

A confidence model for web browsing

Author(s): Prasanta Behera, Naveen Agarwal, Yahoo! Inc

Introduction

The Internet usage model varies a lot - from novice users to highly active users. The user experience plays a critical role in determining how the users use the Internet. Easy access to the Internet, great user interface, and security are some of the aspects that affect the user experience. Many technological strides have been made to create a more secure environment but it is a continuous battle. It would be great to have a simple model that provides a level of confidence to a user when he is navigating to different pages within the domain(s) of a company. The proposal in this paper discussed some ideas but not a concrete model.

Problem

Users navigate the web primarily via links or by typing the URL. The user may make a typing or spelling mistake and thus end up on a different site than intended. Similarly, by various means, a malicious link can take the user to a site/domain different than the one expected in the link.

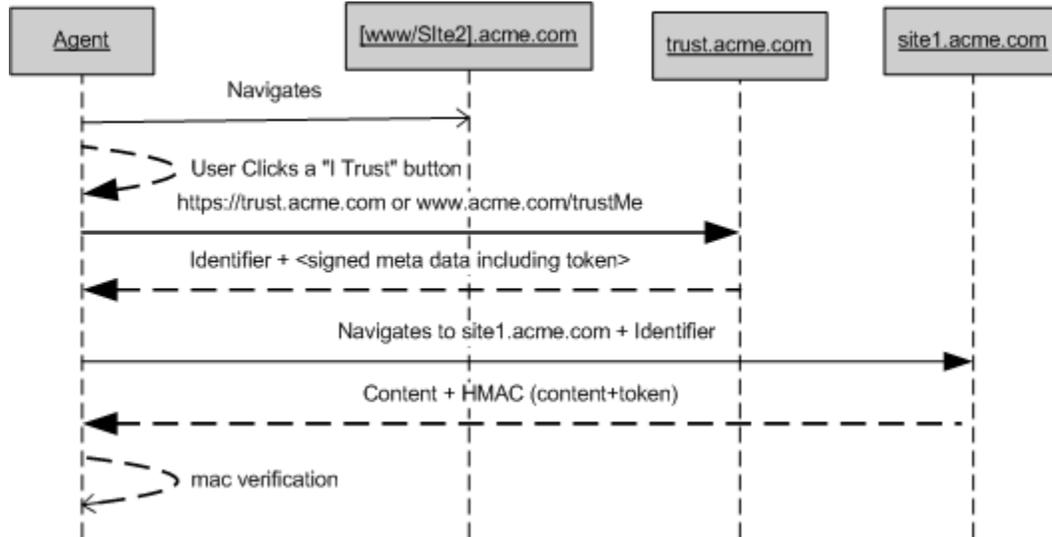
Simple mechanisms are necessary to solve this problem. This can be client based, server based, or a combination of both. There have been a number of proposals focused on the client side so far. They either require manual setup or use a scoring mechanism to provide the user some level of trust. SpoofGuard [1] creates a score based on the aggregated information such as URL, image on login page and user behavior and then alerts the user.

Proposal

We think a mechanism where servers provide security relevant metadata via a standard mechanism/protocol can play an important role along with client-side tools. We propose that page served on a site should contain some information specific to each client. The client uses this information to provide an extra level of confidence to the user.

One approach is that each browser is assigned a domain specific unique identifier. In addition, each domain/site has a predefined location (e.g., trust.acme.com or www.acme.com/trustMe) that the browser contacts when a user indicates to trust the site. Signed metadata is exchanged that contains a

token that is unique to the client and is stored in the client. This exchange must be over SSL. When the user navigates to a site within the domain, based on the unique identified, the token and content, a MAC signature is generated dynamically, which is verified by the browser to make sure that the user is still in the trusted domain. A prominent visual indicator displays the status to the user.



Let's look at a sample metadata returned by the trust.acme.com. The following sample illustrates the idea and is not complete by any means. Further discussions are required.

```

<metadata>
  <token/> // a per browser token generated by the trust.acme.com
  <domain info/> // domain name, etc
  <expiration/> // Expiration rules
  <allowedDomains>
    <d>acme.com/<d>
    <d>acmesports.com</d>
  </allowedDomains>
  <watchListDomains>
    <d>acmee.com</d>
    <d>acmeeee.com</d>
  </watchListDomains>
  <xmld: Signature/> // hash and signature
</metadata>
  
```

The tokens expires based on the values defined and it should be configurable. The <allowedDomains> lists the domains the token is valid. This allows a company to specify other domains that it may use to serve content. The <watchlistDomains> are used indicate that certain domains are trying to spoof (e.g., acmee.com). The browser can process these rules. When the user goes to visit site2.acme.com, the site uses the unique browser identifier and generates the token (using the server key). It uses the token as the shared secret to generate the MAC of the content (excluding header) and appends to the html in a comment. The MAC could also be sent in the header. The client can verify the MAC using the token it has.

```
<html>  
</html>  
<!--MAC="dshdsu#2sjdnss" -->
```

Risks & Caveats

The unique browser id provides a check against global replay attack. User specific replay attack possibilities are there. However, if the token use expiration time, the window of replay attacks can be minimized. The proposed approach may not help protect the user if the machine is compromised. The management of secret used for generation of signature/token for large and hosted sites requires further consideration. The proposal also does not address non-html content. This does not help if the user places his/her trust on a site that that is not what s/he thinks it is.

Conclusion

There is a strong need for a standard to solve some of the issues as discussed. The solution needs to be simple. We think a combination of client and server side solution is required.

Acknowledgement

We would like to thank Drew Dean and Scott Renfro for their helpful insights.

Bibliography

- [1] Neil Chou, Robert Ledesma, Yuka Teraguchi, Dan Boneh, and John C. Mitchell, "Client-side defense against web-based identity theft", Proceedings of the 11th Annual Network and Distributed System Security Symposium, Feb. 2004
- [2] Edward W. Felten, Dirk Balfanz, Drew Dean, and Dan S. Wallach, "Web Spoofing: An Internet Con Game", Proceedings of the 20th National Information Systems Security Conference, Oct. 1997