ACCOUNTABLE HYPER TEXT TRANSFER PROTOCOL (HTTPA)

Miss. Dipali P. Sapkal **Mr. Chandrasekhar D. Badgujar**

Abstract *— As the Internet has gained widespread use, and advanced technologies such as high-speed multi-media technologies and automated digital monitoring have become a reality, privacy is at the greatest risk of all time. At the same time, sophisticated threats from hackers, terrorists, thieves, and others that would abuse privacy highlight the need to find technologies that provide some accountability. A new way of managing how data violations ranging from exposure of browsing habits of on the Web is used with an infrastructure that enables accountability on the Web at the protocol level. A protocol, HTTPA (Accountable Hyper Text Transfer Protocol), which requires that the data producer and the data consumer come to an agreement before an HTTP transaction, takes place. This process makes both parties accountable for the agreement privacy protection for systems like the World Wide Web; they had entered into, especially when reusing the data that was transferred. In HTTPA, the data consumer expresses her intentions of access and usage, and the data producer expresses her usage restrictions. The data transfer only happens when the intentions match the restrictions and the transfer along the agreement is logged. This protocol cannot prevent the unauthorized reuse of data, but rather it can be used to develop accountability mechanisms that will identify violators allowing control systems is that it is the users' responsibility to define them to be held accountable for data they inappropriately consumed and served*

In this paper, we also focus accountability in data mining areas, the interactive technology tools and privacy policy issues of current trends.

Key words *– accountability, Authenticity, Privacy, Restriction.*

I. INTRODUCTION

As of our daily activities are conducted on networked computers, privacy has become harder to maintain. Advances in networking, data storage, and data processing make it easier for information to be retained and accessed. Privacy is an important issue to many people, businesses and Governments through different entities may not even agree on what privacy is. Loosely speaking privacy is the ability to control private information, including identity and identifiers, sensitive information such as medical or financial records, and information about certain kind of personnel, corporate or Government activity. The kind of privacy usually desired in the real world is not hiding all information from all parties, but rather having the ability to disclose selected information to selected parties under certain circumstances, while preventing other disclosure [1].

Alan Westin defines privacy as the ability for people to determine for themselves "when how and to What extent, information about them is communicated to others" This assumes that there are major privacy risks from unauthorized access to information. This focus on controlling information access has been found to be awed when the information is within reasonable bounds of security, it can leak outside these privacy boundaries violating the initial restrictions imposed on the data, as many social media outlets on the Web provide an easy medium for information dissemination at an unprecedented level. The technology press is filled with announcements by social networking sites about their new privacy controls, i.e. new ways for users to define access rules; followed by embarrassment when the choices prove to be inadequate or too complex for people to deal with. For example, Face book's changes to its privacy settings in spring 2010 made news that highlighted how convoluted their privacy policy has become. Tools such as a "Terms of Service Tracker" have led to visualizations of how Face book is sharing more private data than ever before. Also, Face book's Open Graph Protocol's "like" button has led to possible privacy violations ranging from exposure of browsing habits of people on medical sites to pornographic sites being shared with an unanticipated audience.

When access control systems are successful in restricting access to particular users, they are ineffective as privacy protection for systems like the World Wide Web where it is easy to copy or aggregate information. These days, it is also possible to infer sensitive information such as social security numbers (SSN), political affiliations and even sexual orientation from publicly available information. Another problem with using up-front access control systems is that it is

the users' responsibility to define and maintain their privacy policies in every domain they participate in. Government and businesses themselves also have privacy concerns regarding their own sensitive or proprietary information, as well as wanting to avoid poor public relations associated with mishandling of their citizens or customers private information.

Many websites publish privacy policies which are often very verbose, and rarely do users have the time to read them or understand what they really mean. A typical user will click through the privacy policy statements without completely understanding the risks involved. In a pure access restriction system, those who obtain access to the data, legitimately or not, can use the data without restriction. The Controversial whistle-blowing site wiki leaks. This website exposes sensitive data with the aim of making governments and large businesses more transparent and accountable. Their claim is that no-one has been intentionally harmed so far because of the data published on the site. However, due to the sensitive nature of the data published on the site, it's possible for nations, if not individuals, to get harmed at some level and diplomatic relations to deteriorate. In a recent memo, several U.S. agencies have issued a warning saying that the documents published on the site "does not alter the classified status or automatically result in declassification of the documents". Further, the memo states that "classified information, whether or not already posted on public websites or disclosed to the media remains classified and unauthorized federal employees should not look at eked classified data". This usage restriction is inherently faulty because there can be no enforcement nor can employees be held accountable for accepting the restrictions imposed on the sensitive data.

Weitzner et al define information accountability in terms of usage–when information has been used, it should be possible to determine whether the usage was appropriate, identify the violators and hold them accountable. Lampson argues that to be practical, accountability needs an eco-system that makes it easy for senders to become accountable and the receivers to demand it. It is our belief that HTTPA will provide this eco-system. The goal is to develop and deploy technologies and policies that satisfy both legitimate accountability needs and that balance the sometimes conflicting desires of individual and society.

II.HTTPA SCENARIOS

As social media is becoming central to many things ranging from recruiting to personal relationships, the ability to grant and restrict access to personal data is becoming critical. The ubiquity of the Web, the ability to connect data from external sites to the social networking its, and the amount of time people spend interacting with social media are both advancing our freedoms and enabling novel invasions of privacy. It is our belief that users should be aware of and ideally be in control of information about them on the Web [2].

In the scenarios described below, we take a policy-centric view on privacy on the Social Web where policies capture the permissions such as access control, obligations such as terms-of-use, licensing, and other data-handling settings that allow a user to control their interactions with other users. In particular, policies apply privacy settings to the profile and social media frameworks to consistently manage the user expectations of privacy and other obligations. This allows individuals and businesses on the Social Web to share information without any fear of violating user privacy or any regulations within the purview of the intention of use of their audience.

Scenarios assume that Alice is a user of an imaginary social networking site called 'Social Book'. Alice communicates with Social Book using our protocol, and both parties have specified their intensions and usage restrictions using the RMP (Respect My Privacy) ontology. The Provenance Tracker 'Trust Me' is a third party entity trusted by both Alice and Social Book.

A. Upstream Usage Restriction Management

Suppose Alice wants to upload some pictures on Social Book. The default settings on her smart Web client is set with the usage restriction that any HTTP payload carrying data with MIME type such as 'image', or subtypes such as 'image/[bmp, gif, jpeg, png, x-ico, x-tiff]' will only be posted/uploaded if the recipient acknowledges the full ownership of the content to her. However, it appears that Social Book has extremely draconian terms of service that if uploaded to Social Book, the data becomes the property of Social Book. Alice's client examines these two policies, and informs Alice about the mismatch, which then prompts Alice to either stop posting her pictures or to notify Social Book for the potential terms of use mismatch. In the latter case, Trust Me gets a notification of the handshake that happened between the parties. If Social Book decides to modify the terms of use, it will send another request which Alice accepts and the data will be transferred

B. Downstream Usage Restriction Management

Alice has a photo on Social Book with a usage restriction specifying that the photo cannot be used for any commercial purposes. An employee from a large advertising company, Bob, accessed that photo. Bob's smart client confirmed with Social Book and was logged on Trust Me that the intention of accessing the photo was non-commercial, and that he will honor the corresponding usage restriction that Alice has imposed on the photo. However, few weeks later, Alice found out that Bob had used her photo in an online advertisement for his company. Through her Web client Alice Complains to Trust Me by giving the URI of her photo that Bob had allegedly used. Alice in her compliant also says that Bob's advertisement had used her photo, and that it is of commercial-use. Trust Me verifies that Bob had accessed the photo by looking up the accountability logs. Then it looks up the original usage restriction that Bob agreed to verify that it had indeed violated Alice's terms of use, and sends a takedown request to Bob with a proof detailing the violation.

III.HTTPA COMPONENTS

A. Authentication

Authentication is important in the protocol, not just for access control, but also to find the identity of the users who accessed resources should their owners claim that someone violated their usage restrictions on those resources. HTTPA will use the WebID protocol to manage authentication [9].

B. Usage Restrictions Management

HTTPA uses the RMP to describe the usage restrictions and the intentions associated with the data. Some of the terms included are: No Ownership Transfer, No Com mercial/Employment/Financial/Medical/Insurance use of the data.

C. Handshake

HTTPA breaks away from the traditional client-server model of HTTP transactions, to allow clients to act as servers, and vice versa. The sender (server/data provider) conveys usage restrictions, and the receiver (client/data consumer) notifies her intentions on the data. In the current implementation, we define two HTTP Headers: 'X- Usage Restrictions' and 'X-Intentions' for these purposes. If any one of the parties does not agree with the other party's usage restrictions/negotiations, further negotiations can be carried out using the 'X-Negotiate' header

D. Provenance Trackers and Logging

Provenance Trackers are essentially special Web servers that are delegated to handle logging to enable provenance in HTTPA transactions. They are trusted by both parties involved in the data transfer, and the party initiating the transaction can designate the provenance trackers. The logs kept at the provenance trackers have several characteristics: they are immutable except by protocol components encrypted, secure, and readable only by trusted parties involved in the HTTPA transaction, and have all the records pertaining to a particular data transfer and usage such

as what data was accessed, the specified intent of access, and the agreed upon usage restrictions.

E. Accountability Tracking

If a user finds that she was wronged because someone else misused her data by violating the usage restrictions associated with the data, she can take recourse by producing a provenance trail with the help of the provenance tracker.

IV.ACCOUNTABLE IP

Accountability makes certain types of attacks either more traceable or simply more difficult to mount. AIP provide two types of accountability: Control-plane accountability and Source accountability.

Control-plane Accountability

If routers and ASes were accountable for their routing messages then their peers would be able to more easily discover forgeries or errors .This type of accountability uses two authentications:
Origin authentication: which is particularly lacking in the current routing system, becomes automatic because AD numbers are derived from public keys .Ads can extend public keys using separate BGP messages or using a look up services.
Path authentication: It precedes as in S-BGP some router in AD signs the AD path verifies every signature in the update before installing the route in its routing table.

Source Accountability

Today's Internet architecture lacks source accountability: hosts can easily forge the source IP address of data traffic, which makes attacks difficult to track and makes it nearly impossible for network operators to filter based on the source address-the most logical identifier for doing so…

If sources were accountable, then any element in the network that saw a packet could verify that packet's origin. This property eliminates undetectable source address forgery. As a result of preventing such forgery, defenses against DoS could profile using source addresses, routers could implement packets filters or rate limiters using source addresses.

V.ACCOUNTABLE PRIVACY

Accountability is presently accomplished by a combination of entity and message authentication, action/event binding, monitoring and trust infrastructures each of which are as follows-
Entity and Message Authentication Methods: Identity may be traced from an IP address, to a pseudonym, to an account name to the registered name at the ISP, and often through other identifying stages, eventually to some fundamental identifying information such as social security number, a finger print, or possibly a DNA match.
Binding Actions to Events: Few mechanisms for binding actions to events have been proposed and indeed, when such mechanisms have been engineered into the infrastructure, public outrage has purged them from the marketplace e.g. Carnivore and clipper.
Monitoring: To ensure accountability, monitoring must be feasible and effective. Tracing all packets on the internet leads to massive data volumes that challenge even the most sophisticated data mining techniques to analyze.
Trust Infrastructures: Public Key Infrastructures (PKIs) have been proposed and partially deployed as a method for managing trust in security protocols. An important focus is to develop cryptographic infrastructure models, techniques, principals and tools to facilitate accurate, efficient, privacy-balanced accountability [7].

VI. ACCOUNTABILITY IN DATA MINING

Architecture of Data Mining

An information architecture consisting of general-purpose internecine components connected in a manner that provides transparency of inference steps and accountability to rules

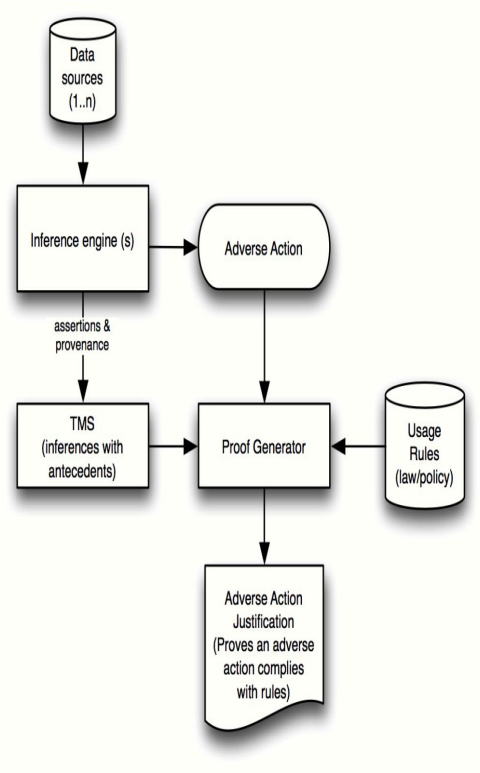

Figure. Functional Architecture

The transparency and accountability architecture depends upon three components as follows:

Inferencing Engine(s): support analysis of data available and assesses compliance with relevant rules.

Truth Maintenance System: a persistent store Fed by the inference engine(s) consisting of proof Antecedents as well as data provenance, used to assess reliability of inferences and d to record Justifications for proof antecedents developed in the course of an investigation.

Proof Generator: constructs proofs that critical transitions and adverse uses of personal information are justified by facts and permissible under applicable rules.

The inference engine provides assistance to the government investigator or analyst in identifying suspicious profiles in those data sets accessible for this purpose. This data would then be processed through an inference engine that provides investigative results.

In addition to these investigatory inferences, a record of the inferences and their justifications will be stored in the Truth Maintenance System (TMS). The TMS combined with a proof generator allows anyone with access to the system to determine whether or not the person al information in the system is being used in compliance with relevant rules and consistent with known facts. At critical stages of the investigation, such as sharing of information across agency boundaries or use of information to support an adverse inference (secondary Screening, criminal indictment, arrest, etc.), the proof generator will attempt to construct a proof that the use proposed for the data at that transition point is appropriate The proof generator would be able to draw o n information collected in the TMS and bring to bear the relevant rule sets.

VII.ACCOUNTABLE PROTOCOLS

P3P (Platform for Privacy Preferences) protocol was developed at the W3C with the intention of communicating the privacy policies of websites to the user-agents who connect with them. The recommendation allows website operators to express their data collection, use, sharing, and Retention practices in a machine-readable format. A user agent can retrieve a machine readable privacy policy from the Web server and respond appropriately (for e.g. display symbols or prompt the user for action). However, P3P has several limitations: a complicated language to express Policies, inability to express preferences on third party data collection, and to specify multiple privacy policies for one web page.

FTC endorsed a 'Do not Track proposal recently to facilitate consumer choice about online tracking, and there are already several implementations that support this proposal. One of the most compelling technical implementations describes sending the user's intention

of not to track online browsing behavior in an HTTP header. Although this approach works for this specific use case, it seems very limited for general purpose usage restrictions matching with intentions

VIII.INTERACTIVE TECHNOLOGY TOOLS

In the digital world of computers, Internet-based interactive technologies such as e-mail, Usenet, lists ergs, and Bulletin Board Systems (BBSs) pre-date the advent of the Web. What invests Web 2.0 with power is not its novelty, but rather the integration of interactivity onto a global electronic platform, where interactivity not only transpires amongst people, but also enables the meshing (or "mashing") of content, applications, and software, among other things.

The explosion of ***Face book*** and ***MySpace*** demonstrates the power and reach of social networking or social media, perhaps the defining Web 2.0 technology. Social media enables individuals to interconnect on the Web, allowing them to follow the updates of family and friends, as well as network to make new connections. Web 2.0 social networking platforms also apply to specific communities, such as *LinkedIn* for professional interactions, and *Just Means* and *Development Crossing* for the CSR community.

Twitter – which imported the concept of text messaging (or Short Message Service – SMS) from the world of cell phones to the World Wide Web – has propelled the popularity of micro blogging, so named after the 140-character limit for content (Milstein 2009). Twitter has proven its usefulness as an efficient, real-time, dispersed communication tool in the 2008 terrorist attacks in Mumbai, India, and as an organizing tool after allegations of election fraud in Iran in June 2009.

Transparency and Accountability in the Current Privacy Policy

Our efforts to structure laws and develop technologies with sensitivity for privacy values should seek guidance from the nearly century-long interplay between ever-growing surveillance capabilities of new technologies and fundamental privacy principles. Historically, we learn that as electronic communications have become more sophisticated and more ubiquitous, communications privacy law has responded to the advance in law enforcement needs privacy threats by trying the growth in surveillance capabilities to gradually expanding privacy protections that kept pace with new intrusion powers [8].

Balancing Privacy and Accountability

To properly analyze the requirement of accountability and Privacy, one must understand and balance a number of potentially conflicting desires [9]. In particular, it is important to consider the conflicts between:

- Anonymity and identification
- Confidentiality and required information disclosure
- Freedom of actions balanced with attribution of actions
- Cyber-Privacy and cyber-forensics
- Free speech and liability/copyright.

VIII. CONCLUSION

HTTPA addresses the limitations of current privacy work and provides the infrastructure to build more privacy-aware systems. The requestor, on data access, will convey what her intention for the data access is. The data provider will determine the compliance/non-compliance of the intention sent by the requestor with the usage restrictions associated with the resources that are being accessed. Their negotiation is being logged by a trusted third party called 'Provenance Trackers' to ensure accountability. If usage restrictions are compliant with the intentions, the data access request will be successful. If it is non-compliant, an explanation as to why the data cannot be transferred will be conveyed to the requestor. The recipient of the data will be held accountable for the usage restrictions she accepted upon the data transfer In other words,

recipients cannot argue after the fact that they did not know the expectations of the data server: for retention or for use of information. Similarly, users cannot claim after the fact that the data server was deceptive or that they had not been informed. This enables market and regulatory forces to punish users who misuse data. We believe that government organizations, academic institutions and businesses will be the early adopters of this accountable. Web protocol with usage restriction management within their intranets. On the longer run, in a similar vein in which the growth of e-commerce Web sites led to the massive adoption of HTTPS, we envision that HTTPA will be accepted by the larger Web community, as privacy problems slowly cripple the growth of the Web.

We can develop this mechanism more prompt as before the violation takes place, it will restrict the violator to violate the data by using real time application or the camera must be used to capture the violator photo also**.**

Acknowledgment

It is a matter of great pleasure for me to present my paper on "Accountable Http". I take this opportunity to express our deep sense of gratitude to my HOD (Deptt.of CO & IT) Prof. P. P. Rewangad and my guide Prof. C. D. Badgujar. As their sincere effort and selfless guidance has been inspiration in all phases of my paper work.

Also I would sincerely like to thank to my parents, friends & all above almighty for the support they provided.

REFERENCES

[1] A. Westin, "Privacy and freedom (Fifth ed.). New York, U.S.A.: Athenaeum," 1968.

[2] L. F. Cramer, "Web privacy with platform for privacy preferences," Oreilly Books, Jan 2002.

[3] J. Mayer and A. Narayanan, "Do Not Track - Universal Web Tracking Opt Out." [Online]. Available: http://donottrack.us

[4] PC World, "Researchers expose security aw in social security numbers," 2009.

[5] Lampson, "Usable security: how to get it," Communications of the ACM, Jan2009.[Online]. Available: http://portal.acm.org/citation.cfm?id=1592761.1592773

[6] T. Kang and L. Kagal, "Enabling privacy-awareness in social networks," in Intelligent Information Privacy management Symposium at the AAAI Spring symposium 2010, March 2010. [Online]. Available: http://dig.csail.mit.edu/2010/Papers/Privacy2010/tkangrmp/paper.pdf

[7] "Webid protocol," WebID 1.0 - Web Identification and Discovery [Online]. Available:http://getwebid.org/spec/drafts/EDwebid20100809/index.html

[8] T. Beth, \Zur Sicherheit der Informationstechnik", Informatik-Spektrum, Vol. 13, Springer-Verlag, 1990, pp. 204-15 (in German).

[9] S. Kent, C. Lynn, and K. Seo. Secure border gateway protocol (S-BGP). IEEE JSAC, 18(4):582–592, Apr. 2000.

AUTHOR'S PROFILE

Miss. Dipali P. Sapkal

M.E. (COMPUTER SCIENCE AND ENGINEERING) from NMU Jalgoan.

B.E. (INFORMATION TECHNOLOGY) Santa Gadage Baba University Amaravati and pursuing research under the guidance of **Mr C. D. Badgujar.**.

Mr.ChandrasekharD.Badgujar

M.Tech (COMPUTER ENGINEER) NMIS,Mumbai.
B.E (INFORMATION TECHNOLOGY) from NMU,Jalgoan and doing research work in Autonomic Computing