WebID + Biometrics with Permuted Disposable Features

Takiva Richardson, Joseph Shelton
Virginia State University
Petersburg, Virginia, USA
tric4112@students.vsu.edu, jshelton@vsu.edu

Yasmin Eady, Kofi Kyei, Albert Esterline
North Carolina Agricultural and Technical State University
Greensboro, North Carolina, USA
yeady@aggies.ncat.edu, kkyei@aggies.ncat.edu, esterlin@ncat.edu

ABSTRACT

For networked communications, cyber security and authentication are critical components. This work deals with the issue of security and authentication as it relates to features of the Semantic Web . The phrase "Semantic Web" alludes to the World Wide Web Consortium's idea of standards to make internet data machine-readable and reusable. WebID, for example, is a technique for managing profile data connected with people and services at self-defined locations. While the WebID protocol alone allows users more control over their connections to online services, biometric authentication is an additional process that can add security and convenience for individuals. The WebID protocol with biometric authentication allows more control for the individual user, but networked connections are still vulnerable to replay attacks (an attack in which the network is compromised and authenticating information is captured and replayed to allow unauthorized access).

Replay attacks on a biometrics authentication system are particularly damaging since biometrics are more difficult to change than passwords. Prior work has been done to develop unique and accurate representations of one's biometric, such that if a bad actor captures any authenticating biometric data, it will not be useful in a replay attack as the system uses a new representation of biometrics after each access attempt. In this paper, we suggest extending previous work to increase the number of unique and reliable data representations in conjunction with WebID identifiers. This paper will provide an overview of the system that is currently under development, as well as additional WebID components.

CCS CONCEPTS

• **Security and privacy** → **Web protocol security**; **Web application security**; **Biometrics**; • **Information systems** → World Wide Web.

KEYWORDS

WebID, Biometrics, Semantic Web, Authentication, Cyber Security

ACMSE 2022, April 18–20, 2022, Virtual Event, USA

ACM ISBN 978-1-4503-8697-5/22/04…$15.00
https://doi.org/10.1145/3476883.3524050

ACM Reference Format:
Takiva Richardson, Joseph Shelton, Yasmin Eady, Kof Kyei, and Albert Esterline. 2022. WebID + Biometrics with Permuted Disposable Features. In *2022 ACM Southeast Conference (ACMSE 2022), April 18–20, 2022, Virtual Event, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3476883.3524050

1 INTRODUCTION

Since the onset of a networked digital landscape, the standard mode for authentication access for users has been through username and password. By using this method to gain access, users must create an account for all online services that they wish to use. There are two types of approaches to registration: casual, such as creating an email account, and formal, such as registering into a workplace space. This approach of registration can cause a variety of problems. Average users have about 25 different passwords to protect their accounts [8]. It is challenging to remember so many passwords for different online services; writing or recording passwords would put the user at risk of their password being stolen. Also, the user's username and passwords can be intercepted during transmission, which is another problem. Attempts can be made to guess a user's login and password using brute force methods. Using a single password for many accounts is an alternative to multiple passwords, but this might represent a serious danger if one password is exposed.

A solution for this issue is the use of WebIDs [24]. With the WedID protocol, it is possible to eliminate the need to remember usernames and passwords by storing all access information on the user's server. This approach will allow access by validating the legitimacy of the user. The WebID specifications are a set of editor's drafts for standardizing identity, identification, and authentication on HTTP-based networks. Solid OIDC, WebID-TLS, and WebID-TLS+Delegation are WebID-based protocols that provide a new way to log into internet services. This approach is a good idea but can lead to another issue. With WebIDs, whoever has control of the server associated with the user has access to all of the owner's accounts and information. As a result, if a malicious actor gains unauthorized access to an owner's device (such as their smart phone), all of the owner's personal information may be viewed and stolen as access can be established through that device. Our solution to this issue is a dual-factor strategy using WebIDs along with biometrics. When these two are used together, security is greatly enhanced since only those with a proven biometric match are allowed to obtain a WebID certificate and thus access the owner's server.

The practice of accepting a user's bodily traits to identify them is known as biometrics [11]. Knowledge-based, token-based, and

biometrics-based authentication are the three methods of authentication. The user uses a PIN or a password to get access to knowledge-based authentication, which is based on the user's knowledge. Token-based authentication verifies the user by using some type of physical identification. Biometrics-based authentication is the final type of authentication. Biometrics is the process of identifying, validating, or recognizing a living individual based on their physical or behavioral features, as previously stated. Physiological biometrics can involve scanning a user's face, iris, or fingerprint, which should be unique to each individual in most situations. Biometrics are practical and generally secure in today's digital world. For example, most mobile phones have a face recognition or even a fingerprint option built into them, and it is difficult to fake a face or fingerprints. For biometric verification , feature extraction is vital for representing a biometric in a manner that can be compared. The work done in this utilizes the Local Binary Patterns (LBP) algorithm [16], which extracts features in the form of multiple histograms. Each histogram correlates to a sub region of some biometric image, and the LBP technique typically segments an entire image into a grid-like structure of subregions for extraction. While LBP is effective, the Genetic and Evolutionary Features (GEFE) technique [21] uses an optimization technique known as evolutionary computations to evolve the optimal locations to extract features from, allowing for overlapping subregions that may not extract from the entire image. Prior work has shown the GEFE approach outperforming the LBP approach in terms of recognition accuracy [21]. This work incorporates WebID with GEFE for added security for authentication as well as a mitigation strategy for biometrics-based replay attacks.

A biometrics-based replay attack occurs when an attacker listens to the communication between some scanning device (camera, fingerprint reader, etc.) and the rest of the system. The attacker will copy packets of data that contain authentication information and resend it to the system to authenticate without using the scanning device. While there are a number of strategies to protect data while in transit through some network, there is no guaranteed effective scheme to fully protect data. GEFE has been used in the past [19] to mitigate replay attacks on a biometric system by creating a number of unique representations of a biometric sample. If data is compromised, it will not have an impact as that biometric representation was used once and will not be used again.

However, there is a limitation in the number of unique representations of biometrics created by GEFE. Unlike GEFE, which uses the same order of histograms to represent a feature vector, we are presenting a scheme that will extend the number of unique representations of one's biometric sample to avoid successful replay attacks [20]. The benefit of our disposable, permuted feature vector approach is that each access attempt of an individual will choose a disposable feature extractor as well as some unique, permuted feature vector from the selected feature extractor. This will enable the system to function as a one-time password authentication scheme using biometrics; the system will be referred to as WebID with the Permuted Disposable Feature Vectors (PDFV) method.

The WebID Protocol relies on a de-centralized approach for stored information. Unlike medium-to-large sized businesses that store its customer's private information on a set of dedicated servers, the WebID protocol allows users to have their own server in which to store authenticating information. While this makes an individual server less of a target than a set of company servers that hold the majority of its customer's information, individual servers may not have the protections of IT management that companies can afford to have. If an individual is targeted, repeated replay attacks are one form of attack that may be executed. The PDFV method will provide multiple one-time representations of an individual for added security to mitigate a successful replay attack.

This work describes an authentication system that inherits the benefits of WebID along with the security and convenience of biometrics for recognition. WebID with the PDFV method in particular will be robust against biometrics-based replay attacks. In Section 2, an overview of the relevant technologies that are used in this proposed system. Section 3 provides a summary of related research efforts and prior work specifically incorporating WebID with Biometrics. Section 4 provides a description of the key components of the proposed WebID system as well as a use case of the proposed system. A discussion of the impacts of this work is provided in Section 5, and the conclusion and future work are presented in Section 6.

2 BACKGROUND

The two major components of WebID with the PDFV method are the WebID protocol as well as biometric recognition for authentication purposes. This section provides an overview of each of these components. In particular, this section provides an overview of the background of the prior technologies that go into the area of WebID. Additionally, this section describes the Local Binary Patterns approach and the Genetic and Evolutionary Feature Extraction technique.

2.1 WebID

WebID is a URI-based protocol of uniquely identifying a person, firm, organization, or other agent. Dan Brickely and Tim Berners-Lee created the phrase "WebID" in 2000 [23]. WebID is built on the architecture of the Semantic Web. The Semantic Web is a collaborative movement led by the World Wide Web Consortium, an international standards body (W3C) [18]. The Semantic Web aims to transform the current web, which is dominated by unstructured and semi-structured documents, into a "web of data" that computers can read directly. The Semantic Web stack is based on the World Wide Web Consortium's Resource Description Framework (RDF). In the semantic web, RDF is a W3C recommendation that provides a data model for annotations. An RDF statement is a subject-predicate-object triple.The subject identifies the resources, whereas the predicate specifies the subject's characteristics or explains the subject's connection with the object. Finally, the object is the value of the resource's predicate to which the subject refers. Users can annotate online resources with named attributes using RDF. These named attributes can take the form of URIrefs to online sites or literals. RDF-annotated resources are identified by uniform resource identifiers (URIs). A URI reference (URIref) is a URI with an optional fragment identifier at the end.

While URIs were initially meant to refer to just documents (as URLs), they are now often used to refer to either logical or physical resources, such as abstract concepts or actual things. CoolURIs [2]

refer to a resource, but subsequently dereference to a page describing that resource through redirection or fragment identifiers. A CoolURI [2] is a URI that uses content-negotiation and redirects or or uses fragment identifiers to dereference to the relevant document that it indicates. The service determines what type of document to serve, which is usually RDF for machine agents and HTML for human agents. The usefulness of WebID is to identify a person, organization, group, or device on the Web. Certificate authorities necessarily require a type of authentication that is based on centralized systems. That is, a user ought to have multiple accounts and identifiers for each service they use. A new registration is required for each service, which can be time-consuming for both the user and the service [6]. By using WebID protocol, WebID refers to a user's WebID profile, which comprises structured data in RDF based on the FOAF ontology (Friend Of A Friend) [4]. FOAF is a semantic web RDF vocabulary for expressing social networks. The profile primarily consists of a FOAF graph, with some triples connecting the subject to their friends via foaf:knows relationships, and others providing subject attributes such as their name.

Figure 1 shows how the WebID Protocol authenticates a person. A private key is stored in a certificate, which is installed on a user's web browser. A public key is also closely linked to the WebID component on a person's server, which is commonly defined in a FOAF file. The certificate includes a URI that points to the FOAF file, which includes the WebID and public key. A WebID and a WebID profile must work together. An RDF document that describes an individual is called a WebID profile. Figure 1 illustrates a use case of the WebID protocol involving two individuals, Romeo and Juliet. Romeo is made aware that Juliet has sent a protected message. Romeo searches Juliet's public FOAF profile for the whereabouts of the message. When Romeo's client tries to dereference Juliet's server to retrieve the message, Romeo is asked for his certificate. Juliet's server will then dereference Romeo's FOAF profile to see if the modulus and exponent from Romeo's certificate match those from Romeo's FOAF profile. After that, Juliet's server will check Juliet's friend graph to see if she can trust Romeo. Romeo will be able to read Juliet's message once this is completed.

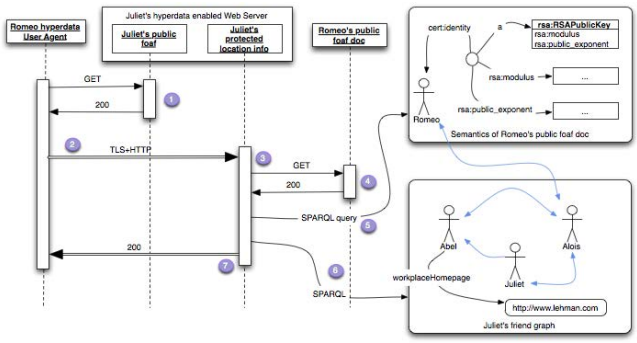

Figure 1: WebID Diagram

2.2 Biometrics

Biometric authentication is the process of using an individual's biometric sample to determine their identity. This can be used in a system that has a scanner to capture biometric samples, a technique to extract features from the biometric sample to be compared, some data storage to store enrolled biometric samples of legitimate users, and some metric to compare features from different samples to determine whether a sample belongs to an individual. The feature extraction technique is vital for accurate representation of the individual, and the technique used in this work is the Local Binary Patterns technique, a simplistic yet robust approach.

The Local Binary Patterns algorithm [16] is a texture-based feature extraction approach that extracts features from a portion of an image. The portion of the image depends on the user, but the portion must be rectangular and must be minimum 3 by 3 pixels. Each portion has a histogram associated with it that consists of the frequency of specific binary strings that can be found in the portion of the image. The binary strings are created by comparing each pixel in the image portion to its immediate and connected surrounding pixels on all sides (meaning a bordering pixel in the image portion will not have a binary string associated with it). Using a 3x3 graph that focuses on a center pixel, each pixel in the image portion is compared to its neighboring pixels. Compared to all its surrounding pixels, the center pixel is used to assess whether the pixel intensities are higher or lower. When the number is equal to or greater than the center region, a 1 will represent the region. Otherwise, it is it will be a 0. Then the resulting 0's and 1's are grouped to construct a binary number between 0000000 and 11111111, which is translated to a decimal number between 0 and 255. For that portion of an image, each decimal number is applied to a sub histogram. Then all the subhistograms of image portions are concatenated in order to create a final feature vector that represents an individual's biometric data. The LBP technique for biometrics typically segments an image into even sized portions.

Genetic and Evolutionary Feature Extraction (GEFE) [21] is an approach that was created after questioning why the entire image is necessary for LBP based feature extraction, and why must an area of an image have features extracted only one time. The GEFE approach uses the optimization technique of genetic algorithms to optimize the locations and dimensions of portions in an image that the LBP algorithm will be applied on to extract features; the only limitations are that the portions cannot be larger than the image itself and there is a limited number of portions that can be created. This allows for overlap, which in turn allows for more weight on more detailed areas of an image. While the GEFE technique was initially created with the sole focus of finding the most salient areas to extract features from for accurate biometric identification, an unintended result was that, due to the non-deterministic nature of genetic algorithms, multiple feature extractors were generated. Each extractor creates different feature vectors that are more accurate for matching than the LBP technique alone. This unintended feature became the crux of a cyber security application for biometric authentication systems by allowing the use of a disposable feature extractor to represent a biometric sample for authentication similar to a one-time password, never to be used again [22]. While experiments showed the practicality of this, there is the limitation of running out of unique and accurate representations of biometrics.

There have been a number of articles that have touched on biometrics as it pertains to the semantic web. Others recognize the importance of a form of identification that is naturally more secure

than passwords in the proposed World Wide Web with machine readable data. Additionally, the proposed WebID with PDFV method is an extension of prior work, which is described in the following section.

3 RELATED WORK

3.1 Similar Work

In Dwivedi et al. [7], the authors introduce the concept of biometrics as a base for authentication systems for improved cyber security, particularly on the semantic web. The authors then proposed solutions to prevent attacks on biometric authentication systems. The authors have proposed Semantic Web Service (SWS) Policy and SWS Security Policy to support concepts of biometric authentication trust levels, biometrics trust for federation, and trust mapping within the Semantic Web services architecture. To protect privacy, BioHashing is used to represent a biometric in such a way that, were it to be compromised, the attacker would not be able to determine the identity of where the biometric sample originated. The authors demonstrate the use of multi-state BioHash to resolve the stolen token problem in semantic web applications.

In Rodríguez et al. [17] the authors considered the potential of the semantic web to enhance biometric authentication. The authors presented a framework for solving multimodal fusion oriented biometric representation. Different biometrics have different metrics, and this requires multiple structures of data storage. This is an important issue due to the heterogeneity problem, keeping the structure of databases created with the aim of being used for identity accreditation and distributed over the Web. The authors add semantics to Web Services to perform a role of entry points for such databases. This allows the proposed framework to enable different biometric identity data to be discovered, located and accessed since they provide formal means of leveraging different vocabularies and terminologies.

3.2 Previous Work with WebID and Biometrics

In Nick et al. [15], the authors built a protocol for federated biometric access. This protocol merged GEFE with the WebID protocol to allow users more control over their cyber authentication credentials. The authors utilized OWL-based policies in place of the W3C vocabulary for increased flexibility. This protocol works similarly to the WebID protocol in that the user can bring their own identity, but their protocol allows for users to bring their own biometrics to a service that has biometrics enabled. The proposed protocol, unlike the WebID protocol, reasons about permissions using policy documents expressed in OWL (the Web Ontology Language). The protocol, unlike the WebID protocol, reasons about permissions using policy documents expressed in OWL (the Web Ontology Language). The authors additionally included an extension for group access control.

Martin et al. [13] proposed to use the WebID protocol process for authentication for the web instead of using the standard process of username and password to gain authentication. The average person has up to 25 password-protected accounts [8]. It will be a challenge to remember different passwords for different websites, but it would be dangerous to write down any password as passwords would be at risk of being stolen. The paper discusses more on the use of biometrics and WebID. When these two are used concurrently, security is greatly enhanced because only a confirmed biometric match is allowed to access a WebID certificate.

Gwyn et al. [9] introduced a simplified validation process for a user, which can be applied through the internet and also protects against intrusion attacks. In the validation process, biometrics of an individual matches the previous enrolled Feature Vectors (FV) of that individual [14]. The paper further discusses how to incorporate biometrics to the WebID protocol and implement an enrollment protocol that has a simplified identity management system and also allows a single sign-on. Gwyn et al. used these libraries to develop and implement the enrollment protocol and an http server application used to provide routing. Gwyn et al. also captured how a client can take videos and picture shots using Angular and JavaScript for web application construction. Gwyn et al. introduced js-objectdetect which is a library in JavaScript in their work. The js-objectdetect library was used to detect and identify the face of the client on the canvas. For enrollment, Gwyn et al. created a homepage through which the user accesses the server and which allows a user to create or register for a WebID account if the user does not have one. After the user acknowledges the WebID account, the access process of the server continues.

Prior work has focused on using GEFE as the disposable representation for an individual. However, the concern is reaching the upper bound on unique representations that are different enough to distinguish between multiple feature vectors. Failure to distinguish different feature vectors could allow for a successful replay attack, though the odds are less as the number of unique feature vectors rises. The different permutations of feature vectors increases the number of unique representations that can be created by a singular feature extractor from 1 to $n!$, where n is the number of regions of a biometric image being extracted from.

4 WEBID WITH PERMUTED DISPOSABLE FEATURE VECTORS

The proposed framework will incorporate components of the WebID protocol in addition to a novel implementation of biometric feature extraction scheme that can create a unique representation of an individual while still remaining suitably accurate. This biometric scheme involves the Local Binary Patterns (LBP) algorithm, Genetic and Evolutionary Feature Extraction (GEFE), and the utilization of the GEFE technique to digitally represent a biometric sample in a number of unique ways. Past instances of GEFE allowed for multiple representations of the LBP algorithm in such a way that each representation was unique from another while the features extracted are unique per individual such that there is a reduced chance of any false positive recognition. These unique representations are referred to as disposable feature extractors. The DPFV technique used in this framework will permute the order in which features are extracted from each unique disposable feature extractor to create a set of unique, permuted feature vectors for individuals. While the combination of WebID and Biometrics simplifies identity management, the proposed approach will allow for increased security against biometrics based replay attacks. This proposed framework and the major components, such as the WebID functionality and the permuted biometric representation, are discussed in this section.

In an effort to avoid reaching the upper bound of unique disposable feature extractors, a permutation of the histograms was proposed that showed promise for a standard biometric authentication system. For facial recognition, the permutation of histograms in GEFE FEs had a significantly more distinct representation than histograms from an LBP feature extractor. GEFE creates feature extractors that extract features from n portions of the image, resulting in n histograms to form the Feature Vector (FV). There are $n!$ possible permutations of the histograms, thus the set of FVs that can be created has a cardinality of $n!$. In order for the FVs in the set to be unique, it must be assumed that all histograms within a FV are significantly different from each other. This is tested by comparing different permutations of FVs from the same extractor, as well as different extractors, and recording the similarity scores between the comparisons of two FVs. This technique will be incorporated into the WebID + Disposable biometrics framework to increase a number of unique representations of an individual to successfully prevent replay attacks while still incorporating WebID components for decentralized authentication.

In Gwyn et al. [9], the number of unique representations of a biometric depended on the number of disposable feature extractors that were created. Assuming there are y applicable feature extractors, there are y unique ways to represent one's biometric. This is based on results in [19] which show that different feature extractors will create distinct FVs to avoid a false acceptance into the system. The limitation is after y sessions, the previously used feature extractors will have to be re-used. Suppose an attacker captured authenticating information from one session. After the feature extractors have been recycled, the odds of a successful replay attack in this case are $1/y$. The proposed PDFV method is based on work that shows the effectiveness of permuted histograms for unique representation [20]. If y feature extractors have z histograms, there are now $y^*z!$ possible unique representations. The PDFV method improves on what was proposed in Gwyn et al. with respect to the number of unique representations that exists. The likelihood of a successful replay attack in the WebID + PDFV system is now $1/(y^*z!)$.

The WebID + PDFV system functions in two parts: 1) The TLS handshake approach of WebIDs that sends the certificate from the client to the resource that is being accessed and 2) the biometric authentication approach that will have a user providing a biometric sample, created with a unique permutation order from a randomly selected feature extractor. This system will be enabled with the addition of the biometric element to WebID profiles, which will allow WebID and biometric authentication for verification and added security. A scenario of this system in action can be seen in Figure 2. The scenario sees student user Taminfi attempting to access their confidential student records from their university's record site. Tamifi's client will first send a TLS request to the University's record providing service. Upon receipt of the TLS request, the university site will request a certificate from Tamifi's client. The certificate will be sent from Tamifi's client, and the University site will read the certificate to pull the URI of Tamifi's profile from Tamifi's dedicated personal server, and de-reference it. The University site will then have to compare the modulus and exponent in Tamifi's profile with the client provided modulus and exponent. Upon matching of the modulus and exponent, the University site will then send the RDF for the specific FE that has been randomly generated. All FEs will have been previously created prior to first-time use of the system. On the client site, a random, one-time permutation order will be used to represent the FV to send to the University site. This permutation will be added to a list of previously used permutations. The permutation that is selected must not be in the previously used list. The University site will de-reference a document on Tamifi's server, and will compare the provided FV from the client side and the previously enrolled FV [with the correct permutation order]. If the biometric FVs are similar enough to fall within the acceptable range of acceptance, Tamifi's client will be notified that access to the University site has been granted. After the biometric FV has been compared with the enrolled sample, the permutation of the selected feature extractor that has been used will be added to the list of previously used permutations. Once all permutations for all feature extractors have been used, a random ordering of new permutations will be generated to represent individuals in future sessions.

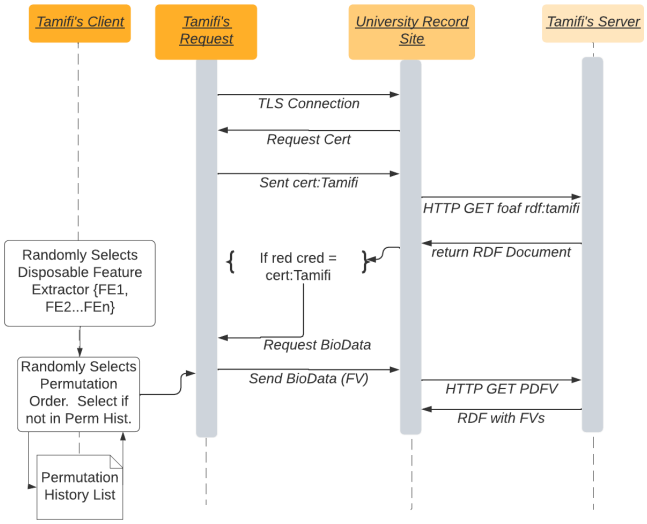

Figure 2: WebID Sequence Diagram

5 DISCUSSION

5.1 Solid

Data manipulation on the internet has become a cause of worry. The internet giants, such as Google and Facebook, make use of their systems to collect data on their users in return for free service. This violates the idea of the web that Tim Berners-Lee proposed. To alleviate this data manipulation phenomenon, Berners-Lee introduced the Social Linked Data (Solid) project [5]. This is to decentralize data and to decouple information using web technologies such as RDF, OWL and SPARQL. Instead of signing in to everything with Google or Facebook, one could sign in with one's favorite Solid provider and would not be tracked. Solid will make users on the web or social networks the owners of their own data and give them control over how applications access their data. RDF plays a major role in Linked Data technologies that are used to reach a high level

of interoperability and serves as part of the Semantic Web technologies data standardization environment [3]. In Solid, WebIDs are used to identify agents (people and organizations). Solid's decentralized data stores called Pods function as secure personal web servers for data. Note that a Solid user may, but need not, self-host, that is, have their data on their own server, but a Solid user could rely on an identity provider (providing their one account, identified by their WebID) and a Pod provider (hosting their Pods). One could also self-host to become an identity provider and Pod provider for a group of users.

Agents can be identified in Solid platforms through the use of WebIDs. In Solid, WebIDs are used to get access to resources through other agents or resources. An Agent or resource can have the privilege to access a particular resource by using an RDF access or authorization mechanism called Web Access Control (WAC) [10]. The WAC framework supports the modes read, write, append, and (to set the access control list) control.

With biometrics incorporated into the identification system of Solid, a Solid user would in a way have their own proxy via their enhanced profile accessed through their identity provider. A user would also in a way have, associated with this proxy, their own online memory in the form of their Pods. While using web standards and software such as browsers, Solid retains a user's autonomy and makes collaboration straightforward and secure.

5.2 Biometrics

Biometric authentication has inherent properties that make the process of recognition more convenient and secure as one does not have to memorize a biometric as one would a passcode, nor is a biometric something that can be stolen or misplaced, such as an access card. The limitation of commercial biometric recognition is the scanning device to capture a biometric sample, but devices such as smartphones and most laptops and tablets have cameras and fingerprint scanners embedded. It would not be a far stretch to say that the future will see more commercial usage of biometric authentication. Within the context of the semantic web, the following factors should be considered: The biometric that should be used, the risk of privacy being compromised, and the trust that can be granted to a biometric sample.

There are a number of biometrics that can be used, with varying pros and cons for each. Facial biometrics are a very convenient biometric as cameras and webcams are prevalent on most devices, and a user simply has to stand in front of a camera. However, recognition using face has a lower average recognition accuracy than other biometrics. There is the risk that individuals who look similar may be mistaken for one another in the system, and that a face's appearance can change drastically with facial hair, makeup, and other features. Iris biometrics have a very high recognition accuracy, they are fairly consistent in appearance over time, and most people have an iris. However, the difficulty in this biometric lies in the cost of an accurate iris scanner. DNA as a biometric sample is nearly perfect for authentication, yet the expense and time it would take to test a DNA sample makes this option nonviable at this time. As technology progresses, more biometrics and combinations of biometrics will be options.

While biometrics have conveniences and are secure in some ways, one risk is loss of anonymity in the event that the provided biometric sample is compromised. Unlike a password or smart card being compromised, privacy can be permanently jeopardized. The PDFV method represents biometrics in such a way that reverse engineering a feature vector into its original image is challenging. Work done in [1] has shown techniques to prevent reverse-engineering feature vectors to determine identity. Another risk is spoofing attempts, where a malicious actor presents a still image of a biometric sample to a biometric scanner pretending to be a legitimate user. There are a number of techniques that can be incorporated to prevent spoofing attacks, such as [12]. The topics for consideration are where in the Semantic Web these techniques should be incorporated, what is the least costly and most effective to incorporate them, and what other strategies could be incorporated as added security.

While the PDFV method is robust against replay attacks, suppose an attacker gains access to a networked session that they should not have access to. Biometric active authentication is one approach that can protect from prolonged unauthorized access. Active authentication is the process of continuously monitoring one's behavior while logged into a session to determine whether the behavior is similar to that of the legitimate user. This is an additional layer for protection, though consideration must be made for what behaviors will be analyzed. This can be device specific, such as mouse movement on a laptop/PC, or finger swipe movements on a mobile device. Biometric-based active authentication has a drawback in that the threshold of the system for determining legitimacy can be challenging to determine. It must allow for a legitimate user to stray from their baseline behavior without triggering multiple false flags, while being strict enough to detect when an illegitimate individual is logged into some system. Ultimately, the ideal system using biometrics will be set in such a way that anyone can provide their biometric sample, it can perfectly authenticate with no error, and it can continuously authenticate even after an access attempt to establish a networked session has been approved.

6 CONCLUSION AND FUTURE WORK

As technology matures, the World Wide Web can evolve to exploit more aspects of the Semantic Web. In addition, biometric authentication is likely to become more commercialized with more devices having tech to function as a biometric scanner. To get the most potential of the maturation of technology, the system presented in this paper addresses this future growth. In this work, we have introduced a system that incorporates the WebID protocol with the permuted disposable feature vector (PDFV) approach for added security and more control to the user for authentication. Replay attacks are devastating due to the information at risk, such as banking information, medical history, or any similar information. The PDFV method improves upon prior research for increasing the number of unique representations for a biometric sample, similar to a one-time passcode.

Future work will be focused on continuous active authentication schemes that integrate into a system using WebID. Hackers can capture and attempt to brute force guess the appropriate permutation order from a captured biometric if they can decrypt the

data. Techniques that can obfuscate any captured data should be considered to prevent this. Storage is also a concern as there are typically multiple servers involved in a system using WebID. We will run tests on a simulated system using biometrics and WebID to determine vulnerabilities depending on where specific data is stored. With the combined WebID representation and biometrics, a user is granted more control over their online representation. This opens up avenues of research that will investigate security vulnerabilities that are open, and how service providers will operate with these security issues.

ACKNOWLEDGMENTS

This research is funded by the National Science Foundation (Award number 1900187, Collaborative Research: HBCU Excellence in Research: Computational Framework and Data Science for Identification).

REFERENCES

[1] Joshua Adams, Gerry Dozier, Kelvin Bryant, Joseph Shelton, Aniesha Alford, Derrick Leflore, and Tamirat Abegaz. 2012. Neurogenetic Reconstruction of Biometric Templates: A New Security Threat?. In *Proceedings of the 2012 IEEE Southeastcon*. IEEE, Orlando, USA, 1–8.

[2] Danny Ayers and Max Völkel. 2008. Cool URI, for the Semantic Web W3C Interest Group Note 03 December 2008. (2008). https://www.w3.org/TR/2008/NOTE-cooluris-20081203/

[3] Tim Berners-Lee. 2009. Linked-data Design Issues. W3C Design Issue Document. *The World-Wide Web Consortium W3C* (2009). https://www.w3.org/DesignIssues/LinkedData.html

[4] Dan Brickley and Libby Miller. 2014. *FOAF Vocabulary Specification 0.99, Namespace Document 14 January 2014-Paddington Edition.* http://xmlns.com/foaf/spec

[5] Sarven Capadisli, Ruben Verborgh, and Kjetil Kjernsmo. 2021. *Solid Protocol Version 0.9.0, Solid Project, 2021.* https://solidproject.org/TR/protocol

[6] Ryan Dellana and Kaushik Roy. 2016. Data Augmentation in CNN-based Periocular Authentication. In *Proceedings of the 2016 6th International Conference on Information Communication and Management (ICICM).* IEEE, Paris, France, 141–145.

[7] Akhilesh Dwivedi, Suresh Kumar, Abhishek Dwivedi, and Manjeet Singh. 2011. Cancellable Biometrics for Security and Privacy Enforcement on Semantic Web. *International Journal of Computer Applications* 975 (2011).

[8] Dinei Florencio and Cormac Herley. 2007. A Large-Scale Study of Web Password Habits. In *Proceedings of the 16th International Conference on World Wide Web*. Alberta, Canada, 657–666.

[9] Tony Gwyn, Taylor Martin, and Albert Esterline. 2020. Increasing Security of WebIDs Through Biometrics. In *Proceedings of the 2020 IEEE SoutheastCon*, Vol. 2. IEEE, online, 1–5.

[10] James Hollenbach, Joe Presbrey, and Tim Berners-Lee. 2009. Using RDF Metadata to Enable Access Control on the Social Semantic Web. In *Proceedings of the Workshop on Collaborative Construction, Management and Linking of Structured Knowledge (CK2009)*, Vol. 514. Washington D.C, USA.

[11] Anil K. Jain, Patrick Flynn, and Arun A. Ross. 2007. *Handbook of Biometrics*. Springer Science & Business Media.

[12] John Jenkins, Kaushik Roy, and Joseph Shelton. 2020. Using Deep Learning Techniques and Genetic-Based Feature Extraction for Presentation Attack Mitigation. *Array* 7 (2020).

[13] Taylor Martin, Yasmin Eady, Justin Zhang, Cory Sabol, Albert Esterline, and Janelle Mason. 2019. The WebID Protocol Enhanced with Biometrics and a Federated Enrollment Protocol. In *Proceedings of the 2019 IEEE SoutheastCon*. IEEE, Huntsville, USA, 1–5.

[14] Taylor Martin, Justin Zhang, William Nick, Cory Sabol, and Albert Esterline. 2018. Implementing WebIDs+ Biometrics. In *Proceedings of the 2018 ACM Southeast Regional Conference*. Richmond, USA.

[15] William Nick, Joseph Shelton, Cory Sabol, and Albert Esterline. 2017. Federated Protocol for Biometric Authentication and Access Control. In *Proceeding of the 2017 IEEE Computing Conference*. IEEE, Napa, USA, 854–862.

[16] Matti Pietikäinen. 2010. Local Binary Patterns. *Scholarpedia* 5, 3 (2010).

[17] Luis Antonio Puente Rodríguez, Maria Jesus Poza, Juan Miguel Gómez, and Belen Ruíz-Mezcua. 2008. Biometric Authentication Devices and Semantic Web Services - An Approach for Multi Modal Fusion Framework. In *Proceedings of the First International Conference on Biomedical Electronics and Devices*.

[18] Cory Sabol, William Nick, Maya Earl, Joseph Shelton, and Albert Esterline. 2016. The WebID Protocol Enhanced With Group Access, Biometrics, and Access Policies. In *Proceedings of the 2016 Modern AI and Cognitive Science Conference (MAICS)*. Dayton, USA.

[19] Joseph Shelton, Kelvin Bryant, Sheldon Abrams, Lasanio Small, Joshua Adams, Derrick Leflore, Aniesha Alford, Karl Ricanek, and Gerry Dozier. 2012. Genetic & Evolutionary Biometric Security: Disposable Feature Extractors For Mitigating Biometric Replay Attacks. *Procedia Computer Science* 8 (2012), 351–360.

[20] Joseph Shelton, Gerry Dozier, Joshua Adams, and Aniesha Alford. 2012. Permutation-based Biometric Authentication Protocols for Mitigating Replay Attacks. In *Proceedings of 2012 IEEE Congress on Evolutionary Computation*. IEEE, Brisbane, Australia, 1–5.

[21] Joseph Shelton, Gerry Dozier, Kelvin Bryant, Lasanio Small, Joshua Adams, Khary Popplewell, Tamirat Abegaz, Aniesha Alford, Damon L. Woodard, and Karl Ricanek. 2011. Genetic and Evolutionary Feature Extraction via X-TOOLSS. In *Proceedings of the International Conference on Genetic and Evolutionary Methods (GEM)*. Las Vegas, USA.

[22] Joseph Shelton, John Jenkins, and Kaushik Roy. 2017. Extending Disposable Feature Templates for Mitigating Replay Attacks. *International Journal of Information Privacy, Security and Integrity* 3, 2 (2017), 96–116.

[23] Henry Story, Bruno Harbulot, Ian Jacobi, and Mike Jones. 2009. FOAF+ SSL: RESTful Authentication for the Social Web. In *Proceedings of the First Workshop on Trust and Privacy on the Social and Semantic Web (SPOT2009)*. Heraklion, Greece.

[24] Sebastian Tramp, Henry Story, Andrei Sambra, Philipp Frischmuth, Michael Martin, and Sören Auer. 2012. Extending the WebID Protocol with Access Delegation. In *Proceedings of the Third International Workshop on Consuming Linked Data (COLD2012)*. CEUR-WS, Boston, USA.