	100	4.77×10^{-7}	100	2.98×10^{-8}	100	9.31×10^{-10}
	1			94.4	$1/13797$	95.2	$1/95325$	96	7.75×10^{-7}	96.7	2.89×10^{-8}
	2					90.5	$1/9039$	92	9.72×10^{-6}	93.3	4.34×10^{-7}
	3							88	$1/12777$	90	4.22×10^{-6}
	4									86.6	$1/33626$

$\mathcal{A}$ = authentication probability when responses are guessed % = percentage of correct answers required

TABLE 3. MINIMUM NUMBER OF CHALLENGE-RESPONSE PAIRS REQUIRED BY THE ADVERSARY TO REACH THE DESIRED ERROR RATE ϵ. (SIZE OF THE KNOWLEDGE BASE 2^{16}, NOISE RATE 10%, AND $\delta = 0.25$)

		Estimated upper bound on secret size		
		2-term monotone 2-DNF	3-term monotone 2-DNF	3-term monotone 3-DNF
Intended upper bound on error	$\epsilon = 0.25$	58	83	121
	$\epsilon = 0.1$	146	209	271

The values in the cells denote the lower bound of how many challenge-response pairs are required given the corresponding parameters. As can be seen in the data, the parameter ϵ (which is chosen by the adversary) heavily influences this bound. A value of ϵ worse than η is likely to cause many wrong responses and this decreases the likelihood of successful impersonation – especially if the user account is at risk of being suspended after few unsuccessful attempts.

There are three main ways to improve the resistance of ZeTA against untrusted devices and observations:

- Teaching/enabling users to cope with larger secrets requires the adversary to make assumptions towards larger upper limits on the secret size.
- Reducing the legitimate user's rate of mistakes requires the adversary to choose a small ϵ to maintain the likeliness to impersonate. A smaller ϵ leads to more required observations.
- When increasing the size of the knowledge base (even though it is known to the adversary) more challenges become available and this significantly improves the security properties.

The conclusion of how many observed authentications ZeTA can withstand depends on the interaction of many parameters, including the number of challenge-response pairs that can be observed at each authentication attempt. Improving security against guessing adversaries by asking more questions negatively influences the security against untrusted devices. Improving the security against untrusted devices by providing fewer examples per observed authentication decreases the security against guessing adversaries. On the other hand, increasing the size of the knowledge base improves the security against untrusted devices and reducing the amount of errors made by legitimate users positively influences the security against both attacks.

5. Pilot Usability Study

The primary usability concern with the proposed protocol is that individuals are not able to respond to issued challenges. The reality is that individuals may struggle to evaluate the logic of arbitrary combinations of elements known to them and those drawn from the knowledge-base.

Consequently, the focus of the usability studies was to assess the ability of users and not to replicate actual usage conditions. The adopted approach was to conduct a series of preliminary studies followed by a primary study, incorporating lessons learnt at each iteration. The main motivation was to ensure the concept of ZeTA was introduced to participants carefully and correctly. Hence, eleven distinct preliminary studies, successively refining instructions and examples. The results of the eleventh iteration with 20 participants, comprising the final pilot study, are reported here.

5.1. Apparatus

The apparatus for the preliminary studies comprised of secret-challenge sets and questionnaire.

5.1.1. Secret-Challenge Sets.

A set of related and unrelated word pairs was analytically determined, i.e. handpicked. The approach is sensitive to idiosyncrasies, potentially biasing the outcome of the preliminarily study. Nonetheless, the procedure is common in computer linguistic research (e.g. see human judgement experiment by Cramer [54]). The set of word pairs constituted the server-side knowledge base.

The client-side secrets consisted of two words, both potentially preceded by **NOT**, concatenated with either **AND** or **OR** such that, from the view point of sentential logic, the answers are assumed to be unambiguous.

5.1.2. Questionnaire.

The questionnaire consisted of four parts:

- **Instructions:** The instructions introduced the purpose of the experiment and explained the task.
- **Examples:** The examples were presented to foster understanding of the task. For example, they were given the pair *'Day OR Cactus'* with the attribute *'Cloudy'* and shown that the answer would be '*Yes*'. Eleven examples were given, five being affirmative and six non-relationships.
- **Challenges:** There was 24 secret pairs, accompanied by an attribute for each. Participants had to respond '*Yes*' or '*No*' for each. Secret words were concatenated with 'AND' six times, there were 6 'OR' secret pairs, 5 'AND NOT' and 6 'OR NOT' secret pairs, and one 'NOT AND NOT' secret pair. The secret-challenge sets were presented in random order on each sheet.
- **Demographics:** The end of the questionnaire contained standard demographic questions. The age was stratified in the following groups: 18-24, 25-34, 35-44, 45-54, 55-64 and 65+.

5.2. Interview Procedure

The procedure and instructions were refined by conducting eleven iterations of the pilot study. The lessons learnt from each iteration fed into the next, the expectation being that it improved the clarity of instructions and procedure. Ten sets of between 10 and 23 participants (188 participants total) helped us to refine our instructions and procedure.

Participants were approach and recruited from a busy public park. We identified ourselves as researchers from the local university. We asked individuals whether they would be willing to participate in a 3 to 5 minute study. Individuals were informed that the purpose of the study was to see whether people were able to confirm or deny relationships between words. The participants were neither paid nor otherwise compensated and were all fluent in the language of the survey.

Participants were handed a clipboard containing the questionnaire and requested to read the instructions and examples before answering questions. Participants were requested, once they had completed the questions, whether the instructions were clear, and how they could be improved. Participants were offered a verbal debriefing that explained the background and reasons for the study, once they had completed the questionnaire. The time taken to read the instructions, and to respond to all the challenges was recorded (in minutes).

5.3. Results

There were 20 participants in the final pilot, with 40% being male. The average participant was 24-35 years old. 70% were currently, or had previously been, students. 19 rated their understanding of the questionnaire language at mother tongue level. The participants gave the correct answers 93.125% of the time. Participants took, on average, 2.15 minutes to read the instructions and 4.2 minutes, on average, to complete the 24 challenge questions. Gender did not impact the correctness of the responses.

Error rates were as follows: AND=1.6%, OR=3.3%, AND NOT=11%, OR NOT=12.5%. From these numbers it becomes clear that the negations led to far more errors than the AND or OR conjunctions.

6. Primary Usability Study

The pilot study helped us to refine our instructions, and we learnt a number of lessons from it:

- It was clear that we should exclude secrets with negations, because participants experienced difficulties coping with these secrets.
- We decided to record how long it took for people to complete the questionnaire in minutes *and seconds* (the pilot study only recorded minutes) since recording only in minutes was not fine grained enough.
- The purpose of the pilot study was to refine the instructions but we realised that giving people 24 'secrets' was not really mirroring an actual ZeTA authentication. We realised that we ought to reduce the number of secrets, and introduce more than one challenge per secret, as ZeTA would.

This usability study would give us greater insight into how people handle ZeTA-type challenges with secret pairs concatenated only with AND or OR.

6.1. Apparatus

6.1.1. Secret-Challenge Pairs.

From the pilot study's secret-challenge pairs, we chose the five worst performing ones. Then, for each secret, we again analytically determined four new challenges. This time, the challenges were chosen such that three of the authors agreed on whether they were related or not. The procedure matches a human judgement task with a three person agreement. We had a total of 5 challenges for each secret word pair. We settled on the final set of secret-challenge sets shown in Table 4.

6.1.2. Questionnaire. The questionnaire consisted of three pages on two sheets.

- **Demographics:** We collected some basic demographics (but no identifying details) and the current time. We did not collect gender since no gender impact was detected in the pilot study. We also explained the purpose of the study, as we did for the pilot study.
- **Instructions:** We explained the task, and explained the concept of semantic similarity.
- **Examples:** Since the task was different from the pilot study we provided only two examples of secret pairs, one AND, and one OR. Each was shown with a number of attributes which are semantically related, and a number that are not semantically related. The AND example was *'Nature AND Edible'*, and one of the semantically related attributes was *'Nut'*. The OR example was *'Politics OR Machine'*. An example of something that is not semantically related was *'Desert'*.
- **The Questionnaire:** Five secrets were presented, each with 5 different attribute challenges. The order of both the challenges and the attributes was randomised.

TABLE 4. SECRET-CHALLENGE SETS

Secret Pair	Challenges				
Glass OR Gift	Wine (y)	Birthday (y)	Toothpick (n)	Window (y)	Tomato (n)
South Pole OR Desert	Penguin (y)	Craftsman (n)	Ice (y)	Sand (y)	Metropolis (n)
Cat OR Memorial	Statue (y)	Cup (n)	Plant (n)	Fur (y)	Pet (y)
Airplane AND Building	Ballpoint Pen (n)	Wedding (n)	Airport (y)	Window (y)	Security Area (y)
Cold AND Beverage	Mulled Wine (n)	Refreshment (y)	Cola (y)	Iced Tea (y)	Cappuccino (n)

6.2. Interview Procedure

We followed a procedure similar to the one described for the pilot study. Participants were, again, able to ask questions before the researcher thanked them and departed.

6.3. Results

42 participants completed our survey, aged 18-54. The mean completion time was 1 minute and 29 seconds (for 5 'secret pairs' each having 5 attribute challenges). The maximum time taken was 2 minutes and 22 seconds[1] and the fastest completion took 52 seconds. 83% of the participants were, or had previously been, students. 41 were mother tongue speakers of the language the questionnaire was in.

All participants answered 25 challenges. Overall, participants answered the challenges correctly 95.24% of the time. The percentages of correct answers for each of the sets is shown in Figure 5. Five participants made one error, one made 4 errors, two made 3 errors.

TABLE 5. SUCCESS RATES FOR SECRET-CHALLENGE SETS

Secret Pair	Challenges				
Glass OR Gift	Wine (90%)	Birthday (95%)	Toothpick (100%)	Window (95%)	Tomato (98%)
South Pole OR Desert	Penguin (95%)	Craftsman (100%)	Ice (98%)	Sand (98%)	Metropolis (100%)
Cat OR Memorial	Statue (100%)	Cup (98%)	Plant (100%)	Fur (100%)	Pet (98%)
Airplane AND Building	Ballpoint Pen (100%)	Wedding (100%)	Airport (100%)	Window (67%)	Security Area (90%)
Cold AND Beverage	Mulled Wine (93%)	Refreshment (81%)	Cola (95%)	Ice Tea (100%)	Cappuccino (90%)

The worst performing was the pair: secret: *'Airplane AND Building'*, challenge: *'Window'*. Only 67% considered these two concepts to be semantically related.

At least one participant did not notice the switch from OR to AND between the two secrets, answering the AND questions as if they were OR questions. The participant was not excluded from the analysis since these kinds of problems are important to include in usability studies.

Two participants were not sure whether penguins lived at the South or North Pole, which caused some confusion for that question but overall people seemed to perform well for that challenge.

7. Discussion

We devised the ZeTA authentication approach because we felt it provided a way to authenticate people over untrusted channels, using untrusted devices, observed by untrustworthy people. The literature suggested that people would find it relatively easy to respond to challenges composed of semantically-related concepts, but we could not be sure, given the need to have multiple secrets with logical operators. We thus ran usability tests, first a pilot then a final test.

The pilot study revealed that people could manage ZeTA challenges if the requirements were explained to them properly. Even so, we noted that they performed most poorly with secrets of the type 'a AND NOT b' and 'a OR NOT b'. This is to be expected, since the human brain cannot store negations [68]. It means that someone needs to search all semantic relationships with a given secret in order to ensure that there is no match for the given challenge. This is cognitively demanding, time consuming and error prone.

We then improved the instructions in preparation for the primary usability study. To test the efficacy of the

[1] This was the only participant who rated his understanding of the language to be only close to native, but made no errors.

refined instructions we chose the weakest performing secrets from the first study, removing negations since it was clear that they imposed too much of a load. We ended up with three 'OR' secrets and two 'AND' secrets. We posed five different challenges for each. Now that we had removed negation we noted that the five worst-performing challenges were all related to secrets of the type 'a AND b'. With two kinds of secrets, the AND combinations were clearly more cognitively demanding than the OR secrets, since they essentially constitute twice the effort required of the OR secrets.

The poor performance of 'AND' secrets could be an artifact of the way we tested the protocol, with the protocol as a stand-alone challenge with participants who did not really care about providing correct answers. It might be different when access to a desired resource is in the balance. On the other hand, participants might have been confused about the task itself, leading to a failure to identify the semantic relationship with both secrets, since there was a switch between 'OR' and 'AND'. They could also have become distracted during completion of a particular challenge.

We now examine the worst performing combinations to see whether any patterns emerge.

- **Airplane AND Building.** The most poorly performing challenges were:
  - *Security Area* — Only buildings of specific types have security areas. For example a home is a building, but it does not really have a security area. Perhaps this confused people.
  - *Window* — Since buildings obviously have windows the problem might have resulted from people not considering the non-opening apertures in airplanes to be windows in the strict sense of the word. They may consider them to be more akin to portholes than windows. On the other hand, this is also the only challenge that tests an attribute of both pairs, as opposed to categories, instances or subparts. We are somewhat at a loss to explain the difficulties people had with this challenge.
- **Cold AND Beverage.** In English the fact that this was the only secret that involved an adjective and a noun might have caused errors. The study was conducted in German though and there was consequently no ambiguity regarding the interpretation of the word *cold* as a noun: the German adjective *kalt* can be clearly distinguished from the noun *Erkältung*. The most poorly performing challenges for this secret were:
  - *Cappuccino* — 10% of participants thought that the answer was positive. We then realised that in many places Cappuccinos are actually served cold. This finding demonstrates the ambiguity of semantic relationships, and emphasises the need for some level of tolerance for disagreement in the protocol.
  - *Refreshment* — One possible explanation might be that people acknowledge that a refreshment can be a beverage, but that no refreshment is needed in a state of cold (be it that they are cold or their surrounding is cold e.g. in winter). Since only a combination of both attributes satisfies the secret such an interpretation would lead to a negative answer. Another explanation might be that people simply consider cold drinks to be less refreshing than hot drinks.

What we learned from our evaluation was that it is non-trivial to choose secret pairs and challenges. Three authors agreed on the secrets and challenges we chose but, even so, some of the semantic relationships were not confirmed by our participants. This is clearly something that needs to be addressed if ZeTA is ever to be a viable authentication protocol. We were well aware, embarking on this research, that semantic relationships can be ambiguous, idiosyncratic and fluid. Hence we have to tolerate a certain number of "incorrect" answers from a legitimate user. We will have to implement ZeTA and test it in order to find out where the sweet spot is with respect to such tolerance. We also need to discover how tolerance levels ought to be mapped to the value of the asset being protected by ZeTA.

There is another sweet spot to be identified: where usability and security are maximised. On either side of this sweet spot either usability or security are sacrificed in order to bolster the other. One can maximise ZeTA's security by requiring people to remember 5 secrets, for example, and then answering 20 'AND' challenges in order to authenticate. The unusability of this scheme would render it useless since people are likely to refuse to use it. On the other hand, reducing the number of secrets and challenges would improve usability but at the expense of security, which would render it useless as an access control mechanism. We need to find the spot at which we can maximise these two essential characteristics.

This kind of authentication is undeniably time consuming, far more so than entering a password. We argued earlier that people might be willing to tolerate such inconvenience if they are clearly uncomfortable with the trustworthiness of their current environment. We merely surmise in this respect and such intuition might be flawed. We would have to test ZeTA in the field to find out whether this is the case or not.

We do feel, however, that ZeTA demonstrates promise. We are keen to carry out further research to see whether such promise delivers in more stringent evaluations.

8. Limitations

A number of limitations must be acknowledged. Similar to text passwords, the protocol is not resistant to humans deliberately divulging their secrets to other people, but ZeTA does not claim to address this flaw which is inherent to nearly all knowledge-based authentication systems.

The usability evaluation, and the pilot usability study, is undeniably not a concrete implementation of the protocol: a ZeTA user would only have one secret per account and successful authentication would provide access to a desired resource. These limitations seem acceptable at this stage because we sought only to test the validity of this protocol by broadening the scope. The switching between secrets makes the task more challenging and thus is not expected to whitewash the results.

The combination of active adversaries and untrusted devices was not specifically replicated in this study. This will be addressed in future work.

9. Conclusion and Future Work

In this paper we reconsider the arguments of Coskun and Herley [15] by suggesting that, instead of challenging people to either conduct mathematical operations or remember long secrets, we challenge them to confirm, or deny, semantically related attributes of their secrets. We presented a security analysis and carried out a usability proof-of-concept showing that people can indeed handle this kind of challenge.

We do not claim that we have a ready-made solution for untrusted environments. What we do believe is that our findings are interesting and suggest that ZeTA might hold potential. We still clearly have to carry out further even more realistic studies, where ZeTA controls access to something of value, to see whether it finds favour with end-users.

As ZeTA, in it's current form, clearly has limitations, future work could evaluate it's applicability to specific areas such as fallback authentication, e.g. recovering access to a resource after the password is lost, and single sign on services, e.g. online password managers. It furthermore could be fruitful to investigate whether users could be enabled to judge the trustworthiness of their environment and either conduct the ZeTA protocol or enter the ZeTA secret 'in the clear' based on their judgement, e.g. given the ZeTA secret 'Red OR Bike' the user could decide to enter the password 'RedBike'.

Investigations into how errors could be minimised could strengthen the security. We envisage three main directions:

- Investigating whether certain kinds of semantic relationships cause more difficulties than others.
- Coming up with a rule for choosing secrets, and sets of secrets. For example, should they be related, as Cold and Beverage were, or should they be completely unrelated, as South Pole and Desert were?
- The user's task in the ZeTA protocol is somewhat abstract. One particular concern is related to how people could be primed on a similar threshold of when a word pair is related and when not. As our studies in Sections 5 and 6 show, people seem to be able to cope with this easily once they grasp the principle. More investigation into variations of instructions and explanations could lead to further improvements.

References

[1] S. Mizuno, K. Yamada, and K. Takahashi, "Authentication using multiple communication channels," in *Proceedings of the 2005 workshop on Digital identity management*. ACM, 2005, pp. 54–62.

[2] M. Al Fairuz and K. Renaud, "Multi-channel, multi-level authentication for more secure ebanking." in *Information Security South Africa (ISSA)*, 2-4 August, Johannesburg, 2010.

[3] N. Saxena, J.-E. Ekberg, K. Kostiainen, and N. Asokan, "Secure device pairing based on a visual channel: Design and usability study," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 1, pp. 28–38, 2011.

[4] E. Waks, K. Inoue, C. Santori, D. Fattal, J. Vuckovic, G. S. Solomon, and Y. Yamamoto, "Secure communication: Quantum cryptography with a photon turnstile," *Nature*, vol. 420, no. 6917, pp. 762–762, 2002.

[5] G. Thomas and R. A. Botha, "Secure mobile device use in healthcare guidance from hipaa and iso17799," *Information Systems Management*, vol. 24, no. 4, pp. 333–342, 2007.

[6] D. Damopoulos, G. Kambourakis, and S. Gritzalis, "isam: an iphone stealth airborne malware," in *Future Challenges in Security and Privacy for Academia and Industry*. Springer, 2011, pp. 17–28.

[7] K. Renaud, P. Mayer, M. Volkamer, and J. Maguire, "Are graphical authentication mechanisms as strong as passwords?" in *Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on*. IEEE, 2013, pp. 837–844.

[8] K. Renaud, M. Volkamer, and J. Maguire, "ACCESS: Describing and contrasting authentication mechanisms," in *Proceedings of the Second International Conference on Human Aspects of Information Security, Privacy, and Trust-Volume 8533*. Springer-Verlag New York, Inc., 2014, pp. 183–194.

[9] K. Renaud, "Guidelines for designing graphical authentication mechanism interfaces," *International Journal of Information and Computer Security*, vol. 3, no. 1, pp. 60–85, 2009.

[10] D. Aspinall and M. Just, ""Give Me Letters 2, 3 and 6!": Partial Password Implementations and Attacks," in *Financial Cryptography and Data Security*. Springer, 2013, pp. 126–143.

[11] R. A. McCarthy, *Semantic knowledge and semantic representations*. Psychology Press, 1995, vol. 3, no. 3-4.

[12] A. Martin and P. Fedio, "Word production and comprehension in alzheimer's disease: The breakdown of semantic knowledge," *Brain and language*, vol. 19, no. 1, pp. 124–141, 1983.

[13] A. C. Lee, K. S. Graham, J. S. Simons, J. R. Hodges, A. M. Owen, and K. Patterson, "Regional brain activations differ for semantic features but not categories," *Neuroreport*, vol. 13, no. 12, pp. 1497–1501, 2002.

[14] E. Du Plessis, "Recognition versus recall." *Journal of Advertising Research*, vol. 34, no. 3, pp. 75–91, 1994.

[15] B. Coskun and C. Herley, "Can "something you know" be saved?" in *Information Security. 11th International Conference, ISC*, T.-C. Wu, C.-L. Lei, V. Rijmen, and D.-T. Lee, Eds. Taipei, Taiwan: Springer, 2008, pp. 421–440.

[16] F. J. Corbató, "On building systems that will fail," in *ACM Turing award lectures*. ACM, 2007, p. 1990.

[17] L. Lamport, "Password authentication with insecure communication," *Communications of the ACM*, vol. 24, no. 11, pp. 770–772, 1981.

[18] T. Simonite, "Watch the ATM Hacker At Work," 2013, http://www.technologyreview.com/view/517621/watch-the-atm-hacker-at-work/.

[19] Krebson Security, "Spike in ATM Skimming in Mexico?" 2015, http://krebsonsecurity.com/2015/07/spike-in-atm-skimming-in-mexico/ Date Accessed: 6 August 2015.

[20] N. Haller, C. Metz, P. Nesser, and M. Straw, "A one-time password system," 1998, iNTERNET STANDARD. [Online]. Available: http://www.rfc-editor.org/info/rfc2289

[21] N. J. Hopper and M. Blum, "Secure human identification protocols," in *Advances in cryptology–ASIACRYPT 2001*. Springer, 2001, pp. 52–66.

[22] T. Matsumoto and H. Imai, "Human identification through insecure channel," in *Advances in Cryptology–EUROCRYPT'91*. Springer, 1991, pp. 409–421.

[23] N. J. Hopper and M. Blum, "A secure human-computer authentication scheme," Carnegie-Mellon University, Tech. Rep., 2000.

[24] H. J. Asghar, "Design and analysis of human identification protocols," Ph.D. dissertation, Macquarie University, 2012.

[25] M. Mannan and P. C. van Oorschot, "Using a personal device to strengthen password authentication from an untrusted computer," in *Financial Cryptography and Data Security*. Springer, 2007, pp. 88–103.

[26] J. Thorpe, P. C. van Oorschot, and A. Somayaji, "Pass-thoughts: authenticating with our minds," in *Proceedings of the 2005 workshop on New security paradigms*. ACM, 2005, pp. 45–56.

[27] J. M. McCune, A. Perrig, and M. K. Reiter, "Seeing-is-believing: Using camera phones for human-verifiable authentication," in *Security and privacy, 2005 IEEE symposium on*. IEEE, 2005, pp. 110–124.

[28] V. Roth, K. Richter, and R. Freidinger, "A PIN-entry method resilient against shoulder surfing," in *Proceedings of the 11th ACM conference on Computer and communications security*. ACM, 2004, pp. 236–245.

[29] C. S. G. Khoo and J. C. Na, "Semantic relations in information science," *Annual review of information science and technology*, vol. 40, pp. 157–228, 2006.

[30] R. Chaffin and D. J. Herrmann, "The similarity and diversity of semantic relations," *Memory & Cognition*, vol. 12, no. 2, pp. 134–141, 1984.

[31] D. J. Herrmann and D. Raybeck, "Similarities and differences in meaning in six cultures," *Journal of Cross-Cultural Psychology*, vol. 12, no. 2, pp. 194–206, 1981.

[32] M. Hudon, "Relationships in multilingual thesauri," in *Relationships in the Organization of Knowledge*. Springer, 2001, vol. 2, pp. 67–80.

[33] A. K. Romney, C. C. Moore, and C. D. Rusch, "Cultural universals: Measuring the semantic structure of emotion terms in english and japanese," *Proceedings of the National Academy of Sciences*, vol. 94, no. 10, pp. 5489–5494, 1997.

[34] D. Raybeck and D. Herrmann, "A cross-cultural examination of semantic relations," *Journal of Cross-Cultural Psychology*, vol. 21, no. 4, pp. 452–473, 1990.

[35] E. Tulving, "Episodic and semantic memory 1," *Organization of Memory. London: Academic*, vol. 381, no. e402, 1972.

[36] W. Kintsch, *The Representation of Meaning in Memory (PLE: Memory)*. Psychology Press, 2014.

[37] K. McRae and M. Jones, "Semantic memory," *The Oxford handbook of cognitive psychology*, pp. 206–219, 2013.

[38] H. Rubenstein and J. B. Goodenough, "Contextual correlates of synonymy," *Communications of the ACM*, vol. 8, no. 10, pp. 627–633, 1965.

[39] G. A. Miller and W. G. Charles, "Contextual correlates of semantic similarity," *Language and cognitive processes*, vol. 6, no. 1, pp. 1–28, 1991.

[40] E. Bruni, N.-K. Tran, and M. Baroni, "Multimodal distributional semantics," *J. Artif. Intell. Res.(JAIR)*, vol. 49, pp. 1–47, 2014.

[41] C. Lofi, "Just ask a human?–controlling quality in relational similarity and analogy processing using the crowd." in *BTW Workshops*, 2013, pp. 197–210.

[42] G. A. Miller, "Wordnet: a lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[43] C. Fellbaum, *WordNet: An Electronic Lexical Database*. MIT Press, 1998.

[44] M. Coltheart, "The MRC psycholinguistic database," *The Quarterly Journal of Experimental Psychology*, vol. 33, no. 4, pp. 497–505, 1981.

[45] M. Wilson, "MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00," *Behavior Research Methods, Instruments, & Computers*, vol. 20, no. 1, pp. 6–10, 1988.

[46] D. L. Nelson, C. L. McEvoy, and T. A. Schreiber, "The University of South Florida word association, rhyme, and word fragment norms." http://web.usf.edu/FreeAssociation/, 1998, accessed: 2015-06-29.

[47] M. Strube and S. P. Ponzetto, "WikiRelate! Computing semantic relatedness using Wikipedia," in *AAAI*, vol. 6, 2006, pp. 1419–1424.

[48] E. Gabrilovich and S. Markovitch, "Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis." in *IJCAI*, vol. 7, 2007, pp. 1606–1611.

[49] T. Zesch, C. Müller, and I. Gurevych, "Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary." in *LREC*, vol. 8, 2008, pp. 1646–1652.

[50] K. Radinsky, E. Agichtein, E. Gabrilovich, and S. Markovitch, "A word at a time: computing word relatedness using temporal semantic analysis," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011, pp. 337–346.

[51] E. H. Huang, R. Socher, C. D. Manning, and A. Y. Ng, "Improving word representations via global context and multiple word prototypes," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics, 2012, pp. 873–882.

[52] M.-T. Luong, R. Socher, and C. D. Manning, "Better word representations with recursive neural networks for morphology," *CoNLL-2013*, vol. 104, 2013.

[53] W. Shen, J. Wang, P. Luo, and M. Wang, "Linden: linking named entities with knowledge base via semantic knowledge," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 449–458.

[54] I. Cramer, "How well do semantic relatedness measures perform?: a meta-study," in *Proceedings of the 2008 Conference on Semantics in Text Processing*. Association for Computational Linguistics, 2008, pp. 59–70.

[55] J. Boyd-Graber, C. Fellbaum, D. Osherson, and R. Schapire, "Adding dense, weighted connections to WordNet," in *Proceedings of the third international WordNet conference*, 2006, pp. 29–36.

[56] P. Mayer, M. Volkamer, and M. Kauer, "Authentication schemes - comparison and effective password spaces," in *Information Systems Security*. Springer, 2014, pp. 204–225.

[57] L. G. Valiant, "A theory of the learnable," *Communications of the ACM*, vol. 27, no. 11, pp. 1134–1142, 1984.

[58] V. N. Vapnik and A. Y. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory of Probability & Its Applications*, vol. 16, no. 2, pp. 264–280, 1971.

[59] J. A. Aslam and S. E. Decatur, "On the sample complexity of noise-tolerant learning," *Information Processing Letters*, vol. 57, no. 4, pp. 189–195, 1996.

[60] N. Littlestone, "Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm," *Machine learning*, vol. 2, no. 4, pp. 285–318, 1988.

[61] A. Ehrenfeucht, D. Haussler, M. Kearns, and L. Valiant, "A general lower bound on the number of examples needed for learning," *Information and Computation*, vol. 82, no. 3, pp. 247–261, 1989.

[62] D. Florêncio, C. Herley, and P. C. Van Oorschot, "An administrator's guide to internet password research," in *USENIX LISA*, 2014.

[63] P. Mayer and M. Volkamer, "Secure and efficient key derivation in portfolio authentication schemes using blakley secret sharing," in *Proceedings of the 31st Annual Computer Security Applications Conference*. ACM, 2015, pp. 431–440.

[64] G. R. Blakley, "Safeguarding cryptographic keys," in *Proceedings of the national computer conference*, vol. 48, 1979, pp. 313–317.

[65] J. Bonneau, “The science of guessing: analyzing an anonymized corpus of 70 million passwords,” in *Security and Privacy (SP), 2012 IEEE Symposium on*. IEEE, 2012, pp. 538–552.

[66] J. Bonneau, S. Preibusch, and R. Anderson, “A birthday present every eleven wallets? The security of customer-chosen banking PINs,” in *Financial Cryptography and Data Security*. Springer, 2012, pp. 25–40.

[67] T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios, A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, and J. Welling, “Never-ending learning,” in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI-15)*, 2015.

[68] P. N. Johnson-Laird, “Mental models and deduction,” *Trends in cognitive sciences*, vol. 5, no. 10, pp. 434–442, 2001.