Forecasting Privacy leackage *

[Extended Abstract] †

Aghiles DJOUDI, Guy PUJOLLE
Sorbonne University
4 Place Jussieu, 75005 Paris, France
firstname.lastname@upmc.fr

ABSTRACT

TODO

Keywords

Privacy score, Trust score

1. INTRODUCTION

TODD aghiles

2. BACKGROUND

The study of privacy measurement applied to information and communication technology (ICT) is very wide and embraces many fields of knowledge from Sociology and Statistics to Cryptography and Artificial Intelligence. To better understand the privacy issues in ICT, we describe different types of privacy threats used by attackers to access users' private information through messaging services. Next, we present four options available to the end users to deal with ICT privacy issues.

2.1 Privacy threats

Four types of privacy threats have been discovered in the literature:

2.1.1 Private information disclosure

When users share information with trusted social network community. they implicitly assume that their information shared through messages would stay within the community destination. However, this assumption is not always valid, an individual messages may be accessed by adversaries. For instance, messages sent to an email-based social network may be stored at a repository and consequently visible to the public, malicious users and applications may follow people through social networks, add-ons and third parties may access users private information, etc.

[**krishnamurthy_leakage_2010**] shows that online social networks and applications leak users personally identifiable information to third parties. The results of their study clearly show that the indirect leakage of PII via OSN identifiers to third-party aggregation servers is happening. OSNs in our study consistently demonstrate leakage of user identifier information to one or more third-parties via Request-URIs, Referer headers and cookies. In addition, two of the OSNs directly leak pieces of PII to third parties with one of the OSNs leaking zip code and email information about users that may not be even publicly available within the OSN itself.

2.1.2 Information aggregation

When users authenticate to their favorite service messaging; generally, online social networks, they voluntarily release different types of personal information: name, screen name, telephone numbers, email addresses, locations, etc. Moreover, when users post messages in forums, blogs and webmails, they also disclose small pieces of private information. However, with the development of information retrieval techniques, private information of the same user may be collected from different sources and aggregated to reveal user privacy [**luo_protecting_2009**].

2.1.3 Inference attacks

Aside from voluntary disclosure of explicit personal information, users information could be inferred from public information items. The privacy literature recognizes two types of private information leakage: identity leakage and attribute leakage, and identity leakage often leads to attribute leakage. Identity disclosure occurs when the adversary is able to determine the mapping from a record to a specific real-world entity (e.g. an individual). Attribute disclosure occurs when an adversary is able to determine the value of a user attribute that the user intended to stay private.

2.1.4 Re-identification and De-anonymization attacks

When social network data sets are published for various legitimate reasons, user identity and some profile information are often removed to protect the user privacy. The privacy literature recognizes two types of privacy mechanisms: interactive and non-interactive. In the interactive mechanism, an adversary poses queries to a data-base and the database provider gives noisy answers. In the non-interactive setting, a data provider releases an anonymized version of the database to meet privacy concerns. Some of the well-known techniques for this purpose includes k-anonymity, l-diversity and t-closeness [**li_tcloseness_2017**].

*note1
†note2

For instance, in a k-anonymized data set, an individual cannot be distinguished by attributes from other k-1 records. However, possibilities of **attribute re-identification attacks** on publicly available data sets have been studied in [**li_new_2014**]. They show that user identities could be recovered from anonymized data sets.

On the other hand, due to the nature of social network data, just anonymizing node attributes is not enough. Graph structure contains significant amount of information which could be utilized to hurt user privacy, i.e. **structural reidentification attacks**. A good survey on structural anonymization and re-identification attacks could be found at [**zhou_brief_2008**]. most works show that node identities could be inferred through graph structure.

2.2 Privacy solutions

When an ICT user accesses a service, in particular, a messaging service, she has to share some information with the service provider, namely identity, type of service required, location, etc. Clearly, the shared information depends on the service but regardless of the exchanged information, to deal with the existing privacy issues, users have to choose between four main options:

2.2.1 Privacy based on trust

This is probably the most common situation. Users tend to trust service providers because, they do not really have alternatives in many cases. Due to the fact that privacy is considered a right, most countries have regulations that oblige companies to guarantee the privacy of their users. Among this regulation, when users data are released to third parties they should be sanitized so as to guarantee users privacy [**chen_privacypreserving_2009**].

2.2.2 Privacy based on Individual User Actions

Despite the legislation, users might prefer to keep some of their private information away from the service provider. In this case, we assume that the user cannot collaborate with the service provider. For example, in the case of sending a query to an Internet search engine such as Google or Yahoo. The user cannot initiate a **collaborative protocol** with the search engine, because the search engine is only able to receive and answer queries.

2.2.3 Privacy based on Collaboration with the Provider

There are situations in which the service provider might collaborate with the user to protect her privacy by running **privacy-aware protocols.** TODO

2.2.4 Privacy based on Collaboration with other Users

This is an evolution of the proposals described in 2.2.2 in which users collaborate to protect their privacy. In this case, users do not want to trust the provider nor other third parties. TODO

3. APPROACHES

Quantifying and measuring privacy is very challenging, mainly because the definition of privacy is very subjective, each individual might have a different opinion about this concept. In our work, we present privacy challenges through three points of view: behavioral, social and technical.

3.1 Behavioral

We consider two types of behavior models in this work, one based on typical behavior from a user perspective and one on typical behavior of a users' messages. Both of these models can be quantified into a probability score.

3.1.1 User behavior

The user's behavior is modeled with respect to their typical service messaging usage, the frequency and type of messages received and sent, and the typical recipients with whom they exchange messages. Known as a behavior profile, this type of model is computed over some training period to learn how the user behaves within the messaging account.

The measurements of the user's message behavior include the frequency of inbound/outbound message traffic, the specific times messages arrive and are sent, the "social cliques" of a user, and the user's response rates when replying to specific senders.

Vidyalakshmi et al. [**vidyalakshmi_privacy_2015**] proposed a privacy scoring using bezier curve. They present a framework for calculating a privacy score metric considering **users personal attitude towards privacy and communication information**. They focus on the rating of the users OSN friends based on their attitudes towards privacy, helping him to make an informed decision of sharing information with them. Bezier curve in its cubic form is used as it has to account for both privacy orientation and communication orientation of the user.

[**alemany_estimation_2018**]

[**zhang_privacypreserving_2017**]

[**liu_framework_2010**] propose a model to compute a privacy score of a user. The privacy score increases based on how sensitive and visible a profile item is and can be used to adjust the privacy settings of friends. Their solution also focused on the privacy settings of users with respect to their profile items. They use Item Response Theory (IRT) to evaluate **sensitivity and visibility of attributes** when evaluating privacy scores. The authors definition of privacy score satisfies the following intuitive properties: the more sensitive information a user discloses, the higher his or her privacy risk. However, their approach do not support personalized privacy view over profile content for each individual in the social network.

3.1.2 Message behavior

The second type of behavior is specific to how spam behaves and how it appears in the message folder among normal messages. In general, spam messages can be easily detected because they appear anomalous with respect to the normal set of messages received and opened by the user. However, ... TODO

3.2 Social

In online social networks, users are sometimes either oblivious about their privacy, or concerned but underestimate the privacy risks. OSN service providers allow users to manage who can access which information and communication (e.g. Facebook and Google+). Researcher studied privacy protection from two directions:

Along the first direction, fundamental changes to the current design of OSN were suggested to enhance users' privacy. Within this direction, Privacy by Design (PbD) is an important approach. For example, in [**baden_persona_2009**], Baden et al. proposed a new type of OSNs by using **attribute-**

based encryption to hide user data, in which symmetric keys are used to encrypt messages and only the designated friend groups can decrypt the messages. In [**erkin_generating_2013**], Erkin et al. proposed to use homomorphic encryption and **multi-party encryption techniques** to hide privacy-sensitive data from the service provider in a recommender system.

The second direction is developing privacy protection tools based on existing OSNs. In our work, we focus on both direction to deal with current and future messaging services.

In [**becker_measuring_2009**], the authors propose to use the amount of information that can be inferred from social networks to quantify the privacy risks. PrivAware detect and report unintended information disclosures through quantifying privacy risk associated with friend relationship in OSNs. PrivAware employs inference model which is based on the fact that information about users can be inferred from their social graph. Privacy score is calculated as total **number of attributes visible to the third party applications** divided by total number of attributes per participant. The measured percentage is then mapped to a letter grade, where A score represents very few attributes being revealed and F score indicates that privacy risk to the threat of a malicious third party application is high.

The authors in [**talukder_privometer_2010**] develop a tool, Privometer, to measure information leakage based on user profiles and their social graph. The leakage is indicated by a probability numerical value. Privometer is based on an augmented inference model where a potentially malicious application installed in the users friend profiles can access substantially more information. It operates in two modes. In online mode, inference is performed based on the friends profile where most frequently value is selected. In offline mode, it uses only immediate friends and "network-only Bayes classifier" to measure the **probability of inference**. The tool can suggest self sanitization actions based on the numerical value.

[**b.s._privacy_2015**] proposed a privacy control framework for information dispersal on social network, they use the quadratic form of bezier curve to arrive at privacy scores for friends, they use the **communication information** for pre-sorting of friends which is lacking in [**vidyalakshmi_privacy_2015**].

Privacy Index (PIDX) proposed in [**nepali_sonet_2013**] is a measure of a users privacy exposure in a social network. PIDX is a numerical value between 0 and 100 with high value indicating high privacy risk in social networks. An **attributes privacy impact factor** is a ratio of its privacy impact to full privacy disclosure. Thus, an attributes privacy impact has a value between 0 and 1. They consider privacy impact factor for full privacy disclosure is 1.

[**akcora_risks_2012**] develop a graph-based approach and a risk model to learn **risk labels of strangers**, the intuition of such an approach is that risky strangers are more likely to violate privacy constraints.

Fang and Le Fevre [**fang_privacy_2010**] proposed a Privacy Wizard to help users grant privileges to their friends. the goal of this tool is to automatically configure a users privacy settings with minimal effort and interaction from the user. The wizard asks users to first assign privacy **labels to selected friends**, and then uses this as input to construct a classifier which classifies friends based on their profiles and automatically assign privacy labels to the unlabeled friends.

In a similar vein, some studies [**maximilien_privacyasaservice_2009**] propose a methodology for quantifying the risk posed by a users privacy settings. A risk score reveals to the user **how far his/her privacy settings are from those of other users**. It provides feedback regarding the state of his/her existing settings. However, it does not help the user refine his/her settings in order to achieve a more acceptable configuration.

Trust metrics can be classified to two main categories: global and local trust metrics. **Local trust metrics**, compute trust values that are dependent on the target user, Local trust metrics take into account the very personal and subjective views of the users, they predict different values of trust for every single user based on their own experience. **Global trust metrics (reputation)**, on the other hand, predict a global reputation value for each node.

3.2.1 Concept of trust

In trust networks users can ask to rate other users, this means that, a user can express her level of trust in another user she has interacted with, i.e. express a trust statement such as "Alice, trust Bob as 0.8 in [0,1]". The system can then aggregate all the trust statements in a single trust networks representing the relationships between users. Trust metrics are algorithms whose goal is to predict, based on the trust network, the trustworthiness of "unknown" users, i.e. users in which a certain user didnt express a trust statement. Their aim is to reduce social complexity by suggesting how much an unknown user is trustworthy. Due to the increased use of OSNs, there is a growing number of studies that focus on using social network data for scoring messages in order to filter unwanted messages in messaging systems. The difference between each study has to do with the way the concept of trust is represented, computed and used.

The concept of trust is used to indicate the relationship between two entities. Trust in an entity is a commitment to an action based on a belief that the future actions of that entity will lead to a good outcome. There are three main properties of trust that are relevant to the development of algorithms for computing it [**wang_trustinvolved_2010**], namely, transitivity, asymmetry, and personalization. The primary property of trust that is used in our work is transitivity. if Alice highly trusts Bob, and Bob highly trusts Chuck, it does not always and exactly follow that Alice will highly trust Chuck. It is also important to note the asymmetry of trust, for two people involved in a relationship, trust is not necessarily identical in both directions. The third property of trust that is important in social networks is the personalization of trust, trust is inherently a personal opinion, two people often have very different opinions about the trustworthiness of the same person.

While much work has focused on tools for understanding and adjusting existing privacy settings, **Protect_U** [**gandouz_protect_20** uses machine learning techniques to recommend privacy settings based on a users personal data and trustworthy friends. Protect_U analyzes user profile contents and ranks them according to four risk levels: Low Risk, Medium Risk, Risky and Critical. The system then suggests personalized recommendations to allow users to make their accounts safer. In order to achieve this, it draws upon two protection models: local and community-based. The first model uses the **users personal data** in order to suggest recommendations, The second model seeks the **users trustworthy friends** to encourage them to help improve the safety of their counter parts account.

Despite the mole of work on social trust, Social Market is the first system to propose the use of **trust relationships** to build a decentralized interest-based marketplace.

Similarly, TAPE [**yongbozeng_study_2015**] is the first attempt to combine explicit and implicit social networks into a single gossip protocol. Zeng et al. [**yongbozeng_study_2015**] approaches the privacy quantification problem from a different angle. First, they consider **how likely a friend reveals others personal information**, by computing the privacy trust score, which is a widely studied research problem [**gundecha_exploiting_2011**]. Furthermore, the proposed work is related to **information diffusion in OSNs** such as [**fang_privacy_2010**]. Finally, TAPE framework differs from other work, in considering information diffusion in the context of privacy protection, which requires different sets of features and considerations.

Ostra [**mislove_ostra_2008**] utilizes trust relationship to thwart unwanted communication, where the number of a users trust relationships is used to limit the amount of unwanted communications he can produce. Ostra utilizes the existing trust relationship among users to charge the senders of unwanted messages and thus block spam. It relies on existing trust networks to connect senders and receivers via **chains of pair-wise trust relationship**, they use a pairwise link-based credit scheme to impose a cost on originator of unwanted communication. Unfortunatly, the scalability of this system stays uncertain as it employs a per-link credit scheme.

Gundecha et al. [**gundecha_exploiting_2011**] propose a feasible approach to the problem of identifying a users vulnerable friends on a social networking site. Vulnerability is somewhat contagious in this context. Their work differs from existing work addressing social networking privacy by introducing a **vulnerability-centered approach to a user** security on a social networking site. On most social networking sites, privacy related efforts have been concentrated on protecting individual attributes only. However, users are often vulnerable through community attributes. Unfriending vulnerable friends can help protect users against the security risks.

In [**zeng_trustaware_2014**], Sun et al proposed a **probability trust model** that uses Beta function to address concatenation propagation and multi-path propagation of trust.

SOAP [**li_soap_2011**] presents a social network based personalized spam filter that integrates **social closeness**, **user (dis)interest** and adaptive **trust** management into a Bayesian filter. SOAP proposed an email scoring mechanism based on an email network augmented with reputation ratings. An email is considered spam if the **reputation score of the email sender** is very low. Different from these social network based methods, SOAP focuses on personal interests in conjunction with social relationship closeness for spam detection. However, several issues with SOAP, including the intrinsic cost of initialization and continuous adaptation of social closeness (between sender and recipient), and social interests (of an individual) in the Bayesian filter, limit its usage.

Relationship between **users trustworthiness** and privacy risk is presented in [**pandey_computing_2015**].

Hameed [**hameed_lens_2011**] proposed LENS, which extends the FoF network by adding trusted users from outside of the FoF networks to mitigate spam beyond social circles. Only emails to a recipient that have been **vouched by the trusted nodes** can be sent into the network. The authors proposed using social networks and trust and reputation systems to combat spam. In contrast, LENS can reject unwanted email traffic during the SMTP time.

SocialEmail [**tran_social_2010**] considers the trust as an integral part of networking rather than working alongside of an existing communication system. SocialEmail leverages **social network trust paths** to rate the messages. The key feature of SocialEmail is that instead of directly connecting the sender and the recipient, messages are routed through existing friendship links. This gives each email recipient control over who can message him/her, In contrast, such social interaction-based methods are not sufficiently effective in dealing with legitimate emails from senders outside of the social network of the receiver.

Social interactions (e.g., **the exchange of messages between users**) have been suggested as an indicator of interpersonal tie strength [**xiang_modeling_2010**]. As a consequence, an unsupervised model has been developed to estimate the **relationship strength** from the interaction activity and the user similarity in the OSN [**xiang_modeling_2010**]. Although interaction-based methods leverage **social relationships** for extracting trust, the applications are not designed to be automated in the sense that the user must explicitly score other users, score messages, create whitelists or adjust the credits.

Fong [**fong_relationshipbased_2011**] formulated this paradigm called a Relationship-Based Access Control (ReBAC) model, it bases authorization decisions on the **relationships between the resource owner and the resource accessor** in an OSN. However, most of these existing work could not model and analyze access control requirements with respect to collaborative authorization management of shared data in OSNs.

3.2.2 Concept of reputation

Trust and reputation concepts are used in order to preserve users privacy while increasing their social capital in OSNs. Reputation concept is used to refer to a more general sense of trust towards a particular entity based on opinions of multiple entities.

A reputation system collects, distributes and aggregates feedback about participants past behavior. Such systems help people decide whom to trust, encourage trustworthy behavior and deter participation by those who are unskilled or dishonest. Various applications use real-time reputation-based systems, including online markets and anti-spam solutions. Anti-spam reputation systems generate a score, or rating, for each incoming message or IP, based on analysis of various parameters: message volume, type of traffic (e.g. sporadic vs continuous), rate of user complaint reports, feedback from spam traps, compliance with regulations, etc. This aggregated information, collected over time, forms the reputation of the sender.

SNARE [**hao_detecting_2009**] infers the reputation of a message sender based on **network-level features**, (e.g. **geodesic distance between sender and recipient, number of recipients**). The most influential feature in the system was the AS number of the sender. Using an automated reputation engine, SNARE classifies message senders as spammers or legitimate with about a 70% detection rate for less than a 0.3% false positive rate, without looking at

the contents of a message. However, lacking authentication and non-repudiation in standard trust and reputation solution make these solutions be subject to identity spoofing, false accusation and collusion attacks. Further, these solutions consume extra valuable resources of messaging servers on message reception and filtering.

In TrustMail, which is a prototype E-mail client, an approach is proposed that makes use of OSN reputation ratings to attribute different scores to E-mails [**golbeck_reputation_2004**]. The actual benefit of this system is that, by using **social network data**, it identifies potentially important and relevant messages even if the recipient does not know the sender [**golbeck_reputation_2004**].

Qian et al. [**qian_networklevel_2010**] addressed this issue by presenting a clustering technique that refines **AS-based and BGP prefix-based clusters**. The authors combined **BGP and DNS information** to identify a cluster of IP addresses within the same administrative boundary, and thus constructed the reputation for an entire cluster. This cluster-based reputation system allowed more accurate identification of the reputation of previously unknown IP addresses, and reduced the false negative rate by 50 percent compared to blacklists, without increasing the false-positive rate [**qian_networklevel_2010**].

Paradesi et al. [**paradesi_integrating_2009**] adopted a multi-agent based reputation model to define **trustworthiness of services**. Moreover, they developed a trust framework to derive trust for a composite service from trust model of component services.

3.2.3 Collaborative management

The trust value assigned to a person in previous work is estimated on the basis of his/her reputation, which can be assessed taking into account the person behaviour. Indeed, it is a matter of fact that people assign to a person with unfair behaviour a bad reputation and, as a consequence, a low level of trust. A possible solution is to estimate the trust level to be assigned to a user in a collaborative community on the basis of his/her reputation, given by his/her behavior with regards to all the other users in the community.

Hu et al. [**hu_detecting_2011**] propose an approach to enable **collaborative privacy management** of shared data in OSNs. In particular, they provide a systematic mechanism to identify and resolve privacy conflicts for collaborative data sharing. their conflict resolution indicates a trade-off between privacy protection and data sharing by quantifying privacy risk and sharing loss

Dealing with **collaborative information sharing**, Hu et al. [**biczok_interdependent_2013**] proposed a method to detect and resolve privacy conflicts.

The collaborative systems, called **COAT** [**ahmad_coat_2012**], do not rely upon semantic analysis but on the community to identify spam messages. Once a message is tagged as spam by one SMTP server, the signature of that message is transmitted to all other SMTP servers. This class requires the collaboration of multiple SMTP servers to implement the system.

SocialFilter [**yang_socialfilter_2009**] proposes a collaborative spam mitigation system that uses social trust embedded in OSN to asses the trustworthiness of Spam reporter. The spammer reports from the SocialFilter nodes are stored at a centralized repository that computes the trust values of the reports and identifies spammers based on IP addresses. However, the SocialFilters effectiveness is doubtful as spammers may use dynamic IPs.

Squicciarini et al. [**hu_detecting_2011**] proposed a solution for collective privacy management for photo sharing in OSNs. This work considered the privacy control of a content that is co-owned by multiple users in an OSN, such that each co-owner may separately specify her/his own privacy preference for the shared content. The Clarke-Tax mechanism was adopted to enable the collective enforcement for shared content. Game theory was applied to evaluate the scheme. However, a general drawback of this solution is the usability issue, as it could be very hard for ordinary OSN users to comprehend the Clarke-Tax mechanism and specify appropriate bid values for auctions. In addition, the auction process adopted in their approach indicates only the winning bids could determine who was able to access the data, instead of accommodating all stakeholders privacy preferences.

3.3 Technical

Nowadays, the Web converged over two main protocols, namely HTTP/HTTPS and SMTP/SMTPS Beside these, DNS still has a central role for reaching almost any Web server. In this section we focus on SMTP privacy, the next section relies on HTTP privacy.

Many anti-spam techniques have been proposed and deployed to counter email Spam from different perspectives. Based on the placement of anti-Spam mechanisms, these techniques can be divided into two main categories: recipient-oriented and sender-oriented.

3.3.1 Recipient-oriented Techniques

This class of techniques either (1) block/delay email Spam from reaching the recipients mailbox or (2) remove/mark Spam in the recipients mailbox. Due to the flourish of techniques in this category, we further divide them into content-based and non-content-based sub-categories.

Content-based Techniques.

The techniques in this sub-category detect and filter spam by analyzing the content of received messages, including both message header and message body.

Email address filters: Email address filters are simply whitelists or blacklists. Whitelists consist of all acceptable email addresses and blacklists are the opposite. Blacklists can be easily broken when spammers forge new email addresses, but using whitelists alone makes the world enclosed.

Heuristic filters: The features that are rare in normal messages but appear frequently in spam, such as nonexisting domain names and spam-related keywords, can be used to distinguish spam from normal email.

Machine learning based filters: Since spam detection can be converted into the problem of text classification, many content-based filters utilize machine-learning algorithms for filtering spam. As these filters can adapt their classification engines with the change of message content, they outperform heuristic filters.

Non-content-based Techniques.

The techniques in this sub-category use non-content spam characteristics, such as source IP address, message sending rate, and violation of SMTP standards, to detect email spam.

DNSBLs: DNSBLs are distributed blacklists, which record IP addresses of spam sources and are accessed via DNS queries. When an SMTP connection is being established, the receiving MTA (Mail Transfer Agent) can verify the sending machines IP address by querying the subscribed DNSBL. Mail server records the number and frequency of the same email sent to multiple destinations from specific IP addresses. If the number and frequency exceed thresholds, the node with the specific IP address is blocked. Even DNSBLs have been widely used, their effectiveness and responsiveness [**jung_empirical_2004**, **ramachandran_can_2006**] are still under study.

MARID: MARID (MTA Authorization Records In DNS) is a class of techniques to counter forged email addresses by enforcing sender authentication. MARID is also based on DNS and can be seen as a distributed whitelist of authorized MTAs. Multiple MARID drafts have been proposed, some of them (SPF and DKIM) are deployed in real world [**spf:_2018**, **BibEntry2014Dec**]. PGP and S/MIME are also

Challenge-Response (CR): CR is used to keep the merit of whitelist without losing important messages. To add a sender email address in the whitelist, senders are requested a challenge that needs to be solved by a human being. After a proper response is received, the senders address can be added into the whitelist.

Cryptographic: Pretty Good Privacy (PGP) [**pgpaghiles2007**] and S/MIME are both cryptographic approaches that sign the message body using public-key cryptography and append the signature in the body. In PGP, Keys are stored in end-user key rings or in public key-servers. Key management uses a peer-to-peer web of trust architecture. Whereas in S/MIME, management follows a hierarchical model similar to SSL and keys are signed by a certificate authority.

Delaying: As a variation of rate limiting, delaying is triggered by an unusually high sending rate. Most delaying mechanisms are applied at receiving MTAs.

3.3.2 Sender-oriented techniques

To effectively deny spam at the source, ISPs and ESPs (Email Service Providers) have taken various measures to manage the usage of email services. For example, message submission protocol [**BibEntry1998Dec**] has been proposed to replace SMTP, when a message is submitted from an MUA (Mail User Agent) to its MTA.

The proposed work in [**ahmad_coat_2012**] differs from the other techniques in a way that all of them categorize mail messages at receiver side, whereas COAT works at the sender side and reduces outgoing spam rather than inbox spam.

Cost-based approaches: Borrowing the idea of postage from regular mail systems, many cost-based techniques attempt to shift the cost of thwarting spam from receiver side to sender side.

All these techniques assume that the average email cost for a normal user is trivial and negligible, but the accumulative charge for a spammer will be high enough to drive them out of business.

Cost concept may have different forms in different proposals. Bonded Sender [**BibEntry2018May**] advocates associating email with real money.

3.4 Technical HTTP

In this section we enumerate existing privacy protection measures available to users and one new protection proposal [**mayer-do-not-track-00**].

Blocking requests to targeted third parties: This block measure includes using an advertisement blocking tool (AdBlock Plus [**adblockplus**]) to syntactically block selected third parties via server/domain name. Another measure blockhidden [**krishnamurthy_privacy_2009**] determines the true source of hidden third-parties by examining their authoritative DNS servers.

Refusing cookies to prevent tracking: Browsers can be set to refuse all cookies (nocook) or just third-party cookies (no3rdcook).

Disabling script execution: JavaScript execution can be disabled (nojs) either permanently via the browser or selectively via a tool such as NoScript [**NoScript**].

Filtering protocol headers: This is done via extensions or at an intermediary and includes the referrer measure available in some browsers to modify or remove the Referer header in an HTTP request.

Anonymizing the user and user actions: One such anon measure is anonymizing users IP address via an anonymizing proxy or by using Tor.

Do-Not-Track HTTP header proposal: Researchers proposed, in early 2010, that browsers add a HTTP DoNot-Track-Header (DNT-Header) [**mayer-do-not-track-00**] to allow users to express their interest in not being tracked by any aggregator or ad network. However, the extent to which third parties would honor such a header is unknown.

4. CONCLUSIONS

TODO

APPENDIX

Year		Factors	Computation Model	Results interpretation
2018	[**alemany_estimation_2018**]	-Closeness Centrality -Degree Centrality	Estimation	**Closeness** have a high degr correlation with **privacy sc**
2017	[**zhang_privacypreserving_2017**]	-Users Credibility	Privacy-preserving	Correlation between identify and **user's credibility** is s
2016	[**li_algorithm_2016**]	-Influence probability -Degree Centrality -Number of seeds neighbors -Number of experts neighbors	Utility degree and Utility privacy cost ratio discount algorithms	Methods evaluations: -Number of seeds activated -Number of expert activated
2015	[**pandey_computing_2015**]	-Attitude Information -Popularity (Page rank) -Privacy Settings: visibility	Users trustworthiness	Relationship between **users** and **privacy risk**
2015	[**yongbozeng_study_2015**]	-Node information diffusion -Link information diffusion -Undesirable Destination	Birnbaum's measure (BM)	**Users trustworthiness**
2015	[**b.s._privacy_2015**]	-Tie-strength -Communication information -Number of mutual friends	Bezier curve	Sorting user's **Friends Priv**
2015	[**vidyalakshmi_privacy_2015**]	-Communication frequency -Privacy setup	Cubic bezier curve	Estimating friends privacy s **users dispositions** to priv
2014	[**caliskanislam_privacy_2014**]		3-class supervised learning	Correlation between **Users** **Friends Privacy** Score
2014	[**zeng_trustaware_2014**]	-Undesirable Destination -Closeness centrality -Diffusion centrality -User's behavior	Probability trust model	-The **information diffusio** depends on the **users beha** -The **information diffusio** depends on the **closeness** b
2013	[**nepali_sonet_2013**]	-Sensitivity, Visibility	Linear model	**Users privacy** exposure
2013	[**biczok_interdependent_2013**]	-User valuation on app -Network valuation on app	Collaborative Interdependent Privacy Game, Nash equilibrium Sub-optimal Equilibrium	Show how network and pers the behavior of social netwo to app usage
2012	[**akcora_risks_2012**]	-Attitude similarity	Baseline estimation Learning Friend Impacts	Risks of friendships can be users attitude towards frien
2012	[**ahmad_coat_2012**]	-Sender address -Sender password -Message body	Collaborative outgoing anti-spam technique	Once a message is tagged as the signature of that messag other SMTP server
2011	[**gundecha_exploiting_2011**]	-Individual index -Community index	Unfriending vulnerable friends	Correlation between unfrien security improvement is sig
2011	[**liu_framework_2010**] [**maximilien_privacyasaservice_2009**]	-User behavior -Sensitivity and visibility	Item response theory	Behavioral, Quantitative an privacy evaluation
2011	[**hu_detecting_2011**]	-Number of privacy conflicts -Trust of an accessor -Sensitivity and Visibility	Collaborative privacy	Quantify **multiparty priva** Evaluate multiparty privacy a resolving score
2010	[**xiang_modeling_2010**]	-Interaction level -User similarity	Unsupervised linear model Link-based latent variable model	The estimated link weights and lead to improved classif
2010	[**tran_social_2010**]	-Trust path	Trust path probability	Message paths trustworthin
2010	[**talukder_privometer_2010**]			
2010	[**fang_privacy_2010**]	-Community membership -Online activity	Privacy-Preference Model as a classifier	**Community structure** of resource when modeling the
2010	[**qian_networklevel_2010**]		Cluster-based reputation	
2009	[**yang_socialfilter_2009**]	-Nodes IP -Reporter Trust -Identity uniqueness	Distributed SocialFilter- repository	Node reputation
2009	[**hao_detecting_2009**]	-Geodesic distance -Number of recipients	Classifier: Supervised Learning	Sender reputation system th email senders based on seve
2009	[**baden_persona_2009**]		Attribute-based encryption	
2009	[**paradesi_integrating_2009**]		Multi-agent reputation	
2008	[**mislove_ostra_2008**]			
2004	[**lixiong_peertrust_2004**]	-Community settings -Number of malicious peers	Trust computation	

Table 1: Social metrics